

Bakalářská práce



České  
vysoké  
učení technické  
v Praze

**F3**

Fakulta elektrotechnická  
Katedra počítačů

## Aplikace pro analýzu veřejně dostupných dat o studentech vysokých škol

**Vojtěch Hořejší**

Vedoucí: Ing. Pavel Náplava, Ph.D.  
Obor: Softwarové inženýrství a technologie  
Studijní program: Business informatics  
Květen 2024



## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Hořejší** Jméno: **Vojtěch** Osobní číslo: **485156**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra počítačů**  
Studijní program: **Softwarové inženýrství a technologie**  
Specializace: **Business informatics**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Aplikace pro analýzu veřejně dostupných dat o studentech vysokých škol**

Název bakalářské práce anglicky:

**Application for analysing publicly available data on university students**

Pokyny pro vypracování:

Navrhněte a vytvořte aplikaci, která umožní provádět „různé“ datové analýzy nad běžně dostupnými daty o studentech a studiiích na vysokých školách. Postupujte následujícím způsobem:

- 1) Vyhledejte volně dostupné datové sady, poskytující informace o studentech vysokých škol, vysokých školách atd.
- 2) Proveďte základní analýzu těchto sad a v případě potřeby proveďte úpravu těchto dat do odpovídající podoby.
- 3) Navrhněte ukázkové hypotézy/otázky, které umožní ověřit kvalitu dat s ohledem na možnosti jejich analýzy.
- 4) Pomocí existujících nástrojů hypotézy/otázky potvrďte/vyvráťte.
- 5) Navrhněte a vytvořte vlastní aplikaci, která umožní uživatelům provádět vlastní analýzy datových sad jak ve vizuální, tak i nevizuální podobě, s možností importu a exportu různých datových sad. Rozsah aplikace koordinujte s vedoucím práce.
- 6) Aplikaci proveďte uživatelsky pomocí dříve stanovených hypotéz, na příkladech demonstруйте další možnosti využití aplikace.

Seznam doporučené literatury:

Navara, M., 2007. Pravděpodobnost a matematická statistika, Praha: ČVUT.  
VanderPlas, J., 2016. Python Data Science Handbook, O'Reilly Media. Dostupné na:  
<https://jakevdp.github.io/PythonDataScienceHandbook/>  
Montgomery, DC 2012, Introduction to Linear Regression Analysis, John Wiley & Sons, Incorporated, Newark. Dostupné na: ProQuest Ebook Central.  
Han, J, Kamber, M, & Pei, J 2011, Data Mining: Concepts and Techniques, Elsevier Science & Technology, San Diego. Dostupné na: ProQuest Ebook Central

Jméno a pracoviště vedoucí(ho) bakalářské práce:

**Ing. Pavel Náplava, Ph.D. Centrum znalostního managementu FEL**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **07.02.2024**

Termín odevzdání bakalářské práce: **24.05.2024**

Platnost zadání bakalářské práce: **21.09.2025**

Ing. Pavel Náplava, Ph.D.  
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.  
podpis děkana(ky)

### III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studenta



## Poděkování

Chtěl bych poděkovat panu Ing. Pavlu Náplavovi, Ph.D., za jeho cenné rady, ochotné vedení a odborné poznatky, které mi byly během práce nesmírně užitečné. Dále chci poděkovat své rodině a přátelům za jejich podporu.

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškerou použitou literaturu.

V Praze, 24. května 2024

## Abstrakt

Tato bakalářská práce se zabývá návrhem a vývojem aplikace pro analýzu dat o studentech vysokých škol, která jsou dostupná z Ministerstva školství ČR. Cílem je umožnit uživatelům provádět statistické analýzy a vizualizace, což přispívá k lepšímu porozumění a využití těchto dat. Práce zahrnuje shromažďování a předzpracování dat, ověřování hypotéz k jejich kvalitě a vývoj interaktivního softwaru. Výsledná aplikace obsahuje několik scénářů, díky kterým si uživatelé mohou osvojit základy datové analýzy.

**Klíčová slova:** datová analýza, vizualizace dat, statistická analýza, veřejně dostupná data, vzdělávací data, aplikace pro analýzu dat, naučná aplikace, python aplikace

**Vedoucí:** Ing. Pavel Náplava, Ph.D.

## Abstract

This bachelor's thesis focuses on designing and developing an application for analyzing data on university students available from the Czech Ministry of Education. The aim is to enable users to perform statistical analyses and visualizations, contributing to a better understanding and utilization of these data. The work includes the collection and preprocessing of data, verification of hypotheses regarding their quality, and the development of interactive software. The resulting application contains several scenarios through which users can learn the fundamentals of data analysis.

**Keywords:** data analysis, data visualization, statistical analysis, publicly available data, educational data, data analysis application, educational application, python application

**Title translation:** Application for analysing publicly available data on university students

# Obsah

<b>1 Úvod</b>	<b>1</b>		
1.1 Úvod do tématu	1		
1.2 Motivace práce	1		
1.3 Cíle práce	2		
1.4 Použité technologie	2		
1.5 Struktura práce	2		
<b>2 Teoretický základ</b>	<b>5</b>		
2.1 Rešerše dostupných zdrojů, nástrojů a metod pro datovou analýzu	5		
2.1.1 Definice cílů rešerše	5		
2.1.2 Existující aplikace pro datovou analýzu	6		
2.1.3 Literatura a související práce	9		
2.1.4 Prozkoumání dostupných návodů a tutoriálů pro datovou analýzu	11		
2.1.5 Zdroje dat	12		
2.1.6 Role umělé inteligence v analýze dat	13		
2.1.7 Vyhodnocení rešerše	14		
2.2 Metody a techniky datové analýzy	15		
2.2.1 Metody pro sběr dat	15		
2.2.2 Metody pro analýzu dat	17		
<b>3 Sběr a zpracování dat</b>	<b>27</b>		
3.1 Definování klíčových pojmů	27		
3.2 Sběr dat	28		
3.3 Průzkum struktury a obsahu tabulek	29		
3.3.1 Ministerstvo školství, mládeže a tělovýchovy (MŠMT)	29		
3.3.2 Statistický úřad	33		
3.4 Zpracování dat	34		
3.4.1 F11 – Souhrn studentů a absolventů podle formy a typu studijního programu	34		
3.4.2 F12 – Souhrn škol, fakult, studentů a absolventů	36		
3.4.3 F13 – Studenti podle měst	37		
3.4.4 Pohyb obyvatel České republiky v letech 1785 až 2022	38		
3.5 Shrnutí kapitoly 3	40		
<b>4 Analýza dat a formulace otázek a hypotéz</b>	<b>41</b>		
4.1 Přístupy k formulaci hypotéz	41		
4.1.1 Náš přístup	41		
4.2 Formulace otázek a hypotéz	42		
4.3 Struktura kapitol s hypotézami	42		
4.4 Hypotéza 1: Trendy ve vývoji počtu studentů v čase	43		
4.4.1 Úvod do problematiky	43		
4.4.2 Analýza dat	43		
4.4.3 Interpretace výsledků	48		
4.4.4 Shrnutí a závěr hypotézy 1	48		
4.5 Hypotéza 2: Vztah mezi národností studenta a typem studia	49		
4.5.1 Úvod do problematiky	49		
4.5.2 Analýza dat	49		
4.5.3 Interpretace výsledků	53		
4.5.4 Shrnutí a závěr hypotézy 2	54		
4.6 Hypotéza 3: Geografické rozložení studentů	54		
4.6.1 Úvod do problematiky	54		
4.6.2 Analýza dat	54		
4.6.3 Interpretace výsledků	58		
4.6.4 Shrnutí a závěr hypotézy 3	59		
4.7 Hypotéza 4: Geografické rozložení zahraničních studentů v České republice	59		
4.7.1 Úvod do problematiky	59		
4.7.2 Analýza dat	59		
4.7.3 Interpretace výsledků	62		
4.7.4 Shrnutí a závěr hypotézy 4	64		
4.8 Hypotéza 5: Rozdíl mezi soukromými a veřejnými školami	64		
4.8.1 Úvod do problematiky	64		
4.8.2 Analýza dat	64		
4.8.3 Interpretace výsledků	67		
4.8.4 Shrnutí a závěr hypotézy 5	67		
4.9 Požadavky na náš nástroj	68		
4.10 Závěr kapitoly	69		
<b>5 Analýza požadavků a návrh aplikace</b>	<b>71</b>		
5.1 Úvod do návrhu systému	71		
5.2 Přejít k požadavkům	71		
5.2.1 Funkční požadavky	72		
5.2.2 Nefunkční požadavky	72		
5.2.3 Use case diagram	72		
5.3 Architektonický návrh	74		
5.3.1 Diagram tříd	74		
5.3.2 Wireframes	76		

5.4 Volba technologií . . . . .	78	8.5.4 Analýza sportovního výkonu	117
5.4.1 Volba programovacího jazyka	79	8.5.5 Analýza zdravotních údajů .	117
5.4.2 Volba GUI knihoven . . . . .	80	8.6 Uzavření kapitoly . . . . .	118
5.4.3 Volba dalších knihoven . . . . .	81	<b>9 Závěr</b>	<b>119</b>
<b>6 Vývoj aplikace</b>	<b>83</b>	9.1 Osobní reflexe . . . . .	121
6.1 Použité technologie . . . . .	83	<b>A Bibliografie</b>	<b>123</b>
6.2 Architektura aplikace . . . . .	84	<b>B Aplikace a repozitáře</b>	<b>127</b>
6.2.1 Modifikace MVC architektury	84	B.1 Odkazy na repozitáře . . . . .	127
6.3 Realizace komponent . . . . .	84	B.2 Excel tabulky pro analýzu . . . .	127
6.3.1 DataScienceApp . . . . .	84	B.3 Spustitelný soubor . . . . .	127
6.3.2 GraphingTools . . . . .	85		
6.3.3 AdvancedAnalytics . . . . .	87		
6.3.4 TutorialManager . . . . .	87		
6.3.5 Další Komponenty Aplikace .	88		
6.4 Shrnutí implementace . . . . .	89		
<b>7 Testování a validace</b>	<b>91</b>		
7.1 Testovací strategie . . . . .	91		
7.1.1 Průběh testování . . . . .	92		
7.2 Výsledky testování . . . . .	92		
7.3 Validace hypotéz pomocí vlastní			
aplikace . . . . .	93		
7.3.1 Hypotéza 1: Trendy ve vývoji			
počtu studentů v čase . . . . .	94		
7.3.2 Hypotéza 2: Vztah mezi			
národností studenta a typem			
studia . . . . .	97		
7.3.3 Hypotéza 3: Geografické			
rozložení studentů . . . . .	101		
7.3.4 Hypotéza 4: Geografické			
rozložení zahraničních studentů v			
České republice . . . . .	105		
7.3.5 Hypotéza 5: Rozdíl mezi			
soukromými a veřejnými školami	107		
7.3.6 Vyhodnocení rozdílů . . . . .	109		
<b>8 Diskuse</b>	<b>111</b>		
8.1 Hodnocení výsledků . . . . .	111		
8.2 Omezení a problémy . . . . .	111		
8.3 Návrhy na zlepšení . . . . .	112		
8.4 Možnosti využití aplikace . . . . .	113		
8.5 Další možnosti využití aplikace	113		
8.5.1 Analýza prodeje pro malé			
podnikatele . . . . .	113		
8.5.2 Osobní analýza výkonnosti při			
učení . . . . .	115		
8.5.3 Plánování dovolené . . . . .	117		

## Obrázky

2.1 Počet pracovních pozic v oblasti data science podle nástrojů. Zdroj: DataForest[3] . . . . .	7
2.2 Udělené tituly z MIT od července 2021 do června 2022. (Zdroj MIT)[10] . . . . .	10
2.3 Příklad spojnicového grafu . . . . .	17
2.4 Příklad sloupcového grafu . . . . .	18
2.5 Příklad bodového grafu . . . . .	19
2.6 Příklad kruhového grafu . . . . .	19
2.7 Příklad histogramu . . . . .	20
2.8 Příklad krabicového grafu . . . . .	21
2.9 Příklad geografického grafu . . . . .	21
2.10 Příklad základních statistických měř . . . . .	22
2.11 Příklad korelační matice . . . . .	24
2.12 Příklad lineární regrese . . . . .	25
3.1 Souhrn studentů a absolventů podle formy a typu studijního programu. . . . .	30
3.2 Souhrn škol, fakult, studentů a absolventů. . . . .	31
3.3 Studenti podle měst. . . . .	32
3.4 Pohyb obyvatel České republiky v letech 1785 až 2022. . . . .	33
3.5 Tabulka se souhrnnými statistikami o studentech. . . . .	36
3.6 Tabulka se souhrnnými statistikami o univerzitách. . . . .	37
3.7 Tabulka se souhrnnými statistikami o studentech v jednotlivých městech. . . . .	38
3.8 Tabulka s odhadem počtu možných studentů v jednotlivých letech. . . . .	40
4.1 Vývoj počtu studentů vysokých škol v letech 2001-2022 . . . . .	44
4.2 Počet narozených dětí v daných letech . . . . .	44
4.3 Celkový počet studentů vs. potenciální noví studenti . . . . .	45
4.4 Potenciální noví studenti vs. nově zapsaní studenti . . . . .	46
4.5 Regresní analýza počtu studentů vysokých škol . . . . .	47
4.6 Regresní analýza potenciálních nových studentů . . . . .	47
4.7 Počet českých studentů ve formě prezenčního a kombinovaného studia 50	
4.8 Počet zahraničních studentů ve formě prezenčního a kombinovaného studia . . . . .	50
4.9 Počet českých studentů ve formě prezenčního studia podle stupně studia . . . . .	51
4.10 Počet českých studentů ve formě kombinovaného studia podle stupně studia . . . . .	51
4.11 Počet zahraničních studentů ve formě prezenčního studia podle stupně studia . . . . .	52
4.12 Počet zahraničních studentů ve formě kombinovaného studia podle stupně studia . . . . .	52
4.13 Počet studentů ve formě prezenčního a kombinovaného studia (čeští a zahraniční studenti dohromady) . . . . .	53
4.14 Geografické rozložení studentů v roce 2001 . . . . .	55
4.15 Počet studentů v jednotlivých městech v roce 2001 . . . . .	55
4.16 Počet studentů v prvních deseti městech v roce 2001 . . . . .	56
4.17 Geografické rozložení studentů v roce 2022 . . . . .	56
4.18 Počet studentů v jednotlivých městech v roce 2022 . . . . .	57
4.19 Počet studentů v prvních deseti městech v roce 2022 . . . . .	58
4.20 Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2001 60	
4.21 Geografické rozložení zahraničních studentů v roce 2001 . . . . .	60
4.22 Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2022 61	
4.23 Geografické rozložení zahraničních studentů v roce 2022 . . . . .	61
4.24 Vývoj počtu zahraničních studentů v letech 2001 až 2022 . . . . .	62

4.25 Důvody pro studium v Česku (Zdroj: DZS - Zpráva o zahraničních studentech na českých vysokých školách 2023, 2023)[30] . . . . .	63	7.10 Počet zahraničních studentů ve formě prezenčního a kombinovaného studia . . . . .	98
4.26 Data o počtu studentů na veřejných vysokých školách . . . . .	65	7.11 Počet českých studentů ve formě prezenčního studia podle stupně studia . . . . .	98
4.27 Počet studentů na veřejných vysokých školách . . . . .	65	7.12 Počet českých studentů ve formě kombinovaného studia podle stupně studia . . . . .	99
4.28 Data o počtu studentů na soukromých vysokých školách . . . . .	65	7.13 Počet zahraničních studentů ve formě prezenčního studia podle stupně studia . . . . .	99
4.29 Počet studentů na soukromých vysokých školách . . . . .	66	7.14 Počet zahraničních studentů ve formě kombinovaného studia podle stupně studia . . . . .	100
4.30 Srovnání počtu studentů na veřejných a soukromých školách . . . . .	66	7.15 Počet studentů ve formě prezenčního a kombinovaného studia (čeští a zahraniční studenti dohromady) . . . . .	100
5.1 Use Case diagram poskytující přehled o funkcích systému z pohledu uživatelů a systému . . . . .	73	7.16 Geografické rozložení studentů v roce 2001 . . . . .	101
5.2 Diagram tříd aplikace . . . . .	75	7.17 Počet studentů v jednotlivých městech v roce 2001 . . . . .	102
5.3 Wireframe zobrazení tabulky a základních funkcí . . . . .	76	7.18 Počet studentů v prvních deseti městech v roce 2001 . . . . .	102
5.4 Wireframe vytváření grafu . . . . .	77	7.19 Geografické rozložení studentů v roce 2022 . . . . .	103
5.5 Wireframe zobrazení záložky pokročilé analýzy . . . . .	77	7.20 Počet studentů v jednotlivých městech v roce 2022 . . . . .	103
5.6 Wireframe zobrazení menu s návodů . . . . .	78	7.21 Počet studentů v prvních deseti městech v roce 2022 . . . . .	104
5.7 Wireframe zobrazení jednoho návodu nebo scénáře . . . . .	78	7.22 Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2001 . . . . .	105
7.1 Sloupcový graf před změnou . . . . .	93	7.23 Geografické rozložení zahraničních studentů v roce 2001 . . . . .	105
7.2 Sloupcový graf po změně . . . . .	93	7.24 Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2022 . . . . .	106
7.3 Vývoj počtu studentů vysokých škol v letech 2001-2022 . . . . .	94	7.25 Geografické rozložení zahraničních studentů v roce 2022 . . . . .	106
7.4 Počet narozených dětí v daných letech . . . . .	94	7.26 Vývoj počtu zahraničních studentů v letech 2001 až 2022 . . . . .	107
7.5 Celkový počet studentů vs. potenciální noví studenti . . . . .	95	7.27 Data o počtu studentů na veřejných vysokých školách . . . . .	107
7.6 Potenciální noví studenti vs. nově zapsaní studenti . . . . .	95	7.28 Počet studentů na veřejných vysokých školách . . . . .	108
7.7 Regresní analýza počtu studentů vysokých škol . . . . .	96		
7.8 Regresní analýza potenciálních nových studentů . . . . .	96		
7.9 Počet českých studentů ve formě prezenčního a kombinovaného studia . . . . .	97		

## Tabulky

7.29 Data o počtu studentů na soukromých vysokých školách . . .	108
7.30 Počet studentů na soukromých vysokých školách . . . . .	108
7.31 Srovnání počtu studentů na veřejných a soukromých školách .	109
8.1 Ukázková data o prodeji produktů . . . . .	113
8.2 Počet prodaných kusů jednotlivých produktů . . . . .	114
8.3 Podíl prodejů jednotlivých produktů na celkovém prodeji . . .	114
8.4 Základní statistiky prodejů . . .	115
8.5 Ukázková data o čase stráveném učením a výsledcích zkoušek . . . .	115
8.6 Základní statistiky studijních dat	116
8.7 Čas strávený učením v různých prostředích . . . . .	116
8.8 Lineární regrese mezi časem stráveným učením a výsledky zkoušky . . . . .	116





# Kapitola 1

## Úvod

### 1.1 Úvod do tématu

Datová analýza je zásadní ve všech aspektech moderního života, umožňuje nám lepší porozumění světu kolem nás. Každodenně se setkáváme s daty prezentovanými skrze grafy, statistiky a zprávy, které jsou aplikovány napříč různými obory. Tyto informace nám pomáhají formovat naše názory a rozhodování. V rámci této práce se zaměříme na analýzu dat specificky z prostředí vysokých škol, čerpaných z veřejně dostupných zdrojů Ministerstva školství, mládeže a tělovýchovy ČR. Vzdělání, a zejména vysokoškolské studium, může hrát klíčovou roli v kariérním růstu a je často diskutováno z hlediska průchodnosti a uplatnitelnosti na trhu práce.

Ačkoli nástroje jako Microsoft Excel poskytují základní možnosti pro analýzu dat, mohou být omezené nebo obtížně využitelné pro složitější analýzy. Na druhé straně pokročilé platformy jako Tableau<sup>1</sup> nebo PowerBI<sup>2</sup> mohou být pro nováčky příliš komplexní a náročné na pochopení.

### 1.2 Motivace práce

Motivací pro tuto práci bylo mé osobní zaujetí pro datovou analýzu, které se prohloubilo během studia, kde jsme právě měli předmět na základy datové analýzy. V předmětu jsme narazili na otázky týkající se průchodnosti a kvality vysokoškolského vzdělání, což jsou diskuse, které rezonují mezi studenty při volbě univerzity.[1] S tímto tématem se často setkávají i veřejné debaty, ve kterých jsou vysoké školy hodnoceny z hlediska náročnosti a uplatnitelnosti vzdělání na trhu práce.

Myšlenka z úvodu nás motivovala k vytvoření uživatelsky přívětivé aplikace, která začátečníkům poskytne jednoduchý nástroj na analýzu dat z prostředí vysokých škol. Tento projekt poté přerostl v myšlenku zahrnout také výuku základům datové analýzy, čímž aplikace poslouží i jako vzdělávací nástroj.

<sup>1</sup>Tableau je software pro vizualizaci dat, umožňující rychlé vytváření grafů a dashboardů s intuitivním ovládáním.

<sup>2</sup>PowerBI od Microsoftu slouží k vizualizaci dat, umožňuje sdílení poznatků a integraci s dalšími Microsoft produkty.

## 1.3 Cíle práce

Prvotním cílem této práce je navrhnout a vytvořit aplikaci pro různé datové analýzy nad běžně dostupnými daty o studentech a studiích na vysokých školách. Tento proces zahrnuje vyhledání volně dostupných datových sad, jejich analýzu a úpravu do lépe zpracovatelného formátu. Dále bude nezbytné navrhnout hypotézy, které ověří kvalitu dat a možnosti jejich analýzy. Následně navrhne a vytvoříme aplikaci pro práci s těmito daty a ověříme její funkčnost na základě navržených hypotéz.

Během vývoje se navíc začneme soustředit na to, aby aplikace byla využitelná i k výuce základům datové analýzy. Naše vytvořené hypotézy budou sloužit jako výukové scénáře, které uživatelům pomohou lépe pochopit základy analýzy dat.

## 1.4 Použité technologie

Bakalářská práce byla tvořena za pomoci několika moderních nástrojů a technologií. Použil jsem **OpenAI chatbota**<sup>3</sup>, který mi pomáhal přeformulovat texty, opravovat gramatiku a dělat rychlou úvodní analýzu textu. Tento nástroj mi značně usnadnil práci s textovými materiály a umožnil mi se soustředit na technické aspekty projektu [2].

Při programování aplikace jsem využíval **GitHub Copilot**<sup>4</sup>, který mi pomáhal doplňovat kód a implementovat metody. Díky GitHub Copilot se zvýšila moje produktivita a rychlost psaní kódu, protože mi doplňoval funkce. Tím jsem se mohl více soustředit na strukturu a obsah aplikace než na samotné psaní kódu. Tento nástroj také snížil počet chyb a překlepů v kódu, což přispělo k vyšší kvalitě výsledného produktu.

## 1.5 Struktura práce

Práce je rozdělena do následujících kapitol:

- **Úvod** – Definiuje problematiku, motivaci a cíle práce, použití technologií a strukturu práce.
- **Rešerše, metody a nástroje pro datovou analýzu** – Zaměřuje se na rešerši dostupných zdrojů, nástrojů a metod pro datovou analýzu. Obsahuje definici cílů rešerše, přehled existujících aplikací, relevantní literaturu, dostupné návody a tutoriály, zdroje dat a roli umělé inteligence v analýze dat. Dále popisuje metody a techniky pro sběr a analýzu dat.
- **Sběr a zpracování dat** – Popisuje metody a techniky sběru dat a jejich předzpracování, včetně detailního průzkumu struktury a obsahu tabulek.

<sup>3</sup><https://chat.openai.com/>

<sup>4</sup><https://github.com/features/copilot>

V této kapitole jsou vyhledány a analyzovány volně dostupné datové sady, poskytující informace o studentech vysokých škol.

- **Analýza dat a formulace otázek a hypotéz pomocí Power BI** – Navrhuje a popisuje ukázkové hypotézy a otázky, které umožní ověřit kvalitu dat a možnosti jejich analýzy. Obsahuje analýzu dat a interpretaci výsledků pro každou formulovanou hypotézu, přičemž využívá nástroj Power BI.
- **Analýza požadavků a návrh aplikace** – Detailní analýza uživatelských a systémových požadavků, návrh architektury aplikace, výběr technologií a vytvoření aplikace pro provádění analýz datových sad.
- **Vývoj aplikace** – Podrobný popis implementace aplikace, včetně realizace jednotlivých komponent, a popis architektury aplikace s důrazem na možnosti importu a exportu různých datových sad.
- **Testování a validace** – Popisuje metody testování aplikace, zahrnující uživatelské testy a validaci výsledků. Testování probíhá na základě dříve stanovených hypotéz, s důrazem na end-to-end a manuální testy. Validace zahrnuje porovnání výsledků naší aplikace s Power BI, což ověřuje její přesnost a spolehlivost.
- **Diskuse** – Diskuze o dosažených výsledcích, identifikovaných omezeních a návrhy na zlepšení. Ukázka dalších možností využití aplikace.
- **Závěr** – Shrnutí hlavních výsledků práce, její přínos a reflexe osobních zkušeností a získaných dovedností.
- **Přílohy** – Obsahuje doplňující materiály a zdroje, včetně technické dokumentace, spustitelného souboru aplikace a odkazu na repozitář se všemi využitými tabulkami, obrázky a zdrojovými kódy aplikace.

Struktura práce byla navržena tak, aby odpovídala požadavkům zadání a cílům práce.



## Kapitola 2

### Teoretický základ

Tato kapitola poskytuje teoretický rámec pro pochopení a analýzu dat v kontextu vysokého školství v České republice. Cílem je představit metody sběru a analýzy dat, které pomáhají lépe porozumět trendům ve vzdělávacím sektoru.

Nejprve se zaměřujeme na rešerši dostupných nástrojů a metod pro datovou analýzu, jako jsou R, Python a Power BI. Prozkoumáváme existující studie a literaturu, které se věnují analýze dat ve vzdělávání, a identifikujeme klíčové techniky používané v této oblasti.

Následně detailně popisujeme metody a techniky datové analýzy, včetně deskriptivní statistiky, korelace, lineární regrese a vizualizačních technik. Tyto metody umožňují efektivní zpracování a interpretaci dat a poskytují podklad pro rozhodování a tvorbu politik ve vzdělávání.

Závěr kapitoly se věnuje diskusi o možnostech dalšího výzkumu a potenciálních aplikacích zjištěných poznatků pro zlepšení vzdělávacího sektoru.

Kapitola poskytuje ucelený přehled o metodách a přístupech k datové analýze relevantních pro oblast vysokého školství, podporujících vědecký výzkum a praktické aplikace.

## 2.1 Rešerše dostupných zdrojů, nástrojů a metod pro datovou analýzu

### 2.1.1 Definice cílů rešerše

Cílem této rešerše je identifikovat a zhodnotit dostupné zdroje, nástroje a metody pro datovou analýzu v kontextu vysokého školství v České republice. Zaměříme se na veřejně dostupná data a prozkoumáme, jaké nástroje a metodiky jsou využívány pro analýzu těchto dat, s důrazem na zjišťování trendů v úspěšnosti studia a zaměstnatelnosti absolventů. Rešerše se zaměřuje na následující klíčové aspekty:

1. **Shromáždění dat:** Lokalizace a agregace veřejně dostupných datových souborů o studentech.

2. **Vyhledání dostupných nástrojů a platforem:** Identifikace a evaluace softwarových řešení a platforem, které umožňují efektivní sběr, zpracování a analýzu dat. Zahrnuje přehled populárních nástrojů jako jsou R, Python, Microsoft Excel, Tableau, Power BI a další, včetně dostupných online návodu a tutoriálů pro jejich použití.
3. **Prozkoumání existujících prací a literatury:** Detailní rešerše a analýza akademických prací, článků a studií zabývajících se tématem datové analýzy ve vzdělávání. Identifikace klíčových zjištění, metodologických přístupů a nevyřešených otázek, které by mohly inspirovat a informovat naši práci.
4. **Analýza a evaluace metod:** Prozkoumání současných metod datové analýzy a statistik používaných pro interpretaci a zpracování vysokoškolských dat. Identifikace vhodných technik a nástrojů pro zpracování komplexních datových souborů.

### 2.1.2 Existující aplikace pro datovou analýzu

V současném dynamickém světě datové analýzy existuje celá řada nástrojů a technologií, které jsou klíčové pro efektivní práci s daty, jejich analýzu a vizualizaci. Tyto nástroje jsou vysoce ceněny jak v akademických kruzích, tak v průmyslovém sektoru pro jejich schopnost zpracovávat rozsáhlé datové sady, poskytovat hluboké analýzy a podporovat rozhodovací procesy na základě dat. Níže (viz 2.1.2) uvádím některé z nejvíce uznávaných a používaných softwarových řešení podle různých internetových článků<sup>1234</sup> a také grafu na obrázku 2.1, který porovnával nejjádanější technické zdatnosti v data science pracích v roce 2022.

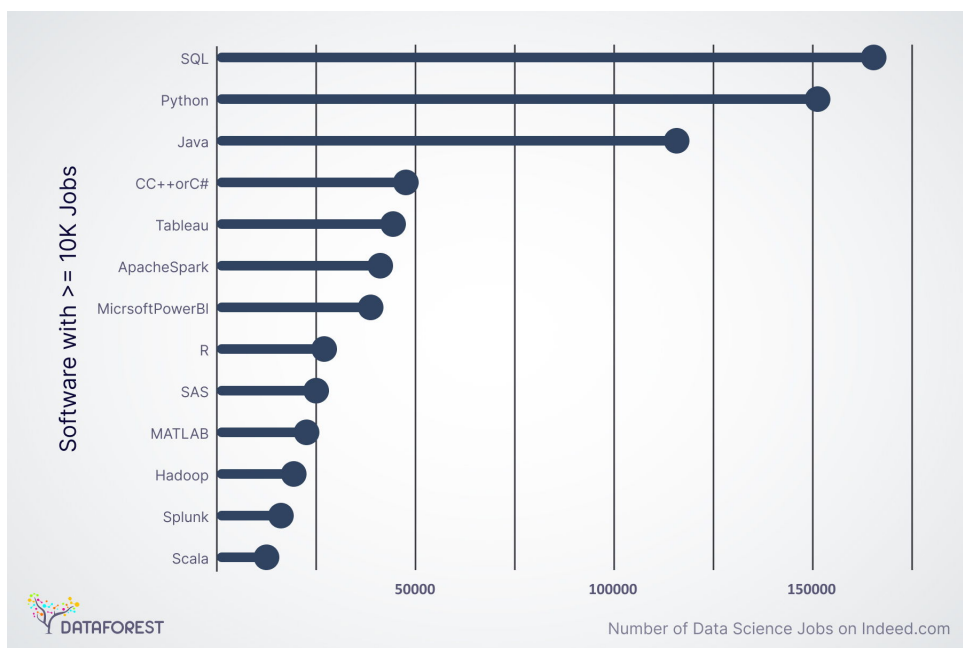
---

<sup>1</sup><https://skillmea.cz/blog/11-najlepsich-nastrojov-na-analyzu-dat>

<sup>2</sup><https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/>

<sup>3</sup><https://dataforest.ai/blog/choosing-data-science-tools-striking-the-balance>

<sup>4</sup><https://www.finereport.com/en/data-analysis/comparison-of-data-analysis-tools-excel-r-python.html>



**Obrázek 2.1:** Počet pracovních pozic v oblasti data science podle nástrojů. Zdroj: DataForest[3]

Výhody a nevýhody těchto nástrojů byly definovány na základě analýzy dostupné literatury, článků a osobních zkušeností s těmito nástroji. Zvláštní pozornost byla věnována kontextu této práce, která se zaměřuje na vzdělávací prostředí a potřeby uživatelů, kteří mohou být začátečníci v oblasti datové analýzy.[4, 5, 6, 3]

## R a RStudio

R<sup>5</sup> je programovací jazyk a prostředí specializované pro statistické výpočty a grafiku. RStudio je IDE<sup>6</sup>, které usnadňuje práci s R.

### Výhody:

- Rozsáhlé knihovny pro statistickou analýzu a vizualizaci.
- Silná podpora komunity.
- Vhodné pro akademický výzkum a složité statistické modely.

### Nevýhody:

- Strmá křivka učení pro začátečníky.
- Méně intuitivní uživatelské rozhraní.

**Cena a dostupnost:** R je open-source software, dostupný zdarma. RStudio nabízí jak zdarma dostupnou verzi RStudio Desktop, tak komerční verzi RStudio Server Pro, která je zpoplatněna na základě licenčního modelu.

<sup>5</sup><https://www.rstudio.com>

<sup>6</sup>Vývojové prostředí





## ■ Microsoft Excel

Excel<sup>13</sup> je široce používaný tabulkový procesor, který nabízí nástroje pro analýzu dat, základní vizualizace a možnost výpočtů pomocí formulí.

### Výhody:

- Snadná dostupnost a znalost většinou uživatelů.
- Flexibilní manipulace s daty a široké možnosti výpočtů.
- Možnost integrace s jinými aplikacemi Microsoft Office.

### Nevýhody:

- Omezené možnosti pro pokročilou analýzu dat.
- Výkon může být omezený při práci s velmi velkými datovými sadami.

**Cena a dostupnost:** Microsoft Excel je součástí balíčku Microsoft Office, který je dostupný buď jako jednorázový nákup nebo předplatné Microsoft 365. Předplatné nabízí další služby a aktualizace.

## ■ Vyhodnocení existujících aplikací

Existuje veliké množství nástrojů pro datovou analýzu, od jednodušších jako Excel, se kterým se složitěji vytvářejí pokročilejší grafy nebo analýzy, přes Power BI a Tableau, což jsou kvalitní vizualizační aplikace, které mohou být díky komplexnosti pro začátečníka náročné na zorientování a používání, až po programovací jazyky jako R a Python, které jsou pro člověka bez znalostí programování nepoužitelné bez návodů či kurzů.

Proto má smysl pokračovat a pokusit se vytvořit aplikaci, která bude jednoduchá pro uživatele a zvládne i pokročilejší úkony.

## ■ 2.1.3 Literatura a související práce

### ■ Existující literatura

Při hledání relevantní literatury jsem se zaměřil na klíčové pojmy spojené s analýzou v této bakalářské práci, jako jsou *studenti*, *absolventi*, *přijímací řízení*, *průchodnost studia*, *vysoké školy*, *trh práce*, *atraktivnost studia*, *platové ohodnocení*. Nalezl jsem několik prací a článků, které se dotýkají částí mého plánu.

Nejčastěji se literatura zabývá nástupními platy vysokoškoláků a atraktivitou jednotlivých zaměstnání. Článek na Lidovkách „*Víme, co se vyplatí studovat. Přinášíme přehled nástupních platů vysokoškoláků*“ se zaměřuje na nástupní mzdy studentů různých oborů a sleduje, ve kterých oblastech došlo k největšímu nárůstu.[8]

Dalším značným tématem jsou data o úspěšnosti a průchodnosti studia. Tyto informace buď poskytují samotné školy, nebo různá online média, jako

<sup>13</sup><https://www.microsoft.com/cs-cz/microsoft-365/excel>

například portál <https://www.vysokeskoly.com>, kde lze nalézt počty nově zapsaných studentů a absolventů, což umožňuje vytvořit si vlastní pohled na situaci.

Zajímavé je, že mnohé analýzy a statistiky provádí také samotné ministerstvo školství, mládeže a tělovýchovy (MŠMT) nebo samy vysoké školy. Tyto instituce často publikují rozsáhlé zprávy a studie, které se zabývají různými aspekty vzdělávacího systému, včetně úspěšnosti studentů, demografických trendů nebo vývoje na trhu práce. Tato data jsou neocenitelná pro hlubší analýzu a poskytují solidní základ pro vědecký výzkum[9].

Statistiky o úspěšnosti jsou populární i u renomovaných zahraničních škol, například na Harvardu nebo MIT. Některé školy, jako je MIT, jsou otevřenější ve sdílení těchto informací a zveřejňují široké spektrum údajů přímo na svých stránkách. Obrázek 2.2 znázorňuje jednu z dostupných statistik přímo na webových stránkách školy. Naopak Harvard poskytuje méně veřejně dostupných informací, ale existuje řada publikací od jiných institucí, které se zaměřují na tuto prestižní školu.

Degrees Awarded at MIT

Degrees Awarded from July 1, 2021 to June 30, 2022:

Ethnicity	Bachelor's	Master's	Doctorate	Total
International (Not US Citizen or Perm. Resident)	86	842	336	1,264
Hispanic or Latino	179	106	45	330
American Indian or Alaskan Native	1	1	0	2
Asian	346	267	71	684
Black or African American	60	65	8	133
Native Hawaiian or Other Pacific Islander	2	0	0	2
White	324	593	211	1,128
Two or more races	82	61	20	163
Race/Ethnicity Unknown	20	55	44	119
Total	1,100	1,990	735	3,825

**Obrázek 2.2:** Udělené tituly z MIT od července 2021 do června 2022. (Zdroj MIT)[10]

Existují také bakalářské a diplomové práce zaměřené na různé aspekty vysokých škol a studijních oborů. Tyto práce často pokrývají specifické otázky, jako jsou úspěšnost studentů, demografické trendy a jejich vliv na vzdělávací systém.

## ■ Omezení existující literatury

Dostupná literatura týkající se dat a podobných analýz je často úzce specializovaná na konkrétní otázky a často chybí širší kontext nebo hlubší úvahy. V článkách a pracích zaměřených na platy a atraktivitu zaměstnání se obvykle kladl důraz na platové aspekty konkrétních oborů, což dokládá jejich atraktivitu. Ovšem tyto práce často opomíjejí faktory, jako je průchodnost studia, náročnost vzdělávacího procesu, a ne vždy zohledňují přeplněnost pracovního trhu v daných odvětvích. Tím zůstává stranou reálná šance absolventů na získání zaměstnání.

V případě statistik o úspěšnosti studia je častým omezením to, že se zaměřují převážně na poměr zapsaných k absolventům. Tyto statistiky někdy nedokážou zachytit hlubší význam těchto dat a nepřihlížejí k různým faktorům, jako jsou opakovaně zapsaní studenti nebo jedinci, kteří studují více oborů současně. To může vytvářet zkreslený obraz o úspěšnosti studijního procesu a jeho skutečném vlivu na absolventy.

## ■ Shrnutí literatury

Celkově existuje mnoho zdrojů a studií zabývajících se jednotlivými aspekty vysokého školství a trhu práce, avšak chybí ucelený pohled na všechny klíčové faktory, které ovlivňují jak úspěšnost studia, tak následné uplatnění absolventů na trhu práce. Z dané literatury můžeme čerpat inspiraci pro formulaci hypotéz a porovnat naše výsledky s existujícími pracemi. Téma vysokých škol a studentů vysokých škol je široce probírané a důležité pro pochopení současných vzdělávacích trendů a jejich dopadů na trh práce. Nikdo se však dosud nezabýval vývojem aplikace pro zpracování těchto dat a jejich jednodušší analýzu, což podtrhuje význam této práce.

