



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA ELEKTROTECHNICKÁ

Katedra teorie obvodů

**Vizualizace a zpracování dat uživatelů
závislých na nikotinu**

**Visualization and Data Processing Users
Addicted to Nicotine**

Diplomová práce

Studijní program: Lékařská elektronika a bioinformatika

Specializace: Lékařská technika

Autor diplomové práce: Bc. Kristýna Míková

Vedoucí diplomové práce: doc. Ing. Daniel Novák, Ph.D.

Praha 2021

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Míková** Jméno: **Kristýna** Osobní číslo: **465471**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra teorie obvodů**
Studijní program: **Lékařská elektronika a bioinformatika**
Specializace: **Lékařská technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Vizualizace a zpracování dat uživatelů závislých na nikotinu

Název diplomové práce anglicky:

Visualization and Data Processing Users Addicted to Nicotine

Pokyny pro vypracování:

Nikotinová závislost je třetí nejtěžší závislostí. Ve spolupráci s 1.LF UK vzniklo webové rozhraní, kde přístup je založen na "evidence based" medicíně. Efektivita léčby byla ověřena klinickou studií. Získaná data jsou velmi užitečná pro další hlubší analýzu chování uživatelů v průběhu terapeutického procesu.

- 1) Seznamte se s problematikou nikotinové závislosti a její léčby.
- 2) Proveďte explorační analýzu a použijte základní NLP metody pro vizualizaci dat.
- 3) Spočítejte digitální fenotypy uživatelů pomocí metod shlukování.

Seznam doporučené literatury:

- [1] Danaher BG, Tyler MS, Crowley RC, Brendryen H, Seeley JR Outcomes and Device Usage for Fully Automated Internet Interventions Designed for a Smartphone or Personal Computer: The MobileQuit Smoking Cessation Randomized Controlled Trial J Med Internet Res 2019;21(6):e13290
- [2] Kulhánek A., Gabrhelík R., Novák D. & Brendren H. (2019). eHealth intervention for smoking cessation for Czech tobacco smokers: Pilot study of user acceptance. Adiktologie, 18(2), 81–85
- [3] Brendryen, H., Drozd, F., & Kraft, P. (2008). A digital smoking cessation program delivered through internet and cell phone without nicotine replacement (happy ending): randomized controlled trial. J Med Internet Res, 10(5), e51. doi:10.2196/jmir.1005

Jméno a pracoviště vedoucí(ho) diplomové práce:

doc. Ing. Daniel Novák, Ph.D., Analýza a interpretace biomedicínských dat FEL

Jméno a pracoviště druhého(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **18.01.2021**

Termín odevzdání diplomové práce: **13.08.2021**

Platnost zadání diplomové práce: **30.09.2022**

doc. Ing. Daniel Novák, Ph.D.
podpis vedoucí(ho) práce

doc. Ing. Radoslav Bortel, Ph.D.
podpis vedoucí(ho) ústavu/katedry

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

III. PŘEVZETÍ ZADÁNÍ

Diplomantka bere na vědomí, že je povinna vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studentky

PROHLÁŠENÍ

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 13. 8. 2021

.....

Kristýna Míková

PODĚKOVÁNÍ

Ráda bych v první řadě poděkovala svému vedoucímu práce doc. Ing. Danielu Novákovi, Ph.D. za profesionální vedení práce, za ochotu, pomoc a jeho čas. Dále bych ráda poděkovala jeho kolegům za pomoc, které se mi dostávalo. Nakonec bych ráda poděkovala své rodině, přátelům a partnerovi za podporu při studiu.

ABSTRAKT

Vizualizace a zpracování dat uživatelů závislých na nikotinu

Kouření a s ním spojená závislost na nikotinu je jeden z vážných problémů společnosti. V České republice pravidelně užívá tabákové výrobky až 25 % populace. Většina kuřáků se pokouší odvyknout samostatně s malou úspěšností. Podporu pacienti nyní mohou nalézt v aplikaci na odvykání kouření, u níž třičtvrtě pacientů potvrdilo, že preferuje tuto virtuální metodu podpory v odvykání před osobními terapiemi s odborníkem. Díky této aplikaci, kterou si stáhlo přes 22 tisíc lidí, vznikla databáze obsahující velké množství anonymních dat od uživatelů, kteří si přejí s kouřením skončit. Data z aplikace byla základními způsoby explorační analýzy zpracována a na otevřené odpovědi byly použity základní NLP metody. Výsledkem je zpracování odpovědi na otázky, které pokládá virtuální terapeut, v závislosti na sociodemografických charakteristikách, které uživatel dobrovolně poskytuje. Kategorie s četnostmi byly mezi sebou porovnány χ^2 testem nezávislosti na hladině významnosti $\alpha = 0,05$. Pro lepší pochopení uživatelů nikotinu a cílenější prevenci závislosti – současně s rozvíjející se oblastí digitálního phenotypingu, byly na základě nasbíraných údajů spočítány digitální fenotypy uživatelů závislých na nikotinu shlukovacím algoritmem k -means. K nalezení rozdílů mezi shluky – fenotypy byly použity testy středních hodnot – jednosměrná ANOVA a Kruskalův–Wallisův test. Pro příznaky se zamítnutou nulovou hypotézou na hladině významnosti $\alpha = 0,05$ byly použity příslušné post-hoc testy. Variabilita ve shlucích při opakované inicializaci shlukovacího algoritmu a nízké průměrné silhouette skóre (0,14) potvrzují vysoký překryv shluků. Výsledné tři fenotypy jsou tedy jednou z variant výsledku – dominantní vlastností fenotypu 0 je ženské pohlaví, fenotyp 1 je reprezentován muži a fenotyp 2 jsou převážně mladí lidé.

Klíčová slova

Závislost na nikotinu, kouření, explorační analýza, shlukovací analýza, digitální fenotyp

ABSTRACT

Visualization and Data Processing Users Addicted to Nicotine

Smoking and the associated addiction to nicotine is one of the serious problems of society. In the Czech Republic, up to 25% of the population regularly uses tobacco products. Most smokers try to quit on their own with little success. Patients can now find support in a smoking cessation application, in which three-quarters of patients have confirmed that they prefer this virtual method of cessation support over personal therapies with a specialist. Thanks to the application, which was downloaded by over 22,000 people, a database was created containing a large amount of anonymous data from users who wish to quit smoking. Data from the application were processed by basic methods of exploratory analysis and basic NLP methods were used for open responses. The result is the processing of answers to the questions asked by the virtual therapist, depending on the socio-demographic characteristics that the user voluntarily provides. Categories with frequencies were compared with each other by χ^2 test of independence at the significance level $\alpha = 0.05$. For a better understanding of nicotine users and more targeted addiction prevention - along with the evolving field of digital phenotyping, digital phenotypes of nicotine-dependent users were calculated using the *k*-means clustering algorithm based on the data collected. Mean value tests - one-way ANOVA and Kruskal – Wallis test were used to find differences between clusters - phenotypes. Appropriate post-hoc tests were used for symptoms with a rejected null hypothesis at a significance level of $\alpha = 0.05$. The variability in the clusters at repeated initialization of the clustering algorithm and the low average silhouette score (0.14) confirms the high overlap of the clusters. The resulting three phenotypes are therefore one of the variants of the result - the dominant characteristic of phenotype 0 is the female sex, phenotype 1 is represented by men, and phenotype 2 is mostly young people.

Keywords

Nicotine addiction, smoking, exploratory analysis, cluster analysis, digital phenotype

Obsah

Seznam tabulek	9
Seznam obrázků	11
1 Úvod	12
1.1 Přehled současného stavu	12
1.1.1 Problematika závislosti na nikotinu v ČR	12
1.1.2 mHealth aplikace	20
1.1.3 Explorační analýza	22
1.1.4 Digitální fenotyp	23
1.2 Motivace	26
1.3 Cíle práce	26
2 Metody	27
2.1 Příprava	27
2.1.1 Seznámení se s aplikací	27
2.1.2 Zdroj dat	27
2.1.3 Jupyter Notebook	31
2.1.4 Výběr a získání dat	32
2.2 Zpracování dat pro explorační analýzu	33
2.2.1 Hlavní tabulka	35
2.2.2 Osobní profil	35
2.2.3 Odpovědi terapie	36
2.3 Shlukovací analýza	37
2.3.1 Příprava	37
2.3.2 Výběr parametrů	40
2.3.3 Redukce dimenze	41
2.3.4 Shlukovací analýza	41
2.3.5 Analýza fenotypů	41
3 Výsledky a diskuse	44
3.1 Explorační analýza	44
3.1.1 Hlavní tabulka	44
3.1.2 Osobní profil	46

3.1.3	Odpovědi terapie	52
3.1.4	Korelace mezi příznaky	60
3.1.5	Shrnutí a diskuse.....	61
3.2	Digitální fenotypy	62
3.2.1	Výběr parametrů pro shlukování	62
3.2.2	PCA	63
3.2.3	Shlukovací analýza	64
3.2.4	Analýza charakteristik fenotypů.....	66
3.2.5	Shrnutí a diskuse.....	75
4	Závěr	76
	Seznam použité literatury	78
	Příloha A: Tabulky a vizualizace	82
	Osobní profil.....	82
	Odpovědi terapie.....	121

Seznam tabulek

Tabulka 1.1 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – z celku všech respondentů [1]	14
Tabulka 1.2 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – ze souboru současných kuřáků [1].....	15
Tabulka 1.3 Věkové rozmezí, kdy začali současní a bývalí kuřáci pravidelně kouřit [1]	16
Tabulka 1.4 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – ze souboru současných kuřáků [1].....	17
Tabulka 1.5 Počet cigaret průměrně vykouřených denními kuřáky za den podle vybraných demografických charakteristik [1].....	18
Tabulka 1.6 Současní uživatelé elektronické cigarety – elektronické cigarety s nikotinem a bez nikotinu [1].....	19
Tabulka 1.7 Přehled forem konzumace tabákových výrobků a užívání elektronických cigaret (EC) u denních konzumentů (uživatelů), podle vybraných demografických charakteristik [1].....	20
Tabulka 2.1 Použité zkratky fází terapie	27
Tabulka 2.2 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 1	29
Tabulka 2.3 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 2.....	30
Tabulka 2.4 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 3.....	30
Tabulka 2.5 Společný seznam kategorií.....	34
Tabulka 2.6 Pracovní zástupné názvy pro parametry.....	38
Tabulka 3.1 Přehled četností v kategoriích podle fáze.....	45
Tabulka 3.2 Přehled procentuálního zastoupení v kategoriích podle fáze	45
Tabulka 3.3 P-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií	46
Tabulka 3.4 Korelace vyšší než 0,3 nebo nižší než -0,3 mezi příznaky.....	61
Tabulka 3.5 Vybrané parametry a jejich četnosti ve výběrech algoritmů pro výběr parametrů.....	63
Tabulka 3.6 Průměrné hodnoty pro každou kategorii uvnitř shluků.....	69

Tabulka 3.7 Tabulka výsledných p-hodnot z jednosměrného ANOVA testu mezi shluky	70
Tabulka 3.8 Tabulka výsledných p-hodnot z Kruskalova-Wallisova testu mezi shluky .	70
Tabulka 3.9 Pořadí parametrů od nejmenších p-hodnot po největší, výsledky Tukeyho testu a Dunnova testu mezi shluky	71
Tabulka 3.10 Dominantní vlastnosti fenotypu 0	72
Tabulka 3.11 Dominantní vlastnosti fenotypu 1	73
Tabulka 3.12 Dominantní vlastnosti fenotypu 2	74

Seznam obrázků

Obrázek 2.1 Ukázka prostředí Jupyter Notebook	32
Obrázek 2.2 Prostředí Jupyter Notebook se skrytým kódem.....	32
Obrázek 2.3 Karta záložek.....	36
Obrázek 3.1 Sloupcový graf procentuálního rozložení věku začátku kouření	46
Obrázek 3.2 Sloupcový graf pro procentuální rozložení u vyzkoušených produktů.....	47
Obrázek 3.3 Sloupcový graf procentuálního rozložení počtu pokusů	48
Obrázek 3.4 Sloupcový graf procentuálního rozložení u zaměstnání	50
Obrázek 3.5 Sloupcový graf procentuálního rozložení u vzdělání.....	50
Obrázek 3.6 Sloupcový graf procentuálního rozložení u vzdělání.....	51
Obrázek 3.7 Vizualizace odpovědí pomocí wordcloud	53
Obrázek 3.8 Vizualizace odpovědí pomocí wordcloud.....	54
Obrázek 3.9 Vizualizace odpovědí pomocí wordcloud	55
Obrázek 3.10 Vizualizace odpovědí pomocí wordcloud	56
Obrázek 3.11 Vizualizace odpovědí pomocí wordcloud	56
Obrázek 3.12 Vizualizace odpovědí pomocí wordcloud	57
Obrázek 3.13 Vizualizace odpovědí pomocí wordcloud	58
Obrázek 3.14 Vizualizace odpovědí pomocí wordcloud	59
Obrázek 3.15 Vizualizace odpovědí pomocí wordcloud	60
Obrázek 3.16 Rozptyl vyjádřený hlavními komponentami	64
Obrázek 3.17 Graf závislosti SSE na počtu shluků	64
Obrázek 3.18 Vizualizace první a druhé hlavní komponenty v prostoru, obarvení na základě přiřazení ke shlukům	65
Obrázek 3.19 Vizualizace shluků pomocí grafického zobrazení párů všech komponent, na diagonále je histogram pro každou komponentu	66
Obrázek 3.20 Vizualizace rozdílů mezi shluky pomocí krabicových grafů, 1. část.....	67
Obrázek 3.21 Vizualizace rozdílů mezi shluky pomocí krabicových grafů, 2. část.....	68

1 Úvod

Podle poslední zprávy z výzkumu o užívání tabáku a alkoholu v České republice z roku 2020 užívá tabákové výrobky 23 % populace [1] a z toho až 70 % dospělých kuřáků si přeje s kouřením přestat. Pacienti mají několik možností, mezi které patří pravidelné intervence od odborníků nebo farmakologická léčba. Většina ale preferuje variantu bez cizí pomoci. Úspěšnost takových pokusů je nízká [2]. V souvislosti s velkým rozvojem moderních technologií v posledních letech a velkým zájmem o mHealth aplikace ze strany uživatelů, vznikla mobilní aplikace, která virtuální terapeutická sezení přináší pacientovi přímo do jeho mobilního zařízení. Tři čtvrtiny uživatelů této aplikace potvrdilo, že preferují spíše tento způsob podpory při odvykání před sezeními s terapeutem [3].

Pro pochopení rozdílného chování nebo například rozdílných reakcí na léčbu se ve zdravotnictví zařazují pacienti do podskupin – fenotypů. Pokud tuto metodu přeneseme do prostředí digitálních technologií, zařazení pacientů do podskupin probíhá na základě interakcí pacienta s vlastním mobilním zařízením. Takzvaným „digital phenotypingem“ jsou pacienti monitorováni skrze senzory v mobilních zařízeních a jejich data jsou okamžitě vyhodnocována na základě znalosti jejich fenotypů. Jedním z problémů, se kterými se tato metoda často potýká, je právě neznalost těchto fenotypů. V oblasti duševního zdraví, kterému se věnuje nejvíce výzkumů v souvislosti s touto metodou, by pak mobilní zařízení samo mohlo predikovat například blížící se záchvat pacienta a upozornit odborníka. Pro sestavení těchto fenotypů lze využít digitální stopy uživatelů, kteří využívají služeb mHealth aplikací. [4] [5] [6]

1.1 Přehled současného stavu

1.1.1 Problematika závislosti na nikotinu v ČR

V České republice je užívání tabákových výrobků jednou z hlavních příčin úmrtnosti a nemocnosti – téměř 20 % úmrtnosti je v souvislosti s kouřením [7]. Více než polovině kuřákům zkracuje život v průměru o 15 let [2]. U kuřáků je větší riziko vzniku kardiovaskulárních, nádorových a respiračních onemocnění. Tato onemocnění pak

nevíce postihují pacienty ve středním a starším věku [7]. Pravidelně kouří až 25 % dospělých lidí. Jiné formy nikotinových výrobků nejsou stále tolik oblíbené jako cigarety. Ve srovnání s evropskými čísly je míra kouření v České republice průměrná. [8]

60–70 % dospělých kuřáků si přeje s kouřením přestat. Většina kuřáků se snaží přestat s kouřením na vlastní pěst, ale míra úspěchu je nízká. Jako vyléčený pacient se považuje takový, který abstinuje 6-12 měsíců s maximálním počtem 5 cigaret za celé období. [2]

Léčba závislosti na nikotinu

Podobně jako u jiných nemocí, lidé mají možnost se s léčbou nikotinové závislosti obrátit na odborníka. Léčba se skládá s psychosociální a behaviorální intervence společně s farmakologickou léčbou. Farmakologická léčba se pak nasazuje pacientům silněji závislým (to platí pro uživatele, kteří kouří více jak 10 cigaret denně, po probuzení si zapálí cigaretu do hodiny nebo kteří se pokoušeli přestat bezúspěšně) a zabraňuje pouze projevům abstinčních příznaků. Takovým lékem je například Vareniklin, který ztrojnásobuje procento úspěšnosti, nebo původní antidepresivum Bupropion, které úspěšnost zdvojnásobuje. Další možností jsou nikotinové náhražky, které jsou na rozdíl od farmak volně prodejné. V ČR taktéž funguje několik center pro závislé na nikotinu, které mají vyšší efektivitu. Často zde pracuje více odborníků, kteří mají vzdělání navíc pro léčbu nikotinově závislých. Centra pak fungují například při pneumologických klinikách a nabízí intenzivnější psychosociální a behaviorální intervence, informují pacienty o užívání tabáku a indikují farmakoterapii. [2]

Statistická data pro ČR

Český zdravotní ústav provádí posledních pár let pravidelné šetření v české populaci ohledně užívání tabákových výrobků a alkoholu. Nejnovější zpráva shrnuje šetření z roku 2020 [1]. Na tyto zprávy se pravidelně odkazuje Ministerstvo zdravotnictví České republiky, vláda ČR a další internetová média. [9] [10]

Důležité výsledky šetření ze zprávy od SZÚ vzhledem k obsahu v této práci jsou v následujících tabulkách. V Tabulce 1.1 je rozdělení všech respondentů (tedy kuřáků i nekuřáků) podle demografických charakteristik. Tabulka ukazuje procentuální zastoupení kuřáků v české populaci. Tabulka 1.2 vybírá pouze aktivní kuřáky. V Tabulce 1.3 je populace kuřáků rozdělena podle věku, kdy začali s kouřením, a v Tabulce 1.4 podle užívaných tabákových výrobků. Tabulka 1.5 rozděluje kuřáky

podle počtu vykouřených cigaret a Tabulka 1.6 ukazuje rozdělení podle druhů užívaných elektronických cigaret. V Tabulce 1.7 je přehled užívaných tabákových výrobků u denních uživatelů.

Tabulka 1.1 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – z celku všech respondentů [1]

Demografická charakteristika	Jakýkoliv kouřený tabákový výrobek	Jakékoliv cigarety ¹	Typ cigarety		Ostatní tabákové výrobky ²
			Komerčně vyrobené	Ručně ubalené	
Procenta (95% CI)					
Celkem (n=1769)	23,1 (21,2–25,2)	21,7 (19,8–23,6)	19,7 (17,9–21,7)	3,1 (2,4–4,0)	3,7 (2,8–4,7)
<i>Pohlaví</i>					
Muži (n=867)	28,3 (25,3–31,4)	26,4 (23,5–29,5)	23,4 (20,6–26,4)	4,5 (3,2–6,1)	4,0 (2,8–5,6)
Ženy (n=902)	18,2 (15,7–20,9)	17,1 (14,7–19,7)	16,2 (13,8–18,8)	1,8 (1,0–2,9)	3,3 (2,3–4,7)
<i>Věková skupina</i>					
15–24 (n=190)	24,2 (18,3–30,9)	20,5 (15,0–27,0)	20,5 (15,0–27,0)	2,1 (0,6–5,3)	8,9 (5,3–13,9)
25–44 (n=603)	26,5 (23,0–30,3)	24,5 (21,2–28,2)	22,4 (19,1–25,9)	3,8 (2,4–5,7)	4,0 (2,6–5,9)
45–64 (n=556)	23,6 (20,1–27,3)	22,7 (19,2–26,4)	20,7 (17,4–24,3)	2,7 (1,5–4,4)	2,5 (1,4–4,2)
65+ (n=420)	17,1 (13,7–21,1)	16,7 (13,2–20,6)	14,3 (11,1–18,0)	3,1 (1,7–5,2)	2,4 (1,1–4,3)
<i>Místo bydliště</i>					
Město (n=1394)	22,3 (20,1–24,6)	20,9 (18,8–23,2)	19,2 (17,2–21,4)	2,7 (1,9–3,7)	4,0 (3,0–5,2)
Venkov (n=375)	26,1 (21,8–30,9)	24,3 (20,0–28,9)	21,6 (17,5–26,1)	4,5 (2,7–7,2)	2,4 (1,1–4,5)
<i>Vzdělání³</i>					
ZŠ a SOU (n=570)	30,9 (27,1–34,8)	30,0 (26,3–33,9)	26,8 (23,2–30,7)	4,9 (3,3–7,0)	2,8 (1,6–4,5)
SŠ s maturitou (n=621)	20,6 (17,5–24,0)	18,8 (15,8–22,1)	16,7 (13,9–19,9)	2,9 (1,7–4,5)	3,2 (2,0–4,9)
VŠ (n=388)	15,2 (11,8–19,2)	14,4 (11,1–18,3)	13,7 (10,4–17,5)	1,3 (0,4–3,0)	3,1 (1,6–5,3)

95% CI – 95% interval spolehlivosti

¹Zahrnuje kuřáctví cigaret, jak vyrobených průmyslově, tak i ručně ubalených

²Zahrnuje kuřáctví dýmek, doutníků, doutniček a vodních dýmek

³Kategorie podle vzdělání je zpracována pro dospělé ve věku 25 let a starší

Poznámka: Současné kuřáctví tabákových výrobků zahrnuje každodenní i příležitostné (méně často než denně) kuřáctví.

Tabulka 1.2 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – ze souboru současných kuřáků [1]

Demografická charakteristika	Jakékoliv cigarety ¹	Typ cigarety		Ostatní tabákové výrobky ²
		Komerčně vyrobené	Ručně ubalené	
Procenta (95% CI)				
Celkem (n=409)	93,6 (90,8–95,8)	85,3 (81,5–88,6)	13,4 (10,3–17,1)	15,9 (12,5–19,8)
<i>Pohlaví</i>				
Muži (n=245)	93,5 (89,6–96,2)	82,9 (77,5–87,4)	15,9 (11,6–21,1)	14,3 (10,2–19,3)
Ženy (n=164)	93,9 (89,1–97,0)	89,0 (83,2–93,4)	9,8 (5,7–15,4)	18,3 (12,7–25,1)
<i>Věková skupina</i>				
15–24 (n=46)	84,8 (71,1–93,7)	84,8 (71,1–93,7)	8,7 (2,4–20,8)	37,0 (23,2–52,5)
25–44 (n=160)	92,5 (87,3–96,1)	84,4 (77,8–89,6)	14,4 (9,3–20,8)	15,0 (9,9–21,5)
45–64 (n=131)	96,2 (91,3–98,7)	87,8 (80,9–92,9)	11,5 (6,6–18,2)	10,7 (6,0–17,3)
65+ (n=72)	97,2 (90,3–99,7)	83,3 (72,7–91,1)	18,1 (10,0–28,9)	13,9 (6,9–24,1)
<i>Místo bydliště</i>				
Město (n=311)	93,9 (90,6–96,3)	86,2 (81,8–89,8)	12,2 (8,8–16,4)	18,0 (13,9–22,7)
Venkov (n=98)	92,9 (85,8–97,1)	82,7 (73,7–89,6)	17,3 (10,4–26,3)	9,2 (4,3–16,7)
<i>Vzdělání³</i>				
ZŠ a SOU (n=176)	97,2 (93,5–99,1)	86,9 (81,0–91,5)	15,9 (10,8–22,2)	9,1 (5,3–14,3)
SŠ s maturitou (n=128)	91,4 (85,1–95,6)	81,3 (73,4–87,6)	14,1 (8,6–21,3)	15,6 (9,8–23,1)
VŠ (n=59)	94,9 (85,9–98,9)	89,8 (79,2–96,2)	8,5 (2,8–18,7)	20,3 (11,0–32,8)

95% CI – 95% interval spolehlivosti

¹Zahrnuje kuřáctví cigaret, jak vyrobených průmyslově, tak i ručně ubalených

²Zahrnuje kuřáctví dýmek, doutníků, doutníčků a vodních dýmek

³Kategorie podle vzdělání je zpracována pro dospělé ve věku 25 let a starší

Poznámka: Současné kuřáctví tabákových výrobků zahrnuje každodenní i příležitostné (méně často než denně) kuřáctví.

Tabulka 1.3 Věkové rozmezí, kdy začali současní a bývalí kuřáci pravidelně kouřit [1]

Demografická charakteristika	14 a méně let	15–19 let	20–24 let	25 a více let
	Procenta (95% CI)			
Celkem (n=680)	8,8 (6,8–11,2)	61,2 (57,4–64,9)	20,6 (17,6–23,8)	9,4 (7,3–11,9)
<i>Pohlaví</i>				
Muži (n=381)	8,7 (6,0–11,9)	65,9 (60,9–70,6)	18,6 (14,9–22,9)	6,8 (4,5–9,8)
Ženy (n=299)	9,0 (6,0–12,9)	55,2 (49,4–60,9)	23,1 (18,4–28,3)	12,7 (9,2–17,0)
<i>Věková skupina</i>				
15–24 (n=62)	9,7 (3,6–19,9)	90,3 (80,1–96,4)	0,0 (0,0–5,8)	–
25–44 (n=248)	11,7 (8,0–16,4)	69,4 (63,2–75,0)	15,3 (11,1–20,4)	3,6 (1,7–6,8)
45–64 (n=218)	6,4 (3,6–10,5)	52,3 (45,4–59,1)	26,1 (20,4–32,5)	15,1 (10,7–20,6)
65+ (n=152)	7,2 (3,7–12,6)	48,7 (40,5–56,9)	29,6 (22,5–37,5)	14,5 (9,3–21,1)
<i>Místo bydliště</i>				
Město (n=542)	8,3 (6,1–11,0)	61,1 (56,8–65,2)	21,8 (18,4–25,5)	8,9 (6,6–11,6)
Venkov (n=138)	10,9 (6,2–17,3)	61,6 (52,9–69,7)	15,9 (10,3–23,1)	11,6 (6,8–18,1)
<i>Vzdělání¹</i>				
ZŠ a SOU (n=261)	9,2 (6,0–13,4)	59,0 (52,8–65,0)	21,5 (16,6–26,9)	10,3 (6,9–14,7)
SŠ s maturitou (n=240)	8,8 (5,5–13,1)	57,1 (50,6–63,4)	23,8 (18,5–29,6)	10,4 (6,9–15,0)
VŠ (n=117)	7,7 (3,6–14,1)	59,0 (49,5–68,0)	23,1 (15,8–31,8)	10,3 (5,4–17,2)

95% CI – 95% interval spolehlivosti

¹Kategorie podle vzdělání je zpracována pro dospělé ve věku 25 let a starší

Poznámka: Tabulka zahrnuje každodenní i příležitostné (méně často než denně) kuřáctví a bývalé kuřáky.

Tabulka 1.4 Současní kuřáci různých druhů tabákových výrobků podle vybraných demografických charakteristik – ze souboru současných kuřáků [1]

Demografická charakteristika	Ostatní tabákové výrobky			
	Dýmky plněné tabákem	Doutníky, doutníčky bez příchuti	Doutníky, doutníčky s příchutí	Sezení vodní dýmky
	Procenta (95% CI)			
Celkem (n=409)	3,4 (1,9–5,7)	4,6 (2,8–7,2)	4,4 (2,6–6,9)	9,3 (6,7–12,5)
<i>Pohlaví</i>				
Muži (n=245)	3,3 (1,4–6,3)	5,3 (2,9–8,9)	4,1 (2,0–7,4)	6,5 (3,8–10,4)
Ženy (n=164)	3,7 (1,4–7,8)	3,7 (1,4–7,8)	4,9 (2,1–9,4)	13,4 (8,6–19,6)
<i>Věková skupina</i>				
15–24 (n=46)	8,7 (2,4–20,8)	6,5 (1,4–17,9)	8,7 (2,4–20,8)	28,3 (16,0–43,5)
25–44 (n=160)	1,9 (0,4–5,4)	3,8 (1,4–8,0)	3,8 (1,4–8,0)	11,3 (6,8–17,2)
45–64 (n=131)	2,3 (0,5–6,5)	4,6 (1,7–9,7)	3,8 (1,3–8,7)	3,1 (0,8–7,6)
65+ (n=72)	5,6 (1,5–13,6)	5,6 (1,5–13,6)	4,2 (0,9–11,7)	4,2 (0,9–11,7)
<i>Místo bydliště</i>				
Město (n=311)	4,5 (2,5–7,4)	4,8 (2,7–7,8)	5,1 (3,0–8,2)	10,9 (7,7–14,9)
Venkov (n=98)	0,0 (0,0–3,7)	4,1 (1,1–10,1)	2,0 (0,2–7,2)	4,1 (1,1–10,1)
<i>Vzdělání¹</i>				
ZŠ a SOU (n=176)	2,8 (0,9–6,5)	4,0 (1,6–8,0)	2,8 (0,9–6,5)	3,4 (1,3–7,3)
SŠ s maturitou (n=128)	2,3 (0,5–6,7)	4,7 (1,7–9,9)	4,7 (1,7–9,9)	9,4 (4,9–15,8)
VŠ (n=59)	3,4 (0,4–11,7)	5,1 (1,1–14,1)	5,1 (1,1–14,1)	11,9 (4,9–22,9)

95% CI – 95% interval spolehlivosti

¹Kategorie podle vzdělání je zpracována pro dospělé ve věku 25 let a starší

Tabulka 1.5 Počet cigaret průměrně vykouřených denními kuřáky za den podle vybraných demografických charakteristik [1]

Demografická charakteristika	Počet cigaret průměrně vykouřených za den ¹				
	< 5	5–9	10–14	15–24	≥ 25
	Procenta (95% CI)				
Celkem (n=285)	13,0 (9,3–17,4)	22,5 (17,7–27,8)	23,9 (19,0–29,2)	31,6 (26,2–37,3)	9,1 (6,0–13,1)
<i>Pohlaví</i>					
Muži (n=177)	13,6 (8,9–19,5)	18,1 (12,7–24,6)	23,2 (17,2–30,1)	35,0 (28,0–42,5)	10,2 (6,1–15,6)
Ženy (n=108)	12,0 (6,6–19,7)	29,6 (21,2–39,2)	25,0 (17,2–34,3)	25,9 (18,0–35,2)	7,4 (3,3–14,1)
<i>Věková skupina</i>					
15–24 (n=24)	20,8 (7,1–42,2)	37,5 (18,8–59,4)	29,2 (12,6–51,1)	4,2 (0,1–21,1)	8,3 (1,0–27,0)
25–44 (n=113)	9,7 (5,0–16,8)	22,1 (14,9–30,9)	22,1 (14,9–30,9)	38,9 (29,9–48,6)	7,1 (3,1–13,5)
45–64 (n=93)	16,1 (9,3–25,2)	17,2 (10,2–26,4)	24,7 (16,4–34,8)	30,1 (21,0–40,5)	11,8 (6,1–20,2)
65+ (n=55)	10,9 (4,1–22,2)	25,5 (14,7–39,0)	23,6 (13,2–37,0)	30,9 (19,1–44,8)	9,1 (3,0–20,0)
<i>Místo bydliště</i>					
Město (n=217)	10,6 (6,8–15,5)	23,0 (17,6–29,2)	25,3 (19,7–31,7)	30,4 (24,4–37,0)	10,6 (6,8–15,5)
Venkov (n=68)	20,6 (11,7–32,1)	20,6 (11,7–32,1)	19,1 (10,6–30,5)	35,3 (24,1–47,8)	4,4 (0,9–12,4)
<i>Vzdělání²</i>					
ZŠ a SOU (n=149)	8,7 (4,7–14,5)	18,1 (12,3–25,3)	24,2 (17,5–31,8)	37,6 (29,8–45,9)	11,4 (6,8–17,6)
SŠ s maturitou (n=79)	12,7 (6,2–22,0)	24,1 (15,1–35,0)	22,8 (14,1–33,6)	32,9 (22,7–44,4)	7,6 (2,8–15,8)
VŠ (n=33)	27,3 (13,3–45,5)	27,3 (13,3–45,5)	21,2 (9,0–38,9)	21,2 (9,0–38,9)	3,0 (0,1–15,8)

95% CI – 95% interval spolehlivosti

¹Denní kuřáci, cigarety zahrnují průmyslově vyráběné i ručně balené

²Kategorie podle vzdělání je zpracována pro dospělé ve věku 25 let a starší

Poznámka: celkový počet denních kuřáků odpovídá respondentům, kteří uvedli počet vykouřených cigaret za den.

Tabulka 1.6 Současní uživatelé elektronické cigarety – elektronické cigarety s nikotinem a bez nikotinu [1]

Demografická charakteristika	Elektronické cigarety s nikotinem	Elektronické cigarety bez nikotinu	Elektronické cigarety občas obsahující nikotin	Nevím
	Procenta (95% CI)			
Celkem (n=133)	45,1 (36,5–54,0)	24,1 (17,1–32,2)	23,3 (16,4–31,4)	7,5 (3,7–13,4)
<i>Pohlaví</i>				
Muži (n=76)	46,1 (34,5–57,9)	23,7 (14,7–34,8)	26,3 (16,9–37,7)	3,9 (0,8–11,1)
Ženy (n=57)	43,9 (30,7–57,6)	24,6 (14,1–37,8)	19,3 (10,0–31,9)	12,3 (5,1–23,7)
<i>Místo bydliště</i>				
Město (n=107)	45,8 (36,1–55,7)	24,3 (16,5–33,5)	24,3 (16,5–33,5)	5,6 (2,1–11,8)
Venkov (n=26)	42,3 (23,4–63,1)	23,1 (9,0–43,6)	19,2 (6,6–39,4)	15,4 (4,4–34,9)
<i>Kuřáctví tabáku v současnosti¹</i>				
Ano (n=65)	50,8 (38,1–63,4)	12,3 (5,5–22,8)	27,7 (17,3–40,2)	9,2 (3,5–19,0)
Ne (n=68)	39,7 (28,0–52,3)	35,3 (24,1–47,8)	19,1 (10,6–30,5)	5,9 (1,6–14,4)

95% CI – 95% interval spolehlivosti

¹Zahrnuje každodenní i příležitostné (méně často než denně) kouření

Tabulka 1.7 Přehled forem konzumace tabákových výrobků a užívání elektronických cigaret (EC) u denních konzumentů (uživatelů), podle vybraných demografických charakteristik [1]

Demografická charakteristika	Konzumenti tabáku celkem ¹	Konzumenti jednotlivých tabákových výrobků				Uživatelé EC
		Kuřáci tabáku ²	Kuřáci cigaret ³	Uživatelé zahříváných tabákových výrobků ⁴	Uživatelé bezdýmných tabákových výrobků ⁵	
<i>Procenta (95% CI)</i>						
Celkem (n=1769)	26,4 (24,4–28,5)	23,1 (21,2–25,2)	21,7 (19,8–23,6)	4,1 (3,2–5,2)	3,4 (2,6–4,4)	4,8 (3,9–5,9)
<i>Pohlaví</i>						
Muži (n=867)	31,9 (28,9–35,2)	28,3 (25,3–31,4)	26,4 (23,5–29,5)	4,3 (3,0–5,8)	5,2 (3,8–6,9)	6,1 (4,6–7,9)
Ženy (n=902)	21,1 (18,4–23,9)	18,2 (15,7–20,9)	17,1 (14,7–19,7)	4,0 (2,8–5,5)	1,8 (1,0–2,9)	3,5 (2,4–5,0)
<i>Věková skupina</i>						
15–24 (n=190)	33,2 (26,5–40,3)	24,2 (18,3–30,9)	20,5 (15,0–27,0)	3,7 (1,5–7,4)	11,6 (7,4–17,0)	10,0 (6,1–15,2)
25–44 (n=603)	30,3 (26,7–34,2)	26,5 (23,0–30,3)	24,5 (21,2–28,2)	7,1 (5,2–9,5)	4,3 (2,8–6,3)	7,3 (5,4–9,7)
45–64 (n=556)	26,1 (22,5–29,9)	23,6 (20,1–27,3)	22,7 (19,2–26,4)	3,1 (1,8–4,9)	1,8 (0,9–3,3)	2,7 (1,5–4,4)
65+ (n=420)	18,1 (14,5–22,1)	17,1 (13,7–21,1)	16,7 (13,2–20,6)	1,4 (0,5–3,1)	0,7 (0,1–2,1)	1,7 (0,7–3,4)

95% CI – 95% interval spolehlivosti

¹Současní konzumenti a uživatelé (denní + příležitostní)

²Zahrnuje kuřáctví tabáku (cigaret, jak vyrobených průmyslově, tak i ručně ubalených; kuřáctví dýmek, doutníků, doutniček a vodních dýmek

³Zahrnuje cigarety vyráběné průmyslově nebo ručně ubalené

⁴Zahrnuje zahříváné tabákové výrobky

⁵Zahrnuje bezdýmné tabákové výrobky – žvýkáci, šňupací a další formy tabáku, které se vkládají do úst, ale nekouří se

1.1.2 mHealth aplikace

V současné době probíhá velký zájem o moderní technologie, které monitorují zdraví uživatele. V první řadě se jedná hlavně o smartphony s takzvanými mHealth aplikacemi, které si uživatel velmi snadno stáhne, a oblíbené jsou i wearable devices, které mnohdy obsahují navíc další senzory pro kontrolu stavu a získané informace dokážou buďto samy zpracovat, nebo se data opět posílají do smartphonů, kde probíhá samotné vyhodnocení. [11] Tyto aplikace pak mohou být prostředníky v komunikaci mezi doktorem a pacientem, nebo slouží pouze pro informovanost samotného pacienta [12]. Mezi nesporné výhody se řadí hlavně jednoduchá a okamžitá komunikace jak s pacientem v jednoduchém grafickém rozhraní, tak se senzory přímo v telefonu. Dnešní velkou výhodou je i výpočetní výkon procesorů v takových zařízeních. Jeden smartphone

nahrazuje několik zařízení najednou – například GPS, fotoaparát, bezdrátové připojení, internetové připojení, reproduktor. [11]

Přestože jsou tato zařízení a aplikace velmi slibnou technologií v oblasti zdraví každého jedince, jejich vysoký potenciál stále nebyl plně využit [11]. Pro diagnózy některých onemocnění jsou například zásadní krevní testy, které jsou tímto způsobem nerealizovatelné, proto je důležité se v tomto směru soustředit převážně na data poskytující senzory, data, které nám poskytne uživatel, a data behaviorálního chování uživatele – například způsob psaní. Nejzásadnějším problémem je ale stále neznalost jasných biomarkerů v návaznosti na znalost pouze těchto dat pro průběh a predikce onemocnění – tedy například nahrazení krevních testů, neznalost složitosti fenotypů, chování a příznaků [4]. Dalším problémem je obrovské množství takových aplikací na trhu, kdy jen hrstka vývojářů dokáže doložit vývoj s pomocí skutečného odborníka nebo prokázat klinickou účinnost [11] [12].

