



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

Katedra biomedicínské techniky

Sběr a zpracování dat o nákupech zdravotnické techniky

**Collection and processing of data on purchases of medical
equipment**

Diplomová práce

Studijní program: Systémová integrace procesů ve zdravotnictví

Vedoucí práce: Ing. Gleb Donin, Ph.D.

Bc. Marek Tošovský

Kladno 2022

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Tošovský** Jméno: **Marek** Osobní číslo: **434147**
Fakulta: **Fakulta biomedicínského inženýrství**
Garantující katedra: **Katedra biomedicínské techniky**
Studijní program: **Systémová integrace procesů ve zdravotnictví**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Sběr a zpracování dat o nákupech zdravotnické techniky

Název diplomové práce anglicky:

Collection and processing of data on purchases of medical equipment

Pokyny pro vypracování:

Cílem diplomové práce je dopracování automatizace sběrů a prvotního zpracování dat o veřejných zakázkách a vypracování deskriptivní analýzy dodavatelů na trhu zdravotnické techniky za pomoci programovacího prostředí pro statistickou analýzu dat (R). Zpracujte data o veřejných zakázkách z otevřených datových sad z portálu „Informační systém o veřejných zakázkách“. Rozšiřte tato data, primárně pomocí formulářů z portálu „Věstník veřejných zakázek“, použitím metody „web scraping“. Vypracujte návod pro práci s vytvořeným datovým setem. Vytvořte a implementujte algoritmus pro filtraci jednotlivých nákupů zdravotnické techniky z veřejných zakázek pomocí klasifikace dle základních skupin zdravotnických přístrojů. Proveďte základní deskriptivní statistickou analýzu z vytvořeného datového setu s důrazem na rozložení dodavatelů zdravotnické techniky.

Seznam doporučené literatury:

- [1] MACEK, Ivo, Zákon o zadávání veřejných zakázek: praktický komentář s judikaturou, ed. 1, Komentátor, 2017, ISBN 978-80-7502-171-7
- [2] BORGNE A.L, HOMOLOVA A. a G. BONTEMPI, OpenTED Browser: Insights into European Public Spendings, 16.09.2016, <https://arxiv.org/abs/1611.02118>
- [3] BALOUNOVÁ Z., Analýza faktorů ovlivňujících postup zadávání veřejných zakázek na zdravotnickou techniku, 2019, Diplomová práce, FBMI ČVUT

Jméno a příjmení vedoucí(ho) diplomové práce:

Ing. Gleb Donin, Ph.D.

Jméno a příjmení konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **15.02.2021**

Platnost zadání diplomové práce: **18.09.2022**

doc. Ing. Martin Rožánek, Ph.D.
vedoucí katedry

prof. MUDr. Jozef Rosina, Ph.D., MBA
děkan

PROHLÁŠENÍ

Prohlašuji, že jsem diplomovou práci s názvem „Sběr a zpracování dat o nákupech zdravotnické techniky“ vypracoval samostatně a použil k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k diplomové práci.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů.

V Kladně dne 12. 5. 2022

.....

Bc. Marek Tošovský

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu této práce, panu doktoru Doninovi za poskytnutí základů, na kterých tato práce stojí, za cenné rady, za pomoc, za trpělivost, kterou se mnou měl, když jsem zpracovával zadané téma. Velice si vážím férového a pozitivního přístupu, který pan doktor při vedení práce nastavil a kterým dovedl tuto práci do cíle.

ABSTRAKT

Sběr a zpracování dat o nákupech zdravotnické techniky:

Cílem práce bylo dopracovat algoritmus pro sběr a prvotní zpracování dat o nákupech zdravotnické techniky z informačního systému o veřejných zakázkách, tyto data rozšířit pomocí veřejně dostupných databází a pomocí metody zpracování přirozeného jazyka, a vytvořit tak datový set vhodný k analýzám. Cílů bylo dosaženo pomocí skriptů psaných v programovacím jazyce R, metody webscrapping a metody fuzzy-matching. Výsledný datový set obsahuje 2650 záznamů. Metoda zpracování přirozeného jazyka byla otestována na 43 vybraných zakázkách. Nejvyšší úspěšnost algoritmu, určit správnou kategorii zdravotnické techniky, které se zakázka týká, je 79 %. Byla provedena deskriptivní a regresní analýza. Se vzrůstajícím počtem přihlášených dodavatelů a se vzrůstajícím počtem nastavených kritérií hodnocení veřejné zakázky stoupá míra úspory.

Klíčová slova

Veřejné zakázky, Zdravotnická technika, Open Data, Zpracování dat

ABSTRACT

Collection and processing of data on purchases of medical equipment:

The aim of this Master's Thesis was to complete the algorithm for the collection and initial processing of data on medical equipment purchases from the public procurement information system, to enrich this data using publicly available databases and natural language processing methods to create a data set suitable for analysis. The goals were achieved using scripts written in the R programming language, webscrapping method and fuzzy-matching method. The resulting data set contains 2650 records. The natural language processing method was tested on 43 selected contracts. The highest success rate of the algorithm, to determine the correct category of medical technology to which the contract relates, is 79%. Descriptive and regression analysis was performed. With the growing number of registered suppliers and with the growing number of set public procurement evaluation criteria, the savings rate increases.

Keywords

Public Procurement, Medical Devices, Open Data, Data Processing

Obsah

Seznam zkratk	9
1 Úvod	10
2 Přehled současného stavu	11
2.1 Zákon o zadávání veřejných zakázek.....	11
2.1.1 Informační systém o veřejných zakázkách.....	11
2.1.2 Uveřejňování formulářů	12
2.1.3 Souhrnné výstupy z uveřejněných formulářů.....	13
2.2 Veřejné zakázky a transparentnost.....	17
2.2.1 Otevřená data.....	17
2.2.2 Analýza dat z veřejných zakázek	18
2.2.3 Analýza otevřených dat v ČR.....	21
2.3 Rozšíření otevřených dat pomocí nature language processing	22
2.3.1 Named entity recognition	22
2.3.2 Kategorizace zdravotnické techniky.....	23
3 Cíle práce	24
3.1 Automatizace sběru a prvotního zpracování dat	24
3.2 Rozšíření datové sady.....	24
3.3 Deskriptivní a regresní analýza	24
3.4 Rozšíření datového setu o kategorii zdravotnické techniky.....	24
4 Metody	25
4.1 Automatizace sběru a prvotního zpracování dat o veřejných zakázkách.....	25
4.1.1 Automatizace stažení a prvotního zpracování open dat veřejných zakázek a jejich import do R.....	25
4.1.2 Filtrace a zpracování se zaměřením na zdravotnickou techniku	27
4.2 Rozšíření datové sady pomocí metody webscrapping	29
4.2.1 Rozšíření z databáze RES.....	29
4.2.2 Rozšíření z databáze formulářů VVZ.....	29
4.3 Metody deskriptivní a regresní analýzy	31
4.3.1 Vytvoření data setu pro účel analýz	31
4.3.2 Lineární regresní analýza	32

4.3.3	Vícenásobný lineární regresní model v R	33
4.4	Algoritmus pro kategorizaci zdravotnické techniky	35
5	Výsledky.....	37
5.1	Datové sady	37
5.1.1	Rozšiřující datová sady REStable	38
5.1.2	Rozšiřující datová sada FORMtable.....	38
5.1.3	Datová sada FNtb pro účel analýz.....	38
5.2	Deskriptivní analýza.....	39
5.2.1	Deskriptivní analýza z databáze RES.....	44
5.3	Regresní analýza	45
5.3.1	Vícenásobný lineární regresní model	46
5.3.2	Ověření modelu	51
5.4	Výstupy algoritmu pro kategorizaci zdravotnické techniky	52
6	Diskuse.....	57
7	Závěr	64
	Seznam použité literatury	65
	Příloha A: Základní typy formulářů.....	70
	Příloha B: Ponechané a vyřazené položky z otevřeného datového setu z portálu ISVZ	74
	Příloha C: Vývojové diagramy zpracování dat.....	80
	Příloha D: Codebook k datovému setu „FNtb“	91

Seznam zkratek

Seznam zkratek

Zkratka	Význam
ZZVZ	Zákon č. 134/2016 Sb., o zadávání veřejných zakázek
MMR	Ministerstvo pro místní rozvoj České republiky
ISVZ	Informační systém o veřejných zakázkách
VVZ	Věstník veřejných zakázek
TED	Tenders Electronic Daily
CPV	Společný slovník pro veřejné zakázky (The Common Procurement Vocabulary)
RDF	Resource description framework
NLP	Natural Language Processing
NER	Named Entity Recognition
UNGM	United Nations Global Marketplace
ÚZIS	Ústav zdravotnických informací a statistiky ČR
NRPZS	Národní registr poskytovatelů zdravotních služeb
VZP	Všeobecná zdravotní pojišťovna
ČSU	Český statistický úřad
GPA	Agreement on Government Procurement
RES	Rejstřík ekonomických subjektů
PZS	Seznam smluvních poskytovatelů zdravotních služeb
RS	Rámcová smlouva
DNS	Dynamický nákupní systém
FN	Fakultní nemocnice

1 Úvod

Zákon č. 134/2016 Sb., o zadávání veřejných zakázek představuje souhrnnou úpravu zadávání veřejných zakázek v ČR [1]. Celkové náklady zemí EU na práci, zboží a služby byli v roce 2017 odhadovány na 2049,8 bilionů euro [2]. V rámci České republiky jsou výdaje za rok 2020 odhadovány na 12,65 % HDP [3]. Veřejné zakázky se tedy podílejí významným dílem na veřejných výdajích.

Proces zadávání veřejných zakázek zažívá sílící tlak ze strany veřejnosti k co největší transparentnosti [4]. Ministerstvo pro místní rozvoj tedy na svém portálu informační systém o veřejných zakázkách zveřejňuje agregovaná data z věstníku veřejných zakázek a to ve formátu otevřených dat [5]. Tyto otevřená data jsou volně dostupná a umožňují hlubší analýzu trhu veřejných zakázek [6].

Ačkoliv existují publikace zabývající se analýzou těchto dat [7–16], pouze jediná se zaměřuje na zdravotnické přístroje v ČR [16].

Cílem této diplomové práce je dopracovat algoritmus pro sběr a prvotní zpracování dat o nákupech zdravotnické techniky z veřejných zakázek. Tuto datovou sadu obohatit o informace z dalších zdrojů a připravit data pro deskriptivní a regresní analýzu. K těmto datům bude vypracován codebook. Dále je cílem provést deskriptivní a regresní analýzu a otestovat algoritmus pro rozšíření dat pomocí techniky zpracování přirozeného jazyka o kategorie zdravotnické techniky.

2 Přehled současného stavu

2.1 Zákon o zadávání veřejných zakázek

Zákon č. 134/2016 Sb., o zadávání veřejných zakázek (ZZVZ) nabyl účinnosti dnem 1. 10. 2016 [1]. Jedná se o čtvrtou generaci právní úpravy v oblasti veřejných zakázek a koncesí. Po deseti letech nahradil původní zákon č. 137/2006 Sb., o veřejných zakázkách a zákon č. 139/2006 Sb., o koncesních smlouvách a koncesním řízení [17]. ZZVZ představuje souhrnnou úpravu zadávání veřejných zakázek v ČR a implementuje tři nové evropské směrnice, v oblasti zadávání veřejných zakázek, které předchází generace právní úpravy nezahrnovala. Jedná se o zadávací směrnici (směrnice Evropského parlamentu a Rady) 2014/24/EU o zadávání veřejných zakázek, 2014/25/EU o zadávání zakázek subjekty působícími v odvětví vodního hospodářství, energetiky, dopravy a poštovních služeb a 2014/23/EU o udělování koncesí. Zákon dále vychází ze čtyř již dříve implementovaných zadávacích směrnic a zadávací směrnice 2014/55/EU o elektronické fakturaci při zadávání veřejných zakázek [1].

V rámci ZZVZ jsou upraveny povinnosti a práva vyplývající ze zadávání všech typů veřejných zakázek, včetně koncesí a sektorových zakázek. Upravuje celkem devět možných druhů zadávacího řízení, a to otevřené řízení, užší řízení, zjednodušené podlimitní řízení, jednací řízení s uveřejněním, jednací řízení bez uveřejnění, řízení se soutěžním dialogem, řízení o inovačním partnerství, řízení pro zadávání veřejné zakázky ve zjednodušeném režimu a koncesní řízení. Zákon se snaží zjednodušit administrativní zátěž pro zadavatele i dodavatele přehlednou systematikou. Uživatelé poté nemusí znát celý zákon, ale jen pro ně příslušné části [1]. ZZVZ se zároveň snaží držet hlavních zásad, které ovládají zadávání veřejných zakázek – transparentnost, přiměřenost, rovné zacházení a zákaz diskriminace [17].

2.1.1 Informační systém o veřejných zakázkách

Podstatnou součástí ZZVZ je i důraz na elektronizaci procesu zadávání veřejných zakázek za účelem zvýšení transparentnosti. Tento požadavek má základ v zadávacích směrnicích 2014/24/EU a 2014/25/EU, které nařizují členským státům zajistit, aby veškerá komunikace, výběr informací a podávání nabídek probíhalo pomocí elektronických komunikačních prostředků [18]. Směrnice umožňuje jiné formy komunikace než elektronické, ale jen u taxativně uvedených výjimek [17]. Povinnost elektronické komunikace pro všechny subjekty, zakotvená v ZZVZ, nabyla účinnosti 18. 10. 2018 [18].

Pro zajištění elektronizace zadávání veřejných zakázek, jejich transparentnosti a dostupnosti informací při realizaci zadávacích řízení byl Ministerstvem pro místní rozvoj ČR (MMR) zřízen informační systém o veřejných zakázkách (ISVZ). ISVZ je dle zákona č. 365/2000 Sb., o informačních systémech veřejné správy stanoven jako informační

system veřejné správy. Dle tohoto zákona může zadávající subjekt pověřit k provádění některých činností spojených s ISVZ externího dodavatele. V případě ISVZ tak MMR není přímo provozovatelem některých jeho částí, ale na konečnou podobu má dopad nepřímo, například schvalováním provozního řádu externího dodavatele. Součástí ISVZ je Věstník veřejných zakázek, seznam kvalifikovaných dodavatelů a další informace potřebné k elektronické komunikaci při zadávání veřejných zakázek [17].

Věstník veřejných zakázek (VVZ) slouží v rámci ISVZ jako modul pro uveřejňování informací o zadávacích řízeních, a to na celostátní úrovni. V rámci snahy o dodržení transparentnosti jsou základní údaje o veřejné zakázce veřejně dohledatelné, stejně tak i profily zadavatelů. Proces uveřejňování ve VVZ je prováděn stanovenými formuláři. Zadavatel je povinen uveřejnit formulář o zahájení zadávacího řízení a formulář oznámení o výsledku zadávacího řízení. Ve specifických případech může použít formulář o předběžném oznámení, čímž dává najevo budoucí záměr o zadání řízení. Dále existují formuláře opravné. Ke konkrétnímu kroku při procesu zadávacího řízení je přiřazen konkrétní formulář [17]. Uveřejňování formulářů v zadávacím řízení veřejných zakázek je striktně povinné [18]. Modul VVZ dále slouží ke komunikaci s Ústředním věstníkem Evropské unie [17] a jeho modulem Tenders Electronic Daily (TED). Tento modul slouží k uveřejňování informací o zadávacích řízeních na úrovni Evropské unie a Evropského hospodářského prostoru a sdružuje tak trh veřejných zakázek napříč dotčenými státy [19].

2.1.2 Uveřejňování formulářů

Uveřejnění je v rámci ZZVZ chápáno jako proces, navázaný na různé druhy zadávacích řízení. Například dle § 56 odst. 1: „Zadavatel zahajuje otevřené řízení odesláním oznámení o zahájení zadávacího řízení k uveřejnění způsobem podle § 212, kterým vyzývá neomezený počet dodavatelů k podání nabídky.“ K uveřejnění je dále povinen dle § 212 odst. 1 použít přímo stanovených formulářů. Tyto formuláře využívá nejen k zahájení, ale i k oznámení o zadání zakázky a v dalších procesech zadávacího řízení [17].

ZZVZ tedy nařizuje zadavateli odeslat uveřejnění dle prováděcího nařízení komise EU č. 2015/1986 kterým se stanoví standardní formuláře pro uveřejňování oznámení v oblasti zadávání veřejných zakázek, pokud se jedná o nadlimitní veřejnou zakázku, do modulu VVZ a zároveň do Ústředního věstníku Evropské unie [17]. Pokud se jedná o podlimitní veřejnou zakázku odesílá zadavatel uveřejnění pomocí formulářů, které stanovuje vyhláška č. 168/2016 Sb. o uveřejňování formulářů pro účely zákona o zadávání veřejných zakázek a náležitostech profilu zadavatele [20]. Tyto formuláře odesílá pouze do VVZ. Zadavatel má možnost podat podlimitní veřejnou zakázku v nadlimitním režimu. Pak ale musí dodržet legislativní předpisy spojené s nadlimitním režimem zadání [17]. Pro přehlednější jsou základní typy formulářů a jejich napojení na jednotlivé režimy zadávání uvedeny v příloze A.

ZZVZ a směrnice EU o zadávání veřejných zakázek tedy požadují odesílání jednotných evropských elektronických formulářů do TED a VVZ pouze u veřejných zakázek v nadlimitním režimu. V případě zakázek podlimitních vychází povinnost uveřejnění jen ze ZZVZ a zadavatel musí uveřejnit formuláře pouze ve VVZ pomocí národních formulářů [17].

2.1.3 Souhrnné výstupy z uveřejněných formulářů

Zadavateli tedy ze ZZVZ vyplývá povinnost použít elektronický nástroj VVZ k celému procesu zadávacího řízení, od vyhlášení zadávacího řízení, přes opravy v zadání až po vyhlášení výsledku, ať už se jedná o zakázku v nadlimitním, nebo podlimitním režimu. Celý proces je zaznamenáván v databázích VVZ právě prostřednictvím jednotných evropských a národních uveřejňovacích formulářů. Tyto formuláře musí být veřejně přístupné nejméně 5 let od jejich uveřejnění [20].

MMR poskytuje otevřená data o veřejných zakázkách ve třech podobách. První variantou jsou otevřené datové sady na portálu ROZZA [21], druhou nové datové sady na portálu ISVZ [22] a třetí variantou jsou datové sady na portálu ISVZ, které jsou strukturované dle formátu starého ISVZ [5].

Portál ROZZA je součástí optimalizace fungování Národní infrastruktury pro elektronické zadávání veřejných zakázek (NIPEZ) dle akčního plánu Strategie elektronizace veřejných zakázek pro období 2016–2020. Vznikl jako snaha sjednotit elektronické nástroje pro zadávání veřejných zakázek, například Tender arena, E-ZAK, TENDERMARKET a další. To poté umožňuje dodavatelům přistupovat k nabídkám z jednoho centrálního místa. Platforma si poté klade za cíl udržovat konkurenční prostředí na trhu a zjednodušit dodavatelům přístup k jednotlivým zakázkám [23]. Data prezentovaná na portálu jsou také dostupná v strukturovaném a úplném formátu, jako otevřená data, pro další použití. Jelikož ale Portál ROZZA agreguje informace z různých portálů, tak i výstupní otevřená data nejsou unifikovaně strukturována. Obecně má záznam hlavičku o veřejné zakázce, která obsahuje stručné informace o zakázce, například druh zakázky, typ zadávacího řízení, název zakázky, popis zakázky a další. Například ale CPV kód zakázky je k dispozici pouze u některých záznamů, v závislosti na tom, ze kterého elektronického nástroje byla informace převzata. Dále záznam obsahuje informaci o zadavateli a poté informace o navázané dokumentaci. Z důvodu nekonzistence uveřejňovaných dat by bylo strojové zpracování této otevřené datové sady přinejmenším nesnadné. U některých záznamů navíc chybí zásadní informace, jako je CPV kód, konečná cena zakázky a další [21]. Zásadním nedostatkem je také neexistující metodika zpracování a uveřejnění těchto otevřených datových sad.

Nové otevřené datové sady na ISVZ jsou vyextrahovaná data z VVZ. Dle metodiky k tomuto setu je cílem: „*poskytnout modulární platformu pro tvorbu statistik, podporující různé dílčí způsoby výpočtu dalších ukazatelů. Ukazatelem je měřitelná veličina, která se*

ve statistikách používá k reprezentaci agregovaných informací.... Vedle ukazatelů jsou pro statistiky dostupné dimenze, to jsou atributy, které slouží k filtrování a rozpadu ukazatelů do různých pohledů a úrovní detailu.[24]. “ V praxi jsou poté vytvořeny datové sady, které poskytují, dle vydavatele, zajímavé informace pro tvorbu statistik. Položky, které obsahují jednotlivé záznamy v jednotlivých otevřených sadách jsou shrnuty v tabulce 2.1.

Tabulka 2.1: Souhrn položek, které obsahují nové otevřené datové sady z portálu ISVZ.

Datová sada				
Zahájené VZ:	podle Režimu	podle Typu	podle Druhu Řízení	podle Druhu
	Režim zakázky	Typ zakázky	Druh řízení	Druh zakázky
		Zadavatel		
		Celkový počet zahájených zakázek		
		Celková korigovaná předpokládaná hodnota v CZK		
		Rok zahájení řízení		
Zadané VZ:	podle Režimu	podle Typu	podle Druhu Řízení	podle Druhu
	Režim zakázky	Typ zakázky	Druh řízení	Druh zakázky
		Zadavatel		
		Celkový počet zadaných zakázek – pohled po letech		
		Celková korigovaná konečná hodnota zakázek v CZK		
		Rok zadání zakázky		

Otevřené datové sady se dají rozdělit na dvě skupiny, „Zahájené VZ“ a „Zadané VZ“. Tyto skupiny mají vždy stejnou vnitřní strukturu, až na jednu proměnnou položku.

Informace v těchto datových sadách jsou velmi omezené a rozhodně nesplňují požadavky stanovené v „Standardech publikace a katalogizace otevřených dat VS ČR“ publikované Ministerstvem Vnitra [6]. Jsou neúplné a velmi zjednodušené. Plusem je, že k datovým sadám existuje metodika [24].

MMR prostřednictvím ISVZ vydává souhrnné otevřené datové sady ve formátu zveřejňovaném na přechodím ISVZ, které obsahují cenné agregované informace o veřejných zakázkách automaticky vyextrahované z dat z přijatých formulářů. Z těchto otevřených sad je možné zjistit zadavatele veřejné zakázky, CPV, nabídky, konečnou cenu zakázky a mnoho dalšího [25]. Datové sady se vždy aktualizují k 1. a 16. dni v měsíci za předcházejících měsíc a jsou dostupné ve formátech pro strojové zpracování (xml, xlsx, csv) [25]. Poskytují možnost pro hlubší analýzu trhu veřejných zakázek, kontrolu a přispívají k transparentnosti a efektivnosti při dispozici s veřejnými financemi [4]. Tyto datové sady byly mezi roky 2016–2020 spravovány firmou ASD Software, s. r.o. Od roku 2021 jsou data poskytována v novém systému RAV a struktura formátu může být mírně rozdílná. Například v názvech položek byla provedena změna z „IDZakazky“ ve starém formátu na „ID_Zakazky“ v novém formátu RAV. Takovýchto drobných změn

je v novém systému více. Poskytovatel si dal závazek tyto nesrovnalosti dále opravovat a data zpětně přegenerovat [25].

Společný slovník pro veřejné zakázky

Společný slovník pro veřejné zakázky (CPV) je klasifikační systém, který si klade za cíl standardizovat popis předmětu zakázky při jejím zadávání. Používání CPV snižuje riziko chybných překladů v databázi TED, protože existují oficiální překlady CPV do všech jazyků Evropské unie [26]. Dále prohlubuje transparentnost při zadávání veřejných zakázek, usnadňuje nalézání příležitostí pro podnikatele zveřejněných v TED a VVZ, zjednodušuje proces zadávání a popisu zakázky pro zadavatele a dále zjednodušuje statistické zpracování dat o veřejných zakázkách, jelikož CPV je mezinárodně uznávaný systém, zejména v rámci OSN [27].

Současná podoba CPV slovníku vychází z nařízení komise (ES) č. 213/2008 o společném slovníku pro veřejné zakázky (CPV) [28]. Použití toto nařízení udává ZZVZ [1] a způsob použití při zadávání veřejné zakázky upravuje vyhláška č. 168/2016 Sb., která definuje podobu a obsah formulářů pro účely zákona o zadávání veřejných zakázek. Zadavatel je tedy povinen při popisu předmětu zakázky vždy použít slovníku CPV a to i v případě použití národních formulářů, jelikož i ty jsou definované s částí pro povinné uvedení kódu CPV [20].

CPV je rozdělen na hlavní slovník, který je používán k specifikaci předmětu veřejné zakázky a doplňkového slovníku, který je možné použít pro upřesnění jiných kvalitativních informací. Hlavní slovník je stromové struktury a jednotlivé kódy jsou sestaveny z devíti číslic a jejich struktura je následující [28, 29]:

- oddíl, určen prvními dvěma číslicemi (33000000-0: Medical equipments, pharmaceuticals and personal care products),
- skupina, určena prvními třemi číslicemi (33100000-1: Medical equipments),
- třída, určena prvními čtyřmi číslicemi (33110000-4: Imaging equipment for medical, dental and veterinary use),
- kategorie, určena prvními pěti číslicemi (33111000-1: X-ray devices).

Každá další číslice slouží k vyššímu členění dané kategorie [28, 29]:

- 33111600-7: Radiography devices,
- 33111610-0: Magnetic resonance unit.

Devátá číslice za pomlčkou slouží ke kontrole a ověřování předešlých číslic v kódu [28]. Doplňkový slovník používá alfanumerický kód, kterému odpovídá slovní popis, jež dále upřesňuje charakteristiku nebo účel zboží, které je předmětem nákupu. Sekce L stanovuje Zbývající lékařské a laboratorní atributy. Například kód LA33-9 označuje použití pro patologii a LA36-8 pro radiodiagnostiku [30].

System CPV kódů má své nedostatky. Evropská komise uvádí v dokumentu „*Revision of CPV* [26]“ z roku 2017 čtyři hlavní problémy spojené s tímto systémem. V rámci problematiky elektronizace, automatického zpracování dat a následné analýzy, je zásadní první problém, a to potíže uživatelů při hledání vhodného kódu pro konkrétní službu, nebo produkt, což vede k nesprávné užívání CPV kódů [26].

CPV kódy zdravotnických přístrojů

Při vypsání veřejné zakázky na nákup zdravotnických přístrojů je zadavatel povinen použít skupinu 3310000-1, zdravotnické přístroje a dále přístroj zařadit do vhodné třídy, kategorie a podkategorií. V rámci tohoto stromu však existují případy, kdy je klasifikační systém CPV buď příliš podrobný, nebo naopak příliš zobecňuje a do jedné podkategorie poté spadají přístroje, které jsou sice technicky podobné, účelem použití se mohou ale zásadně lišit [29]. Toto bylo demonstrováno vyhledáním oznámení o výsledcích zadávacích řízení (formuláře F03 a CZ03, dle vyhlášky 168/2016 Sb.) na VVZ (<https://www.vestnikverejnychzakazek.cz/SearchForm>) pro kód CPV 33151000, přístroje a doplňky pro radioterapii. Bylo nalezeno celkem 8 výsledků. Pokud bylo více výsledků navázáno na jedno zadání, ale z důvodu nevysoutěžení části zakázky, byl proces zadání částečně zopakován, byly tyto výsledky vyfiltrovány. V tabulce 2.2, kde jsou výsledky shrnuty, můžeme nalézt druhově velmi rozdílné přístroje. Pod jednu kategorii spadají lineární urychlovače pro radioterapii, gama nůž, ale i systémy radiační ochrany. Je zřejmé, že toto členění by bylo možné ještě zpřesnit. Určitou možností je použití doplňkových kódů. Ani v jednom případě ale tyto doplňkové kódy nebyly při zadávání použity [31].

V tomto stavu tak nelze pomocí CPV kódů přesně určit, jaký typ zdravotnického přístroje je předmětem zakázky, jelikož jak bylo demonstrováno, některé kategorie pokrývají velmi široký záběr přístrojů a přístrojového vybavení.

Tabulka 2.2: Souhrn výsledků vyhledávání dotazu o výsledcích zadávacích řízení ve VZZ pro kód CPV 33151000 (Přístroje a doplňky pro radioterapii).

CPV	33151000	
Zadavatel	Název zakázky	Přístroj
Krajská nemocnice T. Bati, a. s.	KNTB Zlín – Dovybavení lineárního urychlovače	Lineární urychlovač pro radioterapii
Ústřední vojenská nemocnice – Vojenská fakultní nemocnice Praha	ÚVN – VoFN Praha – Stereotaktické neuroradiochirurgické zařízení – nákup	Leksellův gama nůž
Fakultní nemocnice v Motole	FN Motol – Obměna lineárních urychlovačů - 3 ks	Lineární urychlovač pro radioterapii
Fakultní nemocnice Hradec Králové	FN Hradec Králové – nákup 3 ks lineárních urychlovačů včetně zajištění jejich servisu a provedení úprav stávajících ozařoven	Lineární urychlovač pro radioterapii

Krajská nemocnice T. Bati, a. s.	Měřicí zařízení pro ověřování plánů	Radiologické fantomy pro ověřování ozařovacích plánů
Jihomoravský kraj	Nemocnice Znojmo – obnova a modernizace přístrojového vybavení vybraných pracovišť – rozšíření a modifikace plánovacího systému léčby pro ORKO	Plánovací systém léčby onkologických pacientů

2.2 Veřejné zakázky a transparentnost

Proces zadávání veřejných zakázek zažívá v posledních letech tlak, ze strany veřejnosti i odborné veřejnosti, k co nejširšímu publikování dat z tohoto procesu za účelem zvýšení transparentnosti a efektivity při nakládání s veřejnými financemi [4]. Celkové náklady zemí EU na práci, zboží a služby byli v roce 2017 odhadovány na 2049,8 bilionů euro a mají stoupající tendenci [2]. V rámci České republiky jsou výdaje za rok 2020 odhadovány na 12,65 % HDP [3]. Je tedy patrné, že se veřejné zakázky podílejí významným dílem na veřejných výdajích. Příkladem snahy o zprůhlednění procesu zadávání může být zadávací směrnice EU 2014/24/EU a její implementace do českého systému v podobě zákona č. 134/2016 Sb., které dávají za povinnost shromažďovat a uveřejňovat data z procesu zadávání veřejných zakázek.

2.2.1 Otevřená data

Otevřená data jsou zakotvena v české legislativě zákonem č. 106/1999 Sb. O svobodném přístupu k informacím, konkrétně § 3 odst. 11 udává: „*Otevřenými daty se pro účely tohoto zákona rozumí informace zveřejňované způsobem umožňujícím dálkový přístup v otevřeném a strojově čitelném formátu, jejichž způsob ani účel následného využití není omezen a které jsou evidovány v národním katalogu otevřených dat.*“ Otevřená data se považují za efektivní nástroj k zveřejňování informací z veřejného sektoru, tedy i ze sektoru veřejných zakázek [32]. Dle „Standardů publikace a katalogizace otevřených dat VS ČR“ vydané Ministerstvem vnitra ČR [6] mají být otevřená data:

- úplná, tj. maximální rozsah dat a co nejaktuálnější,
- snadno dostupná, tj. dohledatelná běžnými nástroji a uživateli,
- strojově čitelná, tj. lze je zpracovat pomocí programovacích aplikací,
- používající standardy s volně dostupnou specifikací, tj. používat otevřené standardy, které jsou bezplatně a volně přístupné
- zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení, tj. umožnit využití pro komerční i nekomerční účely.

Otevřená data a veřejné zakázky v ČR

MMR prostřednictvím portálu ISVZ zveřejňuje agregovaná data z VVZ ve formátu zveřejňovaném na přechodím ISVZ a v novém formátu RAV. Agregovaná data ve

formátu zveřejňovaném na přechodím ISVZ splňují podmínky pro definici otevřených dat. Struktura těchto dat je podobná té, kterou publikuje Evropská unie ve svém modulu TED, protože česká legislativa vychází z té evropské a je jí vázána. Struktura těchto dat je ale stále špatně čitelná. Například publikované sety na ISVZ jsou seskupeny do logicky souvisejících sekcí. Tyto sekce jsou na sebe navázané většinou pomocí evidenčního čísla zakázky, ale pokud je například zakázka rozdělena na části, odkazuje se evidenční číslo zakázky na více částí. Dále chybí implementace opravných formulářů. Těchto sekcí je v datové sadě celkem pět a při stažení datového setu je práce s těmito surovými daty a jejich propojení velmi neintuitivní [25]. Tento problém je obecně řešen pomocí ontologií, pomocí kterých jsou data dále zpracována a logicky utříděna [33, 34]. Navíc dále umožňují rozšíření těchto dat o další informace [4, 33, 34].