### ■ 2.1.4 Prozkoumání dostupných návodů a tutoriálů pro datovou analýzu

#### ■ Rozmanitost a specifičnost dostupných tutoriálů

Při zkoumání dostupných návodů a tutoriálů pro nástroje datové analýzy je důležité si uvědomit, že tyto zdroje se značně liší jak ve kvalitě, tak v obsahu a délce. Kurzy na platformách jako Coursera<sup>14</sup> a Udemy<sup>15</sup> mohou trvat od několika hodin po několik týdnů, v závislosti na složitosti a hloubce materiálu.[11]

Některé kurzy jsou navrženy tak, aby byly přístupné úplným začátečníkům, zatímco jiné předpokládají alespoň základní znalost programování nebo předchozí zkušenost s analýzou dat. Například kurzy, které vyučují R nebo Python,

<sup>14</sup>Coursera - online vzdělávací platforma nabízející kurzy od univerzit a korporací po celém světě. <https://www.coursera.org>

<sup>15</sup>Udemy - online vzdělávací platforma pro dospělé, která nabízí kurzy na různá témata od programování po osobní rozvoj. <https://www.udemy.com>



působícím na území České republiky vést matriku studentů, která slouží k evidenci studentů a k rozpočtovým a statistickým účelům.[13] Vzhledem k této zákonné povinnosti by data měla být co nejpřesnější a aktuální.

#### Struktura dat[12]:

- **F1 Souhrny:** Obsahuje ucelené souhrny dat o celkovém počtu studentů, poprvé zapsaných a absolventů od roku 2001. Klíčové tabulky pro analýzu jsou F11, F12 a F13, které se postupně zaměřují na studenty, absolventy a poprvé zapsané podle typu studia.
- **F2 Studenti:** Podrobnější tabulky zaměřené na studenty a různé typy studijních programů.
- **F3 Poprvé zapsaní:** Obsahuje data o poprvé zapsaných studentech, rozdělená podle formy a typu studia na jednotlivých fakultách vysokých škol.
- **F4 Absolventi:** Statistiky týkající se absolventů vysokých škol.
- **F5 Ostatní formy vzdělávání na VŠ:** Informace o studentech v rámci celoživotního vzdělávání s daty dostupnými od roku 2017.

**Ukazatel F:** Pro vysvětlení písmenka F v následujících popiscích – jedná se o Ukazatel F (vzdělávací projekty, programy a záměry) a je zdrojem pro příspěvky novým vysokým školám a pro zvýšení příspěvků. Podporuje vzdělávací projekty a aktivity, které nejsou financovány z jiných ukazatelů. Žádosti o příspěvek posuzuje náměstek pro vysoké školství ministryně školství, mládeže a tělovýchovy. Tento ukazatel slouží jako výkonnostní nástroj MŠMT.[14]

### ■ Statistický úřad

#### Typy dat:

- **Populace ČR:** Zahrnuje informace o populaci České republiky.
- **Narozené děti:** Obsahuje údaje o počtech novorozenců.

**Cíl použití dat:** Slouží k odhadu potenciálního počtu osob, které mohou studovat, a umožňuje porovnání počtu studujících a nestudujících v populaci. Tyto zdroje poskytují komplexní a strukturované informace pro provedení analýzy v oblasti vysokého školství v České republice.

### ■ 2.1.6 Role umělé inteligence v analýze dat

Umělá inteligence (AI) je stále důležitějším nástrojem v analýze dat, ale její schopnosti mají své limity. Přestože AI může efektivně automatizovat některé základní a opakované úlohy a poskytovat rychlé shrnutí dat, není zatím schopna nahradit datové analytiky a vědce v plném rozsahu. Důvodem je, že složitější úkoly vyžadující hluboké porozumění datům, tvůrčí myšlení a

kritické hodnocení zůstávají mimo dosah současných AI systémů. AI je tedy spíše doplňkem, který může výrazně zrychlit a zefektivnit některé procesy, avšak pro komplexní analýzy a rozhodování je stále nezbytná lidská expertíza.

Tuto práci jsem se rozhodl zaměřit na tradiční metody datové analýzy a nevyužívat AI, protože komplexní porozumění a interpretace dat vyžadují lidský přístup. AI může být v budoucnu cestou, jak zefektivnit analýzu dat, ale v současnosti je její role doplňková a nikoliv zcela nahraditelná. Pro budoucí výzkumy by však bylo zajímavé zkoumat, jakým způsobem lze AI nástroje lépe integrovat do procesu datové analýzy.[15]

## ■ Zdokonalení AI technologií

S rozvojem technologií strojového učení a hlubokého učení se AI stále zdokonaluje ve schopnosti identifikovat vzorce a trendy v datech. Například platformy jako OpenAI představují nástroje s technologií GPT-4, které umožňují uživatelům přímo nahrávat Excel soubory a získávat z nich podrobné analýzy bez potřeby pokročilých znalostí v datové vědě.[2]

## ■ Příklady využití AI v datové analýze

Jedním z příkladů využití AI v datové analýze je automatizovaná extrakce informací, kde AI algoritmy procházejí velké množství textových dat k identifikaci a extrakci relevantních informací. Dále AI pomáhá ve vizualizaci dat tím, že automaticky generuje grafy a vizuální reprezentace datových trendů.[16]

## ■ Vliv AI na prediktivní analýzu

AI také významně posiluje prediktivní analýzu, kde modely strojového učení predikují budoucí trendy na základě historických dat. Tato schopnost je zásadní v oblastech jako jsou finance, zdravotní péče a retail, kde firmy mohou předjímat tržní vývoj a lépe reagovat na měnící se potřeby zákazníků.

## ■ Shrnutí AI

AI je na trhu dostupná a neustále se rozvíjí, ale zatím není schopna samostatně nahradit lidskou expertízu v komplexní datové analýze. V této práci AI nepoužíváme, protože tradiční metody poskytují hlubší porozumění a interpretaci dat. AI je však slibnou alternativou pro budoucnost a může se zdokonalit, aby lépe podporovala analytické procesy.[16]

### ■ 2.1.7 Vyhodnocení rešerše

Tato rešerše poskytla komplexní přehled o dostupných zdrojích, nástrojích a metodách pro datovou analýzu v kontextu vysokého školství v České republice. Identifikace a analýza veřejně dostupných datových souborů, nástrojů a platforem ukázaly, že existuje široké spektrum možností, jak efektivně zpracovávat a analyzovat data týkající se studentů a vysokoškolských institucí.

Díky moderním technologiím a metodám, jako je umělá inteligence a strojové učení, je nyní možné dosáhnout hlubších a přesnějších analýz, které mohou významně přispět k lepšímu pochopení a zlepšení vzdělávacích procesů a politik.

Výsledky této rešerše také poukazují na možnost vytvoření aplikace, která by umožňovala provádět různé datové analýzy nad běžně dostupnými daty o studentech a studiiích na vysokých školách. Hlavním cílem této práce je vyvinout jednoduchou a uživatelsky přívětivou aplikaci, která by usnadnila analýzu těchto dat. Aplikace by měla sloužit nejen pro analýzu a vizualizaci dat, ale také pro edukativní účely, pomáhající nováčkům v oblasti datové analýzy.

Tento cíl se rozvinul v průběhu výzkumu, kdy jsme si uvědomili potenciál aplikace nejen jako nástroje pro analýzu dat, ale také jako prostředku pro seznámení nových uživatelů se základy datové analýzy. Taková aplikace by mohla sloužit jako první krok pro ty, kteří mají zájem o datovou analýzu, ale cítí se odrazeni její zdánlivou složitostí nebo náročností dostupných nástrojů. S takovým základním nástrojem by mohli postupně přecházet k pokročilejším aplikacím a technikám, což by mohlo výrazně rozšířit komunitu schopnou efektivně využívat datovou analýzu.

## 2.2 Metody a techniky datové analýzy

Tato kapitola se zabývá metodami a technikami datové analýzy, které byly použity při zpracování dat zmíněných v předchozí kapitole 2.1.5. Popisuje konkrétní postupy pro sběr dat, jejich validaci a integraci s různými nástroji pro analýzu a vizualizaci. Tyto postupy byly použity nejen pro tvorbu této práce a vývoj aplikace, ale také následně pro vytváření výukových materiálů. Zaměříme se na efektivní způsoby manipulace s daty, které nám umožní provádět komplexní analýzy a poskytovat přesné výsledky.

### 2.2.1 Metody pro sběr dat

V této části se zaměříme na metody a nástroje, které jsme použili pro sběr dat nezbytných k analýze. Popíšeme jednotlivé techniky extrakce dat z různých zdrojů, jejich integraci a validaci. Tyto metody nám umožnily získat kvalitní a spolehlivá data potřebná pro další analýzu a vývoj aplikace.

#### Extrahování dat z Excelu

**Popis:** Použití Python knihoven jako Pandas a Openpyxl umožňuje efektivní načítání a zpracování dat přímo z Excelovských tabulek, které poskytuje Ministerstvo školství, mládeže a tělovýchovy (MŠMT). Tento přístup umožňuje automatizaci procesů a snadné manipulace s daty, včetně filtrování, třídění a výpočtů.[17]

**Využití:** Automatizovaný sběr dat z oficiálních zdrojů, jako jsou matriky studentů poskytované ministerstvem, zajišťuje aktuálnost a přesnost informací



potřebných pro analýzu. Podrobnější popis postupů a metod extrakce dat bude poskytnut v kapitole věnované datové analýze, kde budou prezentovány konkrétní techniky a přístupy použité v průběhu zpracování dat.

### ■ Integrace s Power BI

**Popis:** Propojení Python skriptů s Power BI umožňuje vizualizaci a analýzu dat. Tento proces zahrnuje přenos dat z Pythonu do Power BI pomocí přímé integrace nebo exportu datových souborů, což umožňuje dynamické aktualizace a interaktivní analýzy.

**Využití:** Přenos dat z Python skriptů do Power BI pro tvorbu interaktivních dashboardů a reportů nabízí uživatelům nástroje pro lepší vizuální prezentaci a hlubší porozumění datům. Ačkoliv tato metoda bude dostupná, budeme dále pracovat na vývoji vlastní aplikace, která umožní ještě efektivnější a specifické analýzy přizpůsobené našim potřebám.

### ■ Kombinace dat z různých zdrojů

**Popis:** Integrace dat z MŠMT s dalšími zdroji, jako jsou data od statistického úřadu, nabízí možnost provádět komplexnější analýzy. Tato kombinace dat umožňuje srovnávat, kombinovat a analyzovat informace z různých oblastí.

**Využití:** Získání širšího kontextu a podpora rozhodovacích procesů pomocí více zdrojů dat poskytuje bohatší základ pro analýzu a zlepšuje kvalitu výsledků.

### ■ Validace dat

**Popis:** Implementace mechanismů pro kontrolu a validaci dat před jejich analýzou zajišťuje, že výstupy jsou spolehlivé a přesné. To může zahrnovat kontrolu chybějících hodnot, nekonzistencí nebo odlehlých hodnot, které by mohly ovlivnit analýzu.

**Využití:** Zajištění kvality a spolehlivosti dat v analýzách je klíčové pro udržení důvěry ve výsledky a pro zajištění objektivitu rozhodnutí založených na datech.

### ■ Celkově

Zatímco integrace s Power BI a dalšími nástroji poskytuje silnou platformu pro analýzu a vizualizaci dat, naše zaměření na vývoj vlastní aplikace nám umožní specifikovat a optimalizovat analytické procesy pro naše konkrétní účely. V současné fázi se budeme věnovat pouze jednoduchému zahoezení nevalidních hodnot, avšak je důležité zdůraznit, že existuje mnoho dalších metod a způsobů, jak s chybějícími daty pracovat, jako například nahrazování průměrem, mediánem a dalšími složitějšími technikami imputace<sup>16</sup>. [18]

<sup>16</sup>Imputace je proces nahrazování chybějících nebo neúplných dat v datasetu odhadovými hodnotami



## 2.2.2 Metody pro analýzu dat

V této části se zaměříme na různé metody a techniky, které jsou použity pro analýzu dat. Popíšeme konkrétní přístupy k deskriptivní statistice, prediktivní analýze a dalším pokročilým technikám. Tyto metody nám umožňují získat hlubší vhled do dat a poskytují základ pro další interpretaci výsledků.

### Deskriptivní statistika

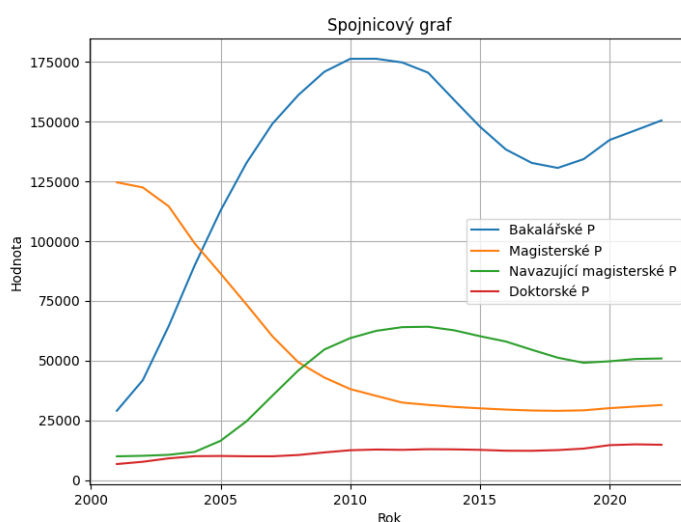
**Popis:** Deskriptivní statistika slouží jako základ pro sumarizaci a vizualizaci základních charakteristik datových sad. Jedná se o prvotní krok v analýze, který poskytuje přehled o rozložení a hlavních vlastnostech souboru dat.

**Využití:** Základní statistické metody umožňují identifikaci trendů a rozložení studentů podle různých kategorií, což je zásadní pro další analýzu a rozhodování.

**Grafické metody.** Spojnicové, sloupcové, bodové, kruhové, histogramy, krabicové a geografické grafy jsou různé typy vizualizací dat, které se používají k zobrazení informací v srozumitelné a vizuálně přitažlivé formě. Každý typ grafu má specifické využití v závislosti na povaze dat a účelu analýzy.[19]

#### Spojnicové grafy (Line Graphs):

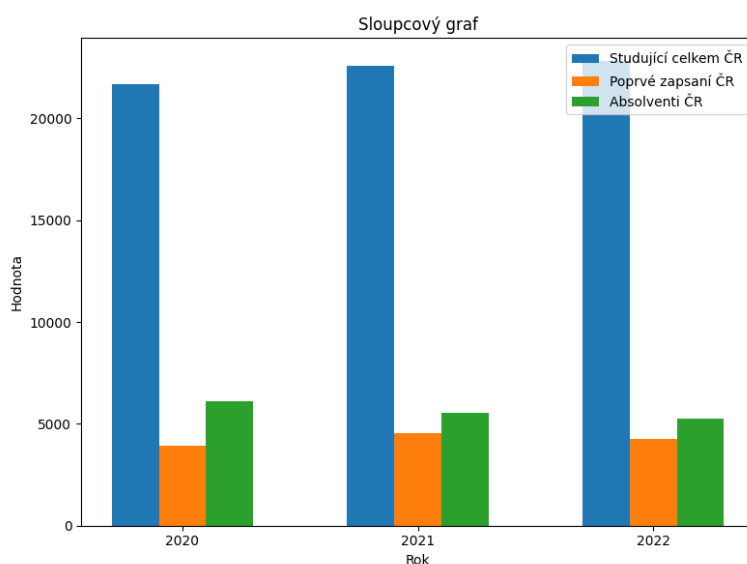
- Používají se k zobrazení dat v průběhu času (časové řady).
- Ideální pro sledování změn nebo trendů.
- Vhodné pro porovnávání více datových řad.



Obrázek 2.3: Příklad spojnicového grafu

### Sloupcové grafy (Bar Graphs):

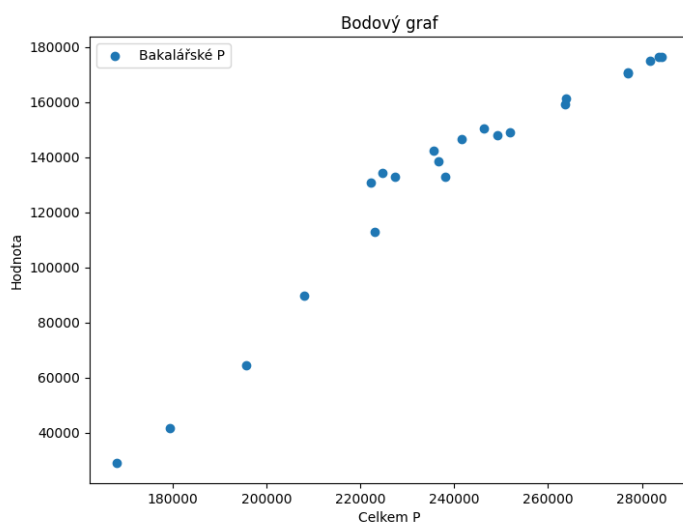
- Umožňují srovnání množství mezi různými skupinami.
- Sloupce mohou být zobrazeny svisle nebo vodorovně.
- Často se používají pro ilustraci porovnání mezi položkami.



Obrázek 2.4: Příklad sloupcového grafu

### Bodové grafy (Scatter Plots):

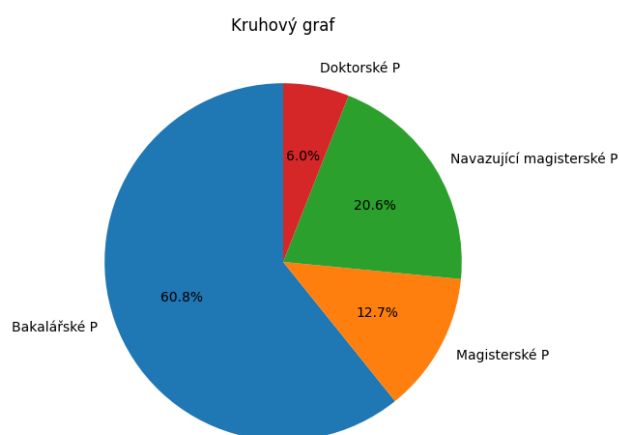
- Zobrazují vztah mezi dvěma proměnnými.
- Každý bod na grafu reprezentuje jednu hodnotu v datasetu.
- Užitečné pro identifikaci korelací nebo odlehlých hodnot.



**Obrázek 2.5:** Příklad bodového grafu

### Kruhové grafy (Pie Charts):

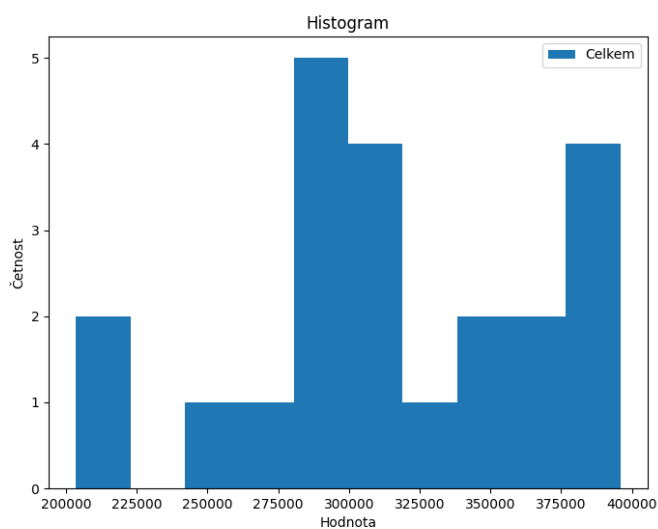
- Zobrazují procentuální nebo relativní podíl částí na celku.
- Vhodné pro zobrazení složení nebo podílů.
- Méně přesné pro detailní srovnání velikostí částí.



**Obrázek 2.6:** Příklad kruhového grafu

### Histogramy (Histograms):

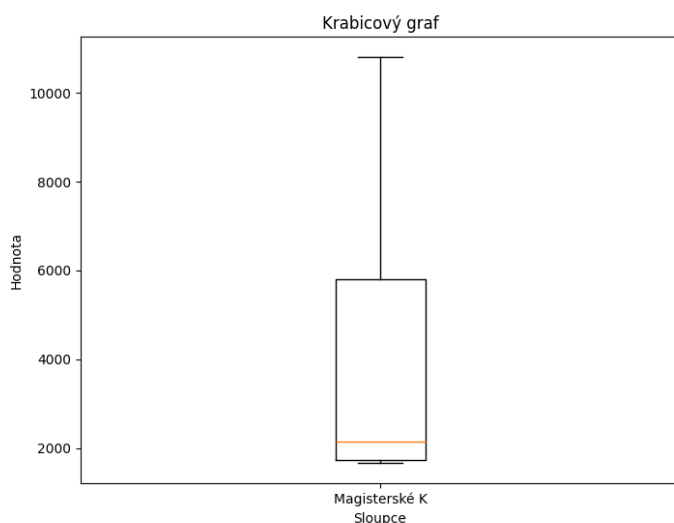
- Používají se k zobrazení frekvence dat v různých intervalových skupinách.
- Ideální pro vizualizaci distribuce dat.
- Pomáhají identifikovat rozložení dat, jako je normalita a symetrie.



Obrázek 2.7: Příklad histogramu

### Krabicové grafy (Box Plots):

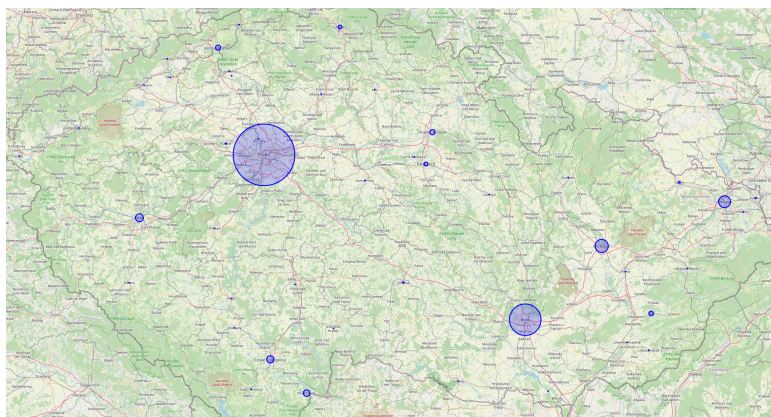
- Znáznorňují distribuci dat skrze kvartily.
- Umožňují rychle identifikovat medián, rozptyl a potenciální odlehlé hodnoty.
- Užitečné pro porovnání rozložení mezi několika skupinami.



**Obrázek 2.8:** Příklad krabicového grafu

### Geografické grafy (Geographic Maps):

- Vizualizují data v geografickém kontextu.
- Umožňují zobrazit, jak se data liší v závislosti na geografické poloze.
- Často se používají v demografii, meteorologii nebo pro analýzu tržních trendů.



**Obrázek 2.9:** Příklad geografického grafu

Každý z těchto grafů má specifické využití a je důležité vybrat typ grafu, který nejlépe odpovídá analýze a prezentaci dat, aby byly informace prezentovány jasně a efektivně. Například, pokud chceme ukázat trendy prodeje v průběhu roku, použijeme spojnicový graf. Pokud chceme porovnat počty studentů mezi různými městy, použijeme sloupcový graf.

### Základní statistické míry.

- **Průměr:** Reprezentuje střední hodnotu datové sady a je vypočítán jako součet všech hodnot dělený počtem hodnot.
- **Medián:** Střední hodnota datové sady, kde polovina hodnot je nižší a polovina vyšší. Je užitečný v případě, že data obsahují odlehlé hodnoty, které by mohly průměr zkreslit.
- **Modus:** Nejčastěji se vyskytující hodnota v datové sadě.
- **Směrodatná odchylka:** Míra variability nebo rozptylu dat kolem průměru. Vyšší hodnota naznačuje větší rozptýlení dat.
- **Minimální a maximální hodnota:** Nejnižší a nejvyšší hodnota v datové sadě.
- **25. a 75. percentil:** Hodnoty, pod kterými spadá 25 % resp. 75 % hodnot datové sady.

**Aplikace deskriptivní statistiky:** Deskriptivní statistika se široce používá v různých oblastech včetně vědy, inženýrství, obchodu a veřejného zdraví pro:

- Počáteční analýzu dat.
- Zpracování a čištění dat.
- Zprávy a prezentace.

**Význam deskriptivní statistiky:** Deskriptivní statistika je klíčová pro efektivní analýzu dat, protože poskytuje základní pochopení datové sady a umožňuje informované rozhodování na základě dat. Je to první krok v jakémkoliv statistickém vyšetřování a základ pro další analytické techniky.[20]

**Příklad základních statistických měř** Uvažujme tabulku, která obsahuje údaje o studentech na vysokých školách. Na základě této tabulky můžeme spočítat a zobrazit základní statistické míry, které nám poskytnou přehled o dané datové sadě.

Název sloupce	Počet	Průměr	Směrodatná odchylka	Min	25. percentil	Medián	75. percentil	Max
Celkem	22.0	315623.2272727273	54200.48482908425	203448.0	289509.5	307781.5	362525.75	395994.0
Celkem P	22.0	239949.4090909091	32794.55559585214	168125.0	223530.75	239865.0	263809.75	284155.0
Bakalářské P	22.0	133290.5	41977.04169429716	29030.0	131176.75	144370.5	160664.25	176340.0
Magisterské P	22.0	53649.181818181816	33638.30125606751	29040.0	30250.0	33858.5	70159.25	124609.0
Navazující magisterské P	22.0	43491.5	15966.75180570603	9955.0	27309.5	50803.5	59079.75	64182.0
Doktorské P	22.0	11675.363636363636	2178.0555667723565	6726.0	10039.75	12405.0	12851.0	14974.0
Celkem K	22.0	78363.13636363637	32383.748966455132	36639.0	64640.5	75164.5	94993.25	116744.0
Bakalářské K	22.0	45834.09090909091	16948.690470399134	14087.0	36442.0	42133.5	58912.75	74364.0
Magisterské K	22.0	4027.090909090909	3158.5316948754007	1663.0	1724.75	2140.0	5794.5	10813.0
Navazující magisterské K	22.0	17819.227272727272	8387.85491581397	2113.0	12697.25	20499.0	23419.0	27826.0
Doktorské K	22.0	10908.0	2485.4103428654865	5748.0	9823.0	11309.5	12747.0	14057.0
Rok	22.0	2011.5	6.493586579592718	2001.0	2006.25	2011.5	2016.75	2022.0

**Obrázek 2.10:** Příklad základních statistických měř

Na obrázku 2.10 vidíme různé statistické míry, jako je počet položek, průměr, směrodatná odchylka, minimum, 25. percentil, medián, 75. percentil a maximum pro různé kategorie dat.

## ■ Korelace a Lineární Regrese

**Korelace.** Korelace je statistická metoda používaná k vyhodnocení vztahu mezi dvěma nebo více proměnnými. Koeficient korelace měří sílu a směr tohoto vztahu. Hodnoty blízké +1 indikují silnou pozitivní korelaci, kde hodnoty jedné proměnné rostou s hodnotami druhé proměnné. Hodnoty blízké -1 označují silnou negativní korelaci, kde hodnoty jedné proměnné klesají s růstem hodnot druhé proměnné. Hodnoty blízké 0 naznačují slabý nebo žádný lineární vztah mezi proměnnými.[21]

**Vzorec pro výpočet Pearsonova korelačního koeficientu[22]:**

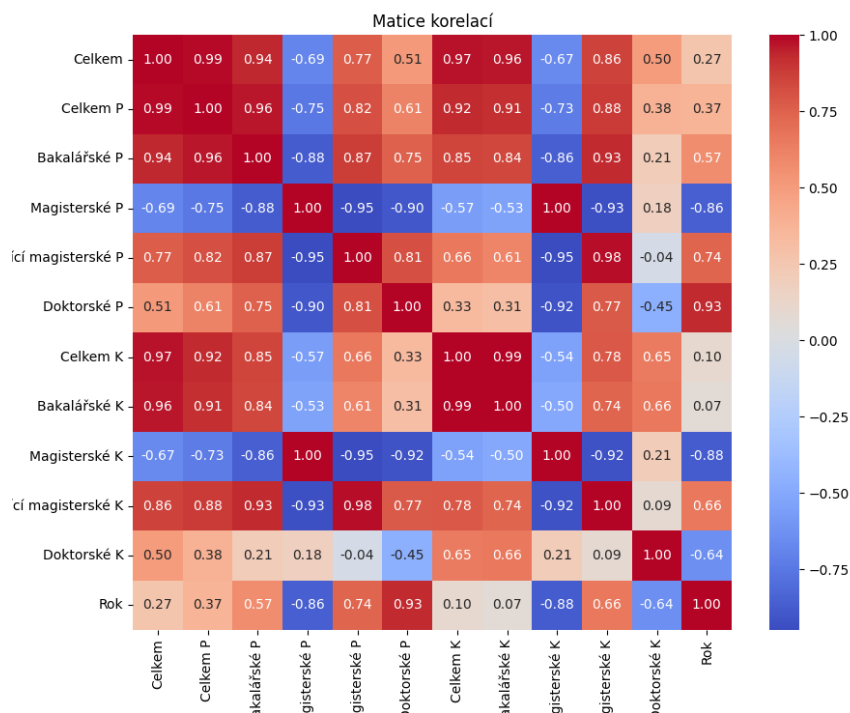
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

kde  $x_i$  a  $y_i$  jsou hodnoty dvou proměnných,  $\bar{x}$  a  $\bar{y}$  jsou průměrné hodnoty těchto proměnných.

### Předpoklady pro korelaci

- Lineární vztah: Korelace efektivně měří pouze lineární vztahy mezi proměnnými.
- Normální rozdělení: Pro Pearsonův korelační koeficient je potřeba, aby obě proměnné měly normální rozdělení.
- Homoskedasticita: Rozptyl jedné proměnné by měl být konstantní na všech úrovních druhé proměnné.
- Nezávislost vzorků: Hodnoty v jednom vzorku by neměly být závislé na hodnotách v jiném vzorku.

**Příklad korelace** Na obrázku 2.11 je korelační matice, která ukazuje vztahy mezi různými proměnnými v datasetu zahrnujícím studenty v letech 2001 až 2022. Barvy v matici reprezentují sílu a směr korelace mezi proměnnými. Červené a modré barvy označují silné pozitivní a negativní korelace, zatímco světlejší odstíny indikují slabší korelace.



Obrázek 2.11: Příklad korelační matice

**Lineární regrese.** Lineární regrese je metoda používaná k modelování vztahu mezi závislou proměnnou ( $Y$ ) a jednou nebo více nezávislými proměnnými ( $X$ ). Cílem je najít lineární rovnici, která nejlépe popisuje závislost mezi proměnnými. Základní forma jednoduché lineární regrese je vyjádřena rovnicí[23]:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

kde:

- $Y$  je závislá proměnná,
- $X$  je nezávislá proměnná,
- $\beta_0$  je intercept (průsečík s osou  $Y$ ),
- $\beta_1$  je směrnice regresní přímky,
- $\epsilon$  je náhodná chyba.

**Předpoklady pro lineární regresi[24]**

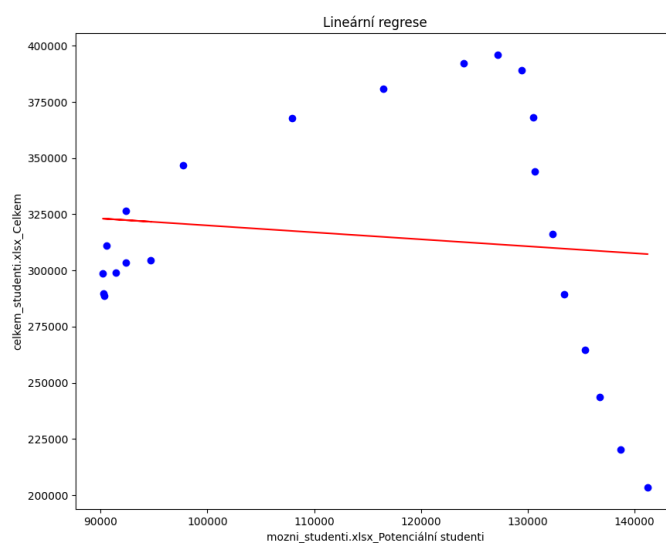
- Linearita vztahu.
- Homoskedasticita.
- Nezávislost reziduí<sup>17</sup>.

<sup>17</sup>Rezidua jsou rozdíly mezi skutečnými a předpovězenými hodnotami v regresním modelu. Tyto rozdíly ukazují, jak dobře model odpovídá pozorovaným datům.



- Normální rozdělení reziduí.
- Absence multikolinearity.

**Příklad lineární regrese** Na obrázku 2.12 je ukázka lineární regrese vztahu mezi jednotlivými roky a celkovým počtem studentů. Modré body představují skutečná data, zatímco červená čára představuje nejlepší linii, která modeluje tento vztah. Tato regrese ukazuje, jak se počet studentů mění v čase, a pomáhá identifikovat dlouhodobé trendy.



Obrázek 2.12: Příklad lineární regrese

## ■ Shrnutí kapitoly 2

V této kapitole jsme prozkoumali různé metody a techniky datové analýzy, které byly použity při zpracování dat v této práci. Popsali jsme postupy pro sběr dat, jejich validaci a integraci s nástroji pro analýzu a vizualizaci. Zaměřili jsme se na deskriptivní statistiku, korelaci a lineární regresi, které jsou klíčové pro získání hlubších vhledů do dat a jejich interpretaci. Tyto metody byly využity nejen pro tvorbu této práce a vývoj aplikace, ale také pro vytváření výukových materiálů.

Popsané techniky nám umožňují efektivně manipulovat s daty a poskytují základ pro komplexní analýzy a přesné výsledky. V další části práce se zaměříme na sběr a zpracování dat, kde aplikujeme tyto metody na konkrétní datové sady, což nám umožní získat konkrétní poznatky a podklady pro další analýzu a rozhodování.



## Kapitola 3

### Sběr a zpracování dat

Tato kapitola se zaměřuje na procesy sběru a zpracování dat, které jsou zásadní pro úspěšnou analýzu a vizualizaci v rámci tohoto projektu. Popisuje metody a techniky použité k získávání dat z různých zdrojů, jejich čištění a přípravu pro následné analýzy. Dále jsou zde definovány klíčové pojmy a struktury dat, které jsou nezbytné pro pochopení a správnou interpretaci informací. Vysvětleny jsou také specifické přístupy k manipulaci s datovými soubory a jejich transformace do formátů vhodných pro analýzu. Celkově kapitola poskytuje ucelený pohled na všechny kroky, od sběru počátečních dat až po jejich finální zpracování, což umožňuje čtenáři lépe pochopit metodiku práce s daty v tomto projektu.

Navazujeme na předchozí kapitolu, kde jsme prozkoumali teoretické základy a metody datové analýzy, které jsou relevantní pro oblast vysokého školství v České republice. V této části se konkrétně zaměříme na praktické aspekty sběru a zpracování dat, která se týkají studentů a vysokoškolských institucí.

#### 3.1 Definování klíčových pojmů

Následující seznam definuje klíčové pojmy související s tabulkami, které budou později zmiňovány ve sběru 3.3 a analýze dat 4. Tyto definice jsou převzaty z dat poskytnutých Ministerstvem školství, mládeže a tělovýchovy (MŠMT)[25] a navíc doplněny o populační křivku.

- **Student** – Student vysoké školy v České republice, který má alespoň jedno aktivní studium, tj. k 31. prosinci daného roku nemá všechna svá neukončená studia přerušena. Jinými slovy, alespoň jedno z jeho studií je v tomto období „aktivní“.
- **Poprvé zapsaný student** – Jedinec, který dosud nenastoupil na žádnou veřejnou nebo soukromou vysokou školu v České republice.
- **Student nově zapsaný** – Jedinec, který ke konci předchozího období, tedy k 31. prosinci předchozího roku, nebyl zapsán na žádné vysoké škole. Tato data monitorují každoroční příliv studentů na vysoké školy.

- **Student bez absolvování** – Student, který nedosáhl vysokoškolského vzdělání. Do celkového počtu studentů bez absolvování jsou zahrnuti pouze studenti bakalářského a magisterského nenavazujícího studia.
- **Student s absolvováním** – Student, který úspěšně dokončil některé ze svých vysokoškolských studií a nadále pokračuje ve studiu.
- **Nestudující s přerušným studiem** – Student, který přerušil všechna svá studia. Údaje o těchto studentech se nezapočítávají do souhrmných statistik počtu studentů.
- **Absolvent** – Student, který úspěšně dokončil všechny požadované studijní povinnosti a obdržel akademický titul.
- **Jinoplátci** – Studenti, kterým studium hradí stipendium nebo obdobné zdroje.
- **Samoplátci** – Studenti, kterým stát již nehradí studium a musí ho platit z vlastních zdrojů.
- **Vysoká škola** – Instituce poskytující terciární vzdělání a udělující akademické tituly.
- **Typ studia** – Charakteristika studia, například prezenční nebo kombinovaná forma, bakalářské nebo magisterské studium.
- **Bakalářské studium** – Studijní program, který vede k získání bakalářského titulu, obvykle trvá tři až čtyři roky.
- **Magisterské studium** – Studijní program, který následuje po bakalářském studiu a vede k získání magisterského titulu, obvykle trvá dva až tři roky.
- **Úspěšnost studia** – Míra, s jakou studenti úspěšně dokončují studium.
- **Populační křivka** – Grafická reprezentace rozdělení věkových skupin.

## 3.2 Sběr dat

Při sběru dat probíhal postupný proces stahování excelových tabulek ze stránek Ministerstva školství, mládeže a tělovýchovy (MŠMT)<sup>1</sup> a následně také ze stránek statistického úřadu<sup>2</sup>. Tento proces zahrnoval následující kroky:

1. **Identifikace zdrojů dat:** Nejprve byly identifikovány relevantní zdroje dat na webových stránkách MŠMT a statistického úřadu, které obsahovaly informace o studentech, absolventech a vysokých školách.

<sup>1</sup>[https://dsia.msmt.cz/vystupy/vu\\_vs.html](https://dsia.msmt.cz/vystupy/vu_vs.html)

<sup>2</sup>[https://www.czso.cz/csu/czso/obyvatelstvo\\_hu](https://www.czso.cz/csu/czso/obyvatelstvo_hu)

2. **Stahování dat:** Následně byly excelové tabulky s daty stahovány postupně z těchto zdrojů. Proces stahování byl opakovaný a postupný, přičemž každá tabulka byla systematicky získávána a ukládána.
3. **Ukládání dat:** Stahované tabulky byly uloženy do vhodně strukturovaných složek nebo databází tak, aby byly snadno přístupné pro další zpracování a analýzu. Ukládání probíhalo postupně, přičemž každá tabulka byla archivována s odpovídajícími metadaty pro identifikaci a usnadnění budoucího použití.

Celý proces sběru dat byl prováděn systematicky a pečlivě s důrazem na správnost a úplnost dat. Tímto způsobem byla zajištěna kvalitní datová sada pro další analýzu a zpracování. Procházel jsem data a hledal způsob, jak je efektivně převádět do použitelných tabulek. Během tvoření skriptů na sběr dat jsem kontroloval jejich správné přenesení a také úplnost tabulek MŠMT.

Zaměřili jsme se na tabulky, které jsou pod označením **F1 Souhrny**. Pro detailní popis jednotlivých tabulek viz následující kapitola 3.3.

## 3.3 Průzkum struktury a obsahu tabulek

V této kapitole provádíme analýzu a popis dat z různých zdrojů, které jsou klíčové pro naši studii. Nejprve se zaměříme na strukturu a obsah tabulek poskytnutých Ministerstvem školství, mládeže a tělovýchovy (MŠMT), konkrétně na **F1 Souhrny**, a poté se přesuneme k popisu souboru dat získaného ze statistického úřadu.

### 3.3.1 Ministerstvo školství, mládeže a tělovýchovy (MŠMT)

V této podkapitole provedeme analýzu a popis dat poskytnutých Ministerstvem školství, mládeže a tělovýchovy.

#### F11 – Souhrn studentů a absolventů podle formy a typu studijního programu

Tabulka F11 poskytuje souhrnné informace o studentech a absolventech, jak můžete vidět na obrázku 3.1, rozdělených podle formy a typu studijního programu. Obsahuje pět podtabulek, které zahrnují celkový souhrn všech vysokých škol, studentů s českým občanstvím, cizinců, veřejných škol a soukromých škol. Detailní informace o celkovém počtu studentů, poprvé zapsaných a absolventech jsou diferencovány podle formy studia a typu studijního programu.