Avšak právě díky velké popularitě smartphonů a aplikací monitorujících zdraví zde dochází k obrovské digitální stopě – sociální, behaviorální a kognitivní po uživatelích používajících tyto aplikace. Jedná se o velmi přirozený způsob sběru dat od uživatelů. Právě tato data pomáhají mnohem lépe porozumět každému uživateli a jsou velmi důležitým zdrojem pro vývoj markerů a fenotypů sloužících právě k diagnostice, monitorování a léčbě nemocí. [4] [5]

Podle serveru Statista.com je k prvnímu kvartálu roku 2021 v Google Play Store k dispozici přes 53 000 mHealth aplikací. [13]

Aplikace na odvykání kouření

Aplikace simuluje reálné intervence odborníka v podobě textových rozhovorů uživatele s virtuálním terapeutem. V sérii sezení se virtuální terapeut sám představí, oslovuje pacienta jménem, empaticky klade otázky, které reflektuje, používá pozdravy a rozloučení tak, jako je to při běžných terapeutických sezeních s odborníkem. Relace jsou navíc přizpůsobeny předchozím odpovědím uživatele, takže je terapie přizpůsobena každému individuálně. Základní délka intervencí je 6 týdnů. Důležitým prvkem aplikace je relapse management systém, který na základě odpovědi uživatele, že si zapálil po takzvaném „dni D“ (počáteční den abstinence), spustí speciální terapeutickou relaci, která se ptá na myšlenky a pocity pacienta a snaží se motivovat uživatele k abstinenci. [3]

Tři čtvrtiny pacientů, kteří se účastnili studie, která zkoumá klinickou stránku aplikace, uvedlo, že preferují spíše tuto (virtuální) formu odvykací metody před hledáním odborné pomoci v podobě adiktologa, terapeuta apod. Jako výhody aplikace uživatelé uvedli například jednoduché a intuitivní ovládání, atraktivita, každodenní podpora, respektující přístup, autentické rozhovory, posílení motivace, časová flexibilita sezení a dostupnost ve vlastním zařízení. [3]

1.1.3 Explorační analýza

Explorační analýza se od klasické analýzy dat se liší například nekonkrétními otázkami. U datové analýzy jsou otázky předem definované a analýzou hledáme jasnou odpověď. Zatímco před začátkem explorační analýzy vzniká mnoho nekonkrétních otázek a mnoho dalších vzniká v průběhu. Často se tedy explorační analýza provádí za účelem seznámení se s daty a snaží se zjistit, co by nám data mohla říct. Z toho důvodu není doména explorační analýzy nijak omezená v rozsáhlosti ani v použitých nástrojích. [14]

NLP metody

Natural Language Processing (NLP) je analytická metoda zabývající se zpracováním přirozené lidské řeči, jejímu porozumění a optimalizaci jejího zpracování. NLP se snaží propojit komunikační bariéru mezi počítačovými technologiemi a přirozenou lidskou řečí, zahrnuje širokou škálu nástrojů od jednoduchých algoritmů pro zpracování textů až sahajících dále za metody strojového učení. Základním úkolem pro NLP je například rychlá extrakce důležitých informací z dlouhého textu bez nutnosti čtení celého textu. [15] [16]

V současné době využíváme každý den spoustu produktů postavených na základě NLP. Je jím například strojový překlad, vyhledávání na internetu, automatické opravy textu, automatické dosazování textu nebo hlasový asistent. Spoustu aplikací NLP najdeme i v komerční sféře, kdy je například potřeba zpracovat spoustu recenzí psaných uživateli nebo se provádí důkladné analýzy trhu na základě NLP. Dále například zpravodajské služby využívají systémy automatického strojového psaní zpráv ihned po skončení události. [15] [16]

Word Clouds

Word Clouds je forma vizualizace, která shrnuje text. Poprvé byla využívána v podobě tag clouds na stránkách jako Flickr pro vizualizaci nejčastěji používaných klíčových slov.

Metoda se poté více rozšířila a má mnoho využití. Nejvíce používanou aplikací je vizualizace nejčastěji použitých slov v textu, což dovoluje udělat si okamžitou představu o nejdiskutovanějších tématech, kterým se text věnuje. Word Clouds jsou založeny na algoritmu pozitivní korelace mezi velikostí písma slova a jeho frekvence v textu. Různě velká slova jsou pak uspořádána do libovolného obrazce a barev. [17]

Analýza sentimentu

Nejen extrakce klíčových slov, ale celková nálada pisatele, která je promítnuta do textu, je důležitým faktorem pro mnoho odvětví. V dnešní době, kdy lze napsat na sociální síť nebo fóra cokoli, co člověka napadne, je takového textu velké množství a pro extrakci důležitých informací je třeba s obsahem rozumně nakládat. Analýza sentimentu dokáže například odhalit, zda si přispěvatelé ve fóru na daný produkt stěžují, nebo ho vyzdvihují. [18] [19]

Analýza sentimentu je převážně klasifikační otázkou. Metod a nástrojů pro sentiment analýzu je velké množství a jejich přesnost je velmi diskutována [18] [19]. V roce 2014 byl vytvořen nástroj VADER, který kombinuje kvalitativní a kvantitativní metody pro hodnocení sentimentu. VADER vychází z reálného hodnocení sentimentu lidmi a aplikuje několik dalších syntaktických a gramatických pravidel. Velmi dobře funguje pro použití u sociálních sítí, tedy u krátkých kontextově řídkých textů, ve kterých se často objevují zkrácené jazykové konvence. [20]

1.1.4 Digitální fenotyp

Fenotyp

Slovo fenotyp má mnoho významů. V biologii má fenotyp význam v podobě pozorovatelných znaků organismu zahrnujících jeho fyziologii, morfologii, nebo expresi genu reagující na podněty v prostředí. V lékařském prostředí pak fenotyp znamená spíše odchylku od normálních vlastností. Pro lékaře je důležité zařadit si pacienty do určitých fenotypových skupin za účelem stanovení diagnózy, což je jeden z nejtěžších úkolů v medicíně. Lékař předepíše pacientovi léčbu a poté pozoruje odpověď jeho organismu, která může a nemusí být pozitivní. Studium fenotypů zahrnuje i fenotypové abnormality spojené s onemocněními – určitý znak může být projevem onemocnění, nebo pouze vlastností pacienta. Pochopení fenotypů pak nahrává včasné a přesné diagnóze. [21]

Digitální fenotyp

V roce 1982 byl představen koncept „rozšířeného fenotypu“ evolučním biologem Richardem Dawkinsem. Myšlenkou je, že fenotypy by se neměly omezovat pouze na biologické procesy, ale předmětem bylo rozšíření fenotypu o všechny účinky, které má daný gen v daném prostředí. Každý jedinec má schopnost úpravy svého vlastního prostředí a tyto modifikace a chování jsou výrazy genomu živých organismů – a součástí rozšířeného fenotypu. Rozšířením Dawkinsovy teorie dostaneme pojem „digitální fenotyp“. Teorie je rozšířena o konfrontaci člověka s digitálními technologiemi. [6] Digitální fenotyp je statickým konstruktem jako genotyp, digital phenotyping zkoumá dynamické chování fenotypu v čase. [22]

„Deep Phenotyping“

„Deep Phenotyping“ (DeP) je součástí precizní medicíny, která je definována jako přesná a komplexní analýza fenotypových abnormalit, ve kterých jsou pozorovány a popisovány jednotlivé složky fenotypu, často pro účely vědeckého zkoumání lidských onemocnění. V praxi znamená DeP členění pacientů do subpopulací například na základě citlivosti na onemocnění, nebo pravděpodobnosti specifické odpovědi na terapii – takových faktorů členění pacientů existuje celá řada a k jejich odhalení vede jejich pečlivá analýza. Pokud se podaří extrahovat přesné fenotypy, lze pak ke každému typu přiřadit i odpovídající terapii a mít k pacientovi individuální přístup, který bude v léčbě úspěšný. [21]

„Digital Phenotyping“

Pojem „Digital Phenotyping“ (DiP) byl definován v roce 2016 skupinou autorů článku *New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research*, kde se objevuje přesná definice: „moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices.“ [4]

DiP je stále vyvíjený klinický nástroj sloužící k porozumění každému pacientu založený na kontinuálním sběru dostupných dat. Hlavním nástrojem pro sběr dat je smartphone nebo wearable device. Zdrojem mohou být senzory daného zařízení (umístění pacienta – GPS, aktivita – rozsvícená obrazovka), hlas a řeč, interakce člověk-počítač (například měření prodlevy při psaní znaků, příspěvky na sociálních sítích, vyhledávání na internetu), sběr dat tímto způsobem je pro pacienty mnohem přirozenější a ukazuje jejich chování v běžném životě, tzv. in-situ. Sbíraná data pro DiP se dělí na aktivní

a pasivní data. Aktivní data představují data, které zaznamenává sám pacient – jedná se například o časové značky – sledování nálady, čas jídla apod. Zatímco pasivní data se sbírají právě díky dostupným sensorům a nevyžadují od pacienta žádnou aktivitu navíc. Data jsou kvantifikována s odkazem na lidský fenotyp. [4]

Cílem je pochopení chování pacientů sběrem pouze těchto dat pro vývoj dynamických modelů zdravotního chování, posunout se dále za tradiční způsoby vyšetření, poskytnout pacientovi stálý dozor tak, aby mohl být predikován vývoj onemocnění, a případně aby mohla být včas poskytnuta pomoc. [6]

Největší rozvoj v oblasti DiP je právě v oblasti duševního zdraví, kdy je snaha o predikci nemocí nebo změn nálad na základě interakce s technologiemi [5]. Bylo například prokázáno, že pomocí údajů z Google vyhledávače lze identifikovat sebevražedné myšlenky [6] nebo například u schizofrenie se potvrdila souvislost relapsů s nekvalitním spánkem – data o spánku lze získat právě ze sensorů ve smartphonu [23].

Vzhledem k pokroku ve sběru dat mobilními zařízeními, přesouvá se problematika DiP spíše k analýzám a modelování. Zásadním problémem v kontinuálním sběru dat pro DiP je jejich objem. Je třeba vybrat takové souhrnné statistiky, aby byly užitečné a co nejlépe reprezentovaly sbíraná data, přestože data budou nezpracovaná a některá budou i chybět. Metody zároveň musí respektovat vědecké otázky v této oblasti, proto se očekává vývoj statistických modelů specifických právě pro oblast DiP. [22]

„Deep Digital Phenotyping“

„Deep Digital Phenotyping“ (DDP) je spojení „Deep Phenotyping“ a „Digital Phenotyping“. DDP se snaží objektivně a automatizovaně poskytovat poznatky o životním stylu, psychologickém stavu, sociodemografických charakteristikách, prostředí jedince a přiblížit se pomocí širší skupiny dat definici precizní medicíny. K identifikaci shluků v podobě fenotypů se díky výpočetním možnostem v dnešní době mohou využívat nástroje, které dokážou zpracovat velké množství dat. Používané metody jsou od jednoduchých algoritmů až po metody hlubokého učení. Běžně se používají algoritmy shlukování, například aglomerativní hierarchické shlukování, kterým předchází redukce dimenze. [24]

1.2 Motivace

Hlavní motivací této práce je podpořit vývoj metody digitálního phenotypingu v adiktologii u pacientů závislých na nikotinu. K dispozici k tomu je rozsáhlý soubor dat, který byl získán díky dlouhodobému fungování aplikace na odvykání kouření. Vzhledem k velikosti daného souboru je třeba se nejdříve s daty seznámit pomocí explorační analýzy, která i sama o sobě může sloužit k lepšímu pochopení závislých pacientů.

1.3 Cíle práce

Hlavním cílem práce je v první řadě přehledné zpracování dat uživatelů závislých na nikotinu v podobě explorační analýzy. Tomu předchází seznámení se s daty, které jsou uloženy v databázi aplikace, seznámení se s aplikací, systémem terapií, formou otázek, na které uživatelé mohou odpovědět. Data poté uspořádat jako četnosti v jednotlivých kategoriích, podle nabízejících se možností, či podle demografických charakteristik. Pro otázky s otevřenou odpovědí provést základní extrakci důležitých informací pomocí základních metod NLP. Cílem je co nejlepší členění dat pro předání co nejkvalitnější informace o lidech závislých na nikotinu jako takových, o klinické kvalitě aplikace, o možnostech budoucí rozsáhlejší analýzy dat. Vše nejlépe ve formě automatizovaného zpracování pro snadnou aktualizaci dat v daném okamžiku.

Dalším cílem je soustředit se na veškeré údaje, které mohou charakterizovat uživatele a jejich chování. Záměr je vybrat několik parametrů, kterými se uživatelé mezi sebou liší, a podle nich najít několik vzorů podobného chování v souvislosti s demografickými údaji a mobilními zařízeními. Pro hledání takzvaných digitálních fenotypů použít algoritmy pro výběr parametrů a následně provést shlukovací analýzu na redukovaném počtu dimenzí. Zpětně přiřadit charakteristiky k jednotlivým skupinám a najít zastupující příznaky každé skupiny.

2 Metody

2.1 Příprava

2.1.1 Seznámení se s aplikací

Před samotnou datovou analýzou bylo třeba se seznámit přímo s aplikací a pochopit formu terapie. Terapie je pro uživatele rozdělena na několik částí. Pro uživatele jsou důležité tři části. V první fázi terapie se uživatel připravuje na takzvaný „den D“ – den, kdy si uživatel již cigaretu nezapálí. Tato fáze trvá obvykle 10 dní a virtuální terapeut s uživatelem probírá každé sezení několik osobních témat a snaží se uživateli uspořádat myšlenky i pomocí vhodných otázek nebo příkladů. Následuje fáze dne D, kdy ne vždy se uživateli podaří přestat s kouřením hned napoprvé. Virtuální terapeut je zde připraven připomenout uživateli, proč se uživatel původně rozhodnul přestat. Po dni D následuje udržovací fáze, kdy terapeut udržuje myšlenky uživatele správným směrem a snaží se předejít relapsům. Tato fáze má pak 15 sezení. Uživatelé také mohou využít ještě jedné možnosti, a to projít přípravnou fází ve zrychleném módu. Možnost se nabízí uživatelům, kteří už přestali kouřit před instalací aplikace a hledají podporu. Za účelem zpracování dat se průběh terapie dá rozdělit na 5 fází (Tabulka 2.1). Nejdůležitější fáze je konečná fáze (FIN), kdy uživatel dokončil všechna sezení.

Tabulka 2.1 Použité zkratky fází terapie

Zkratka fáze	Popis
START	Fáze ihned po instalaci aplikace před prvním sezením
EE	Přípravná fáze, 10 sezení
EQ	Fáze pro den D
FU	Udržovací podpůrná fáze, 15 sezení
FIN	Konečná fáze, do které se uživatel zařazen po skončení terapie

2.1.2 Zdroj dat

Aplikace umožňuje sběr dat, která se dají pro přehlednost rozdělit na 3 kategorie. První kategorií je vyplnění dobrovolného dotazníku, který se nachází mimo terapeutickou část v aplikaci a je na každém uživateli, jestli na otázky odpoví. Dotazník je rozdělen na několik sad otázek a soustřeďuje se převážně na kuřácké zvyklosti a demografické údaje

o uživateli. Přehled otázek, které jsou zpracovány, se nachází v Tabulce 2.2, 2.3, 2.4. Samotné otázky jsou otázkami typickými v klinické praxi.

Tabulka 2.2 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 1

Kuřácké zvyklosti	Typ odpovědí	Možné odpovědi
Od kolika let pravidelně kouříš?	Libovolné číslo	Odpověď byla rozdělena na diskrétní intervaly
Který z následujících tabákových a nikotinových produktů jsi už zkoušel?	Více odpovědí z nabídky	<ol style="list-style-type: none"> 1. Cigarety 2. Doutník 3. Dýmka 4. Vodní dýmka 5. Šňupací tabák 6. Bezdýmový tabák 7. Žvýkací tabák 8. Ecig s nikotinem 9. Ecig bez nikotinu 10. Ecig THC/CBD 11. Zahříváný tabák 12. Nikotinové sáčky
Který z těchto produktů užíváš v současnosti (alespoň jednou týdně)?	Více odpovědí z nabídky	<ol style="list-style-type: none"> 1. Cigarety 2. Doutník 3. Dýmka 4. Vodní dýmka 5. Šňupací tabák 6. Bezdýmový tabák 7. Žvýkací tabák 8. Ecig s nikotinem 9. Ecig bez nikotinu 10. Ecig THC/CBD 11. Zahříváný tabák 12. Nikotinové sáčky
Proč ses rozhodl přestat kouřit?	Více odpovědí z nabídky	<ol style="list-style-type: none"> 1. Zdraví 2. Rodina 3. Finance 4. Tlak partnera, dětí, přátel, rodiny 5. Zaměstnání 6. Zbavení závislosti 7. Obava z koronaviru 8. Jiný
Kolikrát v životě jsi už zkoušel přestat kouřit?	Libovolné číslo	Odpověď byla rozdělena na diskrétní intervaly
Jak dlouho jsi vydržel nekouřit při posledním pokusu?	Jedna odpověď z nabídky	<ol style="list-style-type: none"> 1. Méně než 1 den 2. Méně než 1 týden 3. 1 týden až 1 měsíc 4. 1-3 měsíce 5. 3-6 měsíců 6. Půl roku až rok 7. 1 rok a více
Jak ses o odvykání pokoušel?	Více odpovědí z nabídky	<ol style="list-style-type: none"> 1. Bez pomoci a bez léků 2. S náhradním nikotinem 3. S pomocí léků 4. S pomocí lékaře 5. S pomocí adiktologa, psychologa, terapeuta 6. S pomocí lékárníka 7. S pomocí telefonické linky 8. S pomocí webových infostránek 9. S pomocí mobilní aplikace 10. S pomocí alternativních metod 11. Jiné

Tabulka 2.3 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 2

Osobní informace	Typ odpovědí	Možné odpovědi
Kde momentálně žiješ?	Jedna odpověď z nabídky	1. Víc než 1 m. 2. Nad 100 tis. 3. Nad 10 tis. 4. Do 10 tis. 5. Obec do 2 tis.
Jak jsi na tom se zaměstnáním?	Více odpovědí z nabídky	1. Plný úvazek 2. Částečný úvazek 3. OSVČ živnostník 4. Nezaměstnaný 5. Prezenční student 6. Dálkové nebo kombinované studium 7. Rodičovská dovolená 8. Důchod 9. Pracovní neschopnost 10. Jiné
Nejvyšší dosažené vzdělání?	Jedna odpověď z nabídky	1. Základní 2. Střední 3. VOŠ 4. Bc 5. Mgr 6. Dr a vyšší

Tabulka 2.4 Seznam otázek v dobrovolném dotazníku, typy odpovědí a nabízené možnosti, část 3

Informace o zdraví	Typ odpovědí	Možné odpovědi
Kolik ti je let?	Libovolné číslo	Odpověď byla rozdělena na diskrétní intervaly
Jaký je tvůj aktuální zdravotní stav?	Více odpovědí z nabídky	1. Nic 2. Onemocnění dýchací soustavy 3. Srdečně-cévní onemocnění 4. Diabetes 5. Duševní onemocnění 6. Nádorové onemocnění 7. Jiné
Prodělal jsi onemocnění covid-19?	Jedna odpověď z nabídky	1. Ne 2. Ano, již negativní 3. Ano, léčím se
Užíváš pravidelně nějaké léky?	Jedna odpověď z nabídky	1. Ano 2. Ne

Dalším zdrojem dat jsou odpovědi uživatelů na otázky virtuálního terapeuta. Otázky jsou pokládány přímo v průběhu terapie, uživatel tak na ně musí odpovědět, aby se v terapii posunul dále. Otázky tohoto typu mají osobní a motivační charakter. Mezi možnostmi odpovědí nechybí ani otevřený způsob, kdy uživatel sepisuje slovní reakci. Patří sem i otázka na pohlaví a počet vykouřených cigaret za den – tyto otázky jsou položeny po instalaci aplikace. Otázky jsou často pokládány na základě rozhodovacího

stromu, tedy ne všem uživatelům jsou položeny všechny otázky. Počet odpovědí také záleží na fázi terapie, kdy například ke konečné fázi dojde pouze málo uživatelů.

Posledním možným zdrojem dat jsou pasivní data, která jsou sbírána bez aktivní účasti uživatele. Patří sem data jako fáze, ve které se zrovna nachází, poslední aktivita apod.

2.1.3 Jupyter Notebook

Pro zpracování dat jsem vybrala nástroj Jupyter Notebook, který využívá programovací jazyk Python 3. Výhoda tohoto nástroje oproti psaní python skriptů je, že kód je členěn do buněk, z nichž každá se dá spustit zvlášť v jiný okamžik. Zároveň na sebe buňky mohou navazovat podle toho, v jakém pořadí jsou spuštěny jako v běžném skriptu. Tato forma psaní python skriptu neochuzuje programovací jazyk Python, naopak přidává několik výhod navíc. Další výhodou je zde možnost skrýt veškeré buňky obsahující kód a pomocí markdown formy buněk, kterými lze vkládat popisky nebo nadpisy, upravit vzhled stránky tak, že takový notebook může sloužit přímo jako výstup. Samozřejmostí je i export do pdf nebo html.

Na Obrázku 2.1 je ukázka programovacího prostředí Jupyter Notebook. Na ukázce je vidět členění do buněk, v tomto případě použití předem definované funkce o několik buněk výše, a použití markdown buňky pro popis. Na Obrázku 2.2 už je spuštěný kód se skrytým vstupem.

Jupyter OdpovědiTerapie Last Checkpoint: před minutou (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Voilà

```

1 # EE02.5 Jak důležitý je pro tebe život bez cigaret?
2 ### Trochu jsme se podívali na to, proč chceš přestat kouřit. Jak moc důležité je pro tebe ale momentálně
   kouření?

```

```

In [11]: 1
         2 ind = {'1':'Vůbec', '2':'Trochu', '3':'Dost', '4':'Hodně', '5':'Momentálně nejdůležitější věc na světě'}
         3
         4
         5 vis(table['need'], ttable['need'], ftable['need'], wtable['need'], mtable['need'], ind, agetable)

```

	ALL	TRIAL	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over61
Vůbec	87	84	1	31	56	6	14	3	11	7	3	0
Trochu	609	580	16	315	293	39	101	48	33	10	2	0
Dost	2079	1904	92	892	1184	70	289	158	169	74	16	4
Hodně	3410	2961	249	1656	1754	49	331	226	345	137	36	8
Momentálně nejdůležitější věc na světě	1171	1041	81	609	561	32	89	57	91	41	18	4

	ALL	TRIAL	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over61
Vůbec	1.20	1.30	0.20	0.90	1.50	3.10	1.70	0.60	1.70	2.60	4.00	0.00
Trochu	8.30	8.80	3.60	9.00	7.60	19.90	12.30	9.80	5.10	3.70	2.70	0.00
Dost	28.30	29.00	21.00	25.50	30.80	35.70	35.10	32.10	26.00	27.50	21.30	25.00
Hodně	46.40	45.10	56.70	47.30	45.60	25.00	40.20	45.90	53.20	50.90	48.00	50.00
Momentálně nejdůležitější věc na světě	15.90	15.80	18.50	17.40	14.60	16.30	10.80	11.60	14.00	15.20	24.00	25.00

Obrázek 2.1 Ukázka prostředí Jupyter Notebook

Jupyter OdpovědiTerapie Last Checkpoint: před 13 minutami (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Code Voilà

EE02.5 Jak důležitý je pro tebe život bez cigaret?

Trochu jsme se podívali na to, proč chceš přestat kouřit. Jak moc *důležité* je pro tebe ale momentálně kouření?

	ALL	TRIAL	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over61
Vůbec	87	84	1	31	56	6	14	3	11	7	3	0
Trochu	609	580	16	315	293	39	101	48	33	10	2	0
Dost	2079	1904	92	892	1184	70	289	158	169	74	16	4
Hodně	3410	2961	249	1656	1754	49	331	226	345	137	36	8
Momentálně nejdůležitější věc na světě	1171	1041	81	609	561	32	89	57	91	41	18	4

	ALL	TRIAL	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over61
Vůbec	1.20	1.30	0.20	0.90	1.50	3.10	1.70	0.60	1.70	2.60	4.00	0.00
Trochu	8.30	8.80	3.60	9.00	7.60	19.90	12.30	9.80	5.10	3.70	2.70	0.00
Dost	28.30	29.00	21.00	25.50	30.80	35.70	35.10	32.10	26.00	27.50	21.30	25.00
Hodně	46.40	45.10	56.70	47.30	45.60	25.00	40.20	45.90	53.20	50.90	48.00	50.00
Momentálně nejdůležitější věc na světě	15.90	15.80	18.50	17.40	14.60	16.30	10.80	11.60	14.00	15.20	24.00	25.00

Obrázek 2.2 Prostředí Jupyter Notebook se skrytým kódem

2.1.4 Výběr a zisk dat

Data aplikace jsou uložena na serveru typu PostgreSQL. Nejdříve jsem se musela seznámit se strukturou dat, která kombinovala i jiné formáty, jako například JSON. Pro dotazování, zobrazení a komunikaci jsem nejdříve využívala nástroj DataGrip dokud

jsem veškeré dotazy ve formátu PostgreSQL neodladila. Některé dotazy jsem vytvořila rovnou filtrované už v SQL formátu, například podle fáze terapie uživatelů, někdy jsem rovnou vytvořila tabulku četností.

Pro jednoduchost jsem propojila server databáze přímo v Notebook skriptu pomocí přihlašovacích údajů. Kurzorem jsem pak odesílala předpřipravené dotazy v SQL formátu a ukládala odpovědi do proměnných, se kterými jsem později pracovala.

Pro přehlednost jsem data zpracovala ve třech různých kategoriích – ve třech různých skriptech. Data jsem rozdělila na základní informace o uživatelích – Hlavní tabulka, na odpovědi v dobrovolném dotazníku – Osobní profil a odpovědi na otázky pokládané při terapii – Odpovědi terapie. V každé sekci jsem se dotazovala na data, která byla hlavním tématem sekce, a data společná, například demografické informace. Zpracování a vizualizace byly pak provedeny přímo v Notebooku.

2.2 Zpracování dat pro explorační analýzu

Každý skript má ve většině případů stejné třídění četností podle kategorií zobrazených v Tabulce 2.5.

Tabulka 2.5 Společný seznam kategorií

Kategorie	Popis	Název
Všichni	Zahrnuje všechny možné uživatele	ALL
Pohlaví	Rozdělení podle uvedeného pohlaví na muže a ženy	men women
Míra závislosti	Rozdělení podle uvedeného počtu vykouřených cigaret denně na: Lehčí kuřáci (≤ 10 cigaret), Střední kuřáci (11-20 cigaret), Těžcí kuřáci (> 20 cigaret)	light middle heavy
Věk	Podle uvedeného věku na: věk pod 18 let, věk 18 až 20 let, věk 21 až 25 let, věk 26 až 30 let, věk 31 až 40 let, věk 41 až 50 let, věk 51 až 60 let, věk nad 60 let	under18 f18t20 f21t25 f26t30 f31t40 f41t50 f51t60 over60
Zkrácená terapie	Rozdělení podle výběru zkrácené verze na: normální, 10denní příprava v jeden den	normal bujon

Pro většinu témat jsem zpracovala zvlášť tabulku s četnostmi a celkovým počtem uživatelů, kteří na otázku odpověděli, a doplnila o tabulku procentuálního rozložení četností v každé kategorii podle daných odpovědí. Tabulku s procentuálním rozložením jsem poté doplnila o jednoduchou vizualizaci v podobě podmíněného formátování – barevný odstín pozadí každé buňky, který je závislý na číselné hodnotě. Výsledky jsem poté vizualizovala v podobě sloupcových grafů.

Většinu témat jsem doplnila o χ^2 test nezávislosti mezi dvojicemi všech kategorií. Využila jsem pro tento účel funkci `chi2_contingency` z balíčku SciPy¹. Zobrazila jsem p-hodnoty pro všechny možné dvojice a barevně vyznačila hodnoty, u nichž lze zamítnout nulovou hypotézu na hladině významnosti $\alpha = 0,05$. Pro vyhodnocení jsem použila pouze takové výsledky, které splňují předpoklady provedení χ^2 testu nezávislosti – tedy výsledky, které mají nezávislá pozorování, četnosti jsou v každém poli vyšší než 1 a 80 % pozorovaných četností je v jedné kategorii vyšší než 5 [25].

¹ <https://www.scipy.org/>

2.2.1 Hlavní tabulka

Skript jsem vyhradila převážně pro základní zobrazení četností v jednotlivých kategoriích. Skript má spíše shrnující charakter. Z databáze jsem načetla tabulky četností zpracované přímo v SQL kódu. V SQL kódu jsem rozdělila četnosti podle kategorií z Tabulky 2.5. První tabulka je rozdělena na četnosti uživatelů v jednotlivých fázích (Tabulka 2.1) a celkový počet uživatelů v daných kategoriích, informace jsou doplněny v dalších tabulkách o procentuální rozložení a χ^2 test nezávislosti. Dále jsem se v tomto skriptu věnovala i četnostem uživatelů v jednotlivých fázích EE, která má 10 sezení, a FU, která má 15 sezení. Univerzální hypotézy pro χ^2 test nezávislosti:

- Nulová hypotéza: Průchod uživatelů terapiemi se neliší podle daných kategorií
- Alternativní hypotéza: Průchod uživatelů terapiemi se liší podle daných kategorií

2.2.2 Osobní profil

Tento skript několika způsoby zpracovává odpovědi v dobrovolném dotazníku (Tabulka 2.2, 2.3, 2.4). Pomocí SQL jsem načetla četnosti v kategoriích rozdělených podle otázek. V první části skript zpracovává přehledy četností v kategoriích podle Tabulky 2.5 v závislosti na odpovědích na každou otázku. Data jsem uspořádala pro lepší přehlednost do tzv. tab widgets pomocí ipywidgets². Každé otázce je věnována jedna záložka (Obrázek 2.3). Data jsem doplnila o procentuální rozložení a χ^2 test nezávislosti. Dále jsem se věnovala rozdílům v odpovědích mezi uživateli, kteří terapii dokončili a nedokončili. Rozdíly jsem vizualizovala v podobě tabulky procentuálního rozdělení pro všechny uživatele a pro kategorie z Tabulky 2.5 a v podobě sloupcových grafů.

² <https://ipywidgets.readthedocs.io/en/latest/>

	nums	allreg	men	women	lightsmokers	middlesmokers	heavysmokers
1. Cigarety	01	4444	2279	2165	1163	2363	921
2. Doutník	02	2182	1381	801	621	1142	421
3. Dýmka	03	760	580	180	208	402	150
4. Vodní	04	1189	1099	92	694	1221	375

Obrázek 2.3 Karta záložek

Univerzální hypotézy pro χ^2 test nezávislosti:

- Nulová hypotéza: Odpovědi na otázky se neliší podle daných kategorií
- Alternativní hypotéza: Odpovědi na otázky se liší podle daných kategorií

2.2.3 Odpovědi terapie

Tento skript je věnován zpracování odpovědí na otázky pokládané při terapii. Vzhledem k rozhodovacímu stromu a závislosti fáze terapie na počtu odpovědí jsem vybrala otázky, na které odpovědělo více jak 100 lidí. Každou otázku s předdefinovanými odpověďmi jsem zpracovala podle kategorií v Tabulce 2.5 a rozšířila kategorie o četnosti uživatelů, kteří terapii dokončili a nedokončili. Dále jsem vypočítala procentuální rozdělení pro všechny odpovědi a χ^2 test nezávislosti. Data jsem vizualizovala pomocí sloupcového grafu. Univerzální hypotézy pro χ^2 test nezávislosti:

- Nulová hypotéza: Odpovědi na otázky se neliší podle daných kategorií
- Alternativní hypotéza: Odpovědi na otázky se liší podle daných kategorií

Na otázky s otevřenou odpovědí jsem aplikovala základní metody NLP pro extrakci důležitých informací. Pro jednodušší zpracování a aplikaci funkce pro sentiment analýzu jsem otevřené odpovědi nechala přeložit do anglického jazyka nástrojem Google Translate³. Data jsem načetla do skriptu plně přeložená ve formě tabulky. Data jsem zpracovala pouze pro stejné kategorie jako u otázek s předdefinovanými odpověďmi.

³ <https://translate.google.cz/>

Nejčastěji použitá slova, dvojsloví, trojsloví

Před výpočtem nejčastěji použitých slov, dvojsloví a trojsloví jsem data nejdříve upravila. Z dat jsem odstranila interpunkci a převedla je na malá písmena. Dále jsem odstranila stop slova (stopwords) – slova, která nenesou samostatně žádný význam. Použila jsem balíček slov STOPWORDS pro anglický jazyk z knihovny WordCloud⁴, který jsem rozšířila o výrazy: t, m, s, 1, 2 ,3 ,4, 5, ll, will, ve, re, can, 10, 00. Často jsou to zkrácené tvary slov, které chybějí v původním balíčku, nebo číslovky, které jsou pro tento účel zbytečné. Pomocí funkcí z knihovny nltk⁵ jsem vytvořila tabulky nejčastěji použitých slov, dvojsloví a trojsloví pro kategorie: všichni uživatelé, uživatelé s nedokončenou a dokončenou terapií. Pomocí knihovny WordCloud jsem nejčastější slova vizualizovala.

Sentiment analýza

Díky přeloženému obsahu do anglického jazyka jsem mohla aplikovat sentiment analýzu pomocí funkce VADER⁶, která umí vyhodnotit polaritu vyjádření přirozeným jazykem. Navíc funkce slibuje vysokou úspěšnost u krátkých sdělení na sociálních sítích, takže by se měla hodit i k tomuto účelu. Funkce vrací hodnocení jako soubor čtyř skóre pro každý text. Skóre pozitivní, negativní a neutrální jsou v hodnotách od 0 do 1 a vyjadřují míru sentimentální složky. Celkové skóre je pak od -1 do 1 – pokud se skóre blíží k 1, je text spíše pozitivní, a naopak pokud k -1, je negativní. Pokud se blíží k 0, je text spíše neutrální. Funkcí jsem vypočítala skóre pro každou odpověď a zobrazila jsem průměr pro každou kategorii.

2.3 Shlukovací analýza

2.3.1 Příprava

Z databáze jsem stáhla data, kterým se u digitálního phenotypingu říká aktivní data – tedy data, která poskytuje uživatel v podobě odpovědí na dobrovolný dotazník. Navíc jsem stáhla jejich věk, počet cigaret denně a pohlaví uživatelů. Jako pasivní data jsem vybrala pro každého uživatele postup v aplikaci (fázi terapie), průměrnou denní hodinu, kdy uživatel dělal první 3 sezení, a průměrný počet znaků pro první 3 otevřené odpovědi při

⁴ https://amueller.github.io/word_cloud/index.html

⁵ <https://www.nltk.org/>

⁶ https://www.nltk.org/_modules/nltk/sentiment/vader.html

terapii. Vybrala jsem pouze ty uživatele, kteří měli vyplněny všechny informace, kromě pravidelně užívaných produktů, metody u minulého odvykání a doby trvání poslední abstinence – u těchto uživatelů jsem chybějící odpovědi nahradila hodnotou 0.

Všechna ordinální kategorická data jsem nahradila hodnotou korelující s významem. Takto jsem nahradila fázi terapie hodnotami 0-4 (START: 0, FIN: 4), délku trvání posledního pokusu hodnotami 1-7 (Méně než týden: 1, 1 rok a více: 7), bydliště podle počtu obyvatel hodnotami 1-5 (Více než 1 milion: 1, Obec do 2 tis: 5), vzdělání hodnotami 1-6 (základní: 1, Doktorské a vyšší: 6). Data, u kterých se odpovědi daly zastoupit binárními, jsem nahradila hodnotami 0 a 1, což platí pro pravidelné léky (Ne: 0, Ano: 1) a otázky na prodělání koronaviru, kde jsem obě pozitivní odpovědi spojila do jedné (Ano, léčím se: 1; Ano, již negativní: 1; Ne: 0). Odpovědi u otázek s mnohonásobnými odpověďmi jsem rozdělila na počet parametrů s binárními, podle počtu možných odpovědí. Každá odpověď v těchto otázkách získává svoji vlastní kategorii s binárními hodnotami symbolizujícími ano/ne u odpovědi každého uživatele. Takto byly rozděleny otázky na vyzkoušené produkty, pravidelně užívané produkty, důvody k abstinenci, metody u minulého ovykání, zaměstnání, zdravotní stav. U počtu pokusů jsem všem hodnotám nad 20 pokusů dala hodnotu 20.

Veškeré vizualizace a tabulky jsou pak zobrazeny s pracovními názvy, které zastupují jednotlivé otázky. Seznam těchto zástupných pracovních názvů je v Tabulce 2.6.

Tabulka 2.6 Pracovní zástupné názvy pro parametry

Otázka	Typ	Odpovědi	Pracovní názvy parametrů
Od kolika let pravidelně kouříš?	kvantitativní	číslo	smosince
Který z následujících tabákových a nikotinových produktů jsi už zkoušel?	binární	1. Cigarety 2. Doutník 3. Dýmka 4. Vodní dýmka 5. Šňupací tabák 6. Bezdýmový tabák 7. Žvýkáci tabák 8. Ecig s nikotinem 9. Ecig bez nikotinu 10. Ecig THC/CBD 11. Zahříváný tabák 12. Nikotinové sáčky	protried_cigarettes protried_ecigTheCbd protried_heatedTab protried_NiqSacks protried_cigar protried_pipe protried_hookah protried_snuff protried_nopuff protried_chewy protried_ecigNiq protried_ecigNoNiq

Otázka	Typ	Odpovědi	Pracovní názvy parametrů
Který z těchto produktů užíváš v současnosti (alespoň jednou týdně)?	binární	<ol style="list-style-type: none"> 1. Cigarety 2. Doutník 3. Dýmka 4. Vodní dýmka 5. Šňupací tabák 6. Bezdýmový tabák 7. Žvýkáci tabák 8. Ecig s nikotinem 9. Ecig bez nikotinu 10. Ecig THC/CBD 11. Zahřívaný tabák 12. Nikotinové sáčky 	<p>usage_cigarettes usage_ecigTheCbd usage_heatedTab usage_NiqSacks usage_cigar usage_pipe usage_hookah usage_snuff usage_nopuff usage_chewy usage_ecigNiq usage_ecigNoNiq</p>
Proč ses rozhodl přestat kouřit?	binární	<ol style="list-style-type: none"> 1. Zdraví 2. Rodina 3. Finance 4. Tlak partnera, dětí, přátel, rodiny 5. Zaměstnání 6. Zbavení závislosti 7. Obava z koronaviru 	<p>reasons_health reasons_family reasons_finances reasons_familyPressure reasons_job reasons_addiction reasons_covid</p>
Kolikrát v životě jsi už zkoušel přestat kouřit?	kvantitativní	číslo	attempts
Jak dlouho jsi vydržel nekouřit při posledním pokusu?	ordinální	1-7	lastduration
Jak ses o odvykání pokoušel?	binární	<ol style="list-style-type: none"> 1. Bez pomoci a bez léků 2. S náhradním nikotinem 3. S pomocí léků 4. S pomocí lékaře 5. S pomocí adiktologa, psychologa, terapeuta 6. S pomocí lékárníka 7. S pomocí telefonické linky 8. S pomocí webových infostránek 9. S pomocí mobilní aplikace 10. S pomocí alternativních metod 	<p>methods_no methods_niq methods_meds methods_medic methods_expert methods_pharm methods_tel methods_web methods_app methods_alter</p>
Kde momentálně žiješ?	ordinální	1-5	townsize
Jak jsi na tom se zaměstnáním?	binární	<ol style="list-style-type: none"> 1. Plný úvazek 2. Částečný úvazek 3. OSVČ živnostník 4. Nezaměstnaný 5. Prezenční student 6. Dálkové nebo kombinované studium 7. Rodičovská dovolená 8. Důchod 9. Pracovní neschopnost 	<p>job_fulltime job_parttime job_tradesman job_unemployed job_fullStudent job_combiStudent job_parent job_retirement job_disability</p>
Nejvyšší dosažené vzdělání?	ordinální	1-6	education
Kolik ti je let?	kvantitativní	číslo	age
Jaký je tvůj aktuální zdravotní stav?	binární	<ol style="list-style-type: none"> 1. Nic 2. Onemocnění dýchací soustav 	<p>health_nothing health_respiratory health_vascular</p>

Otázka	Typ	Odpovědi	Pracovní názvy parametrů
		3. Srdečně-cévní onemocnění 4. Diabetes 5. Duševní onemocnění 6. Nádorové onemocnění 7. Jiné	health_diabetes health_psychological health_cancer
Prodělal jsi onemocnění covid-19?	binární	Ne/Ano	covid
Užíváš pravidelně nějaké léky?	binární	Ne/Ano	meds
Pohlaví	binární	Muž/Žena	sex
Počet vykouřených cigaret denně	kvantitativní	číslo	nocig
Fáze terapie	ordinální	0-4	phase
Průměrný počet znaků	kvantitativní	číslo	character_count
Průměrná denní hodina sezení	kvantitativní	číslo	time

2.3.2 Výběr parametrů

Před provedením PCA jsem redukovala počet parametrů pomocí algoritmů na výběr parametrů, které nejlépe reprezentují celou datovou sadu. Celkem jsem použila čtyři algoritmy pro výběr parametrů a vybrala ty, které se ve výběrech objevily dvakrát a více. Podmínka pro výběr těchto algoritmů byla selekce příznaků bez učitele.