Ontologie dat veřejných zakázek

V rámci snahy uspořádat, a ještě zvýšit transparentnost dat z veřejných zakázek vzniklo již několik ontologií k jejich zpracování [4]:

- Open Contracting Data Standard (OCDS)
- TED eSenders
- PPROC ontology
- LOTED2 ontology
- PCO ontology
- MOLDEAS ontology

Společným znakem těchto ontologií je jejich nehomogenita. Každá byla vytvořena pro trochu jiný účel a jejich aplikace není standardizována. Proto vznikají iniciativy pro implementaci již existujících ontologií s dalšími otevřenými daty a jejich aplikování do webového rozhraní, pro snadné použití koncovými uživateli [34]. Příkladem může být iniciativa TheyBuyForYou (<https://theybuyforyou.eu/>), která si klade za cíl vytvořit uzel pro sémantický web, který poskytne zadavatelům, dodavatelům a dalším zájemcům informace o trhu a analýzu trhu [35].

2.2.2 Analýza dat z veřejných zakázek

Při analýze dat získaných z portálů veřejných zakázek nepostupuje každá výzkumná skupina pomocí ontologií. Tento přístup používají hlavně skupiny, které se zaměřují na další obohacení dat jejich implementací do webového rozhraní a jejich standardizaci. Při jednoduché analýze dat postupují někteří autoři cestou hrubé analýzy datových setů, při vybrání konkrétních parametrů, které jsou pro ně zásadní. Souhrn a přehled studií, které využívají otevřená data z veřejných zakázek je v tabulce 2.3.

Tabulka 2.3: Souhrnný přehled studií z databáze Scopus, které používají otevřená data veřejných zakázek k analýze

Název	Rok vydání	Autor	Cíl	Zdroj dat	Zpracování dat
Linked Open Data for Public Procurement	2014	Vojtěch Svátek, et al	Vytvoření ontologie PCO pro webovou aplikaci, která by umožňovala dynamické párování dodavatelů a zadavatelů na základě stanovených kritérií.	TED, VVZ, Věstník UK, Věstník US, Věstník PL	CSV, XML, HTML převedena na RDF (Resource description framework)
PPROC, an Ontology for Transparency in Public Procurement	2016	José Félix, et al	Vytvoření ontologie PPROC pro potřebu španělských samospráv.	Lokální databáze španělských samospráv	JSON, CSV, XML převedeno na RDF
OpenTED Browser: Insights into European Public Spendings	2016	Yann-Ael Le Borgne, et al	Vytvoření webové aplikace OpenTED browser pro interaktivní vyhledávání v databázi TED.	TED + TED contract award notices 2005–2015 (CAN)	CSV, zpracování v Python, Webová aplikace v R Shiny
Pricing of HPV vaccines in European tender-based settings	2018	Venetia Qundri, et al	Ověřuje hypotézu, zda jetli je zadavatel státní, nebo regionální, ovlivňuje výslednou cenu vakcíny.	TED 2007-2018, filtr pomocí výrazů “HPV”, “papilloma”	Systematické ruční filtrování dat na základě zadaných kritérií
TheyBuyForYou platform and knowledge graph: Expanding horizons in public procurement with open linked data	2022	Ahmet Soylu, et al	Zpráva o stavu a pokroku iniciativy TheyBuyForYou a její další směřování.	Data z TED a národních věstníků od třetí strany OpenOpps a OpenCorporates	Zpracování pomocí API a Web scrapping
Regional tenders on biosimilars in Italy: An empirical analysis of awarded prices	2014	Sandro Curtoa, et al	Prozkoumat výslednou cenu zakázky na generická léčiva a z toho vyhodnotit míru konkurence.	TED a regionální věstníky samospráv	Filtrování pomocí CPV kódů a NUTS
DECENTRALIZATION AND PUBLIC PROCUREMENT PERFORMANCE: NEW EVIDENCE FROM ITALY	2020	Raphael Duguay, et al	Zkoumají, jak přístup k datům o veřejných zakázkách ovlivňuje konkurenci, ceny, zadávání zakázek ve státní správě v Itálii.	TED prostřednictvím Digiwhist	Filtrování pomocí kritérií CPV a ceny zakázky
The factors causing delays in public procurement: the Czech Republic versus the UK	2019	Michal Plaček, et al	Porovnání délky zadávacího řízení pro veřejné zakázky na stavební práce v ČR a UK a nalezení ovlivňujících faktorů.	TED 2010-2015 na stavební práce	Další zpracování dat není popsáno

Does the Lowest Bid Price Evaluation Criterion Make for a More Efficient Public Procurement Selection Criterion? (Case of the Czech Republic)	2015	František Ochrana, et. Al	Ověření hypotézy, že při zvolení jen jednoho hodnotícího kritéria, a to nejnižší ceny, dochází k zvýšení ceny veřejné zakázky v ČR	Náhodný výběr zakázek z VVZ a portálu ContractPLUS	Vyfiltrování neúplných zakázek
Another One Bites the Dust: Recurrent Procurement Contracts in the Czech Republic	2019	Michal Plaček, et al	Identifikace faktorů, které ovlivňují opakované zadání veřejné zakázky a zda toto opakované zadání má dopad na konečnou cenu.	VVZ zakázky na stavební práce 2014-1017	Další zpracování dat není popsáno
FACTORS AFFECTING THE LENGTH OF PROCEDURE IN PUBLIC PROCUREMENT: THE CASE OF THE CZECH REPUBLIC	2019	Michal Plaček, et al	Identifikace faktorů, které mají vliv na délku zadávacího řízení veřejné zakázky na stavební práce.	VVZ zakázky na stavební práce 2010-2015	Další zpracování dat není popsáno
Analysis of Factors of Overpricing in Public Procurement: A Study for Low-performing EU Countries	2020	Michal Plaček, et al	Identifikace faktorů, které způsobují předražení veřejné zakázky v 11 "rozvojových" evropských zemích	TED zakázky na stavební práce pro rok 2016 pro 11 dotčených zemí	Další zpracování dat není popsáno
The impact of agreement on government procurement use on the competition in Slovak healthcare sector	2021	Beáta Gavurová, et al	Vliv použití Agreement on Government Procurement (GPA) při zadávání VZ na počet nabídek a úspory pro sektor zdravotní péče na Slovensku	Zadané VZ v sektoru zdravotní péče na Slovensku v roce 2019 z EU Open Data Portal	Další zpracování dat není popsáno
Efficiency of Public Procurement in the Czech and Slovak Health Care Sectors	2021	Juraj Nemeč, et al	Vliv počtu nabídek a zvolených hodnotících kritérií na velikost úspor. Vliv fiktivních úspor na analýzu.	Zadané VZ v sektoru zdravotní péče 2014-2019 z veřejných registrů VZ	Vyloučení výsledků s úsporou menší, nebo větší než 60 %

Přístup autorů k zpracování dat se zásadně liší v závislosti na cíli práce. Autoři zabývající se zpracováním ontologií a jejich implikací [33, 35, 36] upravují data do dále zpracovatelných formátů a rozšiřují možnost jejich použití. Jakousi nadstavbou jsou autoři [35, 37], kteří dále používají vzniklé ontologie, kombinují je a dále je implementují do funkčních aplikací s přidanou hodnotou pro konečného uživatele. Na druhou stranu autoři, kteří se zaměřují primárně na analýzu dat [7–16] požívají jednoduché postupy

ručního třídění dat, případně jednoduchých vyhledávacích výrazů. Nevýhodou je, že takto zpracovaná data prakticky nelze použít v jiném případě, než na která byla vytvořena.

2.2.3 Analýza otevřených dat v ČR

V České republice se analýze trhu veřejných zakázek věnoval tým složený z akademických pracovníků i odborníků ze soukromé sféry. Mezi hlavní členy patří prof. PhDr. František Ochrana, DrSc., který působí na institutu sociologických studií při UK a Doc. Ing. Michal Plaček, Ph.D, MSc, působící pod stejným institutem. Tento tým se primárně zaměřoval na propojené efekty při procesu zadávání veřejných zakázek, přičemž jejich hlavním nástrojem byla regresní analýza [10–12]. Pro analýzu národních veřejných zakázek využívali primárně open data z VVZ. V současné době se tým věnuje spíše otázce managementu v soukromém sektoru se zaměřením na autonomii jednotlivých institucí, konkrétně muzeí [38] a dále například korupci v České postsovětské společnosti [39]. Odklání se tedy od problematiky veřejných zakázek k sociologicko-ekonomickým studiím.

Na české akademické půdě dále působil tým, který vede prof. Ing. Vojtěch Svátek, Dr, profesor aplikované informatiky na univerzitě Ekonomické v Praze. Tento tým stojí za vznikem ontologie PCO [33], který navrhli a využili pro spuštění webu k automatickému přiřazování dodavatelů a zadavatelů na základě zvolených kritérií v roce 2015. Tento tým je ale hlavně zaměřen na ontologie a aplikovanou informatiku. Na ontologickém poli tedy dále vykazují aktivitu, na poli veřejných zakázek již nikoliv.

V roce 2021 publikoval Nemeč et al článek o efektivitě veřejných zakázek v České a Slovenské republice. Zaměřuje se na míru úspory veřejných zakázek v sektoru zdravotní péče v závislosti na počtu podaných nabídek, použitých hodnotících kritériích a na druhu zadávacího řízení. Dále se zabývá otázkou fiktivních úspor u VZ, jejichž cena může být zadavatelem již od začátku nadhodnocena [16].

Ačkoliv existují publikace, které se zabývají analýzou veřejných zakázek v České republice a existuje publikace zabývající se analýzou veřejných zakázek v sektoru zdravotní péče v České a Slovenské republice, neexistují publikace, které by se věnovala analýze trhu veřejných zakázek v České republice v rámci zdravotnické techniky. Dále existují ontologie pro otevřená data veřejných zakázek, jejich rozpracování je ale někdy pro koncového uživatele až zbytečně složité a těžkopádné. To dobře demonstrují práce týmu pod vedením prof. Ing. Vojtěcha Svátka, Dr, který sice pracuje s otevřenými daty z veřejných zakázek, ale ke konkrétním aplikacím využívá jednoduchého principu stažení, a ručního protřídění dat, místo cíleného zpracování pomocí ontologií. Roztříštěnost ontologií a jejich odbornost tak ztěžuje jejich praktickou aplikaci při analýze dat.

2.3 Rozšíření otevřených dat pomocí nature language processing

Natural language processing (NLP) je skupina výpočetních technik pro automatickou analýzu a vyjádření přirozeného jazyka. Automatická analýza textu psaného člověkem je ale ve skutečnosti velmi složitá. Dnešní NLP metody dobře zvládají rozpoznávání slov, výrazů a do určité míry i vět. Mají ale problém při interpretaci větších celků, jako jsou například odstavce. Potenciál NLP je hlavně ve zpracování velkého množství dat, které lze nalézt na WWW. Ten obsahuje asi 20 biliónů stránek, což je potenciálně obrovský zdroj informací [40]. Jednou z technik NLP je named entity recognition (NER).

2.3.1 Named entity recognition

Technika NER označuje jednotlivá slova a zařazuje je do předem vytvořených kategorií. NER pracuje s výrazy v celku a je tedy náročné zvolit vhodný algoritmus, který vyhledá výrazy i v případě, že nejsou „doslovné“. Například pokud bychom měli kategorii „bolest na hrudi“ je potřeba aby algoritmus psanou větu „na hrudi pocíťoval lehkou bolest“ zařadil do této kategorie. Dále musí počítat s různými variantami slov, synonymy, odvozeninami a další [41]. Samotný proces NER se dá rozdělit do tří samostatných kroků a to identifikace hodnotného výrazu, sesbírání identifikovaných výrazů a rozřídění do daných kategorií [42].

Fuzzy-Matching

Aby bylo možné použít techniku NER pro vyhledávání a zařazování slov, které nejsou přesnou kopií databáze, je nutné implementovat některou z metod přibližné identifikace řetězce znaků. Jednou z těchto metod je fuzzy-matching [43]. Základní myšlenkou je spočítání číselné „vzdálenosti“ pro každou potenciální shodu mezi získaným výrazem a výrazy z databáze. Číselně lze tedy vyjádřit míru shody mezi získaným výrazem a výrazem z databáze a tím určit pravděpodobnost potenciální shody. Lze také určit hranici, kdy je číselná „vzdálenost“ mezi výrazy dostatečně malá a tím pádem je dvojice prohlášena za shodnou a naopak kdy příliš velká a dvojice je prohlášena za rozdílnou [44].

Veřejné zakázky a NER

Mehrbod [42, 45] ve svých publikacích používá systém automatického učení k vytvoření jednotného portálu pro TED a United Nations Global Marketplace (UNGM). Rozpoznává entity CPV slovníku a UNSPSC slovníku a jejich kombinaci k zpřesnění vyhledávání zakázek. Slovníky CPV a UNSPSC se totiž v některých skupinách překrývají, v jiných se ale liší. Vytvořením kategorizačního slovníku, založeného na těchto dvou propojených slovnících tak může uživatel nalézt více odpovídajících nabídek v portálech TED a UNGM.

2.3.2 Kategorizace zdravotnické techniky

V České republice je kategorizace zdravotnické techniky řešena Ústavem zdravotnických informací a statistiky ČR. Ta v rámci Pracovní skupiny pro kategorizaci a úhradovou regulaci zdravotnických prostředků vytvořila „Kategorizaci zdravotnické techniky“. Ta byla vyhlášena oficiálním klasifikačním systémem pro sběr statistických dat sdělením 105/2018 Sb. Českého statistického úřadu [46]. Tento klasifikační systém se průběžně upravuje a doplňuje. K datu 1. 5. 2022 je aktuální verze 5.0 z 9. 9. 2021 [47].

Důvodem pro vytvoření této kategorizace byla snaha o zvýšení dynamičnosti a pružnosti kategorizace zdravotnických prostředků s ohledem na rychlý vývoj v tomto odvětví. Zdravotnické prostředky jsou rozděleny do oddílů na základě jejich použití. Ty se dále dělí na skupiny a ty dále na hladiny. Obecně je tento systém velmi podrobný a obsáhlý [47].

3 Cíle práce

Prvním cílem diplomové práce je dopracování automatizace sběru a prvotního zpracování dat o nákupech zdravotnické techniky z veřejných zakázek, pomocí programovacího jazyka R. Druhým cílem je rozšíření této datové sady. Třetím cílem je provedení deskriptivní a regresní analýzy. Čtvrtým cílem je rozšíření dat o VZ o kategorii zdravotnické techniky pomocí techniky NLP.

3.1 Automatizace sběru a prvotního zpracování dat

Prvním cílem je dopracování skriptu pro automatický sběr dat o veřejných zakázkách z portálu ISVZ a dopracování algoritmu pro jejich předzpracování do datového setu, který poté bude možno zpracovat v programovém prostředí RStudio. Dále je cílem pomocí programovacího jazyka R provést filtraci získaného datového setu a omezit jej pouze na veřejné zakázky na zdravotnickou techniku.

3.2 Rozšíření datové sady

Druhým cílem je rozšíření získaného datového setu z prvního cíle a tím celý datový set obohatit. Rozšíření se bude zaměřovat na přidané informace o dodavateli a zadavateli z rejstříku ekonomických subjektů (RES) a dále na informace o samotných zakázkách z formulářů z VVZ. Tyto data budou získána pomocí metody webscrapping.

3.3 Deskriptivní a regresní analýza

Třetím cílem je vytvoření datového setu pro účely provedení deskriptivní a regresní analýzy. Deskriptivní analýza bude provedena se zaměřením na rozdělení trhu a znázornění vztahů dodavatelů a zadavatelů pomocí alluvial diagramů. Regresní analýza bude zaměřena na vztah zvolených nezávislých veličin k závislé veličině úspore zakázky. Analýzy budou provedeny pomocí programovacího jazyka R. Součástí tohoto cíle je také vypracování codebooku k danému datovému setu.

3.4 Rozšíření datového setu o kategorii zdravotnické techniky

Čtvrtým cílem bude prověření možnosti rozšíření datového setu o atribut kategorie zdravotnické techniky. Tento atribut vychází z kategorizace zdravotnické techniky, která je zpracována ÚZIS. Pro rozšíření bude otestována technika NER, konkrétně metoda fuzzy-matching.

4 Metody

V rámci metod jsou popsány postupy získání datových setů a jejich zpracování. Dále jsou popsány metody deskriptivní a regresní analýzy použité pro získání výsledků diplomové práce. Nakonec jsou popsány metody navržené pro aplikaci NER rámci algoritmu pro kategorizaci zdravotnické techniky.

Skripty pro zpracování a analýzu datových sad jsou vytvořeny programovacím jazykem a prostředím pro statistickou analýzu R. R je primárně odvozeno od programovacího jazyku a prostředí S (statistical programming language). R je otevřená platforma a je možné využívat doplňkové balíčky k rozšíření funkcí a usnadnění používání. Tyto balíčky lze bezplatně stáhnout na portálu CRAN mirror. R je freeware který spadá pod licenci GNU GPL v2. [48]

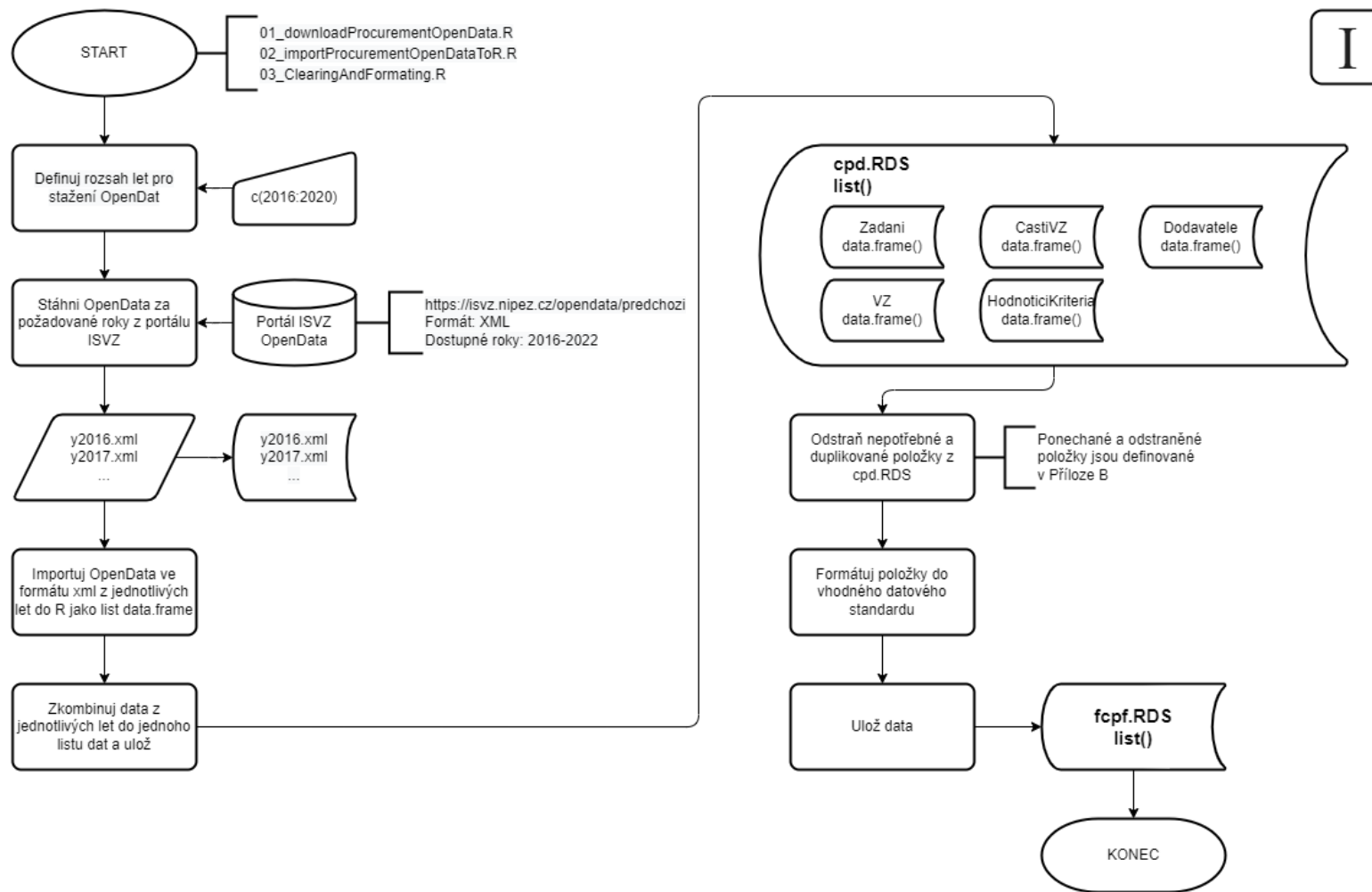
K psaní skriptů v programovacím jazyce R byl použit program RStudio. Jedná se o program, který poskytuje přívětivější podmínky pro používání programovacího jazyka R. [49].

4.1 Automatizace sběru a prvotního zpracování dat o veřejných zakázkách

Prvotní získání dat a jejich import byl řešen vedoucím této práce Ing. Glebem Doninem, Ph.D, v rámci diplomové práce Ing. Balounové [50]. V rámci práce Ing. Balounové byl vytvořen prvotní skript, který jsem ve své práci s dovolením vedoucího práce převzal a dále upravil, aby reflektoval změny v systému vydávání otevřených datových sad na VZ z ISVZ. Filtrace VZ se zaměřením na zdravotnickou techniku je originální prací autora.

4.1.1 Automatizace stažení a prvotního zpracování open dat veřejných zakázek a jejich import do R

Data jsou stažena z portálu ISVZ ve formátu XML. Tato část algoritmu je upravena, aby reflektovala nový systém vydávání datových sad v rámci ISVZ. Algoritmus pro import dat do prostředí RStudia a prvotní formátování je převzat z práce Ing. Balounové [50]. Při importu dojde k převodu dat z XML do data frame a jsou upraveny datové typy u některých položek. Byly přidány nové funkce, jako například automatické určení časového rozmezí datového setu dle aktuálního data a byl vylepšen algoritmus pro správu datových sad. Logika a postup zpracování dat jsou znázorněny ve vývojovém diagramu I, obrázek 4.1. Seznam ponechaných a vyřazených položek ve vytvořeném datovém setu z open dat z ISVZ je k dispozici v příloze B.



Obrázek 4.1: Vývojový diagram k algoritmu pro automatizaci stažení a prvotní zpracování open dat veřejných zakázek a jejich import do R.

Výše zmíněný algoritmus je uskutečněn v rámci tří skriptů:

- 01_downloadProcurementOpenData.R
- 02_importProcurementOpenDataToR.R
- 03_ClearingAndFormating.R

4.1.2 Filtrace a zpracování se zaměřením na zdravotnickou techniku

Datový set je dále filtrován se zaměřením na zdravotnickou techniku. Tato filtrace se skládá ze tří kroků a tyto kroky jsou prováděny v rámci dvou skriptů:

- 04_NRPZS+PZS.R
- 05_filteringForMD.R

Logika a postup zpracování dat jsou znázorněny ve výbojovém diagramu II v příloze C.

Prvním krokem je filtrace pomocí CPV kódů. Do výsledného datového setu jsou filtrovány pouze zakázky, které obsahují alespoň jeden z CPV kódu uvedených v tabulce 4.1. Tento CPV kód může být zadán buď u samotné zakázky, nebo její části. Algoritmus započítává všechny varianty.

Tabulka 4.1: Třídy CPV použité k filtraci zdravotnické techniky z otevřených datových sad z ISVZ. Všechny nižší úrovně vyznačených skupin jsou také zahrnuty do filtrace. Tedy například pro třídu 33150000 budou zahrnuty i všechny kategorie všech příslušných řádů.

Úroveň	CPV	Popis
Oddíl	33000000	Zdravotnické přístroje, farmaceutika a prostředky pro osobní péči
Skupina	33100000	Zdravotnické přístroje
Třída	33110000	Zobrazovací přístroje pro lékařské, stomatologické nebo veterinární účely
Třída	33120000	Záznamové systémy a vyšetřovací přístroje
Třída	33130000	Zubolékařské nástroje a přístroje a nástroje a přístroje podspecializací
Kategorie I	33131000	Zubolékařské ruční nástroje
Kategorie II	33131500	Zubolékařský extrakční nástroj
Kategorie I	33135000	Ortodontické přístroje
Kategorie I	33136000	Rotační a brusné nástroje
Kategorie I	33138000	Pomůcky a přístroje pro zubní náhrady
Třída	33150000	Přístroje pro radioterapii, mechanoterapii, elektroterapii a fyzioterapii
Třída	33160000	Operační technika
Kategorie I	33161000	Elektrochirurgická jednotka
Kategorie I	33162000	Přístroje a nástroje pro operační sály
Kategorie II	33162100	Přístroje pro operační sály
Kategorie I	33164000	Helioskopické přístroje
Kategorie I	33165000	Kryochirurgické a kryoterapeutické přístroje
Kategorie I	33166000	Dermatologické přístroje
Kategorie I	33167000	Chirurgická světla

Kategorie I	33168000	Endoskopie, endochirurgické přístroje
Kategorie I	33169000	Chirurgické nástroje
Kategorie II	33169100	Chirurgický laser
Třída	33170000	Anestézie a resuscitace
Kategorie I	33172000	Anesteziologické a resuscitační přístroje
Třída	33180000	Podpora funkcí
Kategorie I	33181000	Přístroje pro podporu funkce ledvin
Kategorie II	33181100	Přístroje pro hemodialýzu
Kategorie II	33181300	Individuální monitory pro hemodialýzu
Kategorie II	33181400	Vícepokojová hemodialýza
Kategorie I	33182000	Přístroje pro podporu srdeční funkce
Kategorie II	33182100	Defibrilátor
Kategorie II	33182200	Přístroje pro stimulaci srdeční funkce
Kategorie III	33182210	Kardiostimulátor
Kategorie II	33182300	Přístroje pro srdeční chirurgii
Kategorie II	33182400	Kardiologický rentgenový systém
Kategorie I	33185000	Naslouchadla
Kategorie II	33185200	Kochleární implantát
Kategorie II	33185300	Otolaryngologický implantát
Kategorie I	33186000	Jednotka pro mimotělní oběh
Třída	33190000	Různé zdravotnické přístroje a výrobky
Kategorie I	33191000	Sterilizační, desinfekční a hygienické přístroje
Kategorie I	33192000	Lékařský nábytek
Kategorie II	33192100	Lůžka pro použití ve zdravotnictví
Kategorie II	33192200	Lékařské stoly
Kategorie II	33192400	Zubolékařské pracovní stanice
Kategorie I	33194000	Přístroje a nástroje pro transfúzi a infúzi
Kategorie II	33194100	Přístroje a nástroje pro infúzi
Kategorie III	33194110	Infuzní čerpadla
Kategorie II	33194200	Přístroje a nástroje pro transfúzi
Kategorie III	33194210	Přístroje pro transfúzi krve
Kategorie I	33195000	Systémy pro kontrolu pacientů
Kategorie I	33197000	Počítačové vybavení pro zdravotnictví

Druhým krokem je filtrace pomocí číselníku z Národního registru poskytovatelů zdravotních služeb (NRPZS), který vydává ÚZIS a seznamu smluvních poskytovatelů zdravotních služeb (PZS), vydávaného Všeobecnou zdravotní pojišťovnou (VZP). Pokud se IČO zadavatele VZ neshoduje s IČO subjektů z NRPZS anebo PZS, je tato zakázka vyřazena, jelikož by mohlo jít o subjekt, který nebude zdravotnickou techniku používat k poskytování zdravotní péče. Typicky se jedná například o veterinární kliniky, veterinární ambulance a veterinární lékaře. Seznam PZS je doplněním číselníku NRPZS, protože ten obsahuje pouze omezené informace. Obsah NRPZS je totiž dle zákona č. 372/2011 Sb., o zdravotních službách veřejně přístupný na internetových stránkách s výjimkou údajů o:

- poskytovateli, o jehož oprávnění k poskytování zdravotních služeb rozhodlo Ministerstvo obrany, Ministerstvo spravedlnosti nebo Ministerstvo vnitra [51].

Tedy, zdravotnická zařízení provozovaná výše zmíněnými institucemi v NRPZS nenajdeme. Jedná se například o Ústřední vojenskou nemocnici, Vojenskou nemocnici Brno, Vojenské rehabilitační a lázeňské ústavy, Správu státních a hmotných rezerv, zařízení vězeňské služby a další.

Třetím krokem je filtrace zakázek, které mají ukončené řízení. Zakázka je poté buď zadána, nebo zrušena. Každá taková zakázka musí mít formulář CZ03, nebo F03. Zakázky, které tyto formuláře nemají jsou tedy automaticky vyřazeny, protože se nehodí k další analýze z důvodu chybějících informací. Výsledným výstupem je datový set MDcpd, který je dále používám k vytvoření datového setu pro účely analýzy.

4.2 Rozšíření datové sady pomocí metody webscrapping

Metoda webscrapping je použita k obohacení datových sad z RES a formulářů VVZ. V případě RES jde primárně o informace o zadavateli a dodavateli. V případě formulářů VVZ jde o získání potencionálně zajímavých položek, které nejsou součástí otevřených datových setů z ISVZ, například zda byla zakázka financována z EU.

4.2.1 Rozšíření z databáze RES

Pomocí metody webscrapping je na základě IČO subjektu z registru RES získána tabulka REStable pro rozšíření informací o zadavateli a dodavateli zakázky. Tato tabulka je získána pomocí jednoho skriptu:

- 08_ObohaceniRES.R

Logika a postup zpracování dat jsou znázorněny ve výbojovém diagramu III v příloze C.

Z datové sady MDcpd jsou vybrána IČO všech zadavatelů a dodavatelů zakázek. Tato IČO jsou poté automaticky postupně zadávána do portálu RES přes virtuální proces prohlížeče Chrome, který je spuštěn přes RStudio (balíček RSelenium). Virtuální proces prohlížeče je spuštěn v programu Docker. Po přístupu na stránku subjektu jsou z RES pomocí metody webscrapping stažena požadovaná data: ico, adresa, okres, okresKod, zakladniUzmeniJednotka, zakladniUzmeniJednotkaKod, pravniForma, pravnickaFormaKod, sektor, sektorKod, velikost, velikostKod. Ty jsou následně uložena do souhrnné tabulky REStabel. Klíčem k propojení s dalšími tabulkami je IČO subjektu. Tato tabulka je použita k rozšíření datové sady určené k deskriptivní a regresní analýze.

4.2.2 Rozšíření z databáze formulářů VVZ

Stažení stránek zakázek a formulářů zakázek z VVZ bylo řešeno vedoucím této práce Ing. Glebem Doninem, Ph.D, v rámci diplomové práce Ing. Balounové [50]. Tento algoritmus s dovořením vedoucího práce převzat a následně upraven autorem této práce. Algoritmus pro získání dat ze stránek zakázek a formulářů VVZ pomocí metody

webscrapping je originální prací autora. Proces stažení stránek, formulářů a vytěžení informací je součástí dvou skriptů:

- 09_DownloadFormsISVZ.R
- 10_ScrapForms.R

Logika a postup zpracování dat jsou znázorněny ve vývojovém diagramu IV v příloze C.

Po stažení všech formulářů z VVZ k zakázkám z datového setu MDcpd jsou metodou webscrapping získány požadované informace z formulářů typu CZ02, CZ03, CZ04, F02, F03, F14 a F20. Ty jsou dále uloženy do listu FORMtable a příslušných tabulek, pojmenovaných dle typu formuláře, ze kterého informace pocházejí. Obsah tabulek je popsán v tabulce 4.2. Tyto tabulky jsou dále použity pro filtraci a rozšíření datového setu pro účel deskriptivní a regresní analýzy.

Tabulka 4.2: Seznam a popis informací získaných z formulářů VVZ. Všechny tabulky data frame jednotlivých formulářů jsou součástí jednoho listu FORMtable.