Státní občanství Zřizovatel	fyzické osoby celkem										
	Celkem	v prezenčním studiu					v distančním a kombinovaném studiu <sup>4)</sup>				
		celkem	v typu studijního programu				celkem	v typu studijního programu			
			bakalář- ském	magister- ském <sup>2)</sup>	navaz. magister- ském <sup>1)</sup>	doktor- ském		bakalář- ském	magister- ském <sup>2)</sup>	navaz. magister- ském <sup>1)</sup>	doktor- ském
<b>Vysoké školy celkem</b>											
poprvé zapsaní <sup>3)</sup>	81 613	62 468	58 540	6 453	26 578	4 096	19 174	20 610	390	12 066	1 580
z toho ženy	45 943	34 024	31 466	4 134	14 922	1 859	11 938	12 583	355	7 564	649
celkem studentů	395 984	283 517	176 313	38 096	59 440	12 503	1 162 344	73 516	2 656	27 412	13 462
z toho ženy	221 528	156 309	94 817	24 830	32 700	5 609	67 921	43 812	2 274	16 498	5 517
absolventů	88 075	62 061	33 914	8 943	18 792	488	26 113	15 346	730	8 308	1 734
z toho ženy	52 861	36 799	20 361	6 067	10 218	200	16 134	9 839	539	5 079	679
<b>Studující se státním občanstvím ČR</b>											
poprvé zapsaní <sup>3)</sup>	72 676	55 286	52 576	5 084	23 923	3 505	17 416	18 802	389	10 641	1 381
z toho ženy	41 284	30 409	28 540	3 356	13 605	1 613	10 893	11 515	354	6 569	558
celkem studentů	358 506	255 371	161 416	31 722	54 006	10 842	107 169	68 179	2 630	24 534	12 120
z toho ženy	201 728	142 138	87 544	21 182	30 028	4 904	62 172	40 586	2 255	14 588	4 916
absolventů	81 332	57 342	31 707	8 014	17 271	424	24 084	14 339	723	7 450	1 577
z toho ženy	48 995	34 324	19 183	5 529	9 482	175	14 741	9 118	536	4 483	606
<b>Studující s cizím státním občanstvím</b>											
poprvé zapsaní <sup>3)</sup>	8 937	7 182	5 964	1 369	2 655	591	1 758	1 808	1	1 425	199
z toho ženy	4 659	3 615	2 926	778	1 317	246	1 045	1 068	1	995	91
celkem studentů	37 498	28 159	14 904	6 374	5 435	1 662	9 574	5 337	26	2 878	1 344
z toho ženy	19 809	14 178	7 277	3 648	2 673	705	5 750	3 226	19	1 910	601
absolventů	6 743	4 719	2 207	929	1 521	64	2 029	1 007	7	858	157
z toho ženy	3 866	2 475	1 178	538	736	25	1 393	721	3	596	73
<b>Veřejné vysoké školy</b>											
poprvé zapsaní <sup>3)</sup>	69 238	56 377	52 380	6 453	24 899	4 088	12 889	14 057	390	7 310	1 554
z toho ženy	38 060	30 141	27 546	4 134	13 736	1 854	7 938	8 453	355	4 423	639
celkem studentů	339 361	262 551	158 600	38 076	56 175	12 486	80 695	47 610	2 656	17 289	13 404
z toho ženy	186 459	143 354	84 033	24 816	30 520	5 599	45 586	28 015	2 274	9 965	5 482
absolventů	73 085	56 640	29 480	8 916	17 830	488	16 534	9 183	730	4 900	1 726
z toho ženy	42 979	33 124	17 380	6 047	9 542	200	9 920	5 855	539	2 854	674
<b>Soukromé vysoké školy</b>											
poprvé zapsaní <sup>3)</sup>	12 376	6 091	6 160	–	1 679	8	6 285	6 553	–	4 756	26
z toho ženy	7 883	3 883	3 920	–	1 186	5	4 000	4 130	–	3 141	10
celkem studentů	57 322	21 211	17 892	20	3 285	17	36 152	25 956	–	10 145	59
z toho ženy	35 488	13 106	10 889	14	2 196	10	22 405	15 826	–	6 550	35
absolventů	15 007	5 427	4 436	27	964	–	9 580	6 163	–	3 409	8
z toho ženy	9 895	3 680	2 982	20	678	–	6 215	3 984	–	2 226	5

**Obrázek 3.1:** Souhrn studentů a absolventů podle formy a typu studijního programu.

## ■ F12 – Souhrn škol, fakult, studentů a absolventů

Tabulka F12 (viz obrázek 3.2) představuje informace o školách, fakultách, studentech a absolventech. Data jsou rozdělena podle jednotlivých vysokých škol a jejich fakult. Jsou uvedeni studující celkem, ženy ze studijního celku a podrobnější rozdělení pro české studenty a cizince. Pro každou skupinu jsou prezentována data o celkovém počtu studentů, poprvé zapsaných, absolventech a studentech se zvláštním statutem.

### 3.3. Průzkum struktury a obsahu tabulek

Zřizovatel Vysoká škola Fakulta	Studenti (fyzické osoby) ve všech formách studia a typech studijních programů										
	studující celkem	z toho ženy	státní občanství ČR			cizinci			studenti s přerušenými studii	jino- plátcí <sup>2)</sup>	samo- plátcí <sup>3)</sup>
			studující celkem	poprvé zapsaní 1)	absol- venti <sup>2)</sup>	studující celkem	poprvé zapsa- ní <sup>1)</sup>	absol- venti <sup>2)</sup>			
<b>00000</b> všechny veřejné a soukromé VŠ	<b>304 518</b>	<b>169 179</b>	<b>249 767</b>	<b>48 217</b>	<b>50 034</b>	<b>54 770</b>	<b>12 498</b>	<b>8 665</b>	<b>14 371</b>	<b>63</b>	<b>10 256</b>
<b>10000</b> všechny veřejné VŠ	<b>276 632</b>	<b>153 557</b>	<b>226 967</b>	<b>43 979</b>	<b>44 797</b>	<b>49 681</b>	<b>11 193</b>	<b>7 735</b>	<b>11 952</b>	<b>63</b>	<b>10 256</b>
<b>11000</b> Univerzita Karlova	<b>50 663</b>	<b>31 401</b>	<b>39 450</b>	<b>5 966</b>	<b>6 338</b>	<b>11 213</b>	<b>2 067</b>	<b>1 516</b>	<b>2 893</b>	<b>27</b>	<b>4 086</b>
11110 1. lékařská fakulta	4 978	3 385	3 450	556	425	1 528	353	201	94	–	831
11120 3. lékařská fakulta	2 546	1 611	1 680	225	250	866	127	112	76	–	492
11130 2. lékařská fakulta	2 053	1 288	1 445	191	201	608	100	78	88	–	237
11140 Lékařská fakulta v Plzni	2 344	1 452	1 598	225	195	746	118	108	57	–	423
11150 Lékařská fakulta v Hradci Králové	1 791	1 099	1 187	186	190	604	120	81	38	–	322
11160 Farmaceutická fakulta v Hradci Králové	1 471	1 149	935	302	191	536	169	51	77	–	154
11210 Filozofická fakulta	6 280	4 338	5 336	695	888	944	131	104	744	2	41
11220 Právnícká fakulta	3 798	1 734	3 610	445	444	188	36	20	133	–	28
11230 Fakulta sociálních věd	5 351	2 913	3 674	538	795	1 677	193	357	344	–	1 097
11240 Fakulta humanitních studií	3 234	2 436	2 587	489	217	647	149	42	185	–	161
11260 Katolická teologická fakulta	543	311	473	32	65	70	13	7	107	–	–
11270 Evangelická teologická fakulta	366	231	324	44	37	42	8	8	124	–	6
11280 Husitská teologická fakulta	809	618	711	159	89	98	33	4	49	–	–
11310 Přírodovědecká fakulta	5 006	3 115	3 870	517	721	1 136	126	151	138	–	13
11320 Matematicko-fyzikální fakulta	3 050	764	1 908	396	322	1 142	299	109	158	–	177
11410 Pedagogická fakulta	5 534	4 390	5 338	658	977	196	36	36	378	–	13
11510 Fakulta tělesné výchovy a sportu	2 126	954	1 909	315	339	217	56	47	114	25	92
<b>12000</b> Jihočeská univerzita v Českých Budějovicích	<b>8 704</b>	<b>6 009</b>	<b>7 891</b>	<b>1 763</b>	<b>1 678</b>	<b>813</b>	<b>179</b>	<b>82</b>	<b>412</b>	<b>–</b>	<b>–</b>
12110 Zdravotně sociální fakulta	1 478	1 245	1 455	381	353	23	11	4	83	–	–
12210 Filozofická fakulta	851	647	817	211	111	34	5	8	18	–	–
12220 Zemědělská fakulta	1 110	597	1 030	250	221	80	27	2	52	–	–
12260 Teologická fakulta	785	599	741	131	140	24	7	4	77	–	–
12310 Přírodovědecká fakulta	1 137	646	693	125	136	444	48	37	82	–	–
12410 Pedagogická fakulta	2 062	1 530	2 040	357	466	22	9	4	40	–	–
12510 Ekonomická fakulta	1 129	693	1 009	272	224	120	63	15	51	–	–
12520 Fakulta rybářství a ochrany vod	198	70	132	36	27	66	9	8	10	–	–
<b>13000</b> Univerzita J. E. Purkyně v Ústí nad Labem	<b>7 881</b>	<b>5 103</b>	<b>7 535</b>	<b>1 490</b>	<b>1 158</b>	<b>346</b>	<b>136</b>	<b>29</b>	<b>413</b>	<b>–</b>	<b>1</b>
13410 Filozofická fakulta	862	495	815	134	81	47	20	2	81	–	1
13420 Fakulta strojního inženýrství	492	45	459	79	59	33	26	–	28	–	–
13430 Pedagogická fakulta	2 654	2 069	2 638	427	442	16	4	7	179	–	–
13440 Přírodovědecká fakulta	1 015	449	943	246	125	72	41	3	33	–	–
13450 Fakulta zdravotnických studií	988	797	934	228	164	54	16	4	29	–	–
13510 Fakulta sociálně ekonomická	1 216	860	1 183	271	175	33	8	1	31	–	–
13520 Fakulta životního prostředí	308	160	291	64	42	17	5	–	16	–	–
13530 Fakulta umění a designu	370	247	295	41	70	75	16	12	16	–	–
<b>14000</b> Masarykova univerzita	<b>32 335</b>	<b>20 089</b>	<b>24 350</b>	<b>4 312</b>	<b>5 214</b>	<b>7 985</b>	<b>1 757</b>	<b>1 520</b>	<b>1 934</b>	<b>–</b>	<b>1 570</b>
14110 Lékařská fakulta	4 613	3 094	2 636	408	478	1 977	397	244	148	–	825
14160 Farmaceutická fakulta	747	613	527	75	99	220	33	32	38	–	34
14210 Filozofická fakulta	6 184	4 472	4 879	707	986	1 305	275	251	401	–	74
14220 Právnícká fakulta	3 127	1 657	2 948	522	511	179	67	18	291	–	1
14230 Fakulta sociálních studií	3 222	2 096	2 305	370	558	917	211	196	193	–	387
14310 Přírodovědecká fakulta	3 439	2 034	2 482	548	629	957	184	219	160	–	12
14330 Fakulta informatiky	2 197	390	980	208	191	1 217	318	253	55	–	36
14410 Pedagogická fakulta	5 063	4 116	4 834	854	988	229	51	35	409	–	45
14510 Fakulta sportovních studií	1 520	624	1 369	268	302	151	46	27	122	–	–

Obrázek 3.2: Souhrn škol, fakult, studentů a absolventů.

## F13 – Studenti podle měst

Tabulka F13 (viz obrázek 3.3) se zaměřuje na studenty podle měst, ve kterých se vysoké školy nacházejí. Data jsou rozdělena do dvou hlavních kategorií – studenti s českým občanstvím a cizinci. Dále jsou prezentována data podle programu studia (bakalářský, magisterský, doktorský) a jsou uvedeni poprvé zapsaní a absolventi.

Velkou výhodou těchto dat je, že jsou úplná a neobsahují překlepy nebo chybějící záznamy, což velice usnadňuje práci.

## 3. Sběr a zpracování dat

Zřizovatel Město <sup>1)</sup>	Studenti se státním občanstvím ČR						Studující cizinci <sup>5)</sup>
	celkem	v programu studia			poprvé zapsaní <sup>3)</sup>	absol- venti	
		bakalář- ském	magister- ském <sup>2)</sup>	doktor- ském			
<b>VŠ celkem</b>	<b>194 692</b>	<b>41 619</b>	<b>140 400</b>	<b>15 303</b>	<b>-</b>	<b>-</b>	<b>8 763</b>
500496 Olomouc	12 248	2 000	9 317	1 004	2 479	1 838	669
505927 Opava	1 579	817	743	25	313	295	44
532053 Kladno	2	2	-	-	2	-	-
535419 Mladá Boleslav	130	130	-	-	58	-	2
541630 Vsetín	13	12	1	-	-	4	-
544256 České Budějovice	5 832	1 466	4 148	239	1 294	1 113	44
545058 Valašské Meziříčí	265	265	-	-	108	-	3
545881 Jindřichův Hradec	903	296	617	-	198	297	33
550744 Kunovice	537	537	-	-	244	85	36
552046 Tábor	241	-	241	-	61	5	4
554481 Cheb	628	541	87	-	267	156	10
554642 Mariánské Lázně	43	43	-	-	-	13	2
554782 Praha	74 198	11 232	56 557	7 485	13 927	10 691	4 677
554791 Plzeň	12 008	2 442	9 189	506	3 051	1 839	439
554804 Ústí nad Labem	5 575	2 061	3 530	16	1 424	923	22
554821 Ostrava	18 516	3 956	13 493	1 131	4 723	2 711	302
554961 Karlovy Vary	291	291	-	-	154	-	1
555134 Pardubice	4 605	1 912	2 395	304	1 229	648	45
562335 Děčín	213	213	-	-	79	17	1
563889 Liberec	5 915	1 398	4 420	260	1 286	704	344
567027 Most	492	254	239	-	158	55	-
568741 Albrechtůvky	-	-	-	-	-	-	1
569810 Hradec Králové	6 735	2 051	4 388	332	1 501	1 293	251
578347 Litomyšl	59	59	-	-	7	-	4
582786 Brno	39 735	7 256	29 032	3 849	8 879	5 843	1 672
584631 Lednice	478	117	323	38	35	133	15
585068 Zlín	3 229	2 209	895	133	1 214	583	108
589250 Prostějov	69	69	-	-	33	-	-
598917 Karviná	1 683	184	1 504	-	385	297	62
599191 Nový Jičín	22	22	-	-	22	-	-
<b>veřejné VŠ celkem</b>	<b>190 309</b>	<b>37 323</b>	<b>140 248</b>	<b>15 303</b>	<b>-</b>	<b>-</b>	<b>8 467</b>
500496 Olomouc	12 248	2 000	9 317	1 004	2 479	1 838	669
505927 Opava	1 579	817	743	25	313	295	44
541630 Vsetín	13	12	1	-	-	4	-
544256 České Budějovice	5 832	1 466	4 148	239	1 294	1 113	44
545058 Valašské Meziříčí	265	265	-	-	108	-	3
545881 Jindřichův Hradec	903	296	617	-	198	297	33
552046 Tábor	241	-	241	-	61	5	4
554481 Cheb	628	541	87	-	267	156	10
554642 Mariánské Lázně	43	43	-	-	-	13	2
554782 Praha	71 514	8 667	56 404	7 485	12 374	10 654	4 450
554791 Plzeň	12 008	2 442	9 189	506	3 051	1 839	439
554804 Ústí nad Labem	5 575	2 061	3 530	16	1 424	923	22
554821 Ostrava	18 108	3 546	13 493	1 131	4 515	2 711	276
555134 Pardubice	4 605	1 912	2 395	304	1 229	648	45
562335 Děčín	213	213	-	-	79	17	1
563889 Liberec	5 915	1 398	4 420	260	1 286	704	344

Obrázek 3.3: Studenti podle měst.



### 3.3.2 Statistický úřad

#### Pohyb obyvatel České republiky v letech 1785 až 2022

Soubor vyobrazený na obrázku 3.4 obsahuje informace o pohybu obyvatel České republiky od roku 1785 do roku 2022. Tento soubor 3.4 zahrnuje údaje o sňatcích, rozvodech, narozeních, úmrtích a přistěhovaných a vystěhovaných osobách. Naše analýza se však zaměří pouze na údaje o narozených osobách, které mají klíčový význam pro naši studii demografických trendů ve vzdělávacím sektoru.

Pohyb obyvatel České republiky <sup>1)</sup> v letech 1785–2022: absolutní údaje													
Population change of the Czech Republic <sup>1)</sup> in 1785–2022: numbers													
Rok <sup>2)</sup>	Sňatky	Rozvody	Narození Births				Potraty	Abortions	Zemřelí Deaths			Přistěhovaní <sup>5)</sup>	Vystěhovaní <sup>5)</sup>
			celkem	živě	Live births	mrtvě			celkem <sup>4)</sup>	do 1 roku	do 28 dnů		
Year <sup>2)</sup>	Marriages	Divorces	Births	Total	Outside marriage <sup>3)</sup>	Still-births	Total	umělá přerušeni těhotenství Legally induced abortions	Total <sup>4)</sup>	Infant deaths	Neonatal deaths	Immigrants <sup>5)</sup>	Emigrants <sup>5)</sup>
1977	93011	25442	182865	181763	8351	1102	81656	61114	126214	3407	2472	11408	10101
1978	90338	27071	180018	178901	8399	1117	83915	63904	127136	3053	2260	11583	9519
1979	84496	26191	173084	172112	8659	972	83624	64505	127949	2726	1945	11336	8842
1980	78343	27218	154665	153801	8640	864	86503	68930	135537	2592	1735	11401	9545
1981	77453	27608	145186	144438	8443	748	89373	71574	130407	2226	1601	10851	9134
1982	76978	27821	142518	141738	9036	780	91531	74531	130765	2130	1466	11076	9328
1983	80417	29319	138132	137431	9327	701	92033	75037	134474	1997	1368	10745	8362
1984	81714	30514	137587	136941	9918	646	96638	79534	132188	1932	1365	10699	8078
1985	80653	30489	136488	135881	9894	607	99357	83042	131641	1694	1167	9918	7723
1986	81638	29560	133942	133356	9892	586	99452	83564	132585	1639	1121	10712	7699
1987	83773	31036	131469	130921	9466	548	126690	109626	127244	1577	1094	9934	7213
1988	81458	30652	133238	132667	10014	571	129349	113730	125694	1463	1003	9984	7440
1989	81262	31376	128881	128356	10141	525	126507	111683	127747	1280	886	9400	7941
1990	90953	32055	131094	130564	11167	530	126055	111268	129166	1410	1003	12411	11787
1991	71973	29366	129850	129354	12703	496	120050	106042	124290	1343	902	14096	11220
1992	74060	28572	122142	121705	13008	437	109281	94180	120337	1204	749	19072	7291
1993	66033	30227	121470	121025	15323	445	85445	70634	118185	1028	692	12900	7424
1994	58440	30939	106915	106579	15507	336	67434	54836	117373	847	505	10207	265
1995	54956	31135	96397	96097	14947	300	61590	49531	117913	740	475	10540	541
1996	53896	33113	90763	90446	15288	317	59962	48086	112782	547	347	10857	728
1997	57804	32465	90930	90657	16125	273	56973	45022	112744	531	326	12880	805
1998	55027	32363	90829	90535	17209	294	55654	42959	109527	472	289	10729	1241
1999	53523	23657	89774	89471	18426	303	52103	39382	109768	413	261	9910	1136
2000	55321	29704	91169	90910	19792	259	47370	34623	109001	373	231	7802	1263
2001	52374	31586	90978	90715	21276	263	45057	32528	107755	360	212	12918	21469
2002	52732	31758	93047	92786	23459	261	43743	31142	108243	385	251	44679	32389
2003	48943	32824	93957	93685	26713	272	42304	29298	111288	365	221	60015	34226
2004	51447	33060	97929	97664	29839	265	41324	27574	107177	366	224	53453	34818
2005	51829	31288	102498	102211	32409	287	40023	26453	107938	347	206	60294	24065
2006	52860	31415	106130	105831	35259	299	39959	25352	104441	352	246	68183	33463
2007	57157	31129	114947	114632	39537	315	40917	25414	104636	360	235	104445	20500
2008	52457	31300	119842	119570	43457	272	41446	25760	104948	338	217	77817	6027
2009	47862	29133	118667	118348	45954	319	40528	24636	107421	341	194	39973	11629
2010	46746	30783	117446	117153	47164	293	39273	23998	106844	313	196	30515	14867
2011	45137	28113	108990	108673	45421	317	38864	24055	106848	298	186	22590	5701
2012	45206	26402	108955	108576	47088	379	37733	23032	108189	285	175	30298	20005
2013	43499	27895	107117	106751	48000	366	37687	22714	109160	265	151	29579	30876
2014	45575	26764	110252	109860	51267	392	36956	21893	105665	263	172	41625	19964
2015	48191	26083	111162	110764	52976	398	35761	20403	111173	272	165	34922	18945
2016	50768	24996	113083	112663	54733	420	35921	20406	107750	317	192	37503	17439
2017	52567	25755	114789	114405	56091	384	35012	19415	111443	304	203	45957	17684
2018	54470	24313	114419	114036	55338	383	32952	18298	112920	292	187	58148	19519
2019	54870	24141	112633	112231	54093	402	31797	17757	112362	288	175	65571	21301
2020	45415	21734	110631	110200	53408	431	30368	16886	129289	249	172	55661	28734
2021	46778	21107	112197	111793	54203	404	27959	15492	139891	246	152	69201	19232
2022	54820	19846	101676	101299	48872	377	27598	16438	120219	230	125	349548	19806

Obrázek 3.4: Pohyb obyvatel České republiky v letech 1785 až 2022.

## 3.4 Zpracování dat

Za účelem vytvoření analýzy a vizualizací jsme vyvinuli pythonové skripty pro každou excelovskou tabulku. To bylo nezbytné, protože data v těchto tabulkách nemají stejnou strukturu a nejsou jednoduše zpracovatelná pro analýzu pomocí Pythonu, nástrojů Power BI a následně ani pro naši aplikaci. V předchozí kapitole jsme viděli, že každá tabulka obsahuje specifické informace a formát, který vyžaduje individuální zpracování.

Pro zpracování dat jsme využili knihovny Pandas<sup>3</sup> a Openpyxl<sup>4</sup>. Pandas je výkonná knihovna pro analýzu a manipulaci s daty, zatímco Openpyxl umožňuje čtení a zápis dat do a z excelovských souborů.

**Výběr Pythonu:** Python byl vybrán pro svou robustnost a širokou škálu knihoven, které umožňují efektivní zpracování a analýzu dat. Knihovny jako Pandas a Openpyxl jsou velmi flexibilní a umožňují detailní manipulaci s daty, což je klíčové pro tento projekt. Power BI byl zvolen pro vizualizaci díky své schopnosti snadno integrovat data a vytvářet interaktivní dashboardy. Naše aplikace pak umožní specifické analýzy přizpůsobené konkrétním potřebám, které nejsou snadno dostupné v jiných nástrojích.

Nyní se podíváme na jednotlivé tabulky, jak jsme je převáděli pomocí skriptů, a uvidíme také výstupy těchto skriptů.

### 3.4.1 F11 – Souhrn studentů a absolventů podle formy a typu studijního programu

Před zpracováním dat z excelovských tabulek je třeba je efektivně načíst a upravit pro další analýzu. Níže uvedený Python skript vyobrazený na obrázku s ukázkou kódu 3.1 slouží k načítání a zpracování těchto tabulek pomocí knihoven Pandas a Openpyxl. Tento kód načítá data z excelovského souboru `f11.xlsx`, který obsahuje informace o souhrnu studentů a absolventů podle formy a typu studijního programu. Skript načítá relevantní sloupce a pojmenovává je pomocí seznamu `columns`.

Prochází všechny listy v excelovském souboru a načítá pouze první řádek z každého listu, který obsahuje potřebná data. Přidává sloupec s názvem roku a ukládá tyto tabulky do slovníku `dfs`. Nakonec všechny tabulky spojí do jednoho DataFrame `final_df` a uloží je do nových excelovských souborů s názvy odpovídajícími jejich obsahu pomocí funkce `to_excel`. Tento kód a ani žádný další ze skriptů pro zpracování dat neobsahuje kontrolu na chybějící data ani kontrolu na to, zda uživatel má otevřený excelovský soubor.

<sup>3</sup>[https://pandas.pydata.org/docs/user\\_guide/index.html](https://pandas.pydata.org/docs/user_guide/index.html)

<sup>4</sup><https://openpyxl.readthedocs.io/en/stable/tutorial.html>

```

1 import pandas as pd
2
3 def process_data(excel_file_path, start_row,
4                 output_file_name):
5     # Table setup
6     columns = ['Celkem', 'Celkem P', 'Bakalarske P', '
7               Magisterske P', 'Navazujici Magisterske P', 'Doktorske
8               P', 'Celkem K', 'Bakalarske K', 'Magisterske K', '
9               Navazujici magisterske K', 'Doktorske K']
10
11    # Read data from each list for each row
12    dfs = pd.read_excel(excel_file_path, sheet_name=None,
13                       header=None, skiprows=start_row-1, usecols='G:Q')
14
15    # For each list get only the needed line and add the
16    year
17    for sheet_name, df in dfs.items():
18        df = df.iloc[0, :].to_frame().T
19        df.columns = columns
20        df['Rok'] = sheet_name
21        dfs[sheet_name] = df
22
23    # Concatenate all lists into one DataFrame
24    final_df = pd.concat(dfs.values(), ignore_index=True)
25
26    # Save the DataFrame to an Excel file
27    output_file = f'{output_file_name}.xlsx'
28    final_df.to_excel(output_file, sheet_name=
29                      output_file_name, index=False)
30
31 # Define the rows and output file names
32 rows_and_outputs = [(10, 'poprve_zapsani'), (12, '
33                     celkem_studenti'), (19, 'celkem_studenti_cr'), (26, '
34                     celkem_studenti_ciz'), (33, 'celkem_studenti_verejne')
35                     , (40, 'celkem_studenti_soukrome')]
36
37 # Process the data
38 for start_row, output_name in rows_and_outputs:
39     process_data('f11.xlsx', start_row, output_name)
40     print(f"Data saved successfully for {output_name}.")

```

Ukázka kódu 3.1: Python skript pro zpracování dat

Výsledkem je šest nových excelovských souborů, které obsahují zpracovaná data ve formátu vhodném pro další analýzu a vizualizace. Tabulky jsou:

- **poprve\_zapsani:** Obsahuje informace o nově zapsaných studentech.
- **celkem\_studenti:** Poskytuje údaje o celkovém počtu studentů.
- **celkem\_studenti\_cr:** Zahrnuje informace o celkovém počtu studentů z České republiky.

- **celkem\_studenti\_ciz:** Obsahuje informace o celkovém počtu zahraničních studentů.
- **celkem\_studenti\_verejne:** Poskytuje rozdělení studentů ve veřejných školách.
- **celkem\_studenti\_soukrome:** Obsahuje rozdělení studentů ve soukromých školách.

Všechny tyto tabulky mají stejnou strukturu a obsahují data zpracovaná podle definovaných kritérií, která jsou popsána v kódu. Obrázek 3.5 zobrazuje ukázkou tabulky se souhrnným počtem studentů.

Celkem	Celkem P	Bakalářské P	Magisterské P	Navazující magisterské P	Doktorské P	Celkem K	Bakalářské K	Magisterské K	Navazující magisterské K	Doktorské K	Rok
203448	168125	29030	124609		9955	6726	36639	14087	10813	2113	9796 2001
220177	179380	41821	122485		10178	7673	42336	19168	10176	2711	10474 2002
243717	195591	64465	114589		10583	9105	49920	25977	9513	3676	11016 2003
264769	207988	89855	99083		11800	10018	58842	33863	8356	5481	11474 2004
289463	223152	112822	86445		16460	10105	68681	41939	7092	7632	12276 2005
316177	238180	132778	73488		24629	9980	80772	50128	6042	11469	13381 2006
343943	251915	149150	60173		35351	9969	95340	60121	5052	16382	14057 2007
368050	263903	161171	49292		45942	10508	107980	69423	4096	20732	14043 2008
388992	277043	170884	42937		54648	11594	116289	74364	3156	25165	13951 2009
395984	283517	176313	38096		59440	12503	116744	73516	2656	27412	13462 2010
392039	284155	176340	35263		62468	12800	111978	69292	2227	27826	12904 2011
380891	281660	174803	32454		64011	12669	102757	61892	2053	26884	12175 2012
367768	277074	170535	31483		64182	12935	93953	55288	1713	25292	11835 2013
346799	263530	159144	30649		62710	12868	86142	49166	1924	23868	11381 2014
326423	249264	147922	30045		60237	12665	79806	45382	1734	21645	11238 2015
311045	236683	138365	29516		57999	12307	76838	42328	1701	22072	10915 2016
298663	227389	132748	29135		54549	12268	73491	40206	1674	21859	9904 2017
289649	222277	130653	29004		51239	12569	69313	37987	1727	20947	8815 2018
288577	224667	134336	29200		49096	13197	65904	36399	1663	20266	7711 2019
298986	235512	142342	30117		49729	14631	65701	37018	1697	20190	7017 2020
303633	241550	146399	30788		50696	14974	64287	36571	1724	19771	6403 2021
304518	246332	150515	31431		50911	14794	60276	34235	1807	18630	5748 2022

Obrázek 3.5: Tabulka se souhrnnými statistikami o studentech.

### 3.4.2 F12 – Souhrn škol, fakult, studentů a absolventů

Pro zpracování dat ve formátu Excel, který obsahuje informace o školách, fakultách, studentech a absolventech, bylo nejprve nutné řešit složitosti spojené s označováním názvů univerzit a jejich přiřazováním k jednotlivým fakultám. K tomuto účelu jsem využil kontrolu textu tučným písmem, abych správně identifikoval názvy univerzit. Dalším problémem byla korekce dat poskytnutých ministerstvem, kde byla do roku 2019 1. lékařská fakulta nesprávně označena tučným písmem.

Ukázku kódu pro tento konkrétní případ zde již nebudu uvádět, jelikož se jedná o podobný princip jako u předchozích částí. Všechny kódy použité pro zpracování dat jsou k dispozici v příloze této práce.

Výsledkem provedeného kódu je jeden excelový soubor obsahující všechny školy ve všech letech (viz obrázek 3.6). Dále jsou vytvořeny samostatné excelové tabulky pouze s univerzitami v jednotlivých letech. Na fakulty jsem se v tomto procesu nezaměřil, protože by jejich správné strukturování bylo složité a nebylo by je možné jednoduše využít dále v aplikaci. Dále jsou vytvořeny dvě souhrnné tabulky, které obsahují informace o veřejných školách a druhá o soukromých.

Strukturování fakult bylo složité z několika důvodů:

- **Variabilita v pojmenování:** Názvy fakult mohou být různé a nekonzistentní napříč roky a institucemi, což komplikuje jejich jednotné zpracování a kategorizaci.
- **Hierarchická struktura:** Fakulty mají různé podúrovně a specifické organizační struktury, které by vyžadovaly složitější algoritmy pro správnou identifikaci a třídění.
- **Formát dat:** Data o fakultách jsou často prezentována v různých formátech a úrovních detailu, což ztěžuje jejich automatizované zpracování a začlenění do jednotné databáze.
- **Rozdělení podle oborů a zaměření:** Zařazení fakult podle jejich oborů a zaměření by vyžadovalo podrobnou analýzu webových stránek škol a jejich zaměření, což by bylo časově náročné a komplikované pro hlubší využití.

Kvůli těmto výzvám jsem se rozhodl soustředit se na univerzity jako celek, což umožňuje jednodušší a efektivnější zpracování dat pro potřeby této práce a budoucí aplikace.

Škola	Celkem	Celkem ženy	Studující celkem	ČR	první zapsaní	Absolventi	ČR	studující celkem	první zapsaní	solventi	Přerušené studium	Jinoplací	Samoplací	Rok
Univerzita Karlova v Praze	36709	20997	33540	4750	4182	3169	802	248	1307	17	748	2001		
Jihočeská univerzita v Českých Budějovicích	6119	3847	6071	1354	1118	48	7	5	259	0	0	2001		
Univerzita J. E. Purkyně v Ústí nad Labem	5597	3653	5575	1424	923	22	13	2	163	0	0	2001		
Masarykova univerzita	19771	11260	18774	3752	2737	997	316	53	771	0	165	2001		
Univerzita Palackého v Olomouci	12917	8468	12248	2479	1838	669	149	39	330	0	71	2001		
Veterinární a farmaceutická univerzita Brno	1710	1290	1570	271	206	140	35	5	42	0	0	2001		
Ostravská univerzita v Ostravě	5408	3654	5352	1322	1005	56	14	6	117	0	0	2001		
Univerzita Hradec Králové	4509	2732	4494	1122	951	15	5	5	141	0	2	2001		
Slezská univerzita v Opavě	3367	1928	3261	698	592	106	43	3	83	0	0	2001		
České vysoké učení technické v Praze	19243	3060	18705	4175	1963	538	162	24	790	0	34	2001		
Vysoká škola chemicko-technologická v Praze	2759	1457	2687	599	455	72	19	10	67	0	6	2001		
Západočeská univerzita v Plzni	11610	5300	11405	3118	1853	205	21	7	217	3	0	2001		
Technická univerzita v Liberci	6328	3214	5984	1319	704	344	129	18	185	0	37	2001		
Univerzita Pardubice	4650	2213	4605	1229	648	45	17	2	90	0	2	2001		
Vysoké učení technické v Brně	14208	2600	13832	3172	2072	376	156	28	223	0	48	2001		
Vysoká škola báňská - Technická univerzita Ostrava	13914	4981	13689	3493	1804	225	93	7	322	2	21	2001		
Univerzita Tomáše Bati ve Zlíně	3337	1648	3229	1214	583	108	53	2	74	0	0	2001		
Vysoká škola ekonomická v Praze	13210	6833	12257	1771	3255	954	375	48	458	0	2	2001		
Česká zemědělská univerzita v Praze	8158	4102	8055	1763	1188	103	38	11	218	26	8	2001		
Mendelova zemědělská a lesnická univerzita v Brně	5670	2843	5549	1361	864	121	52	6	322	0	1	2001		
Akademie múzických umění v Praze	1196	562	1079	135	315	117	27	28	63	1	18	2001		
Akademie výtvarných umění v Praze	283	107	261	31	30	22	7	1	5	0	0	2001		
Vysoká škola umělecko-průmyslová v Praze	387	169	373	41	58	14	6	3	11	0	0	2001		
Janáčkova akademie múzických umění v Brně	517	290	462	85	99	55	17	6	35	0	0	2001		
Bankovní institut vysoká škola, a.s.	388	173	369	138	0	19	11	0	0	0	0	2001		
Evropský polytechnický institut, s.r.o.	573	286	537	244	85	36	23	4	3	0	0	2001		
Vysoká škola hotelová v Praze, s.r.o.	608	331	579	250	37	29	27	0	0	0	0	2001		
Vysoká škola finanční a správní, o.p.s.	754	377	733	420	0	21	9	0	17	0	0	2001		
Vysoká škola Karlovy Vary, o.p.s.	292	147	291	154	0	1	1	0	0	0	0	2001		
ŠKODA AUTO VŠ, a.s.	132	42	130	58	0	2	1	0	3	0	0	2001		
Literární akademie, s.r.o.	197	117	194	98	0	3	2	0	1	0	0	2001		
Vysoká škola podnikání, a.s.	482	249	456	249	0	26	21	0	3	0	0	2001		

Obrázek 3.6: Tabulka se souhrnnými statistikami o univerzitách.

### 3.4.3 F13 – Studenti podle měst

Pro zpracování dat týkajících se studentů podle měst byl vytvořen skript, který využívá kontrolu tučného písma k identifikaci jednotlivých škol a následně je rozděluje do tří výsledných souborů. Tyto soubory obsahují informace o studentech z vysokých škol v daném městě, a to buď celkový počet studentů

ze všech vysokých škol, pouze z veřejných vysokých škol nebo pouze ze soukromých vysokých škol.

Jeden výsledný soubor obsahuje informace o studentech ze všech vysokých škol v daném městě (viz obrázek 3.7), druhý soubor zahrnuje pouze studenty z veřejných vysokých škol a třetí soubor obsahuje data o studentech ze soukromých vysokých škol.

Město	Celkem	Bakalářské	Magisterské	Doktorské	Poprvé zapsaní	Absolventi	Cizinci	Rok
Olomouc	12248	2000	9317	1004	2479	1838	669	2001
Opava	1579	817	743	25	313	295	44	2001
Kladno	2	2	0	0	2	0	0	2001
Mladá Boleslav	130	130	0	0	58	0	2	2001
Vsetín	13	12	1	0	0	4	0	2001
České Budějovice	5832	1466	4148	239	1294	1113	44	2001
Valašské Meziříčí	265	265	0	0	108	0	3	2001
Jindřichův Hradec	903	296	617	0	198	297	33	2001
Kunovice	537	537	0	0	244	85	36	2001
Tábor	241	0	241	0	61	5	4	2001
Cheb	628	541	87	0	267	156	10	2001
Mariánské Lázně	43	43	0	0	0	13	2	2001
Praha	74198	11232	56557	7485	13927	10691	4677	2001
Plzeň	12008	2442	9189	506	3051	1839	439	2001
Ústí nad Labem	5575	2061	3530	16	1424	923	22	2001
Ostrava	18516	3956	13493	1131	4723	2711	302	2001
Karlovy Vary	291	291	0	0	154	0	1	2001
Pardubice	4605	1912	2395	304	1229	648	45	2001
Děčín	213	213	0	0	79	17	1	2001
Liberec	5915	1398	4420	260	1286	704	344	2001
Most	492	254	239	0	158	55	0	2001
Albrechtův Týn	0	0	0	0	0	0	1	2001
Hradec Králové	6735	2051	4388	332	1501	1293	251	2001
Litomyšl	59	59	0	0	7	0	4	2001
Brno	39735	7256	29032	3849	8879	5843	1672	2001
Lednice	478	117	323	38	35	133	15	2001
Zlín	3229	2209	895	133	1214	583	108	2001
Prostějov	69	69	0	0	33	0	0	2001
Karviná	1683	184	1504	0	385	297	62	2001
Nový Jičín	22	22	0	0	22	0	0	2001

**Obrázek 3.7:** Tabulka se souhrnnými statistikami o studentech v jednotlivých městech.

### 3.4.4 Pohyb obyvatel České republiky v letech 1785 až 2022

Pro vytvoření poslední tabulky jsem nejprve použil skript, který extrahuje data o počtu narozených dětí v každém roce. Tyto extrahované údaje jsem následně využil k vytvoření nového excelového souboru, ve kterém odhaduji teoreticky možný počet nastupujících studentů pro roky 2001 až 2022.