Prvním použitým algoritmem, jehož výsledky jsem použila i v explorační analýze, je analýza korelací mezi všemi dvojicemi parametrů pomocí Pearsonova korelačního koeficientu. Funkce je dostupná s knihovnou pandas⁷. Po vypočtení korelační matice jsem odstranila všechny parametry, které měly oproti ostatním korelace menší než -0,3 a větší než 0,3. Takto jsem vybrala parametry, které jsou na ostatních nejméně závislé a nekorelují.

Druhým algoritmem je analýza principiálních příznaků (PFA) podle [26], pomocí které jsem vybrala 20 hlavních faktorů. Třetím a čtvrtým algoritmem je použití Laplaceova skóre pro výběr parametrů [27] a spektrální selekce příznaků [28]. Oba algoritmy jsem použila v rámci funkcí v knihovně scikit-feature⁸. Oběma algoritmy jsem si nechala vybrat 20 nejlepších příznaků.

⁷ <https://pandas.pydata.org/docs/index.html>

⁸ <https://github.com/jundongl/scikit-feature>

2.3.3 Redukce dimenze

Díky numerickým vyjádřením všech příznaků jsem mohla aplikovat na data analýzu hlavních komponent PCA. Data s vybranými příznaky jsem nejdříve v každé kategorii normalizovala od 0 do 1. Po aplikaci PCA jsem si zobrazila graf závislosti kumulativní vyjádřené odchylky na počtu hlavních komponent a vybrala prvních 10 hlavních komponent, jejichž vyjádření zachovává 80 % informace. Prostor jsem tímto zmenšila na 10 dimenzí, které z 80 % vyjadřují původní prostor pomocí lineární aproximace díky metodě PCA.

2.3.4 Shlukovací analýza

Pro shlukovací analýzu jsem vybrala metodu k -means, kterou jsem shlukovala 10 hlavních komponent z PCA. K -means algoritmus jsem použila z knihovny scikit⁹. Před samotnou analýzou jsem projela algoritmem všechny možné počty shluků od 1 do 10 a vyjádřila si závislost residuální sumy čtverců na počtu shluků. Pomocí funkce `KneeLocator` z knihovny `kneed`¹⁰, která umí určit nejlepší možný počet shluků z residuální sumy čtverců. Funkce určí počet shluků podle místa největšího zakřivení v grafu – tedy nejmenší možná chyba spolu s nejmenším počtem shluků. Výsledkem shlukovací analýzy je zařazení dat do skupin sobě podobných. Pomocí funkcí v knihovně scikit jsem vypočítala Calinski-Harabasz skóre, které dává do poměru rozptyl uvnitř shluku a rozptyl mezi shluky. Čím vyšší je skóre, tím jsou shluky od sebe lépe oddělené [29]. Dále jsem nechala vypočítat celkové silhouette skóre, které je vypočítané pomocí vzdálenosti uvnitř shluku a vzdálenosti od nejbližšího shluku pro každý prvek. Skóre v tomto případě vrací střední hodnotu ze všech prvků [30].

2.3.5 Analýza fenotypů

Po provedení shlukovací analýzy jsou všechna původní data přiřazena k jednotlivým shlukům – fenotypům. Pro porovnání rozdílných vlastností napříč fenotypy jsem vizualizovala rozložení jednotlivých příznaků uvnitř shluků pomocí krabicových grafů. Kromě vyznačeného mediánu jsem přidala průměr, který lépe vystihuje charakter binárních dat. Pro přesnější určení, u kterých příznaků je mezi fenotypy odlišnost, porovnávala jsem shluky mezi sebou pomocí středních hodnot – jednosměrným ANOVA

⁹ <https://scikit-learn.org/stable/index.html>

¹⁰ <https://kneed.readthedocs.io/en/stable/>

testem jako je tomu například v [31]. Hladinu významnosti jsem stanovila na $\alpha = 0,05$. Univerzální hypotézy pro jednosměrný ANOVA test:

- Nulová hypotéza: Mezi shluky není v daném příznaku rozdíl
- Alternativní hypotéza: Mezi shluky je v daném příznaku rozdíl

Mezi příznaky jsou parametry, u kterých nelze předpokládat homoskedasticitu dat, což je předpoklad pro jednosměrný ANOVA test. Nahradila jsem u těchto příznaků použitý test jeho neparametrickou alternativou – Kruskalovým-Wallisovým testem, který testuje shodnost rozdělení [32]. Příznaky, kterých se to týká jsou ordinálního a kvantitativního typu (Tabulka 2.6). Hladinu významnosti jsem stanovila na $\alpha = 0,05$. Univerzální hypotézy pro Kruskalův-Wallisův test:

- Nulová hypotéza: Mezi shluky není v daném příznaku rozdíl
- Alternativní hypotéza: Mezi shluky je v daném příznaku rozdíl

Kruskalův-Wallisův test a jednocestnou ANOVU jsem spočítala pomocí funkcí v knihovně `scipy`. Po provedení obou testů na odpovídajících datech jsem příznaky seřadila podle p-hodnot od nejmenších hodnot – největších rozdílů mezi shluky. U příznaků, u kterých lze zamítnout nulovou hypotézu, jsem dopočítala post-hoc testy, které určují signifikantní rozdíly mezi konkrétními shluky. Pro příznaky, na kterých jsem testovala rozdíly jednocestným ANOVA testem, jsem vybrala Tukeyho HSD test, který nepotřebuje další korekce kvůli mnohonásobnému testování. Pro příznaky testované Kruskalovým-Wallisovým testem jsem použila Dunnův post-hoc test s Holm – Bonferroniho korekcí. Hladinu významnosti jsem u obou testů stanovila na $\alpha = 0,05$. Univerzální hypotézy pro post-hoc testy:

- Nulová hypotéza: Mezi daným a daným shlukem není v daném příznaku rozdíl
- Alternativní hypotéza: Mezi daným a daným shlukem je v daném příznaku rozdíl

Pro vyhodnocení dominantních vlastností jednotlivých fenotypů jsem vytvořila tabulku, mezi kterými shluky, v kterém parametru lze zamítnout nulovou hypotézu. Jako závěrečné vyhodnocení jsem pro každý fenotyp vytvořila tabulku vlastností, kterými se liší:

1. Všechny fenotypy mezi sebou

2. Pouze zkoumaný fenotyp od ostatních fenotypů (je zamítnuta nulová hypotéza mezi zkoumaným fenotypem a ostatními)
3. Pouze zkoumaný fenotyp od vzdálenějšího fenotypu (nulová hypotéza zamítnuta pouze mezi zkoumaným fenotypem a nejvzdálenějším fenotypem ve střední hodnotě)

Parametry jsem seřadila podle důležitosti podle původních p-hodnot z testů středních hodnot.

3 Výsledky a diskuse

3.1 Explorační analýza

3.1.1 Hlavní tabulka

Údaje pro tuto analýzu jsou platné k datu 31. 7. 2021. V Tabulce 3.1 se nachází přehled četností uživatelů v kategoriích podle fáze terapie. Pokud uživatel dále v odvykání nepokračuje, jeho aktuální fáze zůstane stejná a díky tomuto přehledu je možné zjistit, kdy je pacient nejméně motivován, což je jasně vidět v Tabulce 3.2, kde je pozadí buněk tabulky barevně podmíněno hodnotou, kterou buňka obsahuje. V tabulce je vidět, že většinu uživatelů nepřesvědčí fáze EE k dalšímu pokračování. Tabulka 3.1 má poslední sloupec věnovaný celkovému součtu uživatelů v daných kategoriích. V kategorii all tedy najdeme celkový stav uživatelů, kteří si aplikaci stáhli. V Tabulce 3.2 sloupec ALL obsahuje hodnoty, které srovnávají procentuální rozložení pro kategorie pohlaví, míru závislosti, věk a výběr zkrácené délky terapie. Vidíme, že více uživatelů jsou muži, nejvíce uživatelů je ve věku 31-40 let, největší počet je středních kuřáků a více lidí si volí normální nezkrácenou verzi terapie. V Tabulce 3.3 jsou výsledné p-hodnoty pro χ^2 test normality – žlutě jsou označena pole, kde se podle hladiny významnosti dá určit, že rozdělení v daných kategoriích jsou závislé na kategorii, což platí pro téměř všechny dvojice kategorií. Zajímavé nezávislosti jsou například u středních a těžkých kuřáků nebo u blízkých kategorií věku. Uživatelé nad 60 let a pod 18 let věku jsou v průběhu terapie neporovnatelní kvůli nízké četnosti. Celkově u kategorie věk je méně uživatelů z důvodu nepovinného vyplnění osobního profilu, kde se dotaz na věk nachází.

Tabulka 3.1 Přehled četností v kategoriích podle fáze

	START	EE	EQ	FU	FIN	ALL
all	5049	15824	706	604	511	22694
men	2959	8732	330	273	299	12593
women	2088	7072	376	331	212	10079
light	1413	4039	259	237	161	6109
middle	2391	8592	313	260	264	11820
heavy	1047	3516	135	108	117	4923
under18	20	261	17	17	0	315
f18t20	34	458	31	47	8	578
f21t25	40	719	64	88	40	951
f26t30	67	669	57	91	51	935
f31t40	65	807	71	98	76	1117
f41t50	36	351	23	41	54	505
f51t60	15	109	4	7	8	143
over60	1	23	0	1	1	26
bujon	1568	5429	358	316	122	7793
normal	3282	6417	286	269	151	10405

Tabulka 3.2 Přehled procentuálního zastoupení v kategoriích podle fáze

	START	EE	EQ	FU	FIN	ALL
all	22.2	69.7	3.1	2.7	2.3	
men	23.5	69.3	2.6	2.2	2.4	55.5
women	20.7	70.2	3.7	3.3	2.1	44.4
light	23.1	66.1	4.2	3.9	2.6	26.9
middle	20.2	72.7	2.6	2.2	2.2	52.1
heavy	21.3	71.4	2.7	2.2	2.4	21.7
under18	6.3	82.9	5.4	5.4	0	6.9
f18t20	5.9	79.2	5.4	8.1	1.4	12.6
f21t25	4.2	75.6	6.7	9.3	4.2	20.8
f26t30	7.2	71.6	6.1	9.7	5.5	20.5
f31t40	5.8	72.2	6.4	8.8	6.8	24.4
f41t50	7.1	69.5	4.6	8.1	10.7	11.1
f51t60	10.5	76.2	2.8	4.9	5.6	3.1
over60	3.8	88.5	0	3.8	3.8	0.6
bujon	20.1	69.7	4.6	4.1	1.6	42.8
normal	31.5	61.7	2.7	2.6	1.5	57.2

Tabulka 3.3 P-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

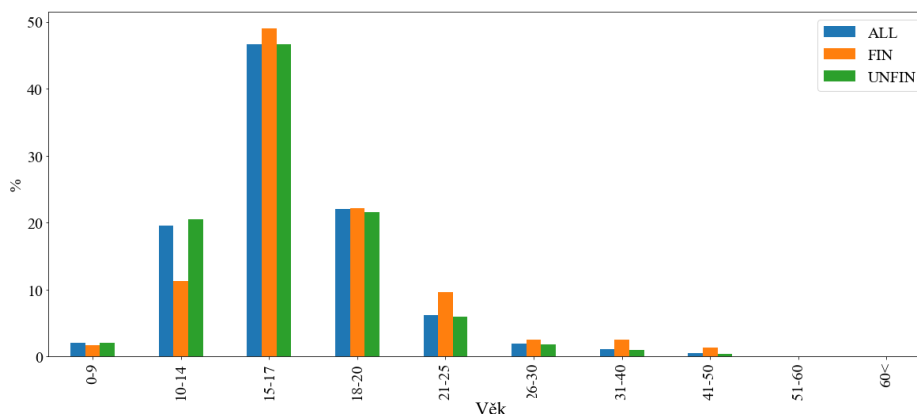
	women	middle	heavy	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	normal
men	0.000										
light		0.000	0.000								
middle			0.553								
under18				0.142	0.000	0.000	0.000	0.000	0.000	0.007	
f18t20					0.009	0.000	0.000	0.000	0.003	0.489	
f21t25						0.036	0.043	0.000	0.003	0.537	
f26t30							0.473	0.005	0.101	0.403	
f31t40								0.040	0.042	0.436	
f41t50									0.102	0.340	
f51t60										0.671	
bujon											0.000

3.1.2 Osobní profil

Veškeré výsledné tabulky a vizualizace jsou kompletní v Příloze A. Výsledky v kapitole Osobní profil jsou platné k datu 31. 7. 2021. Údaje byly sbírané od 30. 5. 2020 do 31. 7. 2021.

Od kolika let pravidelně kouříš?

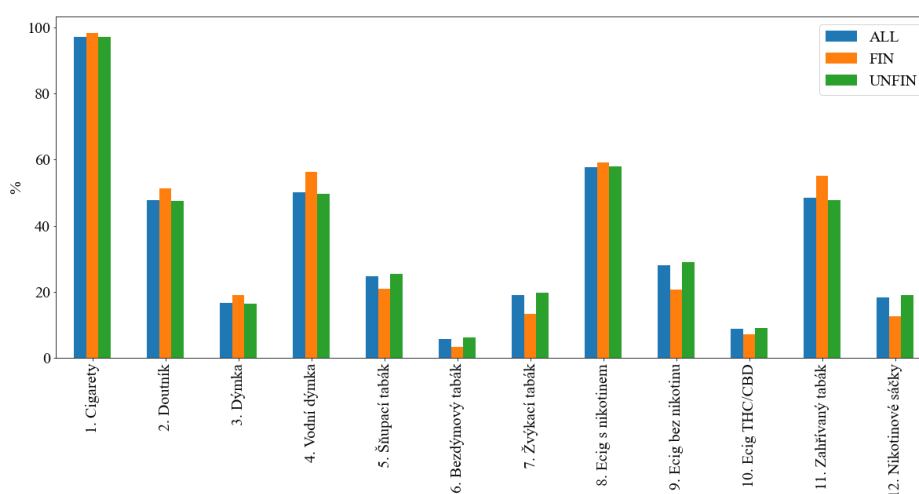
Na tuto otázku odpovědělo 4581 lidí a odpovědí bylo číslo. 68,3 % uživatelů začalo kouřit před 18. rokem a nejvíce lidí začalo kouřit ve věku 15-17 let. Všechny kategorie nelze porovnat na nezávislost. U skupiny, která nedokončila sezení, je vyšší výskyt lidí, kteří začali kouřit mezi 10-14 lety (21 % oproti 11 %), který je vidět na Obrázku 3.1.



Obrázek 3.1 Sloupcový graf procentuálního rozložení věku začátku kouření

Který z následujících tabákových a nikotinových produktů jsi už zkoušel?

Na otázku odpovědělo 4598 lidí a uživatel vybírá jako odpověď několik možností. 97 % uživatelů zkoušelo cigarety. Další nejčastější zkoušené produkty (nad 40 % uživatelů) jsou nikotinové e-cigarety, vodní dýmka, zahříváný tabák, doutník. Nad 60 % uživatelů pod 18 let uvedlo, že zkoušeli výrobky: cigarety, doutník, nikotinové e-cigarety, beznikotinové e-cigarety, nikotinové sáčky – zkušenosti s těmito výrobky s věkem klesají. Rozdělení se mezi kategoriemi neliší u středních a těžkých kuřáků a u uživatelů ve věku mezi 41 a 60 lety. Uživatelé nad 60 let nelze porovnat. Rozdělení podle uživatelů s dokončenou a nedokončenou terapií se liší (Obrázek 3.2).



Obrázek 3.2 Sloupcový graf pro procentuální rozložení u vyzkoušených produktů

Který z těchto produktů užíváš v současnosti (alespoň jednou týdně)?

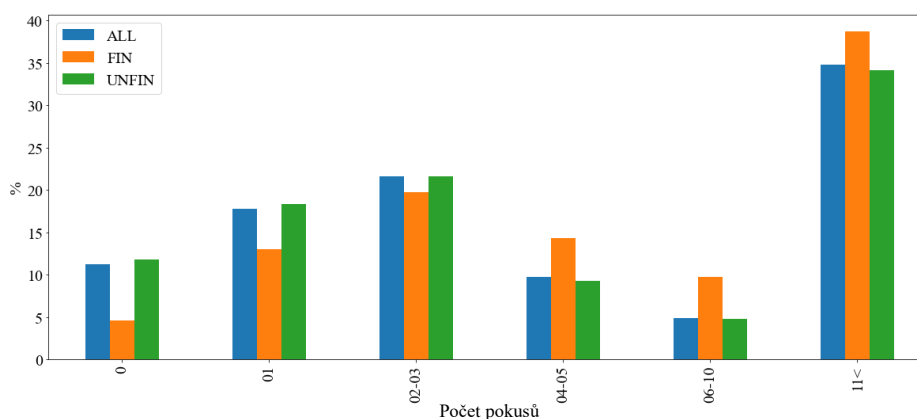
Na tuto otázku odpovědělo 2598 uživatelů a ptá se pouze těch, kteří v současnosti neužívají nikotinové výrobky. Uživatel vybírá několik odpovědí ze seznamu možností. Nejvíce užívají uživatelé pravidelně cigarety (84,9 %) napříč všemi kategoriemi, dále zahříváný tabák (18,6 %) a e-cigarety s nikotinem (10 %). Zahříváný tabák užívají více ženy než muži (22 % oproti 15 %) a ženy užívají méně cigarety (82 % oproti 88 %). Užívání zahříváného tabáku stoupá s mírou závislosti (13 % - lehčí kuřáci oproti 22 % - těžcí kuřáci). Obliba nikotinových sáčků je nejvyšší u uživatelů pod 18 let (19 %), zahříváného tabáku u 21-25letých uživatelů (22 %), e-cigaret s nikotinem u mladších 18 let (19 %). Mezi skupinami jsou rozdíly kromě blízkých věkových skupin vyjma skupin pod 20 let. Uživatelé nad 30 let nelze porovnat. Mezi úspěšnými a neúspěšnými uživateli není rozdíl.

Proč ses rozhodl přestat kouřit?

Na otázku odpovědělo 4582 uživatelů. Otázka nabízí několik důvodů a uživatel jich může vybrat více. 87 % uživatelů uvádí důvod zdraví, 68,2 % uvádí zbavení závislosti a 64,9 % finance. Nejméně uvádějí uživatelé obavu z koronaviru, zaměstnání (oboje 5,1 %) nebo jiný důvod (6,6 %). Mezi kategoriemi není rozdíl mezi středními a těžkými kuřáky a mezi některými blízkými kategoriemi a uživatele nad 60 let a s normální verzí terapie nelze porovnat. Rozdíl není ani mezi úspěšnými a neúspěšnými uživateli.

Kolikrát v životě jsi už zkoušel přestat kouřit?

Na tuto otázku odpovědělo 4560 uživatelů. Nejvíce lidí se snažilo přestat více jak jedenáctkrát (34,8 %), druhá nejpočetnější skupina 2-3 x (21,6 %). Více jak 11x se nejvíce snažili přestat lehčí kuřáci (40 %) a lidé, kteří zvolili krátkou variantu terapie (74 %). V populaci starší 31 let nejsou signifikantní rozdíly v odpovědích stejně tak u uživatelů ve věku 18-25 let. Osoby nad 60 let a s krátkou variantou terapie nelze porovnat. Mezi uživateli, kteří terapii dokončili a nedokončili je signifikantní rozdíl. Uživatelé, kteří terapii dokončili mají v procentuálním zastoupení převahu od 4 pokusů výše (Obrázek 3.3).



Obrázek 3.3 Sloupcový graf procentuálního rozložení počtu pokusů

Jak dlouho jsi vydržel nekouřit při posledním pokusu?

Na tuto otázku odpovědělo 3555 uživatelů a odpovědi byla jedna odpověď z možností. 27,7 % uživatelů vydrželo méně než týden. Druhou nejčastější odpovědí je týden až měsíc (25,1 %). S vyšším věkem stoupá procentuální zastoupení uživatelů, kteří abstinovali více než rok. Mezi skupinami nejsou rozdíly u uživatelů mladších 20 let, 31-60 let věku a mezi

krátkou a normální variantou terapie. Mezi uživateli s dokončenou a nedokončenou terapií není rozdíl.

Jak ses o odvykání pokoušel?

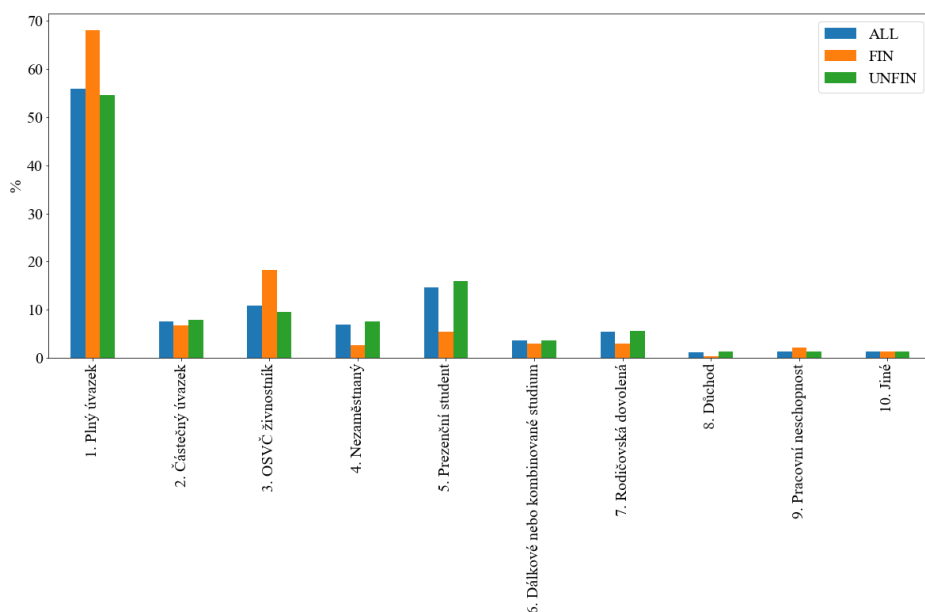
Na otázku odpovědělo 3493 uživatelů a ptá se pouze těch, kteří se o odvykání už pokoušeli. 70,5 % uživatelů se pokoušeli odvyknout bez pomoci a léků, 17 % s náhradním nikotinem a 6,8 % s pomocí mobilní aplikace. U starších uživatelů přibývají pokusy s náhradním nikotinem (30 % u věku 41-50 let) a s pomocí léků (11 % u věku 41-50 let). Kvůli nedostatku četností lze podle věku porovnat pouze uživatele ve věku 31-50, u kterých je rozdíl, a taktéž nelze porovnat lehké kuřáky. Mezi ostatními skupinami jsou rozdíly. Taktéž je rozdíl mezi uživateli ve FIN fázi a ostatními. Uživatelé ve FIN fázi mají menší procentuální zastoupení u odpovědi bez pomoci a léků, a naopak mají mírně vyšší zastoupení u ostatních metod.

Kde momentálně žiješ?

Na otázku odpovědělo 4538 uživatelů a odpovědí byla jedna z možností. Nejvíce uživatelů žije ve městě nad 10000 obyvatel (24,8 %). Signifikantní rozdíly jsou pouze mezi mladšími věkovými kategoriemi pod 20 let a staršími a mezi lidmi s kratší verzí a normální verzí terapie. Mezi uživateli v dokončené fázi a nedokončené fázi není signifikantní rozdíl, pouze procentuální zastoupení uživatelů v dokončené fázi převažuje pro kategorii bydliště s počtem obyvatel větším než 1 milion.

Jak jsi na tom se zaměstnáním?

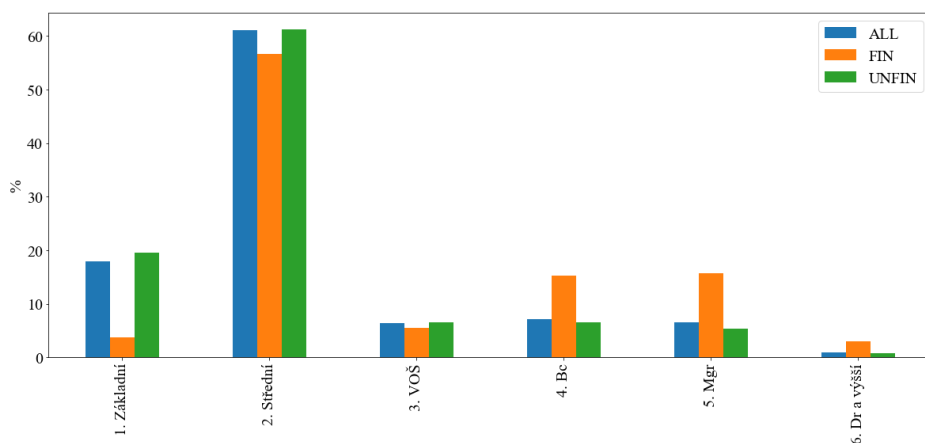
Na otázku odpovědělo 4547 lidí a odpovědí bylo více možností. 55,9 % uživatelů pracuje na plný úvazek, 14 % uživatelů jsou prezenční studenti a 10,8 % živnostníci. Mezi skupinami jsou rozdíly v každé kategorii. Nelze porovnat uživatele pod 18 a nad 50 let. Mezi uživateli s ukončenou a neukončenou terapií je rozdíl. U uživatelů s plným úvazkem, živnostníků a pracovně neschopnými je vyšší zastoupení úspěšných uživatelů, což je vidět na Obrázku 3.4.



Obrázek 3.4 Sloupcový graf procentuálního rozložení u zaměstnání

Nejvyšší dosažené vzdělání?

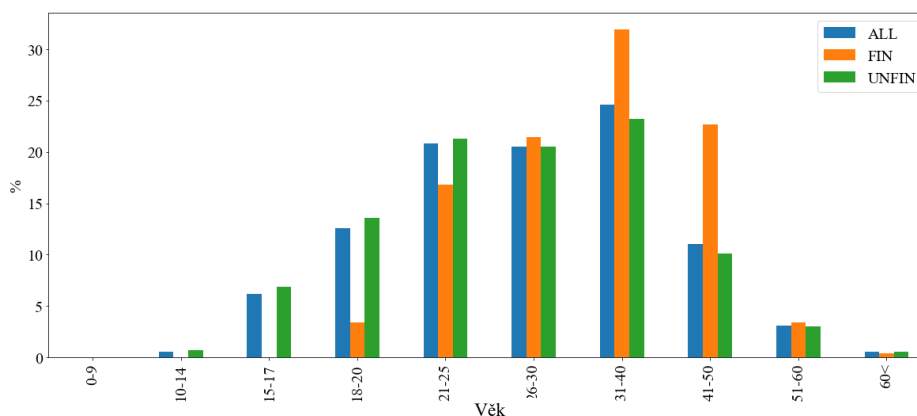
Na otázku odpovědělo 4525 uživatelů a odpovědí byla jedna z možností. Největší zastoupení ve všech kategoriích má středoškolské vzdělání (všichni uživatelé 61 %), kromě mladších uživatelů pod 18 let. Mezi všemi kategoriemi jsou signifikantní rozdíly, kromě středních a těžkých kuřáků. Mezi úspěšnými a neúspěšnými uživateli je signifikantní rozdíl. Procentuální zastoupení úspěšných je vyšší u vysokoškolsky vzdělaných než u uživatelů s nižším vzděláním (Obrázek 3.5).



Obrázek 3.5 Sloupcový graf procentuálního rozložení u vzdělání

Kolik ti je let?

Na tuto otázku odpovědělo 4619 uživatelů a odpovědí bylo číslo. Nejvíce uživatelů je ve věku 31-40 let (24,6 %), ale pokud jsou brány v potaz celé dekády, nejvíce uživatelům je 21-30 let (41,3 %). Nejvíce lehkých kuřáků je věku 21-25 let (22 %) a nejvíce středních a těžkých kuřáků je ve věku 31-40 let (26 % a 31 %). Lidé ve věku 21-25 let nejčastěji volili kratší variantu aplikace a normální variantu volili lidé nejčastěji ve věku 31-40 let. Test nezávislosti byl proveden pouze od kategorie věku 10 a výše. Rozdíly jsou mezi všemi kategoriemi. Mezi úspěšnými a neúspěšnými uživateli je ve věku signifikantní rozdíl (porovnávané skupiny jsou od věku 18 výše kvůli nulovým četnostem u nižších věkových kategoriích). Od věku 26 let je procentuální zastoupení u úspěšných vyšší, což je vidět na Obrázku 3.6.



Obrázek 3.6 Sloupcový graf procentuálního rozložení u vzdělání

Jaký je tvůj aktuální zdravotní stav?

Na otázku odpovědělo 4619 uživatelů a na výběr je k zaškrtnutí více možností. 70,6 % uživatelů je zdravých. Více žen (17 %) má duševní onemocnění oproti mužům (7 %). Mezi 20letými není rozdíl, uživatelé pod 18 let a nad 50 let jsou neporovnatelní. Mezi uživateli v konečné fázi a v nedokončené fázi není rozdíl.

Prodělal jsi onemocnění covid-19?

Na otázku odpovědělo 4592 uživatelů a odpovědí byla jedna z možností. 88,5 % uživatelů ve chvíli, kdy vyplňovali dotazník, neprodělali onemocnění Covid-19. Signifikantní rozdíl mezi kategoriemi je pouze mezi těžkými a lehkými kuřáky, mezi normální a krátkou variantou terapie a uživatelů ve věku 21-25 let oproti skupinám 31-40 let a 41-50 let. Uživatelé pod 18 a nad 60 let jsou neporovnatelní. Mezi úspěšnými a neúspěšnými uživateli není v této kategorii rozdíl.

Užíváš pravidelně nějaké léky?

Na otázku odpovědělo 4587 uživatelů a odpovědí byl výběr ze dvou možností. Většina uživatelů (72,2 %) neužívá žádné léky, ale od věku 51 výše se odpovědi obrací ve prospěch pozitivní odpovědi. Mezi kategoriemi je signifikantní rozdíl mezi středními a těžkými kuřáky, mezi muži a ženami, mezi krátkou a normální verzí i mezi většinou věkových kategorií. Mezi úspěšnými a neúspěšnými lidmi není rozdíl.

Jak dlouho uživatel kouří

Data byla vypočtena na základě věku uživatele a věku, kdy uživatel začal kouřit. Na základě těchto údajů je počet odpovědí 4434. Nejvyšší procentuální zastoupení je u uživatelů, kteří kouří 11-20 let (34,5 %). Lehčí kuřáci mají vyšší zastoupení u kategorií menších než 11 let (27 % u 11-20 let), střední a těžcí kuřáci mají v kategorii 11-20 let vyšší zastoupení (37 % a 38 %). S počtem cigaret denně taktéž stoupá zastoupení uživatelů u hodnoty přes 20 let (lehčí kuřáci 11 %, těžcí kuřáci 30 %). S věkem taktéž roste délka období, kdy uživatel kouřil. Mezi kategoriemi nejsou žádné rozdíly a uživatelé nad 51 let a pod 20 let jsou neporovnatelní.

3.1.3 Odpovědi terapie

Subjektivní výběr několika otázek z EE fáze odvykání, kde na otázky odpovědělo nejvíce lidí. Výsledky jsou platné k datu 4. 8. 2021. Více výsledných tabulek a vizualizací se nachází v Příloze A.

EE 1 Ted' je řada na tobě. Kterým z těchto témat by ses chtěl zabývat během toho, co se budeš snažit přestat kouřit?

Uživatelé se nabízí několik nejčastějších témat. Téma, které zajímá nejvíce lidí je ohledně zvládnutí stresu (33,4 %). U úspěšných osob je nejčastějším tématem zvládnutí smutku,

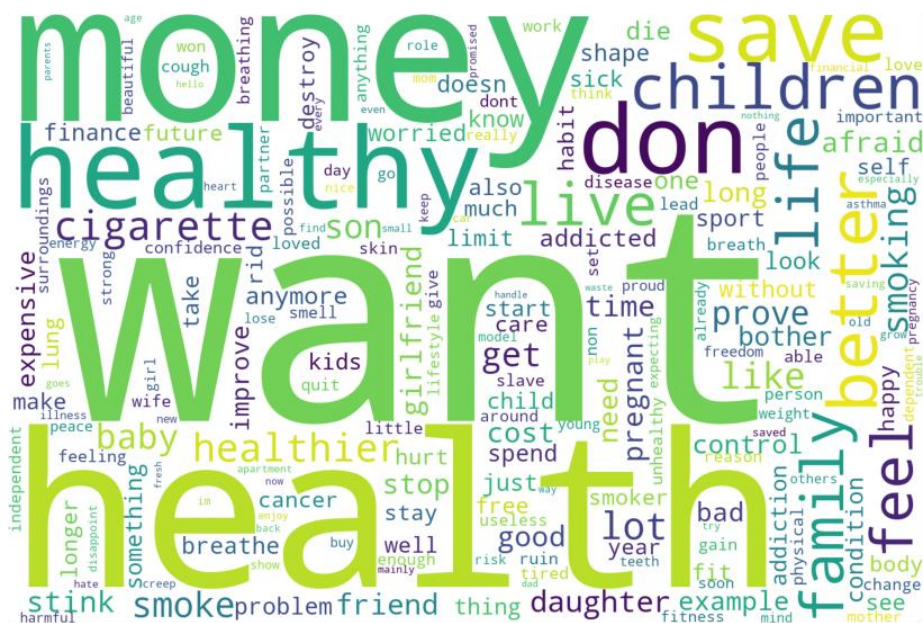
zlosti a trémy (31,8 %). Jak nepřibrat na váze nejvíce zajímá ženy (30,5 %) a uživatele nad 60 let (36 %). Téma ohledně identity ve společnosti zajímá nejvíce uživatele pod 18 let (26,8 %). Mezi kategoriemi jsou signifikantní rozdíly kromě některých věkově blízkých skupin.

EE 2 Několik představ, které lidi uvádějí. Které z nich ti sedí?

U této otázky je nabídnuto několik častých představ o změnách v životě, které mají lidé o abstinenci. Nejvíce odpovědí je u představy ohledně ušetření peněz (18,6 %), druhá nejčastější je odpověď ohledně zdraví (16,9 %) a třetí je na téma lepší kondice a nezadýchávání (16,5 %). Signifikantní rozdíly nejsou u věkových kategorií pod 25 let u a kategorií nad 26 let.

EE 2 Jaký je nejdůležitější důvod, který tě k tomu vede?

Otázka s otevřenou odpovědí, která se ptá na nejdůležitější důvod, který uživatele vede k odvykání. Úspěšní uživatelé měli v průměru více obsáhlé odpovědi. Odpovědi jsou spíše neutrální a nejčastější důvody jsou: touha odvyknout, zdraví, peníze, děti, rodina, život apod. Vizualizace nejčastějších slov je na Obrázku 3.7.



Obrázek 3.7 Vizualizace odpovědí pomocí wordcloud

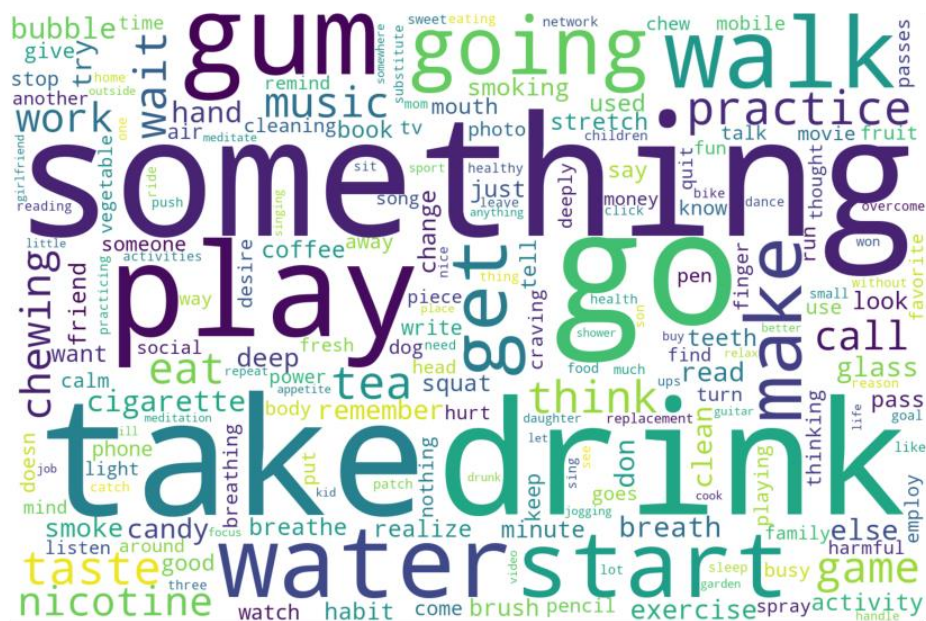
EE 3 Vyber triky, které by ti mohly pomoci

Podívej se na seznam. Když některý označíš, objeví se k němu krátké vysvětlení. Pokud tě trik zaujme, nechej ho zatržený. Když ne, zatržení zruš.

Uživatel vybírá několik tipů, které si myslí, že by mu mohly pomoci při chuti na cigaretu. Nejvíce zodpovídanou odpovědí je ohledně změny chuti v ústech (12,8 %), druhou je uklidnění se (12,2 %). Nejméně zodpovídané odpovědi jsou hledání podpory u přátel a rodiny (5,6 %) a nikotinové náhražky (6,5 %). Signifikantní rozdíly jsou pouze mezi uživateli pod 18 let a jinými věkovými skupinami.

EE 3 Když dostanu chuť kouřit, nezapálím si, ale udělám tohle:

Odpověď na tuto otázku je formou volného textu. Uživatelé ve FIN fázi mají průměrně větší počet znaků a odpovědi jsou spíše neutrální. Nejčastějšími odpověďmi jsou: změnit činnost, pití, hraní her, procházka, cvičení, jídlo, hudba, žvýkání, nádech. Vizualizace je na Obrázku 3.8.



Obrázek 3.8 Vizualizace odpovědí pomocí wordcloud

EE 4 Jak se za své odvykání odměníš?