FORMtable – list()	
CZ02tabel – data.frame()	
ECZ	Evidenční číslo zakázky na VVZ
ECF	Evidenční číslo formuláře na VVZ
typFormulare	Typ formuláře pro účel filtrování formulářů
financovanoEU	Je zakázka financována z EU?
ramcovaSmlouva	Je zakázka rámcovou smlouvou?
ramcovaSmlouvaJeden	<ul style="list-style-type: none"> • S jedním dodavatele?
ramcovaSmlouvaVice	<ul style="list-style-type: none"> • S více dodavateli?
DNS	Je zakázka dynamickým nákupním systémem?
CZ03tabel – data.frame()	
ECZ	Evidenční číslo zakázky na VVZ
ECF	Evidenční číslo formuláře na VVZ
typFormulare	Typ formuláře pro účel filtrování formulářů
financovanoEU	Je zakázka financována z EU?
zavedeniRS	Je formulář oznámení o uzavření RS?
uzavreniPodRS	Je zakázka uzavřena na základě RS?
zavedeniDNS	Je formulář oznámením o zavedení DNS?
uzavreniPodDNS	Je zakázka uzavřena na základě DNS?
CZ04tabel – data.frame()	
Typformulare	Typ formuláře pro účel filtrování formulářů
ECF	Evidenční číslo formuláře na VVZ
zmenaTextu	Propíše, ve kterém oddílu formuláře byl změněn text
zmenaCPV	Propíše, ve kterém oddílu formuláře bylo změněno CPVt
zmenaData	Propíše, ve kterém oddílu formuláře bylo změněno datum
zmenaHodnoty	Propíše, ve kterém oddílu formuláře byla změněna jiná hodnota (např. CPV)
F02table – data.frame ()	
ECZ	Evidenční číslo zakázky na VVZ
ECF	Evidenční číslo formuláře na VVZ
typFormulare	Typ formuláře pro účel filtrování formulářů

financovanoEU	Je zakázka financována z EU?
ramcovaSmlouva	Je zakázka rámcovou smlouvou?
ramcovaSmlouvaJeden	S jedním dodavatele?
ramcovaSmlouvaVice	S více dodavateli?
DNS	Je zakázka dynamickým nákupním systémem?
F14table – data.frame ()	
typFormulare	Typ formuláře pro účel filtrování formulářů
ECF	Evidenční číslo formuláře na VVZ
zmenaTextu	Propíše, ve kterém oddílu formuláře byl změněn text
zmenaCPV	Propíše, ve kterém oddílu formuláře bylo změněno CPVt
zmenaData	Propíše, ve kterém oddílu formuláře bylo změněno datum
zmenaHodnoty	Propíše, ve kterém oddílu formuláře byla změněna jiná hodnota (např. CPV)
F20table – data.frame ()	
typFormulare	Typ formuláře pro účel filtrování formulářů
ECF	Evidenční číslo formuláře na VVZ
textZmena	Komentář ke změně hodnoty zakázky
hodnotaPuvodni	Původní vysoutěžená cena zadání
hodnotaNova	Upravená cena zadání

4.3 Metody deskriptivní a regresní analýzy

Pro účely deskriptivní a regresní analýzy bylo nutné vytvořit agregovaný datový set, který spojuje jednotlivé formuláře a informace do jasně určených jednotlivých zadání. Tato agregovaná tabulka FNtb vznikla propojením všech tabulek data frame z listu MDcpd, rozšířením o informace z tabulek REStable a FORMtable a následnou filtrací z tabulky FORMtable. Poté byla provedena deskriptivní a regresní analýza pomocí dat z tohoto datového setu.

4.3.1 Vytvoření data setu pro účel analýz

Pro vytvoření agregované tabulky je nutné znát vztahy mezi jednotlivými tabulkami v listu MDcpd. Tyto vztahy jsou popsány v tabulce 4.3. Zároveň bylo nutné vyřadit národní formuláře CZ, jelikož neexistuje unikátní klíč, který by propojoval formuláře CZ03 zakázek, které jsou dělené na části, s formulářem CZ02 v části VZ. Prakticky tedy není možné určit předpokládanou hodnotu části VZ, a tedy nelze vypočítat úsporu, což je závislá proměnná pro účel regresní analýzy. Ze stejného důvodu jsou vyřazeny zakázky, které mají formulář F03, ale nemají předcházející formulář F02. Typicky se jedná o zakázky staršího data, kdy formulář F02 byl podán před rokem 2016, tudíž se nepropsal do získaného datového setu z ISVZ. Jsou také vyřazeny zakázky, které jsou zadány v rámci dynamického nákupního systému (DNS) nebo rámcové smlouvy (RS), jelikož nelze určit konečnou hodnotu zadání. Příprava agregovaného datového setu pro účel regresní analýzy je součástí dvou skriptů:

- 13_AgrTable.R
- 14_FinTable.R

Logika a postup zpracování dat jsou znázorněny ve vývojovém diagramu V v příloze C. K finální tabulce FNtb je vytvořen codebook s detailními informacemi o jednotlivých proměnných a jejich zdroji. Tento codebook je součástí přílohy D. Tabulka FNtb je rozšířena pomocí tabulky REStable a FORMtable, viz zmíněná příloha.

Tabulka 4.3: Vztahy mezi jednotlivými data frames v listu MDcpd s důrazem na propojení jednotlivých data frames pomocí unikátních klíčů za účelem vytvoření agregované tabulky FNtb

Zadani – data.frame()	
Unikátní ID	IDZadani*
Podmínky filtrace	Platný formulář Zadání části zakázky Formulář F03
Dodavatel – data.frame()	
Napojení na	Zadani
Klíč	IDZadani
VZ – data.frame()	
Napojení na	Zadani
Klíč	EvidencniCisloVZnaVVZ
Podmínky filtrace	Formulář F02 nebo F03 Platný formulář
CastiVZ – data.frame()	
Napojení na	Zadani
Klíč	NazevCastiVZ** CisloCastiVZ**
Podmínky filtrace	Platný formulář Formulář F02
HodnoticiKriteria – data.frame()	
Napojení na	CastVZ
Klíč	IDCastiVZ

*Unikátní ID, které určuje jednotlivé záznamy v agregované tabulce FNtb. **Nelze propojit přes jediný identifikátor. Jsou tedy porovnány názvy částí VZ. U zakázek, které se takto nepodaří propojit jsou následně porovnány čísla částí VZ. Zakázky, které se nepodaří propojit vůbec jsou vyřazené (vyřazené N=59).

4.3.2 Lineární regresní analýza

Regresní analýza je statistická metoda pro vyšetřování a modelování vztahu mezi proměnnými. Je to jedna z nejvíce využívaných technik k statistické analýze. Určeme vztah, který je na sobě pravděpodobně lineárně závislý. Vytváříme tedy hypotézu, kterou dále testujeme. Jako příklad může posloužit vztah mezi vzdáleností zákazníka od skladu vůči času doručení balíčku. Sesbíráme empirická data z N měření. Čím větší bude N , tím lépe bude model representovat skutečnost. Pokud bychom se pokusili vyjádřit čas doručení jako x a vzdálenost zákazníka od skladu jako y a zároveň předpokládáme lineární vztah těchto dvou veličin, potom můžeme tento vztah zjednodušit do lineární rovnice:

$$y = \beta_0 + \beta_1 x \quad (4.1)$$

kde β_0 je offset, β_1 je sklon, x je nezávislá proměnná a y je závislá proměnná. Model vytvořený na základě rovnice (4.1) je ale nedokonalý. Je jasné, že na čas doručení nemá vliv jen vzdálenost zákazníka od skladu, ale i další náhodné proměnné. V reálu tedy měření nebudou spadat přesně na přímkou, ale do jejího blízkého okolí. Pracujeme tedy s chybou ε . Reálnější model našeho problému lze vyjádřit rovnicí:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (4.2)$$

Rovnice (4.2) vyjadřuje model lineární regrese. Pokud obsahuje pouze jednu závislou veličinu, jedná se jednoduchý lineární regresní model. Obecně ale může být veličina y závislá na k nezávislých veličinách x . Tento vztah poté můžeme vyjádřit jako:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (4.3)$$

Rovnice (4.3) vyjadřuje model vícenásobné lineární regrese. Jednotlivé koeficienty $\beta_1, \beta_2, \dots, \beta_k$ odhadují sílu vlivu nezávislé veličiny na závislou. [52]

Pokud chceme analyzovat data pomocí metody vícenásobné lineární regresní analýzy, musíme splnit určité předpoklady:

- Závisle proměnná musí být metrická.
- Nezávislé proměnné musejí být také metrické, nebo kategorické. Pokud jsou kategorické je potřeba je převést do dichotomického tvaru vytvořením takzvaných „dummy“ proměnných.
- Nezávislé proměnné by mezi sebou neměly příliš korelovat, jelikož by tím porušili podmínku absence multikolinearity. Pokud v modelu existuje multikolinearita, jsou výsledky nespolehlivé a model nestabilní.
- Regresní analýza je citlivá na odlehle hodnoty. Ty musejí být z dat odstraněny.
- Proměnné musejí být v lineárním vztahu.
- Proměnné vykazují normální rozložení. To platí pouze pro metrické proměnné.
- Vztah proměnných nevykazuje heteroskedascitu. Chyba modelu musí být relativně konstantní. [53]

4.3.3 Vícenásobný lineární regresní model v R

Programovací jazyk R umožňuje velmi rychlé a jednoduché sestavení lineárního modelu dat. Následně je možné tento model a přesnost tohoto modelu interpretovat do tabulky a dále do grafů. Dále jsou uvedeny důležité výstupy z modelu a metody ověření jeho přesnosti.

Míra regrese

U jednotlivých nezávislých veličin je vypočtena míra regrese vůči nezávislé veličině. Ta je vyjádřena korelačními koeficienty $\beta_1, \beta_2, \dots, \beta_k$, které jsou interpretovány tabulkou. Naprosto zásadní hodnotou je p-hodnota u každé nezávisle proměnné. Ta určuje, zda je možné prohlásit vztah závislé a nezávislé veličiny za statisticky signifikantní. To samozřejmě závisí na zvolené hladině významnosti testované hypotézy. Pro naše účely volíme hladinu významnosti $\alpha = 0,05$. [54]

Index determinace R^2

Koeficient R^2 je dobrým ukazatelem kvality modelu. Udává z kolika procent variabilita nezávislé proměnné vysvětluje variabilitu závislé proměnné. Obvykle se snažíme nalézt model s hodnotou větší než 0,5. [54]

F-statistika

V případě regresní analýzy nám F-statistika říká, jaká je pravděpodobnost, že jsou všechny korelační koeficienty $\beta_1, \beta_2, \dots, \beta_k$ rovny nule. Tedy, že mezi závislou a nezávislou proměnnou neexistuje žádný korelační vztah. Pokud je p-hodnota tohoto testu nižší než námi určená hranice, můžeme zavrhnout nulovou hypotézu, že se korelační koeficienty rovnají nule. Pro naše účely volíme hladinu významnosti $\alpha = 0,05$. [55]

Shapiro-Wilkův test normality

Pro ověření normality závislé proměnné je možné použít Shapiro-Wilkův test. Ten proloží data regresní přímkou vzhledem k očekávaným hodnotám normálního rozdělení. P-hodnota musí být menší než určená hladina významnosti. Pro naše účely volíme hladinu významnosti $\alpha = 0,05$. [54]

Ověření rozložení chyby ε

Chyba ε by měla splňovat několik podmínek. Měla by být, pokud možno rovnoměrně rozložená okolo lineárního modelu. To znamená že odchylka by měla být konstantní ve všech částech a rozložení těchto odchylek by mělo být normální. Nástroji k ověření těchto předpokladů byly zvoleny tyto grafy:

- rozložení chyby ε vůči předpokládané hodnotě y
- histogram chyb ε
- Q-Q graf který vůči sobě zobrazuje kvantily chyb ε a teoretické kvantily normálního rozdělení. [54]

Pokud jsou oba datové sety normálně rozděleny, měly by body Q-Q grafu sledovat proloženou přímkou. Můžeme tedy vzít kvantily normálního rozdělení u kterých jsme si jisti normalitou a ověřit normalitu druhého datového setu. [56]

4.4 Algoritmus pro kategorizaci zdravotnické techniky

Algoritmus používá techniku NER, konkrétně metodu fuzzy-matching, k určení zařazení daného zadání do kategorie dle KZT. Principem je porovnání knihoven výrazů mezi daným zadáním a danou kategorií. Křížově je vypočítána pravděpodobnost shody jednotlivých slov z knihoven pomocí Jaro-Winklerovi vzdálenosti.

Jaro-Winklerova vzdálenost je číslo mezi 0 a 1, která měří rozdílnost mezi dvěma řetězci znaků. Je rovna 0, pokud mají oba řetězce délku 0 a je rovna 1 pokud se nevyskytují žádné shodné znaky mezi řetězcem A a B. Jaro-Winklerova vzdálenost je definována jako

$$jw = 1 - \left(\frac{1}{3}\right) \left(\frac{w_1 m}{|A|} + \frac{w_2 m}{|B|} + \frac{w_3(m-t)}{m}\right) \quad (4.4)$$

kde $|A|$ značí počet znaků v řetězci A, m je počet znaků, které se shodují z řetězce A vůči řetězci B a t je počet transpozicí shodných znaků, w_i je váha asociovaná se znaky v řetězci A, řetězci B a s transpozicí. Znak c , vybraný ze shod mezi řetězci A a B se shoduje s řetězcem B pokud se znak c nachází v řetězci B a zároveň pokud je rozdíl polohy znaku c menší než

$$\frac{\max(|A|, |B|)}{2} - 1 \quad (4.5)$$

od polohy znaku c v řetězci B (příčemž použijeme integrální dělení). Dva shodné znaky jsou transponovány, pokud se nacházejí v řetězcích A i B, ale na rozdílných pozicích. [57]

Po výpočtu pravděpodobností shody jednotlivých kategorií algoritmus vyhodnotí nejvyšší pravděpodobnost a zapíše ji do tabulky. Vytvoření knihoven výrazů neboli parsing jednotlivých zadání a kategorií a samotný proces fuzzy-matching a zápis výsledků je součástí dvou skriptů:

- 11_NERformatting.R
- 12_NERprocessing.R

Logika a postup zpracování dat pro tvorbu knihoven výrazů, parsing, je znázorněn ve vývojovém diagramu VI v příloze C. Logika a postup zpracování dat pro fuzzy-matching je znázorněn ve vývojovém diagramu VII v příloze C.

Samotné testování algoritmu pro automatickou kategorizaci předmětu VZ dle KZT je provedeno na výběru 43 zadání z tabulky FNtb, kterým byla ručně přiřazena jedna z vybraných kategorií KZT, viz tabulka 4.4. Výsledek algoritmu je poté porovnán s ručně přiřazenou kategorií a je vyhodnoceno, zda algoritmus přiřadil kategorii zadání správně, nebo udělal chybu. Tento algoritmus je v testovací fázi vývoje. Je tedy provedeno několik kol vyhodnocení s iteracemi určitých parametrů, čímž se ověří vliv těchto parametrů na úspěšnost algoritmu.

Tabulka 4.4: Popis kategorií KZT použitých při testování algoritmu pro automatické přiřazení kategorie k VZ.

Název	Zkratka	Kód	Úroveň hierarchie	Poznámka
Rentgenové přístroje	RTG	D.1.1.	2	Diagnostické, vč. CT
Ultrazvukové měřicí/zobrazovací systémy	UZ	D.1.3.	2	Diagnostické
Monitory vitálních funkcí	ANEST	D.4.1.	2	Vč. centrálního monitoringu
Přístroje pro podporu a udržování respiračního systému	MON	T.3.1.	2	Ventilátory
Anesteziologické přístroje	VENT	T.5.4.	2	

5 Výsledky

Výstupem diplomové práce jsou algoritmy, jejichž výstup je pět datových setů, které byly dále využity k vytvoření agregovaného setu, nebo k analýze. Tyto datové sety jsou:

- `list(MDcpd)`
- `data.frame(REStable)`
- `data.frame(FORMtable)`
- `data.frame(NERcpd)`
- `data.frame(FNtb)`

Datový set `data.frame(FNtb)` byl vytvořen kombinací datových setů `list(MDcpd)`, `data.frame(REStable)` a `data.frame(FORMtable)` a byl použit pro deskriptivní a regresní analýzu. K tomuto setu byl vytvořen codebook, viz příloha D. Regresní analýza je zaměřena na vztah mezi úsporou, jako závislou veličinou, a počtem nabídek a počtem kritérií, jako nezávislé veličiny. Bylo provedeno ověření vícenásobného lineárního regresního modelu. Dále bylo provedeno testování algoritmu pro automatickou kategorizaci předmětu veřejných zakázek dle KZT.

5.1 Datové sady

Ze algoritmů byly získány požadovaná data. V jednom skriptu bylo zapotřebí použít virtualizaci. K tomu byl použit program `docker` s obrazem `standalone-chrome`. Kompletní projekt je k dispozici v rámci příloh této diplomové práce. Seznam jednotlivých skriptů je v tabulce 5.1. Balíčky použité v algoritmech pro R jsou vyjmenovány v tabulce 5.2.

Tabulka 5.1: Seznam skriptů použitých pro získání datových sad.

Název skriptu	Poznámka
00_functions.R	Převzato z [50]
01_downloadProcurementOpenData.R	Převzato z [50] a upraveno autorem
02_importProcurementOpenDataToR.R	Převzato z [50] a upraveno autorem
03_clearingAndFormating.R	Převzato z [50] a upraveno autorem
04_NRPZS+PZS.R	Originální práce autora
05_FilteringForMD.R	Originální práce autora
06_KZT.R	Originální práce autora
08_ObohaceniRES.R	Originální práce autora
09_DownloadFormsISVZ.R	Převzato z [50] a upraveno autorem
10_ScrapForms.R	Originální práce autora
11_NERformating.R	Originální práce autora
12_NERprocessing.R	Originální práce autora
13_AgrTable.R	Originální práce autora
14_FinTable.R	Originální práce autora

Tabulka 5.2: Seznam balíčků použitých ve skriptech pro získání datových sad a grafů

Balíčky použité pro získání datových sad
xml2, XML, dplyr, rvest, stringr, readxl, RSelenium, tictoc, udpipe, fuzzyjoin, tidyverse, ggplot2, broom, ggpubr, openxlsx
Balíčky použité pro získání grafů
dplyr, readxl, ggalluvial, openxlsx, RColorBrewer

5.1.1 Rozšiřující datová sady REStable

V rámci rozšíření datových sad byla vytvořena tabulka `data.frame(REStable)`. Položky této tabulky jsou popsány v tabulce 5.3. Tato tabulka byla získána pomocí algoritmu `08_ObohaceniRES.R`. Unikátním ukazatelem je IČO subjektu. Tabulka obsahuje 614 unikátních subjektů. V rámci skriptu byla použita virtualizace pomocí softwaru docker.

Tabulka 5.3: Názvy a popis jednotlivých proměnných v tabulce REStable

Sloupec	Popis
ico	Unikátní identifikátor subjektu
adresa	Úplná adresa subjektu
okres	Okres sídla subjektu
okresKod	Kód pro položku okres
zakladniUzemniJednotka	Základní územní jednotka sídla subjektu
zakladniUzemniJednotkaKod	Kód pro položku zakladniUzemniJednotka
pravniForma	Právní forma subjektu
pravnickaFormaKod	Kód pro položku pravniForma
sektor	Institucionální sektor: dle ESA2010
sektorKod	Kód pro položku sektor
velikost	Velikostní kat. dle počtu zaměstnanců
velikostKod	Kód pro položku velikost

Položka ico je unikátním klíčem tabulky. Byla získána z IČO zadavatelů a dodavatelů VZ z tabulky MDcpd. Zbytek dat pochází z databáze RES.

5.1.2 Rozšiřující datová sada FORMtable

V rámci rozšíření datových sad byl vytvořen list `list(FORMtable)`. Tento list obsahuje 6 tabulek `data.frame(F02table)`, `data.frame(F14table)`, `data.frame(F20table)`, `data.frame(CZ02table)`, `data.frame(CZ03table)`, `data.frame(CZ04table)`. Tento je popsán v tabulce 4.2. Tento list byl získán pomocí algoritmu `10_ScrapForms.R`.

5.1.3 Datová sada FNtb pro účel analýz

Byla vypracována tabulka `data.frame(FNtb)`, která spojuje jednotlivé tabulky z listu MDcpd do formátu použitelného k deskriptivní a regresní analýze. Propojení tabulek je popsáno v tabulce 4.3. Tato tabulka je detailně popsána v rámci vypracovaného

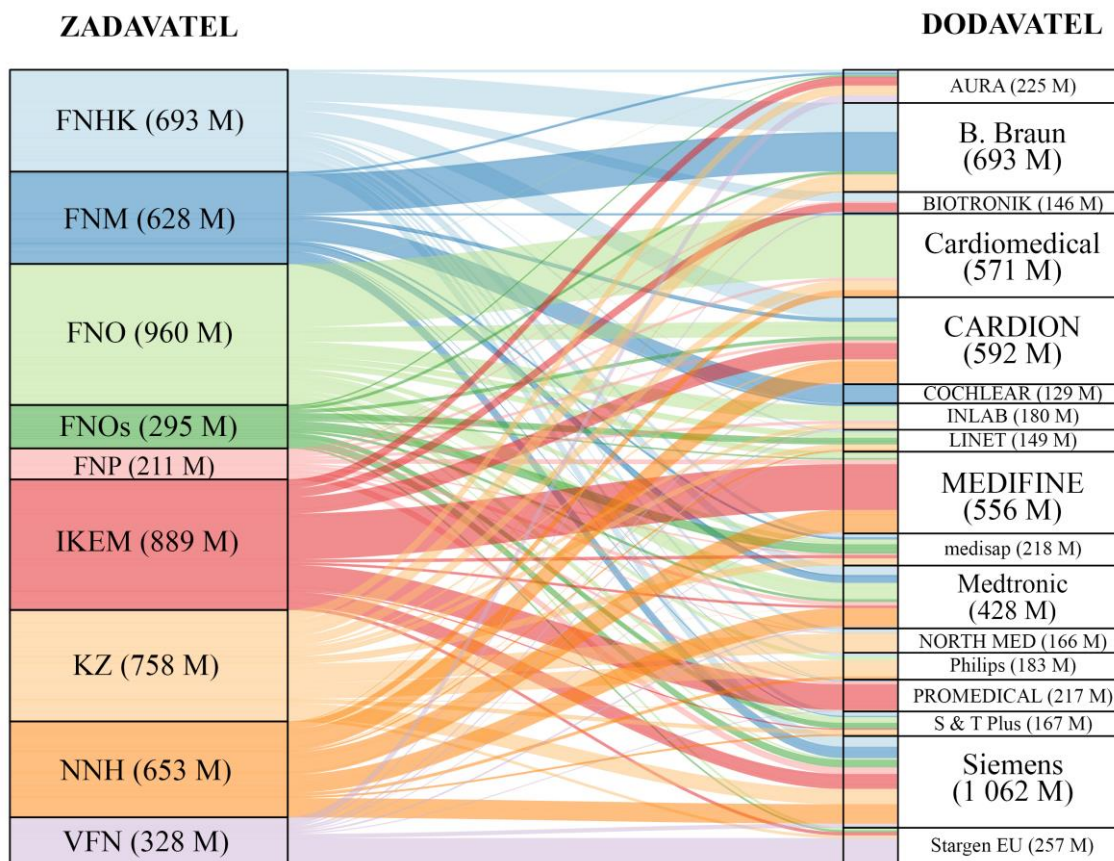
codebooku, viz příloha D. Tato tabulka byla získána pomocí algoritmů 13_AgrTable.R, 14_FinTable.R. Data jsou z let 2016–2020.

5.2 Deskriptivní analýza

Byla provedena deskriptivní analýza se zaměřením na rozdělení trhu jak z pohledu dodavatelů, tak z pohledu zadavatelů. K analýze byl použit datový set FNtb a REStable. Výsledky analýzy jsou prezentovány pomocí alluvial diagramů. Zkratky dodavatelů jsou v tabulce 5.4. Dále byly analyzovány zadavatelé a dodavatelé na trhu se zdravotnickou technikou dle počtu zaměstnanců.

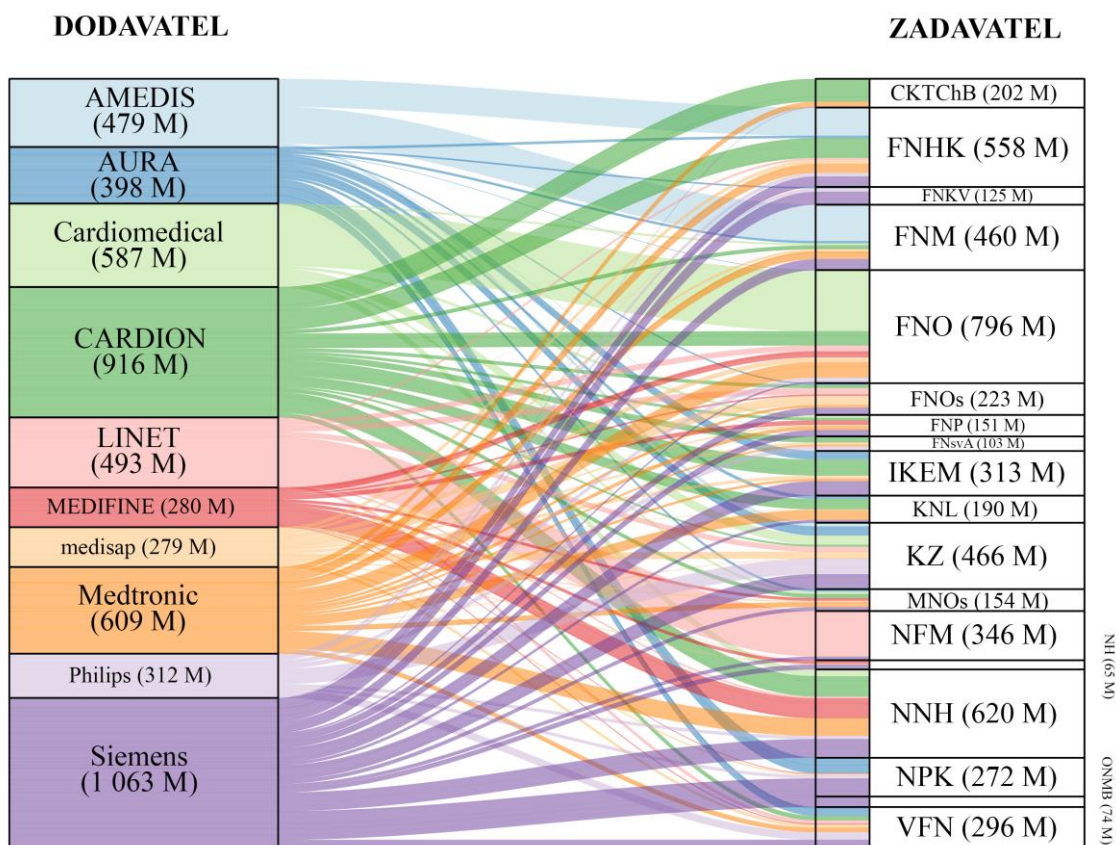
Tabulka 5.4: Zkratky pro subjekty zadavatelů VZ

Zkratka zadavatele	Název zadavatele
CKTChB	Centrum kardiovaskulární a transplantační chirurgie Brno
FNHK	FN Hradec Králové
FNKV	FN Královské Vinohrady
FNM	FN Motol
FNO	FN Olomouc
FNOs	FN Ostrava
FNP	FN Plzeň
FNsvA	FN u svaté Anny
IKEM	Institut klinické a experimentální medicíny
KNL	Krajská nemocnice Liberec
KZ	Krajská zdravotní
MNOs	Městská nemocnice Ostrava
NFM	Nemocnice ve Frýdku-Místku
NH	Nemocnice Havířov
NNH	Nemocnice na Homolce
NPK	Nemocnice Pardubického kraje
NT	Nemocnice Třebíč
ONMB	Oblastní nemocnice Mladá Boleslav
VFN	Všeobecná fakultní nemocnice



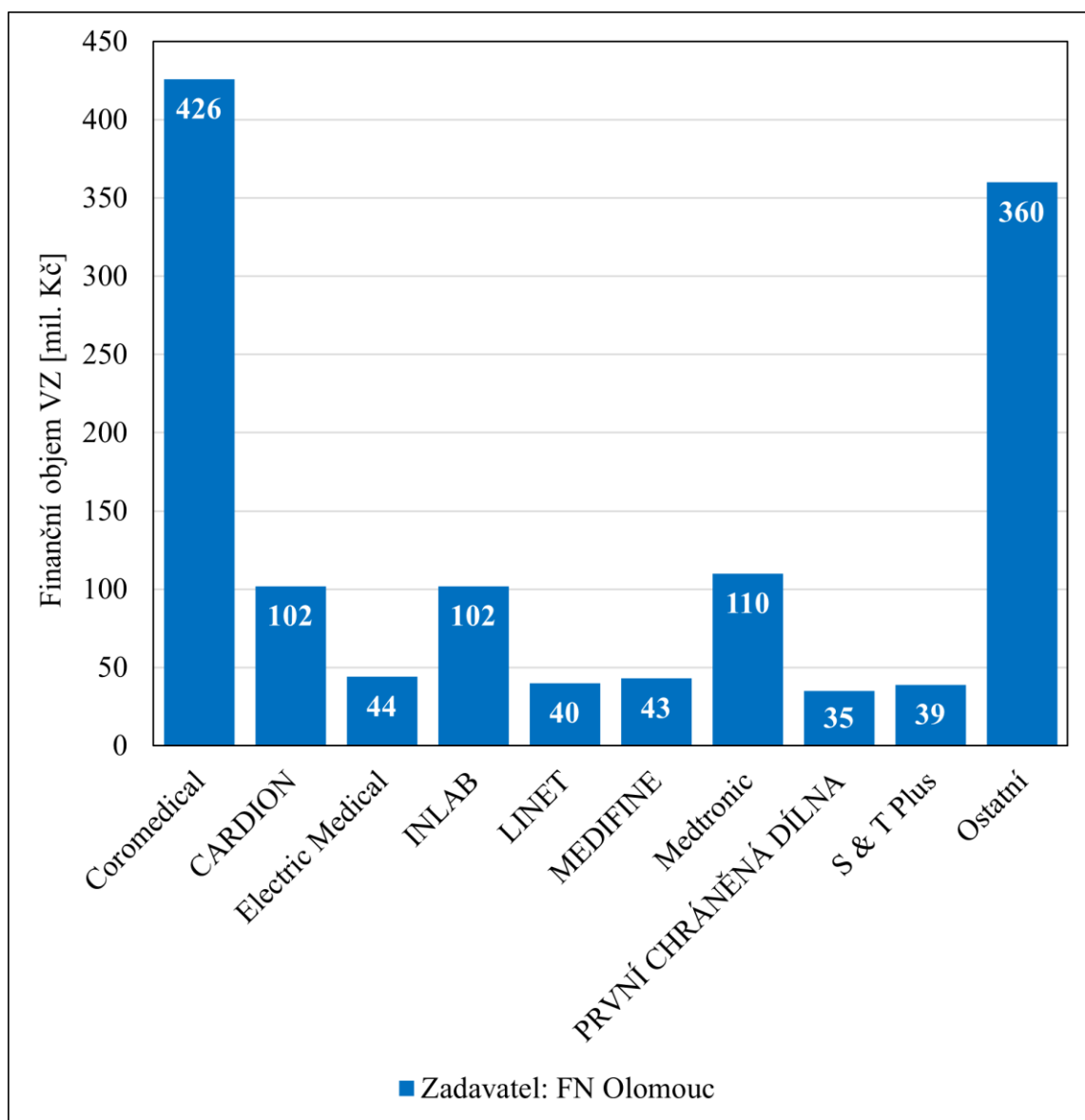
Obrázek 5.1: Rozdělení trhu se zdravotnickou technikou. Je vybráno 9 největších zadavatelů VZ podle finančního objemu zakázek. Pro těchto 9 největších dodavatelů jsou zobrazeny finanční toky vůči 17 největším dodavatelům dle finančního objemu zakázek. V závorkách jsou uvedeny finanční objemy pro jednotlivé subjekty v milionech Kč. Celkový finanční objem zakázek je 5,415 miliard Kč. Plné názvy zadavatelů lze dohledat v tabulce 5.4.

Třemi největšími zadavateli VZ jsou Fakultní nemocnice Olomouc s finančním objemem zakázek 960 milionů Kč, Institutu klinické a experimentální medicíny s finančním objemem zakázek 889 milionů Kč a Krajská zdravotní s finančním objemem zakázek 758 milionů Kč. Celkový finanční objem pro tyto tři dodavatele je 2 607 milionů Kč, což je 48 % z celkového finančního objemu 9 největších zadavatelů dle objemu financí VZ z obrázku 5.1. Dodavatelé s finančním objemem nad 500 milionů korun jsou B. Branu, Cardiomedical, CARDION, MEDIFINE a Siemens. Celkový objem financí zakázek pro těchto 5 dodavatelů je 3 474 milionů Kč, což je 64 % z celkového finančního objemu zakázek vysoutěžených 17 největšími dodavateli.