Pro výpočet tohoto odhadu jsem použil průměr z počtu narozených dětí mezi 18. a 20. rokem věku, což mi umožňuje odhadnout potenciální počet studentů pro rok, ve kterém by měli nastoupit do studia. Tento interval je zvolen, protože většina studentů dokončuje střední školu a nastupuje na vysokou školu právě v tomto věkovém rozmezí. Samozřejmě se do tohoto odhadu nepočítají faktory jako opakování ročníků nebo pozdější nástup na vysokou školu, což činí tento odhad hrubým, ale dostatečně přesným pro

základní analýzu. Například pro rok 2020 byl vypočítán potenciální počet studentů na základě narozených dětí v letech 2000, 2001 a 2002. Ukázka 3.2 zobrazuje část kódu, který je zodpovědný za výpočet průměrného počtu možných studentů.

```
1 # Only want years between 2001 and 2022 for my analysis
2     years_of_study = range(2001, 2023)
3
4     potential_students = []
5     for year in years_of_study:
6         # to get the relevant years of birth - use range
7         # of 18, 19 and 20 years old
8         relevant_birth_years = range(year - 20, year -
9         17)
10        relevant_births = birth_data[birth_data['Rok'].
11        isin(relevant_birth_years)]
12
13        # Calculate the average number and round it
14        avg_births = round(relevant_births['Pocet
15        narozenych'].mean() if not relevant_births.empty else
16        None)
17        potential_students.append({'Rok': year, '
18        Potencialni studenti': avg_births})
```

**Ukázka kódu 3.2:** Python skript pro zpracování dat



Výsledek této části můžete vidět na obrázku 3.8.

Rok	Potenciální studenti
2001	141202
2002	138703
2003	136751
2004	135393
2005	133386
2006	132315
2007	130648
2008	130529
2009	129425
2010	127208
2011	124028
2012	116436
2013	107900
2014	97707
2015	92400
2016	90546
2017	90221
2018	90305
2019	90365
2020	91470
2021	92395
2022	94712

**Obrázek 3.8:** Tabulka s odhadem počtu možných studentů v jednotlivých letech.

## 3.5 Shrnutí kapitoly 3

V této kapitole jsme podrobně popsali procesy sběru a zpracování dat, které jsou klíčové pro analýzu a vizualizaci v rámci našeho projektu. Definovali jsme klíčové pojmy a struktury dat, které jsou nezbytné pro správnou interpretaci informací. Prozkoumali jsme strukturu a obsah tabulek poskytnutých Ministerstvem školství, mládeže a tělovýchovy (MŠMT) a statistickým úřadem, a následně jsme se věnovali zpracování těchto dat pomocí Python skriptů. Tyto kroky nám poskytly kvalitní datovou sadu připravenou pro další analýzu.



## Kapitola 4

### Analýza dat a formulace otázek a hypotéz

V této kapitole se zaměříme na formulaci hypotéz a jejich následnou analýzu na základě dostupných dat. Naším cílem je potvrdit nebo vyvrátit tyto hypotézy pomocí datové analýzy. Všechny grafy a obrázky uvedené v této kapitole byly generovány pomocí Power BI, tento nástroj jsme volili hlavně díky předešlé zkušenosti s ním. V kapitole 7 jsou pak uvedeny obrázky grafů a tabulek vygenerovaných pomocí naší aplikace a popsány případné rozdíly.

#### 4.1 Přístupy k formulaci hypotéz

Existují dva hlavní přístupy k formulaci hypotéz:

- **Hypotézy před daty:** Tento přístup začíná formulací hypotéz na základě teoretických poznatků nebo předchozího výzkumu. Následně se hledají data, která by tyto hypotézy mohla ověřit. Výhodou tohoto přístupu je, že hypotézy jsou často dobře podloženy teorií a existujícím výzkumem, což může zvýšit jejich relevanci a význam. Nevýhodou je, že může být obtížné najít data, která přesně odpovídají formulovaným hypotézám.
- **Hypotézy po datech:** Tento přístup začíná analýzou dostupných dat bez předem stanovených hypotéz. Hypotézy jsou formulovány na základě vzorců a vztahů, které se objeví v datech. Výhodou tohoto přístupu je flexibilita a schopnost objevit nové, nečekané vzorce v datech. Nevýhodou je, že hypotézy mohou být méně teoreticky podloženy a mohou vyžadovat dodatečné testování a ověřování.

##### 4.1.1 Náš přístup

Pro náš projekt jsme se rozhodli použít kombinaci obou přístupů. Na základě teoretických znalostí a předchozího výzkumu jsme formulovali některé hypotézy, které jsme následně ověřovali dostupnými daty. Současně jsme prováděli exploratorní analýzu dat, abychom identifikovali nové vzorce a vztahy, které jsme následně použili k formulaci dalších hypotéz. Tento hybridní přístup nám umožnil efektivně využít dostupné teoretické poznatky a zároveň plně využít potenciál našich dat. Některé hypotézy byly formulovány na základě

předchozích zkušeností a očekávání, zatímco jiné vznikly až během analýzy dat, kdy jsme objevili nové souvislosti a trendy.

## 4.2 Formulace otázek a hypotéz

Nejdříve se zaměříme na formulaci otázek a hypotéz, které budou sloužit jako výchozí bod pro naši analýzu. Tyto příklady vycházejí z teoretických znalostí a předchozího výzkumu, stejně jako z exploratorní analýzy dat. S ohledem na skutečnost, že je to i úvod do analýzy, jsou hypotézy jednodušší:

- **Trendy ve vývoji počtu studentů v čase**  
**Hypotéza:** Počet zapsaných studentů vysokých škol se v průběhu času zvyšuje.
- **Vztah mezi národností studenta a typem studia**  
**Hypotéza:** Zahraniční studenti častěji volí prezenční formu studia než čeští studenti.
- **Geografické rozložení studentů**  
**Hypotéza:** Nejvíce studentů je soustředěno ve velkých městech jako Praha a Brno.
- **Geografické rozložení zahraničních studentů v České republice**  
**Hypotéza:** Jaké jsou změny v geografickém rozložení zahraničních studentů mezi lety 2001 a 2022?
- **Rozdíl mezi soukromými a veřejnými školami**  
**Hypotéza:** Veřejné vysoké školy mají vyšší počet studentů a vyšší úspěšnost než soukromé vysoké školy.

## 4.3 Struktura kapitol s hypotézami

Každá kapitola zaměřená na konkrétní hypotézu je strukturována následovně:

- **Formulace otázky:** Stručný popis zkoumané otázky a formulace hypotézy.
- **Úvod do problematiky:** Kontext a význam zkoumané problematiky, včetně teoretických východisek a praktických důsledků.
- **Analýza dat:** Podrobná analýza dat souvisejících s hypotézou, zahrnující grafy, tabulky a popisy výsledků.
- **Interpretace výsledků:** Diskuze nad zjištěnými výsledky, možné příčiny a souvislosti, které mohou ovlivnit dané výsledky.
- **Shrnutí a závěr:** Krátké shrnutí hlavních zjištění a jejich význam pro potvrzení nebo vyvrácení hypotézy.

Tato struktura umožňuje systematický přístup k analýze každé hypotézy a poskytuje čtenáři jasný a přehledný způsob, jak sledovat a pochopit postup analýzy a dosažené výsledky.

## 4.4 Hypotéza 1: Trendy ve vývoji počtu studentů v čase

**Formulace otázky:** Jaký je trend ve vývoji počtu studentů vysokých škol v České republice v průběhu času?

### 4.4.1 Úvod do problematiky

Pochopení trendů ve vývoji počtu studentů na vysokých školách je klíčové pro mnoho zainteresovaných stran. Pro studenty je důležité vědět, jaké jsou jejich šance na přijetí a kolik dalších lidí mohou očekávat jako své spolužáky či budoucí konkurenty na trhu práce. Pro vysoké školy tato statistika umožňuje plánovat kapacity a rozvoj infrastruktury, což je klíčové pro zajištění kvalitního vzdělávání. Kromě toho mohou vlády a vzdělávací instituce využívat tyto údaje k vytváření politik a strategií, které podporují rozvoj vzdělávacího systému.

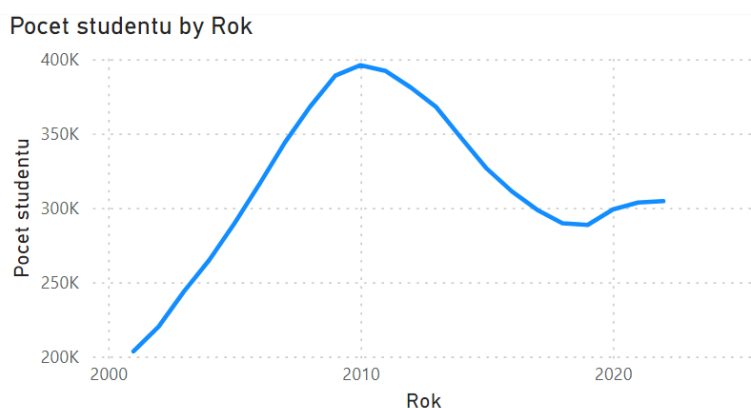
### 4.4.2 Analýza dat

Pro analýzu dat byly použity soubory z Ministerstva školství, mládeže a tělovýchovy (MŠMT), konkrétně tabulky F1 Souhrny, tabulky s celkovými počty studentů a odhady potenciálních studentů vycházející z tabulky o počtu obyvatel ze statistického úřadu. Níže jsou uvedeny jednotlivé kroky analýzy.

Data jsme nahráli do Power BI a následně začali analyzovat.

### Vývoj počtu studentů v letech 2001-2022

Na obrázku 4.1 je zobrazen počet studentů vysokých škol v letech 2001 až 2022. Tento graf ukazuje, že počet studentů rostl do roku 2010, kdy dosáhl vrcholu, a následně začal klesat. Od roku 2018 můžeme pozorovat mírný nárůst, který pokračuje až do roku 2022.

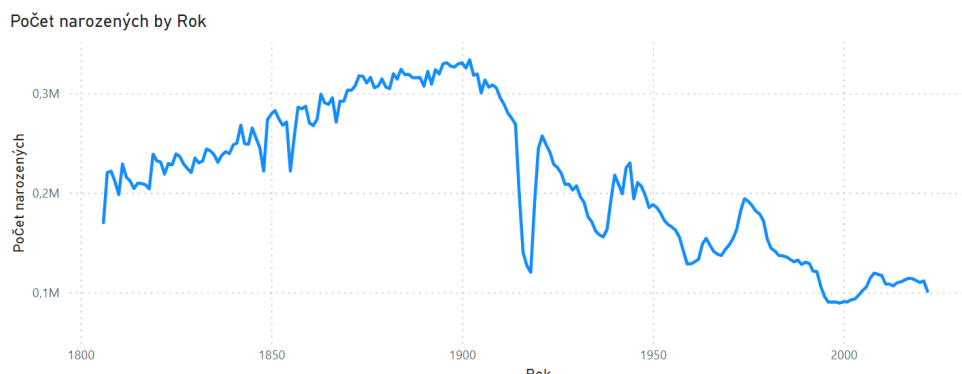


**Obrázek 4.1:** Vývoj počtu studentů vysokých škol v letech 2001-2022

Použili jsme spojnicový graf k vizualizaci těchto dat, což nám umožnilo snadno identifikovat trendy a změny v počtu studentů v průběhu času. Spojnicový graf je vhodný pro zobrazení časových řad, protože jasně ukazuje vývoj a umožňuje sledovat růstové a klesající fáze v datech.

#### ■ Demografické trendy

Pro analýzu demografických trendů jsme použili spojnicový graf, který zobrazuje počet narozených dětí v daných letech. Tento graf poskytuje kontext pro predikci možného počtu potenciálních studentů.



**Obrázek 4.2:** Počet narozených dětí v daných letech

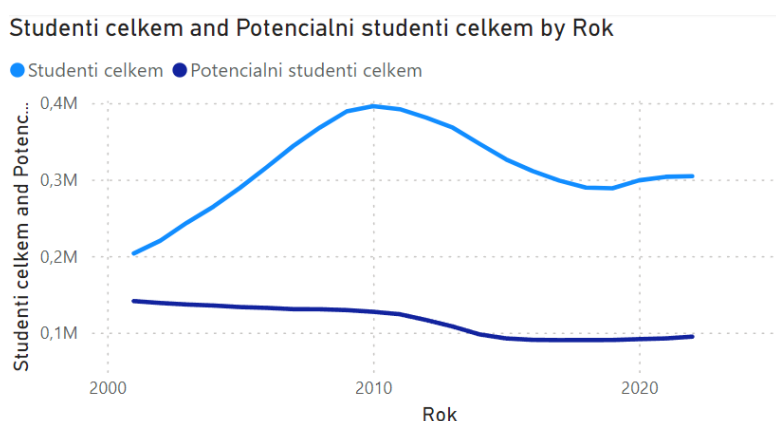
Na obrázku 4.2 je patrný nárůst počtu narozených dětí v 70. letech, známý jako „Husákovy děti“. Tento baby boom byl výsledkem pronatalitní politiky prezidenta Gustáva Husáka, která vedla k výraznému zvýšení porodnosti. „Husákovy děti“ je termín používaný k označení generace dětí narozených v době tzv. baby boomu v 70. letech, kdy československý prezident Gustáv Husák prosadil pronatalitní politiku<sup>1</sup>, známou také jako pro-populační politika, která vedla k výraznému zvýšení porodnosti.[26] Graf také ukazuje výrazné propady během válečných období, kdy se rodilo méně dětí.

<sup>1</sup>Soubor opatření a strategií zaměřených na podporu vyšší porodnosti a růstu populace

Zkoumali jsme tento graf a hledali zajímavé vzorce a události, které ovlivnily počet narozených dětí. Identifikovali jsme významné nárůsty a poklesy a analyzovali jsme důvody, proč k těmto změnám došlo. Tato analýza nám pomohla lépe pochopit demografické faktory, které mohou ovlivnit počet budoucích studentů na vysokých školách.

### ■ Porovnání celkového počtu studentů s odhadem potenciálních nových studentů

Na obrázku 4.3 je porovnán celkový počet studentů s odhadem potenciálních nových studentů, založeným na demografických datech. Tento odhad počítá pouze s možným počtem nových studentů v daném roce a nezahrnuje již studující studenty, což může vést k podhodnocení celkového počtu.



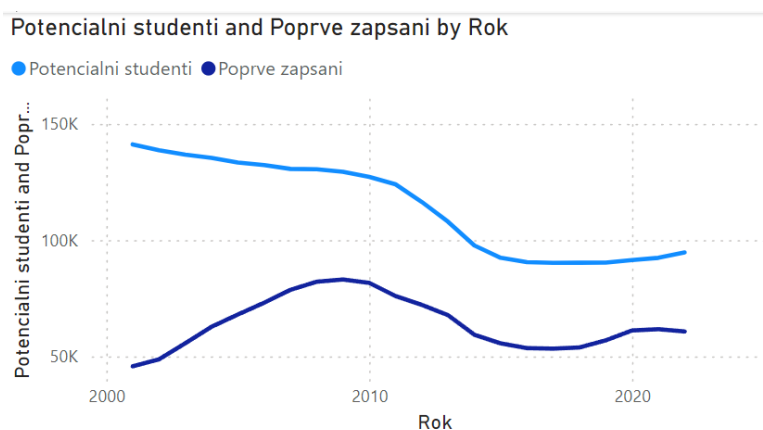
**Obrázek 4.3:** Celkový počet studentů vs. potenciální noví studenti

**Poznámka:** Tato data jsou zaměřena pouze na nově zapsané studenty, což zkrsluje reálný pohled. Pro přesnější analýzu by bylo vhodné vytvořit model, který zahrnuje jak nově zapsané studenty, tak i studenty, kteří již studují.

Pro graf 4.3 jsme opět použili spojnicový graf, který nám umožnil vizualizovat a porovnat trend celkového počtu studentů s odhadem potenciálních nových studentů v daném roce. Graf ukazuje, že odhadovaný počet nových studentů klesal, zatímco celkový počet studentů vykazoval růst do roku 2010 a poté klesal až do roku 2018, kdy opět začal růst.

### ■ Porovnání potenciálních nových studentů s počtem nově zapsaných studentů

Na obrázku 4.4 je zobrazen počet potenciálních nových studentů ve srovnání s počtem studentů, kteří se poprvé zapsali do vysoké školy. Je patrné, že počet nově zapsaných studentů se v průběhu let mění, což může být ovlivněno demografickými změnami a dalšími faktory.



**Obrázek 4.4:** Potenciální noví studenti vs. nově zapsaní studenti

Tento graf 4.4 zobrazuje vztah mezi počtem potenciálních nových studentů a skutečným počtem nově zapsaných studentů, což nám umožňuje lépe pochopit, jak demografické změny ovlivňují skutečný počet nových studentů.

**Poznámka:** Tento graf zobrazuje vztah mezi počtem potenciálních nových studentů a skutečným počtem nově zapsaných studentů, což nám umožňuje lépe pochopit, jak demografické změny ovlivňují skutečný počet nových studentů. Pro tvorbu tohoto grafu jsme použili spojnicový graf k vizualizaci dat z obou datasetů na jedné ose, což umožňuje snadné porovnání a identifikaci trendů v průběhu času.

**Návrh přesnějšiho modelu.** Pro přesnější analýzu by model měl zahrnovat následující faktory:

- Počet nově zapsaných studentů.
- Počet studentů, kteří pokračují ve studiu z předchozích let.
- Počet studentů, kteří ukončili studium (absolventi nebo přerušeni studia).
- Demografické trendy a předpokládané změny v počtu mladých lidí ve věku typickém pro zahájení vysokoškolského studia.

Takový model by umožnil přesnější predikci celkového počtu studentů v budoucnosti a poskytl by lepší základ pro strategické plánování a rozhodování.

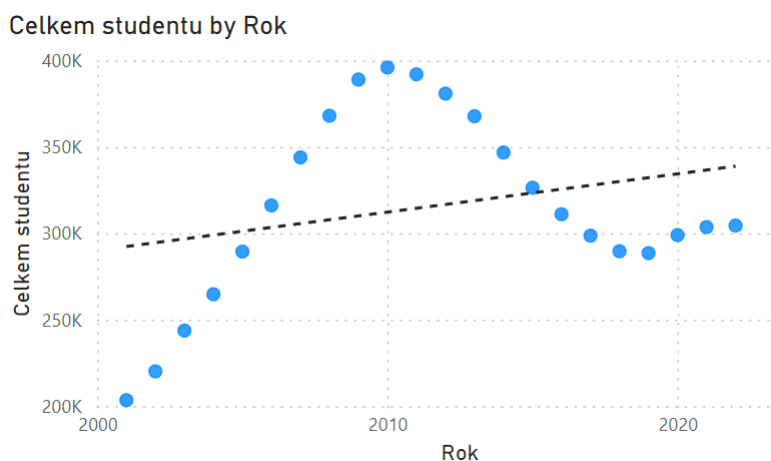
Vytvoření takového modelu by však bylo komplexním úkolem, který by vyžadoval zohlednění počtu studentů v každém ročníku na základě podobného výpočtu, jako byl využit pro odhad počtu nových studentů. Kromě toho by bylo nutné zahrnout faktory jako opakování ročníků, přestupy mezi obory a jiné formy přerušeni nebo pokračování studia. Také by bylo třeba vzít v úvahu různé rychlosti absolvování programů, které se mohou lišit mezi jednotlivými obory a institucemi. Pro naše účely používáme pouze odhad nově zapsaných studentů, což umožňuje vytvořit výukové materiály a demonstrovat základní postupy datové analýzy, i když tento přístup není nejpřesnější.

Problém může být také v dostupných datech. Nepodařilo se mi dohledat dostupná data o možných lidech v daných stářích, protože data ze statistického

úřadu jsou seskupena do širších věkových kategorií, jako například 14-19 let a 20-24 let. Toto seskupení neposkytuje dostatečně detailní informace pro přesnou analýzu. Tento nedostatek v dostupných datech omezuje přesnost našich modelů a predikcí.

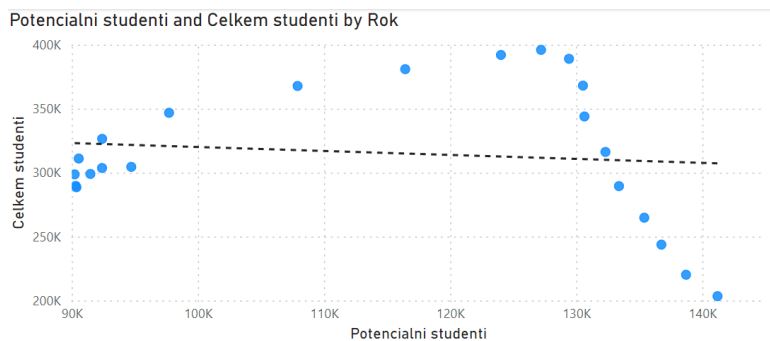
## ■ Regresní analýza

Pro analýzu trendů v datech jsme použili lineární regresní analýzu, která porovnává dvě proměnné. Na obrázku 4.5 je zobrazen graf s lineární regresní čarou, která ukazuje celkový trend mezi rokem a celkovým počtem studentů. Ačkoliv data ukazují vrchol kolem roku 2010, celkový dlouhodobý trend je rostoucí, což naznačuje, že se počet studentů v průběhu času zvyšuje.



**Obrázek 4.5:** Regresní analýza počtu studentů vysokých škol

Regresní analýza potenciálních nových studentů (obrázek 4.6) ukazuje vztah mezi počtem potenciálních studentů a celkovým počtem nově zapsaných studentů. Tento výsledek může indikovat demografické změny a potenciální budoucí výzvy pro vysoké školy v oblasti náboru studentů.



**Obrázek 4.6:** Regresní analýza potenciálních nových studentů

### 4.4.3 Interpretace výsledků

Při analýze dat jsme zjistili, že celkový počet studentů dosáhl vrcholu v roce 2010 a poté klesal, ale od roku 2018 začal opět růst. Tento trend je viditelný jak v grafu celkového počtu studentů, tak v regresní analýze. Odhad potenciálního počtu nových studentů ukazuje, že počet nových uchazečů o studium klesá, což může ovlivnit budoucí trendy v počtu studentů.

Následující interpretace a vysvětlení vycházejí z mých názorů na základě analýzy demografických, ekonomických a sociálních faktorů.

### Možné důvody pro nárůst a pokles počtu zapsaných studentů kolem roku 2010

Zde jsou možné důvody pro nárůst počtu zapsaných studentů kolem roku 2010 a následný pokles, i když křivka možných nástupců v tu dobu klesala:

- **Demografický vývoj:** Jak jsme viděli na grafu počtu narozených dětí, počet potenciálních studentů klesá. Nicméně, počet zapsaných studentů dosáhl vrcholu kolem roku 2010, což může být způsobeno dalšími faktory.
- **Ekonomické faktory:** Po ekonomické krizi v roce 2008[27] mohlo více lidí hledat vysokoškolské vzdělání jako cestu ke zlepšení svých šancí na trhu práce. Pro čerstvé absolventy středních škol také mohlo být atraktivnější pokračovat ve studiu, než se snažit najít zaměstnání v tomto období.
- **Debaty o zavedení školného:** V období kolem roku 2010 se vedly rozsáhlé debaty o zavedení povinného školného, což mohlo motivovat studenty k rychlejšímu zápisu na vysoké školy, aby se vyhnuli případným poplatkům v budoucnu.[28]

Navíc, změny v českém vysokoškolském systému od roku 1999, jako zavedení dvou stupňů studia (bakalářského a navazujícího magisterského místo tradičního pětiletého cyklu), změny v systému akreditací studijních programů, umožnění vzniku soukromých vysokých škol, internacionalizace a větší otevřenost vysokých škol, rovněž přispěly k růstu počtu studentů mezi roky 2001 až 2010.[29]

### 4.4.4 Shrnutí a závěr hypotézy 1

Na základě výše uvedené analýzy můžeme konstatovat, že dlouhodobý trend počtu studentů na vysokých školách je rostoucí, ačkoli v letech 2010-2018 došlo k poklesu. Tento trend je důležitý pro studenty i školy z hlediska plánování a strategického rozhodování. Pro přesnější analýzu by bylo vhodné zahrnout i data o již studujících studentech. Celkový růst počtu studentů naznačuje, že vysokoškolské vzdělání je stále důležitější a poptávka po něm roste.

Hypotéza byla potvrzena, protože dlouhodobý trend ukazuje na růst počtu studentů, i přes období poklesu mezi lety 2010 a 2018.



## 4.5 Hypotéza 2: Vztah mezi národností studenta a typem studia

**Formulace otázky:** Zahraniční studenti častěji volí prezenční formu studia než čeští studenti.

### 4.5.1 Úvod do problematiky

Forma studia může výrazně ovlivnit studijní zkušenost a akademické výsledky studentů. Zatímco čeští studenti mohou preferovat kombinované nebo distanční formy studia, které jim umožňují skloubit studium s prací nebo jinými povinnostmi, zahraniční studenti často volí prezenční formu studia. To může být způsobeno několika faktory, jako je potřeba intenzivní jazykové a kulturní integrace, vyšší náklady na život v cizí zemi nebo specifické požadavky stipendijních programů, které často vyžadují plnou účast na studiu.

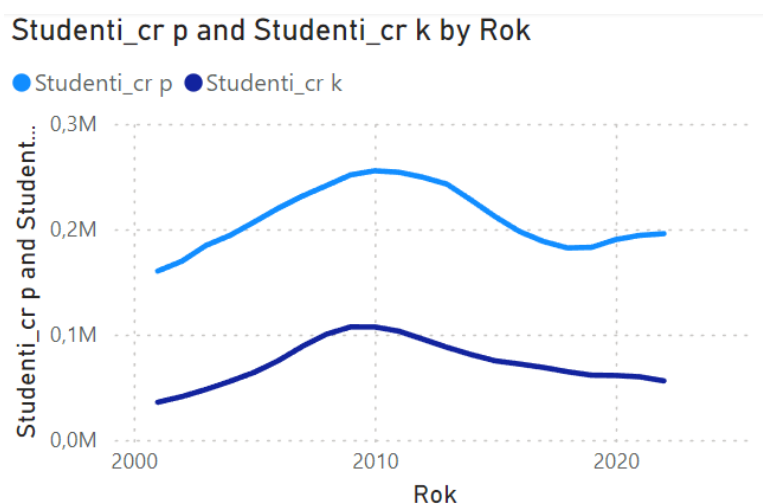
Zahraniční studenti se také často snaží maximalizovat svůj akademický a sociální zážitek během omezené doby pobytu v hostitelské zemi. Prezenční forma studia jim umožňuje více se zapojit do univerzitního života, což může zahrnovat přístup k různým akademickým a sociálním aktivitám, lepší možnosti networking a získání hlubšího porozumění místní kultuře a akademickému prostředí.[30]

### 4.5.2 Analýza dat

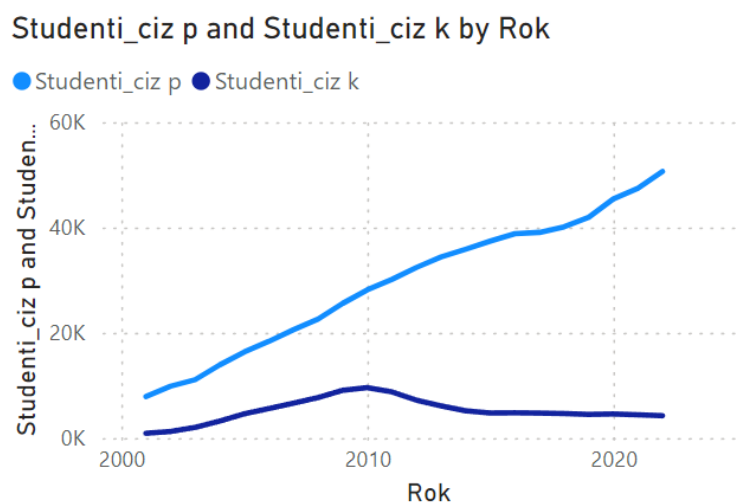
Pro analýzu dat byly použity soubory z Ministerstva školství, mládeže a tělovýchovy (MŠMT), konkrétně tabulky obsahující data o formách studia (prezenční a kombinované) rozdělené podle národnosti studentů. Níže jsou uvedeny jednotlivé kroky analýzy.

#### Data o formách studia českých a zahraničních studentů

Na obrázcích 4.7 a 4.8 je zobrazen počet studentů ve formě prezenčního a kombinovaného studia rozdělený podle českých a zahraničních studentů. Použili jsme spojnicové grafy pro vizualizaci rozdílů mezi těmito dvěma skupinami studentů.



**Obrázek 4.7:** Počet českých studentů ve formě prezenčního a kombinovaného studia

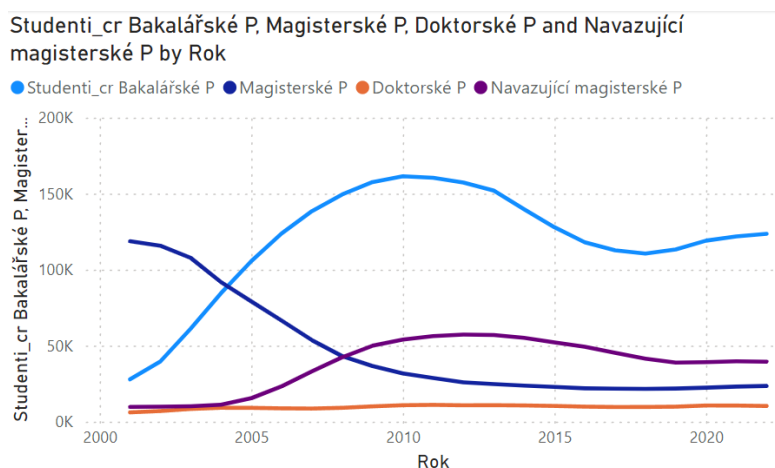


**Obrázek 4.8:** Počet zahraničních studentů ve formě prezenčního a kombinovaného studia

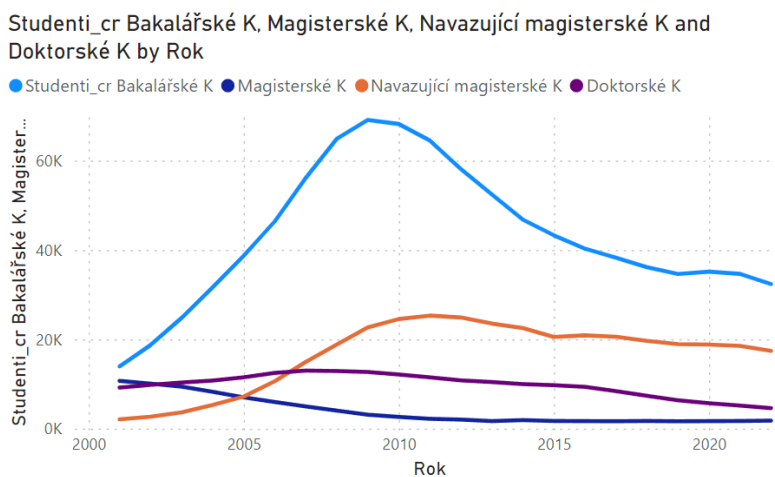
Analyzovali jsme grafy 4.7 a 4.8, abychom identifikovali trendy a rozdíly v preferencích formy studia mezi českými a zahraničními studenty.

## Data o formách studia podle stupně studia

Na obrázcích 4.9 a 4.10 jsou zobrazeny detaily o počtu českých studentů v jednotlivých stupních studia (bakalářské, magisterské, navazující magisterské a doktorské) podle formy studia.

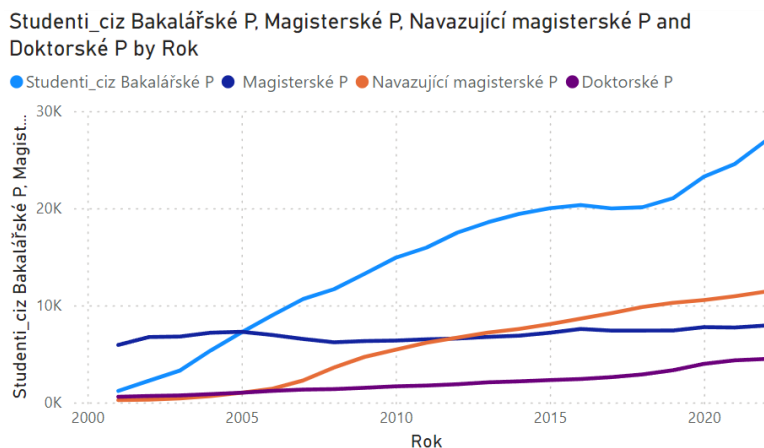


**Obrázek 4.9:** Počet českých studentů ve formě prezenčního studia podle stupně studia

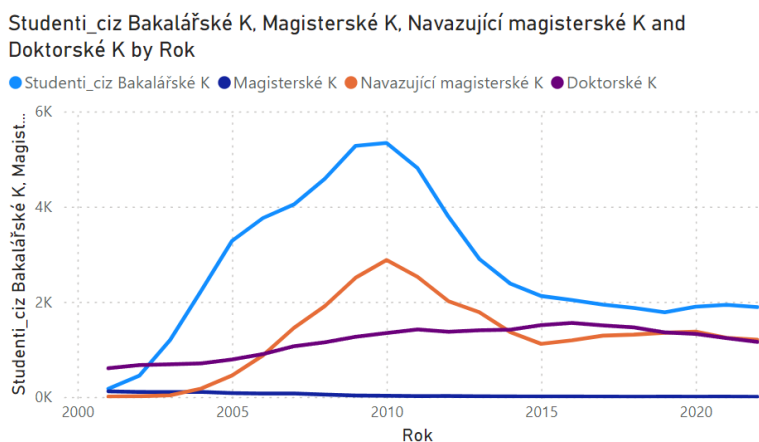


**Obrázek 4.10:** Počet českých studentů ve formě kombinovaného studia podle stupně studia

Na obrázcích 4.11 a 4.12 jsou zobrazeny detaily o počtu zahraničních studentů v jednotlivých stupních studia podle formy studia.



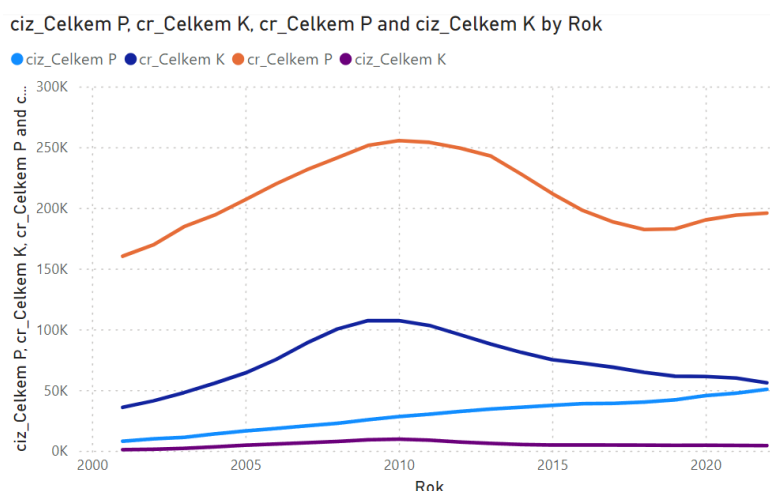
**Obrázek 4.11:** Počet zahraničních studentů ve formě prezenčního studia podle stupně studia



**Obrázek 4.12:** Počet zahraničních studentů ve formě kombinovaného studia podle stupně studia

#### ■ Porovnání prezenčního a kombinovaného studia českých a zahraničních studentů

Na obrázku 4.13 je zobrazen počet studentů ve formě prezenčního a kombinovaného studia pro české i zahraniční studenty dohromady.



**Obrázek 4.13:** Počet studentů ve formě prezenčního a kombinovaného studia (čeští a zahraniční studenti dohromady)

Tyto grafy slouží k vizualizaci počtu studentů v různých formách a stupních studia. Umožňují nám identifikovat a interpretovat rozdíly mezi českými a zahraničními studenty, což je klíčové pro pochopení jejich preferencí a studijních návyků. Pomocí těchto dat můžeme analyzovat, proč některé skupiny studentů preferují určité formy a stupně studia, a hledat důvody těchto rozdílů.

### 4.5.3 Interpretace výsledků

Analýza dat ukazuje, že zahraniční studenti mají tendenci volit prezenční formu studia častěji než čeští studenti. Důvody mohou být různé[30]:

- **Jazyková a kulturní integrace:** Zahraniční studenti často potřebují intenzivní jazykovou praxi a kulturní adaptaci, což jim prezenční studium umožňuje lépe než distanční formy.
- **Finanční a stipendijní podmínky:** Mnoho zahraničních studentů studuje na základě stipendií, která často vyžadují plnou účast na studiu.
- **Maximalizace zážitku:** Zahraniční studenti chtějí během svého omezeného pobytu v hostitelské zemi získat co nejvíce z akademického a sociálního života.
- **Akademické požadavky:** Některé programy, zvláště na úrovni magisterských a doktorských studií, mohou vyžadovat intenzivní laboratorní a výzkumné práce, které jsou lépe realizovatelné v prezenční formě.

Naopak čeští studenti, kteří mají možnost studium kombinovat s prací nebo jinými závazky, často volí kombinované nebo distanční formy studia.

#### 4.5.4 Shrnutí a závěr hypotézy 2

Z analýzy dat vyplývá, že zahraniční studenti preferují prezenční formu studia ve srovnání s českými studenty. Tato preference může být způsobena potřebou intenzivní integrace, finančními a stipendijními podmínkami a snahou maximálně využít pobyt v hostitelské zemi. Tyto poznatky jsou důležité pro univerzity, které mohou na základě těchto informací lépe plánovat a přizpůsobovat své programy a služby jak pro české, tak pro zahraniční studenty.

### 4.6 Hypotéza 3: Geografické rozložení studentů

**Formulace otázky:** Která města mají nejvyšší počet studentů a jaký je jejich podíl na celkovém počtu studentů v letech 2001 a 2022?

#### 4.6.1 Úvod do problematiky

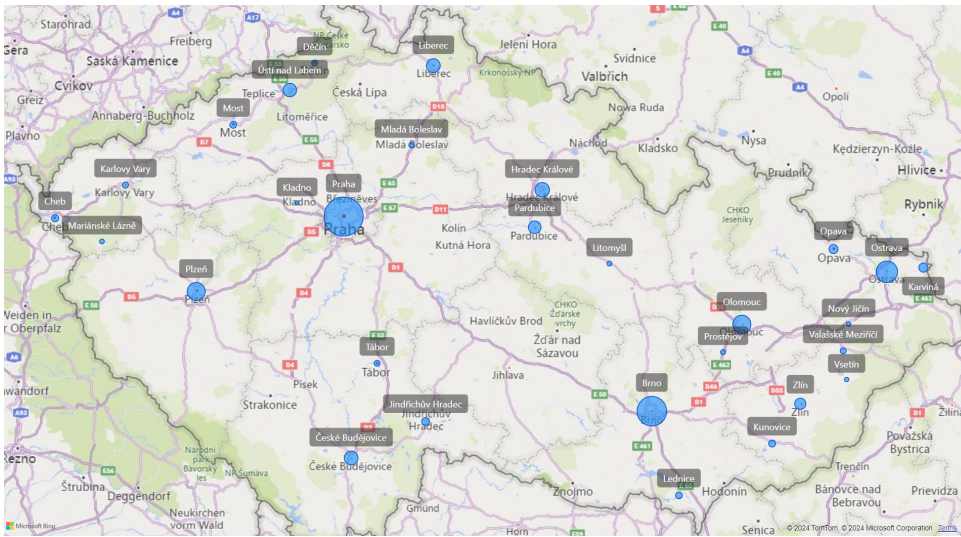
Geografické rozložení studentů je klíčovým faktorem pro plánování a rozvoj vzdělávací infrastruktury. Města s vysokou koncentrací studentů často čelí specifickým výzvám i příležitostem, které ovlivňují nejen vzdělávací instituce, ale i místní ekonomiku a sociální strukturu. Faktory ovlivňující koncentraci studentů ve velkých městech zahrnují dostupnost a kvalitu vysokých škol, životní náklady, možnosti bydlení, pracovní příležitosti a kulturní a společenský život.