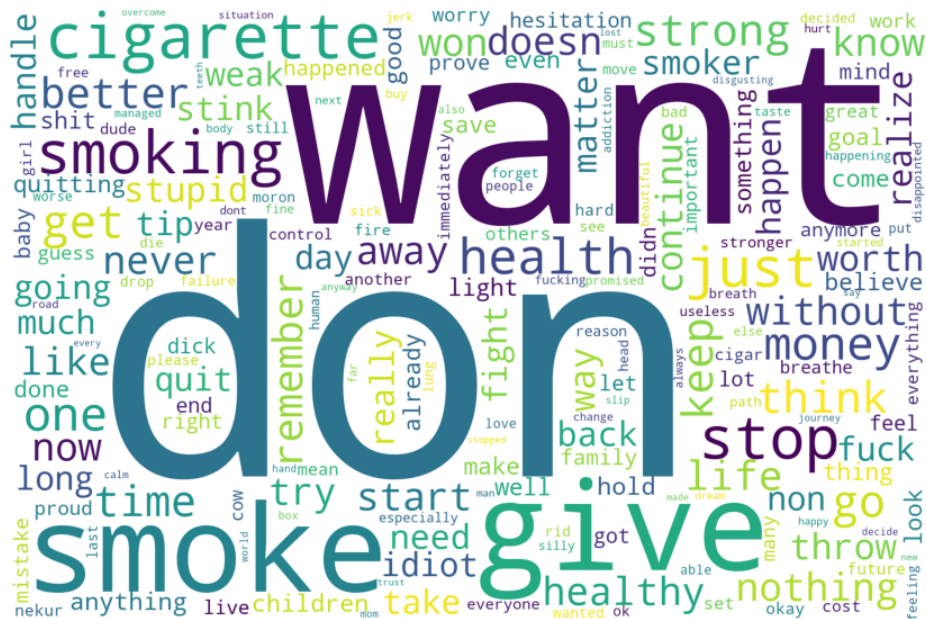
Uživatel vypisuje možné odměny jako volnou odpověď. Uživatelé, kteří terapii dokončili, se v průměru více rozepisují než ostatní a odpovědi jsou spíše neutrální. Nejčastější odměny: pořízení nové věci, ušetřené peníze, dávat peníze stranou, neví. Vizualizace je na Obrázku 3.9



Obrázek 3.10 Vizualizace odpovědí pomocí wordcloud

EE 5 Pokud zaváhám a zapálím si, chci slyšet toto:

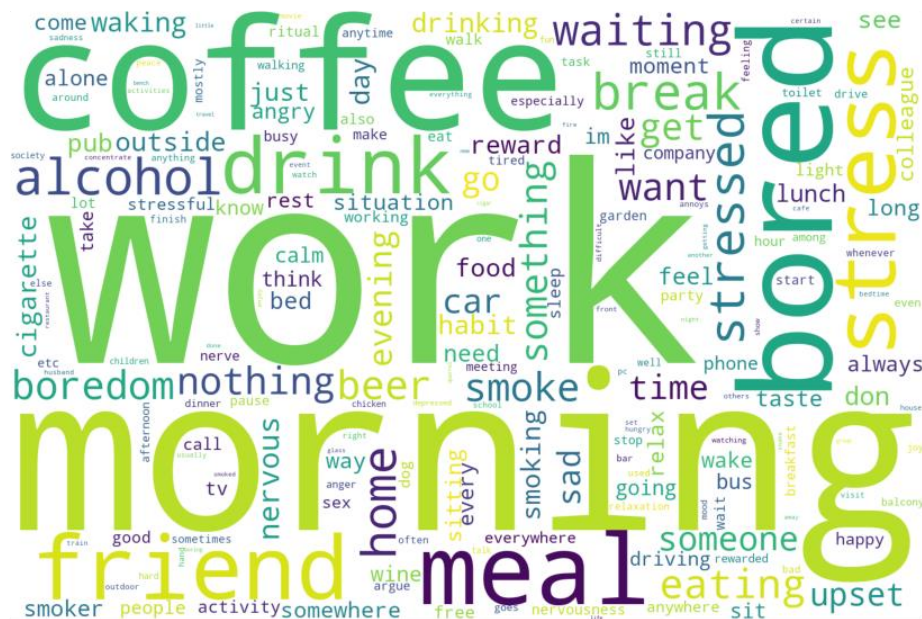
Odovědí je volná odpověď, která je vzkazem pro uživatele do budoucna. Uživatelé ve FIN fázi mají v průměru více obsáhlou odpověď než ostatní a odpovědi jsou spíše neutrální. Nejčastější slova ze vzkazů: nechtít, nevzdat, zdraví, peníze, síla, pamatovat, lepší, vyhodit cigarety, nevádí, přestat. Vizualizace nejčastějších slov je na Obrázku 3.11.



Obrázek 3.11 Vizualizace odpovědí pomocí wordcloud

EE 6 Chuť kouřit dostávám kdy – kde – proč:

Odpovědi na tuto otázku je volná odpověď a odpovědi jsou spíše neutrální. Nejčastěji použitá slova: práce, ráno, káva, nuda, stres, jídlo, přátelé, alkohol, pití, domov, přestávka, čekání, pivo, přestávka v práci, ranní káva, vidět někoho kouřit. Vizualizace nejčastějších slov je na Obrázku 3.12.



Obrázek 3.12 Vizualizace odpovědí pomocí wordcloud

EE 6 Zamysli se, co se změnilo. V čem jsou podmínky pro tento pokus jiné než minule?

Otázka se ptá uživatelů, kteří se pokoušeli odvyknout již dříve, ale nepovedlo se jim to. Odpověď je volná. Úspěšní uživatelé mají průměrně vyšší počet znaků než ostatní a odpovědi jsou spíše neutrální. Nejčastější odpovědi: chtít, zdraví, starší, čas, neví, motivace, peníze, nic, opravdu chtít, větší motivace. Vizualizace nejčastějších slov je na Obrázku 3.13.



Obrázek 3.13 Vizualizace odpovědí pomocí wordcloud

EE 7 Vyber prosím situaci, která se tě skutečně nejvíc podobá:

Vyber si situaci, která nejlépe odpovídá té, během které jsi to nevydržel a začal jsi znovu kouřit.

Otázka se ptá uživatelů, kteří se již v minulosti pokusili přestat. Nejvíce uváděnou odpovědí je odpověď, že se uživatel cítil osamělý (27,3 %), druhou nejčastější je trávení času s kolegy a přáteli (25,4 %) a třetí je večírek nebo oslava (22,5 %). Trávení času s přáteli a kolegy odpovědělo 50 % uživatelů pod 18 let. 21 % uživatelů ve FIN fázi, jako druhou nejčastější odpověď, uvedlo těžké životní období nebo krizi. Signifikantní rozdíly nejsou pouze mezi několika věkovými kategoriemi. Uživatelé pod 18 let a nad 50 let jsou neporovnatelní.

EE 7 Které situace považuješ za nejvíc rizikové?

Označ jednu nebo více situací, které mohou být nejnáročnější pro tebe. Neber si to ale příliš a nezaškrtávej je všechny! Vyber jen ty nejhroživější.

Nejčastější odpovědí byla situace, kdy jsou uživatelé s přáteli, kteří kouří (10,3 %), dále když se rozčílí nebo pohádají (9,6 %) a třetí nejčastější je, když se nutí například při

zodpovídanými hodnotami jsou (pod 2 %): zvědavost a kreativita, umírněnost, oblíbenost, tradice. Rozdíly nejsou pouze mezi některými věkovými kategoriemi a uživatelé nad 50 let jsou neporovnatelní.

EE 9 Zkus popsat, jak kouření ohrožuje tvoje hodnoty:

Odpověď na tuto otázku je volná. Uživatelé v konečné fázi mají průměrný počet znaků vyšší než ostatní a odpovědi jsou spíše neutrální. Nejčastější odpovědi jsou: zdraví, rodina, touha, peníze, svoboda, láska, život, mír, dobré, čas, vnitřní klid, sebeúcta, neví, ohrožení zdraví, dobrá nálada, svoboda pohybu, závislost je opak svobody. Vizualizace nejčastějších slov je na Obrázku 3.15.



Obrázek 3.15 Vizualizace odpovědí pomocí wordcloud

3.1.4 Korelace mezi příznaky

V Tabulce 3.4 jsou dvojice příznaků u kterých je Pearsonův korelační koeficient vyšší než 0,3 nebo nižší než -0,3.

Tabulka 3.4 Korelace vyšší než 0,3 nebo nižší než -0,3 mezi příznaky

Příznak 1	Příznak 2	Korelace
age	smosince	0,42
age	job_fulltime	0,34
meds	health_diabetes	0,31
lastduration	methods_no	0,35
protried_NiqSacks	job_combiStudent	0,35
protried_ecigNoNiq	job_combiStudent	0,31
protried_ecigThcCbd	usage_heatedTab	0,38
protried_heatedTab	usage_heatedTab	0,35
protried_NiqSacks	protried_snuff	0,4
protried_NiqSacks	protried_chewy	0,4
protried_NiqSacks	protried_ecigNoNiq	0,36
protried_NiqSacks	usage_NiqSacks	0,32
protried_cigar	protried_pipe	0,4
protried_cigar	protried_hookah	0,44
protried_cigar	protried_snuff	0,42
protried_cigar	protried_chewy	0,3
protried_cigar	protried_ecigNiq	0,35
protried_cigar	protried_ecigNoNiq	0,3
protried_pipe	protried_snuff	0,35
protried_hookah	protried_snuff	0,39
protried_hookah	protried_ecigNiq	0,31
protried_hookah	protried_ecigNoNiq	0,32
protried_snuff	protried_chewy	0,49
protried_snuff	protried_ecigNoNiq	0,35
protried_chewy	protried_ecigNoNiq	0,31
protried_ecigNiq	protried_ecigNoNiq	0,42
usage_nopuff	usage_chewy	0,45
age	protried_NiqSacks	-0,37
age	protried_hookah	-0,35
age	protried_snuff	-0,34
age	protried_chewy	-0,36
age	protried_ecigNoNiq	-0,38
age	job_combiStudent	-0,45
meds	health_nothing	-0,48
job_fulltime	job_combiStudent	-0,47
health_nothing	health_respiratory	-0,47
health_nothing	health_vascular	-0,4
health_nothing	health_psychological	-0,58
methods_no	methods_niq	-0,47

3.1.5 Shrnutí a diskuse

Explorační analýza dala za vznik mnoha tabulkám a grafům (z nichž většina se nachází v Příloze A). Z celkového počtu přes 22 tisíc uživatelů, kteří aplikaci někdy používali,

celkem přes 4000 vyplnilo osobní profil. Otázky v osobním profilu jsou nejvíce konzistentní, co se týče počtu uživatelů. U otázek přímo při terapii má každý uživatel svou vlastní cestu terapií, a ne všichni odpoví na vše. Skoro u všech otázek byl velmi malý počet uživatelů nad 50 nebo 60 let a také mladých pod 18 let. Z toho důvodu nelze tyto kategorie porovnat u většiny otázek, protože nesplňují požadavky pro χ^2 test normality. Mezi většinou kategorií je v odpovědích signifikantní rozdíl. Co se týče volných odpovědí, u každé otázky měli uživatelé, kteří terapii dokončili, větší průměrný počet znaků u svých odpovědí. Sentiment skóre je u většiny otázek spíše neutrální. Otázka s nejvíce pozitivním průměrným skóre (0,39) byla otázka na zpětnou vazbu týkající se terapií. Velmi časté slovo, které se v odpovědích objevovalo, bylo slovo „chci“. Z nejčastějších slov uživatelů je vidět, že uživatelé abstinenci myslí vážně a spousta z nich myslí na podobné věci a má podobné problémy.

3.2 Digitální fenotypy

Za účelem sestavení digitálních fenotypů byla vybrána data od 1593 uživatelů, která splňovala podmínku kompletní sady u všech příznaků.

3.2.1 Výběr parametrů pro shlukování

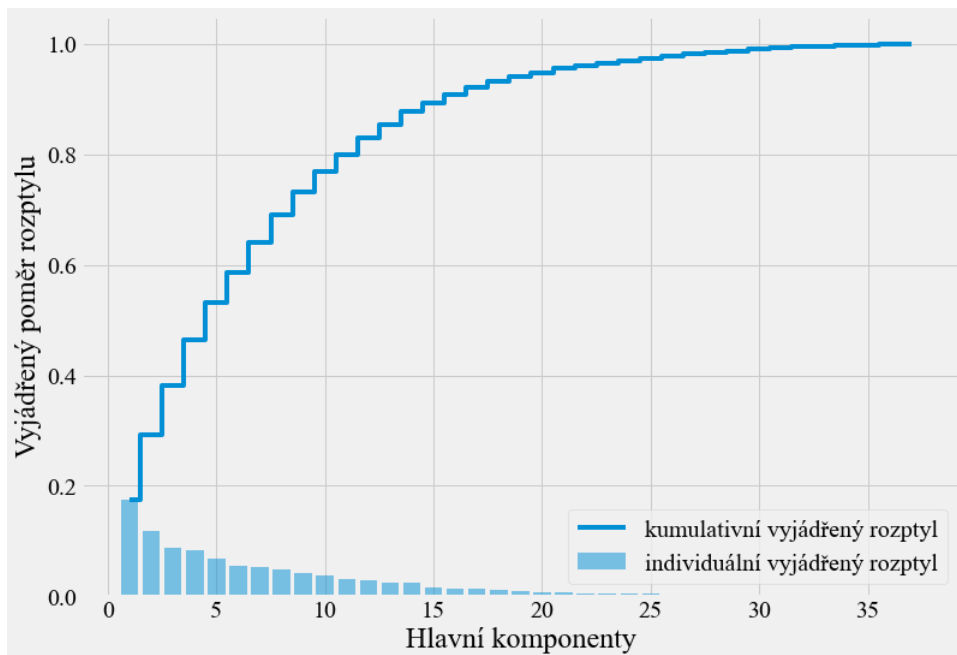
V Tabulce 3.5 je výběr nejlepších parametrů podle korelací mezi parametry (parametry s korelacemi nižšími než 0,3 a vyššími než -0,3), analýzy hlavních parametrů (PFA), Laplaciáského skóre pro výběr parametrů a spektrální analýzy pro výběr parametrů. Výsledkem je 37 parametrů, které se objevují více než dvakrát u výběrů pomocí použitých algoritmů.

Tabulka 3.5 Vybrané parametry a jejich četnosti ve výběrech algoritmů pro výběr parametrů

Pořadí	Název parametru	Četnost u výběrových algoritmů	Pořadí	Název parametru	Četnost u výběrových algoritmů
0	methods_tel	3	19	job_combiStudent	2
1	methods_medic	3	20	job_fulltime	2
2	usage_ecigNoNiq	3	21	protried_snuff	2
3	health_diabetes	3	22	job_retirement	2
4	job_disability	3	23	protried_cigar	2
5	usage_pipe	2	24	protried_heatedTab	2
6	health_cancer	2	25	methods_niq	2
7	methods_pharm	2	26	sex	2
8	methods_expert	2	27	protried_chewy	2
9	usage_hookah	2	28	meds	2
10	usage_snuff	2	29	protried_pipe	2
11	usage_nopuff	2	30	reasons_family	2
12	methods_alter	2	31	phase	2
13	methods_web	2	32	price	2
14	job_parttime	2	33	protried_cigarettes	2
15	usage_ecigThcCbd	2	34	lastduration	2
16	usage_cigar	2	35	protried_ecigThcCbd	2
17	character_count	2	36	protried_nopuff	2
18	job_parent	2			

3.2.2 PCA

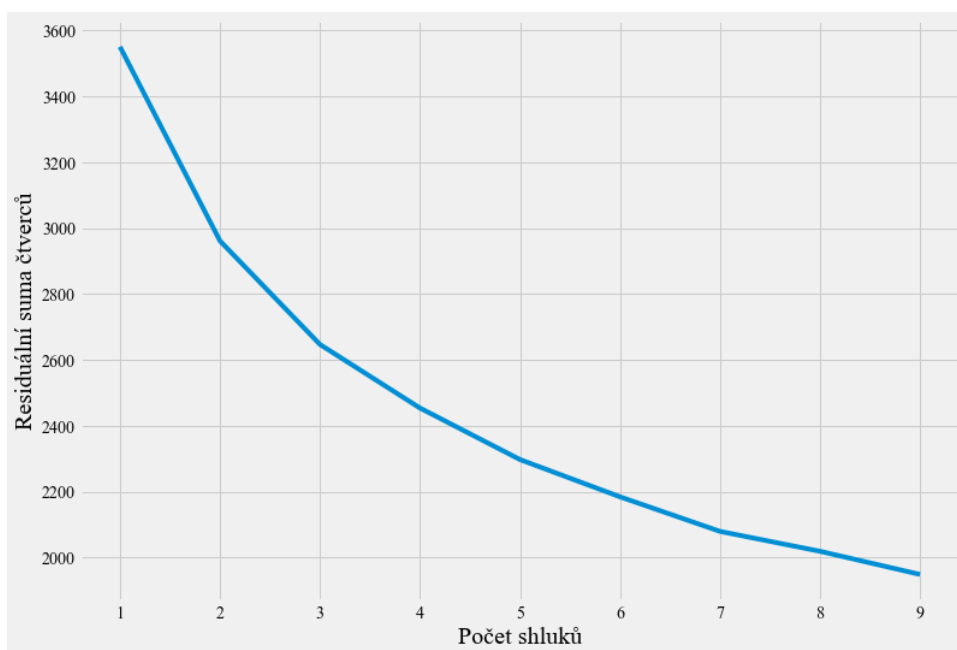
Na Obrázku 3.16 je graf závislosti vyjádřeného poměru rozptylu hlavními komponenty po provedení PCA analýzy na vybraných parametrech z Tabulky 3.5. Deset hlavních komponent odpovídá 80 % zachované informace.



Obrázek 3.16 Rozptyl vyjádřený hlavními komponentami

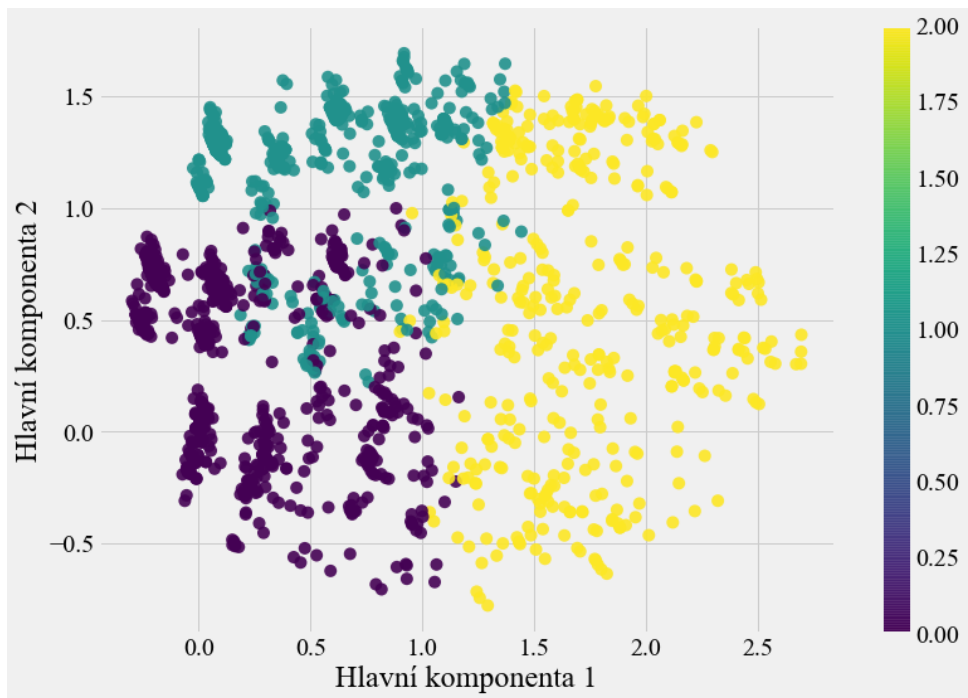
3.2.3 Shlukovací analýza

Na Obrázku 3.17 je graf závislosti residuální sumy čtverců na počtu shluků. Algoritmem pro určení počtu shluků byly vybrány shluky 3, které jsou použity pro *k*-means shlukovací algoritmus.

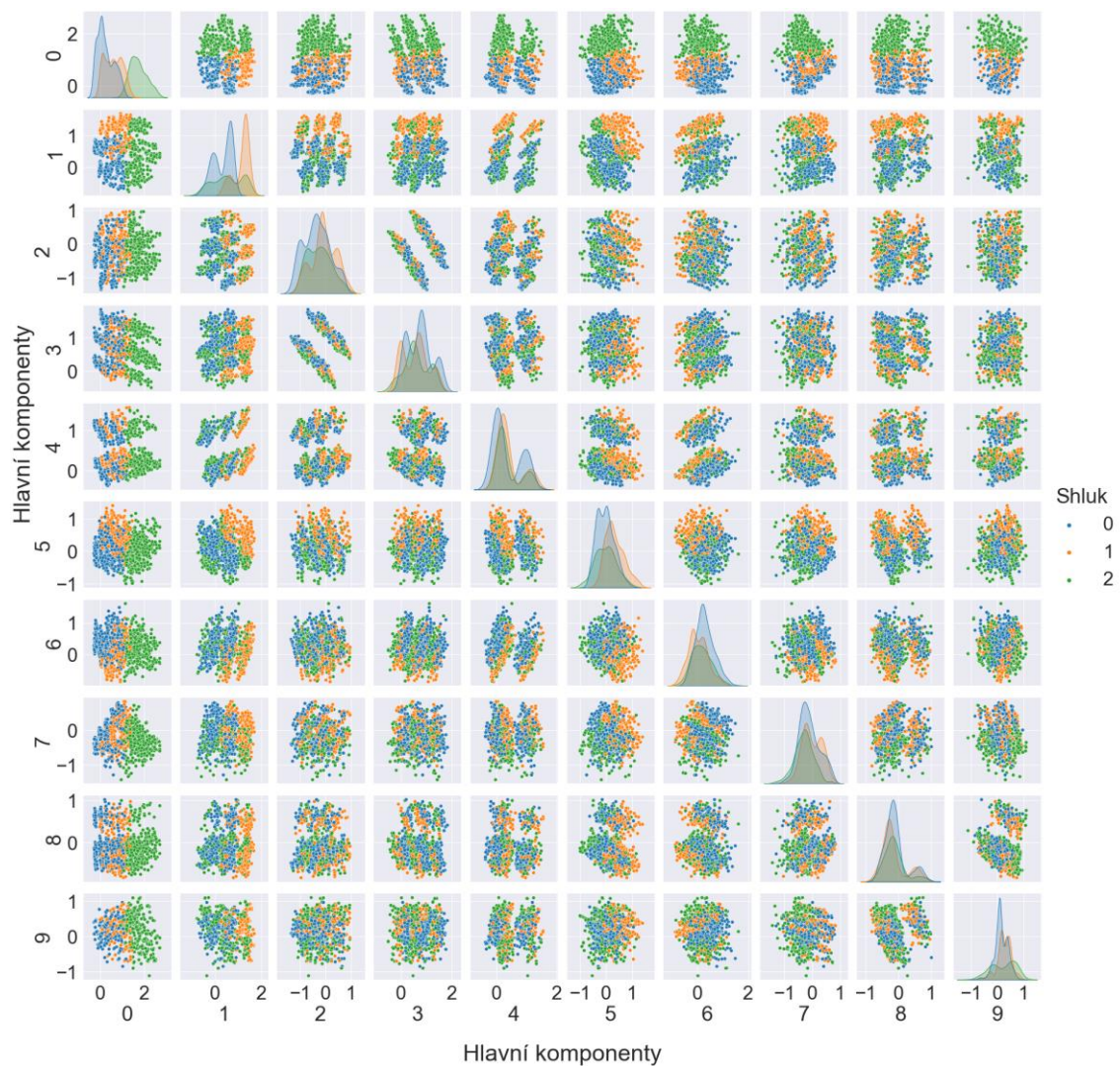


Obrázek 3.17 Graf závislosti SSE na počtu shluků

Na Obrázku 3.18 a Obrázku 3.19 je vizualizace hlavních komponent, které jsou obarveny podle čísla shluku. Calinski-Harabasz skóre je 271 a silhouette skóre je 0,14.



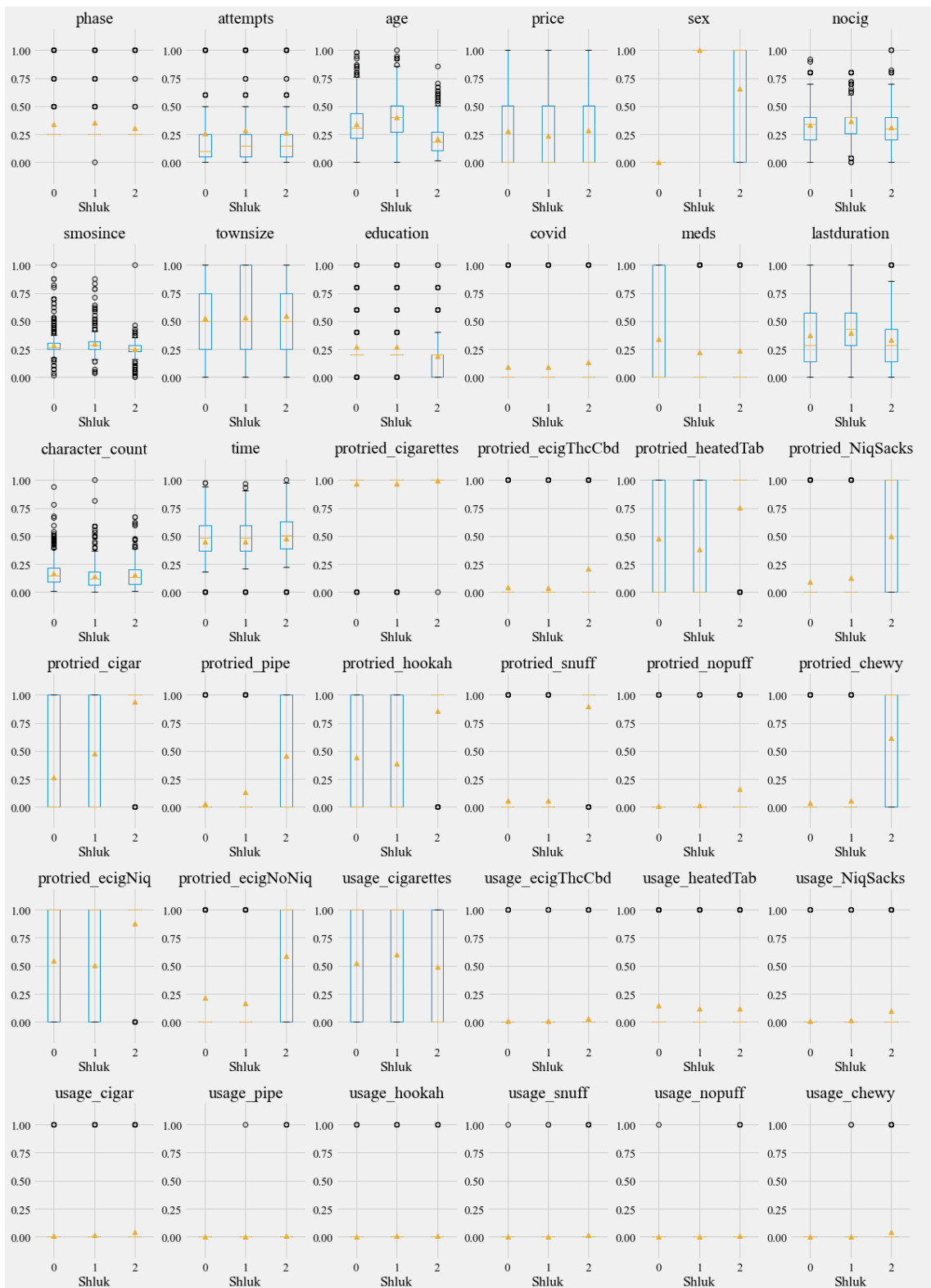
Obrázek 3.18 Vizualizace první a druhé hlavní komponenty v prostoru, obarvení na základě přiřazení ke shlukům



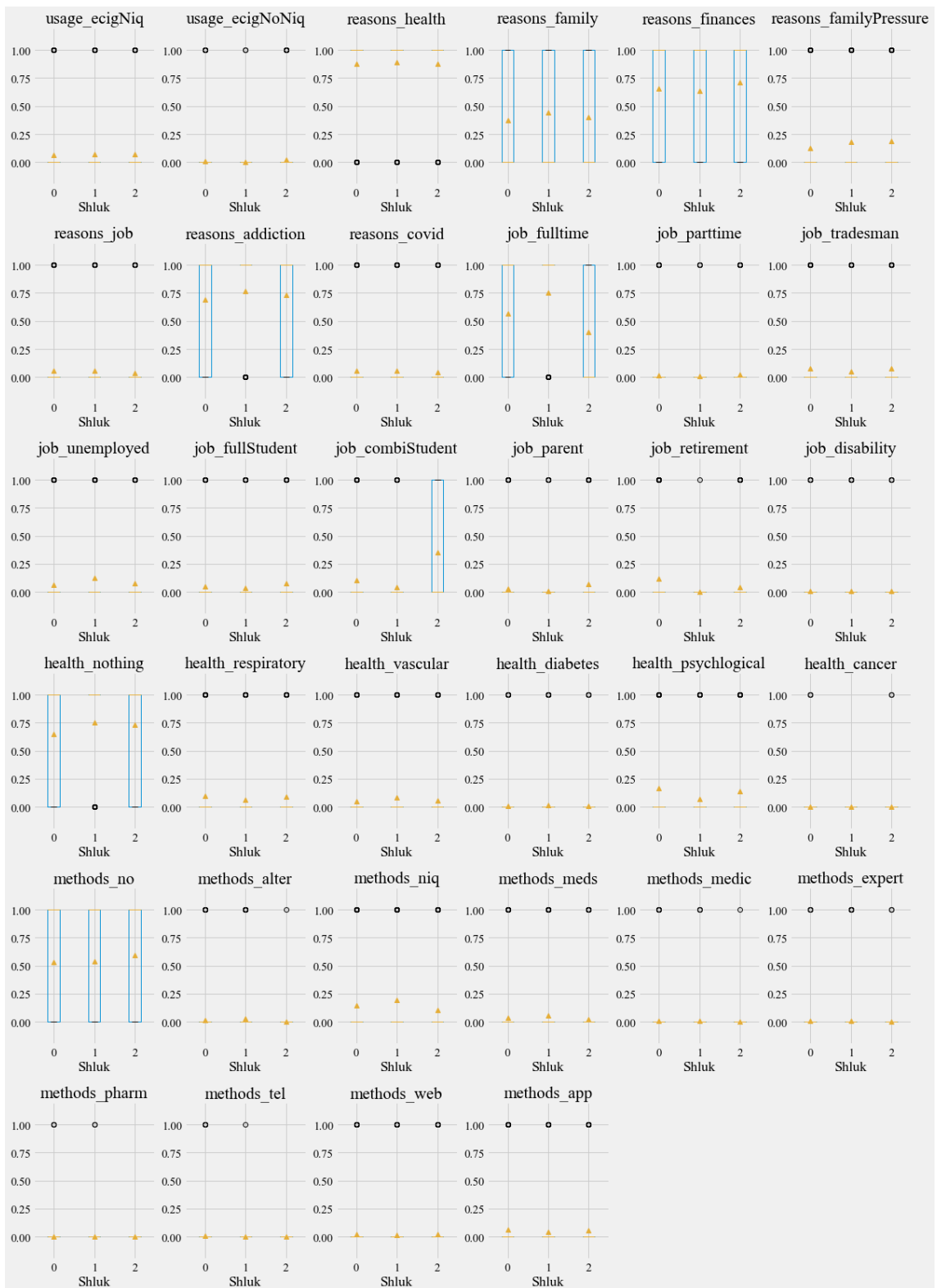
Obrázek 3.19 Vizualizace shluků pomocí grafického zobrazení párů všech komponent, na diagonále je histogram pro každou komponentu

3.2.4 Analýza charakteristik fenotypů

Na Obrázcích 3.20 a 3.21 se nachází vizualizace rozdílů mezi shluky pomocí krabicových grafů. Použitá data jsou původní, která jsou škálována od 0 do 1. Kromě vizualizace pomocí kvantilů byly přidány polohy průměrů v podobě žlutých trojúhelníků, které lépe charakterizují kategorie s binárními hodnotami.



Obrázek 3.20 Vizualizace rozdílů mezi shluky pomocí krabicových grafů, 1. část



Obrázek 3.21 Vizualizace rozdílů mezi shluky pomocí krabicových grafů, 2. část

V Tabulce 3.6 jsou průměrné hodnoty každé kategorie uvnitř shluků. Data nejsou v tomto případě škálována.

Tabulka 3.6 Průměrné hodnoty pro každou kategorii uvnitř shluků

Parametr	shluk 0	shluk 1	shluk 2	Parametr	shluk 0	shluk 1	shluk 2
phase	1.37	1.42	1.23	usage_ecigNiq	0.07	0.07	0.08
attempts	5.16	5.7	5.33	usage_ecigNoNiq	0.01	0	0.03
age	30.66	34.31	23.55	reasons_health	0.88	0.89	0.88
sex	0	1	0.66	reasons_family	0.37	0.44	0.4
nocig	16.6	18.37	15.52	reasons_finances	0.66	0.63	0.71
smosince	16.93	17.76	15.09	reasons_familyPressure	0.13	0.18	0.19
townsize	3.1	3.12	3.2	reasons_job	0.06	0.06	0.03
education	2.36	2.37	1.96	reasons_addiction	0.69	0.76	0.73
covid	0.09	0.09	0.14	reasons_covid	0.05	0.06	0.04
meds	0.34	0.22	0.24	job_fulltime	0.57	0.75	0.4
lastduration	2.64	2.76	2.34	job_parttime	0.01	0.01	0.02
character_count	55.42	47.89	51.01	job_tradesman	0.08	0.05	0.08
time	10.33	10.33	10.98	job_unemployed	0.07	0.13	0.08
protried_cigarettes	0.97	0.97	1	job_fullStudent	0.05	0.04	0.08
protried_ecigThcCbd	0.05	0.04	0.21	job_combiStudent	0.1	0.05	0.36
protried_heatedTab	0.48	0.38	0.75	job_parent	0.03	0.01	0.07
protried_NiqSacks	0.09	0.13	0.5	job_retirement	0.12	0	0.04
protried_cigar	0.27	0.48	0.94	job_disability	0.01	0.01	0.01
protried_pipe	0.03	0.13	0.46	health_nothing	0.65	0.75	0.73
protried_hookah	0.45	0.39	0.86	health_respiratory	0.1	0.06	0.09
protried_snuff	0.06	0.06	0.9	health_vascular	0.05	0.08	0.06
protried_nopuff	0.01	0.01	0.16	health_diabetes	0.01	0.02	0.01
protried_chewy	0.04	0.06	0.61	health_psychological	0.17	0.07	0.14
protried_ecigNiq	0.55	0.51	0.88	health_cancer	0.01	0	0.01
protried_ecigNoNiq	0.22	0.17	0.59	methods_no	0.54	0.54	0.6
usage_cigarettes	0.53	0.6	0.49	methods_alter	0.02	0.03	0
usage_ecigThcCbd	0.01	0.01	0.03	methods_niq	0.14	0.19	0.1
usage_heatedTab	0.15	0.12	0.12	methods_meds	0.04	0.06	0.02
usage_NiqSacks	0.01	0.02	0.1	methods_medic	0.01	0.01	0
usage_cigar	0.01	0.02	0.04	methods_expert	0.01	0.01	0.01
usage_pipe	0	0	0.01	methods_pharm	0.01	0	0
usage_hookah	0.01	0.01	0.01	methods_tel	0.01	0	0
usage_snuff	0	0	0.02	methods_web	0.02	0.02	0.02
usage_nopuff	0	0	0.01	methods_app	0.06	0.05	0.06
usage_chewy	0	0	0.04				

V Tabulce 3.7 jsou výsledky jednosměrného ANOVA testu v podobě p-hodnot. Zvýrazněny jsou ty hodnoty, u kterých lze zamítnout nulovou hypotézu na hladině významnosti $\alpha = 0,05$ a tmavší barvou jsou zvýrazněny hodnoty, které jsou pod hranicí $p = 0,001$. Výsledky ukazují rozdíly ve středních hodnotách podle kategorií mezi všemi shluky pro výběr dominantních vlastností pro charakteristiku shluků – fenotypů. V Tabulce 3.8 jsou výsledky Kruskalova–Wallisova testu pro hodnoty, které nabývají

kvantitativních hodnot a nelze na ně použít jednosměrná ANOVA. P-hodnoty jsou v tabulce vyznačeny stejným způsobem jako v Tabulce 3.7.

Tabulka 3.7 Tabulka výsledných p-hodnot z jednosměrného ANOVA testu mezi shluky

Parametr	P-hodnota	Parametr	P-hodnota	Parametr	P-hodnota
sex	0.00000	usage_pipe	0.00508	job_parent	0.00000
covid	0.03182	usage_hookah	0.56391	job_retirement	0.00000
meds	0.00000	usage_snuff	0.00167	job_disability	0.76194
protried_cigarettes	0.00219	usage_nopuff	0.00539	health_nothing	0.00015
protried_ecigThcCbd	0.00000	usage_chewy	0.00000	health_respiratory	0.07008
protried_heatedTab	0.00000	usage_ecigNiq	0.86529	health_vascular	0.07970
protried_NiqSacks	0.00000	usage_ecigNoNiq	0.00727	health_diabetes	0.71153
protried_cigar	0.00000	reasons_health	0.74507	health_psychological	0.00001
protried_pipe	0.00000	reasons_family	0.06378	health_cancer	0.30308
protried_hookah	0.00000	reasons_finances	0.04363	methods_no	0.09836
protried_snuff	0.00000	reasons_familyPressure	0.01193	methods_alter	0.00696
protried_nopuff	0.00000	reasons_job	0.11786	methods_niq	0.00065
protried_chewy	0.00000	reasons_addiction	0.02428	methods_meds	0.00790
protried_ecigNiq	0.00000	reasons_covid	0.52151	methods_medic	0.40427
protried_ecigNoNiq	0.00000	job_fulltime	0.00000	methods_expert	0.40647
usage_cigarettes	0.00240	job_parttime	0.11381	methods_pharm	0.40670
usage_ecigThcCbd	0.00493	job_tradesman	0.08101	methods_tel	0.10468
usage_heatedTab	0.24953	job_unemployed	0.00040	methods_web	0.75918
usage_NiqSacks	0.00000	job_fullStudent	0.02656	methods_app	0.45120
usage_cigar	0.00079	job_combiStudent	0.00000		

Světle modré zbarvení pro hodnoty $p < 0,05$, tmavě modré pro hodnoty $p < 0,001$

Tabulka 3.8 Tabulka výsledných p-hodnot z Kruskalova-Wallisova testu mezi shluky

Parametr	p-hodnota	Parametr	p-hodnota
phase	0.00262	townsize	0.50546
attempts	0.06128	education	0.00000
age	0.00000	lastduration	0.00533
nocig	0.00000	character_count	0.00000
smosince	0.00000	time	0.05248

V Tabulce 3.9 se nachází pořadí vybraných parametrů u kterých je rozdíl mezi shluky. Pořadí je od nejmenší p-hodnoty, tedy od parametrů, u kterých je největší rozdíl. Zároveň tabulka ukazuje výsledky Tukeyho post-hoc testu na binárních typech parametrů a výsledky Dunnova post-hoc testu s Holm-Bonferroniho korekcí na kvantitativních typech dat. Zvýrazněná pole značí, že lze zamítnout nulovou hypotézu na hladině významnosti $\alpha = 0,05$.