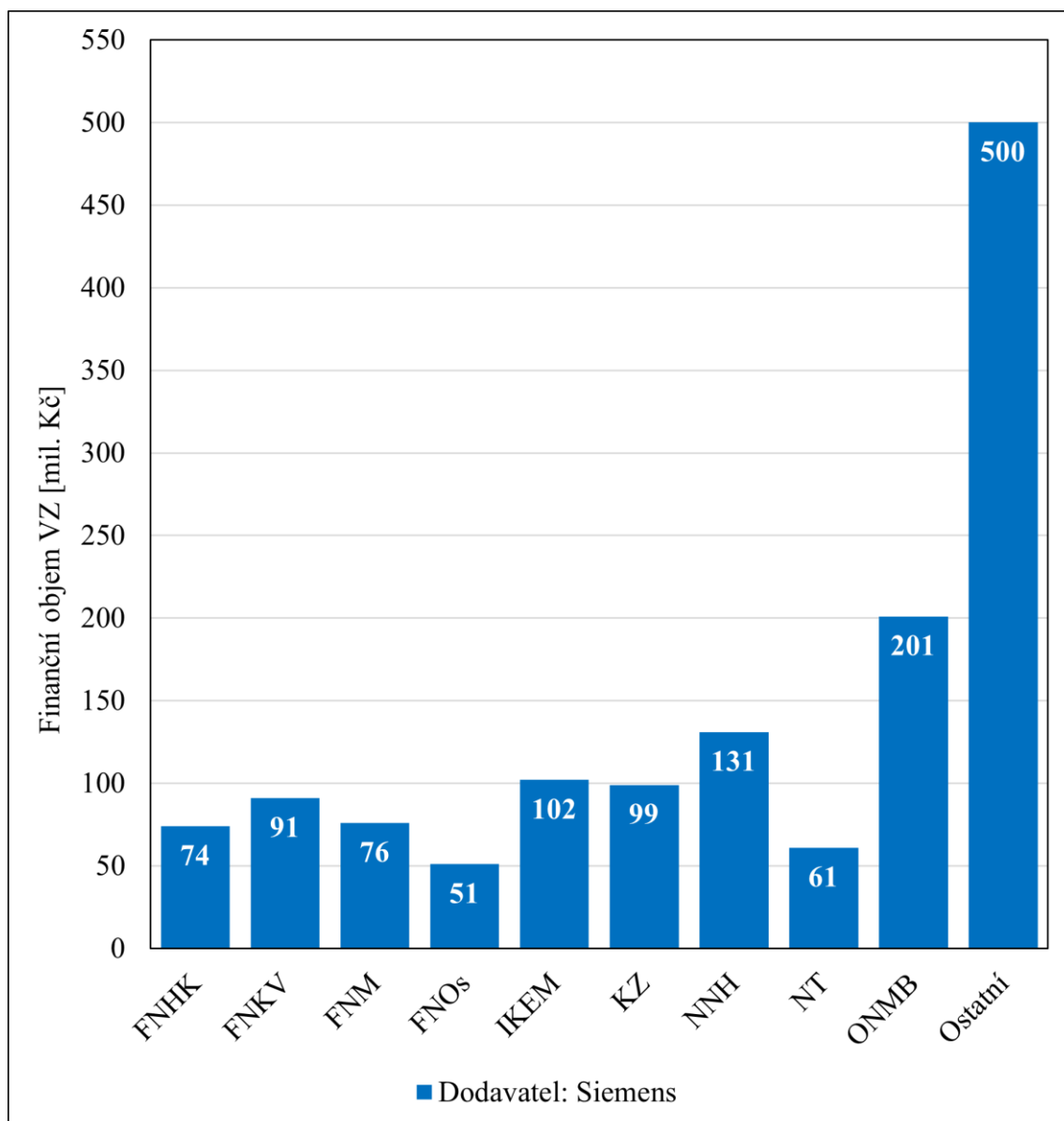
Obrázek 5.2: Rozdělení trhu se zdravotnickou technikou. Je vybráno 10 největších dodavatelů VZ podle finančního objemu zakázek. Pro těchto 10 největších dodavatelů jsou zobrazeny finanční toky vůči 18 největším zadavatelům dle finančního objemu zakázek. V závorkách jsou uvedeny finanční objemy pro jednotlivé subjekty v milionech Kč. Celkový finanční objem zakázek je 5,416 miliard Kč. Plné názvy zadavatelů lze dohledat v tabulce 5.4.

Třemi největšími dodavateli dle finančního objemu zakázek jsou firmy Siemens s finančním objemem zakázek 1 063 milionů, CARDION s finančním objemem zakázek 916 milionů a Medtronic s finančním objemem zakázek 609 milionů. Celkový finanční objem pro tyto tři dodavatele je 2 632 milionů, což je 49 % z celkového finančního objemu 10 největších dodavatelů dle objemu financí VZ z obrázku 5.2. Zadavatelé s finančním objemem zakázek nad 400 milionů jsou FN Hradec Králové, FN Motol, FN Olomouc, Krajská zdravotní a Nemocnice na Homolce. Celkový finanční objem zakázek pro těchto 5 zadavatelů je 2 900 milionů Kč, což je 54 % z celkového finančního objemu zakázek zadaných 18 největšími zadavateli dle objemu financí VZ.



Obrázek 5.3: Zobrazení vztahů mezi největším zadavatelem VZ dle finančního objemu zakázek Fakultní nemocnicí Olomouc a jeho největšími dodavateli dle finančního objemu VZ. Sumy finančních objemů VZ zadaných ostatním dodavatelům, než je 9 největších, jsou označeny jako „Ostatní“. Celkový finanční objem zobrazených VZ je 1 302 milionů Kč.

Fakultní nemocnice Olomouc je největším zadavatelem VZ dle finančního objemu zakázek na zdravotnickou techniku ze zkoumaného vzorku VZ s celkovým finančním objemem zakázek 1 302 milionů Kč. Finanční toky v rámci VZ, mezi tímto zadavatelem a dodavateli, jsou znázorněny na obrázku 5.3. Třemi největšími dodavateli dle finančního objemu zakázek jsou firmy Coromedical s finančním objemem zakázek 426 milionů Kč, Medtronic s finančním objemem zakázek 110 milionů Kč a INLAB s finančním objemem zakázek 102 milionů Kč.



Obrázek 5.4: Zobrazení vztahů mezi největším dodavatelem VZ dle finančního objemu zakázek firmou Siemens a 9 největšími zadavateli dle finančního objemu VZ. Sumy finančních objemů VZ zadaných ostatními zadavateli, než je 9 největších, jsou označeny jako „Ostatní“. Celkový finanční objem zobrazených VZ je 1 388 milionů Kč. Plné názvy zadavatelů lze dohledat v tabulce 5.4.

Siemens je největším dodavatelem VZ dle finančního objemu zakázek na zdravotnickou techniku ze zkoumaného vzorku VZ s celkovým finančním objemem zakázek 1 388 milionů Kč. Finanční toky mezi tímto dodavatelem a zadavateli jsou znázorněny na obrázku 5.4. Tři největší zákazníci dle finančního objemu zakázek jsou Oblastní nemocnice Mladá Boleslav s finančním objemem zakázek 201 milionů Kč, Nemocnice na Homolce s finančním objemem zakázek 131 milionů Kč a Institut klinické a experimentální medicíny s finančním objemem zakázek 102 milionů Kč.

5.2.1 Deskriptivní analýza z databáze RES

Z datového setu FNtb byly vybrány subjekty, jak zadavatelé, tak dodavatelé, které mají 3000 a více zaměstnanců. Jedná se celkem o 22 subjektů. V této skupině se nacházejí pouze zadavatelé VZ. Žádný dodavatel nesplnil toto kritérium. Souhrn těchto subjektů je uveden v tabulce 5.5. Celkem 11 z 22 subjektů jsou fakultní nemocnice, tři subjekty jsou vysoké školy a pět subjektů jsou akciové společnosti, jejichž majoritním vlastníkem je jeden z krajů ČR. Poslední dva subjekty jsou organizační složkou státu. Jedná se o Ministerstvo obrany a Vězeňskou službu České republiky. Původní data o počtech zaměstnanců pocházejí z tabulky REStable.

Tabulka 5.5: Seznam subjektů z datového setu FNtb, které mají 3000 a více zaměstnanců.

IČO	Název	Právní forma	Počet zaměstnanců
27661989	Krajská nemocnice T. Bati	Akciová společnost	3000–3999
00064190	Fakultní Thomayerova nemocnice	Příspěvková organizace	3000–3999
00064173	Fakultní nemocnice Královské Vinohrady	Příspěvková organizace	3000–3999
00159816	Fakultní nemocnice u sv. Anny v Brně	Příspěvková organizace	3000–3999
26068877	Nemocnice České Budějovice	Akciová společnost	3000–3999
61383082	Ústřední vojenská nemocnice	Příspěvková organizace	3000–3999
00064211	Fakultní nemocnice Bulovka	Příspěvková organizace	3000–3999
27283933	Krajská nemocnice Liberec	Akciová společnost	3000–3999
00843989	Fakultní nemocnice Ostrava	Příspěvková organizace	4000–4999
61989592	Univerzita Palackého v Olomouci	Vysoká škola	4000–4999
65269705	Fakultní nemocnice Brno	Příspěvková organizace	5000–9999
60162694	Ministerstvo obrany	Organizační složka státu	5000–9999
00669806	Fakultní nemocnice Plzeň	Příspěvková organizace	5000–9999
00098892	Fakultní nemocnice Olomouc	Příspěvková organizace	5000–9999
27520536	Nemocnice Pardubického kraje	Akciová společnost	5000–9999
00064165	Všeobecná fakultní nemocnice v Praze	Příspěvková organizace	5000–9999
00064203	Fakultní nemocnice v Motole	Příspěvková organizace	5000–9999
00179906	Fakultní nemocnice Hradec Králové	Příspěvková organizace	5000–9999
00216224	Masarykova univerzita	Vysoká škola	5000–9999
00212423	Vězeňská služba České republiky	Organizační složka státu	5000–9999
25488627	Krajská zdravotní, a.s.	Akciová společnost	10 000 a více
00216208	Univerzita Karlova	Vysoká škola (veřejná, státní)	10 000 a více

5.3 Regresní analýza

Regresní analýza bylo provedena z datového setu `data.frame(FNtb)`. Závislou proměnnou byla určena Úspora, viz příloha D, kapitola 3 Úspora. Byly zvoleny 4 kategorické a 4 metrické nezávislé veličiny. Závislé a nezávislé veličiny jsou popsány v tabulce 5.6. Rozsah pro závislou veličinu byl omezen na $\pm 70\%$ (vyřazeno 70 zadání). Počet nabídek byl omezen na menší než 9 (vyřazeny 3 záznamy) a počet kritérií na menší než 6 (vyřazeny 2 záznamy). Datový set po stanovení rozsahů obsahoval 2635 zadání.

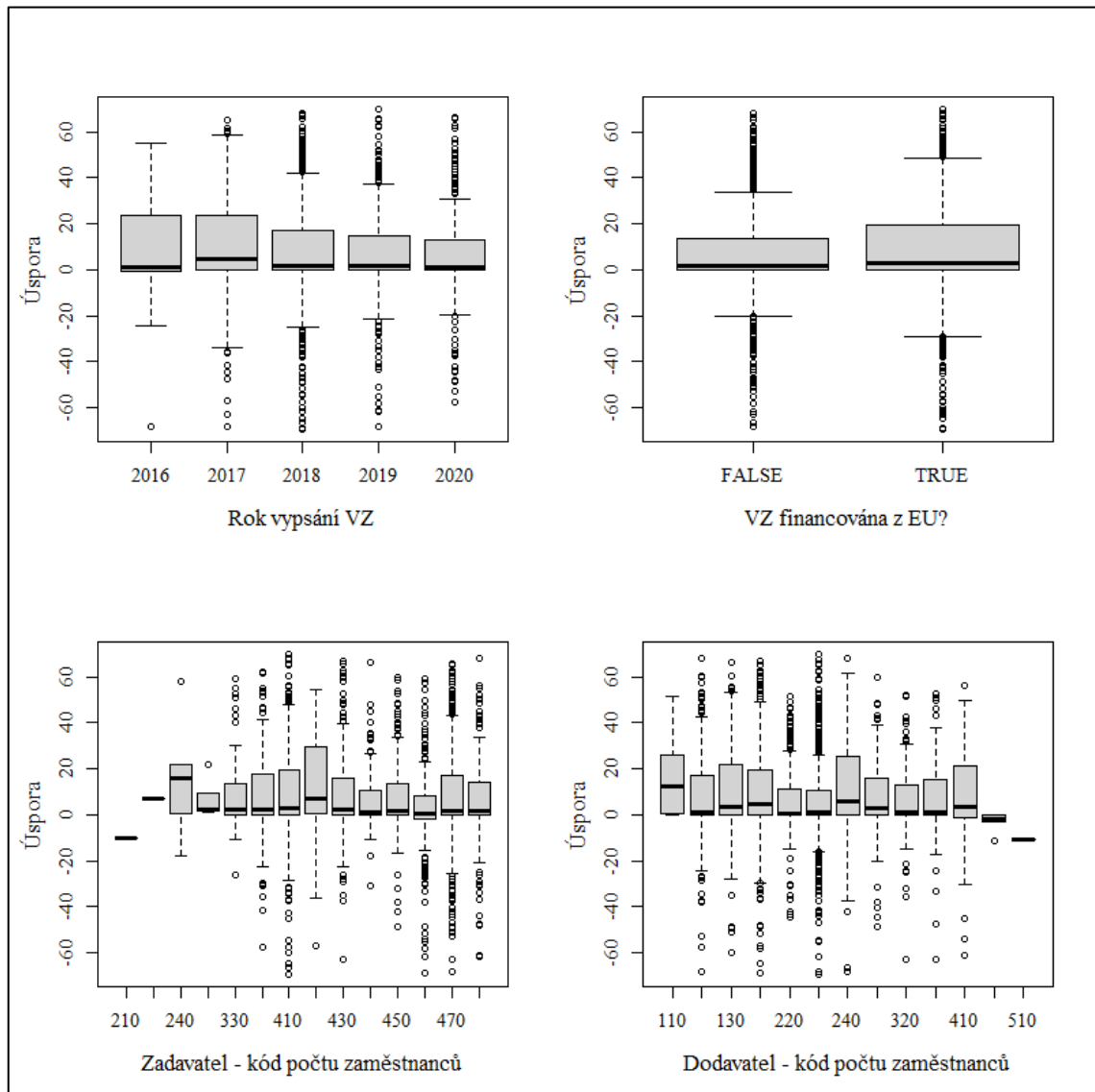
Tabulka 5.6: Popis závislých a nezávislých veličin vstupujících do regresní analýzy

Závislá veličina	
Název	Rozsah
Úspora	$\pm 70\%$
Nezávislé veličiny	
Kategorické	
Název	Proměnné
Rok vypsání VZ	2016–2020
VZ financována z EU?	TRUE/FLASE
Počet zaměstnanců – zadavatel	kód viz příloha D
Počet zaměstnanců – dodavatel	kód viz příloha D
Metrické	
Název	Rozsah
Počet nabídek	1–8
Počet kritérií	1–6
Lhůta pro podání nabídek	16–113 dnů
Doba trvání zadání VZ	22–497 dnů

Byl vytvořen vícenásobný lineární regresní model, ze kterého byly následně převzaty statisticky významné nezávislé proměnné, které byly použity pro tvorbu druhého vícenásobného lineárního regresního modelu. Druhý model byl následně ověřen

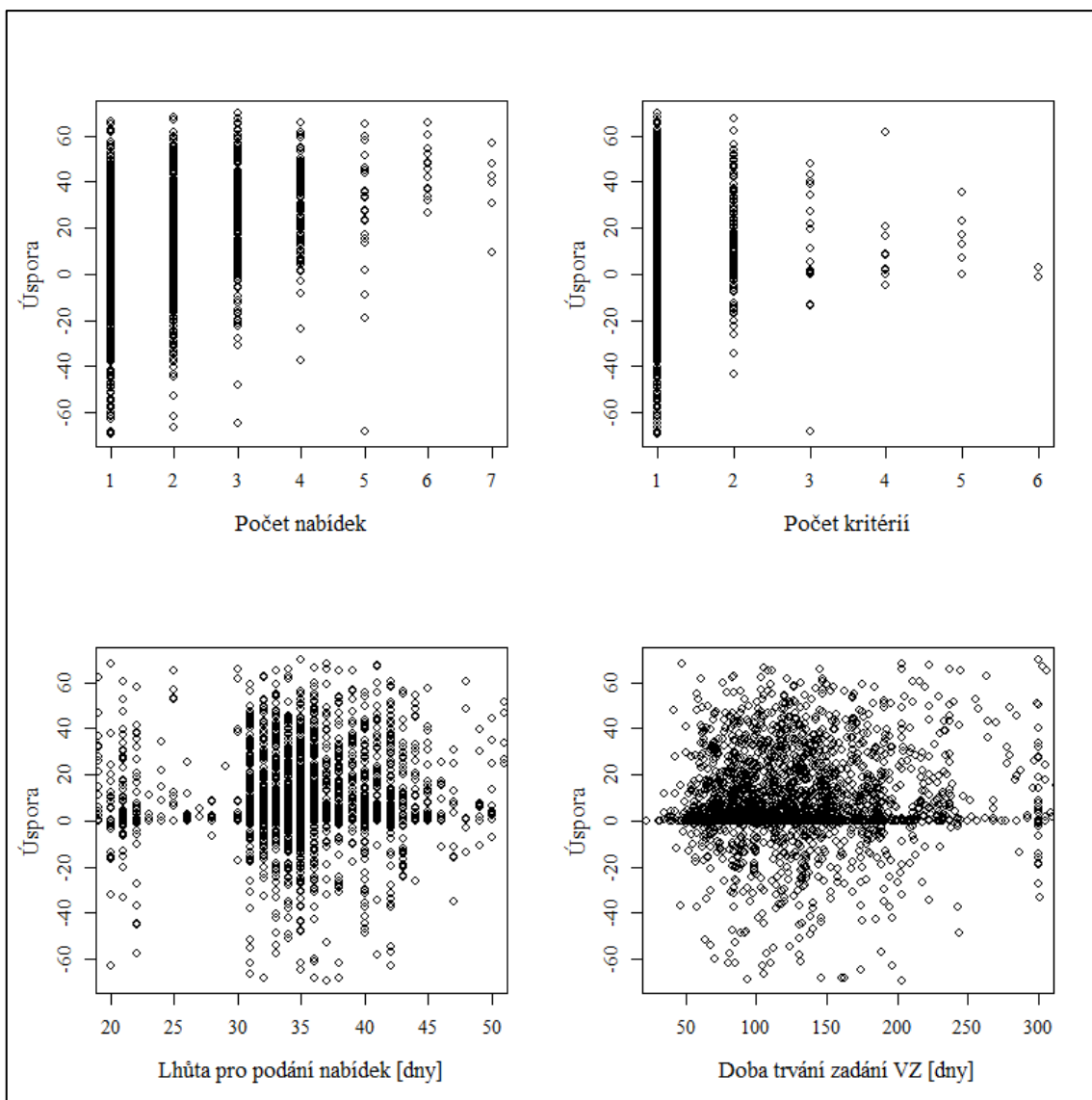
5.3.1 Vícenásobný lineární regresní model

Před provedením regresní analýzy byly nezávislé veličiny zobrazeny do grafů vůči závislé veličině. Kategorické veličiny byly zobrazeny pomocí box grafů, viz obrázek 5.5 a metrické pomocí klasických grafů, viz obrázek 5.6.



Obrázek 5.5: Kategorické nezávislé proměnné vůči závislé proměnné úspoře. Jedná se o první pohled na data, který slouží ke kontrole extrémně odlehlých hodnot a odhadu regrese.

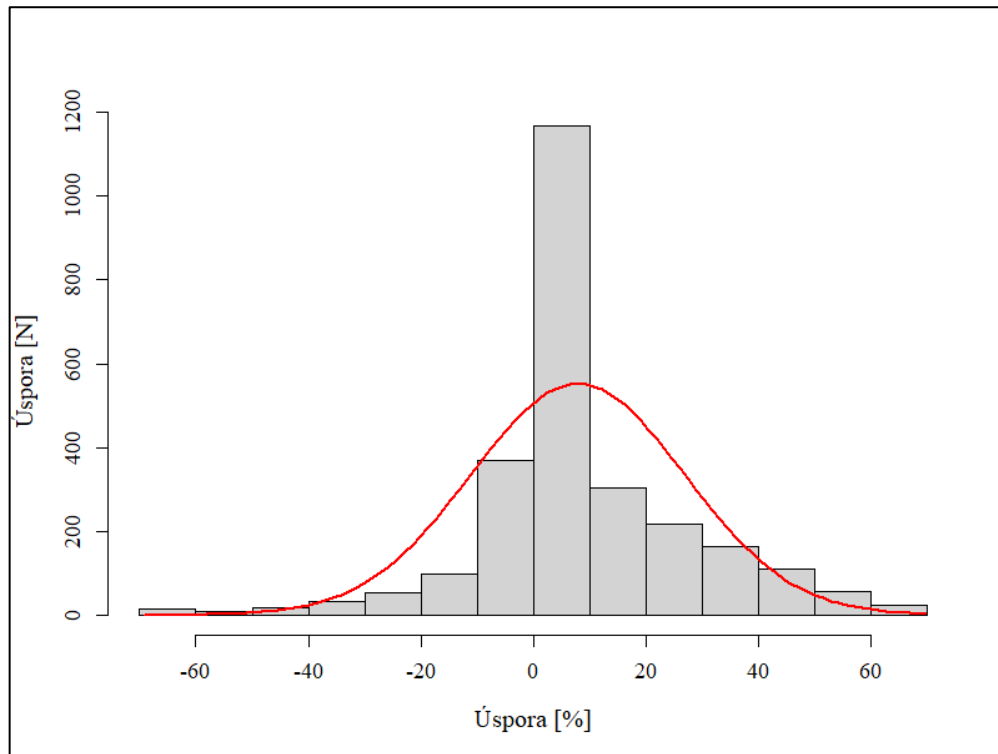
Bylo zkontrolováno, že neexistují extrémně odlehlé hodnoty, které mohli vzniknout chybou zadání parametrů ze strany zadavatelů VZ a bylo přikročeno k provedení regresní analýzy.



Obrázek 5.6: Metrické nezávislé proměnné vůči závislé proměnné úspoře. Jedná se o první pohled na data, který slouží ke kontrole extrémně odlehlých hodnot a odhadu regrese.

Bylo zkontrolováno, že neexistují extrémně odlehlé hodnoty, které mohli vzniknout chybou zadání parametrů ze strany zadavatelů VZ a bylo přikročeno k provedení regresní analýzy.

Závislá proměnná úspora byla zobrazena do histogramu, viz obrázek 5.7, pro kontrolu splnění podmínky normality a byl provede Shapiro-Wilkův test normality, kdy $W=0,905$ ($p\text{-value}<2,2\cdot 10^{-16}$). Závislá proměnná tedy pochází z normálního rozdělení.



Obrázek 5.7: Histogram závislé proměnné Úspora. Červeně proložení křivkou normálního rozložení. Shapiro-Wilkův test normality $W=0,905$ ($p\text{-value}<2,2\cdot 10^{-16}$)

Byl vytvořen první vícenásobný lineární model pro ověření regrese. Byly použity všechny nezávislé veličiny z tabulky 5.6. Kategorické veličiny byly zpracovány metodou „dummy“ proměnných. Kompletní výsledky jsou k dispozici v tabulce 5.7.

Tabulka 5.7: Výsledky regresní analýzy z prvního vícenásobného lineárního modelu. Zahrnuje všechny nezávislé proměnné z tabulky 5.6. Statisticky významné nezávislé proměnné jsou označeny zeleně.

Nezávisle proměnná	Koeficient	Směrodatná odchylka	p-hodnota
Absolutní člen	-19,306	18,983	0,309
Rok_Vypsani2017	2,729	4,158	0,512
Rok_Vypsani2018	1,669	4,109	0,685
Rok_Vypsani2019	0,593	4,135	0,886
Rok_Vypsani2020	0,970	4,178	0,816
Velikost_Kod_Zadavatel310	15,422	19,836	0,437
Velikost_Kod_Zadavatel330	23,434	17,844	0,189
Velikost_Kod_Zadavatel340	13,593	17,745	0,444

Velikost_Kod_Zadavatel410	16,733	17,705	0,345
Velikost_Kod_Zadavatel420	20,791	17,853	0,244
Velikost_Kod_Zadavatel430	18,449	17,716	0,298
Velikost_Kod_Zadavatel440	17,277	17,787	0,331
Velikost_Kod_Zadavatel450	17,342	17,712	0,328
Velikost_Kod_Zadavatel460	11,115	17,721	0,531
Velikost_Kod_Zadavatel470	15,936	17,686	0,368
Velikost_Kod_Zadavatel510	14,093	17,726	0,427
Velikost_Kod_Dodavatel120	-5,814	5,260	0,269
Velikost_Kod_Dodavatel130	-4,386	5,280	0,406
Velikost_Kod_Dodavatel210	-2,347	5,190	0,651
Velikost_Kod_Dodavatel220	-5,991	5,269	0,256
Velikost_Kod_Dodavatel230	-7,308	5,165	0,157
Velikost_Kod_Dodavatel240	-1,252	5,235	0,811
Velikost_Kod_Dodavatel310	-4,309	5,295	0,416
Velikost_Kod_Dodavatel320	-8,152	5,441	0,134
Velikost_Kod_Dodavatel330	-1,752	5,378	0,745
Velikost_Kod_Dodavatel410	-5,827	5,614	0,299
Velikost_Kod_Dodavatel440	-11,113	9,432	0,239
Velikost_Kod_Dodavatel510	-14,992	18,416	0,416
Fin_EUTRUE	1,321	0,908	0,146
Pocet_Nabidek	7,960	0,416	$2,41 \cdot 10^{-75}$
Pocet_Kriterii	2,022	0,850	0,017
Nabidky_Lhuta	-0,024	0,066	0,715
Doba_Zakazka	-0,009	0,007	0,170

Při zvolení hladiny významnosti $\alpha = 0,05$ se jako statisticky významné ukazují dvě veličiny, a to počet nabídek ($p=2,41 \cdot 10^{-75}$) a počet kritérií ($p=0,017$). Index determinance R^2 je pro tento model 0,189 a adjustovaný index determinance R^2 je 0,177. P-hodnota pro F-test je $2,2 \cdot 10^{-16}$ tedy zamítáme nulovou hypotézu o nulových regresních koeficientech. Souhrn testů je zobrazen v tabulce 5.8.

Tabulka 5.8: Souhrn testů prvního vícenásobného lineárního modelu regrese.

Charakteristika modelu	Hodnota
Index determinance R^2	0,189
Adjustovaný index determinance R^2	0,177
F-test	15,32
F-test: p-hodnota	$2,2 \cdot 10^{-16}$

Druhý vícenásobný lineární model pro ověření regrese byl vytvořen pouze ze statisticky významných proměnných z prvního modelu, tedy jako závislá veličina opět vstupuje úspora a nezávislé jsou počet nabídek a počet kritérií. Z tohoto modelu vypadly všechny kategorické proměnné a model pracuje pouze s metrickými proměnnými. Kompletní výsledky tohoto modelu jsou k dispozici v tabulce 5.9.

Tabulka 5.9: Výsledky regresní analýzy z druhého vícenásobného lineárního modelu. Zahrnuje pouze statisticky významné nezávislé proměnné z prvního modelu. Statisticky významné nezávislé proměnné jsou označeny zeleně.

Nezávisle proměnná	Koeficient	Směrodatná odchylka	p-hodnota
Absolutní člen	-7,271	1,133	$1,65 \cdot 10^{-10}$
Pocet_Nabidek	7,984	0,379	$3,65 \cdot 10^{-90}$
Pocet_Kriterii	1,594	0,758	0,036

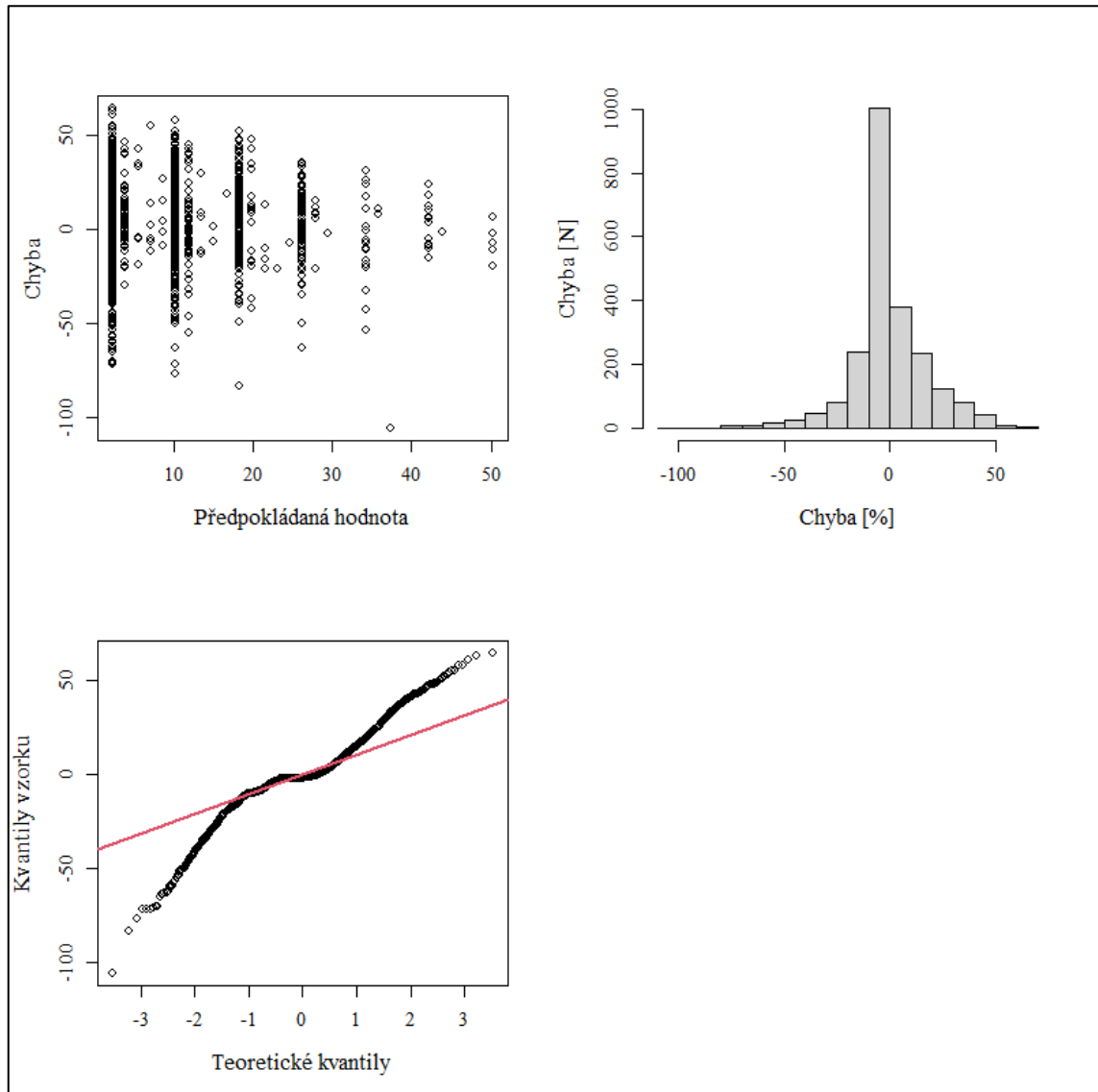
Při zvolení hladiny významnosti $\alpha = 0,05$ se jako statisticky významné ukazují dvě veličiny, a to počet nabídek ($p=3,65 \cdot 10^{-90}$) a počet kritérií ($p=0,036$). Index determinance R^2 je pro tento model 0,163 a adjustovaný index determinance R^2 je 0,163. P-hodnota pro F-test je $2,2 \cdot 10^{-16}$ tedy zamítáme nulovou hypotézu o nulových regresních koeficientech. Souhrn testů je zobrazen v tabulce 5.10.

Tabulka 5.10: Souhrn testů druhého vícenásobného lineárního modelu regrese.

Charakteristika modelu	Hodnota
Index determinance R^2	0,163
Adjustovaný index determinance R^2	0,163
F-test	223,5
F-test: p-hodnota	$2,2 \cdot 10^{-16}$

5.3.2 Ověření modelu

Ověření druhého modelu vícenásobné lineární regrese proběhlo pomocí znázornění vztahů chyb modelu vůči předpokládaným hodnotám, viz obrázek 5.8. Zobrazen je vztah mezi Chybou a předpokládanou hodnotou, histogram chyby a Q-Q graf. Grafy odpovídají předpokladům, které můžeme tvořit z hodnot indexu determinance R^2 a adjustovaného indexu determinance R^2 .



Obrázek 5.8: Grafy ověření kvality druhého modelu vícenásobné lineární regrese. Zaměřují se na rovnoměrnost rozložení chyby ve vytvořeném modelu. Chyba by měla podléhat normálnímu a rovnoměrnému rozložení v rámci celého modelu.