#### 4.6.2 Analýza dat

Data o počtu studentů podle měst byla načtena a zpracována pro roky 2001 a 2022. Grafy zobrazují počet studentů v jednotlivých městech a umožňují identifikovat města s nejvyšší koncentrací studentů v daných letech.

Na obrázku 4.14 je znázorněno geografické rozložení studentů v roce 2001. Tento typ grafu, využívající mapy s kruhy různé velikosti, ukazuje koncentraci studentů v jednotlivých městech. Větší kruhy indikují vyšší počet studentů, což usnadňuje rychlé vizuální porovnání mezi městy. Praha a Brno mají nejvyšší koncentraci studentů, což je snadno viditelné díky velikosti kruhů.

#### 4.6. Hypotéza 3: Geografické rozložení studentů



**Obrázek 4.14:** Geografické rozložení studentů v roce 2001

Podrobnější pohled na počet studentů v jednotlivých městech v roce 2001 poskytuje obrázek 4.15. Tato tabulka uvádí přesný počet studentů pro každé město, což je užitečné pro detailnější analýzu a srovnání. Data ukazují, že Praha, Brno a Olomouc jsou na předních příčkách.

Město	Celkem	Bakalářské	Magisterské	Doktorské	Poprvé zapsaní	Absolventi	Cizinci	Rok
Olomouc	12248	2000	9317	1004	2479	1838	669	2001
Opava	1579	817	743	25	313	295	44	2001
Kladno	2	2	0	0	2	0	0	2001
Mladá Boleslav	130	130	0	0	58	0	2	2001
Vsetín	13	12	1	0	0	4	0	2001
České Budějovice	5832	1466	4148	239	1294	1113	44	2001
Valašské Meziříčí	265	265	0	0	108	0	3	2001
Jindřichův Hradec	903	296	617	0	198	297	33	2001
Kunovice	537	537	0	0	244	85	36	2001
Tábor	241	0	241	0	61	5	4	2001
Cheb	628	541	87	0	267	156	10	2001
Mariánské Lázně	43	43	0	0	0	13	2	2001
Praha	74198	11232	56557	7485	13927	10691	4677	2001
Plzeň	12008	2442	9189	506	3051	1839	439	2001
Ústí nad Labem	5575	2061	3530	16	1424	923	22	2001
Ostrava	18516	3956	13493	1131	4723	2711	302	2001
Karlovy Vary	291	291	0	0	154	0	1	2001
Pardubice	4605	1912	2395	304	1229	648	45	2001
Děčín	213	213	0	0	79	17	1	2001
Liberec	5915	1398	4420	260	1286	704	344	2001
Most	492	254	239	0	158	55	0	2001
Albrechtický	0	0	0	0	0	0	1	2001
Hradec Králové	6735	2051	4388	332	1501	1293	251	2001
Litomyšl	59	59	0	0	7	0	4	2001
Brno	39735	7256	29032	3849	8879	5843	1672	2001
Lednice	478	117	323	38	35	133	15	2001
Zlín	3229	2209	895	133	1214	583	108	2001
Prostějov	69	69	0	0	33	0	0	2001
Karviná	1683	184	1504	0	385	297	62	2001
Nový Jičín	22	22	0	0	22	0	0	2001

**Obrázek 4.15:** Počet studentů v jednotlivých městech v roce 2001



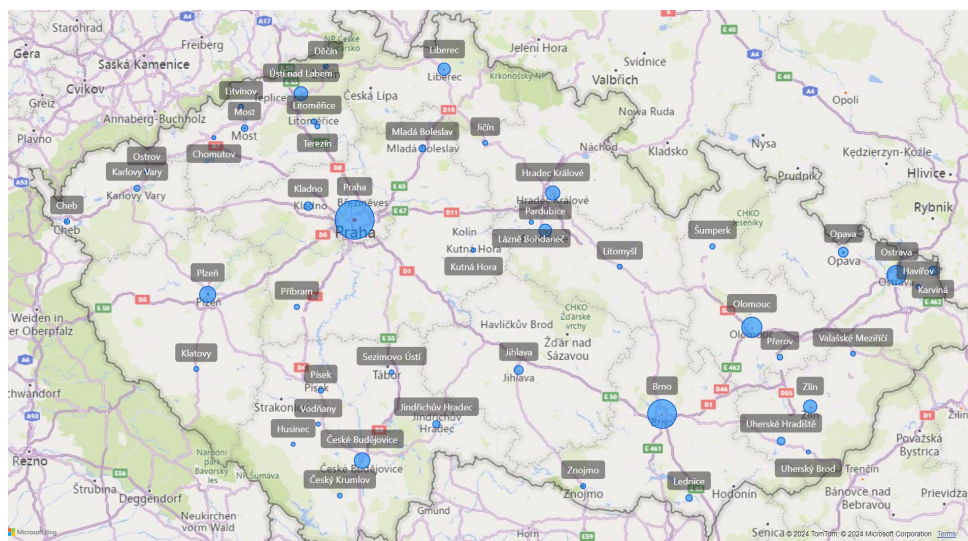
#### 4. Analýza dat a formulace otázek a hypotéz

Obrázek 4.16 zobrazuje počet studentů v prvních deseti městech v roce 2001. Zaměřuje na nejvýznamnější města, což umožňuje lepší pochopení rozdílů mezi nimi. Zde vidíme, že Praha má výrazně vyšší počet studentů než ostatní města.

Město	Celkem	Bakalářské	Magisterské	Doktorské	Poprvé zapsaní	Absolventi	Cizinci	Rok
Praha	74198	11232	56557	7485	13927	10691	4677	2001
Brno	39735	7256	29032	3849	8879	5843	1672	2001
Ostrava	18516	3956	13493	1131	4723	2711	302	2001
Olomouc	12248	2000	9317	1004	2479	1838	669	2001
Plzeň	12008	2442	9189	506	3051	1839	439	2001
Hradec Králové	6735	2051	4388	332	1501	1293	251	2001
Liberec	5915	1398	4420	260	1286	704	344	2001
České Budějovice	5832	1466	4148	239	1294	1113	44	2001
Ústí nad Labem	5575	2061	3530	16	1424	923	22	2001
Pardubice	4605	1912	2395	304	1229	648	45	2001

**Obrázek 4.16:** Počet studentů v prvních deseti městech v roce 2001

Podobně je na obrázku 4.17 zobrazeno geografické rozložení studentů v roce 2022. Opět jsou větší kruhy indikátorem vyššího počtu studentů. Obrázek 4.18 ukazuje počet studentů v jednotlivých městech v roce 2022. Stejně jako v roce 2001, i v roce 2022 dominují Praha a Brno.



**Obrázek 4.17:** Geografické rozložení studentů v roce 2022



#### 4.6. Hypotéza 3: Geografické rozložení studentů

Město	Celkem	Bakalářské	Magisterské	Doktorské	Poprvé zapsaní	Absolventi	Cizinci	Rok
Olomouc	20383	11402	8041	1070	3066	3689	2650	2022
Opava	2734	2265	414	60	606	469	187	2022
Prerov	223	143	80	0	26	82	89	2022
Šumperk	149	115	34	0	64	53	0	2022
Kladno	1649	1107	401	142	246	356	128	2022
Kutná Hora	14	14	0	0	0	30	2	2022
Mladá Boleslav	692	548	144	0	105	157	116	2022
Husinec	2	0	0	2	0	2	0	2022
Příbram	139	139	0	0	56	65	0	2022
České Budějovice	10617	8142	2183	308	2507	2020	885	2022
Nové Hradky	0	0	0	0	0	1	1	2022
Valašské Meziříčí	59	59	0	0	0	25	0	2022
Český Krumlov	67	30	37	0	3	11	0	2022
Jindřichův Hradec	876	397	464	17	124	149	163	2022
Písek	129	91	38	0	29	55	22	2022
Vodňany	7	0	0	7	0	4	23	2022
Sezimovo Ústí	49	0	49	0	0	24	1	2022
Cheb	179	119	60	0	56	7	4	2022
Praha	94132	56487	31329	6981	16798	18389	27767	2022
Plzeň	11818	6268	4965	637	2470	2191	1409	2022
Ústí nad Labem	7532	5366	1929	247	1490	1153	346	2022
Ostrava	17879	11723	5222	1005	3650	3666	2136	2022
Karlovy Vary	361	259	102	0	70	53	8	2022
Havířov	75	64	11	0	6	57	4	2022
Pardubice	5985	4381	1354	252	1339	1169	517	2022
Ostrov	1	1	0	0	0	0	0	2022
Klatovy	90	28	62	0	13	42	0	2022
Děčín	83	73	10	0	19	10	9	2022
Chomutov	1	1	0	0	0	3	0	2022
Liberec	5637	4047	1376	217	1220	951	503	2022
Litoměřice	175	65	110	0	23	100	0	2022
Terezín	80	80	0	0	43	50	2	2022
Most	541	378	163	0	134	140	3	2022
Litvínov	106	99	7	0	26	19	0	2022
Hradec Králové	8214	4521	3310	416	1664	1621	1326	2022
Jičín	85	28	57	0	16	43	0	2022
Lázně Bohdaneč	16	16	0	0	0	1	2	2022
Litomyšl	82	53	25	4	15	17	14	2022
Brno	47832	27875	17142	3238	9417	10648	14829	2022
Lednice	550	361	150	39	39	131	121	2022
Zlín	6794	4641	1906	252	1501	1464	968	2022
Jihlava	1973	1942	31	0	575	301	130	2022
Uherské Hradiště	1171	826	345	0	290	270	38	2022
Uherský Brod	2	2	0	0	0	1	0	2022
Znojmo	98	96	2	0	5	40	1	2022
Karviná	2119	1617	495	9	531	316	146	2022
zahraničí: Slovensko	3	2	1	0	2	0	111	2022
zahraničí: Švýcarská konfederace	0	0	0	0	0	0	152	2022
zahraničí: neuveďeno	1	1	0	0	0	0	99	2022

**Obrázek 4.18:** Počet studentů v jednotlivých městech v roce 2022

Obrázek 4.19 zobrazuje počet studentů v prvních deseti městech v roce 2022. Podobně jako v roce 2001, Praha stále vede s výrazným náskokem před Brnem a dalšími městy.

Město	Celkem	Bakalářské	Magisterské	Doktorské	Poprvé zapsaní	Absolventi	Cizinci	Rok
Praha	94132	56487	31329	6981	16798	18389	27767	2022
Brno	47832	27875	17142	3238	9417	10648	14829	2022
Olomouc	20383	11402	8041	1070	3066	3689	2650	2022
Ostrava	17879	11723	5222	1005	3650	3666	2136	2022
Plzeň	11818	6268	4965	637	2470	2191	1409	2022
České Budějovice	10617	8142	2183	308	2507	2020	885	2022
Hradec Králové	8214	4521	3310	416	1664	1621	1326	2022
Ústí nad Labem	7532	5366	1929	247	1490	1153	346	2022
Zlín	6794	4641	1906	252	1501	1464	968	2022
Pardubice	5985	4381	1354	252	1339	1169	517	2022

Obrázek 4.19: Počet studentů v prvních deseti městech v roce 2022

Tyto grafy a tabulky jsou zásadní pro vizualizaci dat, protože umožňují rychlé porovnání a identifikaci hlavních trendů a rozdílů. Geografické zobrazování koncentrace studentů v různých městech napomáhá při rozhodování o budoucích investicích a rozvoji vzdělávací infrastruktury. Vizualizace dat rovněž usnadňuje komunikaci výsledků a závěrů širokému publiku, což je klíčové pro efektivní prezentaci a pochopení dat.

### 4.6.3 Interpretace výsledků

Na základě analýzy dat bylo zjištěno, že Praha a Brno mají nejvyšší koncentraci studentů v obou sledovaných letech. Tento trend je pravděpodobně způsoben kvalitou a počtem vysokých škol v těchto městech, které přitahují studenty z celé republiky i ze zahraničí. Města jako Olomouc a Plzeň také vykazují významný počet studentů, což naznačuje, že jsou důležitými vzdělávacími centry.

#### Identifikace měst s nejvyšší koncentrací studentů a možných důvodů

- **Praha:** Hlavní město s nejvyšší koncentrací prestižních vysokých škol, jako je Univerzita Karlova, České vysoké učení technické, Vysoká škola ekonomická, Česká zemědělská univerzita a další. Praha také nabízí široké možnosti kulturního a společenského vyžití, což přitahuje studenty.[31]
- **Brno:** Druhé největší město v České republice, domov pro Masarykovu univerzitu, Vysoké učení technické v Brně a další významné instituce. Brno je také známé svou inovativní a technologickou scénou.[32]
- **Olomouc:** Univerzita Palackého v Olomouci je jednou z nejstarších univerzit v České republice a přitahuje studenty svou dlouhou historií a akademickou kvalitou.[31]
- **Plzeň:** Západočeská univerzita v Plzni je významnou institucí v regionu a nabízí širokou škálu studijních programů.[31]

Další faktory, které mohou ovlivňovat koncentraci studentů, zahrnují dostupnost bydlení, možnosti pracovních příležitostí během studia a po jeho ukončení, a také celkovou kvalitu života v daném městě.

### 4.6.4 Shrnutí a závěr hypotézy 3

Na základě výše uvedené analýzy lze konstatovat, že Praha a Brno dominují v počtu studentů jak v roce 2001, tak v roce 2022. Olomouc a Plzeň také patří mezi města s vysokou koncentrací studentů. Tento trend je pravděpodobně způsoben kombinací kvalitních vzdělávacích institucí, bohaté kulturní nabídky a dobrých životních podmínek v těchto městech.

Potvrzuje se tedy hypotéza, že nejvyšší počet studentů je soustředěn ve velkých městech, která nabízejí nejlepší podmínky pro studium a život. Tato informace je klíčová pro plánování rozvoje vysokého školství a infrastruktury v České republice.

## 4.7 Hypotéza 4: Geografické rozložení zahraničních studentů v České republice

**Formulace otázky:** Jaké jsou změny v geografickém rozložení zahraničních studentů mezi lety 2001 a 2022?

### 4.7.1 Úvod do problematiky

Geografické rozložení zahraničních studentů je klíčovým faktorem pro pochopení dynamiky a atraktivity různých měst a regionů v rámci České republiky. Významné změny v počtu zahraničních studentů mohou signalizovat růst atraktivity určitých měst díky kvalitě vzdělání, životním podmínkám nebo ekonomickým a kulturním faktorům.

### 4.7.2 Analýza dat

Pro analýzu byly použity údaje o počtu zahraničních studentů v jednotlivých městech v letech 2001 a 2022. Data zahrnují mapové vizualizace a tabulky top 10 měst s nejvyšším počtem zahraničních studentů.

### Rok 2001

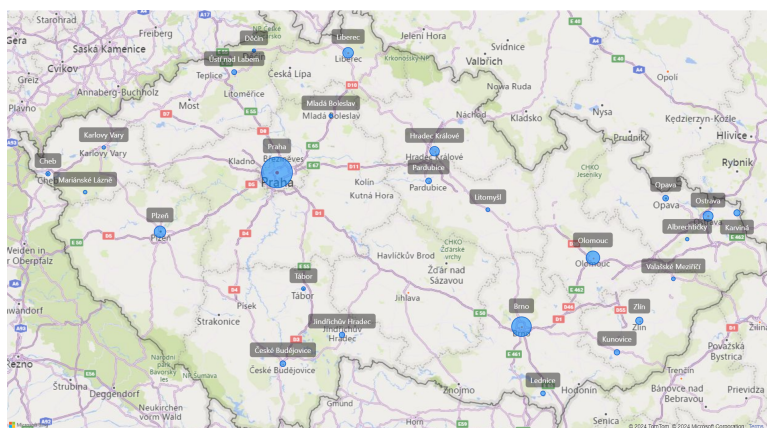
Data pro rok 2001 jsou uvedena v tabulce 4.20 a mapě 4.21. Největší počet zahraničních studentů se v roce 2001 soustředil v Praze (4677), následován Brnem (1672) a Olomoucí (669).

#### 4. Analýza dat a formulace otázek a hypotéz

Město	Cizinci	Rok
Praha	4677	2001
Brno	1672	2001
Olomouc	669	2001
Plzeň	439	2001
Liberec	344	2001
Ostrava	302	2001
Hradec Králové	251	2001
Zlín	108	2001
Karviná	62	2001
Pardubice	45	2001

**Obrázek 4.20:** Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2001

Sloupcový graf 4.20 zobrazuje top 10 měst v roce 2001 podle počtu zahraničních studentů. Pomáhá rychle identifikovat města s nejvyšší koncentrací zahraničních studentů.



**Obrázek 4.21:** Geografické rozložení zahraničních studentů v roce 2001

Mapa 4.21 znázorňuje geografické rozložení zahraničních studentů v roce 2001. Větší kruhy indikují vyšší počet studentů v daných městech, což umožňuje vizuálně porovnat koncentraci zahraničních studentů v různých regionech.

#### Rok 2022

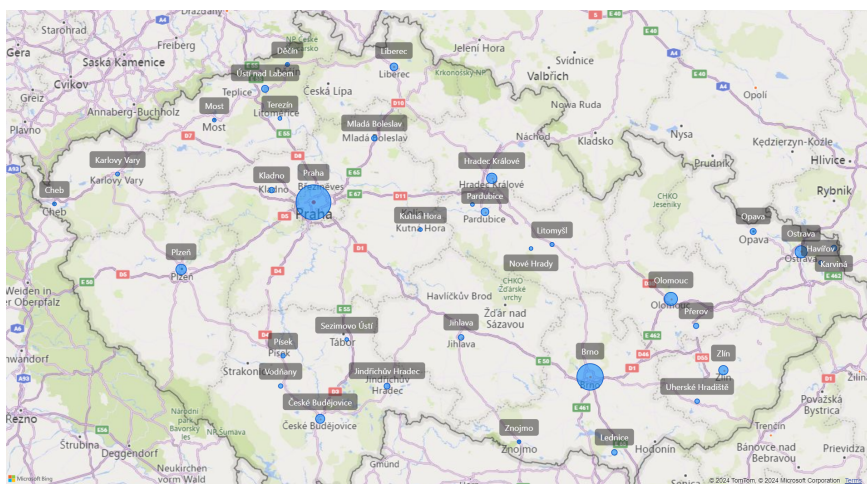
Data pro rok 2022 jsou uvedena v tabulce 4.22 a mapě 4.23. V roce 2022 je nárůst počtu zahraničních studentů patrný v Praze (27767), Brně (14829) a Olomouci (2650).

#### 4.7. Hypotéza 4: Geografické rozložení zahraničních studentů v České republice

Město	Cizinci	Rok
Praha	27767	2022
Brno	14829	2022
Olomouc	2650	2022
Ostrava	2136	2022
Plzeň	1409	2022
Hradec Králové	1326	2022
Zlín	968	2022
České Budějovice	885	2022
Pardubice	517	2022
Liberec	503	2022

**Obrázek 4.22:** Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2022

Tento sloupcový graf 4.22 zobrazuje top 10 měst v roce 2022 podle počtu zahraničních studentů. Umožňuje snadné porovnání s rokem 2001 a identifikaci měst s nejvyšší atraktivitou pro zahraniční studenty.

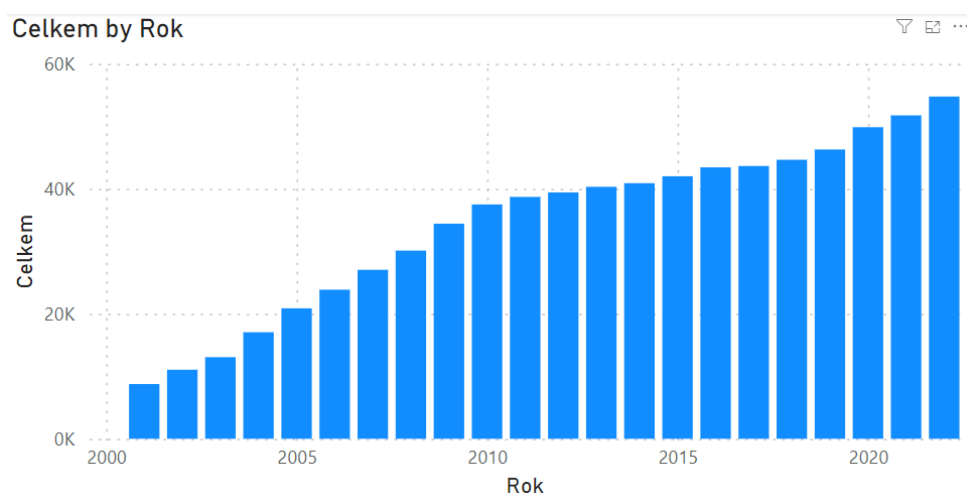


**Obrázek 4.23:** Geografické rozložení zahraničních studentů v roce 2022

Mapa 4.23 znázorňuje geografické rozložení zahraničních studentů v roce 2022. Větší kruhy indikují vyšší počet studentů v daných městech, což umožňuje vizuálně porovnat koncentraci zahraničních studentů v různých regionech a sledovat změny oproti roku 2001.

#### ■ Vývoj počtu zahraničních studentů v letech 2001 až 2022

Graf 4.24 ukazuje vývoj počtu zahraničních studentů v České republice od roku 2001 do roku 2022. Data ukazují stabilní nárůst počtu zahraničních studentů během těchto let, což naznačuje rostoucí atraktivitu českých vysokých škol pro studenty ze zahraničí.



**Obrázek 4.24:** Vývoj počtu zahraničních studentů v letech 2001 až 2022

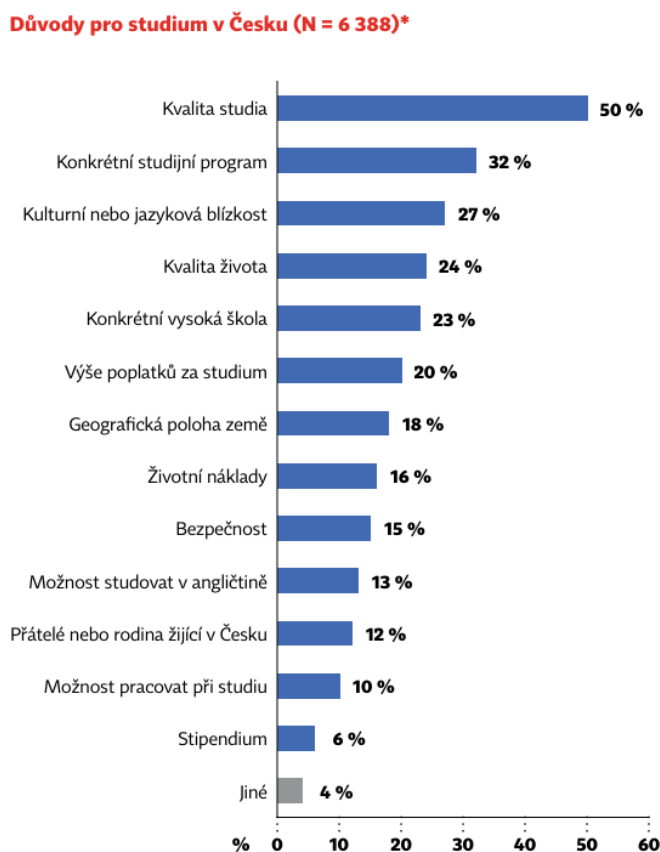
### 4.7.3 Interpretace výsledků

Z výše uvedených grafů a tabulek je zřejmé, že počet zahraničních studentů na českých vysokých školách se od roku 2001 do roku 2022 výrazně zvýšil. V roce 2001 bylo v České republice pouze několik tisíc zahraničních studentů, zatímco v roce 2022 tento počet přesáhl 54 000. Nejvýraznější nárůst byl zaznamenán v Praze a Brně, kde se počet zahraničních studentů výrazně zvýšil. V roce 2022 bylo v Praze více než 27 000 zahraničních studentů, což je více než pětinašobek počtu z roku 2001.

Dalším důležitým městem je Brno, kde se počet zahraničních studentů rovněž významně zvýšil. V roce 2001 bylo v Brně 1672 zahraničních studentů, zatímco v roce 2022 tento počet vzrostl na 14829. Další města, jako jsou Olomouc, Plzeň, Ostrava a Hradec Králové, také zaznamenala nárůst počtu zahraničních studentů, i když ne v takové míře jako Praha a Brno.

## ■ Důvody pro studium v Česku

Podle zprávy DZS<sup>2</sup> (Dům zahraniční spolupráce) z roku 2023 jsou hlavními důvody pro volbu studia v České republice vidět na obrázku 4.25:



\* Bylo možné vybrat až tři možnosti.

**Obrázek 4.25:** Důvody pro studium v Česku (Zdroj: DZS - Zpráva o zahraničních studentech na českých vysokých školách 2023, 2023)[30]

Dle výzkumu zahraničních studentů provedeného Domem zahraniční spolupráce (DZS) je kvalita vzdělání nejdůležitějším kritériem při volbě Česka jako destinace ke studiu. Kvalitu nabízených oborů si kladou za cíl zhodnotit a do určité míry zaručit mezinárodní univerzitní žebříčky. Například dle žebříčku QS EECA (Emerging Europe and Central Asia) University Rankings<sup>3</sup> se v roce 2022 umístily dvě české VŠ v top 100, v žebříčku QS World University Rankings 2023<sup>4</sup> se pak umístily tři české VŠ v top 500.[30]

<sup>2</sup><https://www.dzs.cz/>

<sup>3</sup><https://www.topuniversities.com/university-rankings/eeca-rankings/2022>

<sup>4</sup><https://www.topuniversities.com/world-university-rankings/2023>

#### ■ 4.7.4 Shrnutí a závěr hypotézy 4

Na základě výše uvedené analýzy lze konstatovat, že počet zahraničních studentů na českých vysokých školách významně vzrostl od roku 2001 do roku 2022. Přestože Praha a Brno přitahují největší počet zahraničních studentů, i ostatní města zaznamenávají nárůst. Kvalita vzdělání, konkrétní studijní program a kulturní nebo jazyková blízkost jsou hlavními důvody pro volbu Česka jako studijní destinace.

### ■ 4.8 Hypotéza 5: Rozdíl mezi soukromými a veřejnými školami

**Formulace otázky:** Jaký je rozdíl v počtu studentů a úspěšnosti mezi veřejnými a soukromými vysokými školami?

#### ■ 4.8.1 Úvod do problematiky

Vysoké školy v České republice lze rozdělit na veřejné a soukromé instituce. Rozdíly mezi nimi mohou ovlivnit nejen počet studentů, ale také jejich úspěšnost v dokončení studia. Veřejné školy mají často delší tradici, širší nabídku oborů a jsou finančně podporovány státem. Na druhé straně soukromé školy mohou nabízet inovativní programy, flexibilní studijní podmínky a individuálnější přístup ke studentům. Faktory jako finanční náklady, prestiž, kvalita vzdělání a možnosti kariérního uplatnění mohou ovlivnit volbu mezi soukromou a veřejnou vysokou školou.[33]

#### ■ 4.8.2 Analýza dat

Pro analýzu byly použity údaje z tabulky F12, která obsahuje data o počtu studentů na veřejných a soukromých vysokých školách v letech 2020 až 2022. Předchozí roky nebyly rozděleny na veřejné a soukromé školy, a proto se zaměřujeme na data z těchto let. Byly vytvořeny srovnávací grafy pro porovnání počtu studentů a absolventů na těchto institucích. Pro lepší přehlednost byly zmenšeny oblasti v tabulkách a grafech, které nejsou relevantní.

#### ■ Veřejné školy

Data pro veřejné školy jsou uvedena v tabulce 4.26 a grafu 4.27. Z grafů a tabulek je zřejmé, že počet studentů na veřejných vysokých školách je výrazně vyšší než na soukromých školách. Počet studujících českých studentů se pohybuje kolem 230 000, zatímco počet poprvé zapsaných studentů je kolem 45 000 ročně. Počet absolventů se pohybuje kolem 46 000 ročně.

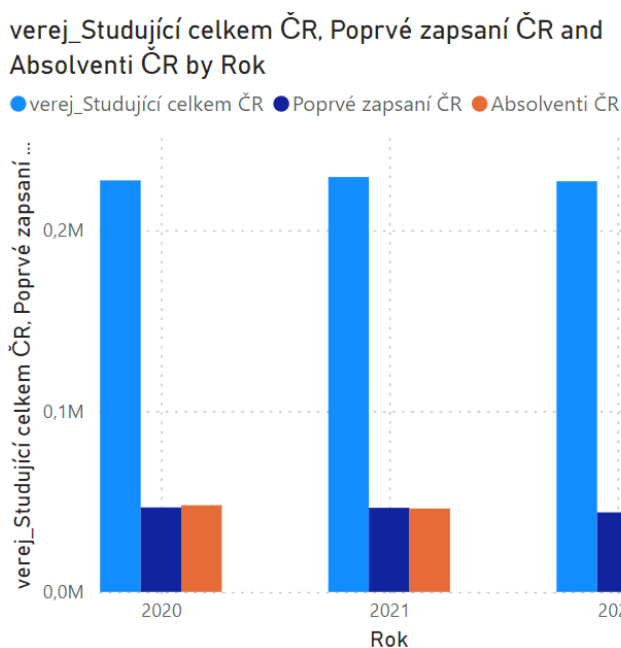


#### 4.8. Hypotéza 5: Rozdíl mezi soukromými a veřejnými školami

Škola	Celko	Celko	Studující celkem ČR	Poprvé zapsaní ČR	Absolventi ČR	Stud	Popr	Absc	Přeri	Jinoy	Samy	Rok
Veřejné školy	272088	150966	227440	46707	47920	44658	9537	7794	10419	138	8919	2020
Veřejné školy	276247	153665	229314	46512	46114	46942	9540	7876	11355	74	9569	2021
Veřejné školy	276632	153557	226967	43979	44797	49681	11193	7735	11952	63	10256	2022

**Obrázek 4.26:** Data o počtu studentů na veřejných vysokých školách

Tabulka 4.26 obsahuje data o počtu studentů na veřejných vysokých školách v letech 2020 až 2022, včetně počtu poprvé zapsaných a absolventů.



**Obrázek 4.27:** Počet studentů na veřejných vysokých školách

Sloupcový graf 4.27 zobrazuje počet studentů, nově zapsaných a absolventů na veřejných vysokých školách v letech 2020 až 2022.

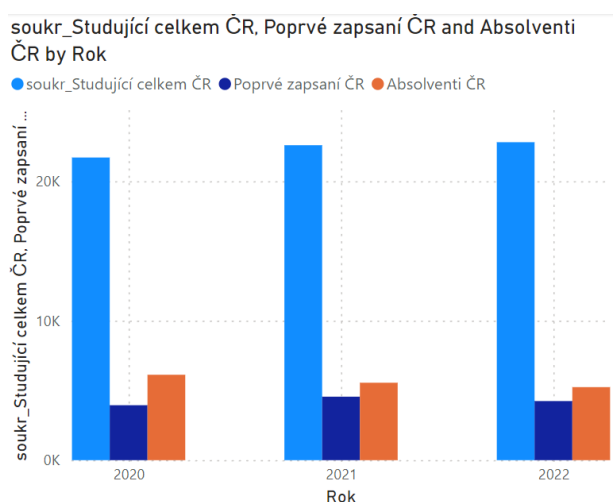
#### ■ Soukromé školy

Data pro soukromé školy jsou uvedena v tabulce 4.28 a grafu 4.29. Počet studentů na soukromých vysokých školách je výrazně nižší. Počet studujících českých studentů se pohybuje kolem 22 000, počet poprvé zapsaných studentů je kolem 4 000 ročně, a počet absolventů je kolem 5 000 ročně.

Škola	Celko	Celko	Studující celkem ČR	Poprvé zapsaní ČR	Absolventi ČR	Stud	Popr	Absc	Přeri	Jinoy	Samy	Rok
Soukromé školy	26898	14992	21701	3940	6125	5203	967	1122	2266	0	0	2020
Soukromé školy	27386	15275	22583	4552	5553	4807	1092	1052	2604	0	0	2021
Soukromé školy	27886	15622	22800	4238	5237	5089	1305	930	2419	0	0	2022

**Obrázek 4.28:** Data o počtu studentů na soukromých vysokých školách

Tabulka 4.28 obsahuje data o počtu studentů na soukromých vysokých školách v letech 2020 až 2022, včetně počtu poprvé zapsaných a absolventů.

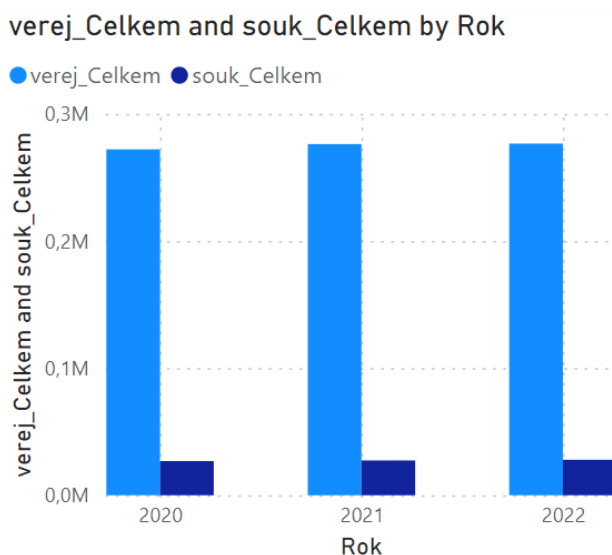


**Obrázek 4.29:** Počet studentů na soukromých vysokých školách

Sloupcový graf 4.29 zobrazuje počet studentů, nově zapsaných a absolventů na soukromých vysokých školách v letech 2020 až 2022.

#### ■ Srovnání celkových počtů studentů

Celkové počty studentů na veřejných a soukromých školách jsou srovnány v grafu 4.30. Je zřejmé, že počet studentů na veřejných školách je výrazně vyšší než na soukromých školách.



**Obrázek 4.30:** Srovnání počtu studentů na veřejných a soukromých školách

Graf 4.30 porovnává celkový počet studentů na veřejných a soukromých školách v letech 2020 až 2022. Umožňuje snadné srovnání a identifikaci rozdílů mezi těmito typy škol.

### 4.8.3 Interpretace výsledků

Na základě analýzy dat je zřejmé, že veřejné vysoké školy mají výrazně vyšší počet studentů než soukromé školy. Možné důvody pro vyšší počet studentů a úspěšnost na veřejných školách mohou zahrnovat:

- Tradiční povahu a delší historii veřejných škol.
- Širší nabídku studijních programů.
- Nižší finanční náklady díky státní podpoře.
- Vyšší prestiž některých veřejných škol.

Naopak soukromé školy mohou být atraktivní pro studenty hledající specifické obory, individuální přístup a flexibilní studijní podmínky.

Z grafů a tabulek také můžeme odvodit následující poměry:

- Veřejné školy:
  - Poměr absolventů vůči studujícím (2020-2022): přibližně 20.5
  - Poměr absolventů vůči nově zapsaným (2020-2022): přibližně 102
- Soukromé školy:
  - Poměr absolventů vůči studujícím (2020-2022): přibližně 22.7
  - Poměr absolventů vůči nově zapsaným (2020-2022): přibližně 125

### 4.8.4 Shrnutí a závěr hypotézy 5

Na základě výše uvedené analýzy lze konstatovat, že veřejné vysoké školy mají výrazně vyšší počet studentů než soukromé školy. Nicméně, pro přesné vyhodnocení úspěšnosti absolventů nejsou k dispozici potřebná data. Informace o absolventech neobsahují údaje o typech studia (bakalářské, magisterské, doktorské), opakování ročníků ani jiné podrobnosti. Navíc, do roku 2019 nebyla v tabulce F12 sbírána data o veřejných a soukromých školách zvlášť, ale pouze dohromady. Data v tabulce F12 jsou dostupná k jednotlivým školám, ale pro shrnující informaci o všech veřejných a soukromých školách se data začala uvádět až od roku 2020. Tato analýza tedy poskytuje zjednodušený pohled na rozdíly mezi veřejnými a soukromými školami.

Pro plné potvrzení nebo vyvrácení této hypotézy by bylo zapotřebí přesnějších údajů, zejména o jednotlivých oborech a opakování ročníků. Tabulka F12 bohužel neposkytuje všechny potřebné informace, takže by bylo nutné dohledat jiné zdroje dat pro detailnější analýzu.

Například touto problematikou se zabývá aplikace [dropout.pef.czu.cz](https://dropout.pef.czu.cz)<sup>5</sup>. Tato aplikace vznikla v rámci projektu „Stanovení postupu výpočtu ukazatele propadavosti studentů českých vysokých škol“ podpořeného Technologickou

<sup>5</sup><https://dropout.pef.czu.cz/Uspesnost.aspx>

agenturou ČR v letech 2016-2017 na České zemědělské univerzitě v Praze. Provozovatelem aplikace je Česká zemědělská univerzita v Praze, anonymní data obsažená v aplikaci jsou poskytována Ministerstvem školství, mládeže a tělovýchovy z databáze Sdružených informací matrik studentů.[34]

Dalším zdrojem je projekt Česko v datech<sup>6</sup>. Jedná se o unikátní projekt na poli datové žurnalistiky, který na základě analýzy aktuálních dat odhaluje zajímavá a mnoha lidem dosud neznámá fakta o naší zemi. S těmito výstupy následně seznamuje veřejnost a dává je k dispozici médiím. Ta naše analýzy využívají pro vlastní články, pořady a reportáže. Naší ambicí je vyhledávat témata, která jsou přitažlivá nebo důležitá pro celou společnost. Chceme objevit souvislosti, které se v datech skrývají, a potvrdit či vyvrátit tvrzení, která se v české společnosti tradují, ale bez práce s rozsáhlými daty jim chybí důkaz.[35]

## 4.9 Požadavky na náš nástroj

Na základě analýzy a formulace hypotéz jsme identifikovali klíčové požadavky na náš nástroj, které jsou nezbytné pro efektivní datovou analýzu. Tyto požadavky vycházejí z potřeby načítání dat, provádění analýz a základní vizualizace:

- **Nahrávání dat z Excel a CSV souborů:** Uživatelé musí mít možnost snadno nahrát data z různých zdrojů ve formátu Excel (.xls, .xlsx) a CSV<sup>7</sup> (.csv).
- **Vizualizace dat:** Nástroj musí poskytovat možnosti vizualizace dat pomocí grafů a tabulek.
- **Filtrování dat:** Uživatelé musí být schopni filtrovat data podle různých kritérií.
- **Analýza dat:** Nástroj musí umožňovat základní statistické analýzy, jako jsou grafy a statistiky, lineární regrese a korelace.
- **Práce s grafy:** Možnost kombinovat tabulky a ukládat grafy.
- **Uživatelské návody:** Poskytnutí návodů a tutoriálů pro uživatele, aby mohli efektivně využívat nástroj.

Identifikované požadavky zajišťují, že nástroj bude nejen schopný načítat data, ale také provádět jejich analýzu a vizualizaci, čímž poskytne uživatelům komplexní přehled a praktické dovednosti v oblasti datové analytiky. Tyto požadavky budou podrobněji popsány v následující kapitole, která se zaměří na návrh a implementaci nástroje.

<sup>6</sup><https://www.ceskovdatech.cz/graphs/vs2.php>

<sup>7</sup>CSV (Comma-Separated Values) soubor je formát pro ukládání tabulkových dat ve formě prostého textu. Každý řádek v CSV souboru představuje jeden záznam nebo řádek tabulky, a jednotlivé hodnoty v řádku jsou odděleny čárkami nebo jinými oddělovači (například středníky nebo tabulátory).