Tabulka 3.9 Pořadí parametrů od nejmenších p-hodnot po největší, výsledky Tukeyho testu a Dunnova testu mezi shluky

Pořadí	Parametr	0 a 1	0 a 2	1 a 2	Pořadí	Parametr	0 a 1	0 a 2	1 a 2
0	protried_snuff	ne	ano	ano	22	meds	ano	ano	ne
1	sex	ano	ano	ano	23	health_psychological	ano	ne	ano
2	protried_chewy	ne	ano	ano	24	health_nothing	ano	ano	ne
3	protried_cigar	ano	ano	ano	25	job_unemployed	ano	ne	ano
4	protried_pipe	ano	ano	ano	26	usage_cigar	ne	ano	ano
5	protried_NiqSacks	ne	ano	ano	27	usage_snuff	ne	ano	ano
6	age	ano	ano	ano	28	character_count	ano	ano	ano
7	protried_hookah	ne	ano	ano	29	protried_cigarettes	ne	ano	ano
8	protried_ecigNoNiq	ne	ano	ano	30	usage_cigarettes	ano	ne	ano
9	job_combiStudent	ano	ano	ano	31	phase	ne	ano	ano
10	protried_ecigNiq	ne	ano	ano	32	usage_ecigThcCbd	ne	ano	ano
11	protried_nopuff	ne	ano	ano	33	usage_pipe	ne	ano	ano
12	protried_heatedTab	ano	ano	ano	34	lastduration	ne	ne	ano
13	job_fulltime	ano	ano	ano	35	usage_nopuff	ne	ano	ano
14	protried_ecigThcCbd	ne	ano	ano	36	methods_alter	ne	ne	ano
15	job_retirement	ano	ano	ano	37	usage_ecigNoNiq	ne	ne	ano
16	smosince	ano	ano	ano	38	methods_meds	ne	ne	ano
17	usage_NiqSacks	ne	ano	ano	39	reasons_familyPressure	ne	ano	ne
18	usage_chewy	ne	ano	ano	40	reasons_addiction	ano	ne	ne
19	education	ne	ano	ano	41	job_fullStudent	ne	ne	ano
20	nocig	ano	ano	ano	42	covid	ne	ano	ne
21	job_parent	ne	ano	ano	43	reasons_finances	ne	ne	ano

Fenotyp 0

V Tabulce 3.10 jsou vybrané vlastnosti pro fenotyp 0 a jejich vyhodnocení. Vlastnosti jsou seřazeny na základě výsledné p-hodnoty od nejmenší po provedení jednosměrného ANOVA testu. Výše v tabulce jsou tedy vlastnosti s výraznějšími rozdíly mezi shluky.

Nejvíce dominantní vlastností pro fenotyp 0 je pohlaví – fenotyp 0 je zastoupen pouze ženami, jejichž kuřácké zvyklosti a demografické charakteristiky jsou podobné s ostatními fenotypy, z toho důvodu je výběr vlastností menší. Fenotyp má nejvyšší průměrný počet znaků a v průměru bere pravidelné léky nejvíce.

Tabulka 3.10 Dominantní vlastnosti fenotypu 0

Vlastnost	Vyhodnocení	Odlišnost oproti ostatním shlukům	Střední hodnota
Pohlaví	ženy	pouze jedno pohlaví	0
Vyzkoušen doutník	spíše ne	nejnižší četnost	0,27
Vyzkoušena dýmka	ne	nejnižší četnost	0,03
Věk	průměr 31 let	střední průměr	30,66
Kombinované studium	ne	nejnižší četnost	0,1
Vyzkoušen zahřívaný tabák	středně	střední četnost	0,48
Plný úvazek	středně	střední četnost	0,57
Důchod	ne	nevyšší četnost	0,12
Věk začátku kouření	průměr 17 let	střední průměr	16,93
Počet cigaret denně	průměr 17	střední průměr	16,6
Léky	spíše ne	nejvyšší průměr	0,34
Žádné onemocnění	spíše ano	nejnižší četnost	0,65
Počet znaků	průměr 55 znaků	nejvyšší průměr	55,42
Důvod tlak rodiny	ne	nejnižší četnost	0,13
Důvod zbavení se závislosti	spíše ano	nejnižší četnost	0,69
Důvod koronavirus	ne	střední četnost	0,05

Fenotyp 1

V Tabulce 3.11 jsou vybrané vlastnosti pro fenotyp 1 a jejich vyhodnocení. Navíc dominantní vlastností tohoto fenotypu je pohlaví, které je opačné oproti fenotypu 0 – muži. Mají nejvyšší průměrný věk a nejvyšší průměr cigaret denně. Mají nejvyšší četnost ze všech fenotypů četnost u práce na plném úvazku. Mají nejvyšší motivaci ve zbavení se závislosti a nejnižší u financí oproti ostatním.

Tabulka 3.11 Dominantní vlastnosti fenotypu 1

Vlastnost	Vyhodnocení	Odlišnost oproti ostatním shlukům	Střední hodnota
Pohlaví	muži	pouze jedno pohlaví	1
Vyzkoušen doutník	středně	střední četnost	0,48
Vyzkoušena dýmka	ne	střední četnost	0,13
Věk	průměr 34 let	nejvyšší průměr	34,31
Kombinované studium	ne	střední četnost	0,05
Vyzkoušen zahříváný tabák	spíše ne	nejnižší četnost	0,38
Plný úvazek	spíše ano	nejvyšší četnost	0,75
Důchod	ne	nejnižší četnost	0
Věk začátku kouření	průměr 18 let	nejvyšší průměr	17,76
Počet cigaret denně	průměr 18	nejvyšší průměr	18,37
Duševní onemocnění	ne	nejnižší četnost	0,07
Nezaměstnaný	ne	nejvyšší četnost	0,13
Počet znaků	průměr 48 znaků	nejnižší průměr	47,89
Trvání posledního pokusu	spíše kratší	nejvyšší průměr	2,76
Užívání cigaret	střední	nejvyšší četnost	0,6
Alternativní metody odvykání	ne	nejvyšší četnost	0,03
Užívání beznikotinových e-cigaret	ne	nejnižší četnost	0
Léky	spíše ne	střední četnost	0,22
Důvod zbavení závislosti	spíše ano	nejvyšší četnost	0,76
Prezenční student	ne	nejnižší četnost	0,04
Důvod finance	spíše ano	nejnižší četnost	0,63

Fenotyp 2

V Tabulce 3.12 jsou vybrané vlastnosti pro fenotyp 2 a jejich vyhodnocení. Fenotyp 2 je nejbohatší co se týče kuřáckých zvyklostí, které se liší od ostatních fenotypů. Pohlaví je namíchané, ale jedná se spíše o mladší uživatele okolo 24 let, kteří se liší hlavně množstvím vyzkoušených produktů od ostatních fenotypů. Jejich nejvíce dominantní vlastností je, že vyzkoušeli šňupací tabák. Jejich poslední abstinenční pokus byl spíše kratší a jejich fáze v terapii je v průměru nejvíce na začátku.

Tabulka 3.12 Dominantní vlastnosti fenotypu 2

Vlastnost	Vyhodnocení	Odlišnost oproti ostatním shlukům	Střední hodnota
Vyzkoušen šňupací tabák	ano	nejvyšší četnost	0,9
Pohlaví	spíše muži	střední četnosti	0,66
Vyzkoušen žvýkáci tabák	spíše ano	nejvyšší četnost	0,61
Vyzkoušen doutník	ano	nejvyšší četnost	0,94
Vyzkoušena dýmka	střední	nejvyšší četnost	0,46
Vyzkoušeny nikotinové sáčky	střední	nejvyšší četnost	0,5
Věk	průměr 24 let	nejnižší průměr	23,55
Vyzkoušena vodní dýmka	ano	nejvyšší četnost	0,86
Vyzkoušeny beznikotinové e-cigarety	střední	nejvyšší četnost	0,59
Kombinované studium	spíše ne	nejvyšší četnost	0,36
Vyzkoušeny nikotinové e-cigarety	ano	nejvyšší četnost	0,88
Vyzkoušen bezdýmový tabák	ne	nejvyšší četnost	0,16
Vyzkoušen zahříváný tabák	spíše ano	nejvyšší četnost	0,75
Plný úvazek	střední	nejnižší četnost	0,4
Vyzkoušeny THC/CBD e-cigarety	spíše ne	nejvyšší četnost	0,21
Důchod	ne	střední četnost	0,04
Věk začátku kouření	průměr 15 let	nejnižší průměr	15,09
Počet cigaret denně	průměr 16	nejnižší průměr	15,52
Užívání nikotinových sáčků	ne	nejvyšší četnost	0,1
Užívání žvýkacího tabáku	ne	nejvyšší četnost	0,04
Vzdělání	spíše nízké	nejnižší průměr	1,96
Rodičovská dovolená	ne	nejvyšší četnost	0,07
Užívání doutníků	ne	nejvyšší četnost	0,04
Užívání šňupacího tabáku	ne	nejvyšší četnost	0,02
Počet znaků	průměr 51 znaků	střední průměr	51,01
Vyzkoušeny cigarety	ano	nejvyšší četnost	1
Trvání posledního pokusu	spíše krátké	nejnižší průměr	2,34
Užívání THC/CBD e-cigaret	ne	nejvyšší četnost	0,03
Užívání dýmek	ne	nejvyšší četnost	0,01
Užívání bezdýmového tabáku	ne	nejvyšší četnost	0,01
Alternativní metody odvykání	ne	nejnižší četnost	0
Užívání beznikotinových e-cigaret	ne	nejvyšší četnost	0,03
Fáze terapie	spíše začáteční	nejnižší průměr	1,23
Odvykání pomocí léků	ne	nejnižší četnost	0,02
Prezenční student	ne	nejvyšší četnost	0,08
Důvod koronavirus	ne	nejnižší četnost	0,04
Důvod finance	spíše ano	nejvyšší četnost	0,71

3.2.5 Shrnutí a diskuse

Ačkoliv ze shlukovací analýzy bylo vytvořeno několik vzorů chování uživatelů závislých na nikotinu, překryv ve vlastnostech mezi shluky – fenotypy je velký. Silhouette skóre, které nabývá hodnot od -1 do 1 podle toho, jak dobře jsou shluky oddělené, má v tomto případě hodnotu 0,14. Pokud se silhouette skóre blíží 0, skóre značí překrývající se shluky [30]. Protože jsem chtěla shlukovací analýzu provést co nejlépe, vyzkoušela jsem dále algoritmy aglomerativního hierarchického shlukování a shlukování pomocí Gaussovských směsí. Obě shlukování dávala nižší skóre, jak silhouette, tak Calinski-Harabasz. Dále jsem zkoušela shlukování pouze parametrů, která se ve výběrech pomocí výběrových algoritmů pro parametry objevila třikrát, ale skóre pro porovnání na stejných datech bylo nižší. Další jev, který značí špatnou separovatelnost vzorů, je náhoda ve výběru příznaků a u shlukovacího algoritmu. Na datech jsem algoritmy použila několikrát, abych dosáhla co nejlepšího výsledku, ale pokaždé byly shluky mírně rozdílné. Pro další analýzy by toto bylo možné řešit například bootstrapp algoritmem, který data navzorkuje a pro každý vzorek se přepočítá shlukovací analýzou.

Výsledkem jsou tedy 3 digitální fenotypy, které můžou sloužit pro základní pochopení vzorů chování u nikotinově závislých pacientů. Výsledné skupiny jsou jedním z příkladů rozdělení závislých pacientů. Vzhledem k obsáhlosti a zajímavosti získaného souboru dat by bylo zajímavé se v budoucnosti věnovat tomuto souboru více a přejít k například pokročilejším metodám pro získání fenotypů, rozdělit data na více fenotypů nebo se zaměřit na metody učení s učitelem, a to třeba podle více pasivních dat získaných digitálním phenotypingem.

4 Závěr

Za účelem seznámení se s daty, která byla shromážděna od uživatelů nikotinových výrobků, jsem vypracovala explorační analýzu v podobě aktualizovatelných skriptů pomocí nástroje Jupyter Notebook. Data uživatelů jsem rozdělila na otázky, které pokládá virtuální terapeut, a v závislosti na sociodemografických charakteristikách jsem podle četností spočítala procentuální rozložení odpovědí v každé kategorii. Pro volné odpovědi jsem použila základní metody NLP, mezi které patří nejčtenější slova, dvojsloví a trojsloví, průměrný počet znaků, sentiment analýza s pomocí knihovny VADER. Pro vizualizaci jsem použila barevné formátování tabulek v závislosti na hodnotách uvnitř buněk, sloupcové grafy a u volných odpovědí tzv. wordclouds – způsob vizualizace nejčastějších slov. Pro porovnání četností napříč jednotlivými kategoriemi jsem spočítala χ^2 test nezávislosti s hladinou významnosti $\alpha = 0,05$. Mezi většinou kategorií byly nalezeny signifikantní rozdíly v četnostech odpovědí. Zpracovaná data jsou vzhledem do problematiky závislosti na nikotinu. Mohou sloužit k cílenější prevenci, lepšímu pochopení pacientů nebo ke zpětné vazbě. Především pak ukazují, které oblasti dat by se hodily k podrobnějším analýzám, například charakterizovat ideálního úspěšného pacienta, zaměřit se a lépe analyzovat pacienty s vysokým počtem minulých pokusů o odvykání, nebo se zaměřit na motivaci mladých lidí, kteří mají v aplikaci nižší úspěšnost.

V další části práce jsem se věnovala problematice digitálního phenotypingu – vytvořením digitálních fenotypů. Pomocí explorační analýzy jsem vybrala kategorie, které jsou vhodné k analýze. Data jsem upravila a pomocí několika algoritmů pro výběr příznaků jsem redukovala jejich dimenzionalitu, kterou jsem následně ještě zmenšila pomocí PCA na 10 komponent. Na těchto 10 komponentech jsem provedla shlukovací analýzu algoritmem k -means a vytvořila 3 shluky – fenotypy. Výsledné tři shluky jsem mezi sebou testovala podle příznaků na shodnost středních hodnot – jednosměrným ANOVA testem a Kruskalovým-Wallisovým testem s hladinou významnosti $\alpha = 0,05$, abych našla odlišné vlastnosti fenotypů. Pro příznaky, u nichž jsem zamítla nulovou hypotézu, jsem provedla navazující post-hoc testy – Tukeyho HSD test a Dunnův test s Holm-Bonferroniho korekcí. Hladina významnosti byla stanovena u obou testů na $\alpha = 0,05$. Výsledné rozdíly mezi fenotypy jsem sepsala jako dominující vlastnosti pro každý z nich. Nejvýraznější vlastnost fenotypu 0 je pouze ženské pohlaví, fenotypu 1

pouze mužské pohlaví a fenotyp 2 jsou v průměru mladí lidé. Ačkoliv je výsledkem soubor několika vlastností pro každý fenotyp, překryv všech shluků je velký a výsledné fenotypy jsou jedním z příkladů možného výsledku.

Seznam použité literatury

- [1] CSÉMY, Ladislav, Zuzana DVOŘÁKOVÁ, Alena FIALOVÁ, Miloslav KODL, Marek MALÝ a Miroslava SKÝVOVÁ. *Národní výzkum užívání tabáku a alkoholu v České republice 2020* [online]. 1. Praha: Státní zdravotní ústav, 2021 [cit. 2021-07-06]. ISBN 978-80-7071-405-8. Dostupné z: http://www.szu.cz/uploads/documents/szu/aktual/nauta_2020.pdf
- [2] KRÁLÍKOVÁ, Eva, Richard ČEŠKA, Alexandra PÁNKOVÁ, Lenka ŠTĚPÁNKOVÁ, Kamila ZVOLSKÁ, Vladislava FELBROVÁ, Stanislava KULOVANÁ a Miroslav ZVOLSKÝ. Doporučení pro léčbu závislosti na tabáku. *Vnitřní lékařství*. 2015, **61**(5), 4–115.
- [3] KULHÁNEK, A., GABRHELÍK, R., NOVÁK, D., BURDA, V., BRENDRYEN, H. EHealth Intervention for Smoking Cessation for Czech Tobacco Smokers: Pilot Study of User Acceptance. *Adiktologie*. 2018, **18**(2), 81–85.
- [4] ONNELA, JP. Opportunities and Challenges in the Collection and Analysis of Digital Phenotyping Data. *Neuropsychopharmacol.* 2021, (46), 45–54. Dostupné z: doi:<https://doi.org/10.1038/s41386-020-0771-3>
- [5] STROUD, Clare, Jukka-Pekka ONNELA a Husseini MANJI. Harnessing digital technology to predict, diagnose, monitor, and develop treatments for brain disorders. *Npj Digit. Med.* 2019, **44**(2). Dostupné z: doi:<https://doi.org/10.1038/s41746-019-0123-z>
- [6] JAIN, Sachin, Brian POWERS, Jared HAWKINS a John BROWNSTEIN. The digital phenotype. *Nature Biotechnology* [online]. **33**(5), 462-463 [cit. 2021-07-05]. Dostupné z: <https://www.nature.com/articles/nbt.3223.pdf>
- [7] CSÉMY, Ladislav, Zuzana DVOŘÁKOVÁ, Alena FIALOVÁ, Miloslav KODL, Marek MALÝ a Miroslava SKÝVOVÁ. *Užívání tabáku a alkoholu v České republice 2019*. Státní zdravotní ústav, 2020.
- [8] *Národní strategie prevence a snižování škod spojených se závislostním chováním 2019-2027*. Praha: Úřad vlády České republiky, 2019. Koordinace. ISBN 978-80-7440-231-9.
- [9] Národní protidrogová koordinátorka: Počty kuřáků se snižují, daří se vytvářet zdravější prostředí pro odvykání. *Vláda České republiky* [online]. [cit.

2021-07-06]. Dostupné z: <https://www.vlada.cz/scripts/detail.php?id=188719&tmplid=50>

- [10] Výsledky národního výzkumu SZÚ o užívání tabáku v ČR potvrdily, že českých kuřáků ubývá, povědomí o zdravotních rizicích kouření se zvýšilo. *Ministerstvo zdravotnictví České republiky* [online]. [cit. 2021-07-06]. Dostupné z: <https://www.mzcr.cz/tiskove-centrum-mz/vysledky-narodniho-vyzkumu-szu-o-uzivani-tabaku-v-cr-potvrdily-ze-ceskych-kuraku-ubyva-povedomi-o-zdravotnich-rizicich-koureni-se-zvysilo/>
- [11] BAUER, Michael, Tasha GLENN, John GEDDES et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. *Int J Bipolar Disord.* 2020, **8**(2). Dostupné z: [doi:https://doi.org/10.1186/s40345-019-0164-x](https://doi.org/10.1186/s40345-019-0164-x)
- [12] LARSON, Richard S. A Path to Better-Quality mHealth Apps. *JMIR mHealth and uHealth* [online]. 2018, **6**(7) [cit. 2021-07-05]. ISSN 2291-5222. Dostupné z: [doi:10.2196/10414](https://doi.org/10.2196/10414)
- [13] Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 1st quarter 2021. *Statista* [online]. [cit. 2021-07-05]. Dostupné z: <https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/>
- [14] ANDRIENKO, Natalia a Gennady ANDRIENKO. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. ISBN 978-354-0311-904.
- [15] KANG, Yue, Zhao CAI, Chee-Wee TAN, Qian HUANG a Hefu LIU. Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics* [online]. 2020, **7**(2), 139-172 [cit. 2021-07-28]. ISSN 2327-0012. Dostupné z: [doi:10.1080/23270012.2020.1756939](https://doi.org/10.1080/23270012.2020.1756939)
- [16] EISENSTEIN, Jacob. *Introduction to natural language processing*. Cambridge: The MIT Press, 2019. ISBN 978-0-262-04284-0.
- [17] HEIMERL, Florian, Steffen LOHMANN, Simon LANGE a Thomas ERTL. Word Cloud Explorer: Text Analytics Based on Word Clouds. In: *2014 47th Hawaii International Conference on System Sciences* [online]. IEEE, 2014, s. 1833-1842 [cit. 2021-08-02]. ISBN 978-1-4799-2504-9. Dostupné z: [doi:10.1109/HICSS.2014.231](https://doi.org/10.1109/HICSS.2014.231)
- [18] DANG, Nhan Cach, María N. MORENO-GARCÍA a Fernando DE LA PRIETA. Sentiment Analysis Based on Deep Learning: A Comparative Study.

- Electronics* [online]. 2020, **9**(3), 483 [cit. 2021-08-02]. ISSN 2079-9292. Dostupné z: doi:10.3390/electronics9030483
- [19] PRABOWO, Rudy a Mike THELWALL. Sentiment analysis: A combined approach. *Journal of Informetrics* [online]. 2009, **3**(2), 143-157 [cit. 2021-08-02]. ISSN 17511577. Dostupné z: doi:10.1016/j.joi.2009.01.003
- [20] HUTTO, C. a E. GILBERT. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *ICWSM*. 2014, **8**(1), 216-225.
- [21] ROBINSON, Peter N. Deep phenotyping for precision medicine. *Human Mutation* [online]. 2012, **33**(5), 777-780 [cit. 2021-08-02]. ISSN 10597794. Dostupné z: doi:10.1002/humu.22080
- [22] ONNELA, Jukka-Pekka a Scott L RAUCH. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*. 2016, **41**(7), 1691-1696. ISSN 0893-133X. Dostupné z: doi:10.1038/npp.2016.7
- [23] HENSON, Philip, Ian BARNETT, Matcheri KESHAVAN a John TOROUS. Towards clinically actionable digital phenotyping targets in schizophrenia. *Npj Schizophrenia* [online]. 2020, **6**(1) [cit. 2021-07-05]. ISSN 2334-265X. Dostupné z: doi:10.1038/s41537-020-0100-1
- [24] FAGHERAZZI, Guy. Deep Digital Phenotyping and Digital Twins for Precision Health: Time to Dig Deeper. *Journal of Medical Internet Research* [online]. 2020, **22**(3) [cit. 2021-08-02]. ISSN 1438-8871. Dostupné z: doi:10.2196/16770
- [25] MCHUGH, Mary L. The Chi-square test of independence. *Biochemia Medica*. 2013, **23**(2), 143-149. ISSN 18467482. Dostupné z: doi:10.11613/BM.2013.018
- [26] LU, Yijuan, Ira COHEN, Xiang Sean ZHOU a Qi TIAN. Feature selection using principal feature analysis. In: *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07* [online]. New York, New York, USA: ACM Press, 2007, s. 301- [cit. 2021-08-08]. ISBN 9781595937025. Dostupné z: doi:10.1145/1291233.1291297
- [27] HE, XIAOFEI; CAI, DENG; NIYOGI, PARTHA. Laplacian score for feature selection. *Advances in neural information processing systems*. 2005, **18**.
- [28] ZHAO, Zheng a Huan LIU. Spectral feature selection for supervised and unsupervised learning. In: *Proceedings of the 24th international conference on*

Machine learning - ICML '07 [online]. New York, New York, USA: ACM Press, 2007, s. 1151-1157 [cit. 2021-08-08]. ISBN 9781595937933. Dostupné z: doi:10.1145/1273496.1273641

- [29] CALINSKI, T. a J. HARABASZ. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods* [online]. 1974, **3**(1), 1-27 [cit. 2021-08-08]. ISSN 0361-0926. Dostupné z: doi:10.1080/03610927408827101
- [30] ROUSSEEUW, Peter J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* [online]. 1987, **20**, 53-65 [cit. 2021-08-08]. ISSN 03770427. Dostupné z: doi:10.1016/0377-0427(87)90125-7
- [31] ZANCANARO, Massimo, Tsvi KUFLIK, Zvi BOGER, Dina GOREN-BAR a Dan GOLDWASSER. Analyzing Museum Visitors' Behavior Patterns. CONATI, Cristina, Kathleen MCCOY a Georgios PALIOURAS, ed. *User Modeling 2007* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, s. 238-246 [cit. 2021-08-08]. Lecture Notes in Computer Science. ISBN 978-3-540-73077-4. Dostupné z: doi:10.1007/978-3-540-73078-1_27
- [32] KRUSKAL, William H. a W. Allen WALLIS. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* [online]. 1952, **47**(260), 583-621 [cit. 2021-08-08]. ISSN 0162-1459. Dostupné z: doi:10.1080/01621459.1952.10483441

Příloha A: Tabulky a vizualizace

Osobní profil

Od kolika let pravidelně kouříš?

Tabulka 0.1 Četnosti odpovědí podle kategorií

Věk	0-9	10-14	15-17	18-20	21-25	26-30	31-40	41-50	51-60	60<	ALL
all	91	896	2140	1008	283	87	49	23	2	2	4581
men	42	410	1062	578	172	49	19	11	0	1	2344
women	49	486	1078	430	111	38	30	12	2	1	2237
light	43	229	563	247	67	28	18	7	0	0	1202
middle	39	468	1140	550	159	39	17	9	2	0	2423
heavy	9	199	441	210	57	20	14	7	0	2	959
under18	10	195	102	1	0	0	0	0	0	0	308
f18t20	24	119	345	75	2	0	0	0	0	1	566
f21t25	36	189	492	182	29	0	0	0	0	0	928
f26t30	8	176	420	214	68	7	0	2	0	0	895
f31t40	3	141	493	292	113	34	22	1	0	0	1099
f41t50	2	41	182	158	48	33	16	12	0	1	493
f51t60	0	9	46	50	8	6	6	7	2	0	134
over60	0	2	6	9	1	1	2	1	0	0	22
bujon	88	380	1013	322	105	27	20	5	1	2	1963
normal	3	508	1111	670	170	60	29	18	1	0	2570

Tabulka 0.2 Procentuální rozložení odpovědí podle kategorií

Věk	0-9	10-14	15-17	18-20	21-25	26-30	31-40	41-50	51-60	60<
all	2	19.6	46.7	22	6.2	1.9	1.1	0.5	0	0
men	1.8	17.5	45.3	24.7	7.3	2.1	0.8	0.5	0	0
women	2.2	21.7	48.2	19.2	5	1.7	1.3	0.5	0.1	0
light	3.6	19.1	46.8	20.5	5.6	2.3	1.5	0.6	0	0
middle	1.6	19.3	47	22.7	6.6	1.6	0.7	0.4	0.1	0
heavy	0.9	20.8	46	21.9	5.9	2.1	1.5	0.7	0	0.2
under18	3.2	63.3	33.1	0.3	0	0	0	0	0	0
f18t20	4.2	21	61	13.3	0.4	0	0	0	0	0.2
f21t25	3.9	20.4	53	19.6	3.1	0	0	0	0	0
f26t30	0.9	19.7	46.9	23.9	7.6	0.8	0	0.2	0	0
f31t40	0.3	12.8	44.9	26.6	10.3	3.1	2	0.1	0	0
f41t50	0.4	8.3	36.9	32	9.7	6.7	3.2	2.4	0	0.2
f51t60	0	6.7	34.3	37.3	6	4.5	4.5	5.2	1.5	0
over60	0	9.1	27.3	40.9	4.5	4.5	9.1	4.5	0	0
bujon	4.5	19.4	51.6	16.4	5.3	1.4	1	0.3	0.1	0.1

Věk	0-9	10-14	15-17	18-20	21-25	26-30	31-40	41-50	51-60	60<
normal	0.1	19.8	43.2	26.1	6.6	2.3	1.1	0.7	0	0

Tabulka 0.3 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light			0.000	0.000											
middle				0.056											
heavy															
under18						0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f18t20							0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f21t25								0.000	0.000	0.000	0.000	0.000	0.000		
f26t30									0.000	0.000	0.000	0.000	0.000		
f31t40										0.000	0.000	0.000	0.000		
f41t50											0.084	0.000	0.000		
f51t60												0.000	0.000		
over60															
bujon															0.000
normal															

Tabulka 0.4 Četnosti odpovědí podle kategorií

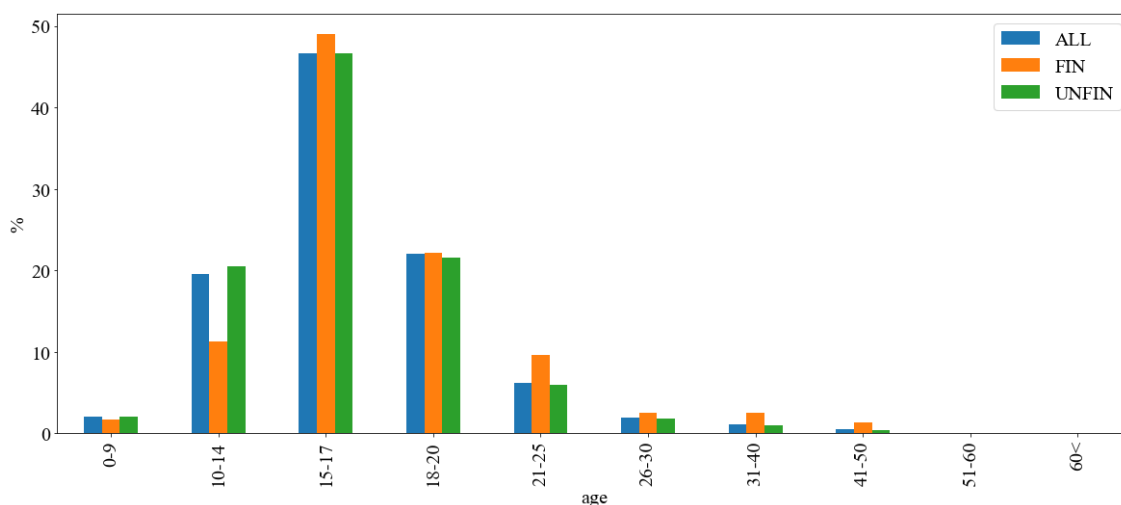
	ALL	FIN	UNFIN
0-9	91	4	86
10-14	896	27	836
15-17	2140	117	1905
18-20	1008	53	882
21-25	283	23	241
26-30	87	6	72
31-40	49	6	42
41-50	23	3	17
51-60	2	0	2
60<	2	0	2

Tabulka 0.5 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.004
UNFIN		

Tabulka 0.6 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
0-9	2	1.7	2.1
10-14	19.6	11.3	20.5
15-17	46.7	49	46.6
18-20	22	22.2	21.6
21-25	6.2	9.6	5.9
26-30	1.9	2.5	1.8
31-40	1.1	2.5	1
41-50	0.5	1.3	0.4
51-60	0		0
60<	0		0



Obrázek 0.1 Sloupcový graf procentuálního rozdělení v kategoriích

Který z následujících tabákových a nikotinových produktů jsi už zkoušel?

Tabulka 0.7 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
all	4461	2188	763	2299	1138	265	872	2658	1290	405	2229	848	4598
men	2285	1384	582	1194	729	205	562	1399	673	248	1112	526	2349
women	2176	804	181	1105	409	60	310	1259	617	157	1117	322	2249

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
light	1169	622	209	697	367	99	290	710	414	122	561	293	1205
middle	2372	1147	403	1228	578	121	429	1412	662	209	1221	411	2434
heavy	923	421	151	376	195	45	153	538	214	74	449	144	962
under18	298	209	65	168	176	52	153	252	223	55	163	192	308
f18t20	547	352	116	389	285	78	244	416	300	93	330	235	566
f21t25	900	479	162	640	293	63	229	591	334	94	547	167	930
f26t30	886	418	145	521	157	39	119	500	200	68	459	108	906
f31t40	1079	466	160	427	155	20	83	564	147	68	470	86	1105
f41t50	480	174	67	84	37	5	19	212	43	12	169	32	493
f51t60	129	34	19	19	7	1	4	56	13	3	33	7	133
over60	21	8	6	5	3	1	1	9	2	0	6	2	22
bujon	1924	961	326	1138	536	143	435	1183	630	209	1127	419	1979
normal	2491	1208	430	1141	592	121	431	1451	648	193	1082	424	2572

Tabulka 0.8 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
all	97	47.6	16.6	50	24.7	5.8	19	57.8	28.1	8.8	48.5	18.4
men	97.3	58.9	24.8	50.8	31	8.7	23.9	59.6	28.7	10.6	47.3	22.4
women	96.8	35.7	8	49.1	18.2	2.7	13.8	56	27.4	7	49.7	14.3
light	97	51.6	17.3	57.8	30.5	8.2	24.1	58.9	34.4	10.1	46.6	24.3
middle	97.5	47.1	16.6	50.5	23.7	5	17.6	58	27.2	8.6	50.2	16.9
heavy	95.9	43.8	15.7	39.1	20.3	4.7	15.9	55.9	22.2	7.7	46.7	15
under18	96.8	67.9	21.1	54.5	57.1	16.9	49.7	81.8	72.4	17.9	52.9	62.3
f18t20	96.6	62.2	20.5	68.7	50.4	13.8	43.1	73.5	53	16.4	58.3	41.5
f21t25	96.8	51.5	17.4	68.8	31.5	6.8	24.6	63.5	35.9	10.1	58.8	18
f26t30	97.8	46.1	16	57.5	17.3	4.3	13.1	55.2	22.1	7.5	50.7	11.9
f31t40	97.6	42.2	14.5	38.6	14	1.8	7.5	51	13.3	6.2	42.5	7.8
f41t50	97.4	35.3	13.6	17	7.5	1	3.9	43	8.7	2.4	34.3	6.5
f51t60	97	25.6	14.3	14.3	5.3	0.8	3	42.1	9.8	2.3	24.8	5.3
over60	95.5	36.4	27.3	22.7	13.6	4.5	4.5	40.9	9.1	0	27.3	9.1
bujon	97.2	48.6	16.5	57.5	27.1	7.2	22	59.8	31.8	10.6	56.9	21.2
normal	96.9	47	16.7	44.4	23	4.7	16.8	56.4	25.2	7.5	42.1	16.5

Tabulka 0.9 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light			0.000	0.000											
middle				0.147											
heavy															

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
under18							0.000	0.000	0.000	0.000	0.000	0.000	0.001		
f18t20								0.000	0.000	0.000	0.000	0.000	0.004		
f21t25									0.000	0.000	0.000	0.000	0.077		
f26t30										0.000	0.000	0.000	0.419		
f31t40											0.000	0.001	0.699		
f41t50												0.907	0.816		
f51t60													0.836		
over60															
bujon															0.000
normal															

Tabulka 0.10 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Cigarety	4461	234	3979
2. Doutník	2188	122	1946
3. Dýmka	763	45	675
4. Vodní dýmka	2299	134	2028
5. Šňupací tabák	1138	50	1041
6. Bezdýmový tabák	265	8	250
7. Žvýkací tabák	872	32	809
8. Ecig s nikotinem	2658	141	2376
9. Ecig bez nikotinu	1290	49	1188
10. Ecig THC/CBD	405	17	372
11. Zahříváný tabák	2229	131	1951
12. Nikotinové sáčky	848	30	782
13. Počet uživatelů	4598	238	4101

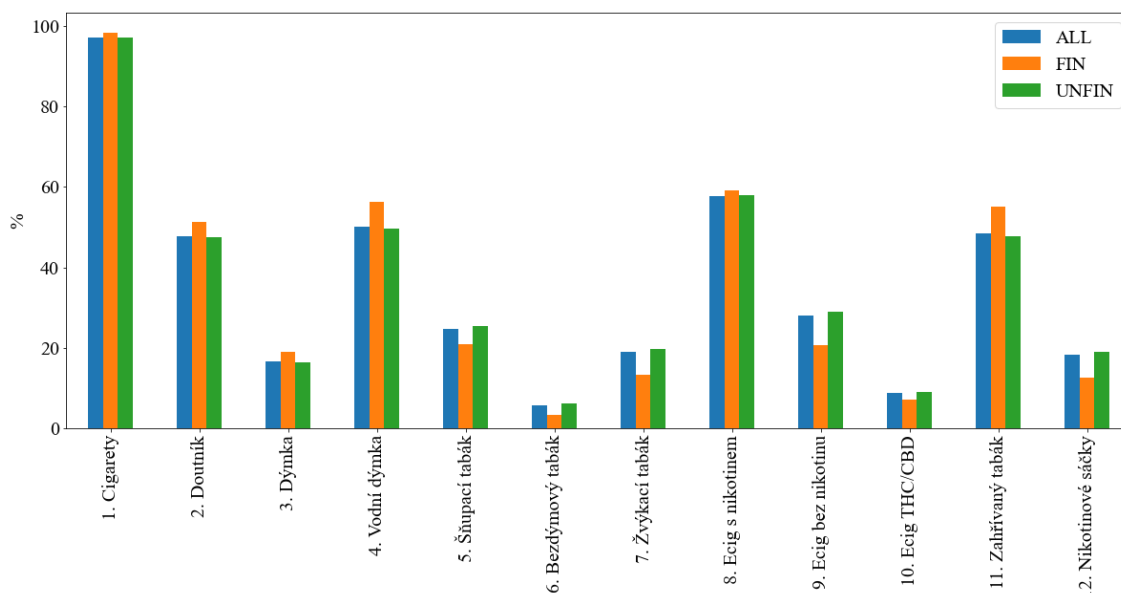
Tabulka 0.11 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.004
UNFIN		

Tabulka 0.12 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Cigarety	97	98.3	97

	ALL	FIN	UNFIN
2. Doutník	47.6	51.3	47.5
3. Dýmka	16.6	18.9	16.5
4. Vodní dýmka	50	56.3	49.5
5. Šňupací tabák	24.7	21	25.4
6. Bezdýmový tabák	5.8	3.4	6.1
7. Žvýkací tabák	19	13.4	19.7
8. Ecig s nikotinem	57.8	59.2	57.9
9. Ecig bez nikotinu	28.1	20.6	29
10. Ecig THC/CBD	8.8	7.1	9.1
11. Zahříváný tabák	48.5	55	47.6
12. Nikotinové sáčky	18.4	12.6	19.1



Obrázek 0.2 Sloupcový graf procentuálního rozdělení v kategoriích

Který z těchto produktů užíváš v současnosti (alespoň jednou týdně)?