5.4 Výstupy algoritmu pro kategorizaci zdravotnické techniky

Bylo provedeno testování algoritmu pro automatickou kategorizaci předmětu VZ dle KZT. Pro test bylo vybráno 5 kategorií KZT viz tabulka 4.4. Postup je popsán v kapitole 4.4. U jednotlivých průběhů byly měněny vybrané parametry pro otestování jejich vlivu na správnost určení KZT. Měněnými parametry byly:

- Způsob tvorby knihovny
- Zvolená vzdálenost pro shodu znaků v řetězci
- Minimální hodnota p indexu pro prohlášení shody slov
- Konstanta síly efektu prefixu

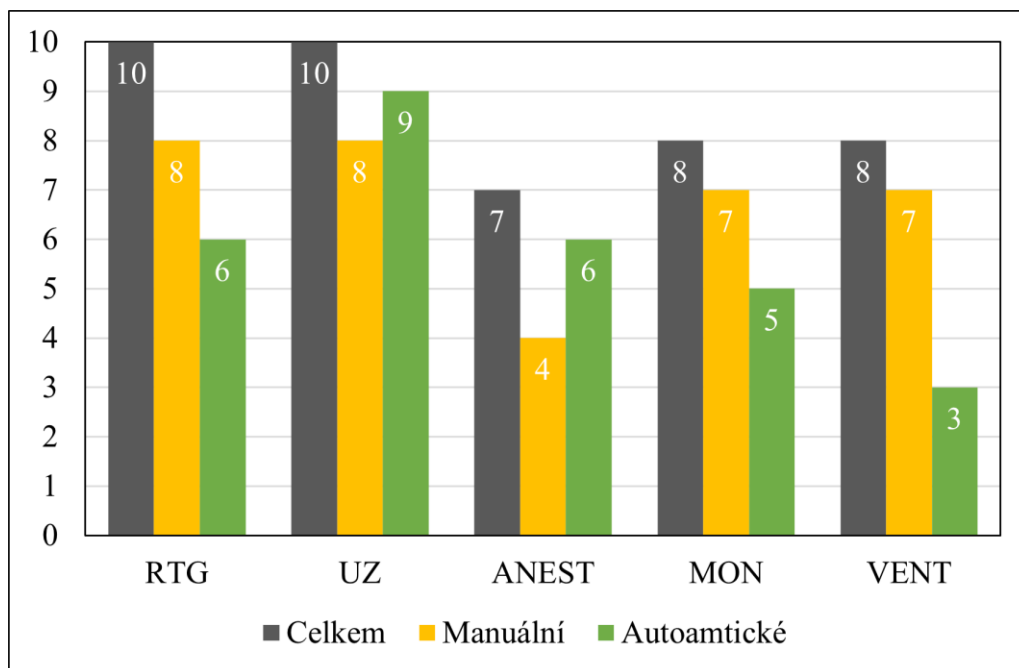
Ve všech případech se za výchozí nastavení algoritmu považuje použití:

- Způsob tvorby knihovny KZT = manuální
- Vzdálenost pro shodu znaků v řetězci = 5
- P index pro prohlášení shody slov = 0,8
- Konstanta síly efektu prefixu = 0,1

Všechny iterace jsou tedy vztahovány k tomuto výchozímu nastavení a je měněn vždy pouze jeden příslušný parametr. Souhrn úspěšnosti algoritmu při iteraci jednotlivých parametrů je k dispozici v tabulce 5.11.

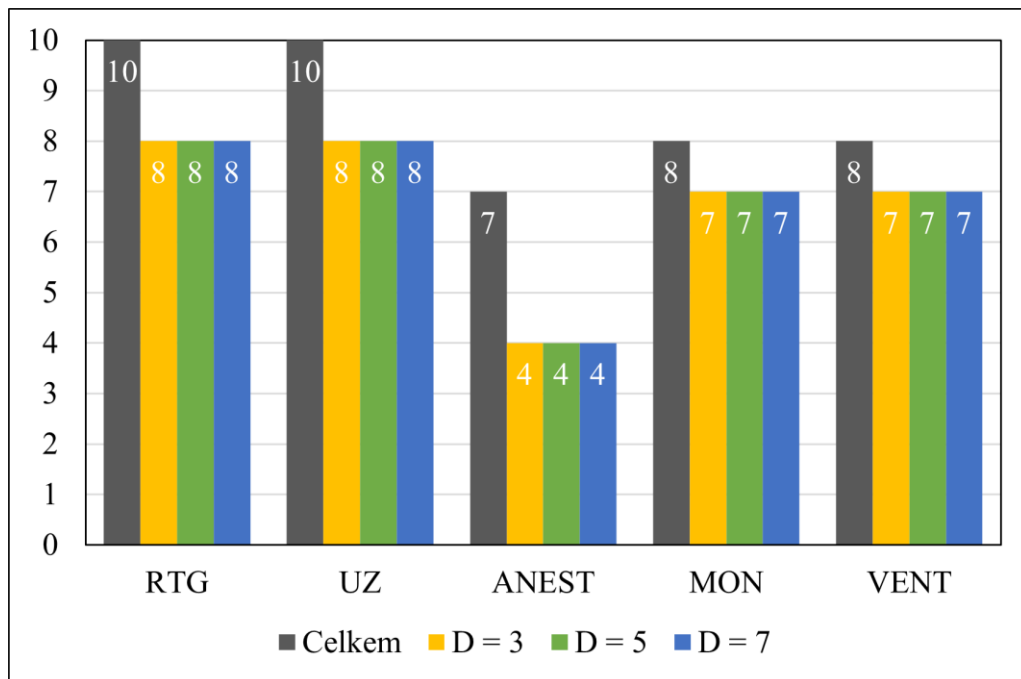
Tabulka 5.11: Počet správně určených kategorií VZ dle KZT v jednotlivých iteracích. Zeleně jsou vyznačeny parametry s největší úspěšností.

Způsob tvorby knihovny	Počet	Úspěšnost
Manuální	34	79 %
Automatický	29	67 %
Zvolená vzdálenost pro shodu znaků v řetězci		
3 znaky	34	79 %
5 znaků	34	79 %
7 znaků	34	79 %
Minimální hodnota p indexu pro prohlášení shody slov		
$p_i = 0,7$	33	77 %
$p_i = 0,8$	34	79 %
$p_i = 0,9$	34	79 %
Konstanta síly efektu prefixu		
$p = 0,25$	30	70 %
$p = 0,1$	34	79 %



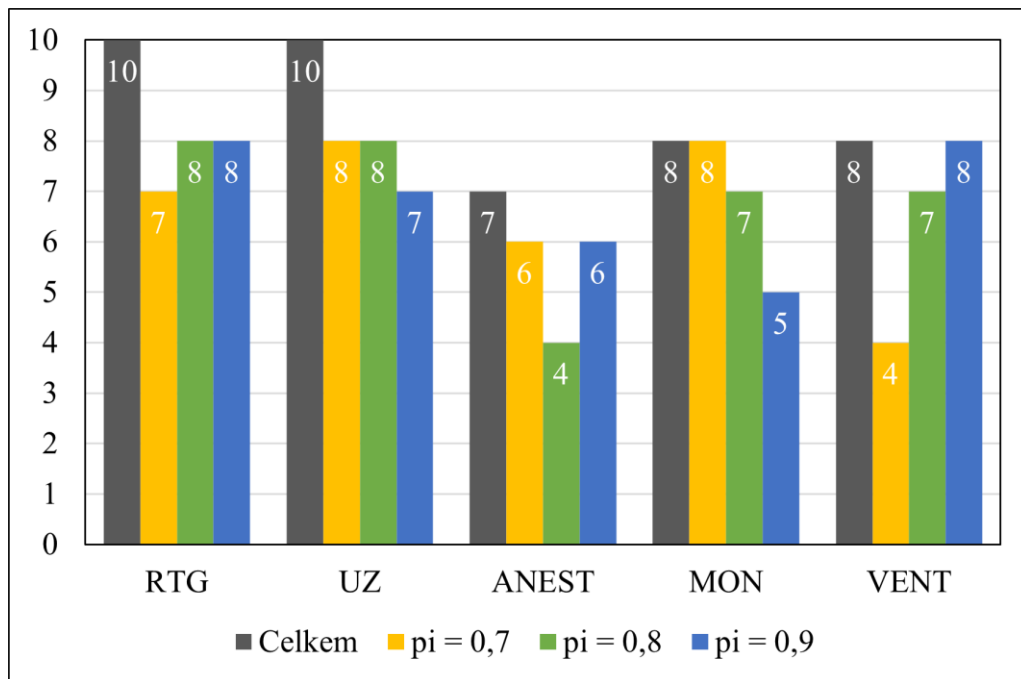
Obrázek 5.9: Správně přiřazené kategorie KZT k VZ při zvolení manuální nebo automatické knihovny výrazů KZT. Položka „Celkem“ určuje celkový počet zakázek v dané kategorii, které byly podrobeny přiřazení pomocí algoritmu.

První iterací v průběhu algoritmů byla úprava knihoven výrazů KZT. V prvním průběhu byl zvolen automatický způsob vytvoření knihovny z číselníku KZT, kde byly vzaty výrazy z názvu kategorie, její synonyma, a to až do třetí hierarchie. Z těchto výrazů byla vytvořena knihovna pro fuzzy-matching. V druhém průběhu byla tato automaticky vytvořená knihovna ručně upravena o další vhodné výrazy a nevhodné a generické výrazy byly odstraněny. Následně byly vyhodnoceny správně přiřazené kategorie, viz obrázek 5.9. S automaticky vytvořenou knihovnou vyhodnotil algoritmus správně 67 % zakázek. S upravenou knihovnou vyhodnotil algoritmus správně 79 % zakázek.



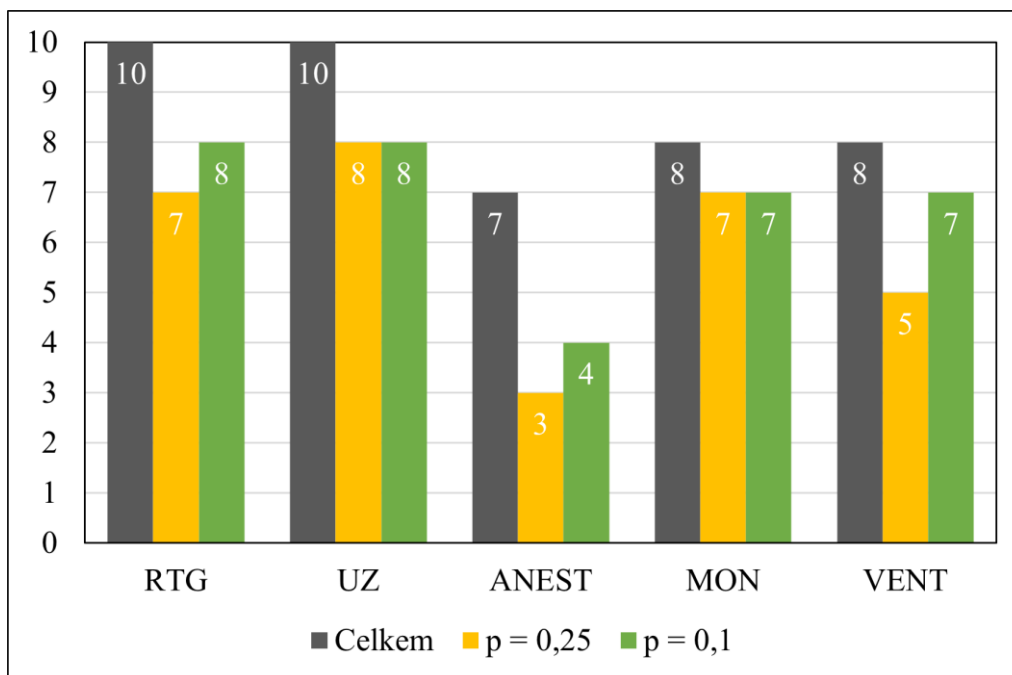
Obrázek 5.10: Správně přiřazené kategorie KZT k VZ při změně parametru vzdálenosti pro shodu znaků v řetězci. Položka „Celkem“ určuje celkový počet zakázek v dané kategorii, které byly podrobeny přiřazení pomocí algoritmu.

Druhou iterací v průběhu algoritmu byla změna vzdálenosti pro shodu znaků v řetězci při fuzzy-matching. Všechny iterace vykazují stejný výsledek úspěšného určení kategorie 79 %, viz obrázek 5.10. Změna tohoto parametru neměla pro tento výběr zakázek žádný dopad na úspěšnost algoritmu.



Obrázek 5.11: Správně přiřazené kategorie KZT k VZ při změně parametru minimální hodnoty p-indexu pro prohlášení shody slov. Položka „Celkem“ určuje celkový počet zakázek v dané kategorii, které byly podrobeny přiřazení pomocí algoritmu.

Třetí iterací v průběhu algoritmu byla změna parametru minimální hodnoty p-indexu pro prohlášení shody slov. P-index slouží k stanovení hranice, kdy můžeme prohlásit dvě slova za shodná. Byly vyhodnoceny správně přiřazené kategorie dle KZT, viz obrázek 5.11. Při nastavení parametru indexu na 0,7 byla úspěšnost algoritmu 77 %. Pro index 0,8 byla úspěšnost 79 % a pro index 0,9 také 79 %. Největší rozdíl v úspěšnosti algoritmu při použití indexu 0,7 nebo 0,9 vidíme v kategorii VENT, kdy s indexem 0,7 má algoritmus úspěšnost pouze 50 %, zatímco s indexem 0,9 je úspěšnost 100 %.



Obrázek 5.12: Správně přiřazené kategorie KZT k VZ při změně parametru konstanta síly efektu prefixu. Položka „Celkem“ určuje celkový počet zakázek v dané kategorii, které byly podrobeny přiřazení pomocí algoritmu

Čtvrtou iterací v průběhu algoritmu byla změna parametru konstanta síly efektu prefixu. Ten upravuje pravděpodobnost shody slov na základě shodných znaků na začátku řetězce. Byly vyhodnoceny správně přiřazené kategorie dle KZT, viz obrázek 5.12. Pokud byla konstanta rovna 0,1 byla úspěšnost algoritmu vyhodnotit správnou kategorii KZT 79 %. Při použití konstanty 0,25 byla úspěšnost algoritmu 70 %.

6 Diskuse

Byl vypracován algoritmus pro získání agregovaného datového setu, obohaceného z databází VVZ a RES, který je vhodný pro provedení deskriptivní a regresní analýzy. Tento set obsahuje celkem 2635 zadání. K tomuto setu byl vypracován codebook. Byla provedena deskriptivní analýza pomocí alluvial grafů. Dále byla provedena regresní analýza mezi závislou veličinou úspora a dalšími nezávislými veličinami. Statisticky signifikantní regresi ($\alpha = 0,05$) vykazují dvě nezávislé veličiny, a to počet nabídek ($p=3,65 \cdot 10^{-90}$) a počet kritérií ($p=0,036$). Index determinance R^2 pro vytvořený model je 0,163 a adjustovaný index determinance R^2 je 0,163. P-hodnota pro F-test je $2,2 \cdot 10^{-16}$. Byl vypracován algoritmus pro automatickou kategorizaci předmětu VZ dle KZT. Nejvyšší dosažená úspěšnost algoritmu je 79 %.

Datové sady

Otevřené datové sady z ISVZ

Primárním zdrojem dat byly otevřené datové sady z ISVZ. Tyto datové sady jsou vydávány od roku 2016, po úpravě legislativy, do současnosti. V roce 2021 ale došlo ke změně v poskytování systému těchto datových sad z ISVZ na RAV systém. Byla renovována stránka ISVZ, data byla přesunuta a po dobu zhruba jednoho roku nebyla vůbec vydávána. To se změnilo v roce 2022, kdy byla tato data opět zpřístupněna na novém webu ISVZ. Problémem je ale nekontinuita ve formátování. Ačkoliv se jedná o menší úpravy, jako je například, vynechání pomlčky v názvu sloupce, pro účel automatizace toto představuje zásadní problém. Vytvořený skript je sice připraven stahovat data k aktuálnímu datu, není ale připraven pro jejich zpracování. Bylo tedy možné použít pouze data mezi roky 2016–2020, která mají jednotný formát. Sjednocení formátu datových setů a následné rozšíření na další roky by mohlo být námětem budoucí práce. [5]

Samotné zpracování dat proběhlo bez problému. Je ovšem vidět, že strukturu otevřených dat po delší dobu nikdo nerevidoval. V datovém setu se nacházejí proměnné, které se nepropisují z VVZ, například zda je zakázka financována z EU, a dále proměnné, které nabývají různých hodnot pro různé typy formulářů. Ačkoliv existuje dokumentace k této datové sadě, je popsán velice střídme a neobsahuje všechny informace. Příkladem může být chybějící informace o propojení jednotlivých vydávaných tabulek v rámci datového setu. Tato informace je přitom naprosto zásadní pro práci s celým datovým setem. Tato dokumentace je v podstatě jen výpisem jednotlivých kategorií v daných tabulkách. V dokumentu také chybí číselníky. Je tedy zřejmé, že tento dokument nelze použít pro plnou orientaci v datových sadách a uživatel, který tyto data zpracovává musí potřebné vztahy a informace dohledat, což je vzhledem k myšlence otevřených dat,

naprosto nevhodné. Poslední výtkou, která již jen dokresluje přístup MMR k otevřeným datovým sadám, je ten, že při navigování na stránkách ISVZ do kategorie „OpenData“ se uživatel dostane na stránku, která poskytuje opět jiný formát výstupu dat z VVZ. Výstupy v těchto sadách jsou extrémně ořezané výstupy z VVZ. Prakticky se jedná o informaci, kdo zadal zakázku, komu a za kolik. Je zde pokus o jakési třídění, ale uživatelé jsou v podstatě předloženy hotové přefiltrované informace tak, jak uznal provozovatel ISVZ za vhodné. K těmto datovým sadám se snaží provozovatel vydávat metodiku tvorby statistik, jejíž informační hodnota je ovšem nulová. Na data, která jsou ve formátu původně vydávaných datových sad, se uživatel dostane až proklikem přes malý prolink v záhlaví stránky. Opět, všechny tyto skutečnosti jsou naprosto v rozporu s principy otevřených dat a ztěžují přístup uživatelů k těmto datovým sadám. Na závěr je nutné podotknout, že nový vydavatel Datera s. r. o. se snaží o nápravu některých výše zmíněných výtek, ale výsledky jsou ke květnu 2022 přinejmenším nepřesvědčivé.

Filtrace VZ na zdravotnickou techniku

Pro filtrování předmětu VZ byly použity CPV kódy a IČO zadavatelů. V případě CPV kódu byly ručně vybrány všechny kódy, které se týkají zdravotnické techniky a pokud tento kód zadání zakázky neobsahovalo, byly tyto zadání vyřazeny. To se ukázalo jako vhodný přístup, ale je zde několik limitací, které je nutné zmínit. První je, že CPV kódy zadává zadavatel sám. Je tedy nutné počítat s nejednotným formátováním. Algoritmus si s tímto problémem umí do značné míry poradit, ale pokud se uživatel přepíše, není možné tuto chybu opravit. Stejně tak může uživatel špatně odhadnout předmět zakázky a přiřadit nesprávný CPV kód. V rámci filtru jsou také zahrnuty všechny zakázky s nespécifickými kódy na úrovni oddílu, jako 33000000 (Zdravotnické přístroje, farmaceutika a prostředky pro osobní péči), nebo skupin, 33100000 (Zdravotnické přístroje). Zadavatelé totiž mnohdy specifikují pouze generickou skupinu bez bližšího upřesnění. Například pouze výše zmíněné obecné CPV kódy udává 89 z 2710 zadání, což je 2,27 %. Toto číslo se zdá poměrně malé, ale když si uvědomíme, že se jedná skutečně o tu nejhorší variantu, tak hrozba, že se tyto opomenutí promítají i do nižších hladin číselníku, je poměrně velká. Algoritmus také počítá s tím, že všechny zdravotnická technika spadá pod oddíl 33000000, „Zdravotnické přístroje, farmaceutika a prostředky pro osobní péči“. Je ovšem pravdou, že například laboratorní přístroje mohou být zařazeny jak do tohoto oddílu, tak do oddílu 38000000, „Laboratorní, optické a přesné přístroje a zařízení (mimo skel)“. Který oddíl použít pro filtraci není zřejmé, jelikož použití vychází z praxe. CPV kódy ale nadále zůstávají silným nástrojem pro filtraci předmětu veřejné zakázky.

Filtrací z databáze NRPZS a PZS byla snaha odstranit zadavatele, kteří sice nakupují zdravotnickou techniku, ale nevyužívají ji k poskytování zdravotní péče. Jako typické příklady byly uvedeny vysoké školy a veterinární kliniky. Pro filtraci subjektů veterinárních klinik tento filtr fungoval, ale problém nastal u vysokých škol. Jelikož

vzdělávání lékařů v ČR zajišťují vysoké školy, je běžné, že jsou také tyto subjekty poskytovatelem zdravotní péče. Není tedy pravdou, že by veškerou péči směřovali ke svým fakultním nemocnicím. Příkladem může být Karlova univerzita. Není tedy možné vysoké školy ze seznamu poskytovatelů vyloučit. Nedořešeným problémem je také časová posloupnost těchto datových setů. Pro filtraci jsou použity aktuální sety z roku 2022, ale zakázky mohou být až z roku 2016. Pokud tedy existuje zadavatel, který byl v roce 2016 poskytovatelem zdravotních služeb a zadal zakázku na zdravotnickou techniku, ale v roce 2022 jím již není, tak tuto zakázku algoritmus chybně vyloučí. Tento nedostatek by mohl být námětem pro budoucí práci.

Obohacení datových sad

Pro obohacení datových sad byly zvoleny dva zdroje. Prvním zdrojem byly informace z databáze RES. Algoritmus byl úspěšně implementován a tato přidaná data byla zařazena do regresní analýzy. Druhým zdrojem byli stažené formuláře zakázek z VVZ. Snahou bylo získat informace, které se nepropisují do otevřených datových sad. I tato implementace byla úspěšná. Oba zdroje byly zpracovány pomocí metody webscrapping, pomocí adresování html elementů. Ačkoliv tato metoda potřebuje určité ošetření pro chybějící data, nebo nedostatky v databázi, ukázala se být velmi silným nástrojem pro získávání dat.

Datová sada pro potřeby analýzy

Tabulka použitá pro potřeby analýz byla vytvořena spojením tabulek stažených z ISVZ a obohacením z dalších zdrojů. Jak už ale bylo zmíněno, nedostatek dokumentace a nejednotná metodika v uveřejňování těchto dat vedla k značným limitacím výsledné tabulky. Bylo nutné vyřadit všechny národní formuláře, tedy prakticky vyřadit všechny podlimitní zakázky. Pokud budeme brát všechny zadané zakázky za rok 2019, tak to znamená vyřadit zakázky za 65 miliard korun, což je 18,9 % z celkového finančního objemu. Celkem se jednalo o 6453 zakázek, což je 47,9 % z celkového počtu zakázek zadaných v roce 2019 [58]. Důvodem byla nemožnost navázat tabulku Část VZ na VZ, protože neexistuje unikátní identifikátor pro napojení a bez tohoto propojení nelze spočítat úsporu. Aby byla zakázka podlimitní zakázkou, musí splňovat finanční limit do částky 3 653 000 Kč [59]. V rámci přístrojové techniky je to částka, která je rozhodně relevantní a opomenutím těchto zakázek vnášíme do analýzy chybu. Na problém propojení tabulek VZ a části VZ narážíme i u evropských formulářů. V datovém setu z ISVZ neexistuje unikátní identifikátor, kterým by toto propojení bylo možné provést. Tyto dvě tabulky tedy byly propojeny pomocí textového popisu části VZ anebo čísla části VZ. Ideální by bylo použít číslo části VZ, ale jedná se o položku vyplňovanou zadavatelem a bylo zjištěno, že není vyplňována konzistentně a obsahuje mnoho chyb. Proto jde až o druhou možnost. Z důvodu tohoto problému bylo nutné vyřadit 114 zakázek, které nebylo možné propojit ani jednou metodou.

OpenData vs. webscrapping

Zásadní otázkou zůstává vhodnost zvolené metodiky k získání dat o VZ. Tato práce pracuje s otevřenými daty z ISVZ, zároveň ale bylo demonstrováno, jak účinná je metoda webscrapping pro získání cílových dat. Kombinací těchto metod by bylo možné získat datové sady, které by byly ještě spolehlivější a rozsáhlejší než při použití těchto metod samostatně. Z otevřených datových sad je možné získat čísla zakázek na VVZ a čísla formulářů na VVZ a dále stáhnout tyto formuláře z VVZ ve formátu html. To bylo provedeno i v rámci této práce. Další data z otevřených datových sad jsou ale nepřehledná, jak bylo uvedeno v kapitole 6.1.1. Řešením by mohla být těžba dat přímo z formulářů. Tento přístup má jednu zásadní výhodu a to je, že je přesně známý formát i tvorba dat, protože podoba formulářů je obsahem Vyhlášky č. 168/2016 Sb. o uveřejňování formulářů pro účely zákona o zadávání veřejných zakázek a náležitostech profilu zadavatele. Zároveň k těmto formulářům existuje rozsáhlá dokumentace a metodické pokyny na stránkách VVZ a jakákoliv změna je uvedena v dané vyhlášce. Otevřené datové sady z ISVZ navíc z těchto formulářů vycházejí, ale díky neexistující metodice neznáme přesné zdroje všech dat a tím vznikají nejistoty při zpracování. Navíc metoda webscrapping, jak bylo demonstrováno, umožňuje využití i dalších datových zdrojů a jejich napojení na existující data. Pokud by k otevřeným datům z ISVZ existovala jednotná a jasná metodika, byly by jasným favoritem pro zdroj dat na analýzu VZ. Jelikož ale tato metodika neexistuje, je vhodné zamyslet se nad alternativními cestami a jednou z těchto cest by mohla být právě metoda webscrapping. Bylo by také možné do analýzy zahrnout národní formuláře a tím rozšířit počet zahrnutých zakázek.

Při případném budoucím zpracování dat o VZ by také možná bylo vhodné volit cestu položení výzkumné otázky a na tuto otázku tvořit datový set. Snaha o vytvoření univerzálního datového setu není ideální, jelikož problematika je velice komplexní a aby byl výsledný datový set opravdu kvalitní, je zapotřebí použít kvalitní ontologii. Tvorba ontologie je však samostatné komplikované téma, kterým se zabývají výzkumné týmy po mnoho let. Například tým, který stojí za platformou TheyBuyForYou pracuje na ontologii pro danou databázi již od roku 2016 [36]. Jak bylo popsáno výše vytvořený datový set má mnoho limitací a je tedy otázkou do jaké míry provedené analýzy opravdu reprezentují skutečnost.

Deskriptivní analýza

K popsání finančních toků mezi zadavateli a dodavateli byly použity alluvial grafy. Ty dobře reprezentují zastoupení největších hráčů na poli VZ na zdravotnickou techniku. Z 9 největších zadavatelů dle finančního objemu VZ je 5 subjektů FN. Je ale zajímavé, že se zde vyskytují 4 subjekty, které do této kategorie nespádají. Jedním z těchto subjektů je i Krajská zdravotní, jehož majoritním vlastníkem je Ústecký kraj. Tedy skupina zdravotnických zařízení, provozována jako akciová společnost, vlastněná jedním z krajů

je jedním z hlavních hráčů na poli VZ na zdravotnickou techniku. Zároveň co do počtu zaměstnanců spadá do kategorie 10 000 a více, čímž se řadí na první místo jako největší zaměstnavatel v oboru zdravotnictví. Karlova univerzita není zahrnuta, jelikož se primárně věnuje činnosti vzdělávací a výzkumné.

Obrázek 5.2 naznačuje poměrně rovnoměrné rozložení toku financí mezi 10 největšími dodavateli a 18 největšími zákazníky. Dodavatelé tak mají poměrně diverzifikované zdroje příjmu. Výjimku tvoří společnosti Cardiomedical a LINET.

Největší zadavatel VZ dle finančního objemu FN Olomouc zadal 9 největším dodavatelům zakázky za 942 milionů Kč. To představuje 72 % celkového finančního objemu zakázek. Bylo by možné uvažovat nad diverzifikací dodavatelů, což by mohlo snížit náklady z důvodu konkurence, ale na druhou stranu zvýšit náklady z důvodu nutnosti správy více zařízení od více výrobců. Největší dodavatel VZ dle finančního objemu, firma Siemens, dodala 9 největším zadavatelům zdravotnickou techniku za 888 milionů Kč, což představuje 64 % z celkového objemu zakázek. Tento dodavatel má dobré rozložení finančních toků mezi své zákazníky, což je patrné i z obrázku 5.2.

Datový set obsahuje 22 zadavatelů, kteří mají 2500 zaměstnanců a více. Do tohoto seznamu ale vstupují i subjekty, které nemají jako svou hlavní činnost poskytování zdravotní péče, například Karlova univerzita, nebo Ministerstvo obrany.

Z vytvořených datových setů je možné získat nepřehledné množství informací o rozložení trhu zdravotnické techniky. Zajímavé by mohlo být například zobrazení finančních toků mezi fakultními nemocnicemi a dodavateli. Tato práce si bere za cíl pouze demonstrovat tuto možnost. Další výstupy z datových setů mohou být námětem pro budoucí práci.

Regresní analýza

Na hodnotu úspory má statisticky významný vliv počet přihlášených dodavatelů a počet zadaných kritérií. Obě tyto proměnné zvyšují míru úspory. Tyto výsledky se shodují i s výsledky prací autorů Plaček a spol. [10, 12] a Nemeč a spol. [16]. Stejnou závislost popisuje i Ing. Balounová v rámci své diplomové práce, která vychází ze stejného primárního zdroje dat, jako tato práce [50]. Výsledky této regresní analýzy splnili předpoklady. Nezávislá proměnná prošla Shapiro-Wilkovým testem normality, p-hodnota $< 2,2 \cdot 10^{-16}$. Závislá proměnná tedy pochází z normálního rozdělení.

Na druhou stranu je nutné uvést limity vytvořeného modelu. Index determinance R^2 pro druhý vytvořený vícenásobný lineární regresní model nabývá hodnoty 0,163, stejně tak adjustovaný R^2 . Testování modelu naznačuje, že nebyl zvolen vhodný model pro danou závislost. Ačkoliv p-hodnota F-testu dosahuje statisticky významné hodnoty $2,2 \cdot 10^{-16}$, tudíž můžeme zamítnout nulovou hypotézu o nulových korelačních koeficientech, nemůžeme s jistotou určit míru korelace mezi úsporou a nezávislými

veličinami, právě z důvodu nevhodného modelu. Tento fakt je dobře demonstrován i v grafech testování chyby ε viz obrázek 5.8. Chyba vykazuje nerovnoměrné rozložení v okolí modelu a Q-Q graf ukazuje jasnou ne-normalitu chyby v extrémních hodnotách. Z práce Ing. Balounové [50] vyplývá, že ani zvolením exponenciálního vztahu mezi závislou a nezávislými veličinami nedosáhneme ideálního modelu. Pravděpodobně se bude jednat o kombinaci, tedy o nelineární regresi. Upravení a vyladění regresního modelu může být námětem pro budoucí práci.

Zvolení závislé veličiny úspory má také své limitace. Ta je totiž vypočtena, mimo jiné, z hodnoty předpokládané ceny zakázky, která je určena zadavatelem na základě kvalifikovaného odhadu, nebo na základě před-tržní poptávky. V tomto případě je spekulativní, zda zadavatel úmyslně nenadhodnocuje tuto částku, aby si vytvořil jakousi rezervu pro alokaci financí na VZ. Tomu napovídá fakt, že z 2650 zadání má zápornou hodnotu úspory pouze 445 z nich. Tedy zadavatel podhodnotil cenu zakázky pouze v 16,79 % případech, zatímco nadhodnocení se dopustil ve zbylých 83,21 % případech. I kdyby vytvořený model prošel testováním, není jisté, že vypočtený model by byl správný, právě díky nadhodnocení předpokládané ceny zakázky. Míra nadhodnocení předpokládané ceny zakázky může být námětem pro budoucí práci.

Algoritmus pro automatické určení předmětu zakázky dle KZT

Nejvyšší dosažená úspěšnost algoritmu při určení předmětu VZ dle KZT byla 79 %. Při iteracích parametrů měla největší vliv úprava knihoven výrazů. Zajímavé je, jaký měli vliv jednotlivé iterace na úspěšnost v jednotlivých kategoriích KZT. V případě úpravy knihovny výrazů byla při použití automatické knihovny úspěšnost v kategorii VENT pouze 37,5 %, zatímco s upravenou knihovnou 87,5 %. U kategorie ANEST se poměr obrací a při použití automatické knihovny dosahuje algoritmus úspěšnosti 85,7 %, zatímco při použití upravené 57,1 %. Nastavení knihoven pro různé kategorie je tedy zásadní pro úspěšnost algoritmu. Je také nutné zdůraznit, že algoritmus byl testován pouze na 5 kategoriích KZT. Pokud by měl být tento algoritmus aplikován na všechna získaná zadání, množství knihoven by bylo o mnoho větší, aby ideálně pokrylo veškerou zdravotnickou techniku. Odladění každé knihovny by poté bylo klíčovým úkolem pro zajištění úspěšnosti. Je ale jasné, že automatická tvorba knihovny výrazů není vhodným přístupem.