## ■ 4.10 Závěř kapitoly

V této kapitole jsme ukázali, že na základě dostupných dat je možné provádět datovou analýzu. Formulovali jsme hypotézy, shromáždili relevantní data a na jejich základě tyto hypotézy buď potvrdili, nebo vyvrátili. Demonstrace jednotlivých analýz byla zaměřena také na začátečníky, kterým jsme ukázali metodický přístup k datové analýze. Definovali jsme základní požadavky na náš nástroj, které blíže popíšeme v další kapitole. Tyto analýzy a naučné materiály budou integrovány do aplikace, aby uživatelům poskytly ucelený přehled a praktické dovednosti v oblasti datové analytiky.



## Kapitola 5

### Analýza požadavků a návrh aplikace

Tato kapitola se zaměřuje na analýzu požadavků a návrh aplikace pro datovou analýzu, která umožní uživatelům efektivně pracovat s datovými sadami o studentech a studiích na vysokých školách a provádět různé datové analýzy. Aplikace bude navržena tak, aby sloužila nejen pro analýzu dat, ale také jako naučný program pro uživatele bez hlubokých technických znalostí. Podrobně se věnujeme jak funkčním, tak nefunkčním požadavkům systému, a předkládáme návrh uživatelského rozhraní a architektury aplikace.

#### 5.1 Úvod do návrhu systému

V současné době, kdy objem dat neustále roste a stává se klíčovým prvkem v mnoha oblastech, je nezbytné mít nástroje, které umožňují tato data efektivně zpracovávat a analyzovat. Cílem tohoto projektu je vytvořit systém pro datovou analýzu, který bude schopen zvládnout širokou škálu úloh od základních statistických výpočtů až po složité analytické modelování. Systém bude navržen tak, aby byl uživatelsky přívětivý a umožnil uživatelům různých úrovní dovedností provádět datové analýzy bez potřeby hlubokých technických znalostí.

Aplikace bude zaměřena na práci s datovými sadami o studentech a studiích na vysokých školách, což je klíčové pro vzdělávací instituce a studenty. Kromě toho bude sloužit jako naučný nástroj, který pomůže uživatelům porozumět základním principům datové analýzy a rozvíjet jejich dovednosti v této oblasti.

Primárními uživateli systému budou jednotlivci, kteří potřebují nástroj pro rychlou a intuitivní analýzu dat. Systém bude poskytovat možnosti pro import dat z různých zdrojů, jejich čištění, transformaci, analýzu a vizualizaci. Důraz bude kladen na flexibilitu a rozšiřitelnost, aby systém mohl růst a adaptovat se na měnící se požadavky uživatelů.

#### 5.2 Přejít k požadavkům

Na základě těchto cílů, které byly definovány v úvodu a v kapitole 4, jsou formulovány následující klíčové požadavky na systém:





pomáhá lépe pochopit funkčnost a požadavky systému. Tyto nástroje jsou užitečné pro komunikaci mezi vývojáři a ostatními zainteresovanými stranami v projektu.[36]

Z formulace funkčních a nefunkčních požadavků vychází náš diagram na obrázku 5.1. Tento diagram popisuje funkční požadavky z pohledu uživatele a systému, které jsou detailně popsány v předchozích částech. Funkční požadavky identifikují, jaké činnosti a funkce má systém provádět, aby splnil požadavky uživatelů.

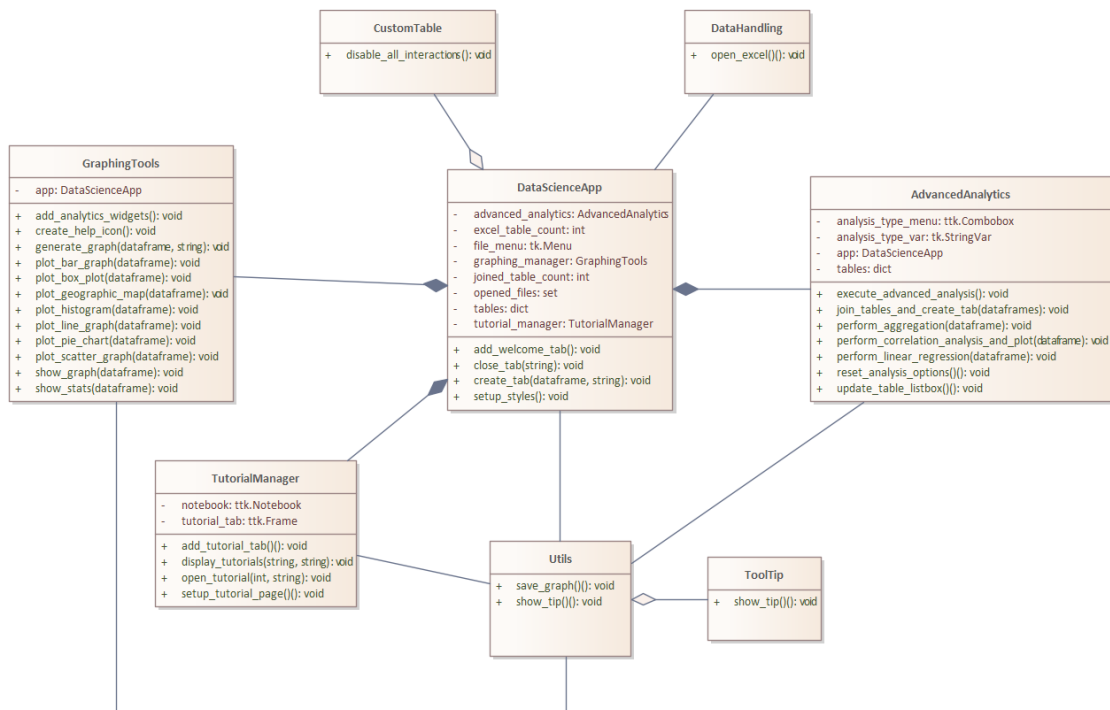


**Obrázek 5.1:** Use Case diagram poskytující přehled o funkcích systému z pohledu uživatelů a systému.

Tento Use Case diagram poskytuje přehled o funkcích systému z pohledu uživatelů a systému a slouží jako základ pro další detailní návrh a implementaci systému [37]. Byl vytvořen pomocí nástroje Enterprise Architect<sup>1</sup>, který je běžně používán pro modelování UML diagramů.

<sup>1</sup><https://sparxsystems.com/>





Obrázek 5.2: Diagram tříd aplikace.

Diagram tříd na obrázku 5.2 zobrazuje strukturu a návrh naší aplikace pro datovou analýzu. Následuje detailní popis klíčových tříd a jejich rolí v systému:

**DataScienceApp** - Jako hlavní třída aplikace přebírá koordinační a řídicí funkce. Odpovídá za inicializaci, správu záložek a ostatních component, a stává se tak centrálním bodem pro interakce s dalšími moduly.

**GraphingTools** - Specializuje se na vytváření statistických vizualizací. Nabízí metody pro generování grafů a statistik, což usnadňuje interpretaci dat.

**AdvancedAnalytics** - Tato třída rozšiřuje možnosti aplikace o pokročilé analytické funkce jako je lineární regrese a korelační analýza.

**TutorialManager** - Spravuje interaktivní návody a scénáře, které nejen usnadňují uživatelům orientaci v aplikaci, ale také pomáhají naučit se základem datové analýzy. Tímto způsobem zvyšuje uživatelský komfort a snižuje čas potřebný na naučení se ovládnutí systému.

**DataHandling** - Podpůrná třída zjednodušující import dat.

**CustomTable** - Umožňuje efektivní prohlížení a analýzu dat v tabulkové formě.

**Utils** - Třída obsahující pomocné funkce, jako jsou ukládání grafů a zobrazování nápověd. Tyto funkce podporují hlavní operace systému a usnadňují běžné úlohy.

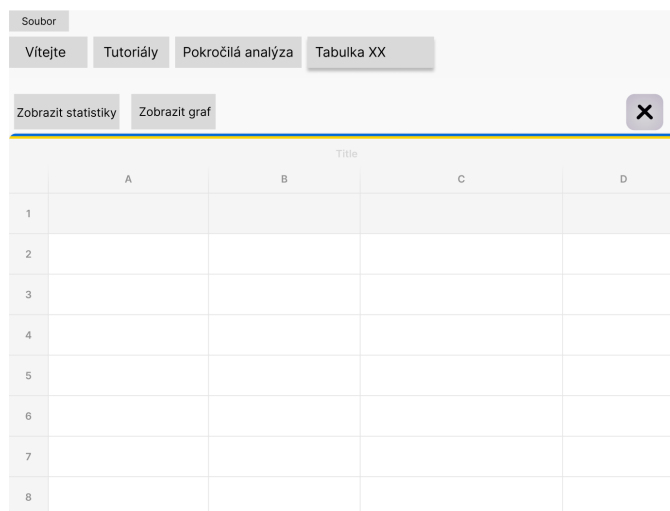
### 5.3.2 Wireframes

Pro návrh wireframů byl použit nástroj Figma<sup>2</sup>, který umožňuje vytvářet low fidelity návrhy. Low fidelity wireframy jsou základní náčrty uživatelského rozhraní, které se zaměřují na rozložení a funkčnost, aniž by se zabývaly grafickým designem nebo detaily. Tyto návrhy jsou užitečné pro rychlou iteraci a diskusi o základních prvcích a struktuře uživatelského rozhraní.

Wireframy, které jsou představeny v této sekci na obrázcích 5.3 až 5.7, jsou navrženy s ohledem na klíčové funkční požadavky systému a zaměřují se na jednoduchost rozvržení, aby byly co nejvíce uživatelsky přívětivé. Tato základní vizuální reprezentace uživatelského rozhraní ukazuje, jak by mohly jednotlivé komponenty aplikace vypadat a fungovat.

Použití wireframů v této fázi návrhu vychází z potřeby rychle iterovat a testovat různé návrhy rozvržení a funkčnosti uživatelského rozhraní. Vycházeli jsme také z analýzy existujících nástrojů a snahy, aby naše aplikace byla intuitivní a snadno použitelná pro uživatele.

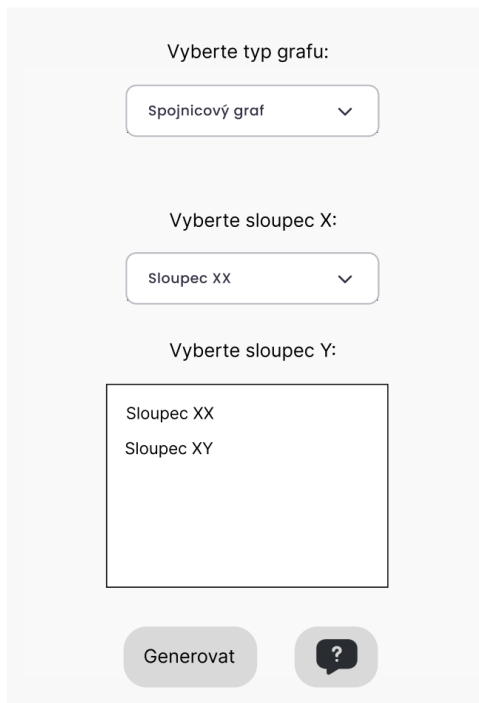
První wireframe (obrázek 5.3) ukazuje záložku s otevřenou tabulkou, kde budou data po nahrání z Excelu nebo CSV. Uživatel zde uvidí data, bude si je moci seřadit nebo upravit, a poté po kliknutí na tlačítko „Zobrazit graf“ nebo „Zobrazit statistiky“ vytvořit různé typy grafů nebo zobrazit základní statistiku.



Obrázek 5.3: Wireframe zobrazení tabulky a základních funkcí.

<sup>2</sup>[urlhttps://www.figma.com/](https://www.figma.com/)

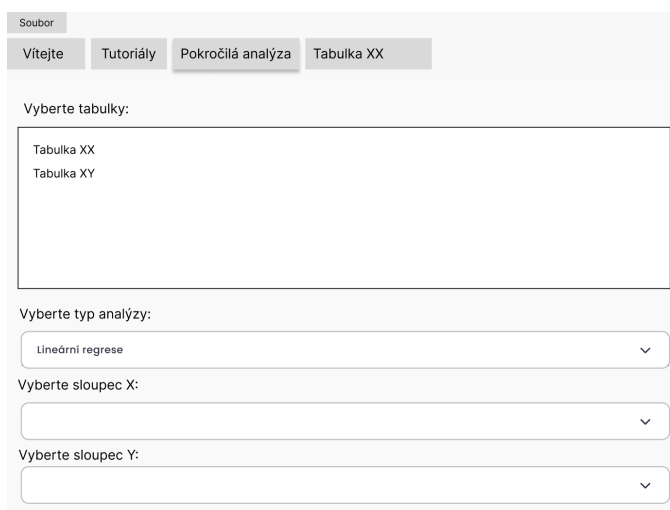
Druhý wireframe (obrázek 5.4) ukazuje popup okno pro tvorbu grafu, kde si uživatel vybere typ grafu a následně sloupce pro osu X a Y. Po výběru může generovat graf nebo zobrazit nápovědu ke grafům.



The wireframe shows a vertical form for creating a graph. It starts with the label 'Vyberte typ grafu:' followed by a dropdown menu with 'Spojnicový graf' selected. Below this is 'Vyberte sloupec X:' with a dropdown menu showing 'Sloupec XX'. The next section is 'Vyberte sloupec Y:' with a list box containing 'Sloupec XX' and 'Sloupec XY'. At the bottom, there are two buttons: 'Generovat' and a help icon (a speech bubble with a question mark).

**Obrázek 5.4:** Wireframe vytváření grafu.

Třetí wireframe (obrázek 5.5) ukazuje záložku pokročilé analýzy. Uživatel zde uvidí všechny aktuálně nahrané soubory, na kterých může provádět pokročilejší analýzy jako lineární regrese a korelace nebo spojovat tabulky. V závislosti na výběru akce se mu bude upravovat volba sloupců.



The wireframe shows a web interface for advanced analysis. At the top, there is a navigation bar with tabs: 'Soubor', 'Vítejte', 'Tutoriály', 'Pokročilá analýza', and 'Tabulka XX'. The main content area is titled 'Vyberte tabulky:' and contains a list box with 'Tabulka XX' and 'Tabulka XY'. Below this is 'Vyberte typ analýzy:' with a dropdown menu showing 'Lineární regrese'. The next two sections are 'Vyberte sloupec X:' and 'Vyberte sloupec Y:', each with an empty dropdown menu.

**Obrázek 5.5:** Wireframe zobrazení záložky pokročilé analýzy.



### ■ 5.4.1 Volba programovacího jazyka

Pro tento projekt byl vybrán **Python**, což je vysokoúrovňový programovací jazyk, který je široce preferován pro svou čitelnost, jednoduchost a obrovský ekosystém knihoven. Python je ideální pro rychlý vývoj a prototypování, což je zásadní v dynamickém prostředí datové analýzy. Jak bylo uvedeno v úvodní části 2.1.2, Python je vhodný nástroj pro datovou analýzu díky své schopnosti rychle zpracovávat data a poskytovat rozsáhlé možnosti vizualizace a analýzy dat.[39]

Výběr jazyka jsem především omezil na **Python** a **Java**, jelikož s těmito jazyky mám nejvíce zkušeností. Kromě těchto hlavních možností jsem zvážil i další alternativy, ale nakonec byl preferován Python, protože jsme ho používali v předmětu Základy datových analýz na naší univerzitě. Python se ukázal být obzvláště vhodný pro manipulaci s daty a jejich analýzu, což bylo klíčové pro úspěch tohoto projektu.[40]

*Výhody:*

- Velká komunita a podpora.
- Rozsáhlé knihovny pro vědecké výpočty a datovou analýzu.
- Jednoduchý syntax, který zjednodušuje učení a vývoj.

Tento výběr zajišťuje, že vývoj bude nejen efektivní, ale i přístupný díky široce dostupným zdrojům a komunitní podpoře, což umožňuje rychlé řešení problémů a integraci nových funkcí.

*Alternativy[41]:*

- **R**: Speciálně zaměřen na statistické výpočty, ale s menší flexibilitou v oblasti obecného softwarového vývoje.
- **Java**: Vhodnější pro velké, distribuované systémy, ale s větší složitostí vývoje a pomalejším prototypováním.
- **C++**: Vysoce výkonný jazyk vhodný pro systémové programování a aplikace vyžadující intenzivní výpočty, ale s komplexnějším syntaxem a vyššími nároky na správu paměti.





### ■ 5.4.3 Volba dalších knihoven

Projekt využívá širokou škálu Python knihoven, které poskytují robustní nástroje pro datovou analýzu a vizualizaci.[43]

- **Pandas:** Nezbytná pro manipulaci s daty a analýzu.
- **NumPy:** Základ pro numerické výpočty v Pythonu.
- **Matplotlib a Seaborn:** Pro výkonné vizualizace dat.
- **Scikit-learn:** Pro lineární regrese.
- **GeoPandas a Contextily:** Pro práci s geografickými daty a mapami.
- **GeoPy:** Pro geokódování světových lokací z názvů měst.
- **Pillow (PIL):** Pro zpracování obrazu.

*Výhody:*

- Tyto knihovny jsou dobře udržované a mají silnou komunitní podporu.
- Společně poskytují silný základ pro téměř jakoukoli potřebu datové analýzy a vizualizace.

*Alternativy:*

- Pro datovou analýzu mohou být alternativou balíčky v R, jako jsou **ggplot2** pro vizualizace a **dplyr** pro manipulaci s daty.
- Pro interaktivní vizualizace může být alternativou **Plotly**.

Tyto technologické volby poskytují robustní základ pro vývoj našeho projektu, zajišťují efektivitu vývoje, udržitelnost a rozšiřitelnost aplikace.



# Kapitola 6

## Vývoj aplikace

V této kapitole se věnuji procesu implementace aplikace, která byla navržena na základě požadavků a návrhu uvedených v předchozích kapitolách. Popíši použité technologie a programovací nástroje, které byly zvoleny pro vývoj backendové a frontendové části aplikace, a dále se zaměřím na architekturu a klíčové komponenty systému.

### 6.1 Použité technologie

Vývoj aplikace byl založen na programovacím jazyku **Python**, což bylo zvoleno kvůli jeho vynikající podpoře pro různé knihovny, které jsou klíčové pro datovou analýzu a zpracování dat, jak bylo popsáno v předchozí kapitole 5.4. Pro backend aplikace byly využity specifické Python knihovny, které umožňují efektivní manipulaci s daty a jejich analýzu. Mezi hlavní použité knihovny patří **Pandas** pro manipulaci s daty, **NumPy** pro numerické výpočty, a **Matplotlib** a **Seaborn** pro vizualizace dat.

Pro vývoj frontendu byl použit **Python** s grafickými knihovnami **Tkinter** a **TTK**, které poskytují nástroje pro tvorbu desktopových aplikací s grafickým uživatelským rozhraním. Tyto knihovny byly zvoleny pro jejich jednoduchost a efektivitu ve vytváření standardních GUI komponent, což usnadňuje rychlý vývoj a implementaci uživatelského rozhraní.

Celý vývojový proces byl spravován pomocí verzovacího systému **Git**<sup>1</sup>, s kódem hostovaným na platformě **GitLab**<sup>2</sup>, což umožňuje efektivní správu verzí a spolupráci na projektu. Použití GitLabu také podporuje kontinuální integraci a nasazování, což je zásadní pro udržení kvality kódu a efektivitu vývojových cyklů.

Pro vývoj bylo využito vývojové prostředí **JetBrains PyCharm**<sup>3</sup>, které poskytuje rozsáhlé nástroje pro psaní Python kódu, ladění, testování a správu projektů. PyCharm byl zvolen pro jeho pokročilé funkce, které značně zvyšují produktivitu vývoje, včetně inteligentního dokončování kódu, analýzy kvality kódu a integrace s verzovacími systémy.

<sup>1</sup><https://git-scm.com>

<sup>2</sup><https://gitlab.com>

<sup>3</sup><https://www.jetbrains.com/pycharm/>

Tato kombinace technologií byla zvolena pro jejich osvědčené schopnosti v oblasti rychlého vývoje, spolehlivosti a podpory pro realizaci požadovaných funkcionalit aplikace.

## 6.2 Architektura aplikace

Aplikace využívá architektonický vzor, který je inspirován modelem Model-View-Controller (MVC), ale byl upraven tak, aby lépe vyhovoval potřebám a specifikám datové analýzy. Tento přizpůsobený přístup umožňuje efektivnější manipulaci s daty a jejich vizualizaci.

### 6.2.1 Modifikace MVC architektury

Tradiční MVC vzor odděluje logiku aplikace (kontroler), uživatelské rozhraní (pohled) a datovou vrstvu (model) do striktně oddělených komponent.[44] V našem případě tuto roli zčásti plní třída **DataScienceApp**, která funguje převážně jako hlavní kontroler, ale zároveň obsahuje prvky modelu a view. Tato třída zajišťuje manipulaci s daty a jejich zobrazování v různých uživatelských záložkách, což umožňuje centralizované řízení aplikace.

Třídy jako **GraphingTools**, **TutorialManager**, a **AdvancedAnalytics** jsou primárně zaměřeny na vizualizaci dat (view), ale také zahrnují určité aspekty modelu, protože přímo manipulují s daty nebo obsahují logiku pro interpretaci datových setů.

Tato struktura a rozdělení funkcionalit jsou vizuálně reprezentovány v class diagramu, který je zobrazen na obrázku 5.2. Diagram ukazuje, jak jednotlivé komponenty interagují a spolupracují k dosažení funkcionalit aplikace.

## 6.3 Realizace komponent

Každá třída v diagramu představuje specifickou část aplikace a má své unikátní odpovědnosti, které jsou propojeny s požadavky na systém a zajišťují efektivní fungování aplikace.

### 6.3.1 DataScienceApp

**DataScienceApp** funguje jako centrální kontrolní jednotka celé aplikace. Je navržena tak, aby sloužila jako hlavní kontroler, koordinující všechny hlavní operace mezi uživatelským rozhraním a datovými moduly. Tato třída drží reference na všechny ostatní moduly, jako jsou **GraphingTools**, **AdvancedAnalytics**, a **TutorialManager**, a řídí jejich interakce a správu dat.

## ■ Klíčové funkce

Mezi klíčové funkce patří:

- **Inicializace a správa modulů:** **DataScienceApp** inicializuje všechny potřebné moduly při spuštění aplikace a udržuje jejich reference pro snadné volání jejich funkcí.
- **Vytváření a zavírání záložek:** Umožňuje dynamické vytváření a zavírání záložek v uživatelském rozhraní, což zvyšuje flexibilitu a uživatelský komfort při práci s aplikací.
- **Delegace operací:** Jako hlavní kontroler deleguje **DataScienceApp** specifické úkoly ostatním modulům, například vyžádání generování grafu od **GraphingTools** nebo zahájení analýzy od **AdvancedAnalytics**.

## ■ Výzvy a rozhodnutí

Při implementaci bylo klíčové správně rozhodnout, jaká data a podpory potřebují jednotlivé moduly pro jejich efektivní funkci. **DataScienceApp** musí efektivně zpracovávat a distribuovat data mezi moduly, zajišťujíc tím, že každý modul má přístup k informacím potřebným pro jeho operace. Hlavní výzvou bylo zajištění modularity a rozšiřitelnosti bez narušení celkové stability aplikace.

## ■ Integrace s ostatními moduly

**DataScienceApp** slouží jako hlavní most mezi uživatelským rozhraním a backendovými operacemi, zajišťujíc, že všechny komponenty aplikace mohou efektivně komunikovat a synchronizovat své činnosti. Tato integrace je klíčová pro hladký běh aplikace a její schopnost reagovat na uživatelské vstupy v reálném čase.

### ■ 6.3.2 GraphingTools

Komponenta **GraphingTools** představuje jednu z nejsložitějších částí aplikace, zaměřující se na generování různých typů grafů a statistických vizualizací. Tato třída zahrnuje široké spektrum funkcí od jednoduchých lineárních grafů až po složité geografické vizualizace.

## ■ Modularizace a rozvoj

Původní implementace byla založena na jedné centrální funkci, která řídila generování všech typů grafů. Tento přístup vedl k nepřehlednosti kódu a komplikoval údržbu a rozvoj aplikace. V reakci na tuto situaci byl kód přepracován do modularnější struktury, kde každý typ grafu má nyní svou dedikovanou metodu. Tyto metody jsou následně koordinovány prostřednictvím jedné nadřazené metody, která zajišťuje správný výběr a generování požadovaného grafu v závislosti na uživatelském vstupu.

## ■ Výzvy s geografickými grafy

Zejména realizace geografických grafů představovala významnou výzvu. Bylo nutné implementovat funkcionalitu, která umožňuje převod názvů měst na geografické souřadnice zeměpisné šířky a délky, formát vyžadovaný knihovnou **pyplot** pro správnou vizualizaci. Dále se ukázalo jako náročné najít vhodnou knihovnu, která by poskytla mapové podklady, na které mohou být data efektivně vykreslena. Původní implementace bez této integrace vedla k zobrazení grafů na bílém pozadí bez jakýchkoli mapových orientačních bodů, což snižovalo informativní hodnotu vizualizací.

V rámci této implementace jsem nejprve zkoušel využít knihovnu **Plotly** a její funkci `scatter_map`, avšak narazil jsem na několik problémů. Mapa nevykreslovala jednotlivé země zvlášť, ale spíše zobrazovala celý svět nebo velké regiony, což nebylo vhodné pro vizualizaci měst. Tato omezení vedla k nedostatečné úrovni detailů grafů.

Následně jsem se rozhodl použít knihovnu **tkintermapview**<sup>4</sup>, která umožňuje snadné vykreslování bodů na mapě přímo v aplikaci Tkinter. Nicméně, tato knihovna neumožňuje vytváření bublin, pouze bodů. I když toto nebylo ideální, zjistil jsem, že by bylo lepší mít interaktivnější a vizuálně přitažlivější funkci.

Proto jsem přešel na knihovnu **Folium**, která umožňuje vytváření map v HTML a nabízí pokročilé možnosti vizualizace, včetně bublin a interaktivních prvků. Nicméně integrace HTML výstupu do aplikace Tkinter se ukázala jako problematická, protože aplikace nedokázala správně zobrazit HTML soubor přímo v rozhraní Tkinter. Toto vedlo k prázdnému nebo bílému oknu.

Jako nejrozsáhlejší řešení jsem nakonec zvolil otevření vygenerované mapy ve webovém prohlížeči. Toto řešení, i když není ideální, poskytuje nejlepší funkčnost a uživatelskou přívětivost z pohledu kvality a přehlednosti vizualizovaných dat. Uživatel má možnost interagovat s mapou přímo v prohlížeči, což zajišťuje lepší zážitek a přehled o geografických datech.

Díky využívání bezplatného API pro geokódování trvá převod geografických dat delší dobu, což způsobuje prodlevu při načítání dat do aplikace. Alternativní možnosti zahrnovaly použití **Google Geocoding API** nebo offline databáze jako SimpleMaps<sup>5</sup>. Google API je však placené a verze zdarma od SimpleMaps obsahuje pouze zlomek měst, což by omezilo funkcionalitu aplikace.

## ■ Optimalizace a údržba

Rozdělení funkcionality tvorby grafů do jednotného kontroleru a specifických metod pro každý typ grafu usnadnilo údržbu kódu, jeho organizaci a testování. Tento přístup také zjednodušil integraci nových typů grafů a nástrojů do komponenty **GraphingTools**, což vedlo k významnému zlepšení její funkčnosti.

<sup>4</sup><https://github.com/TomSchimansky/TkinterMapView>

<sup>5</sup><https://simplemaps.com/data/world-cities>

### 6.3.3 AdvancedAnalytics

Komponenta **AdvancedAnalytics** je zásadní pro provádění pokročilých analytických operací, jako jsou regresní analýzy a korelace. Tato část aplikace je navržena tak, aby uživatelům poskytovala hlubší statistické náhledy a umožňovala jim vytvářet složitější modely datových vztahů.

#### Hlavní funkce

- **Regresní analýzy:** Umožňují uživatelům modelovat vztahy mezi proměnnými a predikovat hodnoty závislých proměnných.
- **Korelační analýzy:** Poskytují metody pro výpočet statistických korelací, což pomáhá identifikovat vztahy mezi různými veličinami.

#### Výzvy a řešení

Jednou z klíčových výzev při implementaci statistických funkcí bylo zajištění, že systém správně zpracovává a vizualizuje data. Bylo zásadní implementovat robustní kontrolu vstupů, aby se předešlo situacím, kdy by nečíselné nebo nesprávně formátované hodnoty mohly narušit analytické procesy. Tato kontrola je zavedena, protože práce s nečíselnými hodnotami je komplexní a vyžaduje pokročilé metody pro správné zpracování dat. Řešení tedy spočívá ve filtraci vstupních dat tak, aby byly do analýz přijímány pouze validní a správně formátované hodnoty, což značně zjednodušuje další analytické zpracování a zvyšuje spolehlivost výsledků.

#### Využití

**AdvancedAnalytics** představuje klíčový prvek aplikace, který značně rozšiřuje její analytické schopnosti. Tato komponenta poskytuje uživatelům mocné nástroje pro rozbor a interpretaci dat, zatímco zůstává přístupná a snadno použitelná.

### 6.3.4 TutorialManager

**TutorialManager** je klíčová komponenta aplikace, která se zabývá správou a zobrazením návodů a scénářů. Jejím hlavním účelem je usnadnit uživatelům orientaci v aplikaci a poskytnout jim praktické návody na její používání.

#### Správa obsahu

Texty k návodům a scénářům jsou uloženy ve vlastním souboru, což umožňuje snadnější správu a údržbu těchto materiálů. Tato centralizace obsahu zjednodušuje aktualizace a změny, jelikož veškeré úpravy se provádějí na jednom místě, což zvyšuje efektivitu a snižuje riziko nekonzistence v návodech.

## ■ Funkce a realizace

**TutorialManager** načítá texty návodů a scénářů při spuštění aplikace a poskytuje uživatelům přístup k těmto informacím přes uživatelské rozhraní.

## ■ Výzvy při implementaci

Obsahová náročnost byla jednou z hlavních výzev při vývoji **TutorialManager**. Příprava textů a podkladů, které byly nezbytné pro návody a scénáře, vyžadovala pochopení datové analýzy a funkcionality aplikace. Texty musely být srozumitelné, informativní a užitečné, aby skutečně přispívaly ke vzdělávání uživatelů.

## ■ Využití

**TutorialManager** zlepšuje použitelnost aplikace tím, že novým a stávajícím uživatelům poskytuje návody k používání aplikace. Tato komponenta hraje zásadní roli v procesu výuky základům datové analýzy.

## ■ 6.3.5 Další Komponenty Aplikace

Kromě hlavních analytických a vizualizačních modulů, aplikace obsahuje několik dalších klíčových komponent, které zajišťují její funkčnost a uživatelský komfort.

### ■ DataHandling

Komponenta **DataHandling** je zodpovědná za import a validaci dat. Tato třída umožňuje uživatelům nahrávat data z excelovských tabulek a CSV souborů a zajišťuje jejich validaci, aby byla zajištěna jejich správnost a konzistence. Efektivní správa dat je základem pro správnou funkcionality celé aplikace.

### ■ Utils

Komponenta **Utils** obsahuje soubor pomocných funkcí, které jsou využívány napříč různými moduly aplikace. Tyto funkce zahrnují například ukládání grafů, formátování dat a další operace, které jsou nezbytné pro správnou funkci a udržitelnost aplikace. Umístění těchto funkcí do samostatného modulu zvyšuje modularitu a udržitelnost kódu.

### ■ ToolTip

Třída **ToolTip** je navržena tak, aby poskytovala kontextové nápovědy uživatelům během používání aplikace. Tato funkce je klíčová pro zvýšení uživatelské přívětivosti a pomáhá novým uživatelům rychleji se seznámit s funkcemi aplikace. Nápovědy jsou zobrazovány dynamicky v závislosti na kontextu, ve kterém se uživatel nachází.



## ■ CustomTable

Komponenta **CustomTable** snižuje možnosti práce s tabulkami v knihovně **PandasTable** jako například filtraci dat. Nicméně, díky **CustomTable** je uživatel schopen přesouvat sloupce v tabulce, řadit tabulku podle abecedy v případě textových buněk nebo podle hodnot v případě číselných buněk, a to buď od nejvyšší po nejnižší, nebo od nejnižší po nejvyšší. Tato komponenta však neumožňuje pokročilejší filtraci dat, jako je selektování specifických hodnot nebo aplikace složitějších filtračních kritérií.

## ■ 6.4 Shrnutí implementace

V rámci implementace bylo úspěšně dosaženo většiny z počátečně stanovených cílů projektu. Aplikace nyní umožňuje nahrávání dat, jejich zobrazení v tabulkové formě, provádění statistických analýz a ukládání výsledků analýz ve formě grafů.

Byly však zaznamenány značné obtíže při pokusu o implementaci dynamického filtrování dat přímo v aplikaci. Kvůli omezením v čase a technickým výzvám spojeným s integrací této funkcionality do stávajícího systému bylo rozhodnuto, že tato funkce nebude přidána. Technické výzvy zahrnovaly složitost dokumentace funkcionalit tabulek v knihovně Pandas, kde se nedařilo správně propojit různé funkce, což vedlo k nesprávnému řazení dat a dalším chybám. To vyžaduje, aby uživatelé prováděli předzpracování a filtrování dat samostatně před jejich nahráním do aplikace. Ačkoliv to není ideální, aplikace i bez této funkce funguje bez větších problémů.

Dalším významným omezením, se kterým se aplikace potýká, je rychlost a optimalizace při zpracování geografických dat. Problémy s latencí, které se objevují při načítání velkých geografických dat z externích knihoven, mohou negativně ovlivnit uživatelský zážitek, zejména při manipulaci s rozsáhlými datovými sadami. Alternativní variantou bylo využití offline databáze, avšak bezplatná verze obsahuje pouze několik hlavních měst, což není využitelné. Placená verze, například z SimpleMaps<sup>6</sup>, se pohybuje od 200 do 500 dolarů. Další možností bylo využití Google Geo API<sup>7</sup>, které si účtuje poplatky na základě počtu volání. Při malém využití do 100 000 žádostí měsíčně to vychází na 5 dolarů za 1000 žádostí, a při větším využívání cena za 1000 volání klesá. Vzhledem k tomu, že se jedná o jednoduchý nástroj, toto omezení není zásadní a aplikace zůstává funkční a užitečná.

Navíc bylo zavedeno omezení na načítání maximálně pěti excelovských a CSV souborů současně. Toto omezení bylo implementováno, aby se zajistilo, že uživatelé nepřekročí paměťový limit jejich počítačů, což by mohlo vést k nestabilitě nebo pádu aplikace pod velkým objemem dat. Tato opatření pomáhají udržovat stabilitu a spolehlivost aplikace i při práci s velkými sady dat.

<sup>6</sup><https://simplemaps.com/data/world-cities>

<sup>7</sup><https://developers.google.com/maps/documentation/geocoding/usage-and-billing>

Tyto komponenty dohromady tvoří robustní základ, který podporuje hlavní analytické a vizualizační funkce aplikace a zároveň zvyšuje její celkovou uživatelskou přívětivost a flexibilitu. Vzhledem k tomu, že aplikace je zaměřena na základní analytické úkony, uvedená omezení zásadně neovlivňují její celkovou funkčnost a použitelnost.

# Kapitola 7

## Testování a validace

Tato kapitola se zaměřuje na testování a validaci systému pro zajištění jeho kvality a spolehlivosti. Kombinujeme end-to-end (E2E) testování s manuálním uživatelským testováním, aby byla ověřena funkčnost systému. Uživatelé procházejí scénáři pokrývající hlavní funkce aplikace, jako je nahrávání dat, analýza, vizualizace a export výsledků. Testování je založeno na hypotézách uvedených v kapitole 4 a zahrnuje i testování pokročilých analýz.

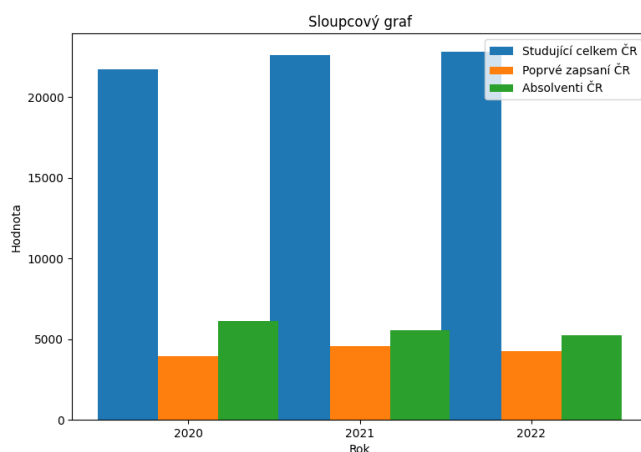
Výsledky testování poskytují klíčové informace o uživatelské zkušenosti a identifikovaných chybách, které jsou následně řešeny. Tato kapitola rovněž porovnává výsledky naší aplikace s výsledky z Power BI, aby byla ověřena její přesnost a spolehlivost.

### 7.1 Testovací strategie

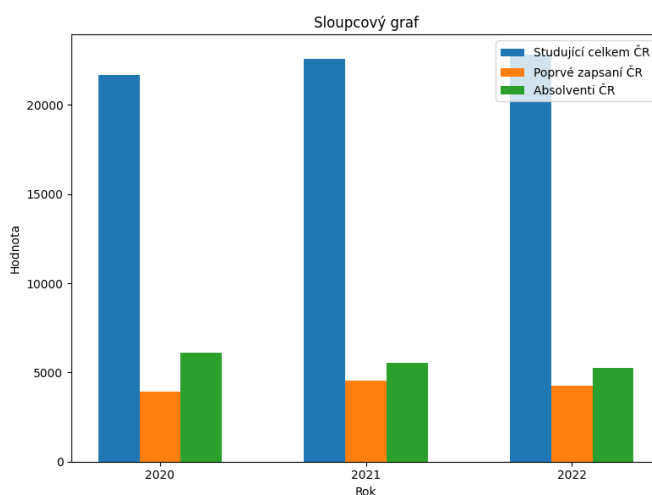
End-to-end (E2E) testování je technika, která testuje celý softwarový produkt od začátku do konce, aby bylo zajištěno, že aplikační tok se chová podle očekávání a že všechny integrované části pracují společně, jak mají. Hlavním účelem E2E testování je simulace reálných uživatelských scénářů a validace systému z pohledu koncového uživatele. Softwarové systémy jsou složité a propojené s mnoha subsystémy, a pokud některý z těchto subsystémů selže, může to způsobit selhání celého systému. Toto riziko lze minimalizovat pomocí end-to-end testování, které zajišťuje integritu a spolehlivost celého systému [45].

Naše testovací strategie je založena na kombinaci end-to-end testování a manuálního uživatelského testování. Automatizované testy nebyly prováděny, protože výstupy aplikace jsou především grafy a tabulky, které bylo jednodušší kontrolovat ručně. Místo toho jsme prováděli manuální uživatelské testy, při kterých skuteční uživatelé prošli celou aplikaci podle vytvořených scénářů pro datovou analýzu a poté měli možnost aplikaci prozkoumat podle svého uvážení. Tento přístup nám umožnil konzultovat funkčnost, design a další aspekty přímo s uživateli, což poskytlo cennou zpětnou vazbu a umožnilo odhalit a opravit potenciální problémy v systému, což vedlo ke zlepšení celkové kvality a uživatelské přívětivosti aplikace.





Obrázek 7.1: Sloupcový graf před změnou



Obrázek 7.2: Sloupcový graf po změně

Proběhla také úprava některých textů v návodech. Více chyb při testování nebylo odhaleno.