Tabulka 0.13 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
all	2207	84	15	52	20	18	49	261	71	63	483	128	2598
men	1212	59	12	36	14	13	38	153	38	41	212	79	1382
women	995	25	3	16	6	5	11	108	33	22	271	49	1216
light	489	27	2	12	6	6	19	55	24	13	78	38	582
middle	1181	36	5	22	7	4	22	151	28	37	268	61	1381
heavy	538	21	8	18	7	8	8	55	19	13	137	29	636
under18	159	16	1	10	6	6	21	34	23	13	20	36	184
f18t20	192	14	1	6	2	6	13	29	10	6	37	32	222

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
f21t25	328	12	1	13	4	1	8	47	15	12	86	20	397
f26t30	438	15	0	10	3	2	5	52	5	13	112	13	520
f31t40	611	19	9	11	2	2	0	55	9	16	146	11	737
f41t50	324	2	2	1	2	1	0	29	5	3	59	13	367
f51t60	101	0	0	0	0	0	1	7	2	0	15	1	107
over60	12	3	1	1	1	0	0	3	1	0	0	0	14
bujon	1	0	0	0	0	0	0	0	0	0	0	0	1
normal	2172	83	14	52	20	18	49	257	71	63	479	125	2560

Tabulka 0.14 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
all	84.9	3.2	0.6	2	0.8	0.7	1.9	10	2.7	2.4	18.6	4.9
men	87.7	4.3	0.9	2.6	1	0.9	2.7	11.1	2.7	3	15.3	5.7
women	81.8	2.1	0.2	1.3	0.5	0.4	0.9	8.9	2.7	1.8	22.3	4
light	84	4.6	0.3	2.1	1	1	3.3	9.5	4.1	2.2	13.4	6.5
middle	85.5	2.6	0.4	1.6	0.5	0.3	1.6	10.9	2	2.7	19.4	4.4
heavy	84.6	3.3	1.3	2.8	1.1	1.3	1.3	8.6	3	2	21.5	4.6
under18	86.4	8.7	0.5	5.4	3.3	3.3	11.4	18.5	12.5	7.1	10.9	19.6
f18t20	86.5	6.3	0.5	2.7	0.9	2.7	5.9	13.1	4.5	2.7	16.7	14.4
f21t25	82.6	3	0.3	3.3	1	0.3	2	11.8	3.8	3	21.7	5
f26t30	84.2	2.9	0	1.9	0.6	0.4	1	10	1	2.5	21.5	2.5
f31t40	82.9	2.6	1.2	1.5	0.3	0.3	0	7.5	1.2	2.2	19.8	1.5
f41t50	88.3	0.5	0.5	0.3	0.5	0.3	0	7.9	1.4	0.8	16.1	3.5
f51t60	94.4	0	0	0	0	0	0.9	6.5	1.9	0	14	0.9
over60	85.7	21.4	7.1	7.1	7.1	0	0	21.4	7.1	0	0	0
bujon	100	0	0	0	0	0	0	0	0	0	0	0
normal	84.8	3.2	0.5	2	0.8	0.7	1.9	10	2.8	2.5	18.7	4.9

Tabulka 0.15 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.000	0.007										
middle					0.013										
heavy															
under18						0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.081		
f18t20							0.001	0.000	0.000	0.000	0.000	0.000	0.019		
f21t25								0.108	0.000	0.000	0.024	0.001			
f26t30									0.084	0.007	0.000	0.000			
f31t40										0.000	0.054	0.000			
f41t50											0.575	0.000			
f51t60												0.000			

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
over60															
bujon															1.000
normal															

Tabulka 0.16 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Cigarety	2207	120	1970
2. Doutník	84	3	78
3. Dýmka	15	3	12
4. Vodní dýmka	52	2	48
5. Šňupací tabák	20	0	19
6. Bezdýmový tabák	18	0	17
7. Žvýkací tabák	49	0	49
8. Ecig s nikotinem	261	19	234
9. Ecig bez nikotinu	71	1	69
10. Ecig THC/CBD	63	3	54
11. Zahříváný tabák	483	27	420
12. Nikotinové sáčky	128	3	123
13. Počet uživatelů	2598	143	2315

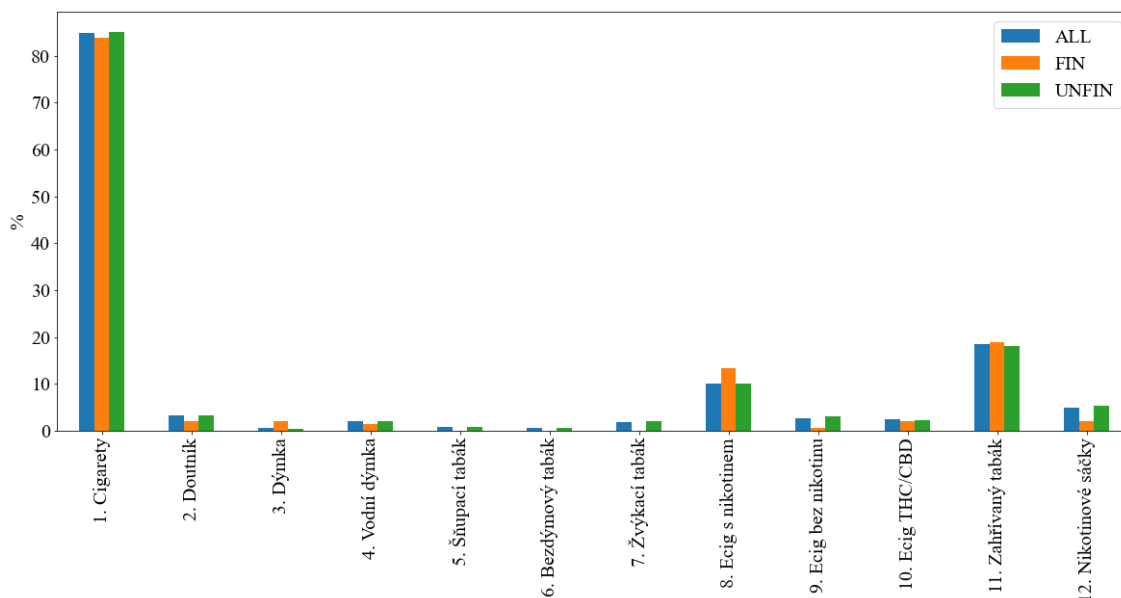
Tabulka 0.17 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.067
UNFIN		

Tabulka 0.18 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Cigarety	84.9	83.9	85.1
2. Doutník	3.2	2.1	3.4
3. Dýmka	0.6	2.1	0.5
4. Vodní dýmka	2	1.4	2.1
5. Šňupací tabák	0.8	0	0.8
6. Bezdýmový tabák	0.7	0	0.7
7. Žvýkací tabák	1.9	0	2.1
8. Ecig s nikotinem	10	13.3	10.1
9. Ecig bez nikotinu	2.7	0.7	3
10. Ecig THC/CBD	2.4	2.1	2.3

	ALL	FIN	UNFIN
11. Zahříváný tabák	18.6	18.9	18.1
12. Nikotinové sáčky	4.9	2.1	5.3



Obrázek 0.3 Sloupcový graf procentuálního rozdělení v kategoriích

Proč ses rozhodl přestat kouřit?

Tabulka 0.19 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.
all	3985	1805	2973	764	233	3126	233	303	4582
men	2059	951	1527	381	122	1646	120	132	2339
women	1926	854	1446	383	111	1480	113	171	2243
light	1069	451	708	240	61	837	73	117	1203
middle	2107	961	1614	382	123	1652	117	151	2425
heavy	810	392	653	142	49	639	43	37	957
under18	247	116	184	66	9	210	16	32	308
f18t20	491	169	396	126	21	394	20	65	565
f21t25	826	309	707	167	59	660	35	75	930
f26t30	794	375	617	152	59	628	45	57	904
f31t40	963	554	677	159	48	746	71	46	1102
f41t50	432	203	246	57	21	323	28	11	492
f51t60	108	35	62	18	10	85	13	3	133
over60	18	7	11	6	0	12	1	0	22
bujon	1868	733	1281	360	96	1339	86	301	1979
normal	2077	1054	1667	400	135	1758	145	0	2557

Tabulka 0.20 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.
all	87	39.4	64.9	16.7	5.1	68.2	5.1	6.6
men	88	40.7	65.3	16.3	5.2	70.4	5.1	5.6
women	85.9	38.1	64.5	17.1	4.9	66	5	7.6
light	88.9	37.5	58.9	20	5.1	69.6	6.1	9.7
middle	86.9	39.6	66.6	15.8	5.1	68.1	4.8	6.2
heavy	84.6	41	68.2	14.8	5.1	66.8	4.5	3.9
under18	80.2	37.7	59.7	21.4	2.9	68.2	5.2	10.4
f18t20	86.9	29.9	70.1	22.3	3.7	69.7	3.5	11.5
f21t25	88.8	33.2	76	18	6.3	71	3.8	8.1
f26t30	87.8	41.5	68.3	16.8	6.5	69.5	5	6.3
f31t40	87.4	50.3	61.4	14.4	4.4	67.7	6.4	4.2
f41t50	87.8	41.3	50	11.6	4.3	65.7	5.7	2.2
f51t60	81.2	26.3	46.6	13.5	7.5	63.9	9.8	2.3
over60	81.8	31.8	50	27.3	0	54.5	4.5	0
bujon	94.4	37	64.7	18.2	4.9	67.7	4.3	15.2
normal	81.2	41.2	65.2	15.6	5.3	68.8	5.7	0

 Tabulka 0.21 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.135													
women															
light				0.000	0.000										
middle					0.312										
heavy															
under18							0.243	0.009	0.019	0.000	0.000	0.002	0.796		
f18t20								0.037	0.000	0.000	0.000	0.000	0.690		
f21t25									0.024	0.000	0.000	0.001	0.539		
f26t30										0.001	0.000	0.015	0.595		
f31t40											0.139	0.008	0.495		
f41t50												0.121	0.448		
f51t60													0.592		
over60															
bujon															0.000
normal															

Tabulka 0.22 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Zdraví	3985	216	3532
2. Rodina	1805	118	1558
3. Finance	2973	144	2675

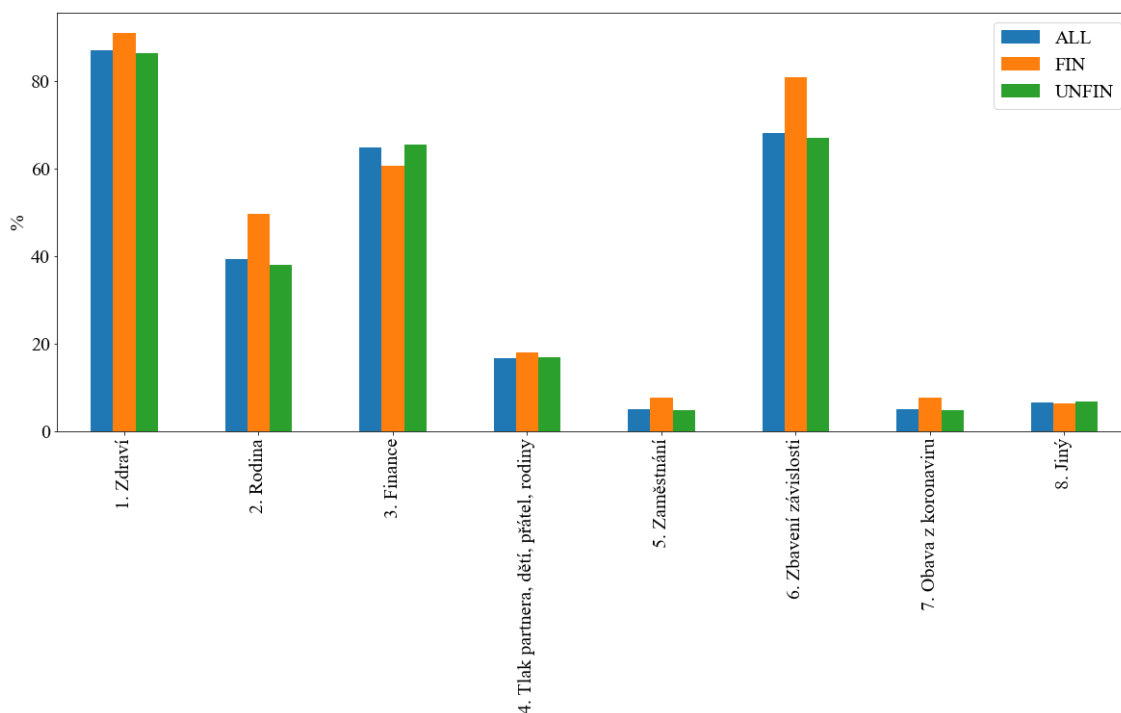
	ALL	FIN	UNFIN
4. Tlak partnera, dětí, přátel, rodiny	764	43	691
5. Zaměstnání	233	18	200
6. Zbavení závislosti	3126	192	2737
7. Obava z koronaviru	233	18	197
8. Jiný	303	15	275
9. Počet uživatelů	4582	237	4086

Tabulka 0.23 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.061
UNFIN		

Tabulka 0.24 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Zdraví	87	91.1	86.4
2. Rodina	39.4	49.8	38.1
3. Finance	64.9	60.8	65.5
4. Tlak partnera, dětí, přátel, rodiny	16.7	18.1	16.9
5. Zaměstnání	5.1	7.6	4.9
6. Zbavení závislosti	68.2	81	67
7. Obava z koronaviru	5.1	7.6	4.8
8. Jiný	6.6	6.3	6.7



Obrázek 0.4 Sloupcový graf procentuálního rozdělení v kategoriích

Kolikrát v životě jsi už zkoušel přestat kouřit?

Tabulka 0.25 Četnosti odpovědí podle kategorií

Počet pokusů	0	1	2-3	4-5	6-10	11<	ALL
all	510	813	985	443	224	1585	4560
men	223	379	542	244	133	810	2331
women	287	434	443	199	91	775	2229
light	116	223	242	90	53	472	1196
middle	255	422	533	234	125	845	2414
heavy	139	168	212	118	46	270	953
under18	55	74	67	15	14	83	308
f18t20	55	136	86	34	17	237	565
f21t25	91	176	160	68	30	403	928
f26t30	113	161	204	76	43	295	892
f31t40	121	146	283	145	69	332	1096
f41t50	47	80	136	70	34	126	493
f51t60	18	14	32	24	12	33	133
over60	1	2	3	1	0	15	22
bujon	0	489	18	0	0	1456	1963
normal	507	316	954	436	219	119	2551

Tabulka 0.26 Procentuální rozložení odpovědí podle kategorií

Počet pokusů	0	1	2-3	4-5	6-10	11<
all	11.2	17.8	21.6	9.7	4.9	34.8
men	9.6	16.3	23.3	10.5	5.7	34.7
women	12.9	19.5	19.9	8.9	4.1	34.8
light	9.7	18.6	20.2	7.5	4.4	39.5
middle	10.6	17.5	22.1	9.7	5.2	35
heavy	14.6	17.6	22.2	12.4	4.8	28.3
under18	17.9	24	21.8	4.9	4.5	26.9
f18t20	9.7	24.1	15.2	6	3	41.9
f21t25	9.8	19	17.2	7.3	3.2	43.4
f26t30	12.7	18	22.9	8.5	4.8	33.1
f31t40	11	13.3	25.8	13.2	6.3	30.3
f41t50	9.5	16.2	27.6	14.2	6.9	25.6
f51t60	13.5	10.5	24.1	18	9	24.8
over60	4.5	9.1	13.6	4.5	0	68.2
bujon	0	24.9	0.9	0	0	74.2
normal	19.9	12.4	37.4	17.1	8.6	4.7

Tabulka 0.27 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.036	0.000										
middle					0.000										
heavy															
under18							0.000	0.000	0.008	0.000	0.000	0.000	0.004		
f18t20								0.272	0.000	0.000	0.000	0.000	0.237		
f21t25									0.000	0.000	0.000	0.000	0.323		
f26t30										0.000	0.000	0.001	0.032		
f31t40											0.285	0.302	0.010		
f41t50												0.300	0.001		
f51t60													0.003		
over60															
bujon															0.000
normal															

Tabulka 0.28 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
0	510	11	480
01	813	31	744
02-03	985	47	880

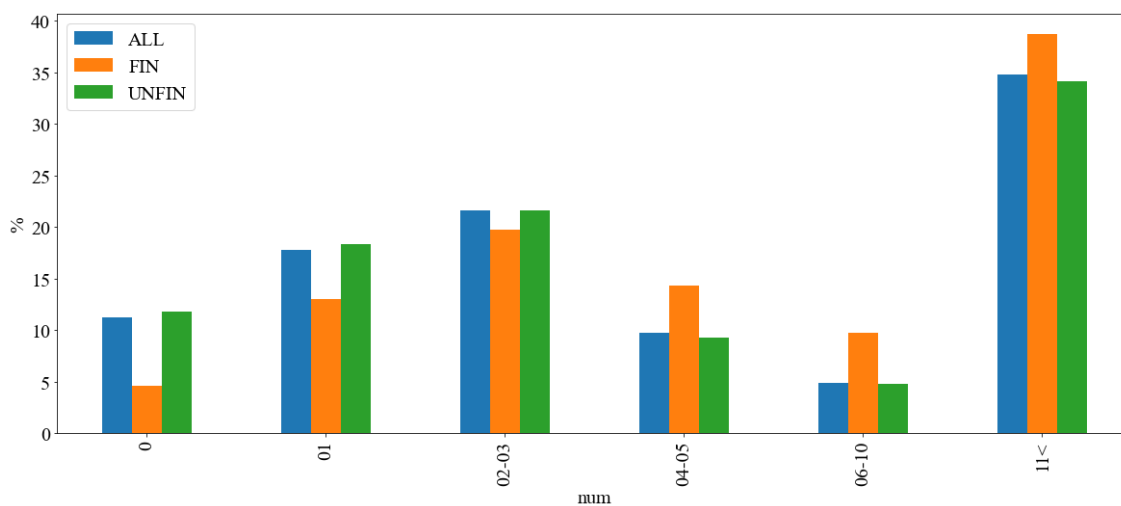
	ALL	FIN	UNFIN
04-05	443	34	380
06-10	224	23	197
11<	1585	92	1385

Tabulka 0.29 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.30 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
0	11.2	4.6	11.8
01	17.8	13	18.3
02-03	21.6	19.7	21.6
04-05	9.7	14.3	9.3
06-10	4.9	9.7	4.8
11<	34.8	38.7	34.1



Obrázek 0.5 Sloupcový graf procentuálního rozdělení v kategoriích

Jak dlouho jsi vydržel nekouřit při posledním pokusu?

Tabulka 0.31 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	ALL
all	486	985	894	508	193	200	289	3555
men	231	536	501	285	113	90	142	1898
women	255	449	393	223	80	110	147	1657
light	68	224	288	151	54	59	77	921
middle	267	550	450	269	102	111	158	1907
heavy	151	211	158	89	37	30	54	730
under18	41	71	69	17	6	4	3	211
f18t20	61	150	117	47	13	8	5	401
f21t25	102	229	178	94	31	42	31	707
f26t30	92	175	165	117	50	30	46	675
f31t40	112	211	219	145	55	71	98	911
f41t50	44	95	99	60	27	24	73	422
f51t60	13	24	17	19	7	17	18	115
over60	3	6	4	1	1	0	6	21
bujon	188	379	379	230	82	91	123	1472
normal	292	596	503	272	109	107	164	2043

Tabulka 0.32 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.
all	13.7	27.7	25.1	14.3	5.4	5.6	8.1
men	12.2	28.2	26.4	15	6	4.7	7.5
women	15.4	27.1	23.7	13.5	4.8	6.6	8.9
light	7.4	24.3	31.3	16.4	5.9	6.4	8.4
middle	14	28.8	23.6	14.1	5.3	5.8	8.3
heavy	20.7	28.9	21.6	12.2	5.1	4.1	7.4
under18	19.4	33.6	32.7	8.1	2.8	1.9	1.4
f18t20	15.2	37.4	29.2	11.7	3.2	2	1.2
f21t25	14.4	32.4	25.2	13.3	4.4	5.9	4.4
f26t30	13.6	25.9	24.4	17.3	7.4	4.4	6.8
f31t40	12.3	23.2	24	15.9	6	7.8	10.8
f41t50	10.4	22.5	23.5	14.2	6.4	5.7	17.3
f51t60	11.3	20.9	14.8	16.5	6.1	14.8	15.7
over60	14.3	28.6	19	4.8	4.8	0	28.6
bujon	12.8	25.7	25.7	15.6	5.6	6.2	8.4
normal	14.3	29.2	24.6	13.3	5.3	5.2	8

Tabulka 0.33 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.002													
women															
light				0.000	0.000										
middle					0.002										
heavy															
under18						0.614	0.003	0.000	0.000	0.000	0.000	0.000	0.000		
f18t20							0.002	0.000	0.000	0.000	0.000	0.000	0.000		
f21t25								0.004	0.000	0.000	0.000	0.000	0.000		
f26t30									0.010	0.000	0.000	0.000	0.013		
f31t40										0.044	0.059	0.138			
f41t50											0.034	0.573			
f51t60												0.292			
over60															
bujon															0.114
normal															

Tabulka 0.34 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Méně než 1 den	486	26	435
2. Méně než 1 týden	985	57	876
3. 1 týden až 1 měsíc	894	53	780
4. 1-3 měsíce	508	36	440
5. 3-6 měsíců	193	9	172
6. Půl roku až rok	200	12	173
7. 1 rok a více	289	19	249

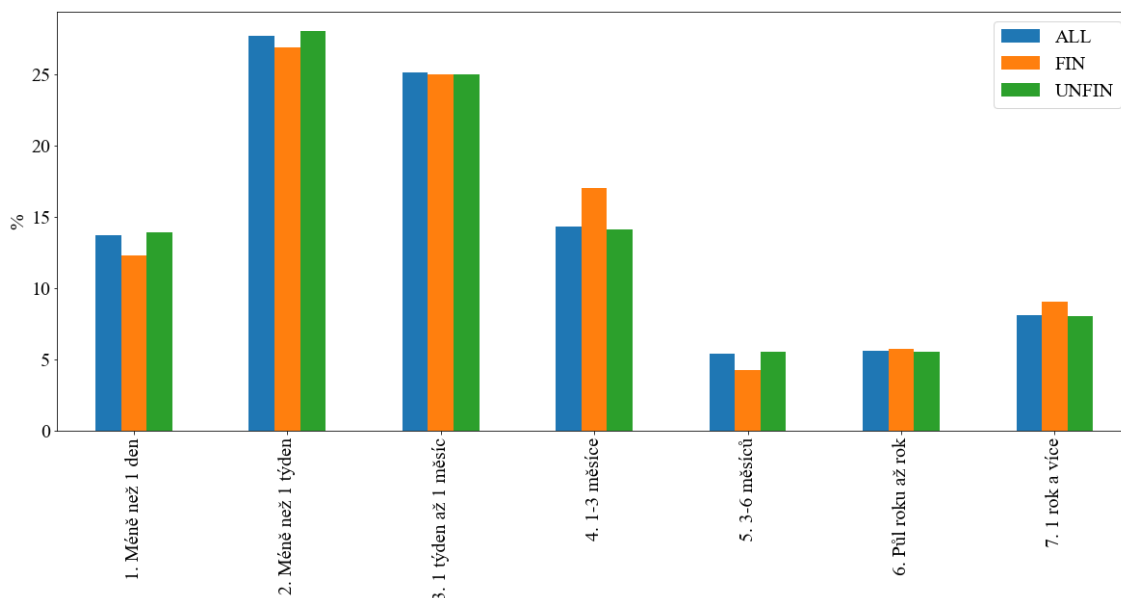
Tabulka 0.35 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.870
UNFIN		

Tabulka 0.36 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Méně než 1 den	13.7	12.3	13.9
2. Méně než 1 týden	27.7	26.9	28
3. 1 týden až 1 měsíc	25.1	25	25
4. 1-3 měsíce	14.3	17	14.1

	ALL	FIN	UNFIN
5. 3-6 měsíců	5.4	4.2	5.5
6. Půl roku až rok	5.6	5.7	5.5
7. 1 rok a více	8.1	9	8



Obrázek 0.6 Sloupcový graf procentuálního rozdělení v kategoriích

Jak ses o odvykání pokoušel?

Tabulka 0.37 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
all	2464	593	145	39	33	14	30	77	237	74	43	3493
men	1323	323	89	21	20	7	12	39	106	42	14	1867
women	1141	270	56	18	13	7	18	38	131	32	29	1626
light	703	102	17	6	4	4	4	20	61	8	13	902
middle	1318	318	82	22	15	7	19	43	132	39	22	1881
heavy	442	175	46	11	14	3	7	14	46	27	8	713
under18	165	18	0	0	1	0	0	6	14	0	5	203
f18t20	326	29	1	1	1	2	2	9	20	2	8	394
f21t25	559	72	9	4	5	4	2	13	45	7	5	698
f26t30	475	123	17	2	3	2	10	16	55	8	7	672
f31t40	575	184	57	16	8	3	8	22	75	26	14	892
f41t50	218	125	45	12	12	3	6	8	21	19	1	413
f51t60	73	21	9	3	2	0	2	1	4	7	1	111
over60	9	6	4	1	0	0	0	0	1	2	0	20
bujon	1077	212	59	13	12	8	13	26	92	30	43	1482
normal	1372	365	83	26	21	6	17	50	139	42	0	1974

Tabulka 0.38 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
all	70.5	17	4.2	1.1	0.9	0.4	0.9	2.2	6.8	2.1	1.2
men	70.9	17.3	4.8	1.1	1.1	0.4	0.6	2.1	5.7	2.2	0.7
women	70.2	16.6	3.4	1.1	0.8	0.4	1.1	2.3	8.1	2	1.8
light	77.9	11.3	1.9	0.7	0.4	0.4	0.4	2.2	6.8	0.9	1.4
middle	70.1	16.9	4.4	1.2	0.8	0.4	1	2.3	7	2.1	1.2
heavy	62	24.5	6.5	1.5	2	0.4	1	2	6.5	3.8	1.1
under18	81.3	8.9	0	0	0.5	0	0	3	6.9	0	2.5
f18t20	82.7	7.4	0.3	0.3	0.3	0.5	0.5	2.3	5.1	0.5	2
f21t25	80.1	10.3	1.3	0.6	0.7	0.6	0.3	1.9	6.4	1	0.7
f26t30	70.7	18.3	2.5	0.3	0.4	0.3	1.5	2.4	8.2	1.2	1
f31t40	64.5	20.6	6.4	1.8	0.9	0.3	0.9	2.5	8.4	2.9	1.6
f41t50	52.8	30.3	10.9	2.9	2.9	0.7	1.5	1.9	5.1	4.6	0.2
f51t60	65.8	18.9	8.1	2.7	1.8	0	1.8	0.9	3.6	6.3	0.9
over60	45	30	20	5	0	0	0	0	5	10	0
bujon	72.7	14.3	4	0.9	0.8	0.5	0.9	1.8	6.2	2	2.9
normal	69.5	18.5	4.2	1.3	1.1	0.3	0.9	2.5	7	2.1	0

Tabulka 0.39 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.014													
women															
light				0.000	0.000										
middle					0.000										
heavy															
under18						0.826	0.205	0.004	0.000	0.000	0.000	0.000	0.000		
f18t20							0.233	0.000	0.000	0.000	0.000	0.000	0.000		
f21t25								0.000	0.000	0.000	0.000	0.000	0.000		
f26t30									0.001	0.000	0.000	0.000	0.000		
f31t40										0.000	0.379	0.242			
f41t50											0.342	0.925			
f51t60												0.000			
over60															
bujon															0.000
normal															

Tabulka 0.40 Četnosti odpovědí podle kategorií

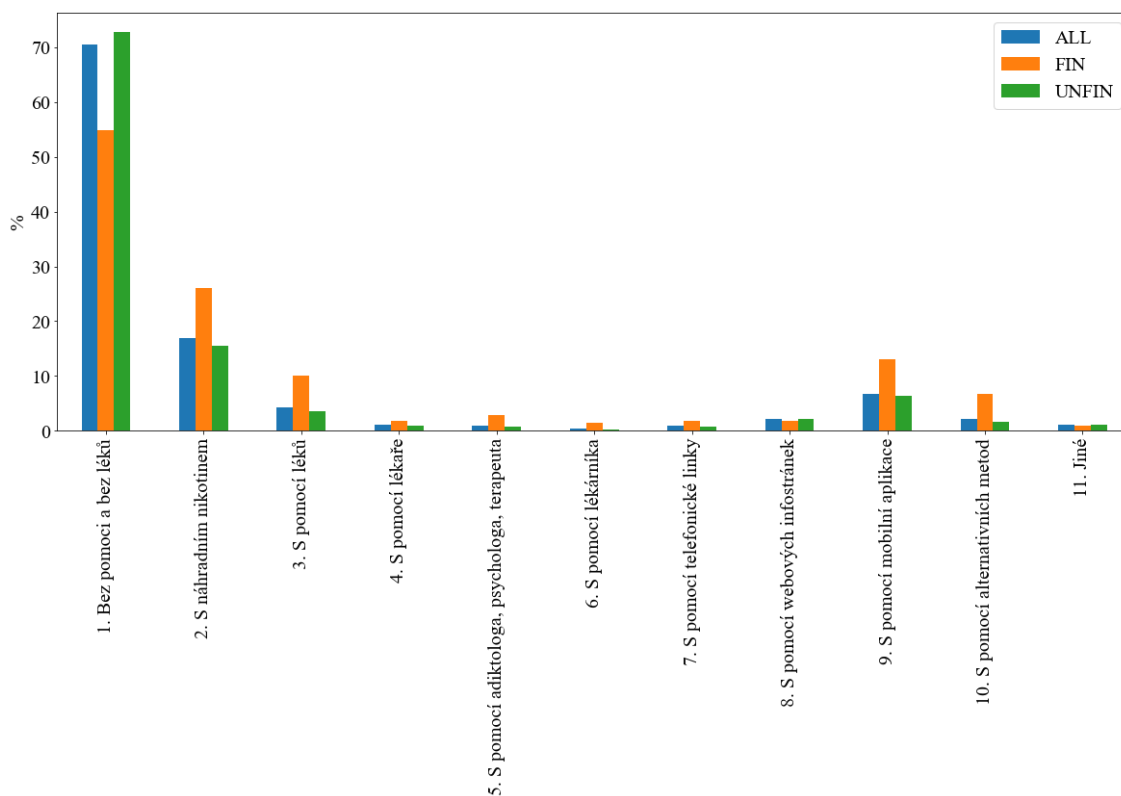
	ALL	FIN	UNFIN
1. Bez pomoci a bez léků	2464	114	2233
2. S náhradním nikotinem	593	54	478
3. S pomocí léků	145	21	108
4. S pomocí lékaře	39	4	31
5. S pomocí adiktologa, psychologa, terapeuta	33	6	24
6. S pomocí lékárníka	14	3	10
7. S pomocí telefonické linky	30	4	21
8. S pomocí webových infostránek	77	4	67
9. S pomocí mobilní aplikace	237	27	197
10. S pomocí alternativních metod	74	14	49
11. Jiné	43	2	35
12. Počet uživatelů	3493	208	3071

Tabulka 0.41 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.42 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Bez pomoci a bez léků	70.5	54.8	72.7
2. S náhradním nikotinem	17	26	15.6
3. S pomocí léků	4.2	10.1	3.5
4. S pomocí lékaře	1.1	1.9	1
5. S pomocí adiktologa, psychologa, terapeuta	0.9	2.9	0.8
6. S pomocí lékárníka	0.4	1.4	0.3
7. S pomocí telefonické linky	0.9	1.9	0.7
8. S pomocí webových infostránek	2.2	1.9	2.2
9. S pomocí mobilní aplikace	6.8	13	6.4
10. S pomocí alternativních metod	2.1	6.7	1.6
11. Jiné	1.2	1	1.1



Obrázek 0.7 Sloupcový graf procentuálního rozdělení v kategoriích

Kde momentálně žiješ?

Tabulka 0.43 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	ALL
all	824	815	1126	723	1050	4538
men	410	407	583	349	554	2303
women	414	408	543	374	496	2235
light	242	221	308	187	244	1202
middle	411	445	580	391	572	2399
heavy	176	149	238	145	233	941
under18	27	40	81	60	100	308
f18t20	72	86	160	95	145	558
f21t25	178	184	223	167	185	937
f26t30	179	190	213	140	186	908
f31t40	233	192	261	155	257	1098
f41t50	90	78	133	79	113	493
f51t60	27	30	32	14	33	136
over60	5	2	3	4	10	24
bujon	411	362	475	288	392	1928
normal	401	445	642	430	646	2564

Tabulka 0.44 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.
all	18.2	18	24.8	15.9	23.1
men	17.8	17.7	25.3	15.2	24.1
women	18.5	18.3	24.3	16.7	22.2
light	20.1	18.4	25.6	15.6	20.3
middle	17.1	18.5	24.2	16.3	23.8
heavy	18.7	15.8	25.3	15.4	24.8
under18	8.8	13	26.3	19.5	32.5
f18t20	12.9	15.4	28.7	17	26
f21t25	19	19.6	23.8	17.8	19.7
f26t30	19.7	20.9	23.5	15.4	20.5
f31t40	21.2	17.5	23.8	14.1	23.4
f41t50	18.3	15.8	27	16	22.9
f51t60	19.9	22.1	23.5	10.3	24.3
over60	20.8	8.3	12.5	16.7	41.7
bujon	21.3	18.8	24.6	14.9	20.3
normal	15.6	17.4	25	16.8	25.2

 Tabulka 0.45 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.344													
women															
light				0.052	0.121										
middle					0.324										
heavy															
under18							0.100	0.000	0.000	0.000	0.000	0.000	0.193		
f18t20								0.000	0.000	0.000	0.176	0.027	0.190		
f21t25									0.690	0.040	0.189	0.232	0.077		
f26t30										0.188	0.122	0.560	0.090		
f31t40											0.365	0.584	0.214		
f41t50												0.240	0.190		
f51t60													0.192		
over60															
bujon															0.000
normal															

Tabulka 0.46 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Víc než 1 m.	824	56	699
2. Nad 100 tis.	815	39	737
3. Nad 10 tis.	1126	60	1012

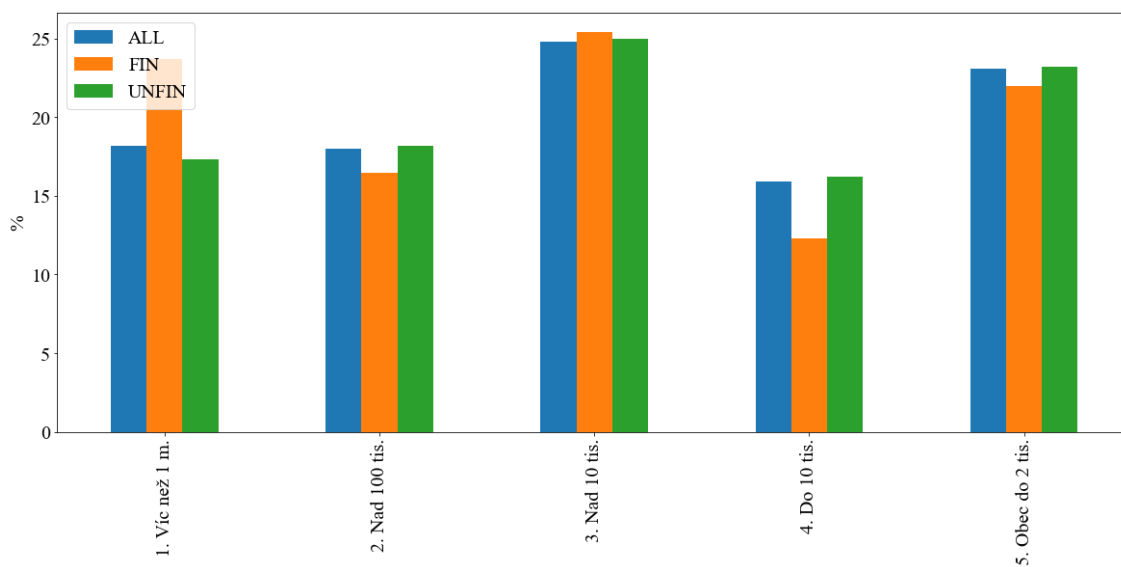
	ALL	FIN	UNFIN
4. Do 10 tis.	723	29	657
5. Obec do 2 tis.	1050	52	939

Tabulka 0.47 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.094
UNFIN		

Tabulka 0.48 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Víc než 1 m.	18.2	23.7	17.3
2. Nad 100 tis.	18	16.5	18.2
3. Nad 10 tis.	24.8	25.4	25
4. Do 10 tis.	15.9	12.3	16.2
5. Obec do 2 tis.	23.1	22	23.2



Obrázek 0.8 Sloupcový graf procentuálního rozdělení v kategoriích

Jak jsi na tom se zaměstnáním?

Tabulka 0.49 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
all	2543	346	491	316	669	163	248	56	63	61	4547
men	1459	140	342	135	289	63	3	30	22	25	2303
women	1084	206	149	181	380	100	245	26	41	36	2244
light	562	119	85	83	318	66	64	11	11	27	1204
middle	1435	158	273	170	291	72	121	23	38	28	2405
heavy	548	71	133	63	60	27	63	22	14	6	942
under18	2	21	3	37	206	47	3	2	0	18	308
f18t20	144	75	25	53	273	42	13	1	4	9	560
f21t25	554	84	78	74	151	44	62	2	11	16	938
f26t30	600	58	120	46	19	16	89	7	17	5	917
f31t40	751	61	162	72	9	12	74	13	15	4	1101
f41t50	358	33	74	20	1	2	4	7	9	6	491
f51t60	93	9	17	6	0	0	0	8	3	1	135
over60	6	2	0	2	0	0	0	15	1	0	24
bujon	1014	146	256	124	364	82	103	11	22	61	1943
normal	1499	197	229	190	300	80	143	45	40	0	2559

Tabulka 0.50 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
all	55.9	7.6	10.8	6.9	14.7	3.6	5.5	1.2	1.4	1.3
men	63.4	6.1	14.9	5.9	12.5	2.7	0.1	1.3	1	1.1
women	48.3	9.2	6.6	8.1	16.9	4.5	10.9	1.2	1.8	1.6
light	46.7	9.9	7.1	6.9	26.4	5.5	5.3	0.9	0.9	2.2
middle	59.7	6.6	11.4	7.1	12.1	3	5	1	1.6	1.2
heavy	58.2	7.5	14.1	6.7	6.4	2.9	6.7	2.3	1.5	0.6
under18	0.6	6.8	1	12	66.9	15.3	1	0.6	0	5.8
f18t20	25.7	13.4	4.5	9.5	48.8	7.5	2.3	0.2	0.7	1.6
f21t25	59.1	9	8.3	7.9	16.1	4.7	6.6	0.2	1.2	1.7
f26t30	65.4	6.3	13.1	5	2.1	1.7	9.7	0.8	1.9	0.5
f31t40	68.2	5.5	14.7	6.5	0.8	1.1	6.7	1.2	1.4	0.4
f41t50	72.9	6.7	15.1	4.1	0.2	0.4	0.8	1.4	1.8	1.2
f51t60	68.9	6.7	12.6	4.4	0	0	0	5.9	2.2	0.7
over60	25	8.3	0	8.3	0	0	0	62.5	4.2	0
bujon	52.2	7.5	13.2	6.4	18.7	4.2	5.3	0.6	1.1	3.1
normal	58.6	7.7	8.9	7.4	11.7	3.1	5.6	1.8	1.6	0

Tabulka 0.51 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.000	0.000										
middle					0.000										
heavy															
under18						0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f18t20							0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f21t25								0.000	0.000	0.000	0.000	0.000	0.000		
f26t30									0.027	0.000	0.000	0.000	0.000		
f31t40										0.000	0.000	0.000	0.000		
f41t50											0.213	0.000	0.000		
f51t60												0.000	0.000		
over60															
bujon															0.000
normal															

Tabulka 0.52 Četnosti odpovědí podle kategorií

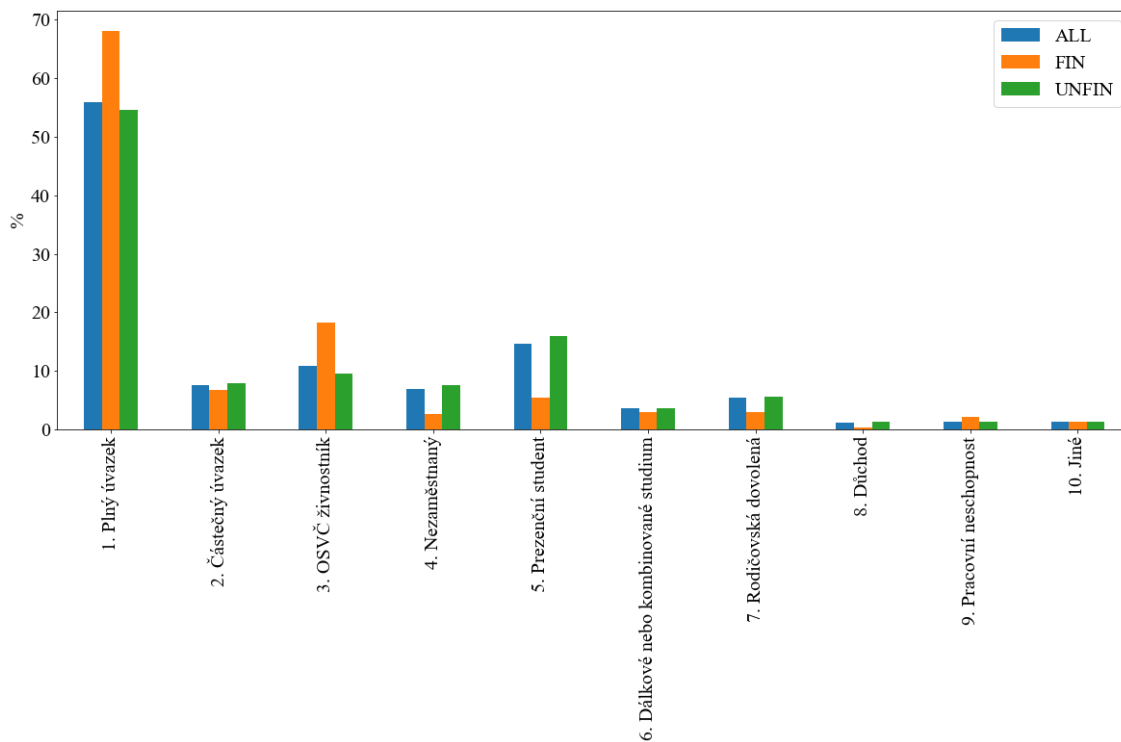
	ALL	FIN	UNFIN
1. Plný úvazek	2543	160	2211
2. Částečný úvazek	346	16	319
3. OSVČ živnostník	491	43	391
4. Nezaměstnaný	316	6	308
5. Prezenční student	669	13	644
6. Dálkové nebo kombinované studium	163	7	146
7. Rodičovská dovolená	248	7	225
8. Důchod	56	1	52
9. Pracovní neschopnost	63	5	56
10. Jiné	61	3	56
11. Počet uživatelů	4547	235	4053

Tabulka 0.53 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.54 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Plný úvazek	55.9	68.1	54.6
2. Částečný úvazek	7.6	6.8	7.9
3. OSVČ živnostník	10.8	18.3	9.6
4. Nezaměstnaný	6.9	2.6	7.6
5. Prezenční student	14.7	5.5	15.9
6. Dálkové nebo kombinované studium	3.6	3	3.6
7. Rodičovská dovolená	5.5	3	5.6
8. Důchod	1.2	0.4	1.3
9. Pracovní neschopnost	1.4	2.1	1.4
10. Jiné	1.3	1.3	1.4



Obrázek 0.9 Sloupcový graf procentuálního rozdělení v kategoriích

Nejvyšší dosažené vzdělání?