Parametr vzdálenost pro shodu znaků v řetězci nemá při zvolených iteracích žádný vliv na úspěšnost algoritmu. Naopak parametr minimální hodnoty p-indexu pro prohlášení shody slov má jasný dopad na úspěšnost algoritmu, a to v různé míře pro různé kategorie. Největší rozdíl je opět vidět v kategorii VENT, kdy při hodnotě indexu 0,7 je úspěšnost 50 % a při hodnotě indexu 0,9 100 %. Index 0,7 vykazuje velmi slabé výsledky v kategorii VENT, ale průměrné, až nadprůměrné ve všech ostatních. Index 0,8 vykazuje

slabé výsledky v kategorii ANEST, ale průměrné, maximálně lehce podprůměrné ve všech ostatních kategoriích a nakonec index 0,9 vykazuje silně podprůměrné výsledky v kategorii MON, ale nadprůměrné nebo průměrné výsledky ve všech ostatních kategoriích. V těchto výsledcích není možné nalézt nějaký trend. Je ale zajímavé, že žádná iterace nebyla schopna správně určit více než 8 zakázek z kategorie UZ a RTG. To ukazuje na možnost výběru zakázky, která byla chybně popsána, a proto ji algoritmus nemůže rozlišit. Pro ověření tohoto parametru by bylo vhodné provést testy na větším objemu dat.

Shoda algoritmu je nejvyšší při použití konstanty síly efektu prefixu s hodnotou 0,1 a to napříč všemi kategoriemi. Úspěšnost algoritmu při nastavení konstanty na 0,1 je 79 %, zatímco při nastavení konstanty na 0,25 je pouze 70 %.

Při automatickém přiřazení KZT musíme brát v potaz, že strom KZT je poměrně obsáhlý a obsahuje hodně detailů. Tyto detaily není možné zpracovat v plném rozsahu, proto se ukázalo vhodným, pro úspěšnost tohoto algoritmu, zvolit kategorii KZT 2. úrovně. Je zde ovšem limitace a to, že ne vždy je 2. úroveň dostatečně podrobná. Například pod pojmem RTG se skrývají i CT přístroje. Naopak v případě kategorie Monitorovací systémy je toto třídění dostatečné. Námětem pro další práci by tedy mohlo být určení vhodných kategorií pro vytvoření knihoven. Tyto knihovny optimalizovat a následně nový algoritmus vyzkoušet na větším data setu. Metoda fuzzy-matching pro přiřazení kategorie KZT vykazuje slibné výsledky a s případným odladěním a implementací by mohla být funkční metodou pro určení předmětu VZ dle KZT.

Byl také vyzkoušen přístup přiřazení kategorie zdravotnické techniky dle CPV kódu. Výsledek tohoto algoritmu pro vybrané kategorie KZT je v tabulce FNtb, viz codebook, příloha D. Výsledky tohoto algoritmu nejsou ale v této práci uváděny, z důvodů nespolehlivého zadávání CPV kódů zadavateli, jak bylo popsáno výše. Tato metoda určení kategorie KZT tedy není spolehlivá, mohla by být ale doplňujícím prvkem pro metodu fuzzy-matching, například pro křížovou kontrolu výsledků.

7 Závěr

Byl vypracován algoritmus pro získání datové sady o veřejných zakázkách na zdravotnickou techniku. Tento datový set vznikl spojením a úpravou tabulek z ISVZ a obohacením o informace z databáze RES a formulářů z VVZ. K obohacení byla použita metoda webscrapping. K této datové sadě byl vypracován codebook, který je součástí této práce, viz příloha D. Nedostatky tohoto datového setu vycházejí z principu vydávání dat na ISVZ. Možným řešením by mohla být kombinace dat z ISVZ s metodou webscrapping.

Pomocí alluvial grafů byly popsány finanční toky mezi největšími dodavateli a zadavateli VZ dle finančního objemu zakázek. Dále byla provedena regresní analýza, která potvrdila, že na míru úspory má vliv počet přihlášených uchazečů ve výběrovém řízení a počet nastavených kritérií. Tyto parametry zvyšují míru úspory. Výsledky odpovídají výstupům již provedených studií. Není ale možné určit přesnou míru regrese, jelikož byl použit nevhodný regresní model. Námětem pro budoucích práci by mohla být optimalizace regresního modelu.

Byla otestována metoda fuzzy-matching pro přiřazení kategorie zdravotnické techniky dle číselníku KZT. Algoritmus byl otestován s 5 vybranými kategoriemi na vzorku 43 zakázek. Tato metoda vykazuje slibné výsledky, je ale nutné další ověření na větším vzorku zakázek.

Otevřená data jsou obrovskou příležitostí, ale i výzvou. Obsahují obrovské množství potencionálně užitečných dat. Získání těchto dat, jejich zpracování a prezentace však není triviálním úkolem. Metody automatizace jsou správným přístupem ke zpracování těchto objemných setů, stejně tak snaha o rozšiřování a propojování s dalšími zdroji. Pokud zpracování doplníme o metody NLP, můžeme dostat kvalitní a ucelený datový setu pro potřeby analýz, s velkou přidanou hodnotou. Zdroje dat a jejich zpracování musí být, ale řádně ošetřeny a všechny limity by měli být řádně popsány. Bez splnění tohoto předpokladu nelze očekávat spolehlivé a kvalitní výstupy z prováděných analýz.

Seznam použité literatury

- [1] ŠERAFIN, Petr. *Zákon o zadávání veřejných zakázek*. Praha: Informační centrum ČKAIT, 2016. ISBN 978-80-87438-80-0.
- [2] EUROPEAN COMMISSION. *Public Procurement Indicators 2017* [online]. B.m.: Evropská komise. 2019 [vid. 2022-05-08]. Dostupné z: <https://ec.europa.eu/docsroom/documents/38003/attachments/1/translations/en/renditions/native>
- [3] MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR. *Výroční zpráva o stavu veřejných zakázek v České republice za rok 2020* [online]. B.m.: Ministerstvo pro místní rozvoj ČR. 2021 [vid. 2022-05-08]. Dostupné z: https://komora.cz/files/uploads/2021/05/69_Materi%C3%A1l.docx
- [4] SIMPERL, Elena, Oscar CORCHO, Marko GROBELNIK, a ET AL. Towards a Knowledge Graph Based Platform for Public Procurement. *Metadata and Semantic Research* [online]. 2019, **2018**(846), 317–323. Dostupné z: doi:10.1007/978-3-030-14401-2
- [5] MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR. *OpenData - ISVZ* [online]. 2022 [vid. 2022-03-27]. Dostupné z: <https://isvz.nipez.cz/opendata/predchozi>
- [6] MINISTERSTVO VNITRA ČESKÉ REPUBLIKY. *Standardy publikace a katalogizace otevřených dat VS ČR* [online]. B.m.: MVČR. 2015 [vid. 2022-05-08]. Dostupné z: https://opendata.gov.cz/_media/standardy_publikace_a_katalogizace_otevrenych_dat_vs_cr.pdf
- [7] QENDRI, Venetia, Johannes A. BOGAARDS a Johannes BERKHOF. Pricing of HPV vaccines in European tender-based settings. *The European Journal of Health Economics* [online]. 2019, **20**(2), 271–280. ISSN 1618-7601. Dostupné z: doi:10.1007/s10198-018-0996-9
- [8] CURTO, Sandro, Simone GHISLANDI, Katelijne VAN DE VOOREN, a ET AL. Regional tenders on biosimilars in Italy: An empirical analysis of awarded prices. *Health Policy* [online]. 2014, **116**(2–3), 182–187. ISSN 18726054. Dostupné z: doi:10.1016/j.healthpol.2014.02.011
- [9] CHIAPPINELLI, Olga. Decentralization and Public Procurement Performance: New Evidence From Italy. *Economic Inquiry* [online]. 2020, **58**(2), 856–880. ISSN 14657295. Dostupné z: doi:10.1111/ecin.12871
- [10] PLAČEK, Michal, Juraj NEMEC, František OCHRANA, a ET. AL. Analysis of Factors of Overpricing in Public Procurement: A Study for Low-performing EU Countries. *International Journal of Public Administration* [online]. 2020, **43**(4), 350–360. ISSN 15324265. Dostupné z: doi:10.1080/01900692.2019.1636393
- [11] PLAČEK, Michal, Martin SCHMIDT, František OCHRANA, a ET. AL. Factors affecting the length of procedure in public procurement: The case of the Czech Republic. *Prague Economic Papers* [online]. 2019, **28**(3), 313–329. ISSN 2336730X. Dostupné z: doi:10.18267/j.pep.692
- [12] PLAČEK, Michal, Jana SOUKOPOVÁ, Gabriela VACEKOVÁ, a ET AL. Another One Bites the Dust: Recurrent Procurement Contracts in the Czech

- Republic. *Public Works Management and Policy* [online]. 2019. ISSN 15527549. Dostupné z: doi:10.1177/1087724X19876647
- [13] PLÁČEK, Michal, František OCHRANA, Martin SCHMIDT, a ET AL. The factors causing delays in public procurement: the Czech Republic versus the UK. *Public Money and Management* [online]. 2020, **40**(2), 131–139. ISSN 14679302. Dostupné z: doi:10.1080/09540962.2019.1651034
- [14] OCHRANA, František a Kristýna HRNČÍŘOVÁ. Does the Lowest Bid Price Evaluation Criterion Make for a More Efficient Public Procurement Selection Criterion? (Case of the Czech Republic). *NISPAcee Journal of Public Administration and Policy* [online]. 2015, **8**(1), 41–59. ISSN 1338-4309. Dostupné z: doi:10.1515/nispa-2015-0003
- [15] GAVUROVÁ, Beáta, Jaroslav BELAS, Zuzana ROWLAND, a ET AL. The impact of agreement on government procurement use on the competition in Slovak healthcare sector. *ADMINISTRATIE SI MANAGEMENT PUBLIC* [online]. 2021, 102–115. Dostupné z: doi:10.24818/amp/2021.36-06
- [16] NEMEC, Juraj, Matúš KUBÁK, Gleb DONIN, a ET AL. Efficiency of Public Procurement in the Czech and Slovak Health Care Sectors. *Transylvanian Review of Administrative Sciences* [online]. 2021, **17**(62), 115–133. ISSN 1842-2845. Dostupné z: doi:10.24193/tras.62E.7
- [17] PODEŠVA, Vilém, Lukáš SOMMER, Jiří VOTRUBEC, Martin FLAŠKÁR, Jiří HARNACH, Jan MĚKOTA a Martin JANOUŠEK. *Zákon o zadávání veřejných zakázek. Zákon o registru smluv. Komentář*. Praha: Wolter Kluwer ČR, 2016. ISBN 978-80-7552-102-6.
- [18] POREMSKÁ, Michaela. *Veřejné zakázky stručně a prakticky*. Praha: Wolters Kluwer ČR, 2017. ISBN 978-80-7552-697-7.
- [19] ÚŘAD PRO PUBLIKACE. *Ted-tenders electronic daily* [online]. 2020 [vid. 2022-08-05]. Dostupné z: <https://ted.europa.eu/TED/search/search.do>
- [20] ČESKO. *Vyhláška č. 168/2016 Sb. o uveřejňování formulářů pro účely zákona o zadávání veřejných zakázek a náležitostech profilu zadavatele* [online]. 2016 [vid. 2022-05-08]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2016-168>
- [21] ROZZA [online]. [vid. 2022-03-27]. Dostupné z: <https://rozza.cz/zakazky>
- [22] MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR. *OpenData - ISVZ* [online]. 2022 [vid. 2022-03-27]. Dostupné z: <https://isvz.nipez.cz/opendata>
- [23] SPOJENÉ NÁSTROJE ELEKTRONICKÉ. *Rozza* [online]. 2022 [vid. 2022-03-27]. Dostupné z: <https://rozza.cz/#filtr-zakladni>
- [24] MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR. *Metodika tvorby statistik - ISVZ* [online]. 2022 [vid. 2022-03-27]. Dostupné z: <https://isvz.nipez.cz/centrum-podpory/napoveda/metodika-tvorby-statistik>
- [25] PROKOP, Pavel a Jan DRDA. *Tvorba OpenDat o veřejných zakázkách* [online]. B.m.: Datera s.r.o. 2022 [vid. 2022-03-27]. Dostupné z: <https://isvz.nipez.cz/sites/default/files/content/opendata-dokumentace.pdf>
- [26] COSINEX. *Revision of CPV* [online]. B.m.: European Commission. 2017 [vid. 2022-05-08]. Dostupné

z: <https://ec.europa.eu/docsroom/documents/27821/attachments/1/translations/en/renditions/pdf>

- [27] EVROPSKÁ KOMISE. *Příručka ke společnému slovníku pro veřejné zakázky* [online]. B.m.: Evropská komise. 2008 [vid. 2022-05-08]. Dostupné z: https://simap.ted.europa.eu/documents/10184/36234/cpv_2008_guide_cs.pdf
- [28] ÚŘAD PRO PUBLIKACE. *CPV* [online]. 2020 [vid. 2022-05-08]. Dostupné z: <https://simap.ted.europa.eu/cs/cpv>
- [29] PUBLICTENDERING. List of the CPV codes. *PublicTendering* [online]. 2021 [vid. 2022-05-08]. Dostupné z: <https://www.publictendering.com/cpv-codes/list-of-the-cpv-codes/>
- [30] EVROPSKÁ KOMISE. *Narizení Komise (ES) č. 213/2008 ze dne 28. listopadu 2007, kterým se mění nařízení Evropského parlamentu a Rady (ES) č. 2195/2002 o společném slovníku pro veřejné zakázky (CPV) a směrnice Evropského parlamentu a Rady 2004/17/ES a 2004/18/ES o postupech při zadávání zakázek, pokud jde o přezkum CPV (Text s významem pro EHP)* [online]. 2008 [vid. 2022-05-08]. Dostupné z: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:074:0001:0375:CS:PDF>
- [31] NESS CZECH. *VĚSTNÍK VEŘEJNÝCH ZAKÁZEK* [online]. 2020 [vid. 2022-05-08]. Dostupné z: <https://www.vestnikverejnychzakazek.cz/>
- [32] MÍŠEK, Jakub. *Co jsou otevřená data?* [online]. 2021 [vid. 2022-05-08]. Dostupné z: <https://opendata.gov.cz/informace:start>
- [33] SVÁTEK, Vojtěch, Jindřich MYNARZ, Jakub KLÍMEK, a ET AL. Linked Open Data for Public Procurement. In: *Linked Open Data -- Creating Knowledge Out of Interlinked Data* [online]. 2014, s. 196–213. ISBN 978-3-319-09845-6. Dostupné z: [doi:10.1007/978-3-319-09846-3_10](https://doi.org/10.1007/978-3-319-09846-3_10)
- [34] SOYLU, Ahmet, Oscar CORCHO, Elena SIMPERL, a ET AL. Towards integrating public procurement data into a semantic knowledge graph. *CEUR Workshop Proceedings*. 2018, **2262**(780247), 1–4. ISSN 16130073.
- [35] SOYLU, Ahmet, Oscar CORCHO a Brian ELVESÆTER. TheyBuyForYou platform and knowledge graph: Expanding horizons in public procurement with open linked data. *Semantic Web* [online]. 2022, **13**(2), 265–291. ISSN 22104968, 15700844. Dostupné z: [doi:10.3233/SW-210442](https://doi.org/10.3233/SW-210442)
- [36] MUÑOZ-SORO, José Félix, Guillermo ESTEBAN, Oscar CORCHO, a ET AL. PPROC, an ontology for transparency in public procurement. *Semantic Web* [online]. 2016, **7**(3), 295–309. ISSN 22104968. Dostupné z: [doi:10.3233/SW-150195](https://doi.org/10.3233/SW-150195)
- [37] LE BORGNE, Yann Aël, Adriana HOMOLOVA a Gianluca BONTEMPI. OpenTED browser: Insights into European Public Spendings. *CEUR Workshop Proceedings*. 2016, **1831**, 1–12. ISSN 16130073.
- [38] PLAČEK, M., F. OCHRANA, M. PŮČEK, a ET AL. Autonomy and performance of local museums: the case of Czech Republic. *Museum Management and Curatorship* [online]. 2021, **36**(5), 504–522. ISSN 0964-7775. Dostupné z: [doi:10.1080/09647775.2020.1803112](https://doi.org/10.1080/09647775.2020.1803112)

- [39] PLAČEK, M., F. OCHRANA, G. VACEKOVÁ, a ET AL. Shedding Light on Three “Moral Cancer” Cases From the Czech Republic: Corruption in Communist and Post-Communist Regimes. *Public Integrity* [online]. 2021. ISSN 1099-9922. Dostupné z: doi:10.1080/10999922.2021.1988407
- [40] CHOWDHARY, K.R. *Fundamentals of Artificial Intelligence* [online]. New Delhi: Springer India, 2020. ISBN 978-81-322-3970-3. Dostupné z: doi:10.1007/978-81-322-3972-7
- [41] NADKARNI, Prakash M, Lucila OHNO-MACHADO a Wendy W CHAPMAN. Natural language processing: an introduction. *Journal of the American Medical Informatics Association* [online]. 2011, **18**(5), 544–551. ISSN 1067-5027. Dostupné z: doi:10.1136/amiajnl-2011-000464
- [42] MEHRBOD, Ahmad, Aneesh ZUTSHI, António GRILO, a ET AL. Application of a semantic product matching mechanism in open tendering e-marketplaces. *Journal of Public Procurement* [online]. 2018, **18**(1), 14–30. ISSN 21506930. Dostupné z: doi:10.1108/JOPP-03-2018-002
- [43] JANLTA INTERACTIVE. *What is Fuzzy Matching? - Definition from Techopedia* [online]. [vid. 2022-03-21]. Dostupné z: <http://www.techopedia.com/definition/24183/fuzzy-matching>
- [44] R-PROJECT. *Fuzzy Matching* [online]. [vid. 2022-03-21]. Dostupné z: <https://cran.r-project.org/web/packages/fedmatch/vignettes/Fuzzy-matching.html>
- [45] MEHRBOD, Ahmad a António GRILO. Tender calls search using a procurement product named entity recogniser. *Advanced Engineering Informatics* [online]. 2018, **36**(June 2017), 216–228. ISSN 14740346. Dostupné z: doi:10.1016/j.aei.2018.04.005
- [46] ÚZIS. *Kategorizace zdravotnické techniky* [online]. 2022 [vid. 2022-05-08]. Dostupné z: <https://www.uzis.cz/index.php?pg=registry-sber-dat--klasifikace--kategorizace-zdravotnicke-techniky#o-klasifikaci>
- [47] PRACOVNÍ SKUPINA URČENÁ PRO REALIZACI KATEGORIZAČNÍHO SYSTÉMU ZDRAVOTNICKÉ PŘÍSTROJOVÉ TECHNIKY. *Metodika kategorizace zdravotnické techniky* [online]. B.m.: ÚZIS, 2021 [vid. 2022-05-08]. Dostupné z: <https://www.uzis.cz/res/file/klasifikace/kzt/kzt-metodika-5-0.pdf>
- [48] THE R FOUNDATION. *R: The R Project for Statistical Computing* [online]. [vid. 2022-05-01]. Dostupné z: <https://www.r-project.org/>
- [49] RSTUDIO. *About RStudio* [online]. [vid. 2022-05-01]. Dostupné z: <https://www.rstudio.com/about/>
- [50] BALOUNOVÁ, Zlata. *ANALÝZA VEŘEJNÝCH ZAKÁZEK NA ZDRAVOTNICKOU TECHNIKU* [online]. B.m., 2019 [vid. 2022-05-08]. Diplomová práce. České vysoké učení technické v Praze. Dostupné z: <https://dspace.cvut.cz/handle/10467/91433>
- [51] ČESKO. *Zákon č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování (zákon o zdravotních službách)* [online]. 2011 [vid. 2022-05-01]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2011-372>

- [52] MONTGOMERY, Douglas C., Elizabeth A. PECK a G. Geoffrey VINING. *Introduction to linear regression analysis*. 5th ed. Hoboken, NJ: Wiley, 2012. Wiley series in probability and statistics, 821. ISBN 978-0-470-54281-1.
- [53] RABUŠIC, Ladislav. *Mnohonásobná lineární regrese: Studijní materiály k předmětu SOC418 Analýza kategorizovaných dat* [online]. 2004 [vid. 2022-05-03]. Dostupné z: https://is.muni.cz/el/1423/podzim2004/SOC418/multipl_regres_1.pdf
- [54] JIANG, Jinhang. *Linear Regression Analysis in R* [online]. 30. květen 2021 [vid. 2022-05-03]. Dostupné z: <https://towardsdatascience.com/linear-regression-analysis-in-r-fdd59295d4a8>
- [55] STATISTICS HOW TO. *F Statistic / F Value: Simple Definition and Interpretation* [online]. 2022 [vid. 2022-05-03]. Dostupné z: <https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/>
- [56] FORD, Clay. *Understanding Q-Q Plots* [online]. 2015 [vid. 2022-05-03]. Dostupné z: <https://data.library.virginia.edu/understanding-q-q-plots/>
- [57] LOO, Mark van der, Jan van der LAAN, R CORE TEAM, a ET AL. *stringdist: Approximate String Matching, Fuzzy Text Search, and String Distance Functions* [online]. 2021 [vid. 2022-05-03]. Dostupné z: <https://CRAN.R-project.org/package=stringdist>
- [58] MINISTERSTVO PRO MÍSTNÍ ROZVOJ ČR. *Výroční zpráva o stavu veřejných zakázek v České republice za rok 2019* [online]. B.m.: Ministerstvo pro místní rozvoj ČR. 2020 [vid. 2022-05-08]. Dostupné z: <http://www.portal-vz.cz/getmedia/bdb1b032-f6cf-4de2-aadd-76491f13f954/Vyrocní-zpráva-za-rok-2019-f.pdf>
- [59] ČESKO. *Narizení vlády č. 172/2016 Sb., o stanovení finančních limitů a částek pro účely zákona o zadávání veřejných zakázek* [online]. 2016 [vid. 2022-05-07]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2016-172/zneni-20220101>

Příloha A: Základní typy formulářů

EVROPSKÉ FORMULÁŘE DLE PROVÁDĚCÍHO NAŘÍZENÍ KOMISE EU/2015/1986 PRO NADLIMITNÍ ZAKÁZKY		
Číslo formuláře	Název formuláře	Účel formuláře
F01	Předběžné oznámení	<ul style="list-style-type: none"> • Předběžná informace o zadání zakázky, • zkrácení lhůty pro doručení nabídek, • výzva k účasti v soutěži.
F02	Oznámení o zahájení zadávacího řízení	<ul style="list-style-type: none"> • Zahájení otevřeného řízení, užšího řízení, jednacího řízení s uveřejněním, řízení se soutěžním dialogem, řízení o inovačním partnerství. • Zřízení a prodloužení dynamického nákupního systému.
F03	Oznámení o výsledku zadávacího řízení	<ul style="list-style-type: none"> • Oznámení o zadání zakázky v otevřeném řízení, užším řízení, jednacím řízení s uveřejněním, jednacím řízení bez uveřejněním, soutěžním dialogu, řízení o inovačním partnerství • Uzavření smlouvy na základě rámcové dohody, nebo na základě dynamického nákupního systému • V případě zakázky rozdělené na části: postupné uveřejnění částí, jak jsou známy výsledky částí zadávacích řízení
F04	Pravidelné předběžné oznámení – veřejné služby	<ul style="list-style-type: none"> • Jako F01, ale v případě sektorové veřejné zakázky
F05	Oznámení o zahájení zadávacího řízení – veřejné služby	<ul style="list-style-type: none"> • Jako F02, ale v případě sektorové veřejné zakázky
F06	Oznámení o výsledku zadávacího řízení – veřejné služby	<ul style="list-style-type: none"> • Jako F031, ale v případě sektorové veřejné zakázky
F07	Systém kvalifikace – veřejné služby	<ul style="list-style-type: none"> • Zavedení systému kvalifikace v případě sektorové veřejné zakázky

F08	Oznámení na profilu kupujícího	<ul style="list-style-type: none"> • Zadavatel vyplňuje formulář v případě, že hodlá nahradit uveřejnění předběžného oznámení nebo pravidelného předběžného oznámení uveřejněním těchto oznámení na svém profilu.
F12	Oznámení soutěže o návrh	<ul style="list-style-type: none"> • Oznamuje zahájení soutěže o návrh a tím dává možnost přihlášení pro dodavatele.
F13	Výsledky soutěže o návrh	<ul style="list-style-type: none"> • Oznamuje výherce soutěže o návrh.
F14	Oprava	<ul style="list-style-type: none"> • Oprava formulářů F01-F8, F12, F13, F15 a F21-F25 • Formuláře F14 je možné navázat jen na formuláře jiného čísla než F14 a F20.
F15	Oznámení o dobrovolné průhlednosti ex ante	<ul style="list-style-type: none"> • Zadavatel v případech, kdy nebyl povinen uveřejnit oznámení o zahájení zadávacího řízení, může uveřejnit oznámení o záměru uzavřít smlouvu prostřednictvím formuláře Oznámení o dobrovolné průhlednosti ex ante. • Oznámení se podává do VVZ i TED
F16	Oznámení předběžných informací – obrana a bezpečnost	<ul style="list-style-type: none"> • Jako F01, v případě zakázky v oblasti obrany a bezpečnosti.
F17	Oznámení o zakázce – obrana a bezpečnost	<ul style="list-style-type: none"> • Jako F02, v případě zakázky v oblasti obrany a bezpečnosti.
F18	Oznámení o zadání zakázky – obrana a bezpečnost	<ul style="list-style-type: none"> • Jako F03, v případě zakázky v oblasti obrany a bezpečnosti.
F19	Oznámení o subdodávce – obrana a bezpečnost	<ul style="list-style-type: none"> • Formulář pro oznámení o poddodávce dle ZZVZ § 208.
F20	Oznámení o změně	<ul style="list-style-type: none"> • Oznámení o změně § 222 odst. 8 ZZVZ. Formulář F20 slouží k oznámení o změně plnění veřejné zakázky. Například vznik víceprací. Tato úprava se dokládá tímto formulářem. • Formulář F20 není možné upravovat formulářem F14. Je nutné založit nový a navázat.

F21	Sociální a jiné zvláštní služby – veřejné zakázky	<ul style="list-style-type: none"> • Předběžné oznámení veřejné zakázky ve zjednodušeném režimu • Výzva k účasti v soutěži veřejné zakázky ve zjednodušeném režimu • Oznámení o zahájení zadávacího řízení veřejné zakázky ve zjednodušený režim • Oznámení o výsledku zadávacího řízení veřejné zakázky ve zjednodušený režim • NEPOUŽÍVAT <ul style="list-style-type: none"> ○ Koncesní řízení (formulář F23) ○ Sektorové zakázky (formulář F22)
F22	Sociální a jiné zvláštní služby – veřejné služby	<ul style="list-style-type: none"> • Používá se pro sektorové zakázky na sociální a jiné zvláštní služby • Předběžné oznámení • Výzva k účasti v soutěži • Zavedení systému kvalifikace • Oznámení o zahájení zadávacího řízení • Oznámení o výsledku zadávacího řízení
F23	Sociální a jiné zvláštní služby – koncese	<ul style="list-style-type: none"> • Používá se pro koncesní řízení na sociální a jiné zvláštní služby • Oznámení o zahájení koncesního řízení • Oznámení o výsledku koncesního řízení
F24	Oznámení o zahájení koncesního řízení	<ul style="list-style-type: none"> • Pro zahájení koncesního řízení
F25	Oznámení o výsledku koncesního řízení	<ul style="list-style-type: none"> • Pro vyhlášení výsledku koncesního řízení

NÁRODNÍ FORMULÁŘE DLE VYHLÁŠKY č. 168/2016 Sb. PRO PODLIMITNÍ ZAKÁZKY

Číslo formuláře	Název formuláře	Účel formuláře
CZ01	Předběžné oznámení zadávacího řízení v podlimitním režimu	<ul style="list-style-type: none">• Předběžná informace• Zkrácení lhůt pro doručení nabídek• Výzva k účasti v soutěži
CZ02	Oznámení o zahájení zadávacího řízení v podlimitním režimu	<ul style="list-style-type: none">• Oznámení o zahájení otevřeného řízení, užšího řízení, jednacího řízení s uveřejněním, řízení se soutěžním dialogem a řízení o inovačním partnerství.• Zavedení dynamického nákupního systému• Změna doby trvání dynamického nákupního systému
CZ03	Oznámení o výsledku zadávacího řízení v podlimitním režimu	<ul style="list-style-type: none">• Oznámení o zadání zakázky v otevřeném řízení, užším řízení, jednacím řízení s uveřejněním, jednacím řízení bez uveřejněním, soutěžním dialogem, řízení o inovačním partnerství.
CZ04	Oprava národního formuláře	<ul style="list-style-type: none">• Oprava národního formuláře
CZ05	Oznámení profilu zadavatele	<ul style="list-style-type: none">• Zadavatel použije k uveřejnění informací ve Věstníku veřejných zakázek, které se týkají profilu zadavatele, národní formulář CZ05 pro oznámení profilu zadavatele.
CZ06	Zrušení/zneaktivnění profilu zadavatele	<ul style="list-style-type: none">• Zadavatel použije k uveřejnění informací ve Věstníku veřejných zakázek, které se týkají profilu zadavatele, národní formulář CZ 06 pro zrušení či zneaktivnění profilu zadavatele.

Příloha B: Ponechané a vyřazené položky z otevřeného datového setu z portálu ISVZ

PONECHANÉ POLOŽKY	
Tabulka „VZ“	
Položka	Popis
ID_Zakazky	Unikátní ID zakázky
EvidencniCisloVZnaVVZ	Evidenční číslo zakázka na VVZ
CisloFormulareNaVVZ	Evidenční číslo formuláře na VVZ
DruhFormulare	Typ formuláře (národní – CZ01, ..., evropský – F01, ...)
PlatnyFormular	Pouze aktuální platný formulář
VZdelenaNaCasti	VZ dělená na části. Pro národní formuláře CZ03 (Ano/Ne), CZ01 a CZ02 (Ano/-). Pro evropské formuláře F01-F02 (Ano/Ne)
ZadavatelICO	IČO zadavatele
LimitVZ	Limit VZ (nadlimitní/podlimitní/VZMR)
DatumOdeslaniFormulareNaVVZ	Datum odeslání formuláře provozovateli Věstníku VZ k uveřejnění.
DatumUverejneni	Datum zveřejnění na Věstníku VZ
ZadavatelUredniNazev	Úřední název zadavatele. U obchodních společností položka obsahuje název obchodní firmy zapsané v obchodním rejstříku, u fyzických osob obsahuje jméno a příjmení fyzické osoby.
ZadavatelDruh	Druh veřejného zadavatele. Položka nabývá jednu ze 7 nabízených možností podle § 2 ZVZ.
ZakazkaZahrnujeSpolecneZadavaniZakazek	Společné zadání zakázek více zadavateli
ZakazkuZadavaCentralniZadavatel	Centrální zadavatel (například v rámci nákupů v DNS)
NazevVZ	Název přidělený veřejné zakázce
DruhVZ	Druh zakázky (dodávky, služby, stavební práce)
StrucnyPopisVZ	Textový popis charakteristiky předmětu veřejné zakázky
CPVhlavni	Charakteristika VZ pomocí Společného slovníku pro veřejné zakázky (CPV – Common Procurement Vocabulary), který nejlépe popisuje hlavní předmět veřejné zakázky
CPVdoplnkovy1	Doplňkový slovník 1
CelkovaKonecnaHodnotaVZ	Celková konečná hodnota VZ, včetně všech zakázek, částí zakázek, obnovení a opcí.
CelkovaKonecnaHodnotaVZmena	Celková konečná hodnota VZ – Měna

DruhRizeni	Druh řízení dle ZVZ
BylaPouzitaElektronickaDrazba	Byla použita elektronická dražba jako prostředek pro hodnocení nabídek.
OdhadovanaHodnotaVZbezDPH	Předpokládaná hodnota VZ
OdhadovanaHodnotaVZmena	Měna předpokládané hodnoty VZ
LhutaProDoruceniNabidek	Lhůta pro doručení nabídek nebo žádostí o účast
OteviraniNabidekDatumCas	Datum a čas otevírání nabídek

Tabulka „Části VZ“

Položka	Popis
ID_Zakazky	Unikátní ID zakázky
EvidencniCisloVZnaVVZ	Evidenční číslo zakázka na VVZ
CisloFormulareNaVVZ	Evidenční číslo formuláře na VVZ
DruhFormulare	Typ formuláře (národní – CZ01, ..., evropský – F01, ...)
ID_CastiVZ	Unikátní ID části zakázky
CisloCastiVZ	Číslo části VZ udané zadavatelem
NazevCastiVZ	Název části VZ
PopisCastiVZ	Popis části VZ
PredpokladanaCelkovaHodnotaCastiVZ	Předpokládaná celková hodnota části VZ
PredpokladanaCelkovaHodnotaCastiVZMena	Měna předpokládané celkové hodnoty části VZ
HlavniMistoPlneniNUTS	Hlavní místo plnění – NUTS kód(y)
PlatnyFormular	Pouze aktuální platný formulář
CPVkod	CPV kód části zakázky

Tabulka „Hodnotící kritéria“

Položka	Popis
ID_Zakazky	Unikátní ID zakázky
ID_PopisuCastiZakazky	Unikátní ID části zakázky (odpovídá položce "ID_CastiVZ" z tabulky "Části VZ")
DruhFormulare	Typ formuláře (národní – CZ01, ..., evropský – F01, ...)
CisloCastiVZ	Číslo části VZ udané zadavatelem
NazevCastiVZ	Název části VZ
PoradiKriteria	Pořadí kritérií
KriteriumZadaniZakazky	Kritérium zadání zakázky – výběr z číselníku
NazevKriteria	Název kritéria – volný text
VahaKriteria	Váha kritéria v procentech

Tabulka „Zadání“

Položka	Popis
ID_Zakazky	Unikátní ID zakázky
EvidencniCisloVZnaVVZ	Evidenční číslo zakázka na VVZ
CisloFormulareNaVVZ	Evidenční číslo formuláře na VVZ

DruhFormulare	Typ formuláře (národní – CZ01, ..., evropský – F01, ...)
ID_Zadani	Unikátní ID zadání
CisloCastiZadaniVZ	Číslo části VZ udané zadavatelem
NazevCastiVZ	Název části VZ
PlatnyFormular	Pouze aktuální platný formulář
ZadaniCastiZakazky	Zadání části zakázky (ano/ne)
InformaceONezadaniCastiZakazky	Informace o nezadání části zakázky
DatumZadaniVZ	Datum zadání VZ, uvádí datum uzavření smlouvy s vybraných uchazečem
PocetObdrzenychNabidek	Počet obdržených nabídek (jen pro evropské formuláře F03)
PuvodniOdhadovanaCelkovaHodnotaVZ	Původní odhadovaná celková hodnota části zakázky.
PuvodniOdhadovanaCelkovaHodnotaVZMena	Původní odhadovaná celková hodnota části zakázky – Měna
CelkovaKonecnaHodnotaVZzaZadani	Celková konečná hodnota části zakázky
CelkovaKonecnaHodnotaVZmenaZaZadani	Celková konečná hodnota části zakázky – Měna

Tabulka „Dodavatelé“

Položka	Popis
ID_Zakazky	Unikátní ID zakázky
DruhFormulare	Typ formuláře (národní – CZ01, ..., evropský – F01, ...)
ID_Zadani	Unikátní ID zadání
DodavatelICO	IČO dodavatele, kterému byla zakázka zadána.
DodavatelNazev	Název dodavatele, kterému byla zakázka zadána.
DodavatelPostovniAdresa	Poštovní adresa dodavatele, kterému byla zakázka zadána.
DodavatelObec	Obec dodavatele, kterému byla zakázka zadána.
DodavatelPSC	PSC dodavatele, kterému byla zakázka zadána.
DodavatelStat	Stát dodavatele, kterému byla zakázka zadána.