## 7.3 Validace hypotéz pomocí vlastní aplikace

V této části se zaměříme na ověření, zda naše aplikace dokáže vytvořit stejné nebo podobné výsledky jako Power BI. Použijeme dříve stanovené hypotézy a provedeme jejich analýzu v naší aplikaci. Výsledky z naší aplikace budou následně porovnány s výsledky získanými pomocí Power BI, aby se ověřila přesnost a spolehlivost naší aplikace. Nezaměříme se zde na obsah samotné analýzy, který byl podrobně popsán v kapitole 4, ale na vygenerované obrázky grafů a tabulek vytvořené naší aplikací a na závěr každé hypotézy popíšeme,

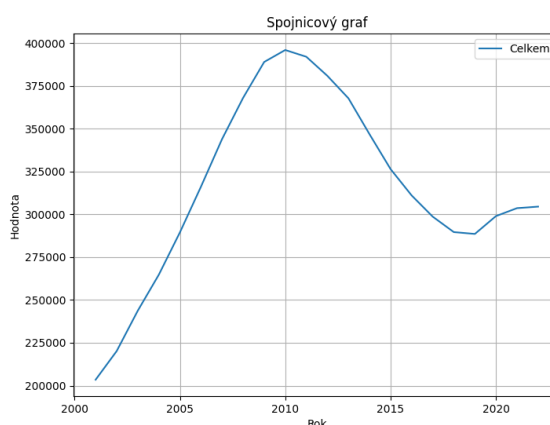
zda se podařilo dosáhnout podobných grafů, tabulek a zobrazení jako v Power BI, nebo zda se v něčem výrazně liší.

Analýza dat vychází ze stejných tabulek, které byly použity v kapitole 4. Tím je zajištěno, že porovnáváme stejné datové sady a můžeme přesně posoudit schopnosti naší aplikace ve srovnání s Power BI.

### 7.3.1 Hypotéza 1: Trendy ve vývoji počtu studentů v čase

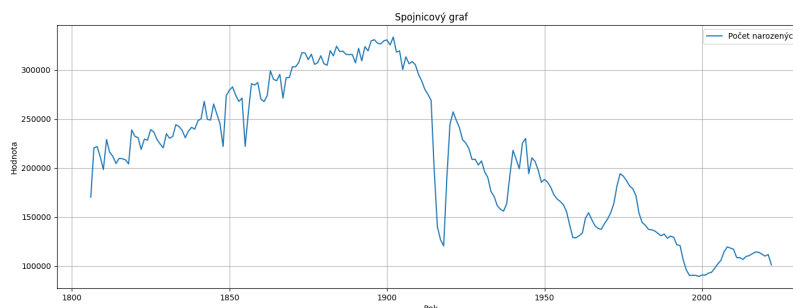
#### Grafy a tabulky vytvořené naší aplikací

Na obrázku 7.3 je zobrazen počet studentů vysokých škol v letech 2001 až 2022, který byl vytvořen pomocí naší aplikace. Tento graf ukazuje, že počet studentů rostl do roku 2010, kdy dosáhl vrcholu, a následně začal klesat. Od roku 2018 můžeme pozorovat mírný nárůst, který pokračuje až do roku 2022.



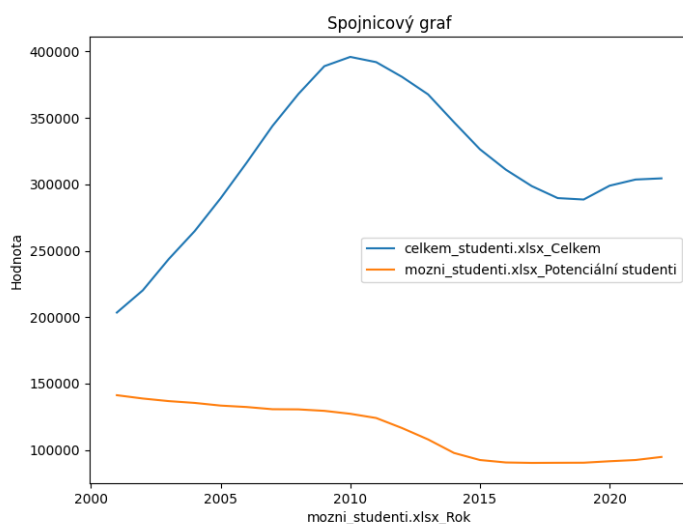
Obrázek 7.3: Vývoj počtu studentů vysokých škol v letech 2001-2022

Na obrázku 7.4 je zobrazen počet narozených dětí v daných letech, vytvořený pomocí naší aplikace. Tento graf poskytuje kontext pro predikci možného počtu potenciálních studentů.



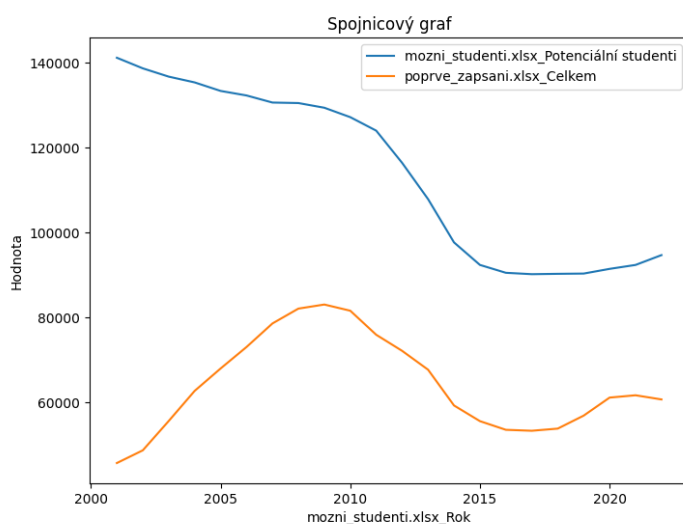
Obrázek 7.4: Počet narozených dětí v daných letech

Na obrázku 7.5 je porovnán celkový počet studentů s odhadem potenciálních nových studentů, založeným na demografických datech. Tento graf byl vytvořen pomocí naší aplikace.



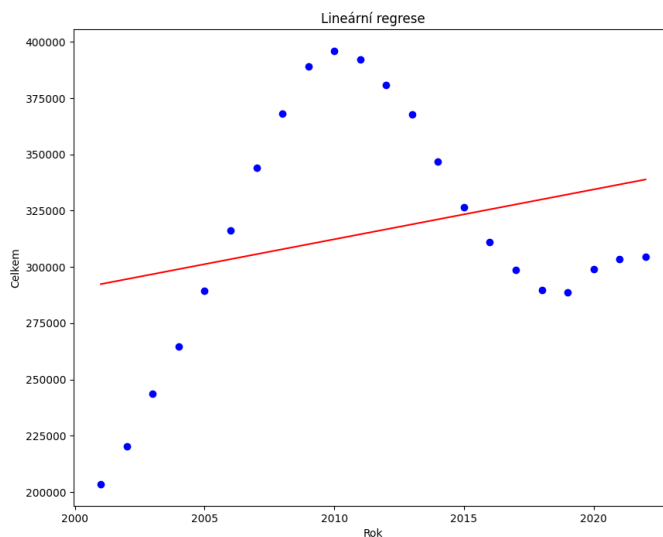
**Obrázek 7.5:** Celkový počet studentů vs. potenciální noví studenti

Na obrázku 7.6 je zobrazen počet potenciálních nových studentů ve srovnání s počtem studentů, kteří se poprvé zapsali do vysoké školy. Tento graf byl vytvořen pomocí naší aplikace.



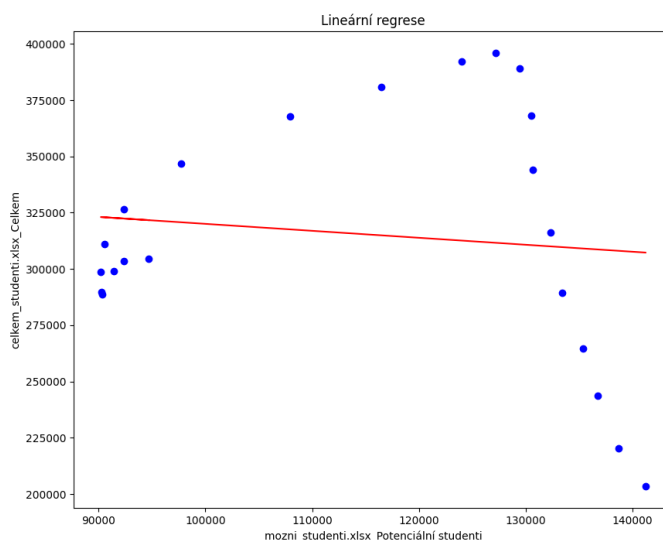
**Obrázek 7.6:** Potenciální noví studenti vs. nově zapsaní studenti

Na obrázku 7.7 je zobrazen graf s lineární regresní čarou, která ukazuje celkový trend mezi rokem a celkovým počtem studentů. Tento graf byl vytvořen pomocí naší aplikace.



**Obrázek 7.7:** Regresní analýza počtu studentů vysokých škol

Na obrázku 7.8 je zobrazen graf regresní analýzy potenciálních nových studentů, který byl vytvořen pomocí naší aplikace.



**Obrázek 7.8:** Regresní analýza potenciálních nových studentů

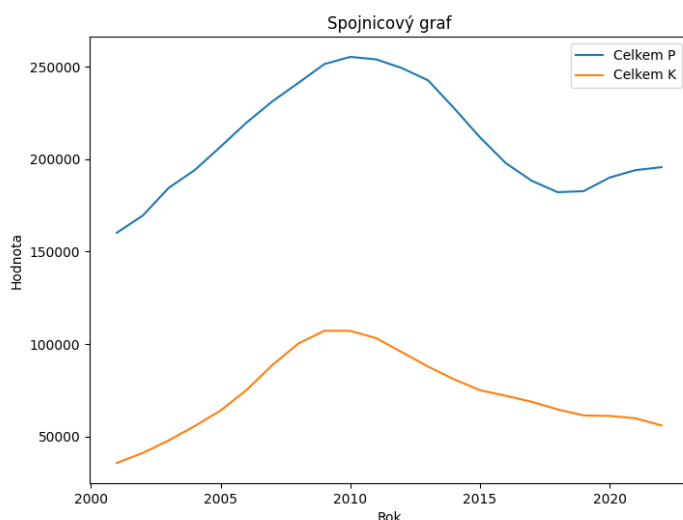


## ■ Porovnání s Power BI

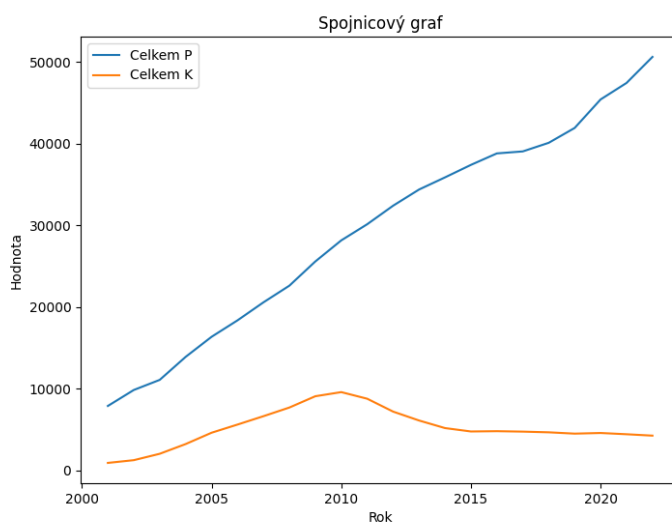
Grafy a tabulky vytvořené naší aplikací vykazují stejné trendy a vzorce jako grafy vytvořené v Power BI. Hlavní trendy, jako je nárůst nebo pokles počtu studentů v průběhu let, byly zachyceny s vysokou přesností v obou nástrojích. Mírné rozdíly v detailech mohou být způsobeny různými metodami vizualizace dat, ale celkově jsou výsledky velmi podobné. To potvrzuje, že pomocí naší aplikace jsme schopni replikovat hypotézu 1.

### ■ 7.3.2 Hypotéza 2: Vztah mezi národností studenta a typem studia

Na obrázcích 7.9 a 7.10 je zobrazen počet studentů ve formě prezenčního a kombinovaného studia rozdělený podle českých a zahraničních studentů. Použili jsme spojnicové grafy pro vizualizaci rozdílů mezi těmito dvěma skupinami studentů.

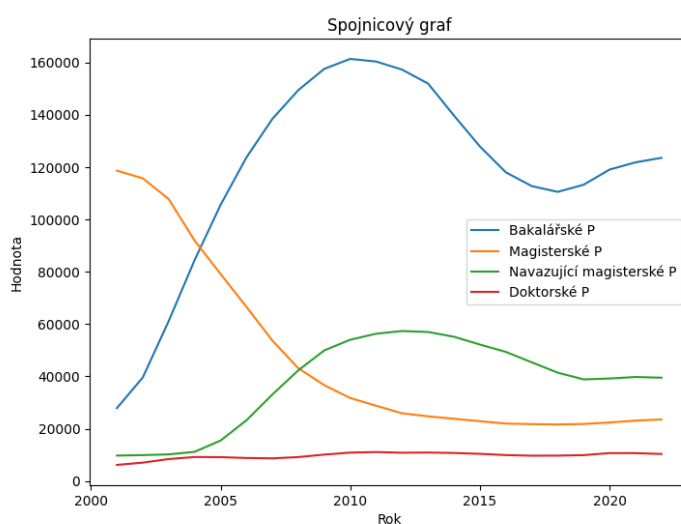


**Obrázek 7.9:** Počet českých studentů ve formě prezenčního a kombinovaného studia

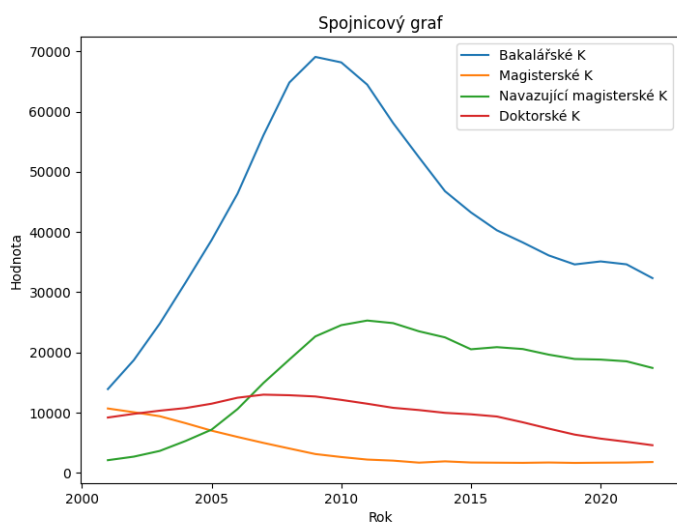


**Obrázek 7.10:** Počet zahraničních studentů ve formě prezenčního a kombinovaného studia

Na obrázcích 7.11 a 7.12 jsou zobrazeny detaily o počtu českých studentů v jednotlivých stupních studia (bakalářské, magisterské, navazující magisterské a doktorské) podle formy studia.

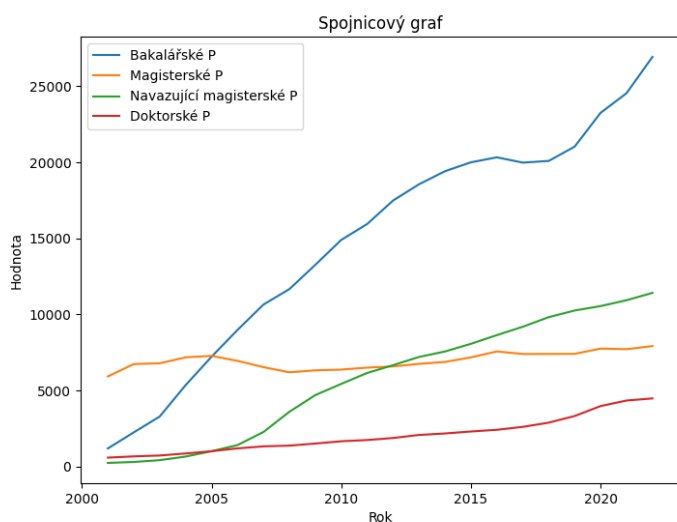


**Obrázek 7.11:** Počet českých studentů ve formě prezenčního studia podle stupně studia

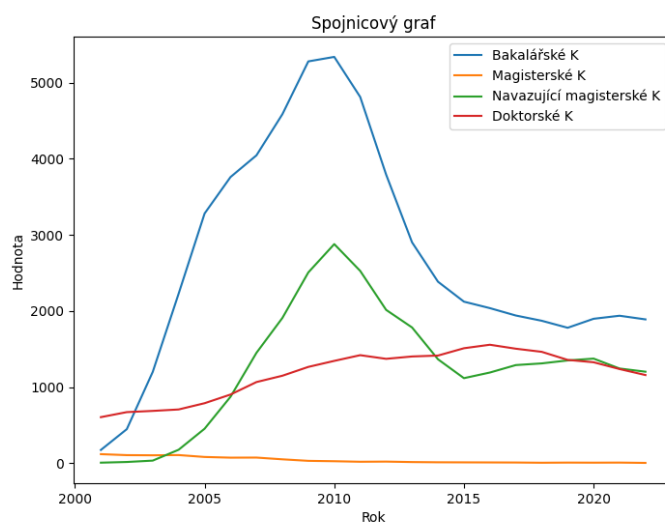


**Obrázek 7.12:** Počet českých studentů ve formě kombinovaného studia podle stupně studia

Na obrázcích 7.13 a 7.14 jsou zobrazeny detaily o počtu zahraničních studentů v jednotlivých stupních studia podle formy studia.

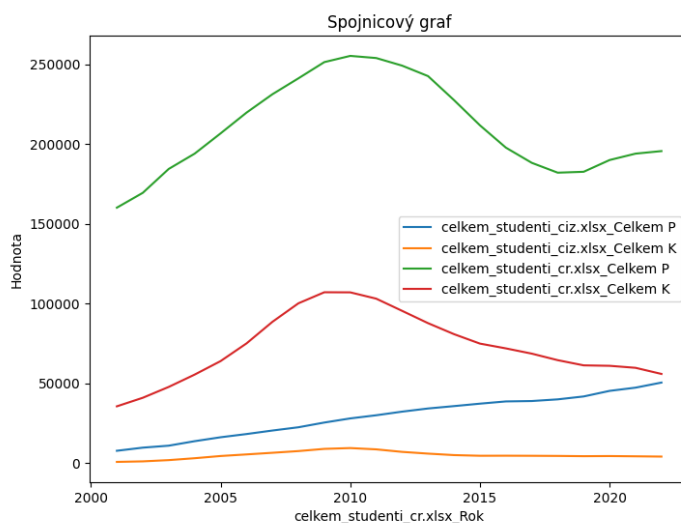


**Obrázek 7.13:** Počet zahraničních studentů ve formě prezenčního studia podle stupně studia



**Obrázek 7.14:** Počet zahraničních studentů ve formě kombinovaného studia podle stupně studia

Na obrázku 7.15 je zobrazen počet studentů ve formě prezenčního a kombinovaného studia pro české i zahraniční studenty dohromady.



**Obrázek 7.15:** Počet studentů ve formě prezenčního a kombinovaného studia (čeští a zahraniční studenti dohromady)

## ■ Porovnání s Power BI

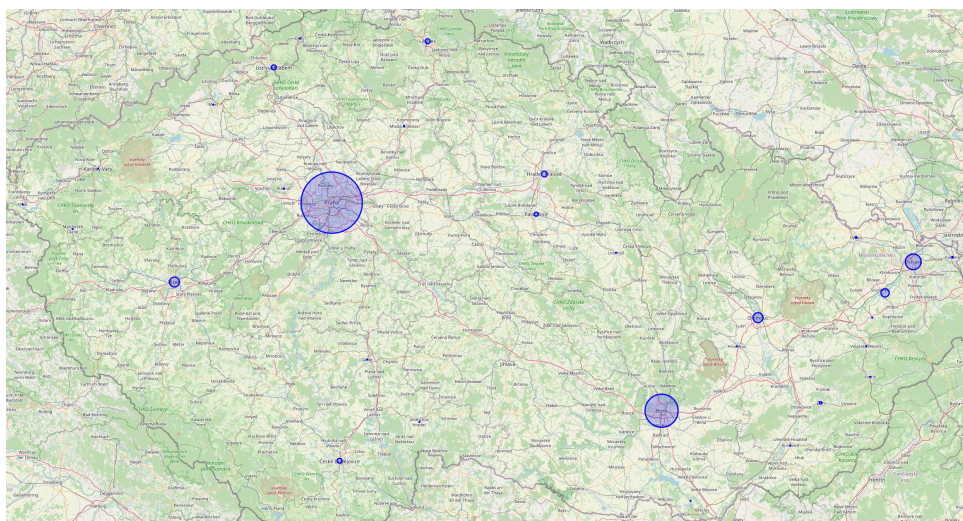
Při porovnání grafů vytvořených v Power BI a v naší aplikaci je patrné několik rozdílů a podobností:

- Grafy z aplikace vykazují podobné trendy jako grafy z Power BI. Například počet zahraničních studentů ve formě prezenčního studia vykazuje vzestupný trend, což je konzistentní v obou sadách grafů.
- V aplikaci jsou grafy prezentovány s čistšími a jednoduššími vizualizacemi, což může usnadnit interpretaci výsledků. Naopak Power BI poskytuje více možností pro detailní úpravy a interaktivitu.
- Některé rozdíly v prezentaci dat mohou být způsobeny odlišným přístupem ve volbě vizualizačních technik. Například v některých grafech jsou mírné odlišnosti ve škálování os a použití barev.

Celkově však obě platformy poskytují konzistentní výsledky, které podporují hypotézu, že zahraniční studenti preferují prezenční formu studia ve srovnání s českými studenty.

## ■ 7.3.3 Hypotéza 3: Geografické rozložení studentů

Na obrázku 7.16 je znázorněno geografické rozložení studentů v roce 2001.



**Obrázek 7.16:** Geografické rozložení studentů v roce 2001

Podrobnější pohled na počet studentů v jednotlivých městech v roce 2001 poskytuje obrázek 7.17.

Město	Celkem	Bakalářsk	Magisters	Doktorské	Poprvé za	Absolventi	Cizinci	Rok
Olomouc	12248	2000	9317	1004	2479	1838	669	2001
Opava	1579	817	743	25	313	295	44	2001
Kladno	2	2	0	0	2	0	0	2001
Mladá Boleslav	130	130	0	0	58	0	2	2001
Vsetín	13	12	1	0	0	4	0	2001
České Budějovice	5832	1466	4148	239	1294	1113	44	2001
Valašské Meziříčí	265	265	0	0	108	0	3	2001
Jindřichův Hradec	903	296	617	0	198	297	33	2001
Kunovice	537	537	0	0	244	85	36	2001
Tábor	241	0	241	0	61	5	4	2001
Cheb	628	541	87	0	267	156	10	2001
Mariánské Lázně	43	43	0	0	0	13	2	2001
Praha	74198	11232	56557	7485	13927	10691	4677	2001
Plzeň	12008	2442	9189	506	3051	1839	439	2001
Ústí nad Labem	5575	2061	3530	16	1424	923	22	2001
Ostrava	18516	3956	13493	1131	4723	2711	302	2001
Karlovy Vary	291	291	0	0	154	0	1	2001
Pardubice	4605	1912	2395	304	1229	648	45	2001
Děčín	213	213	0	0	79	17	1	2001
Liberec	5915	1398	4420	260	1286	704	344	2001
Most	492	254	239	0	158	55	0	2001
Albrechtův Týn	0	0	0	0	0	0	1	2001
Hradec Králové	6735	2051	4388	332	1501	1293	251	2001
Litomyšl	59	59	0	0	7	0	4	2001
Brno	39735	7256	29032	3849	8879	5843	1672	2001
Lednice	478	117	323	38	35	133	15	2001
Zlín	3229	2209	895	133	1214	583	108	2001
Prostějov	69	69	0	0	33	0	0	2001
Karviná	1683	184	1504	0	385	297	62	2001
Nový Jičín	22	22	0	0	22	0	0	2001

**Obrázek 7.17:** Počet studentů v jednotlivých městech v roce 2001

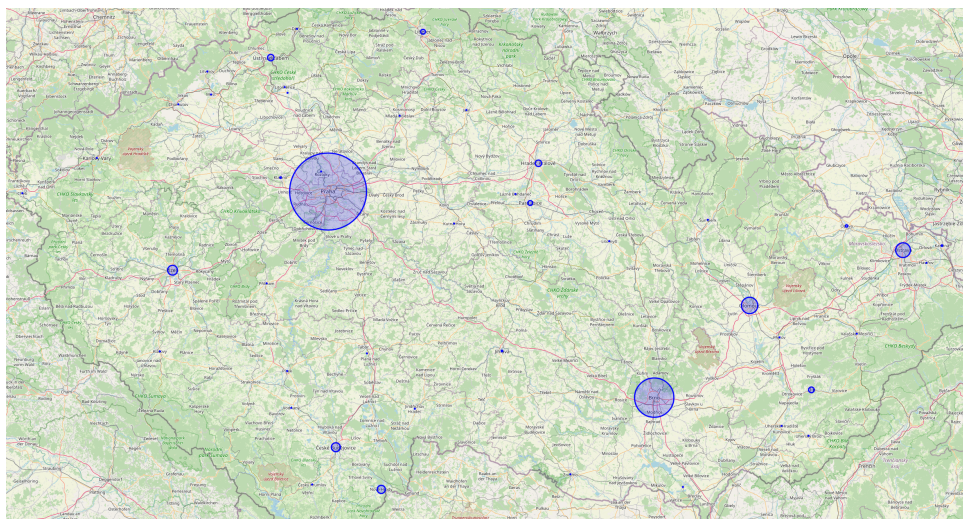
Obrázek 7.18 zobrazuje počet studentů v prvních deseti městech v roce 2001.

	Město	Celkem	Bakalářsk	Magisters	Doktorské	Poprvé za	Absolventi	Cizinci	Rok
1	Praha	74198	11232	56557	7485	13927	10691	4677	2001
2	Brno	39735	7256	29032	3849	8879	5843	1672	2001
3	Ostrava	18516	3956	13493	1131	4723	2711	302	2001
4	Olomouc	12248	2000	9317	1004	2479	1838	669	2001
5	Plzeň	12008	2442	9189	506	3051	1839	439	2001
6	Hradec Králové	6735	2051	4388	332	1501	1293	251	2001
7	Liberec	5915	1398	4420	260	1286	704	344	2001
8	České Budějovice	5832	1466	4148	239	1294	1113	44	2001
9	Ústí nad Labem	5575	2061	3530	16	1424	923	22	2001
10	Pardubice	4605	1912	2395	304	1229	648	45	2001

**Obrázek 7.18:** Počet studentů v prvních deseti městech v roce 2001



Podobně je na obrázku 7.19 zobrazeno geografické rozložení studentů v roce 2022.



**Obrázek 7.19:** Geografické rozložení studentů v roce 2022

Obrázek 7.20 ukazuje počet studentů v jednotlivých městech v roce 2022.

Město	Celkem	Bakalářsk	Magisters	Doktorské	Poprvé za	Absolventi	Cizinci	Rok
Olomouc	20383	11402	8041	1070	3066	3689	2650	2022
Opava	2734	2265	414	60	606	469	187	2022
Přerov	223	143	80	0	26	82	89	2022
Šumperk	149	115	34	0	64	53	0	2022
Kladno	1649	1107	401	142	246	356	128	2022
Kutná Hora	14	14	0	0	0	30	2	2022
Mladá Boleslav	692	548	144	0	105	157	116	2022
Husinec	2	0	0	2	0	2	0	2022
Příbram	139	139	0	0	56	65	0	2022
České Budějovice	10617	8142	2183	308	2507	2020	885	2022
Nové Hradky	0	0	0	0	0	1	1	2022
Valašské Meziříčí	59	59	0	0	0	25	0	2022
Český Krumlov	67	30	37	0	3	11	0	2022
Jindřichův Hradec	876	397	464	17	124	149	163	2022
Písek	129	91	38	0	29	55	22	2022
Vodňany	7	0	0	7	0	4	23	2022
Sezimovo Ústí	49	0	49	0	0	24	1	2022
Cheb	179	119	60	0	56	7	4	2022
Praha	94132	56487	31329	6981	16798	18389	27767	2022
Pízeň	11818	6268	4965	637	2470	2191	1409	2022
Ústí nad Labem	7532	5366	1929	247	1490	1153	346	2022
Ostrava	17879	11723	5222	1005	3650	3686	2136	2022
Karlový Vary	361	259	102	0	70	53	8	2022
Havířov	75	64	11	0	6	57	4	2022
Paroubice	5985	4381	1354	252	1339	1169	517	2022
Ostrov	1	1	0	0	0	0	0	2022
Klatovy	90	28	62	0	13	42	0	2022
Děčín	83	73	10	0	19	10	9	2022
Chomutov	1	1	0	0	0	3	0	2022
Liberec	5637	4047	1376	217	1220	951	503	2022
Litoměřice	175	65	110	0	23	100	0	2022
Terežín	80	80	0	0	43	50	2	2022
Most	541	378	163	0	134	140	3	2022
Litvínov	106	99	7	0	26	19	0	2022
Hradec Králové	8214	4521	3310	416	1664	1621	1326	2022
Jičín	85	28	57	0	16	43	0	2022
Lázně Bohdaneč	16	16	0	0	0	1	2	2022
Litomyšl	82	53	25	4	15	17	14	2022
Brno	47832	27875	17142	3238	9417	10648	14829	2022
Lednice	550	361	150	39	39	131	121	2022
Zlín	6794	4641	1906	252	1501	1464	968	2022
Jihlava	1973	1942	31	0	575	301	130	2022
Uherské Hradiště	1171	826	345	0	290	270	38	2022
Uherský Brod	2	2	0	0	0	1	0	2022
Znojmo	98	96	2	0	5	40	1	2022
Karviná	2119	1617	495	9	531	316	146	2022
zahraničí: Slovensko	3	2	1	0	2	0	111	2022
zahraničí: Švýcarská konfederace	0	0	0	0	0	0	152	2022
zahraničí: neuvedeno	1	1	0	0	0	0	99	2022

**Obrázek 7.20:** Počet studentů v jednotlivých městech v roce 2022





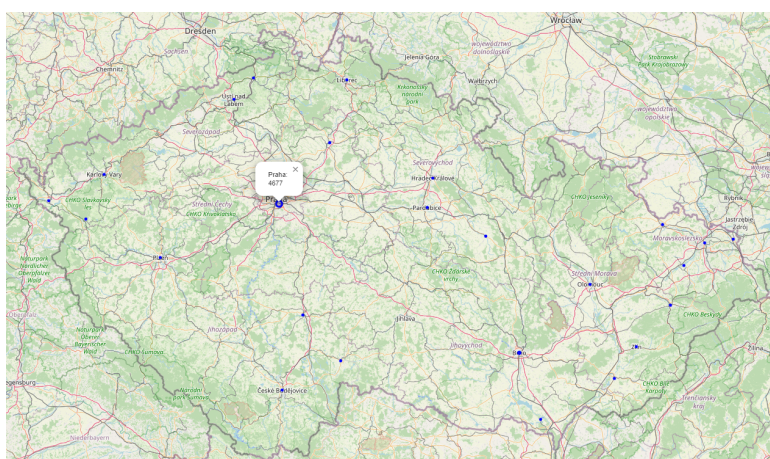
### 7.3.4 Hypotéza 4: Geografické rozložení zahraničních studentů v České republice

Obrázek 7.22 zobrazuje top 10 měst v roce 2001 podle počtu zahraničních studentů.

	Město	Cizinci	Rok
1	Praha	4677	2001
2	Brno	1672	2001
3	Olomouc	669	2001
4	Plzeň	439	2001
5	Liberec	344	2001
6	Ostrava	302	2001
7	Hradec Králové	251	2001
8	Zlín	108	2001
9	Karviná	62	2001
10	Pardubice	45	2001

**Obrázek 7.22:** Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2001

Obrázek 7.23 znázorňuje geografické rozložení zahraničních studentů v roce 2001.



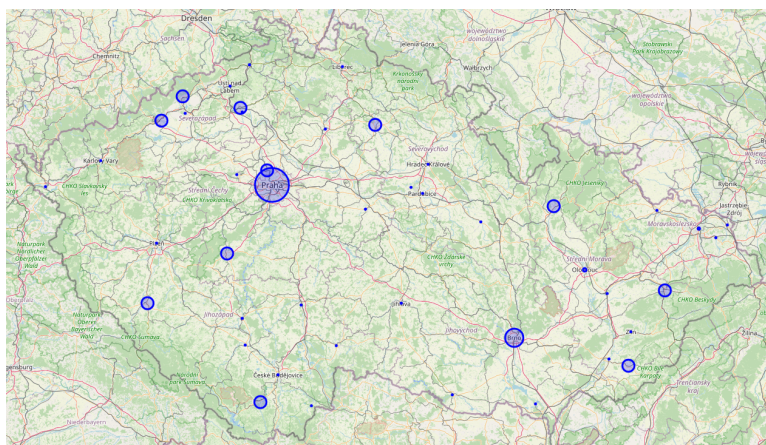
**Obrázek 7.23:** Geografické rozložení zahraničních studentů v roce 2001

Obrázek 7.24 zobrazuje top 10 měst v roce 2022 podle počtu zahraničních studentů.

	Město	Cizinci	Rok
1	Praha	27767	2022
2	Brno	14829	2022
3	Olomouc	2650	2022
4	Ostrava	2136	2022
5	Plzeň	1409	2022
6	Hradec Králové	1326	2022
7	Zlín	968	2022
8	České Budějovice	885	2022
9	Pardubice	517	2022
10	Liberec	503	2022

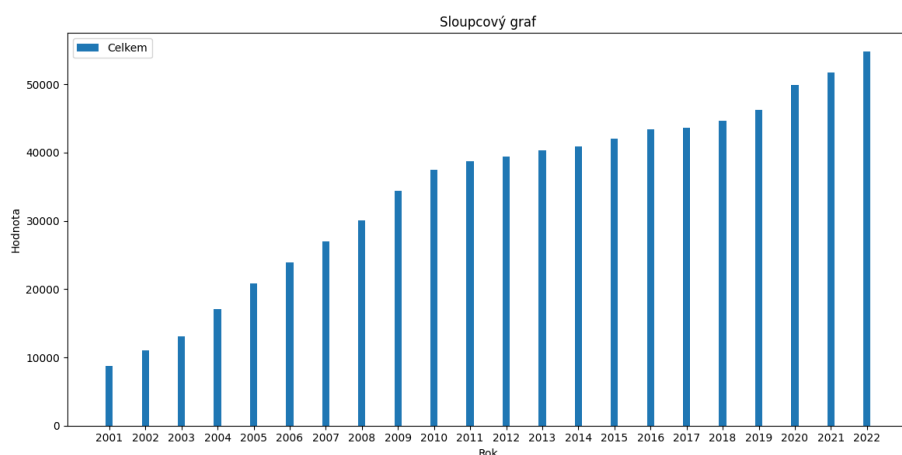
**Obrázek 7.24:** Top 10 měst s nejvyšším počtem zahraničních studentů v roce 2022

Obrázek 7.25 znázorňuje geografické rozložení zahraničních studentů v roce 2022.



**Obrázek 7.25:** Geografické rozložení zahraničních studentů v roce 2022

Obrázek 7.26 ukazuje vývoj počtu zahraničních studentů v České republice od roku 2001 do roku 2022.



**Obrázek 7.26:** Vývoj počtu zahraničních studentů v letech 2001 až 2022

## ■ Porovnání s Power BI

Při porovnání vizualizací mezi Power BI a naší aplikací se ukázalo, že data jsou identická, což potvrzuje správné fungování naší aplikace. Všechny grafy a tabulky obsahují stejné hodnoty a trendy, což svědčí o konzistentnosti a spolehlivosti datových zdrojů a analytických postupů používaných v obou nástrojích.

Co se týká geografických grafů, byly zaznamenány stejné rozdíly jako ty, které jsou zmíněny v Hypotéze 7.3.3. To znamená, že Praha a Brno dominují v počtu zahraničních studentů jak v roce 2001, tak v roce 2022, s významným nárůstem v počtu studentů v průběhu let. Tento trend je v obou nástrojích zobrazen konzistentně, což potvrzuje správnost a přesnost naší aplikace při vizualizaci geografického rozložení studentů.

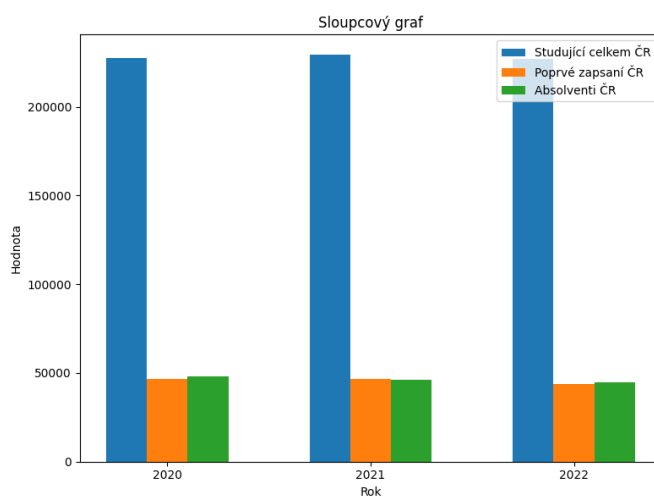
## ■ 7.3.5 Hypotéza 5: Rozdíl mezi soukromými a veřejnými školami

Na obrázku 7.27 je tabulka, která zobrazuje data o počtu studentů na veřejných vysokých školách.

	Cel	Cel	Studující celkem ČR	Poprvé zapsaní ČR	Absolventi ČR	Studu	Pop	Abs	Přer	Jino	Sam	Rok
1	272	150	227440	46707	47920	446	953	77	104	138	8919	2020
2	276	153	229314	46512	46114	469	954	78	113	74	9569	2021
3	276	153	226967	43979	44797	496	111	77	119	63	1025	2022

**Obrázek 7.27:** Data o počtu studentů na veřejných vysokých školách

Obrázek 7.28 znázorňuje počet studentů na veřejných vysokých školách. Tento graf umožňuje vizuálně sledovat vývoj počtu studentů v průběhu času.



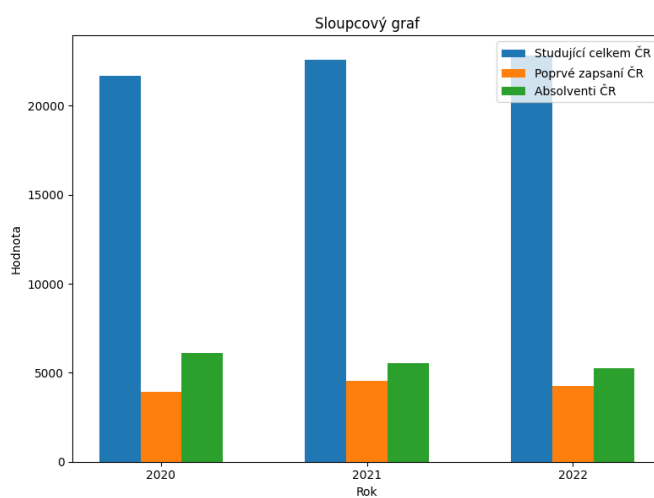
**Obrázek 7.28:** Počet studentů na veřejných vysokých školách

Na obrázku 7.29 je tabulka, která zobrazuje data o počtu studentů na soukromých vysokých školách.