Tabulka 0.55 Četnosti odpovědí podle kategorií

	1. Základní	2. Střední	3. VOŠ	4. Bc	5. Mgr	6. Dr a vyšší	ALL
all	811	2759	288	322	300	45	4525
men	363	1463	162	137	136	32	2293
women	448	1296	126	185	164	13	2232
light	262	661	64	106	90	14	1197

	1. Základní	2. Střední	3. VOŠ	4. Bc	5. Mgr	6. Dr a vyšší	ALL
middle	394	1496	158	167	161	17	2393
heavy	155	602	68	49	51	14	939
under18	228	73	6	0	0	0	307
f18t20	210	308	23	12	1	4	558
f21t25	138	659	48	75	14	2	936
f26t30	87	581	46	96	90	6	906
f31t40	94	679	88	92	125	18	1096
f41t50	27	338	43	27	50	7	492
f51t60	10	71	23	10	14	7	135
over60	6	10	4	2	1	1	24
bujon	339	1140	110	165	158	15	1927
normal	467	1592	177	153	134	29	2552

Tabulka 0.56 Procentuální rozložení odpovědí podle kategorií

	1. Základní	2. Střední	3. VOŠ	4. Bc	5. Mgr	6. Dr a vyšší
all	17.9	61	6.4	7.1	6.6	1
men	15.8	63.8	7.1	6	5.9	1.4
women	20.1	58.1	5.6	8.3	7.3	0.6
light	21.9	55.2	5.3	8.9	7.5	1.2
middle	16.5	62.5	6.6	7	6.7	0.7
heavy	16.5	64.1	7.2	5.2	5.4	1.5
under18	74.3	23.8	2	0	0	0
f18t20	37.6	55.2	4.1	2.2	0.2	0.7
f21t25	14.7	70.4	5.1	8	1.5	0.2
f26t30	9.6	64.1	5.1	10.6	9.9	0.7
f31t40	8.6	62	8	8.4	11.4	1.6
f41t50	5.5	68.7	8.7	5.5	10.2	1.4
f51t60	7.4	52.6	17	7.4	10.4	5.2
over60	25	41.7	16.7	8.3	4.2	4.2
bujon	17.6	59.2	5.7	8.6	8.2	0.8
normal	18.3	62.4	6.9	6	5.3	1.1

Tabulka 0.57 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.000	0.000										
middle					0.072										
heavy															
under18							0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f18t20								0.000	0.000	0.000	0.000	0.000	0.000		
f21t25									0.000	0.000	0.000	0.000	0.000		

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
f26t30										0.011	0.000	0.000	0.003		
f31t40											0.046	0.001	0.026		
f41t50												0.002	0.001		
f51t60													0.177		
over60															
bujon															0.000
normal															

Tabulka 0.58 Četnosti odpovědí podle kategorií

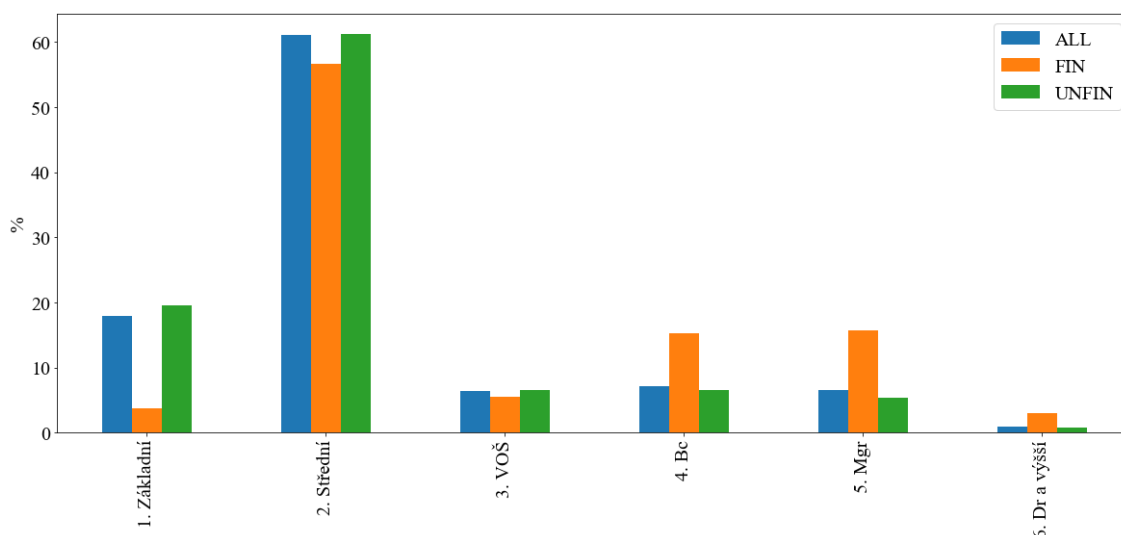
	ALL	FIN	UNFIN
1. Základní	811	9	788
2. Střední	2759	133	2468
3. VOŠ	288	13	262
4. Bc	322	36	266
5. Mgr	300	37	217
6. Dr a vyšší	45	7	31

Tabulka 0.59 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.60 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Základní	17.9	3.8	19.5
2. Střední	61	56.6	61.2
3. VOŠ	6.4	5.5	6.5
4. Bc	7.1	15.3	6.6
5. Mgr	6.6	15.7	5.4
6. Dr a vyšší	1	3	0.8



Obrázek 0.10 Sloupcový graf procentuálního rozdělení v kategoriích

Kolik ti je let?

Tabulka 0.61 Četnosti odpovědí podle kategorií

Věk	0-9	10-14	15-17	18-20	21-25	26-30	31-40	41-50	51-60	60<	ALL
all	1	29	286	581	961	949	1134	508	144	26	4619
men	1	11	131	277	445	458	624	304	77	14	2342
women	0	18	155	304	516	491	510	204	67	12	2277
light	1	17	143	245	265	237	206	85	22	4	1225
middle	0	9	110	264	545	534	633	259	77	14	2445
heavy	0	3	33	72	153	178	296	165	45	8	953
under18	1	29	286	0	0	0	0	0	0	0	316
bujon	1	21	105	348	537	381	368	127	27	8	1923
normal	0	8	181	231	417	551	756	372	115	17	2648

Tabulka 0.62 Procentuální rozložení odpovědí podle kategorií

Věk	0-9	10-14	15-17	18-20	21-25	26-30	31-40	41-50	51-60	60<
all	0	0.6	6.2	12.6	20.8	20.5	24.6	11	3.1	0.6
men	0	0.5	5.6	11.8	19	19.6	26.6	13	3.3	0.6
women	0	0.8	6.8	13.4	22.7	21.6	22.4	9	2.9	0.5
light	0.1	1.4	11.7	20	21.6	19.3	16.8	6.9	1.8	0.3
middle	0	0.4	4.5	10.8	22.3	21.8	25.9	10.6	3.1	0.6
heavy	0	0.3	3.5	7.6	16.1	18.7	31.1	17.3	4.7	0.8
under18	0.3	9.2	90.5	0	0	0	0	0	0	0
bujon	0.1	1.1	5.5	18.1	27.9	19.8	19.1	6.6	1.4	0.4
normal	0	0.3	6.8	8.7	15.7	20.8	28.5	14	4.3	0.6

Tabulka 0.63 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	bujon	normal
men		0.000					
women							
light				0.000	0.000		
middle					0.000		
heavy							
bujon							0.000
normal							

Tabulka 0.64 Četnosti odpovědí podle kategorií

Věk	ALL	FIN	UNFIN
0-9	1	0	1
10-14	29	0	28
15-17	286	0	284
18-20	581	8	561
21-25	961	40	879
26-30	949	51	844
31-40	1134	76	958
41-50	508	54	418
51-60	144	8	124
60<	26	1	25

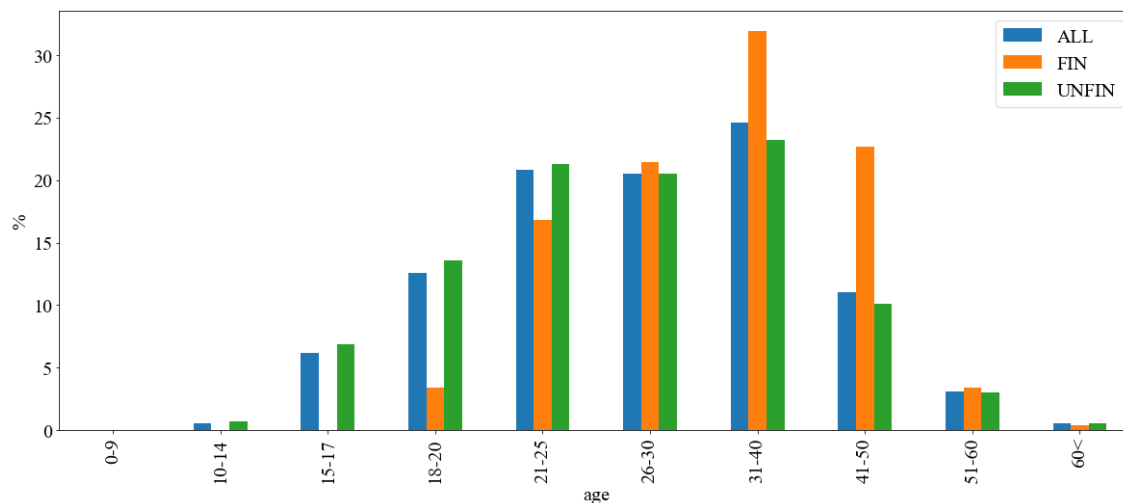
Tabulka 0.65 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.66 Procentuální rozložení odpovědí podle kategorií

Věk	ALL	FIN	UNFIN
0-9	0		0
10-14	0.6		0.7
15-17	6.2		6.9
18-20	12.6	3.4	13.6
21-25	20.8	16.8	21.3
26-30	20.5	21.4	20.5
31-40	24.6	31.9	23.2
41-50	11	22.7	10.1

Věk	ALL	FIN	UNFIN
51-60	3.1	3.4	3
60<	0.6	0.4	0.6



Obrázek 0.11 Sloupcový graf procentuálního rozdělení v kategoriích

Jaký je tvůj aktuální zdravotní stav?

Tabulka 0.67 Četnosti odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.	8.
all	3263	396	276	87	552	30	156	4619
men	1767	166	171	45	168	15	53	2338
women	1496	230	105	42	384	15	103	2281
light	883	87	48	17	152	12	65	1223
middle	1756	205	153	42	276	8	68	2446
heavy	628	104	75	28	124	10	23	954
under18	232	27	10	3	54	3	8	316
f18t20	384	71	30	4	97	10	32	577
f21t25	744	62	23	8	89	4	35	960
f26t30	712	78	34	14	96	3	36	954
f31t40	807	85	69	23	134	3	28	1138
f41t50	306	52	71	19	57	2	11	506
f51t60	72	17	28	12	18	3	4	141
over60	6	4	11	4	7	2	2	26
bujon	1480	136	76	22	191	11	156	1938
normal	1747	257	197	64	358	19	0	2634

Tabulka 0.68 Procentuální rozložení odpovědí podle kategorií

	1.	2.	3.	4.	5.	6.	7.
all	70.6	8.6	6	1.9	12	0.6	3.4
men	75.6	7.1	7.3	1.9	7.2	0.6	2.3
women	65.6	10.1	4.6	1.8	16.8	0.7	4.5
light	72.2	7.1	3.9	1.4	12.4	1	5.3
middle	71.8	8.4	6.3	1.7	11.3	0.3	2.8
heavy	65.8	10.9	7.9	2.9	13	1	2.4
under18	73.4	8.5	3.2	0.9	17.1	0.9	2.5
f18t20	66.6	12.3	5.2	0.7	16.8	1.7	5.5
f21t25	77.5	6.5	2.4	0.8	9.3	0.4	3.6
f26t30	74.6	8.2	3.6	1.5	10.1	0.3	3.8
f31t40	70.9	7.5	6.1	2	11.8	0.3	2.5
f41t50	60.5	10.3	14	3.8	11.3	0.4	2.2
f51t60	51.1	12.1	19.9	8.5	12.8	2.1	2.8
over60	23.1	15.4	42.3	15.4	26.9	7.7	7.7
bujon	76.4	7	3.9	1.1	9.9	0.6	8
normal	66.3	9.8	7.5	2.4	13.6	0.7	0

 Tabulka 0.69 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.000	0.000										
middle					0.000										
heavy															
under18						0.086	0.012	0.045	0.042	0.000	0.000	0.000	0.000		
f18t20							0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f21t25								0.351	0.000	0.000	0.000	0.000	0.000		
f26t30									0.042	0.000	0.000	0.000	0.000		
f31t40										0.000	0.000	0.000	0.000		
f41t50											0.038	0.000	0.000		
f51t60												0.050	0.000		
over60															
bujon															0.000
normal															

Tabulka 0.70 Četnosti odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Nic	3263	161	2926
2. Onemocnění dýchací soustav	396	22	350
3. Srdečně-cévní onemocnění	276	16	243

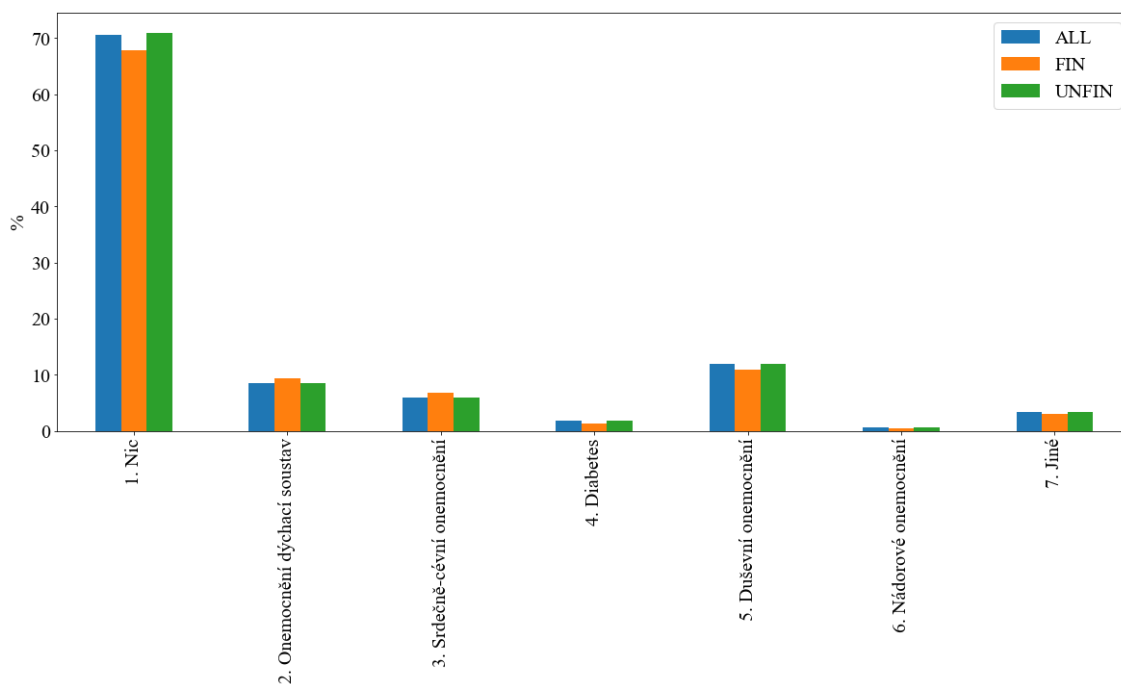
	ALL	FIN	UNFIN
4. Diabetes	87	3	79
5. Duševní onemocnění	552	26	496
6. Nádorové onemocnění	30	1	29
7. Jiné	156	7	140
8. Počet uživatelů	4619	237	4121

Tabulka 0.71 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.955
UNFIN		

Tabulka 0.72 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Nic	70.6	67.9	71
2. Onemocnění dýchací soustav	8.6	9.3	8.5
3. Srdečně-cévní onemocnění	6	6.8	5.9
4. Diabetes	1.9	1.3	1.9
5. Duševní onemocnění	12	11	12
6. Nádorové onemocnění	0.6	0.4	0.7
7. Jiné	3.4	3	3.4



Obrázek 0.12 Sloupcový graf procentuálního rozdělení v kategoriích

Prodělal jsi onemocnění covid-19?

Tabulka 0.73 Četnosti odpovědí podle kategorií

	1. Ne	2. Ano, již negativní	3. Ano, léčím se	ALL
all	4065	470	57	4592
men	2079	228	21	2328
women	1986	242	36	2264
light	1054	145	17	1216
middle	2154	246	30	2430
heavy	861	79	10	950
under18	279	35	2	316
f18t20	503	65	6	574
f21t25	821	122	13	956
f26t30	827	100	14	941
f31t40	1021	95	14	1130
f41t50	459	43	5	507
f51t60	130	9	2	141
over60	24	1	1	26
bujon	1657	240	24	1921
normal	2366	225	33	2624

Tabulka 0.74 Procentuální rozložení odpovědí podle kategorií

	1. Ne	2. Ano, již negativní	3. Ano, léčím se
all	88.5	10.2	1.2
men	89.3	9.8	0.9
women	87.7	10.7	1.6
light	86.7	11.9	1.4
middle	88.6	10.1	1.2
heavy	90.6	8.3	1.1
under18	88.3	11.1	0.6
f18t20	87.6	11.3	1
f21t25	85.9	12.8	1.4
f26t30	87.9	10.6	1.5
f31t40	90.4	8.4	1.2
f41t50	90.5	8.5	1
f51t60	92.2	6.4	1.4
over60	92.3	3.8	3.8
bujon	86.3	12.5	1.2
normal	90.2	8.6	1.3

Tabulka 0.75 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.061													
women															
light				0.226	0.017										
middle					0.245										
heavy															
under18							0.816	0.413	0.495	0.235	0.411	0.214	0.131		
f18t20								0.601	0.708	0.143	0.295	0.215	0.223		
f21t25									0.346	0.005	0.037	0.093	0.243		
f26t30										0.195	0.297	0.293	0.354		
f31t40											0.907	0.702	0.372		
f41t50												0.660	0.294		
f51t60													0.620		
over60															
bujon															0.000
normal															

Tabulka 0.76 Četnosti odpovědí podle kategorií

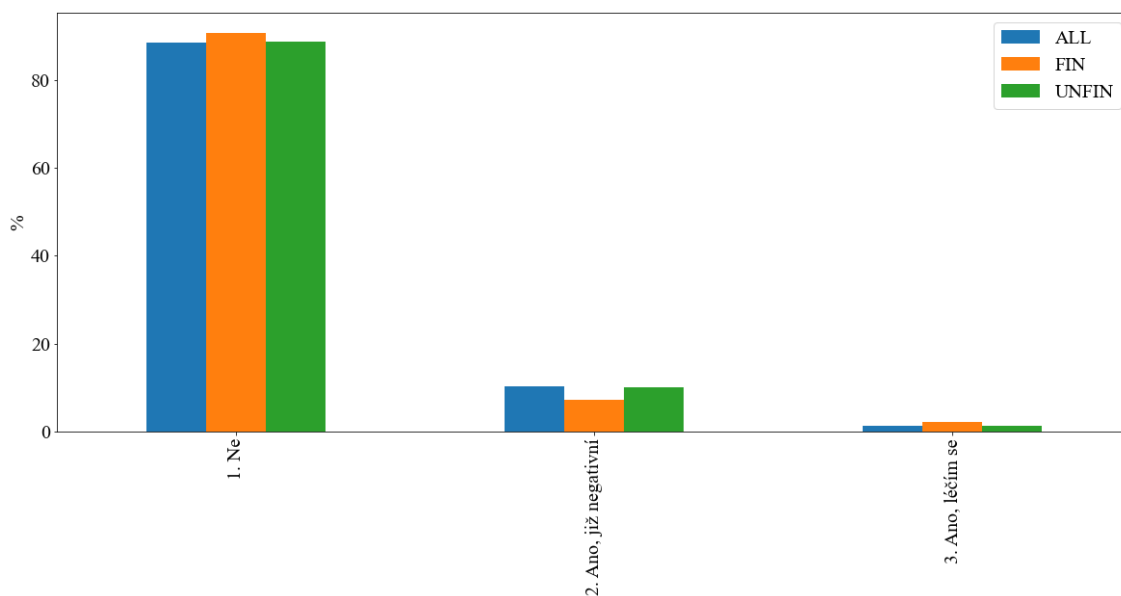
	ALL	FIN	UNFIN
1. Ne	4065	215	3630
2. Ano, již negativní	470	17	415
3. Ano, léčím se	57	5	51

Tabulka 0.77 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.185
UNFIN		

Tabulka 0.78 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Ne	88.5	90.7	88.6
2. Ano, již negativní	10.2	7.2	10.1
3. Ano, léčím se	1.2	2.1	1.2



Obrázek 0.13 Sloupcový graf procentuálního rozdělení v kategoriích

Užíváš pravidelně nějaké léky?

Tabulka 0.79 Četnosti odpovědí podle kategorií

	1. Ano	2. Ne	ALL
all	1275	3312	4587
men	466	1858	2324
women	809	1454	2263
light	346	868	1214
middle	638	1790	2428

	1. Ano	2. Ne	ALL
heavy	293	656	949
under18	82	234	316
f18t20	171	403	574
f21t25	211	745	956
f26t30	203	735	938
f31t40	312	816	1128
f41t50	197	310	507
f51t60	76	65	141
over60	22	4	26
bujon	497	1424	1921
normal	764	1855	2619

Tabulka 0.80 Procentuální rozložení odpovědí podle kategorií

	1. Ano	2. Ne
all	27.8	72.2
men	20.1	79.9
women	35.7	64.3
light	28.5	71.5
middle	26.3	73.7
heavy	30.9	69.1
under18	25.9	74.1
f18t20	29.8	70.2
f21t25	22.1	77.9
f26t30	21.6	78.4
f31t40	27.7	72.3
f41t50	38.9	61.1
f51t60	53.9	46.1
over60	84.6	15.4
bujon	25.9	74.1
normal	29.2	70.8

Tabulka 0.81 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.166	0.249										
middle					0.008										
heavy															
under18							0.255	0.179	0.133	0.595	0.000	0.000	0.000		
f18t20								0.001	0.000	0.387	0.002	0.000	0.000		
f21t25									0.865	0.004	0.000	0.000	0.000		
f26t30										0.002	0.000	0.000	0.000		

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
f31t40											0.000	0.000	0.000		
f41t50												0.002	0.000		
f51t60													0.007		
over60															
bujon															0.016
normal															

Tabulka 0.82 Četnosti odpovědí podle kategorií

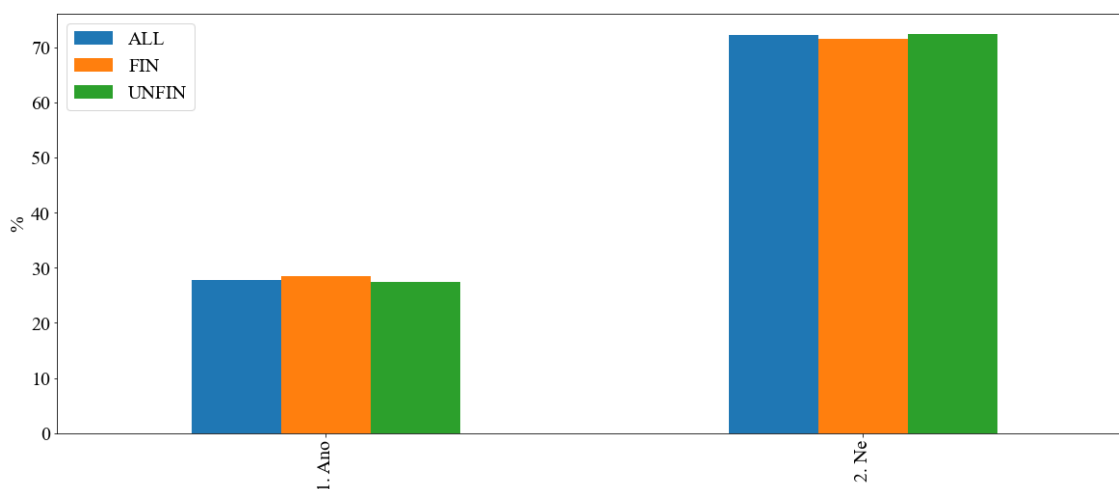
	ALL	FIN	UNFIN
1. Ano	1275	67	1125
2. Ne	3312	169	2967

Tabulka 0.83 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.822
UNFIN		

Tabulka 0.84 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. Ano	27.8	28.4	27.5
2. Ne	72.2	71.6	72.5



Obrázek 0.14 Sloupcový graf procentuálního rozdělení v kategoriích

Jak dlouho uživatel kouří

Tabulka 0.85 Četnosti odpovědí podle kategorií

	1. 0-2	2. 3-5	3. 6-10	4. 11-20	5. 20<	ALL
all	411	689	944	1540	869	4453
men	202	327	435	788	513	2265
women	209	362	509	752	356	2188
light	217	253	246	324	133	1173
middle	159	345	530	869	462	2365
heavy	35	91	170	347	275	918
under18	179	101	18	8	0	306
f18t20	150	325	65	21	0	561
f21t25	42	206	570	111	1	930
f26t30	12	32	236	619	6	905
f31t40	16	17	46	718	307	1104
f41t50	10	6	5	58	412	491
f51t60	2	2	4	4	122	134
over60	0	0	0	1	21	22
bujon	190	373	473	638	221	1895
normal	220	314	459	887	634	2514

Tabulka 0.86 Procentuální rozložení odpovědí podle kategorií

	1. 0-2	2. 3-5	3. 6-10	4. 11-20	5. 20<
all	9.2	15.5	21.2	34.6	19.5
men	8.9	14.4	19.2	34.8	22.6
women	9.6	16.5	23.3	34.4	16.3
light	18.5	21.6	21	27.6	11.3
middle	6.7	14.6	22.4	36.7	19.5
heavy	3.8	9.9	18.5	37.8	30
under18	58.5	33	5.9	2.6	0
f18t20	26.7	57.9	11.6	3.7	0
f21t25	4.5	22.2	61.3	11.9	0.1
f26t30	1.3	3.5	26.1	68.4	0.7
f31t40	1.4	1.5	4.2	65	27.8
f41t50	2	1.2	1	11.8	83.9
f51t60	1.5	1.5	3	3	91
over60	0	0	0	4.5	95.5
bujon	10	19.7	25	33.7	11.7
normal	8.8	12.5	18.3	35.3	25.2

Tabulka 0.87 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	men	women	light	middle	heavy	under18	f18t20	f21t25	f26t30	f31t40	f41t50	f51t60	over60	bujon	normal
men		0.000													
women															
light				0.000	0.000										
middle					0.000										
heavy															
under18							0.000	0.000	0.000	0.000	0.000	0.000	0.000		
f18t20								0.000	0.000	0.000	0.000	0.000	0.000		
f21t25									0.000	0.000	0.000	0.000	0.000		
f26t30										0.000	0.000	0.000	0.000		
f31t40											0.000	0.000	0.000		
f41t50												0.018	0.692		
f51t60													0.827		
over60															
bujon															0.000
normal															

Tabulka 0.88 Četnosti odpovědí podle kategorií

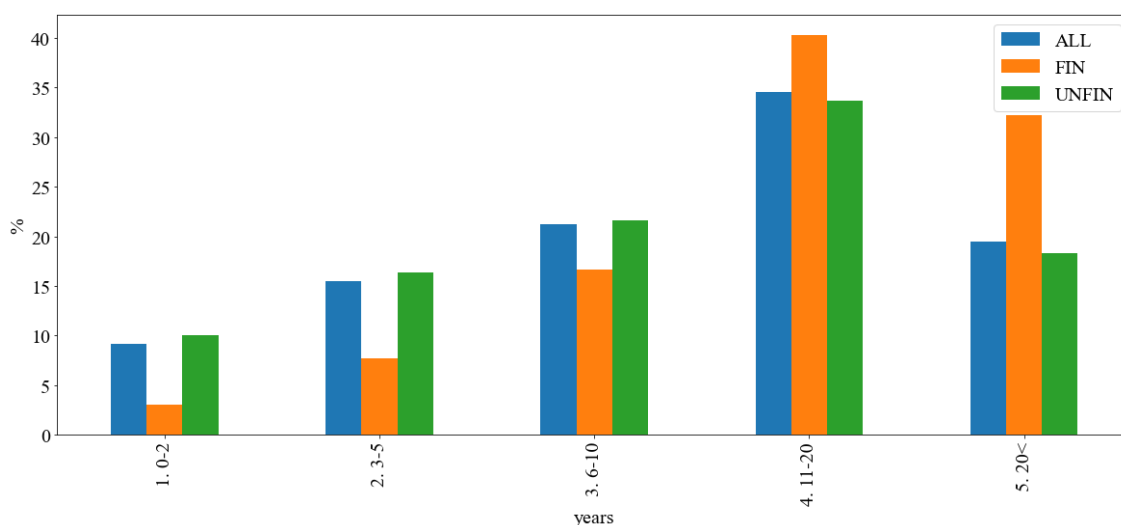
	ALL	FIN	UNFIN
1. 0-2	411	7	398
2. 3-5	689	18	649
3. 6-10	944	39	856
4. 11-20	1540	94	1335
5. 20<	869	75	725

Tabulka 0.89 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	FIN	UNFIN
FIN		0.000
UNFIN		

Tabulka 0.90 Procentuální rozložení odpovědí podle kategorií

	ALL	FIN	UNFIN
1. 0-2	9.2	3	10
2. 3-5	15.5	7.7	16.4
3. 6-10	21.2	16.7	21.6
4. 11-20	34.6	40.3	33.7
5. 20<	19.5	32.2	18.3



Obrázek 0.15 Sloupcový graf procentuálního rozdělení v kategoriích

Odpovědi terapie

EE 1 Ted' je rada na tobě. Kterým z těchto témat by ses chtěl zabývat během toho, co se budeš snažit přestat kouřit?

Tabulka 0.91 Četnosti odpovědí podle kategorií

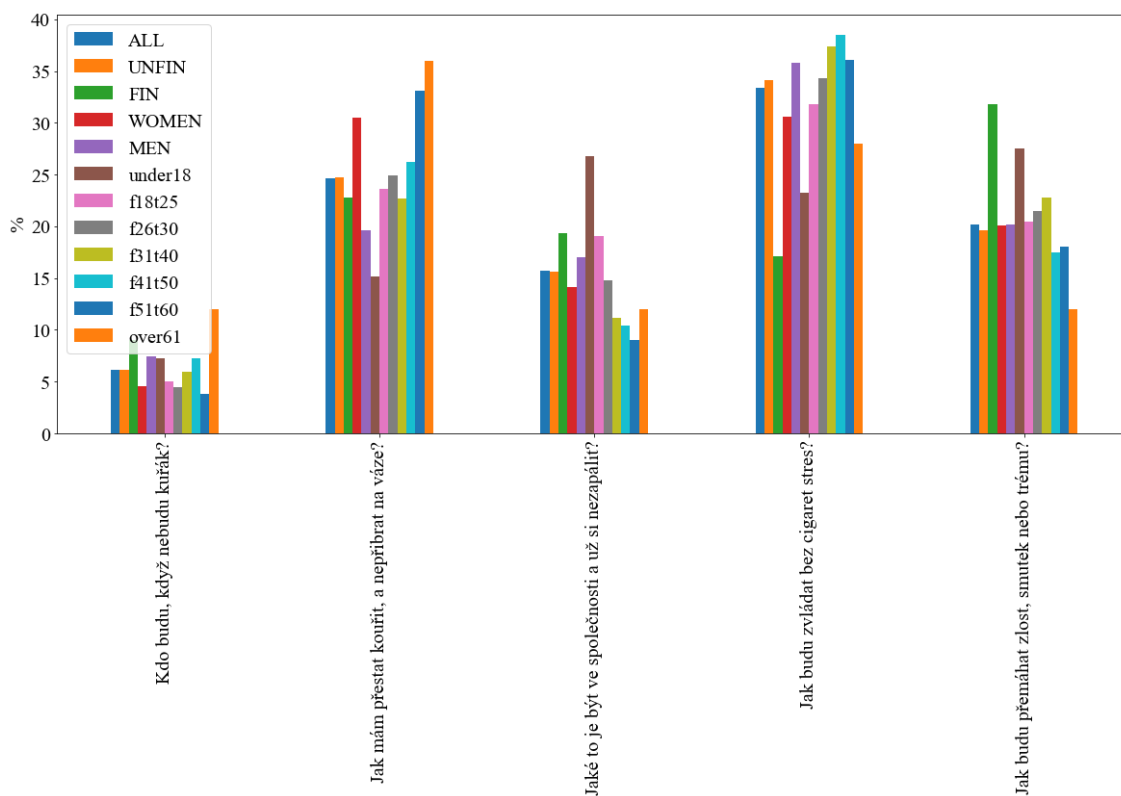
	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Kdo budu, když nebudu kuřák?	1080	1011	46	373	707	22	75	41	65	35	5	3
Jak mám přestat kouřit, a nepřibrat na váze?	4326	4117	116	2460	1865	46	353	226	247	126	44	9
Jaké to je být ve společnosti a už si nezapálit?	2759	2611	98	1135	1623	81	286	134	122	50	12	3
Jak budu zvládat bez cigaret stres?	5882	5688	87	2467	3411	70	475	311	408	185	48	7
Jak budu přemáhat zlost, smutek nebo trému?	3549	3274	162	1622	1924	83	306	195	248	84	24	3

Tabulka 0.92 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Kdo budu, když nebudu kuřák?	6.1	6.1	9	4.6	7.4	7.3	5	4.5	6	7.3	3.8	12
Jak mám přestat kouřit, a nepřibrat na váze?	24.6	24.7	22.8	30.5	19.6	15.2	23.6	24.9	22.7	26.2	33.1	36
Jaké to je být ve společnosti a už si nezapálit?	15.7	15.6	19.3	14.1	17	26.8	19.1	14.8	11.2	10.4	9	12
Jak budu zvládat bez cigaret stres?	33.4	34.1	17.1	30.6	35.8	23.2	31.8	34.3	37.4	38.5	36.1	28
Jak budu přemáhat zlost, smutek nebo trému?	20.2	19.6	31.8	20.1	20.2	27.5	20.5	21.5	22.8	17.5	18	12

Tabulka 0.93 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.000									
FIN											
WOMEN				0.000							
MEN											
under18					0.000	0.000	0.000	0.000	0.000	0.000	0.028
f18t25						0.085	0.000	0.000	0.012		0.235
f26t30							0.046	0.010	0.157		0.239
f31t40								0.118	0.086		0.267
f41t50									0.389		0.607
f51t60											0.417
over60											



Obrázek 0.16 Sloupcový graf procentuálního rozdělení v kategoriích

EE 2 Několik představ, které lidi uvádějí. Které z nich ti sedí?

Tabulka 0.94 Četnosti odpovědí podle kategorií

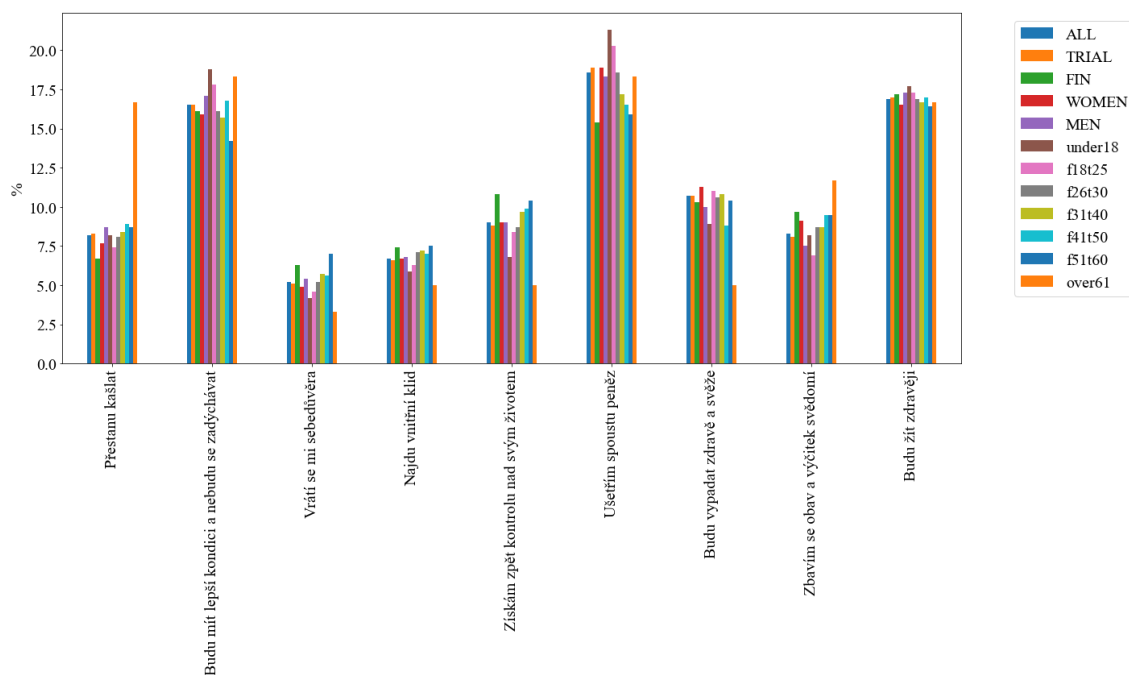
	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Přestanu kašlat	3401	3098	162	1565	1836	77	363	253	331	136	35	10
Budu mít lepší kondici a nebudu se zadýchávat	6849	6189	387	3230	3617	176	871	506	618	257	57	11
Vrátí se mi sebedůvěra	2156	1914	151	1006	1149	39	226	163	225	86	28	2
Najdu vnitřní klid	2794	2480	179	1361	1432	55	309	221	282	108	30	3
Získám zpět kontrolu nad svým životem	3726	3289	260	1825	1901	64	410	272	382	152	42	3
Ušetřím spoustu peněz	7725	7085	371	3847	3878	200	991	584	677	253	64	11
Budu vypadat zdravě a svěže	4436	4005	249	2304	2130	83	538	331	425	135	42	3
Zbavím se obav a výčitek svědomí	3450	3038	234	1860	1588	77	338	273	344	145	38	7
Budu žít zdravěji	7041	6356	413	3369	3671	166	847	531	660	260	66	10

Tabulka 0.95 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Přestanu kašlat	8.2	8.3	6.7	7.7	8.7	8.2	7.4	8.1	8.4	8.9	8.7	16.7
Budu mít lepší kondici a nebudu se zadýchávat	16.5	16.5	16.1	15.9	17.1	18.8	17.8	16.1	15.7	16.8	14.2	18.3
Vrátí se mi sebedůvěra	5.2	5.1	6.3	4.9	5.4	4.2	4.6	5.2	5.7	5.6	7	3.3
Najdu vnitřní klid	6.7	6.6	7.4	6.7	6.8	5.9	6.3	7.1	7.2	7	7.5	5
Získám zpět kontrolu nad svým životem	9	8.8	10.8	9	9	6.8	8.4	8.7	9.7	9.9	10.4	5
Ušetřím spoustu peněz	18.6	18.9	15.4	18.9	18.3	21.3	20.3	18.6	17.2	16.5	15.9	18.3
Budu vypadat zdravě a svěže	10.7	10.7	10.3	11.3	10	8.9	11	10.6	10.8	8.8	10.4	5
Zbavím se obav a výčitek svědomí	8.3	8.1	9.7	9.1	7.5	8.2	6.9	8.7	8.7	9.5	9.5	11.7
Budu žít zdravěji	16.9	17	17.2	16.5	17.3	17.7	17.3	16.9	16.7	17	16.4	16.7

Tabulka 0.96 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.000									
FIN											
WOMEN				0.000							
MEN											
under18					0.278	0.074	0.001	0.010	0.013		0.524
f18t25						0.018	0.000	0.000	0.024		0.154
f26t30							0.709	0.301	0.630		0.305
f31t40								0.603	0.965		0.278
f41t50									0.896		0.472
f51t60											0.330
over60											



Obrázek 0.17 Sloupcový graf procentuálního rozdělení v kategoriích

EE 2 Jaký je nejdůležitější důvod, který tě k tomu vede?