VYŘAZENÉ POLOŽKY

Tabulka „VZ“

Položka	Popis
TypFormulare	Nepropisující se položka z VVZ
ZadavatelHlavniPredmetCinnosti	Nekonzistentní vyplňování, informace bez přidané hodnoty
ZadavatelProfilURL	Nekonzistentní vyplňování, informace bez přidané hodnoty

ZadavatelProfilURLPlatnost	Nekonzistentní vyplňování, informace bez přidané hodnoty
ZadavatelProfilZruseniTyp	Položka se týká formulářů CZ06 o zrušení/zneaktivnění profilu zadavatele, nepotřebná informace
ZadavatelProfilURLNova	Položka se týká formulářů CZ06 o zrušení/zneaktivnění profilu zadavatele, nepotřebná informace
DatumZruseniCiZneaktivneniProfilu	Položka se týká formulářů CZ06 o zrušení/zneaktivnění profilu zadavatele, nepotřebná informace
NaVZseVztahujeGPA	Nepropisující se položka z VVZ
NejnizsiNabidkaVzataVuvahu	Nekonzistentní vyplňování
NejnizsiNabidkaVzataVuvahuMena	Nekonzistentní vyplňování
NejvyssiNabidkaVzataVuvahu	Nekonzistentní vyplňování
ZadavatelKontaktniOsoba	Přímý kontakt na zadávající osobu, nepotřebná informace
ZadavatelEmail	Přímý kontakt na zadávající osobu, nepotřebná informace
ZadavatelTelefon	Přímý kontakt na zadávající osobu, nepotřebná informace
OteviraniNabidekMisto	Nekonzistentní vyplňování, ruční vyplnění uživatelem, nepotřebná informace
OteviraniNabidekOpravneneOsobyDalsiInfo	Nekonzistentní vyplňování, ruční vyplnění uživatelem, nepotřebná informace
UchazecVazanNabidkouDo	Nepotřebná informace
UchazecVazanNabidkouDobaMesice	Nepotřebná informace
PravidelnePredbezneOznameni	Nepotřebná informace, pouze pro formuláře F22 – Sociální a jiné zvláštní služby – veřejné služby
PravidelnePredbezneOznameniVyzvaKSoutez i	Nepotřebná informace, pouze pro formuláře F22 – Sociální a jiné zvláštní služby – veřejné služby
PredbezneOznameni	Nepotřebná informace, pouze pro formuláře F21 – Sociální a jiné zvláštní služby – veřejné zakázky a F23 – Sociální a jiné zvláštní služby – koncese
PredbezneOznameniVyzvaKSoutezi	Nepotřebná informace, pouze pro formuláře F21 – Sociální a jiné zvláštní služby – veřejné zakázky
SystemKvalifikace	Nepotřebná informace, pouze pro formuláře F22 – Sociální a jiné zvláštní služby – veřejné služby
SystemKvalifikaceVyzvaKSoutezi	Nepotřebná informace, pouze pro formuláře F22 – Sociální a jiné zvláštní služby – veřejné služby

OznameníOZahajeníZR	Nepotřebná informace, pouze pro formuláře F21 – Sociální a jiné zvláštní služby – veřejné zakázky a F22 – Sociální a jiné zvláštní služby – veřejné služby
OznameníOVysledkuZR	Nepotřebná informace, pouze pro formuláře F21 – Sociální a jiné zvláštní služby – veřejné zakázky a F22 – Sociální a jiné zvláštní služby – veřejné služby
OznameníOVysledkuKS	Nepotřebná informace, pouze pro formuláře „F23 – Sociální a jiné zvláštní služby – koncese“

Tabulka „Části VZ“

Položka	Popis
HlavniMistoPlneni	Vyplňováno uživatelem do textového pole, nejednotný formát, nevhodné pro další zpracování.
ZakazkaSeVztahujeKprojektuFinZes	Nepropisující se položka z VVZ
ProjektyCiProgramyFinZes	Nepropisující se položka z VVZ
DatumOdeslaniFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"

Tabulka „Hodnotící kritéria“

Položka	Popis
EvidencniCisloVZnaVVZ	Duplikovaný identifikátor z tabulky "VZ"
CisloFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"
DruhFormulare	Duplikovaný identifikátor z tabulky "VZ"
PlatnyFormular	Není použito k navázání na jinou tabulku
DatumOdeslaniFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"

Tabulka „Zadání“

Položka	Popis
HodnotaNejnizsiNabidky	Nekonzistentní vyplňování
HodnotaNejnizsiNabidkyMena	Nekonzistentní vyplňování
HodnotaNejvyssiNabidky	Nekonzistentní vyplňování
SubdodavkyHodnotaBezDPH	Informuje, zda je zadání řešeno dále subdodávku a v jakém rozsahu. Nekonzistentní vyplňování, nepotřebná informace
SubdodavkyMena	Informuje, zda je zadání řešeno dále subdodávku a v jakém rozsahu. Nekonzistentní vyplňování, nepotřebná informace
SubdodavkyPomer	Informuje, zda je zadání řešeno dále subdodávku a v jakém rozsahu. Nekonzistentní vyplňování, nepotřebná informace
DatumOdeslaniFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"

Tabulka „Dodavatelé“

Položka	Popis
EvidencniCisloVZnaVVZ	Duplikovaný identifikátor z tabulky "VZ"
CisloFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"
CisloCastiZadaniVZ	Duplikovaný identifikátor z tabulky "Části VZ"
NazevCastiVZ	Duplikovaný identifikátor z tabulky "Části VZ"
DodavatelWww	Nekonzistentní vyplňování, nepotřebná informace
PlatnyFormular	Není použito k navázání na jinou tabulku
DatumOdeslaniFormulareNaVVZ	Duplikovaný identifikátor z tabulky "VZ"

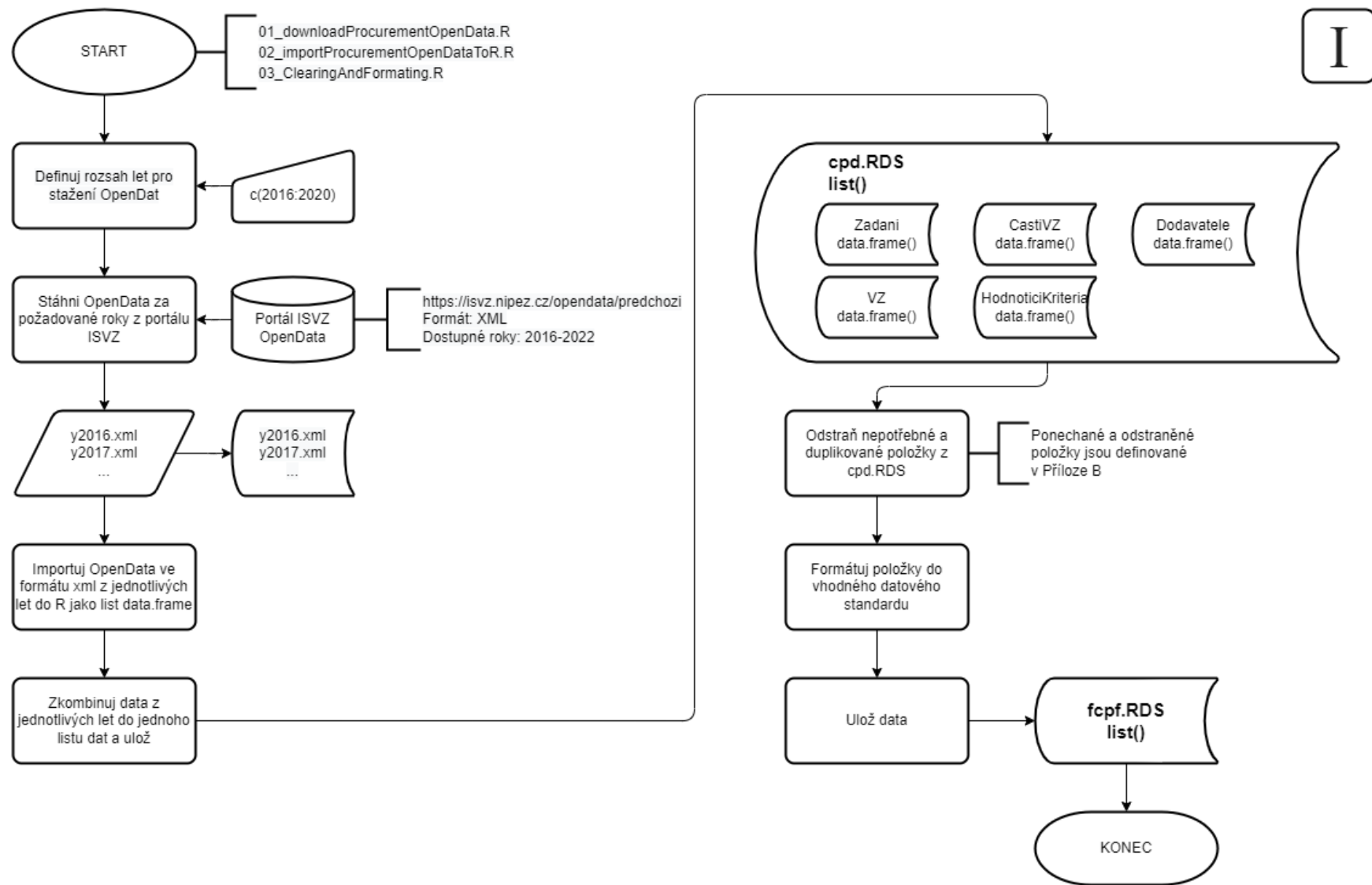
Položky a popis byl převzat z nápovědy k OpenData z ISVZ [25]. Popis byl dále upraven autorem. V případě vyřazených položek je popis vytvořen autorem.

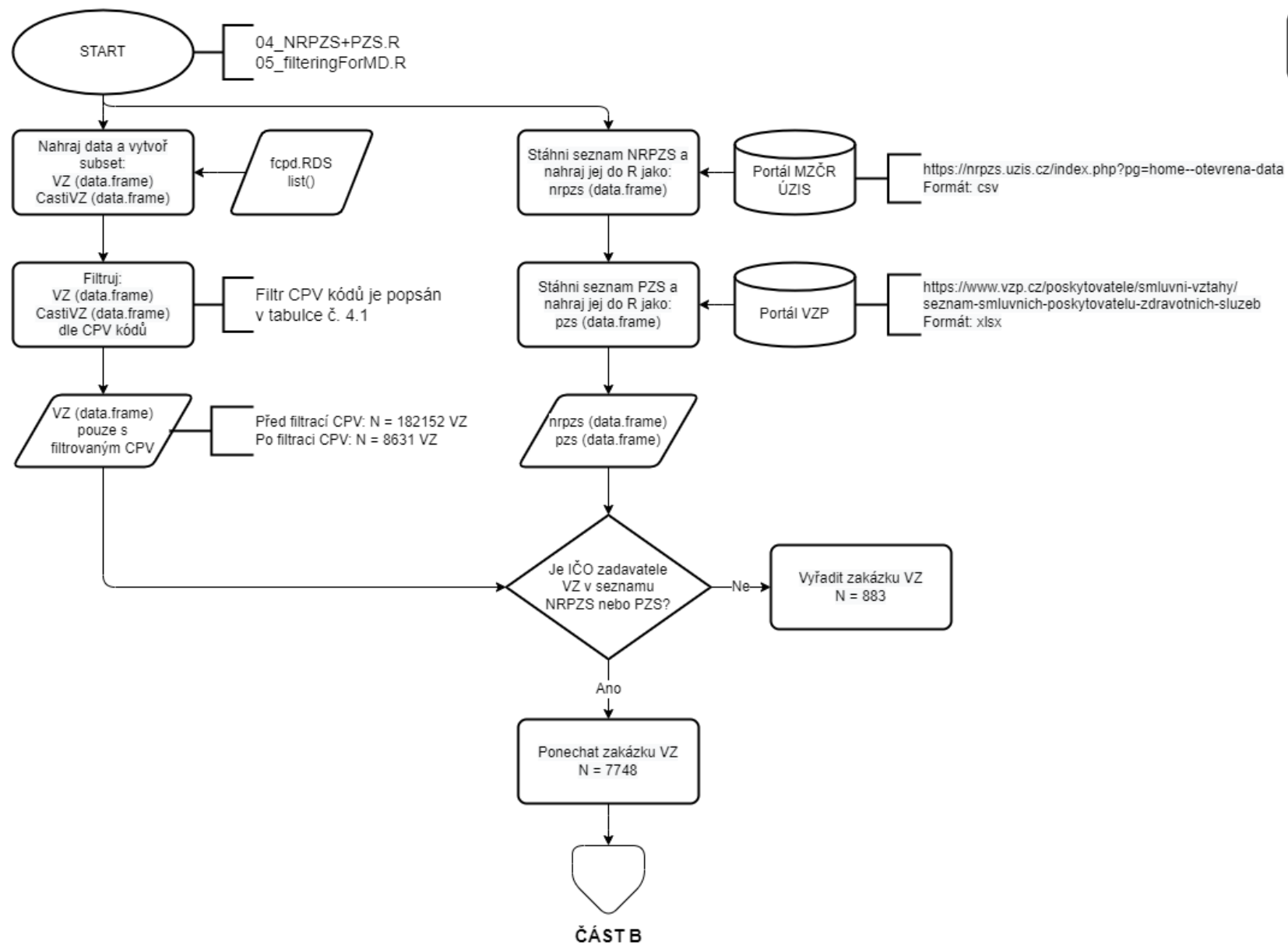
Příloha C: Vývojové diagramy zpracování dat

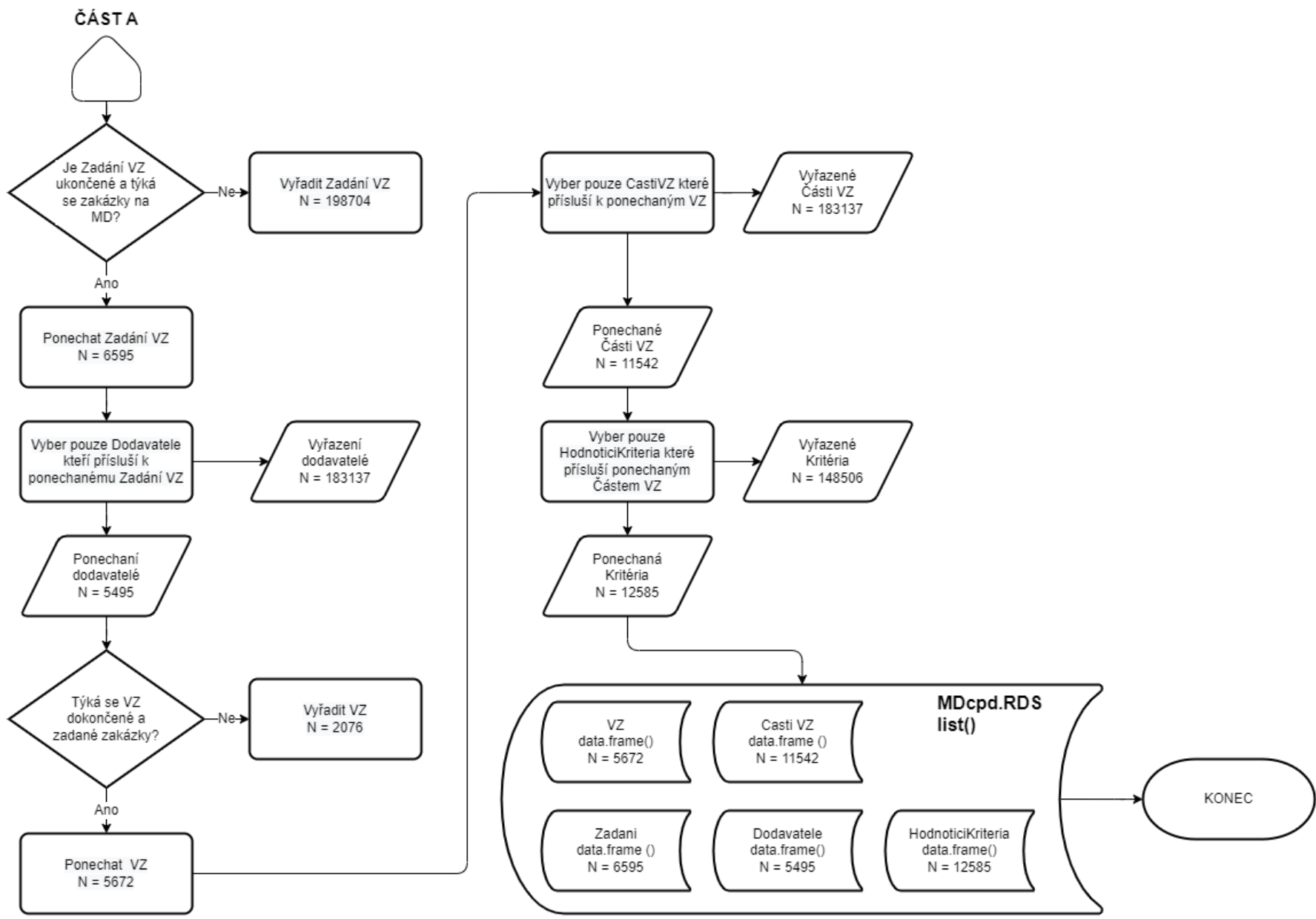
Z důvodu zachování čitelnosti má každý vývojový diagram vyčleněnou svou vlastní stránku. Označení a popis obrázků vývojových diagramů s komentářem je v tabulce C.1. Vývojové diagramy jsou označeny římskými číslicemi v jejich pravém horním rohu.

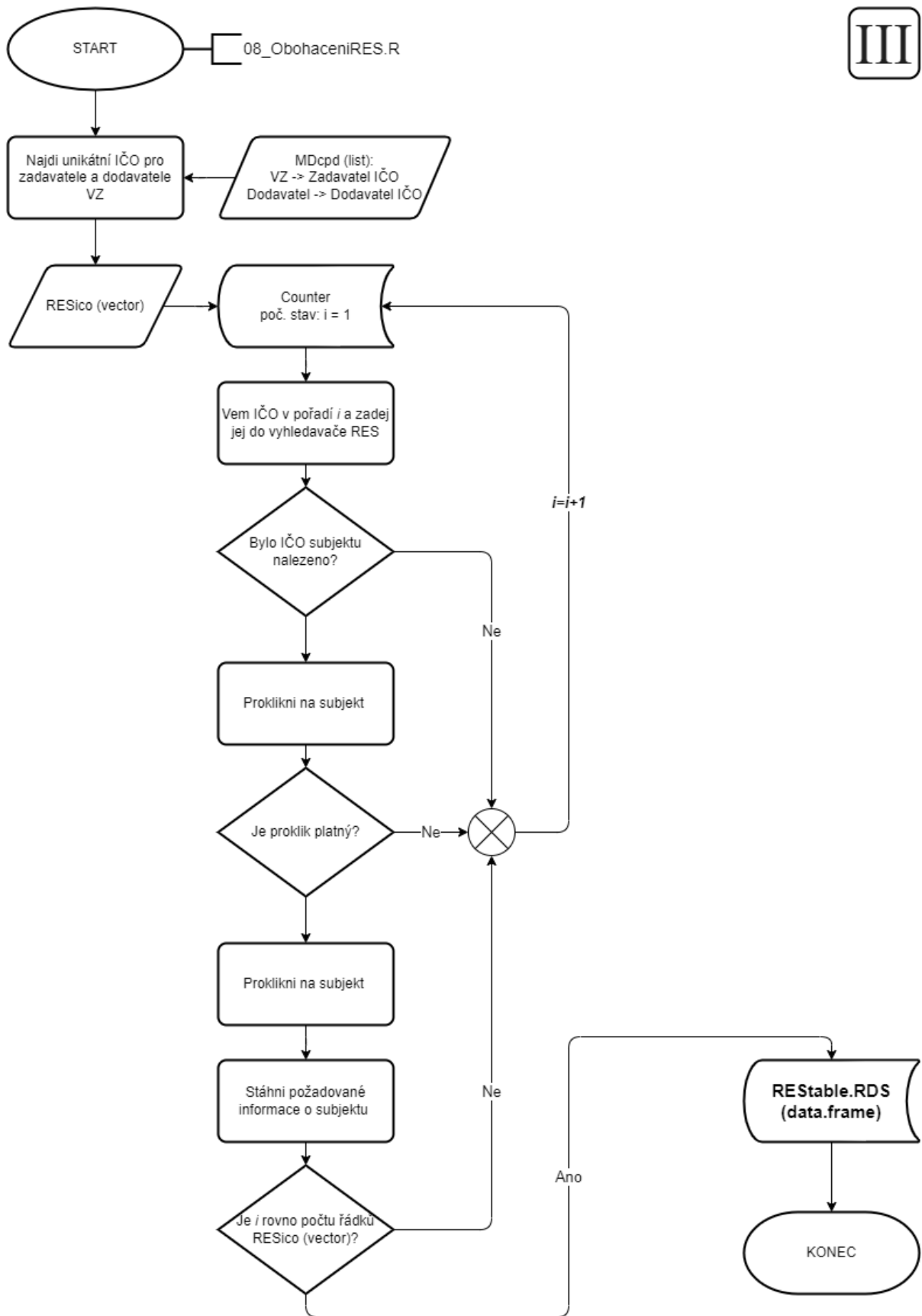
Tabulka C.1: Označení a popis obrázků vývojových diagramů s komentářem

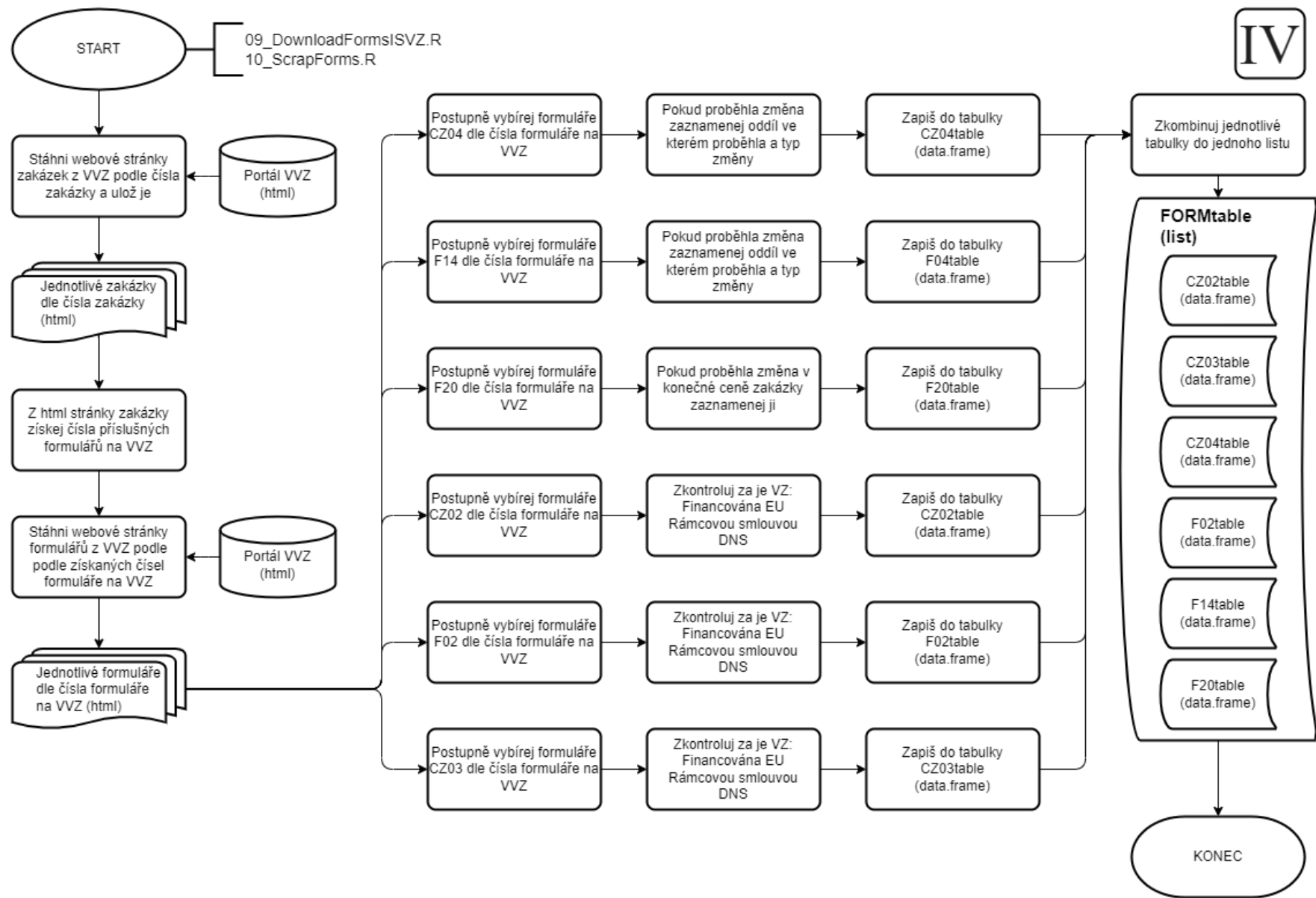
Označení	Proces	Komentář
I	Import a první zpracování Open dat z ISVZ	Import dat z XML do list()
II	Filtrování na VZ na zdravotnickou techniku	CPV filtr, dokončené zakázky, filtrace dle NRPZS a PZS
III	Stažení dat o zadavatelích a dodavatelích z databáze RES	webscrapping z webu RES
IV	Stažení dat o VZ z formulářů z VVZ	webscrapping z formulářů z VVZ
V	Agregace tabulek z listu MDcpd do sloučené tabulky	Výsledná tabulka je dále upravena a použita k analýzám
VI	Metoda NER – parsing knihoven pro jednotlivá zadání a kategorie KZT	
VII	Metoda NER – fuzzy-matching jednotlivých zadání vůči kategoriím KZT	Vyhodnocení pravděpodobnosti shody knihoven a určení kategorie.