	Cel	Cel	Studující celkem ČR	Poprvé zapsaní ČR	Absolventi ČR	Studu	Popr	Absol	Přer	Ji	Sar	Rok
1	268	149	21701	3940	6125	520	967	112	226	0	0	2020
2	273	152	22583	4552	5553	480	109	105	260	0	0	2021
3	278	156	22800	4238	5237	508	130	930	241	0	0	2022

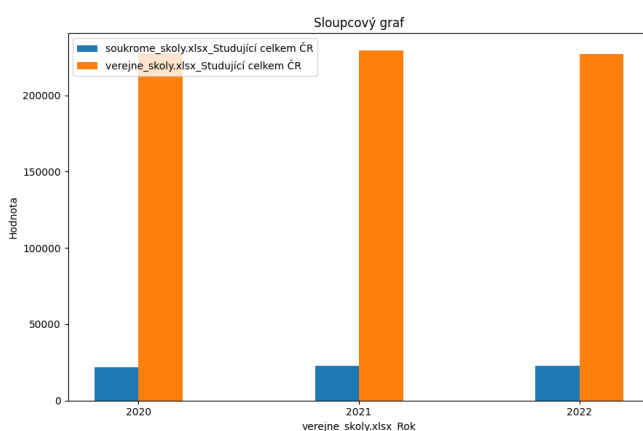
**Obrázek 7.29:** Data o počtu studentů na soukromých vysokých školách

Obrázek 7.30 znázorňuje počet studentů na soukromých vysokých školách.



**Obrázek 7.30:** Počet studentů na soukromých vysokých školách

Obrázek 7.31 zobrazuje srovnání počtu studentů na veřejných a soukromých školách.



**Obrázek 7.31:** Srovnání počtu studentů na veřejných a soukromých školách

## ■ Porovnání s Power BI

Při porovnání vizualizací mezi Power BI a naší aplikací se ukázalo, že data jsou identická, což potvrzuje správné fungování naší aplikace.

Všechny grafy a tabulky obsahují stejné hodnoty a trendy, což svědčí o konzistentnosti a spolehlivosti datových zdrojů a analytických postupů používaných v obou nástrojích.

### ■ 7.3.6 Vyhodnocení rozdílů

Výsledky grafů a tabulek z naší aplikace byly shodné s výsledky generovanými v Power BI, což potvrzuje správnost a spolehlivost naší aplikace při zpracování a vizualizaci dat. Nicméně, byly zaznamenány některé rozdíly v prezentaci grafů:

- **Barvy:** Naše aplikace používá odlišné barevné schéma ve srovnání s Power BI.
- **Číselná ohraničení:** Nastavení číselných ohraničení na osách grafů se lišilo mezi oběma nástroji.

Tyto rozdíly však nemají vliv na interpretaci výsledků. Naše aplikace správně zpracovává data a poskytuje spolehlivé vizualizace, které odpovídají výsledkům z Power BI.

Naše aplikace se osvědčila jako spolehlivý nástroj pro analýzu a vizualizaci dat, s výsledky konzistentními s Power BI. I když má omezené možnosti pro úpravy grafů, stále poskytuje přesné a srozumitelné vizualizace, což z ní činí užitečný nástroj pro základní a středně pokročilé datové analýzy.

Naše aplikace má být jednoduchým nástrojem, takže ani nemá konkurovat Power BI s možnostmi úprav grafů a dalších pokročilých funkcí.



# Kapitola 8

## Diskuse

Tato kapitola se zaměřuje na zhodnocení výsledků dosažených při vývoji aplikace, identifikaci omezení a problémů, se kterými jsme se setkali, a návrhy na zlepšení do budoucna. Kapitola poskytuje přehled o tom, jak úspěšně byla aplikace vytvořena, jaké oblasti potřebují další práci a jaké kroky by měly být podniknuty pro optimalizaci a rozšíření funkcionality aplikace.

### 8.1 Hodnocení výsledků

Aplikaci se v rámci možností podařilo úspěšně vytvořit, ačkoli má některé nedostatky. Podařilo se implementovat základní funkcionality, jako je nahrávání dat, jejich zobrazení v tabulce, základní třídění sloupců, vytváření grafů a provádění složitějších analytických operací, jako je lineární regrese a korelace. Tyto výsledky lze exportovat do obrázků, což uživatelům umožňuje snadné sdílení a prezentaci analýz.

Výukové materiály integrované do aplikace poskytují uživatelům základní přehled o datové analýze a mohou sloužit jako první krok pro ty, kteří se chtějí této oblasti věnovat hlouběji. Aplikace tak úspěšně funguje jako mezystupeň mezi jednoduchými nástroji, jako je Excel, a složitějšími analytickými aplikacemi nebo nástroji.

### 8.2 Omezení a problémy

Během vývoje a testování byly identifikovány některé omezení a problémy:

- **Chybějící pokročilý filtr dat:** Aplikace postrádá pokročilé možnosti filtrování dat. Uživatelé musí data předzpracovávat a filtrovat externě před jejich nahráním do aplikace, což může být pro některé uživatele nepohodlné.
- **Rychlost zpracování geografických grafů:** Načítání a vykreslování geografických grafů je pomalé kvůli použití veřejného API a neparalelizaci aplikace. Při načítání velkých datových sad aplikace zamrzne, což znemožňuje uživateli provádět jiné úkony, dokud se data nenačtou.

- **Nepřehlednost v některých grafech:** Uživatelé si stěžovali na nepřehlednost sloupcových grafů, kde se při více vstupních veličinách hodnoty vykreslovaly příliš blízko sebe. Tento problém byl částečně vyřešen oddálením hodnot, což zlepšilo čitelnost grafů.
- **Design aplikace:** Design aplikace byl hodnocen rozporuplně. Některým uživatelům přišel jednoduchý a přehledný, jiným až příliš strohý a nehezký. Pokusy o vylepšení designu nebyly uspokojujivé, a tak byl ponechán v původní verzi.
- **Nedostatek zpětné vazby:** Bohužel jsme nestihli vyzkoušet aplikaci s větším počtem uživatelů. To omezuje možnost plně zhodnotit, zda aplikace opravdu splňuje potřeby širokého spektra uživatelů a zda je schopna u uživatele vzbudit zájem o datovou analýzu, nebo mu ukázat, že datová analýza pro něj není ta správná cesta.

### 8.3 Návrhy na zlepšení

Na základě zjištěných nedostatků a uživatelské zpětné vazby jsou navrhována následující vylepšení:

- **Pokročilé filtrování dat:** Implementace pokročilejších možností filtrování a manipulace s daty přímo v aplikaci by výrazně zvýšila její užitnou hodnotu.
- **Optimalizace výkonu:** Paralelizace aplikace a optimalizace načítání geografických grafů by mohly výrazně zlepšit rychlost a plynulost aplikace při práci s velkými datovými sadami.
- **Zlepšení designu:** Přehodnocení designu aplikace s ohledem na moderní uživatelské rozhraní by mohlo zvýšit její atraktivitu a použitelnost.
- **Rozšíření uživatelských testů:** Provedení rozsáhlejších uživatelských testů s větším počtem respondentů by poskytlo cenné informace o tom, jak dobře aplikace plní svůj účel a jaké další úpravy by mohly být potřebné. Důraz by měl být kladen na to, zda aplikace skutečně je schopna pomoci uživateli rozhodnout, zda má zájem o datovou analýzu nebo naopak ne.
- **Rozšíření funkcionality:** Přidání dalších analytických funkcí a možností vizualizace by mohlo aplikaci přiblížit k více pokročilým analytickým nástrojům, aniž by ztratila svou jednoduchost a přehlednost.

Celkově lze říci, že aplikace splnila většinu stanovených cílů a poskytuje uživatelům užitečný nástroj pro základní datovou analýzu. Díky výukovým materiálům obsaženým v aplikaci mohou uživatelé získat základní představu o datové analýze a rozhodnout se, zda se chtějí v této oblasti dále vzdělávat a využívat pokročilejší nástroje.



Pro ověření dlouhodobé užitečnosti a efektivity aplikace by bylo nutné provést další testování a získat více zpětné vazby od širšího spektra uživatelů. Teprve poté bude možné definitivně zhodnotit, zda aplikace skutečně naplňuje svůj potenciál jako nástroj pro začínající datové analytiky.

## 8.4 Možnosti využití aplikace

Naše aplikace má široké možnosti využití v různých oblastech. Primárně byla navržena pro analýzu dat týkajících se vysokých škol, což zahrnuje:

- **Analýza dat z vysokých škol:** Aplikace umožňuje zpracování a vizualizaci dat o studentech, studijních programech a dalších relevantních informacích. To je užitečné pro plánování kapacit, hodnocení efektivity studijních programů a další strategické rozhodování.
- **Výukový materiál:** Aplikace může sloužit jako nástroj pro výuku základů datové analýzy. Jedinci, kteří se chtějí naučit pracovat s daty, mohou pomocí naší aplikace získat praktické zkušenosti a základy, které jim umožní přejít na pokročilejší nástroje, jako je Power BI nebo Tableau.

## 8.5 Další možnosti využití aplikace

Aplikace má potenciál být užitečným nástrojem pro různé další oblasti mimo vzdělávání a analýzu dat o studentech.

### 8.5.1 Analýza prodeje pro malé podnikatele

Malí podnikatelé, kteří nemají přístup k složitým analytickým nástrojům, mohou využít aplikaci pro analýzu prodejních dat. Na základě excelovské tabulky s daty o prodeji a cenách zboží mohou podnikatelé:

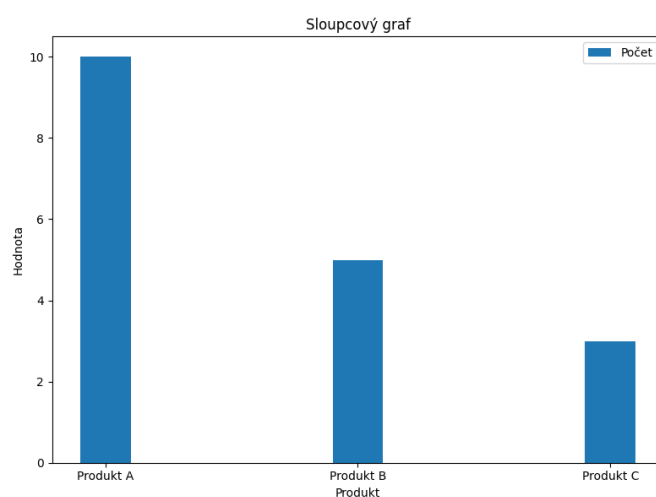
- Analyzovat, které produkty jsou nejprodávanější.
- Určit, které produkty jsou nejvýnosnější.
- Vytvářet grafy pro vizualizaci prodejních trendů.

**Příklad dat:**

	Datum	Produkt	Počet	Cena za kus	Celková cena
1	2024-05-01 00:00:00	Produkt A	10	100	1000
2	2024-05-02 00:00:00	Produkt B	5	200	1000
3	2024-05-03 00:00:00	Produkt C	3	300	900

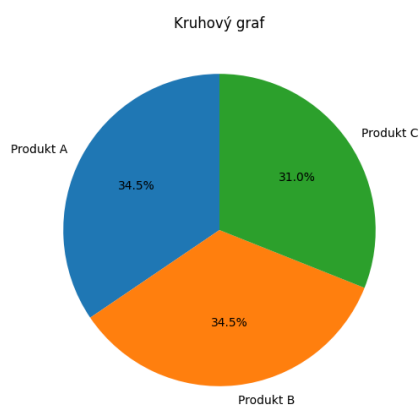
**Obrázek 8.1:** Ukázková data o prodeji produktů

Na obrázku 8.1 jsou znázorněna ukázková data o prodeji produktů. Tabulka obsahuje datum prodeje, název produktu, počet prodaných kusů, cenu za kus a celkovou cenu.



**Obrázek 8.2:** Počet prodaných kusů jednotlivých produktů

Obrázek 8.2 zobrazuje sloupcový graf počtu prodaných kusů jednotlivých produktů. Tento graf umožňuje rychle identifikovat, který produkt byl nejprodávanější.



**Obrázek 8.3:** Podíl prodejů jednotlivých produktů na celkovém prodeji

Obrázek 8.3 zobrazuje kruhový graf podílu prodejů jednotlivých produktů na celkovém prodeji. Tento graf poskytuje vizuální přehled o tom, které produkty přispívají nejvíce k celkovým tržbám.

Název sloupce	Počet	Průměr	Min	25. percentil	Medián	75. percentil	Max
Datum	3	2024-05-02 00:00:00	2024-05-01 00:00:00	2024-05-01 12:00:00	2024-05-02 00:00:00	2024-05-02 12:00:00	2024-05-03 00:00:00
Počet	3.0	6.0	3.0	4.0	5.0	7.5	10.0
Cena za kus	3.0	200.0	100.0	150.0	200.0	250.0	300.0
Celková cena	3.0	966.6666666666666	900.0	950.0	1000.0	1000.0	1000.0

Obrázek 8.4: Základní statistiky prodejů

Obrázek 8.4 zobrazuje základní statistiky prodejů, včetně průměru, mediánu, minimální a maximální hodnoty, a percentilů. Tyto statistiky poskytují hlubší vhled do prodejních trendů a pomáhají při rozhodování o cenové strategii.

Pomocí těchto grafů a výpočtů může malý podnikatel rychle a snadno analyzovat své prodejní data a identifikovat klíčové produkty, které generují nejvíce tržeb. Výpočty průměrné ceny a mediánu prodaných kusů poskytují hlubší vhled do prodejních trendů a pomáhají při rozhodování o cenové strategii. Naše aplikace umožňuje vytvářet tyto grafy a provádět výpočty rychleji a jednodušeji než v Excelu, což z ní činí užitečný nástroj pro podnikatele.

### 8.5.2 Osobní analýza výkonnosti při učení

Studenti mohou aplikaci využít k analýze své výkonnosti při učení na zkoušky. Na základě záznamů o svém studiu mohou studenti:

- Porovnávat, jak různé studijní metody ovlivňují jejich výkon.
- Analyzovat, které faktory (např. čas strávený učením, prostředí) nejvíce přispívají k jejich úspěchu.
- Vytvářet korelační grafy pro identifikaci klíčových faktorů úspěchu.

**Příklad dat:**

	Datum	Čas strávený učením	Prostředí	Výsledek
1	2024-05-01 00:00:00	3	Doma	85
2	2024-05-02 00:00:00	2	Knihovna	78
3	2024-05-03 00:00:00	4	Doma	90
4	2024-05-04 00:00:00	1	Kavárna	70

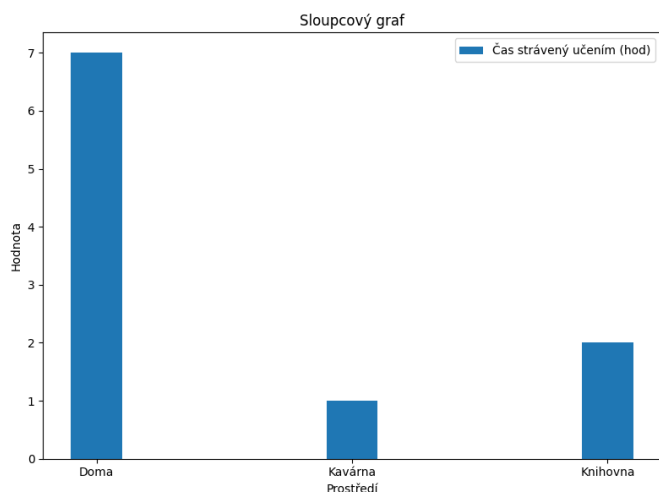
Obrázek 8.5: Ukázková data o čase stráveném učením a výsledcích zkoušek

Obrázek 8.5 ukazuje příklad tabulky s daty o čase stráveném učením, prostředí a výsledcích zkoušek. Studenti mohou analyzovat tato data pro zjištění, jak různé faktory ovlivňují jejich výkon.

Základní statistiky							
Název sloupce	Počet	Průměr	Min	25. percentil	Medián	75. percentil	Max
Datum	4	2024-05-02 12:00:00	2024-05-01 00:00:00	2024-05-01 18:00:00	2024-05-02 12:00:00	2024-05-03 06:00:00	2024-05-04 00:00:00
Čas strávený učením (hod)	4.0	2.5	1.0	1.75	2.5	3.25	4.0
Výsledek zkoušky (%)	4.0	80.75	70.0	76.0	81.5	86.25	90.0

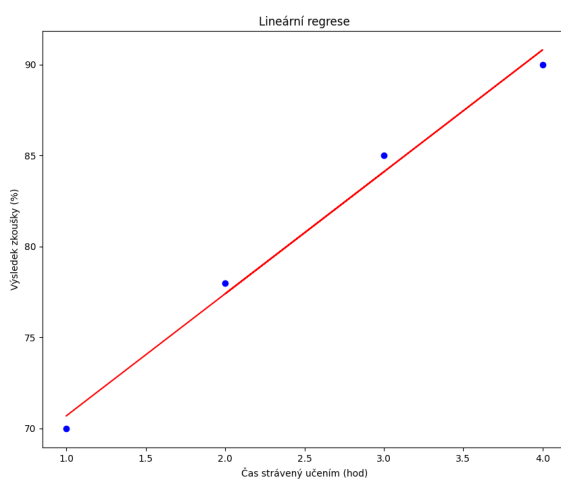
**Obrázek 8.6:** Základní statistiky studijních dat

Obrázek 8.6 zobrazuje základní statistiky zadaných studijních dat, včetně počtu záznamů, průměru, mediánu a percentilů, což poskytuje studentům rychlý přehled o jejich studijních návycích a výsledcích.



**Obrázek 8.7:** Čas strávený učením v různých prostředích

Obrázek 8.7 ukazuje sloupcový graf času stráveného učením v různých prostředích. Tento graf pomáhá studentům vizualizovat, kde tráví nejvíce času učením, a jak to může ovlivnit jejich výsledky.



**Obrázek 8.8:** Lineární regrese mezi časem stráveným učením a výsledky zkoušky

Obrázek 8.8 zobrazuje lineární regresi mezi časem stráveným učením a výsledky zkoušky. Tento graf ukazuje, jaký vliv má délka studia na úspěšnost studentů, a pomáhá identifikovat optimální studijní návyky.

Studenti mohou pomocí těchto grafů a výpočtů analyzovat, jak různé studijní metody a prostředí ovlivňují jejich výsledky. Korelační graf pomáhá identifikovat klíčové faktory úspěchu, zatímco výpočty průměrného času a mediánu výsledků poskytují cenné informace o studijních návycích. Naše aplikace umožňuje rychlé a snadné vytváření těchto grafů a provádění výpočtů, což je mnohem efektivnější než v Excelu. Díky tomu se aplikace stává užitečným nástrojem pro studenty, kteří chtějí optimalizovat své studijní metody a dosáhnout lepších výsledků.

### 8.5.3 Plánování dovolené

Aplikace může pomoci lidem plánovat dovolenou na základě historických dat o počasí a cenách. Pomocí excelovských tabulek obsahujících informace o počasí a cenách mohou uživatelé:

- Analyzovat, kdy je nejlepší doba na dovolenou z hlediska počasí.
- Porovnávat ceny ubytování a dalších nákladů v různých obdobích.
- Vytvářet grafy, které vizualizují nejvhodnější čas pro cestování.

### 8.5.4 Analýza sportovního výkonu

Sportovci mohou aplikaci využít k analýze svého tréninku a výkonu. Pomocí dat z tréninkových deníků mohou:

- Analyzovat, jak různé tréninkové metody ovlivňují jejich výkon.
- Identifikovat, které faktory (např. intenzita tréninku, odpočinek) nejvíce přispívají k jejich zlepšení.
- Vytvářet grafy a korelační analýzy pro optimalizaci tréninkových plánů.

### 8.5.5 Analýza zdravotních údajů

Aplikace může být použita i pro analýzu osobních zdravotních údajů. Jednotlivci mohou:

- Sledovat a analyzovat své zdravotní ukazatele, jako je krevní tlak, hladina cukru v krvi, váha a další.
- Identifikovat trendy a vzory ve svých zdravotních údajích.
- Vytvářet grafy pro lepší vizualizaci zdravotních změn a účinků různých intervencí.

## 8.6 Uzavření kapitoly

Naše aplikace má široké možnosti využití v různých oblastech. Primárně byla navržena pro analýzu dat týkajících se vysokých škol, což zahrnuje analýzu dat o studentech, studijních programech a dalších relevantních informacích. Díky výukovým materiálům může aplikace sloužit jako nástroj pro výuku základů datové analýzy a pomoci uživatelům získat praktické zkušenosti, které jim umožní přejít na pokročilejší nástroje, jako je Power BI nebo Tableau.

Kromě toho má aplikace potenciál být užitečným nástrojem pro různé další oblasti, jako je analýza prodeje pro malé podnikatele, osobní analýza výkonnosti při učení, plánování dovolené, analýza sportovního výkonu a analýza zdravotních údajů. Pomocí aplikace mohou uživatelé snadno vytvářet grafy a provádět výpočty, což díky grafickému klikátku je rychlejší a jednodušší než v Excelu, a tím získat užitečné informace pro své rozhodovací procesy.

Celkově lze říci, že aplikace splnila většinu stanovených cílů a poskytuje uživatelům užitečný nástroj pro základní datovou analýzu. Pro ověření dlouhodobé užitečnosti a efektivity aplikace by bylo nutné provést další testování a získat více zpětné vazby od širšího spektra uživatelů. Teprve poté bude možné definitivně zhodnotit, zda aplikace skutečně naplňuje svůj potenciál jako nástroj pro začínající datové analytiky.

## Kapitola 9

### Závěr

Tato práce se zabývá vývojem aplikace pro základní datovou analýzu, zaměřenou na data z vysokých škol. Hlavním cílem bylo vytvořit nástroj, který by byl uživatelsky přívětivý a snadno použitelný pro začátečníky, kteří nemají předchozí technické znalosti v oblasti datové analýzy.

Během vývoje aplikace jsme narazili na několik výzev, zejména v oblasti pokročilého filtrování dat a optimalizace rychlosti zpracování geografických grafů. Přesto se podařilo implementovat základní funkcionality, které umožňují nahrávání dat, jejich zobrazení v tabulce, třídění sloupců, vytváření základních grafů a provádění složitějších analytických operací, jako jsou lineární regrese a korelace.

Aplikace byla testována s pěti respondenty, jejichž zpětná vazba byla klíčová pro identifikaci a opravu nedostatků. Design aplikace byl hodnocen rozporuplně, což naznačuje potřebu dalšího vývoje a vylepšení.

Navržená vylepšení zahrnují implementaci pokročilejších možností filtrování dat, optimalizaci výkonu aplikace, přehodnocení designu a rozšíření uživatelských testů. S dalšími vylepšeními a optimalizacemi by aplikace mohla být ještě užitečnějším a efektivnějším nástrojem pro široké spektrum uživatelů.

Ambicí aplikace není konkurovat vyspělým nástrojům, ale být vstupenkou do světa datových analýz. Z tohoto pohledu nevadí, že ne všechno se podařilo implementovat. Vzhledem k této skutečnosti nemá smysl aplikaci výrazně rozšiřovat, ale spíše přejít k existujícím pokročilejším nástrojům, pokud uživatelé potřebují komplexnější funkce.

Aplikace byla navržena a vytvořena tak, aby umožnila provádět různé datové analýzy nad běžně dostupnými daty o studentech a studiích na vysokých školách. Tento cíl byl naplněn prostřednictvím následujících kroků:

1. Vyhledání volně dostupných datových sad poskytujících informace o studentech vysokých škol a vysokých školách.
2. Provedení základní analýzy těchto sad a jejich úprava do odpovídající podoby.
3. Návrh ukázkových hypotéz k ověření kvality dat s ohledem na možnosti jejich analýzy.
4. Potvrzení nebo vyvrácení hypotéz pomocí existujících nástrojů.

5. Vytvoření aplikace umožňující uživatelům provádět vlastní analýzy datových sad ve vizuální i nevizuální podobě s možností importu a exportu různých datových sad.
6. Uživatelské testování aplikace pomocí dříve stanovených hypotéz a ukázání dalších možností využití aplikace.

Aplikace má potenciál být užitečná i v jiných oblastech, například:

- **Analýza prodeje pro malé podnikatele:** Malí podnikatelé mohou analyzovat prodejní data, identifikovat nejprodávanější a nejvýnosnější produkty a vizualizovat prodejní trendy.
- **Plánování dovolené:** Uživatelé mohou analyzovat data o počasí a cenách, aby zjistili nejlepší dobu pro dovolenou z hlediska počasí a nákladů.
- **Osobní analýza výkonnosti při učení:** Studenti mohou sledovat svou studijní výkonnost, identifikovat efektivní studijní metody a optimalizovat své studijní návyky.
- **Analýza sportovního výkonu:** Sportovci mohou analyzovat tréninková data, identifikovat klíčové faktory pro zlepšení výkonu a optimalizovat tréninkové plány.
- **Analýza zdravotních údajů:** Jednotlivci mohou sledovat své zdravotní ukazatele, identifikovat trendy a vizualizovat zdravotní změny.

Tato práce mi poskytla cenné zkušenosti a znalosti v oblasti vývoje softwaru a datové analýzy. Doufám, že aplikace, kterou jsem vytvořil, bude užitečná nejen pro mě, ale i pro ostatní, kteří chtějí získat základní dovednosti v datové analýze a postupně přecházet k pokročilejším nástrojům.



## 9.1 Osobní reflexe

Během práce na tomto projektu jsem se setkal s mnoha výzvami, které mi pomohly rozvinout mé technické dovednosti a schopnosti řešení problémů. Spolupráce s **OpenAI chatbotem** a využití **GitHub Copilot** mi umožnily zefektivnit proces psaní a programování. Celkově tato zkušenost mi poskytla hlubší porozumění tomu, jak lze tyto nástroje efektivně využít v reálných projektech.

Práci jsem zpracovával průběžně s různými přestávkami, což mi umožnilo rozumně zvládnout jednotlivé úkoly a části projektu. Díky tomuto přístupu jsem měl dostatek času na řešení problémů a implementaci většiny funkcionalit. Pravidelné schůzky s vedoucím práce byly nesmírně důležité, protože mi poskytovaly cennou zpětnou vazbu a umožnily mi lépe plánovat a organizovat jednotlivé fáze práce. Zároveň mně nutily pracovat opravdu průběžně, což zpětně hodnotím velice kladně.

Celkově jsem s výsledkem práce spokojen a věřím, že aplikace, kterou jsem vytvořil, splňuje z většiny stanovené cíle. Tento projekt mi poskytl cenné zkušenosti a znalosti, které mi budou užitečné v mé budoucí kariéře.



# Příloha A

## Bibliografie

1. *Motivace, priority a kvalita uchazečů o VŠ studium* [online]. [cit. 2024-05-13]. Dostupné z: [https://download.scio.cz/analyzy/souhrn\\_poznatku\\_Vektor\\_uchazeci.pdf](https://download.scio.cz/analyzy/souhrn_poznatku_Vektor_uchazeci.pdf).
2. *Chat GPT-4 – All You Need To Know in 2024* [online]. 2024. [cit. 2024-04-15]. Dostupné z: <https://www.geeksforgeeks.org/chat-gpt-4-all-you-need-to-know/>.
3. *Choosing Data Science Tool in 2024: Striking the Balance* [online]. 2023. [cit. 2024-05-15]. Dostupné z: <https://dataforest.ai/blog/choosing-data-science-tools-striking-the-balance>.
4. *11 nejlepších nástrojů pro analýzu dat* [online]. 2021. [cit. 2024-03-15]. Dostupné z: <https://skillmea.cz/blog/11-najlepsich-nastrojov-na-analyzu-dat>.
5. *Comparison of Data Analysis Tools: Excel, R, Python and BI* [online]. [cit. 2024-05-17]. Dostupné z: <https://www.finereport.com/en/data-analysis/comparison-of-data-analysis-tools-excel-r-python-and-bi.html>.
6. HILLIER, Will. *The 11 Best Data Analytics Tools for Data Analysts in 2024* [online]. 2023. [cit. 2024-05-15]. Dostupné z: <https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/>.
7. *Power BI vs Tableau: Co si vybrat v roce 2023?* [online]. 2023. [cit. 2024-03-09]. Dostupné z: <https://skillmea.cz/blog/power-bi-vs-tableau-co-si-vybrat-v-roce-2023>.
8. SMÍTALOVÁ, Petra. *Víme, co se vyplatí studovat. Přinášíme přehled nástupních platů vysokoškoláků* [online]. 2023. [cit. 2020-03-08]. Dostupné z: [https://www.lidovky.cz/orientace/vzdelavani/work-life-balance-volno-mzda-penize-prace-zamestnani-firma-hodnoty.A230911\\_153213\\_ln-vzdelavani\\_ape](https://www.lidovky.cz/orientace/vzdelavani/work-life-balance-volno-mzda-penize-prace-zamestnani-firma-hodnoty.A230911_153213_ln-vzdelavani_ape).
9. *ANALÝZY - Ministerstvo v rámci své činnosti zpracovává různé analýzy.* [online]. [cit. 2024-05-09]. Dostupné z: <https://www.msmt.cz/vzdelavani/vysoke-skolstvi/analyzy>.

10. MASSACHUSETTS INSTITUTE OF TECHNOLOGY. *Udělené tituly z MIT od července 2021 do června 2022*. 2023. Dostupné z: <https://ir.mit.edu/more-student-data>.
11. HILLIER, WILL. *9 Free Data Analytics Courses for Beginners* [online]. 2024. [cit. 2024-05-01]. Dostupné z: <https://careerfoundry.com/en/blog/data-analytics/free-data-analytics-courses/>.
12. *DATA O STUDENTECH, poprvé zapsaných a absolventech vysokých škol* [online]. 2022. [cit. 2020-03-15]. Dostupné z: <https://www.msmt.cz/vzdelavani/skolstvi-v-cr/statistika-skolstvi/data-o-studentech-poprve-zapsanych-a-absolventech-vysokych>.
13. ČESKO. *Zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách)*. [online]. 1998. [cit. 2020-03-21]. Dostupné z: <https://www.zakonyprolidi.cz/cs/1998-111>.
14. *PRAVIDLA PRO POSKYTOVÁNÍ PŘÍSPĚVKŮ A DOTACÍ VEŘEJNÝM VŠ* [online]. [cit. 2024-05-16]. Dostupné z: <https://www.msmt.cz/vzdelavani/vysoke-skolstvi/pravidla-pro-poskytovani-prispevku-a-dotaci-verejnym-vs>.
15. *Will AI Replace Data Analysts? Be Prepared Now!* [online]. 2024. [cit. 2024-05-03]. Dostupné z: <https://www.ccslearningacademy.com/will-ai-replace-data-analysts/>.
16. TAKYAR, Akash. *AI USE CASES & APPLICATIONS ACROSS MAJOR INDUSTRIES* [online]. 2024. [cit. 2024-04-29]. Dostupné z: <https://www.leewayhertz.com/ai-use-cases-and-applications/>.
17. SELVARAJ, Natassha. *Python Excel Tutorial: The Definitive Guide* [online]. 2023. [cit. 2024-04-07]. Dostupné z: <https://www.datacamp.com/tutorial/python-excel-tutorial>.
18. ŠNÁBL, Ivo. *Chybějící hodnoty* [online]. [cit. 2024-05-01]. Dostupné z: <https://portal.matematickabiologie.cz/index.php?pg=analiza-hodnoceni-biologickych-dat--vicerozmerne-metody-pro-analyzu-dat--uvod-do-vicerozmerne-analyzy-dat--mozne-problemy-vicerozmernych-dat-a-jejich-reseni--chybejici-hodnoty>.
19. <https://www.umimeinformatiku.cz/> [online]. [cit. 2024-05-05]. Dostupné z: <https://www.umimeinformatiku.cz/book/cviceni-vizualizace-dat-typy-grafu>.
20. HAYES, ADAM. *Descriptive Statistics: Definition, Overview, Types, and Example* [online]. 2024. [cit. 2024-05-08]. Dostupné z: [https://www.investopedia.com/terms/d/descriptive\\_statistics.asp](https://www.investopedia.com/terms/d/descriptive_statistics.asp).
21. AVCONTENTTEAM. *What is the Difference Between Covariance and Correlation?* [online]. 2023. [cit. 2024-05-05]. Dostupné z: <https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/>.

22. ŠNÁBL, Ivo. *Výpočet Pearsonova korelačního koeficientu* [online]. [cit. 2024-05-10]. Dostupné z: <https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--biostatistika-pro-matematickou-biologii--zaklady-korelacni-analyzy--pearsonuv-korelacni-koeficient--vypocet-pearsonova-korelacniho-koeficientu>.
23. MONTGOMERY, Douglas C. *Introduction to Linear Regression Analysis* [online]. 1. vyd. John Wiley & Sons, Incorporated, 2012 [cit. 2024-04-28]. ISBN 9781118627365. Dostupné z: <https://ebookcentral.proquest.com/lib/cvut/>.
24. MALI, KAVITA. *Everything you need to Know about Linear Regression!* [online]. 2024. [cit. 2024-05-10]. Dostupné z: <https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>.
25. *DATA O STUDENTECH – Popis metodiky výpočtu jednotlivých tabulek* [online]. 2022. [cit. 2020-03-21]. Dostupné z: <https://www.msmt.cz/file/21927/download/>.
26. STRAŠILOVÁ, Gabriela. *„Husákovy“ versus „Havlovy děti“* [online]. [cit. 2024-04-20]. Dostupné z: <https://www.czso.cz/csu/czso/52002e2055>.
27. *Dopady světové finanční a hospodářské krize na ekonomiku ČR* [online]. 2011. [cit. 2024-05-14]. Dostupné z: <https://www.czso.cz/documents/10180/20534938/115610j.pdf/b487dd3c-0ad7-4ccd-b62d-8fc9bf917b95?version=1.0>.
28. NĚMEC, Jan. *Školné na VŠ bude, shodly se ODS, TOP 09 a VV* [online]. 2010. [cit. 2024-05-13]. Dostupné z: <https://zpravy.aktualne.cz/domaci/skolne-na-vs-bude-shodly-se-ods-top-09-a-vv/r~i:article:670295/>.
29. *KONCEPCE REFORMY VYSOKÉHO ŠKOLSTVÍ V ČR* [online]. 2004. [cit. 2024-05-15]. Dostupné z: [https://www.msmt.cz/file/102\\_1\\_1/](https://www.msmt.cz/file/102_1_1/).
30. *STUDIUM A ŽIVOT V ČESKU POHLEDEM ZAHRANIČNÍCH STUDENTŮ II* [online]. 2023. [cit. 2024-05-15]. Dostupné z: [https://www.dzs.cz/sites/default/files/2023-12/DZS\\_zprava\\_o\\_zahranicnich\\_studentech\\_2023\\_WEB.pdf](https://www.dzs.cz/sites/default/files/2023-12/DZS_zprava_o_zahranicnich_studentech_2023_WEB.pdf).
31. DAŇKOVÁ, Julie. *Praha, Brno, Olomouc: jaké je město, kde budete studovat?* [online]. 2016. [cit. 2024-05-15]. Dostupné z: <https://www.jobs.cz/poradna/jake-je-univerzitni-mesto-kde-budete-studovat/>.
32. *Žijete ve městě, které patří studentům* [online]. [cit. 2024-05-13]. Dostupné z: <https://www.muni.cz/uchazeci/navazujici-magisterske-studium/10-duvodu/brno-je-studentske-mesto>.

33. HUSOVSKÁ, Christiana. *Vzdělání nebo trh práce? Motivace ke studiu u studentů soukromých a veřejných vysokých škol* [online]. Praha, 2022 [cit. 2024-04-28]. Dostupné z: <https://dspace.cuni.cz/handle/20.500.11956/173427>. Bakalářská práce. Univerzita Karlova, Fakulta sociálních věd, Katedra sociologie.
34. *Dropout.pef.czu.cz* [online]. [cit. 2024-05-16]. Dostupné z: <https://dropout.pef.czu.cz/>.
35. *Šance na dostudování českých VŠ (počet zapsaných/počet absolventů)* [online]. [cit. 2024-05-06]. Dostupné z: <https://www.ceskovdatech.cz/graphs/vs2.php>.
36. KOMÁREK, Martin. *Přednáška - ANALÝZA A DOKUMENTACE POŽADAVKŮ* [online]. ČVUT, 2022 [cit. 2024-05-05]. Dostupné z: <https://moodle.fel.cvut.cz/mod/resource/view.php?id=227415>.
37. KOMÁREK, Martin. *Přednáška - UML diagramy případů užití* [online]. 2022. [cit. 2024-05-11]. Dostupné z: <https://moodle.fel.cvut.cz/mod/resource/view.php?id=227413>.
38. KOMÁREK, Martin. *Přednáška - Pokročilé UML digramy tříd a aktivit* [online]. 2020. [cit. 2024-05-09]. Dostupné z: <https://moodle.fel.cvut.cz/mod/resource/view.php?id=227419>.
39. *What is Python? Executive Summary* [online]. [cit. 2024-04-28]. Dostupné z: <https://www.python.org/doc/essays/blurb/>.
40. LUNA, Javier Canales. *Top programming languages for data scientists in 2023* [online]. 2022. [cit. 2024-05-12]. Dostupné z: <https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022>.
41. *Nine top programming languages for data science* [online]. 2024. [cit. 2024-05-07]. Dostupné z: <https://www.edx.org/resources/9-top-programming-languages-for-data-science>.
42. KUMMARIKUNTLA, Teja. *Choosing the Right Python GUI Framework: A Complete Guide* [online]. 2023. [cit. 2024-03-12]. Dostupné z: <https://blog.tooljet.com/python-gui-framework/>.
43. *Top 26 Python Libraries for Data Science in 2024* [online]. 2024. [cit. 2024-03-16]. Dostupné z: <https://www.datacamp.com/blog/top-python-libraries-for-data-science>.
44. FAIGL, Jan; VOKŘÍNEK, Jiří. *GUI v Javě a událostmi řízené programování* [online]. 2017. [cit. 2024-05-12]. Dostupné z: [https://cw.fel.cvut.cz/b212/\\_media/courses/b0b36pjb/prednasky/lecture06-slides.pdf](https://cw.fel.cvut.cz/b212/_media/courses/b0b36pjb/prednasky/lecture06-slides.pdf).
45. FRAJTÁK, Karel. *AUTOMATED END-TO-END TESTING* [online]. [cit. 2024-05-01]. Dostupné z: <https://moodle.fel.cvut.cz/mod/resource/view.php?id=233313>.

## Příloha B

### Aplikace a repozitáře

#### B.1 Odkazy na repozitáře

Na stránce GitLab FEL<sup>1</sup> nebo v případě chybějícího přístupu do FEL GitLabu také na GitHub<sup>2</sup> najdete všechny soubory spojené s aplikací a bakalářskou prací.

#### B.2 Excel tabulky pro analýzu

Excel tabulky pro analýzu jsou zpracované tabulky, které lze využívat pro datovou analýzu v aplikaci. Tyto tabulky doporučuji stáhnout a používat v aplikaci.

#### B.3 Spustitelný soubor

V hlavní složce aplikace najdete soubor `DataStart.exe`. Jedná se o spustitelný soubor, ke kterému byste neměli potřebovat další závislosti. Pokud aplikace nefunguje, může být problém s chybějícími Visual C++ Redistributables nebo jinými podobnými komponentami.

Aplikace by měla být spustitelná pouhým stažením a dvojklikem na ni. V aplikaci je na úvodní stránce krátký popis a případné vysvětlení, co a jak dělat.

---

<sup>1</sup><https://gitlab.fel.cvut.cz/horejvo1/bakalarska-prace-datova-analyza>

<sup>2</sup>[https://github.com/vojtee/Bakalarska\\_prace\\_datova\\_analyza](https://github.com/vojtee/Bakalarska_prace_datova_analyza)