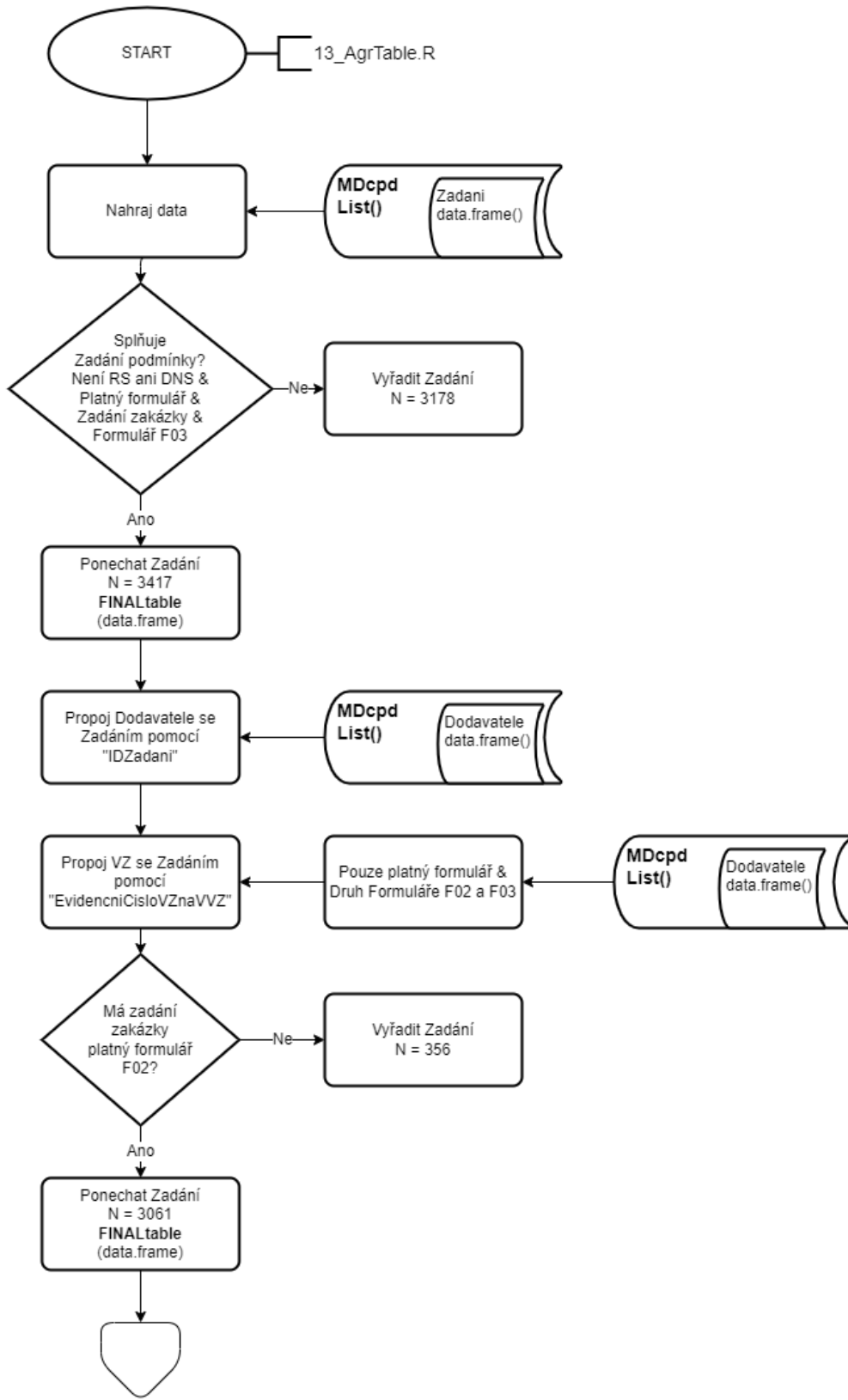




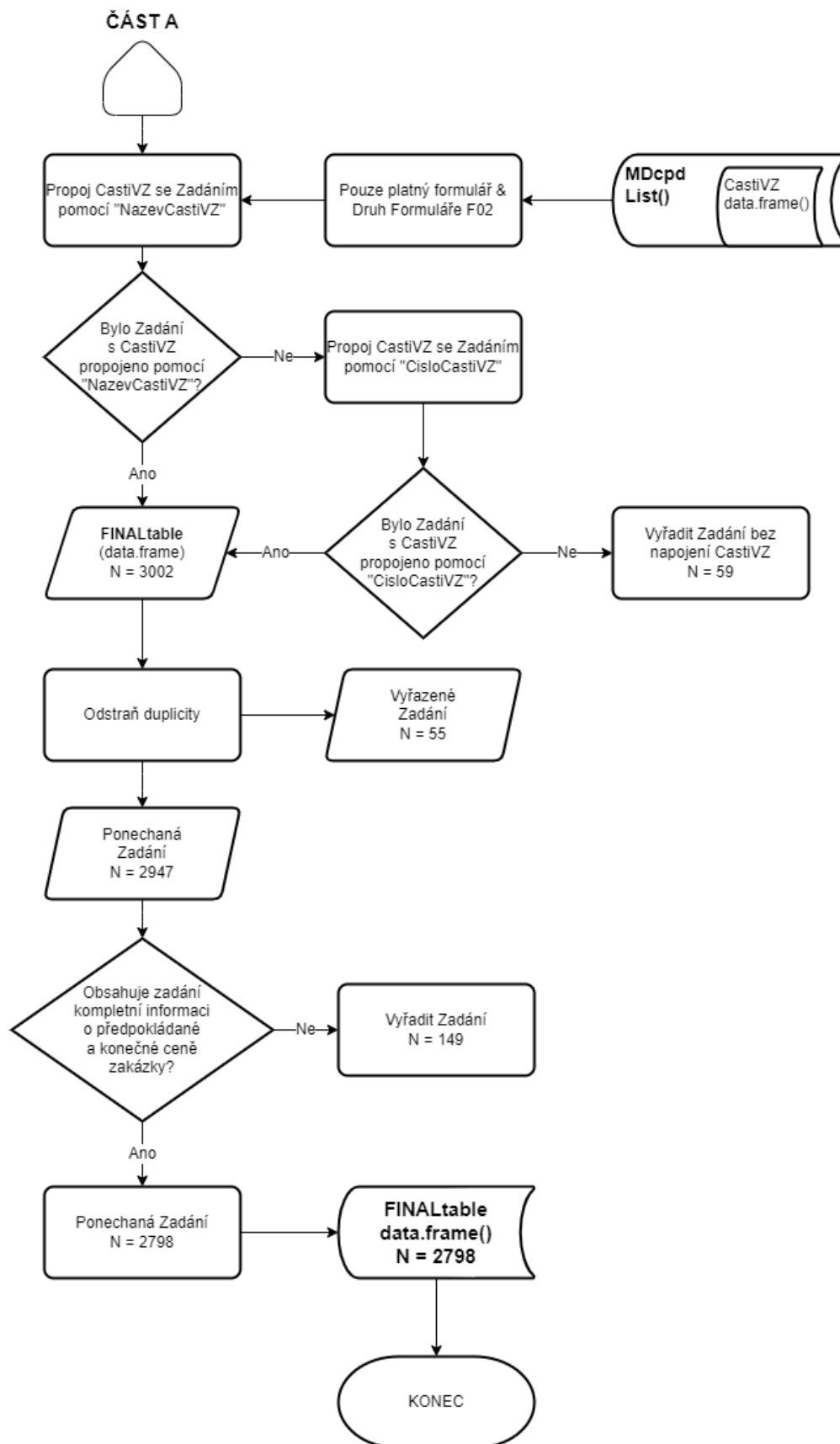


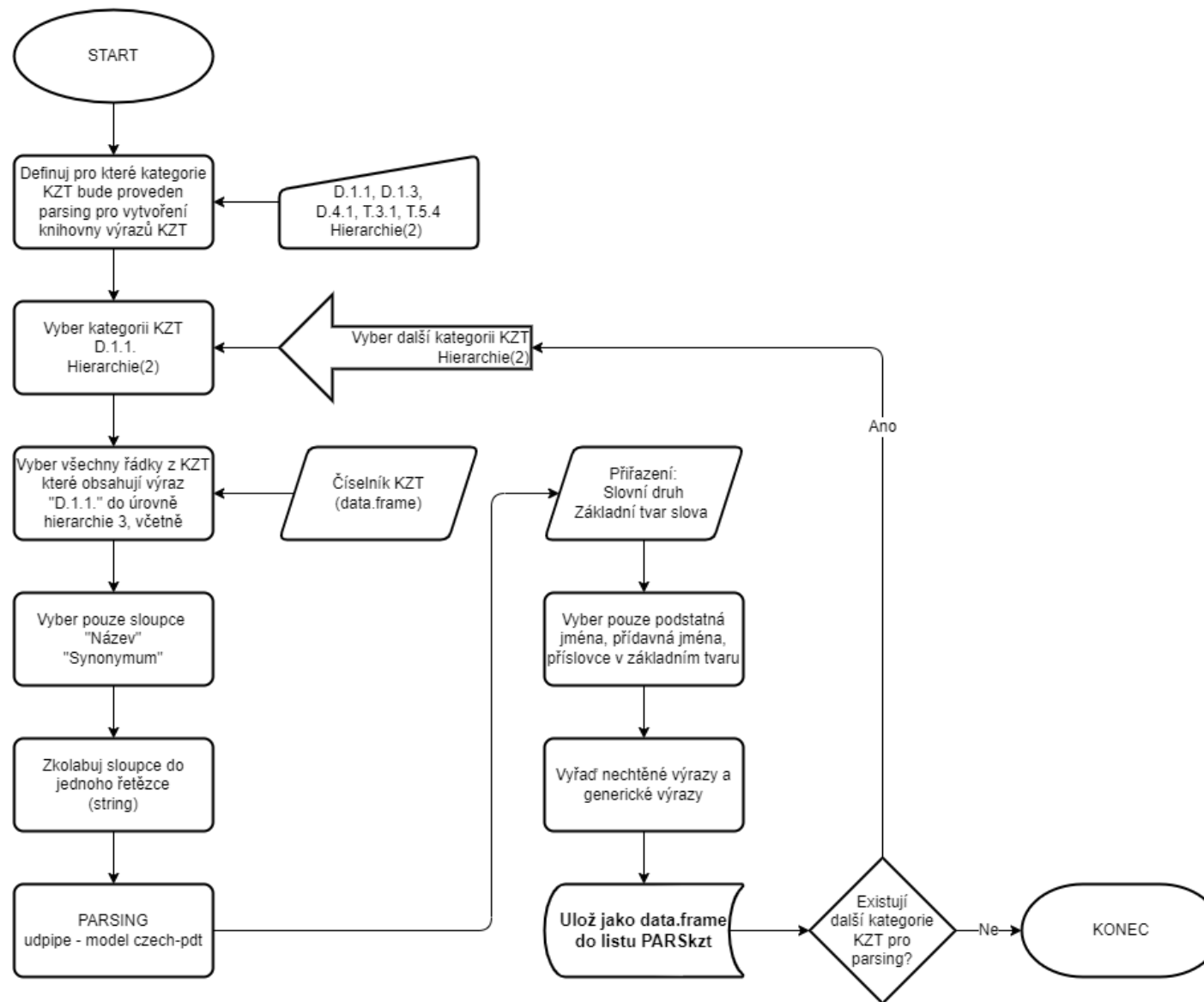


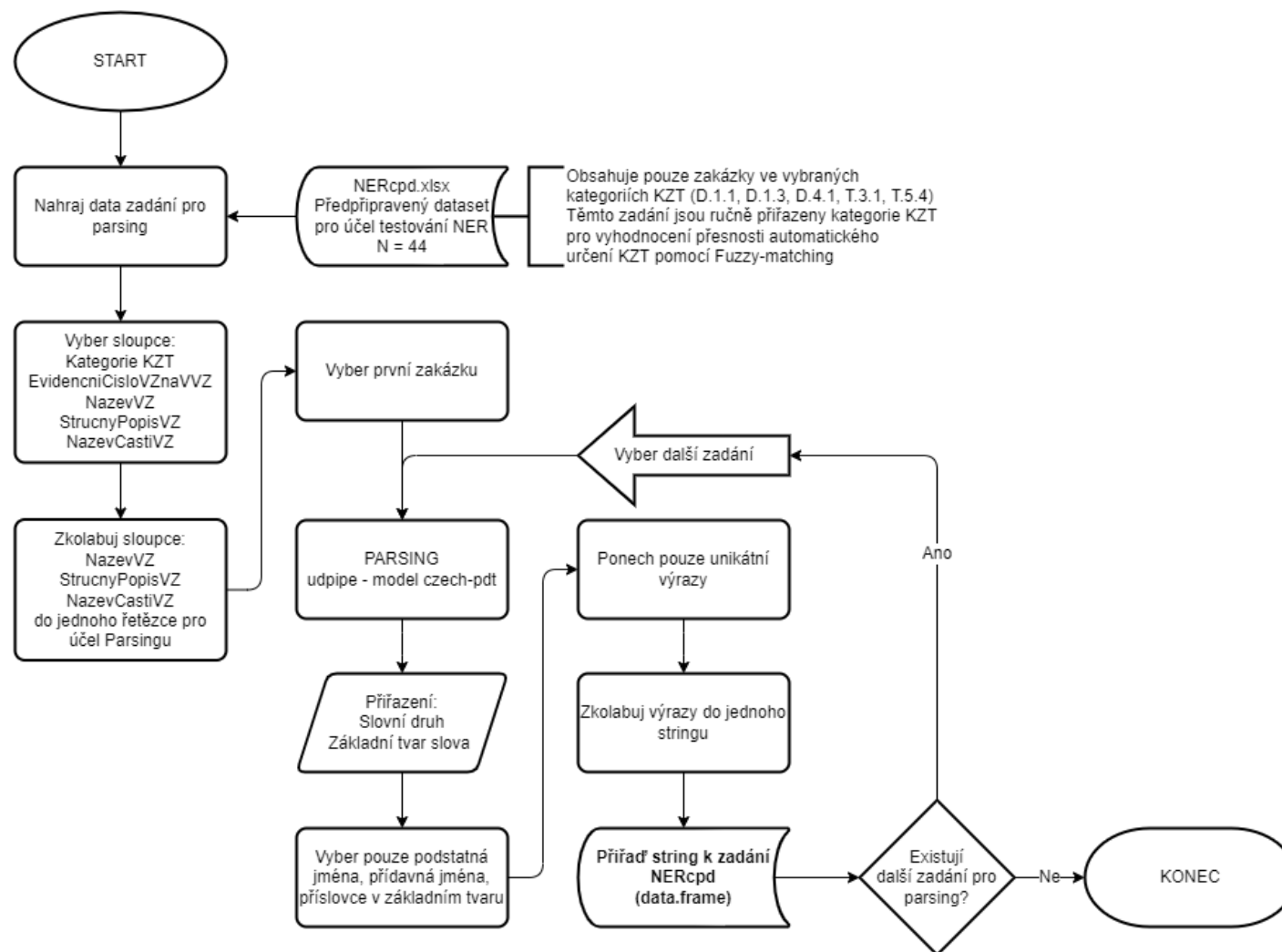


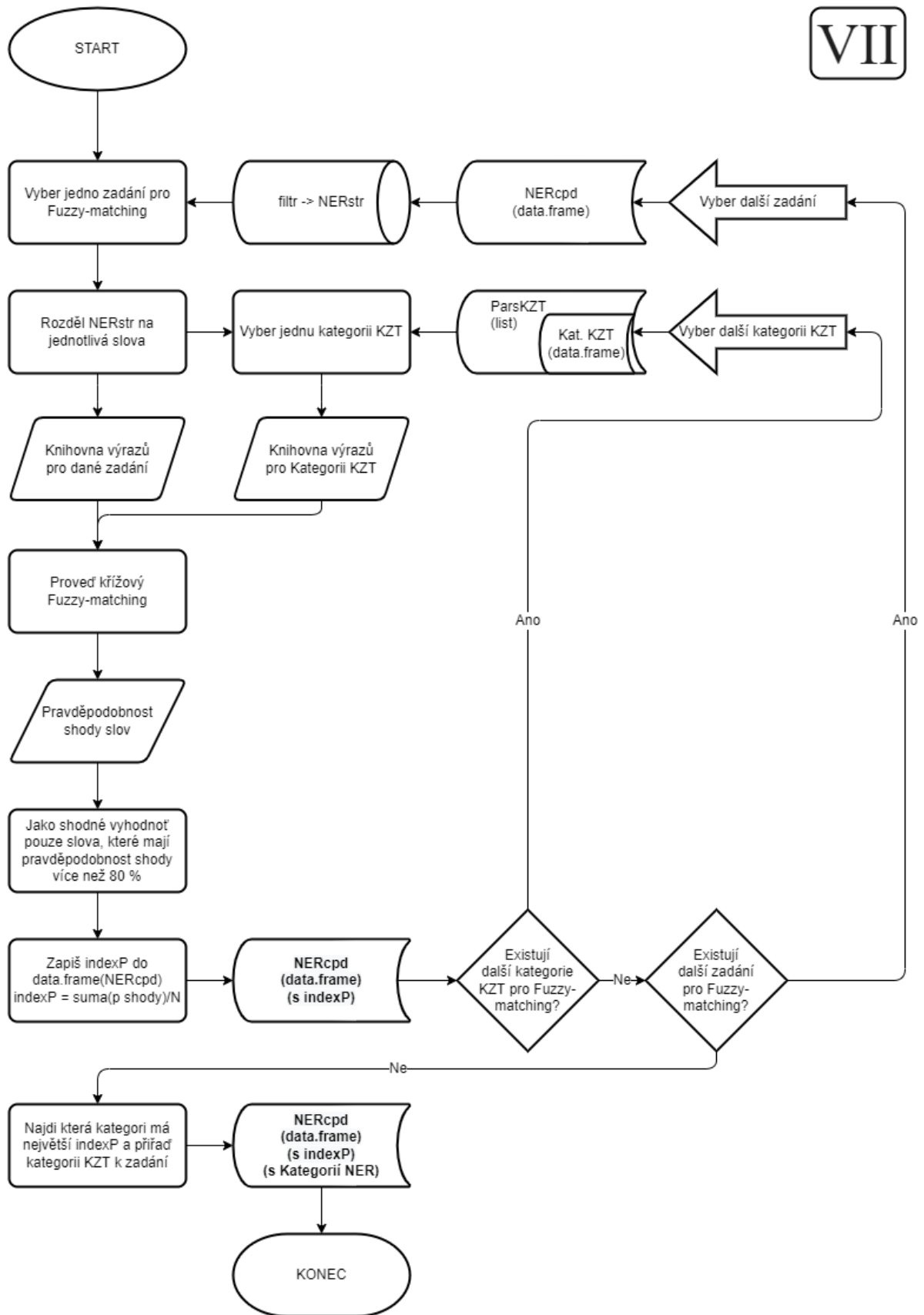


ČÁST B









Příloha D: Codebook k datovému setu „FNtb“

Codebook: Data.frame(FNtb)

Tabulka FNtb je souhrn agregovaných dat pro potřebu regresní analýzy veřejných zakázek na zdravotnickou techniku. V tomto Codebooku je možné nalézt informaci o struktuře a původu dat tohoto data framu.

FNtb

- *Datový typ:* data.frame()
- *Počet sloupců:* 20
- *Počet řádků:* 2710

Tabulka 1.D: Sloupce v tabulce FNtb

Pořadí	Název sloupce
1	ID
2	Zadavtel_ICO
3	Uspora
4	Rok_Vypsani
5	Rok_Zadani
6	Forma_Kod_Zadavatel
7	Velikost_Kos_Zadavatel
8	CPV
9	Druh_Rizeni
10	Pocet_Nabidek
11	Pocet_Kriterii
12	Fin_EU
13	Nabidky_Lhuta
14	Dodavatel_ICO
15	NUTS_Plneni
16	Forma_Kod_Dodavatel
17	Velikost_Kod_Dodavatel
18	Doba_Zakazka
19	Doba_Vyber
20	KZT

1. ID

- *Popis:* Unikátní identifikátor „Zadání“ veřejné zakázky. Zároveň unikátní identifikátor v tomto data setu.
- *Zdroj:* list(FINALtable) \$ IDZadani

2. Zadavatel_ICO

- *Popis:* IČO subjektu zadávajícího veřejnou zakázku
- *Zdroj:* list(FINALtable) \$ ZadavatelICO

3. Uspora

- *Popis:* Závislá proměnná pro účel regresní analýzy
- *Zdroj:* Výpočet viz vzorec D.1
- *Jednotka:* Uspora [%]
- *Rozsah:* (-918,18 % – 99,97 %)
- *Průměr:* 5,78 %
- *Medián:* 2,01 %
- *Směrodatná odchylka:* 43,12 %

Vzorec pro výpočet úspory:

$$\text{Uspora} = \frac{\text{konečná cena} - \text{předpokládaná cena}}{\text{předpokládaná cena}} \quad (\text{D.1})$$

Pokud je předpokládaná a konečná cena v jiné měně než CZK, je před výpočtem proveden převod měn dle kurzu platného k 10. 4. 2022. Kurz je nutné nastavit ručně pro každou cizí měnu, která se v data setu vyskytne ve skriptu 14_FinTable.R. Zdroje a popis proměnných ze vzorce D.1 jsou popsány v tabulce 2.D.

Tabulka 2.D: Popis a zdroje pro proměnné předpokládaná cena a konečná cena ze vzorce D.1

Předpokládaná cena	
Popis	Předpokládaná hodnota veřejné zakázky určena zadavatelem na základě kvalifikovaného odhadu nebo průzkumu trhu před vypsáním veřejné zakázky
Zdroj	<i>VZ je na části:</i> list(MDcpd) \$ data.frame(CastiVZ) \$ CelkovaPredpokladanaHodnotaCastiVZ <i>VZ není na části:</i> list(MDcpd) \$ data.frame(VZ) \$ OdhadovanaHodnotaVZbezDPH
Konečná cena	
Popis	Konečná hodnota veřejné zakázky, kterou nabídl dodavatel a se kterou uspěl ve výběrovém řízení
Zdroj	list(MDcpd) \$ data.frame(Zadani) \$ CelkovaKonecnaHodnotaVZzaZadani

4. Rok_Vypsani

- *Popis:* Rok, kdy bylo vypsáno zadání zakázky
- *Zdroj:* list(MDcpd) \$ data.frame(VZ) \$ DatumUverejneni

5. Rok_Zadani

- *Popis:* Rok zadání zakázky určenému dodavateli
- *Zdroj:* list(MDcpd) \$ data.frame(VZ) \$ DatumZadaniVZ

6. Forma_Kod_Zadavatel

- *Popis:* Kód právnické formy zadavatele dle RES
- *Zdroj:* data.frame(RESable) \$ pravnickaFormaKod

Tabulka 3.D: Klíč ke kódům právnické formy dle RES a četnost výskytu v data setu.

Kód	Text zkrácený	Text	Četnost
112	Společnost s r.o.	Společnost s ručením omezeným	25
121	Akciová společnost	Akciová společnost	799
325	Org. složka státu	Organizační složka státu	168
331	Příspěvková organizace	Příspěvková organizace	1546
601	Vysoká škola (veřejná, státní)	Vysoká škola (veřejná, státní)	172

7. Velikost_Kod_Zadavatel

- *Popis:* Kód kategorie velikosti zadavatele dle počtu zaměstnanců dle RES
- *Zdroj:* data.frame(RESable) \$ velikostKod

Tabulka 4.D: Klíč ke kódům velikosti zadavatele dle počtu zaměstnanců dle RES a četnost výskytu v data setu.

Kód	Text zkrácený	Text	Četnost
110	Bez zaměstnanců	Bez zaměstnanců	0
120	1–5 zaměstnanců	1–5 zaměstnanců	0
130	6–9 zaměstnanců	6–9 zaměstnanců	0
210	10–19 zaměstnanců	10–19 zaměstnanců	1
220	20–24 zaměstnanci	20–24 zaměstnanci	0
230	25–49 zaměstnanců	25–49 zaměstnanců	1
240	50–99 zaměstnanců	50–99 zaměstnanců	5
310	100–199 zaměstnanců	100–199 zaměstnanců	5
320	200–249 zaměstnanců	200–249 zaměstnanců	0
330	250–499 zaměstnanců	250–499 zaměstnanců	66
340	500–999 zaměstnanců	500–999 zaměstnanců	217
410	1000–1499 zaměstnanců	1000–1499 zaměstnanců	623
420	1500–1999 zaměstnanců	1500–1999 zaměstnanců	93
430	2000–2499 zaměstnanců	2000–2499 zaměstnanců	210
440	2500–2999 zaměstnanců	2500–2999 zaměstnanců	131
450	3000–3999 zaměstnanců	3000–3999 zaměstnanců	260

460	4000–4999 zaměstnanců	4000–4999 zaměstnanců	175
470	5000–9999 zaměstnanců	5000–9999 zaměstnanců	699
510	10 000 a více zaměstnanců	10 000 a více zaměstnanců	224

8. CPV

- *Popis:* Řetězec CPV kódů přiřazených k danému zadání veřejné zakázky. Jednotlivé CPV kódy jsou v řetězci odděleny mezerou.
- *Zdroj:* `list(MDcpd) $ data.frame(VZ) & CPVhlavni list(MDcpd) $ data.frame(CastiVZ) $ CPVkod`

Aktuální číselník CPV kódů je k dispozici na webových stránkách Evropské unie, věnovaných veřejným zakázkám (*URL: <https://simap.ted.europa.eu/cpv>*)

9. Druh_Rizeni

- *Popis:* Druh řízení, ve kterém je veřejná zakázka soutěžena
- *Zdroj:* `list(MDcpd) $ data.frame(VZ) $ DruhRizeni`

Tabulka 5.D: Souhrn položek pro sloupec Druh_rizeni a četnost výskytu jednotlivých proměnných

Druh_rizeni	Četnost
Otevřené řízení	2708
Užší řízení	2

10.Pocet_Nabidek

- *Popis:* Počet nabídek, které byly podány při soutěžení daného zadání
- *Zdroj:* `list(MDcpd) $ data.frame(Zadani) $ PocetObdrzenychNabidek`

Tabulka 6.D: Souhrn položek pro sloupec Pocet_Nabidek, četnost výskytu jednotlivých proměnných a průměrný počet nabídek na jedno zadání

Počet nabídek	Četnost
1	1622
2	704
3	242
4	95
5	23
6	15
7	6
9	1
12	1
16	1
Průměrný počet nabídek na zadání = 1,631	

11.Pocet_Kriterii

- *Popis:* Počet hodnotících kritérií pro výběr dodavatele, které stanovil zadavatel při vypsání výběrového řízení.
- *Zdroj:* list(MDcpd) \$ data.frame(HodnoticiKriteria) \$ (viz níže)

Počet hodnotících kritérií je získán pomocí funkce, která propojuje jednotlivá zadání s příslušnými kritérii. Následně spočte četnost těchto kritérií a výslednou hodnotu přiřadí k danému zadání. Tento algoritmus je funkcí s názvem *HKfnv* ve skriptu 13_AgrTable.R.

Tabulka 7.D: Souhrn položek pro sloupec Pocet_Kriterii a četnost výskytu jednotlivých proměnných v data setu.

Počet použitých kritérií	Četnost
1	2102
2	204
3	27
4	10
5	9
6	2
7	1
15	1

12.Fin_EU

- *Popis:* Byla zakázka financována za přispění EU? (minimálně 50 % z ceny zakázky)
- *Hodnoty:* TRUE / FALSE
- *Zdroj:* list(FORMtable) \$ data.frame(F02table) \$ financovanoEU
- *Propojení:* Před Evidenční číslo zakázky na VVZ

13.Nabidky_Lhuta

- *Popis:* Jak dlouho mohli účastníci soutěže podávat nabídky?
- *Jednotka:* Nabidky_Lhuta [dny]
- *Zdroj:* Výpočet viz vzorec D.2

Vzorec pro výpočet Nabidky_Lhuta:

$$\text{Nabidky}_{\text{Lhuta}} = \text{Lhůta pro doručení nabídek} - \text{Datum odeslání formuláře} \quad (\text{D.2})$$

Tabulka 8.D: Zdroje pro proměnné Lhůta pro doručení nabídek a Datum odeslání formuláře ze vzorce D.2

Datum odeslání formuláře na VVZ	
Zdroj	list(MDcpd) \$ data.frame(VZ) \$ DatumOdeslaniFormulareNaVVZ
Lhůta pro doručení nabídek	
Zdroj	list(MDcpd) \$ data.frame(VZ) \$ LhutaProDoruceniNabidek

14.Dodavatel_ICO

- *Popis:* IČO subjektu dodavatele pro dané zadání
- *Zdroj:* list(MDcpd) \$ data.frame(Dodavatele) \$ DodavatelICO

15.NUTS_Plneni

- *Popis:* Hlavní místo plnění zakázky vyjádřené kódem CZ-NUTS
- *Zdroj:* list(MDcpd) \$ data.frame(CastiVZ) \$ HlavniMistoPlneniNUTS

Tabulka 9.D: Klíč ke kódům NUTS_Plneni dle číselníku Českého statistického úřadu (CZ-NUTS)

Kód	NUTS 0 stát	NUTS 1 území	NUTS 2 region soudržnosti	NUTS 3 kraj
CZ	Česko			
CZ0		Česko		
CZ01			Praha	
CZ010				Hlavní město Praha
CZ02			Střední Čechy	
CZ020				Středočeský kraj
CZ03			Jihozápad	
CZ031				Jihočeský kraj
CZ032				Plzeňský kraj
CZ04			Severozápad	
CZ041				Karlovarský kraj
CZ042				Ústecký kraj
CZ05			Severovýchod	
CZ051				Liberecký kraj
CZ052				Královéhradecký kraj
CZ053				Pardubický kraj
CZ06			Jihovýchod	
CZ063				Kraj Vysočina
CZ064				Jihomoravský kraj
CZ07			Střední Morava	
CZ071				Olomoucký kraj
CZ072				Zlínský kraj
CZ08			Moravskoslezsko	
CZ080				Moravskoslezský kraj
CZZ		Extra-Regio		
CZZZ			Extra-Regio	
CZZZZ				Extra-Regio

16.Forma_Kod_Dodavatel

- *Popis:* Kód právnické formy dodavatele dle RES
- *Zdroj:* data.frame(RES) \$ pravnickaFormaKod

Tabulka 10.D: Klíč ke kódům právnické formy dodavatele dle RES a četnost výskytu v data setu.

Kód	Text zkrácený	Text	Četnost
101	Fyz. os. podnikající dle živ. zákona	Fyzická osoba podnikající dle živnostenského zákona	22
111	Veřejná obch.společnost	Veřejná obchodní společnost	1
112	Společnost s r.o.	Společnost s ručením omezeným	2240
121	Akciová společnost	Akciová společnost	306
421	Odšt.záv.zahr.práv.os.	Odštěpný závod zahraniční právnické osoby	5
932	Evropská společnost	Evropská společnost	6

17. Velikost_Kod_Dodavatel

- *Popis:* Kód kategorie velikosti dodavatele dle počtu zaměstnanců dle RES
- *Zdroj:* data.frame(RESdata) \$ velikostKod

Tabulka č. 11.D: Klíč ke kódům velikosti dodavatele dle počtu zaměstnanců dle RES a četnost výskytu v data setu.

Kód	Text zkrácený	Text	Četnost
110	Bez zaměstnanců	Bez zaměstnanců	12
120	1–5 zaměstnanců	1–5 zaměstnanců	212
130	6–9 zaměstnanců	6–9 zaměstnanců	195
210	10–19 zaměstnanců	10–19 zaměstnanců	457
220	20–24 zaměstnanci	20–24 zaměstnanci	217
230	25–49 zaměstnanců	25–49 zaměstnanců	638
240	50–99 zaměstnanců	50–99 zaměstnanců	281
310	100–199 zaměstnanců	100–199 zaměstnanců	219
320	200–249 zaměstnanců	200–249 zaměstnanců	99
330	250–499 zaměstnanců	250–499 zaměstnanců	122
410	1000–1499 zaměstnanců	1000–1499 zaměstnanců	65
440	2500–2999 zaměstnanců	2500–2999 zaměstnanců	5
510	10 000 a více zaměstnanců	10 000 a více zaměstnanců	1

18.Doba_Zakazka

- *Popis:* Doba od vypsání veřejné zakázky po zadání zakázky dodavateli
- *Jednotka:* Doba_Zakzka [dny]
- *Zdroj:* Výpočet viz vzorec D.3

Vzorec pro výpočet Doba_Zakzka:

$$Doba_{zakazka} = Datum\ zadání\ VZ - Datum\ odeslání\ formuláře \quad (D.3)$$

Tabulka 12.D: Zdroje pro proměnné Datum zadání VZ a Datum odeslání formuláře ze vzorce D.3

Datum odeslání formuláře na VVZ	
Zdroj	list(MDcpd) \$ data.frame(VZ) \$ DatumOdeslaniFormulareNaVVZ
Datum zadání VZ	
Zdroj	list(MDcpd) \$ data.frame(Zadani) \$ DatumZadaniVZ

19.Doba_Vyberu

- *Popis:* Doba od uzavření možnosti podávat nabídky po zadání VZ
- *Jednotka:* Doba_Vyberu [dny]
- *Zdroj:* Výpočet viz vzorec D.4

Vzorec pro výpočet Doba_Vyberu:

$$Doba_{Vyberu} = Datum\ zadání\ VZ - Lhůta\ pro\ doručení\ nabídek \quad (D.4)$$

Tabulka 13.D: Zdroje pro proměnné Datum zadání VZ a Lhůta pro doručení nabídek ze vzorce D.4

Lhůta pro doručení nabídek	
Zdroj	list(MDcpd) \$ data.frame(VZ) \$ DatumOdeslaniFormulareNaVVZ
Datum zadání VZ	
Zdroj	list(MDcpd) \$ data.frame(Zadani) \$ DatumZadaniVZ

20.KZT

- *Popis:* Kategorizace zdravotnické techniky, která je předmětem zakázky pomocí CPV kódu a číselníku KZT
- *Zdroj:* Kategorizace zdravotnické techniky (ÚZIS ČR)

Přiřazení kategorie KZT k zadání probíhá pomocí algoritmu ve skriptu 14_FinTable.R dle klíče, který je popsán v tabulce 14.D. Do algoritmu byly vybrány kategorie KZT, které mají relativně jasný protějšek v CPV kódu. Touto metodou jsou tedy kategorizované pouze vybrané zakázky s vybraným CPV kódem.

Tabulka 14.D: Návaznost CPV kódů ze zadání veřejné zakázky na kategorie v KZT

Kategorie	Kategorie KZT	CPV název	CPV kód
D.1.1.4.	Výpočetní tomografy	Tomografické přístroje	33115000
D.1.1.4.	Výpočetní tomografy	CT skener	33115100-0
D.1.1.4.	Výpočetní tomografy	CAT skener	33115200-1
D.1.1.5.	Rentgenové přístroje mamografické	Mamografické přístroje	33111650-2
D.1.2.	Magneticko-rezonanční zobrazovací systémy	Jednotka magnetické resonance	33111610-0

D.1.2.	Magneticko-rezonanční zobrazovací systémy	Zobrazovací zařízení magnetické resonance	33113000
D.1.2.	Magneticko-rezonanční zobrazovací systémy	Skenery pracující na základě magnetické resonance	33113100-6
D.1.2.	Magneticko-rezonanční zobrazovací systémy	Skenery pracující na základě nukleární magnetické resonance	33113110-9
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Echokardiografické, ultrazvukové a Dopplerovy zobrazovací přístroje	33112000
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Ultrazvukový snímač srdce	33112100-9
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Ultrazvukové jednotky	33112200-0
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Ultrazvukové skenery	33112300-1
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Dopplerův ultrazvuk	33112310-4
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Dopplerovo zařízení	33112320-7
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Echoencefalograf	33112330-0
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	Echokardiograf	33112340-3
D.1.4.1.1.	Gama kamery	Gama kamery	33111620-3
D.2.1.1.	Elektrokardiografy	Elektrokardiogram	33121500-9
D.2.1.1.	Elektrokardiografy	elektrokardiografické přístroje	33123200-0
D.4.1.	Monitory vitálních funkcí	Systémy pro kontrolu pacientů	33195000-3
T.3.2.3.	Defibrilátory	Defibrilátor	33182100-0
T.5.3.1.	Operační svítidla	Chirurgická světla	33167000
T.5.4.	Anesteziologické přístroje	Anesteziologické přístroje	33172100-7
T.5.6.1.	Generátory elektrochirurgické	Elektrochirurgická jednotka	33161000

Tabulka 15.D: Četnost jednotlivých kategorií KZT v data setu

Kategorie	Kategorie KZT	Četnost
D.1.1.4.	Výpočetní tomografie	31
D.1.1.5.	Rentgenové přístroje mamografické	8
D.1.2.	Magneticko-rezonanční zobrazovací systémy	31
D.1.3.	Ultrazvukové měřicí/zobrazovací systémy	139
D.1.4.1.1.	Gama kamery	6
D.2.1.1.	Elektrokardiografy	22
D.4.1.	Monitory vitálních funkcí	43
T.3.2.3.	Defibrilátory	74
T.5.3.1.	Operační svítidla	20
T.5.4.	Anesteziologické přístroje	48
T.5.6.1.	Generátory elektrochirurgické	33
Celkem kategorizovaných Zadání: 455/2710		