Tabulka 0.97 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	30.4	29.4	40.1	36.8	32.1	38.8	38.4	37.0	34.9	33.8	31.2	37.5
Počet odpovědí	7272	6469	449	3468	3797	193	813	486	656	278	75	16

Tabulka 0.98 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.24	0.24	0.26	0.15	0.12	0.13	0.16	0.15	0.15	0.15	0.10	0.11
Negativní průměrné skóre	0.08	0.08	0.09	0.05	0.05	0.07	0.05	0.04	0.05	0.05	0.05	0.02

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Neutralní												
průměrné skóre	0.67	0.68	0.65	0.80	0.83	0.80	0.79	0.80	0.80	0.80	0.85	0.86
Celkové												
průměrné skóre	0.14	0.13	0.17	0.11	0.07	0.09	0.12	0.11	0.11	0.10	0.01	0.08

Tabulka 0.99 Nejčastější nalezená slova

	Slova	Četnost
0	want	3289
1	health	1364
2	money	1119
3	healthy	1051
4	don	732
5	save	595
6	better	348
7	children	272
8	family	248
9	live	244
10	life	235
11	feel	230
12	healthier	207
13	lot	194
14	cigarettes	185
15	like	178
16	smoking	176
17	smoke	162
18	baby	161
19	son	149

Tabulka 0.100 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(want, healthy)	722
1	(don, want)	557

EE 3 Vyber triky, které by ti mohly pomoci

Podívej se na seznam. Když některý označíš, objeví se k němu krátké vysvětlení. Pokud tě trik zaujme, nechej ho zatržený. Když ne, zatržení zruš.

Tabulka 0.102 Četnosti odpovědí podle kategorií

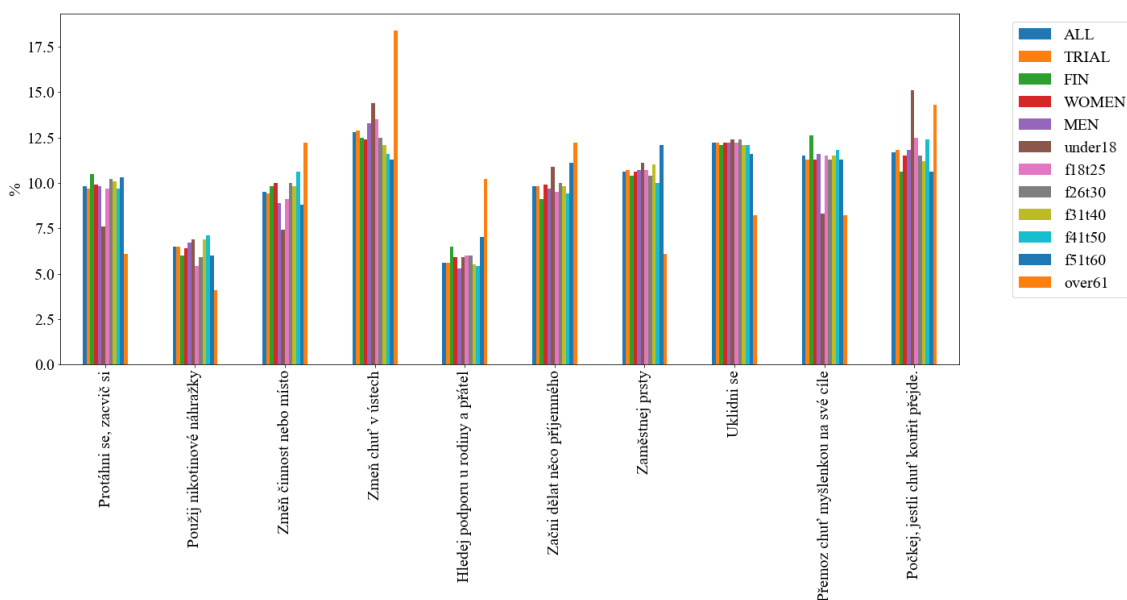
	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Protáhni se, zacvič si	3437	2919	310	1800	1636	53	414	306	371	146	41	3
Použij nikotinové náhražky	2276	1955	177	1165	1110	48	232	176	252	107	24	2
Změň činnost nebo místo	3300	2820	291	1816	1483	52	389	299	360	159	35	6
Zmeň chuť v ústech	4474	3889	369	2258	2216	101	577	373	443	174	45	9
Hledej podporu u rodiny a přátel	1957	1671	191	1071	886	41	259	179	202	82	28	5
Začni dělat něco příjemného	3421	2951	269	1810	1610	76	409	298	361	141	44	6
Zaměstnej prsty	3715	3215	307	1925	1790	78	458	310	404	151	48	3
Uklidni se	4264	3663	358	2234	2030	87	522	372	446	182	46	4
Přemoz chuť myšlenkou na své cíle	4001	3406	374	2062	1938	58	493	337	422	177	45	4
Počkej, jestli chuť kouřit přejde.	4073	3555	315	2103	1970	106	535	344	413	187	42	7

Tabulka 0.103 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Protáhni se, zacvič si	9.8	9.7	10.5	9.9	9.8	7.6	9.7	10.2	10.1	9.7	10.3	6.1
Použij nikotinové náhražky	6.5	6.5	6	6.4	6.7	6.9	5.4	5.9	6.9	7.1	6	4.1
Změň činnost nebo místo	9.5	9.4	9.8	10	8.9	7.4	9.1	10	9.8	10.6	8.8	12.2
Zmeň chuť v ústech	12.8	12.9	12.5	12.4	13.3	14.4	13.5	12.5	12.1	11.6	11.3	18.4
Hledej podporu u rodiny a přátel	5.6	5.6	6.5	5.9	5.3	5.9	6	6	5.5	5.4	7	10.2
Začni dělat něco příjemného	9.8	9.8	9.1	9.9	9.7	10.9	9.5	10	9.8	9.4	11.1	12.2
Zaměstnej prsty	10.6	10.7	10.4	10.6	10.7	11.1	10.7	10.4	11	10	12.1	6.1
Uklidni se	12.2	12.2	12.1	12.2	12.2	12.4	12.2	12.4	12.1	12.1	11.6	8.2
Přemoz chuť myšlenkou na své cíle	11.5	11.3	12.6	11.3	11.6	8.3	11.5	11.3	11.5	11.8	11.3	8.2
Počkej, jestli chuť kouřit přejde.	11.7	11.8	10.6	11.5	11.8	15.1	12.5	11.5	11.2	12.4	10.6	14.3

Tabulka 0.104 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN FIN		0.056									
WOMEN MEN				0.004							
under18					0.041	0.008	0.004	0.016	0.198	0.762	
f18t25						0.712	0.084	0.191	0.832	0.727	
f26t30							0.885	0.773	0.945	0.677	
f31t40								0.934	0.936	0.539	
f41t50									0.718	0.576	
f51t60										0.656	
over60											



Obrázek 0.19 Sloupcový graf procentuálního rozdělení v kategoriích

EE 3 Když dostanu chuť kouřit, nezapálím si, ale udělám tohle:

Tabulka 0.105 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	34.6	33.2	42.7	47.6	40.2	38.6	42.3	46.5	49.9	51.2	43.9	44.4
Počet odpovědí	12703	10582	1286	3235	3207	173	801	512	636	277	71	15

Tabulka 0.106 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.08	0.08	0.08	0.06	0.05	0.08	0.06	0.06	0.06	0.04	0.04	0.05
Negativní průměrné skóre	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Neutralní průměrné skóre	0.88	0.88	0.88	0.93	0.94	0.91	0.93	0.93	0.93	0.95	0.94	0.93
Celkové průměrné skóre	0.08	0.07	0.08	0.08	0.06	0.09	0.08	0.07	0.09	0.08	0.05	0.06

Tabulka 0.107 Nejčastější nalezená slova

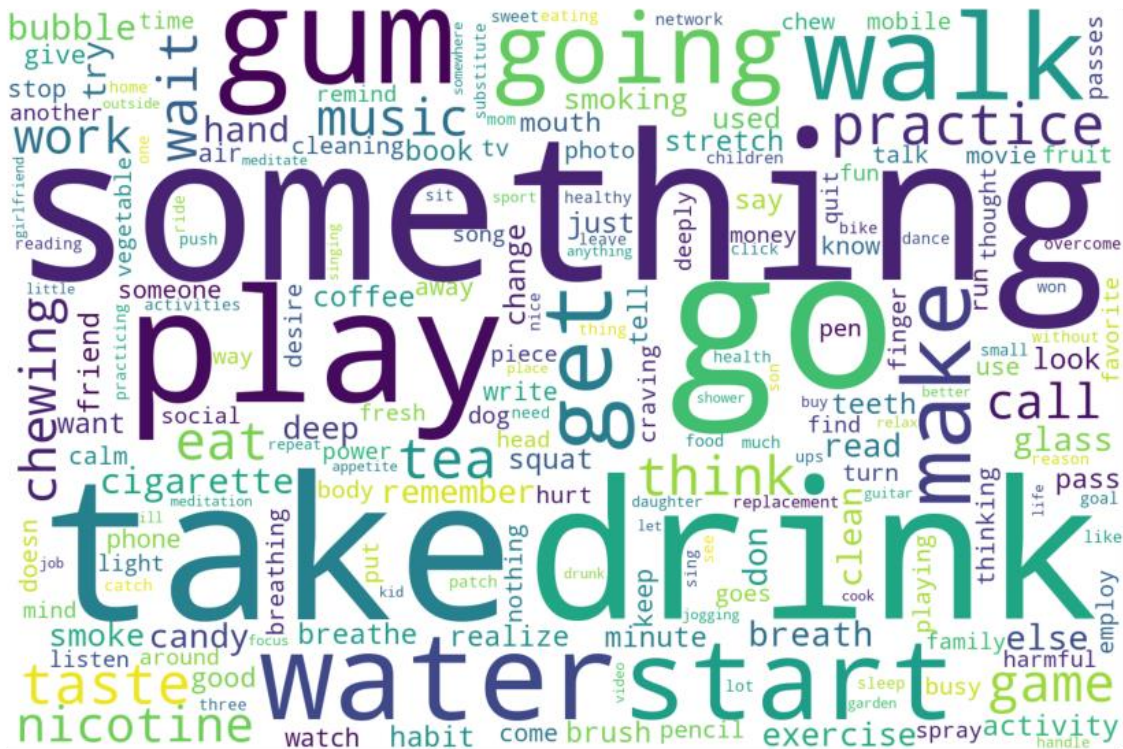
	Slova	Četnost
0	something	1342
1	take	1277
2	go	1222
3	drink	1188
4	play	1116
5	gum	950
6	water	762
7	walk	731
8	start	650
9	going	516
10	get	500
11	make	483
12	practice	458
13	taste	449
14	tea	448
15	wait	428
16	think	422
17	eat	404
18	music	396
19	chewing	372

Tabulka 0.108 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(chewing, gum)	343
1	(drink, water)	340
2	(bubble, gum)	330
3	(something, else)	293
4	(play, game)	293
5	(make, tea)	281
6	(glass, water)	281
7	(go, walk)	271
8	(play, music)	269
9	(get, used)	202

Tabulka 0.109 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(drink, glass, water)	186
1	(take, deep, breath)	180
2	(take, chewing, gum)	117



Obrázek 0.20 Vizualizace nejčastějších slov

EE 4 Jak se za své odvykání odměníš?

Tabulka 0.110 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	47.4	44.4	60.1	58.8	46.3	43.9	53.3	55.5	48.4	54.2	48.7	50.9
Počet odpovědí	2910	2273	392	1367	1274	71	315	198	289	127	35	8

Tabulka 0.111 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.20	0.19	0.21	0.13	0.11	0.10	0.13	0.14	0.12	0.14	0.11	0.09
Negativní průměrné skóre	0.03	0.04	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00
Neutralní průměrné skóre	0.76	0.76	0.76	0.86	0.87	0.89	0.86	0.85	0.87	0.85	0.88	0.91
Celkové průměrné skóre	0.27	0.25	0.33	0.25	0.18	0.17	0.22	0.24	0.19	0.23	0.17	0.19

Tabulka 0.112 Nejčastější nalezená slova

	Slova	Četnost
0	buy	956
1	money	899
2	something	699
3	save	386
4	put	298
5	cigarettes	294

	Slova	Četnost
6	nice	282
7	aside	246
8	reward	225
9	day	223
10	saved	214
11	every	206
12	good	203
13	don	180
14	go	177
15	new	166
16	week	164
17	know	158
18	days	98
19	month	93

Tabulka 0.113 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(buy, something)	524
1	(something, nice)	208
2	(save, money)	193
3	(money, saved)	128
4	(put, money)	127
5	(every, day)	114
6	(don, know)	113
7	(aside, money)	106
8	(money, cigarettes)	84
9	(money, buy)	80

Tabulka 0.114 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(buy, something, nice)	179
1	(put, aside, money)	59
2	(put, money, aside)	57



Obrázek 0.21 Vizualizace nejčastějších slov

EE 5 Tvůj názor

Zpětná vazba na aplikaci

Tabulka 0.115 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	50.4	41.4	79.1	60.5	45.3	25.3	52.5	56.4	56.6	60.6	63.5	96.0
Počet odpovědí	2971	2166	543	1442	1399	36	298	221	288	129	23	2

Tabulka 0.116 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.39	0.41	0.31	0.16	0.13	0.11	0.16	0.16	0.16	0.14	0.24	0.03

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Negativní												
průměrné skóre	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.00
Neutralní												
průměrné skóre	0.55	0.52	0.64	0.82	0.85	0.88	0.83	0.82	0.83	0.84	0.74	0.97
Celkové												
průměrné skóre	0.36	0.34	0.42	0.26	0.20	0.14	0.27	0.25	0.24	0.21	0.39	0.10

Tabulka 0.117 Nejčastější nalezená slova

	Slova	Četnost
0	like	406
1	good	355
2	agree	306
3	great	216
4	smoking	213
5	far	190
6	therapy	166
7	know	164
8	help	159
9	application	157
10	don	156
11	think	131
12	nice	124
13	really	118
14	support	116
15	yet	115
16	see	113
17	friend	104
18	someone	103
19	quit	102

Tabulka 0.118 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(don, know)	71
1	(quit, smoking)	71
2	(stop, smoking)	40
3	(far, good)	40
4	(feel, like)	38
5	(know, yet)	32
6	(really, like)	30
7	(agree, opinions)	29
8	(every, day)	27
9	(like, talking)	27

Tabulka 0.119 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(don, know, yet)	11
1	(real, collaboration, therapist)	10
2	(feel, like, alone)	10



Obrázek 0.22 Vizualizace nejčastějších slov

EE 5 Pokud zaváhám a zapálím si, chci slyšet toto:

Tabulka 0.120 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	60.5	53.6	79.8	73.8	55.0	88.2	87.6	83.7	78.6	72.3	78.5	172.0
Počet odpovědí	2826	2062	506	1422	1314	32	273	190	254	113	26	1

Tabulka 0.121 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.10	0.10	0.12	0.08	0.06	0.05	0.09	0.08	0.07	0.10	0.09	0.32
Negativní průměrné skóre	0.14	0.15	0.12	0.07	0.08	0.09	0.08	0.07	0.09	0.10	0.05	0.15
Neutralní průměrné skóre	0.75	0.74	0.76	0.85	0.87	0.85	0.82	0.85	0.84	0.80	0.87	0.52
Celkové průměrné skóre	-0.00	-0.01	0.03	0.04	-0.01	-0.03	0.05	0.02	-0.01	0.05	0.09	0.74

Tabulka 0.122 Nejčastější nalezená slova

	Slova	Četnost
0	don	720
1	want	370
2	give	269
3	smoke	263
4	smoking	218
5	stop	216

	Slova	Četnost
6	just	187
7	health	147
8	money	127
9	one	127
10	get	126
11	cigarette	124
12	go	120
13	cigarettes	116
14	think	114
15	strong	114
16	keep	110
17	better	108
18	remember	101
19	life	99

Tabulka 0.123 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(don, want)	107
1	(don, give)	104
2	(don, smoke)	88
3	(throw, away)	55
4	(want, smoke)	40
5	(doesn, matter)	39
6	(keep, going)	37
7	(non, smoker)	34
8	(stop, smoking)	31
9	(want, healthy)	30

Tabulka 0.124 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(don, want, smoke)	37
1	(non, smoker, smoke)	8
2	(come, long, way)	8

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Negativní												
průměrné skóre	0.14	0.14	0.13	0.09	0.08	0.04	0.10	0.09	0.08	0.10	0.04	0.15
Neutralní												
průměrné skóre	0.80	0.79	0.81	0.86	0.89	0.86	0.86	0.87	0.87	0.86	0.95	0.77
Celkové												
průměrné skóre	-0.06	-0.06	-0.05	-0.06	-0.04	0.13	-0.07	-0.06	-0.03	-0.06	-0.01	-0.40

Tabulka 0.127 Nejčastější nalezená slova

	Slova	Četnost
0	work	704
1	morning	630
2	coffee	605
3	bored	399
4	stress	279
5	meal	257
6	friends	240
7	alcohol	212
8	drink	212
9	stressed	202
10	home	187
11	break	180
12	waiting	179
13	eating	156
14	beer	150
15	boredom	148
16	nothing	137
17	something	134
18	want	133
19	get	126

EE 6 Zamysli se, co se změnilo. V čem jsou podmínky pro tento pokus jiné než minule?

Tabulka 0.130 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	42.0	37.9	48.5	51.7	41.3	29.7	48.4	61.7	49.4	50.8	90.9	148.0
Počet odpovědí	1710	1140	388	829	742	14	145	119	177	85	15	1

Tabulka 0.131 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.18	0.18	0.18	0.09	0.07	0.15	0.09	0.11	0.08	0.09	0.06	0.12
Negativní průměrné skóre	0.07	0.07	0.07	0.05	0.03	0.05	0.06	0.04	0.05	0.05	0.09	0.27
Neutralní průměrné skóre	0.74	0.74	0.74	0.86	0.89	0.79	0.85	0.86	0.88	0.86	0.85	0.62
Celkové průměrné skóre	0.09	0.08	0.11	0.05	0.05	0.09	0.05	0.10	0.06	0.07	-0.13	-0.49

Tabulka 0.132 Nejčastější nalezená slova

	Slova	Četnost
0	want	244
1	health	165
2	don	164
3	older	126
4	time	112

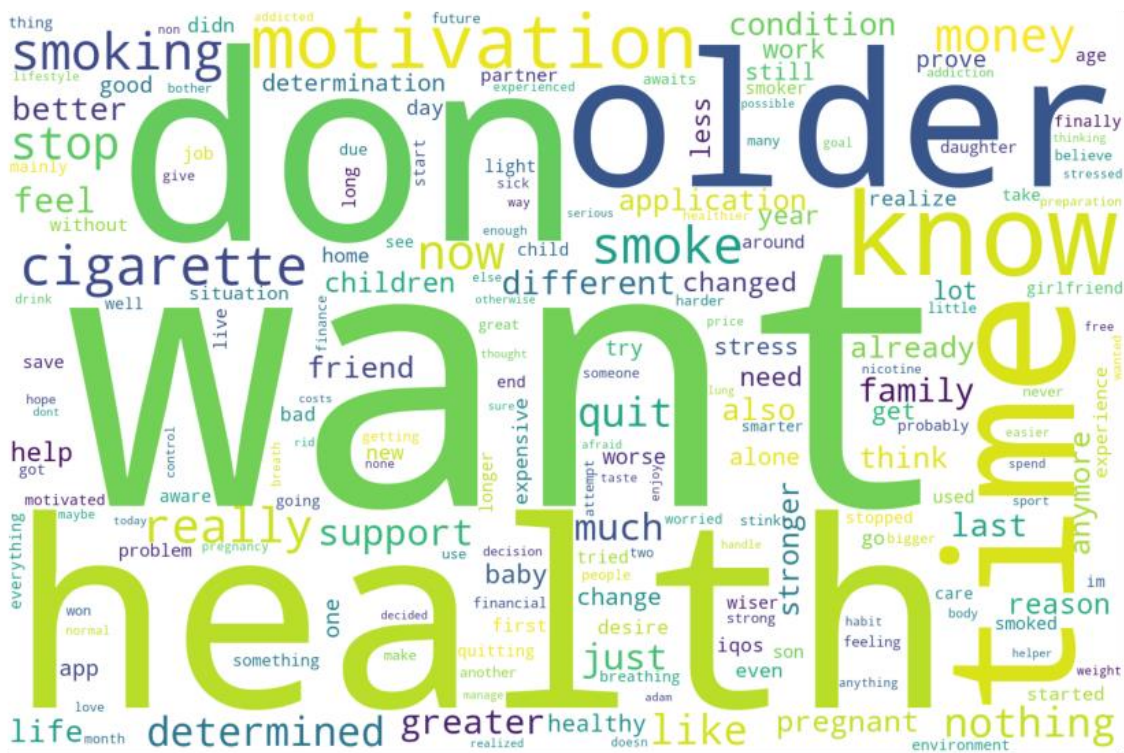
	Slova	Četnost
5	know	112
6	motivation	110
7	smoking	90
8	smoke	85
9	now	79
10	really	77
11	money	75
12	stop	71
13	cigarettes	67
14	like	55
15	nothing	55
16	quit	54
17	determined	54
18	much	47
19	just	45

Tabulka 0.133 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(don, want)	47
1	(really, want)	37
2	(don, know)	36
3	(want, stop)	27
4	(last, time)	25
5	(greater, motivation)	21
6	(already, know)	18
7	(want, quit)	16
8	(don, like)	15
9	(feel, like)	14

Tabulka 0.134 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(don, want, smoke)	11
1	(really, want, stop)	10
2	(greater, desire, quit)	6



Obrázek 0.25 Vizualizace nejčastějších slov

EE 7 Vyber prosím situaci, která se té skutečné nejvíc podobá:

Vyber si situaci, která nejlépe odpovídá té, během které jsi to nevydržel a začal jsi znovu kouřit.

Tabulka 0.135 Četnosti odpovědí podle kategorií

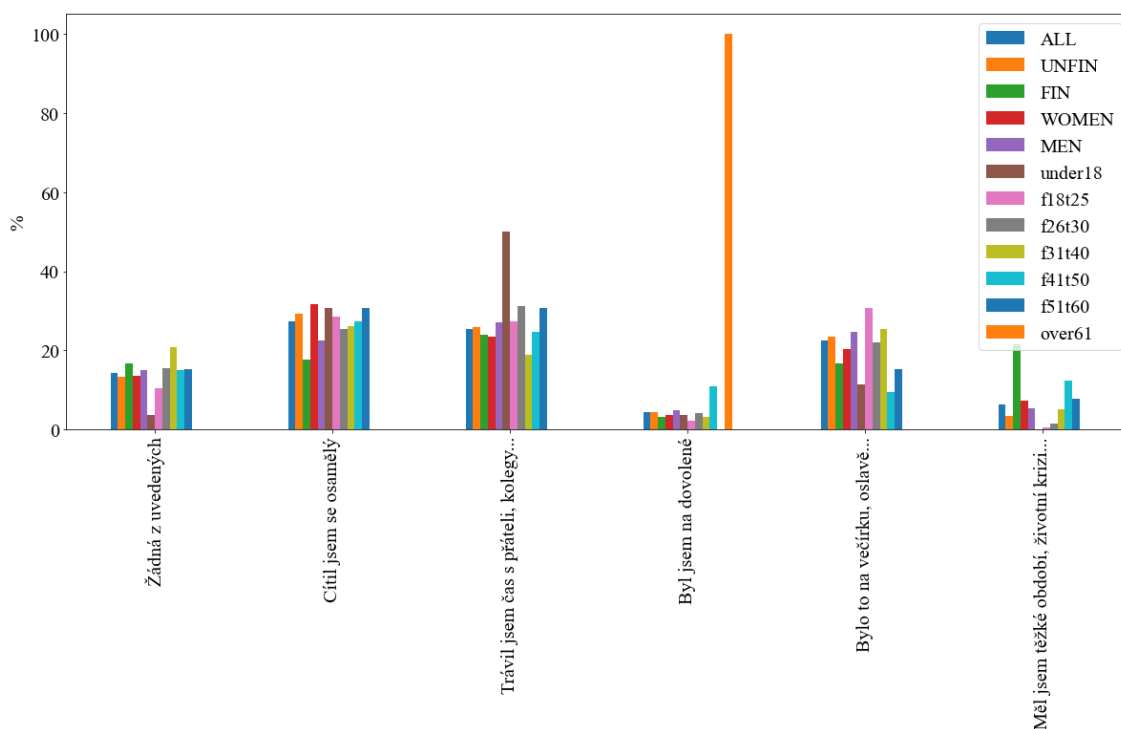
	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Žádná z uvedených	182	147	21	90	92	1	18	19	32	11	2	0
Cítil jsem se osamělý	349	322	22	210	139	8	49	31	40	20	4	0
Trávil jsem čas s přáteli, kolegy...	324	287	30	157	167	13	47	38	29	18	4	0
Byl jsem na dovolené	55	49	4	24	31	1	4	5	5	8	0	1
Bylo to na večírku, oslavě...	287	258	21	135	152	3	53	27	39	7	2	0
Měl jsem těžké období, životní krizi...	81	39	27	48	33	0	1	2	8	9	1	0

Tabulka 0.136 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Žádná z uvedených	14.2	13.3	16.8	13.6	15	3.8	10.5	15.6	20.9	15.1	15.4	0
Cítil jsem se osamělý	27.3	29.2	17.6	31.6	22.6	30.8	28.5	25.4	26.1	27.4	30.8	0
Trávil jsem čas s přáteli, kolegy...	25.4	26	24	23.6	27.2	50	27.3	31.1	19	24.7	30.8	0
Byl jsem na dovolené	4.3	4.4	3.2	3.6	5	3.8	2.3	4.1	3.3	11	0	100
Bylo to na večírku, oslavě...	22.5	23.4	16.8	20.3	24.8	11.5	30.8	22.1	25.5	9.6	15.4	0
Měl jsem těžké období, životní krizi...	6.3	3.5	21.6	7.2	5.4	0	0.6	1.6	5.2	12.3	7.7	0

Tabulka 0.137 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.000									
FIN											
WOMEN				0.004							
MEN											
under18					0.144	0.283	0.007	0.066	0.436	0.000	
f18t25						0.349	0.008	0.000	0.201	0.000	
f26t30							0.153	0.004	0.723	0.001	
f31t40								0.006	0.825	0.000	
f41t50									0.822	0.198	
f51t60										0.016	
over60											



Obrázek 0.26 Sloupcový graf procentuálního rozdělení v kategoriích

EE 7 Které situace považuješ za nejvíc rizikové?

Označ jednu nebo více situací, které mohou být nejnáročnější pro tebe. Neber si to ale příliš a nezaškrťavej je všechny! Vyber jen ty nejhroživější.

Tabulka 0.138 Četnosti odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Po jídle	1101	794	211	570	531	7	148	87	119	55	12	2
Když se rozčílím nebo se s někým pohádám	1535	1104	298	876	659	20	189	139	180	72	13	1
Při kávě	982	706	176	503	479	7	117	85	113	51	12	1
Když si dám pivo, víno nebo panáka	1394	1034	253	708	686	19	188	131	160	60	6	1
Při nákupu	100	76	13	48	52	4	13	4	12	3	1	0
Na dovolené	385	267	76	210	174	2	32	30	50	25	6	1
Když se budu veselit na večírku	1327	1012	213	679	648	23	201	123	133	45	5	1
Když jsem v kavárně (vyřazena)	320	223	71	161	159	0	6	1	11	6	0	0
V hospodě nebo v baru	1227	908	221	586	640	11	183	120	127	41	8	1
Když budu s přáteli, kteří kouří	1646	1249	279	875	771	32	226	154	175	74	9	1

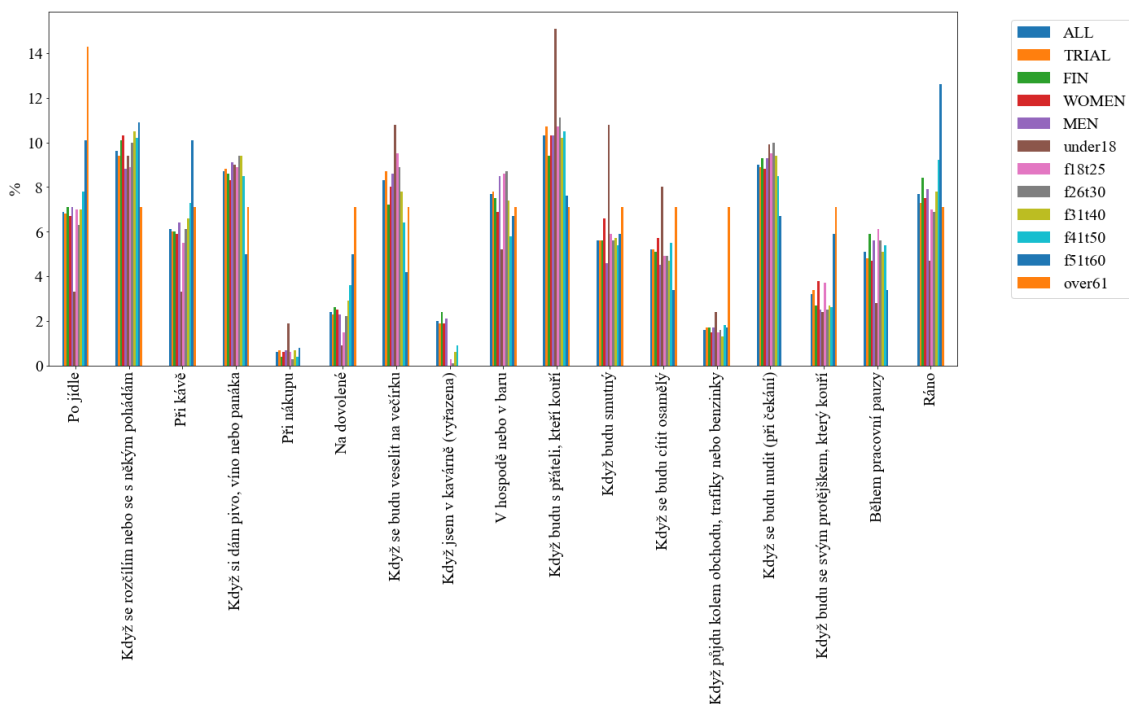
	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Když budu smutný	901	655	165	559	342	23	126	77	97	38	7	1
Když se budu cítit osamělý	828	608	150	487	341	17	104	68	81	39	4	1
Když půjdu kolem obchodu, trafiky nebo benzinky	260	197	50	130	130	5	32	22	23	13	2	1
Když se budu nudit (při čekání)	1445	1039	276	747	697	21	201	139	161	60	8	0
Když budu se svým protějškem, který kouří	513	399	79	327	186	5	78	34	46	18	7	1
Během pracovní pauzy	818	563	175	400	418	6	130	77	87	38	4	0
Ráno	1226	850	247	635	589	10	148	96	134	65	15	1

Tabulka 0.139 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Po jídle	6.9	6.8	7.1	6.7	7.1	3.3	7	6.3	7	7.8	10.1	14.3
Když se rozčilím nebo se s někým pohádám	9.6	9.4	10.1	10.3	8.8	9.4	8.9	10	10.5	10.2	10.9	7.1
Při kávě	6.1	6	6	5.9	6.4	3.3	5.5	6.1	6.6	7.3	10.1	7.1
Když si dám pivo, víno nebo panáka	8.7	8.8	8.6	8.3	9.1	9	8.9	9.4	9.4	8.5	5	7.1
Při nákupu	0.6	0.7	0.4	0.6	0.7	1.9	0.6	0.3	0.7	0.4	0.8	0
Na dovolené	2.4	2.3	2.6	2.5	2.3	0.9	1.5	2.2	2.9	3.6	5	7.1
Když se budu veselit na večírku	8.3	8.7	7.2	8	8.6	10.8	9.5	8.9	7.8	6.4	4.2	7.1
Když jsem v kavárně (vyřazena)	2	1.9	2.4	1.9	2.1	0	0.3	0.1	0.6	0.9	0	0
V hospodě nebo v baru	7.7	7.8	7.5	6.9	8.5	5.2	8.6	8.7	7.4	5.8	6.7	7.1
Když budu s přáteli, kteří kouří	10.3	10.7	9.4	10.3	10.3	15.1	10.7	11.1	10.2	10.5	7.6	7.1
Když budu smutný	5.6	5.6	5.6	6.6	4.6	10.8	5.9	5.6	5.7	5.4	5.9	7.1
Když se budu cítit osamělý	5.2	5.2	5.1	5.7	4.5	8	4.9	4.9	4.7	5.5	3.4	7.1
Když půjdu kolem obchodu, trafiky nebo benzinky	1.6	1.7	1.7	1.5	1.7	2.4	1.5	1.6	1.3	1.8	1.7	7.1
Když se budu nudit (při čekání)	9	8.9	9.3	8.8	9.3	9.9	9.5	10	9.4	8.5	6.7	0
Když budu se svým protějškem, který kouří	3.2	3.4	2.7	3.8	2.5	2.4	3.7	2.5	2.7	2.6	5.9	7.1
Během pracovní pauzy	5.1	4.8	5.9	4.7	5.6	2.8	6.1	5.6	5.1	5.4	3.4	0
Ráno	7.7	7.3	8.4	7.5	7.9	4.7	7	6.9	7.8	9.2	12.6	7.1

Tabulka 0.140 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.014									
FIN											
WOMEN				0.000							
MEN											
under18					0.003	0.001	0.001	0.000	0.000	0.000	0.000
f18t25						0.618	0.037	0.001	0.014	0.859	
f26t30							0.399	0.026	0.022	0.881	
f31t40								0.903	0.218	0.910	
f41t50									0.585	0.974	
f51t60										0.000	
over60											



Obrázek 0.27 Sloupcový graf procentuálního rozdělení v kategoriích

EE 7 Co můžu udělat, když se dostanu do rizikové situace:

Tabulka 0.141 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	38.0	33.4	49.8	46.3	38.9	32.9	48.2	48.3	47.2	43.3	43.3	30.5
Počet odpovědí	2934	2042	650	1293	1230	37	300	214	277	127	22	2

Tabulka 0.142 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.08	0.07	0.07	0.04	0.04	0.06	0.04	0.04	0.04	0.05	0.04	0.0
Negativní průměrné skóre	0.09	0.09	0.07	0.04	0.03	0.02	0.04	0.05	0.03	0.05	0.07	0.0
Neutralní průměrné skóre	0.81	0.80	0.84	0.92	0.93	0.92	0.91	0.91	0.92	0.90	0.90	1.0
Celkové průměrné skóre	0.03	0.02	0.04	0.03	0.02	0.04	0.02	0.03	0.03	0.05	-0.03	0.0

Tabulka 0.143 Nejčastější nalezená slova

	Slova	Četnost
0	don	287
1	something	267
2	go	236
3	leave	214
4	take	188
5	think	165
6	drink	163
7	change	154
8	avoid	153

	Slova	Četnost
9	get	137
10	smoke	134
11	try	129
12	teeth	124
13	else	124
14	cigarette	124
15	just	123
16	activity	108
17	away	105
18	gum	103
19	smoking	102

Tabulka 0.144 Nejčastější nalezená dvojsloví

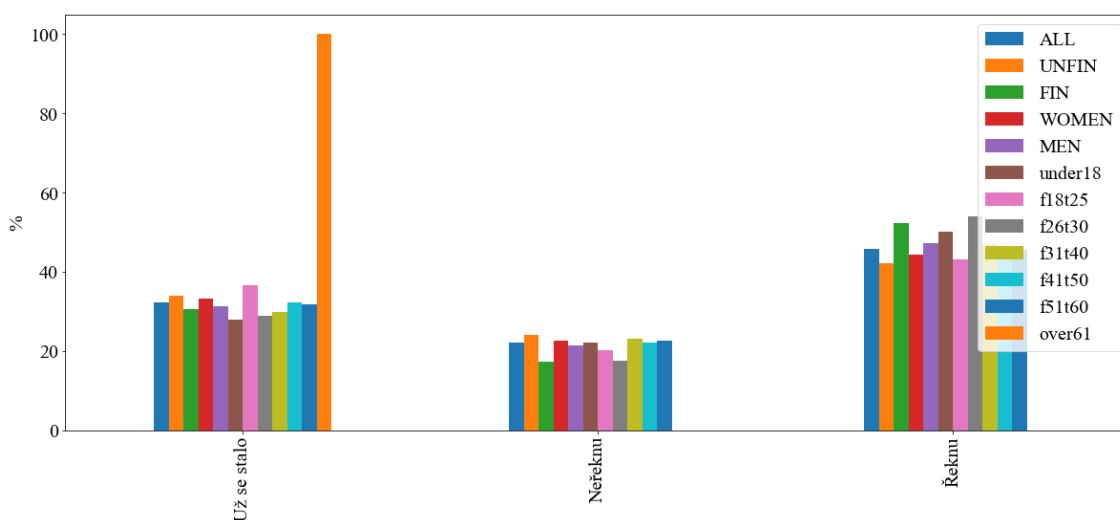
	Slova	Četnost
0	(something, else)	115
1	(grit, teeth)	77
2	(don, know)	67
3	(take, deep)	61
4	(deep, breath)	59
5	(change, activity)	57
6	(chewing, gum)	44
7	(drink, water)	41
8	(start, something)	39
9	(go, walk)	36

Tabulka 0.145 Nejčastější nalezená trojsloví

	Slova	Četnost
0	(take, deep, breath)	57
1	(think, something, else)	27
2	(start, something, else)	25

Tabulka 0.148 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.000									
FIN											
WOMEN				0.417							
MEN											
under18					0.574	0.777	0.941	0.863	0.934	0.102	
f18t25						0.052	0.218	0.684	0.895	0.181	
f26t30							0.200	0.313	0.721	0.089	
f31t40								0.887	0.982	0.100	
f41t50									0.997	0.131	
f51t60										0.162	
over60											



Obrázek 0.29 Sloupcový graf procentuálního rozdělení v kategoriích

EE 9 Označ 4–6 hodnot

Tabulka 0.149 Četnosti odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Zodpovědnost, spolehlivost	719	462	188	323	396	10	125	89	107	42	7	1
Rodina	1409	922	346	737	672	22	220	165	224	97	14	1
Svoboda jednání, samostatnost	716	447	208	357	359	15	113	79	104	61	6	0
Zdraví	1505	979	377	776	729	22	224	165	240	105	21	2

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pohoda, dobrá nálada	887	562	233	453	434	15	138	115	138	60	10	1
Vnitřní klid	916	568	250	496	420	16	133	109	139	67	9	1
Upřímnost k sobě	491	334	106	274	217	10	92	52	62	28	4	2
Láska	1116	748	268	650	466	22	210	136	153	65	10	0
Zvídavost, kreativita	291	165	93	148	143	6	50	35	34	16	4	0
Znalosti, vědomosti	442	261	124	198	244	4	75	50	73	27	7	0
Odvaha	384	267	82	187	197	10	77	34	43	21	4	0
Příroda	549	339	137	295	254	7	100	61	92	43	5	0
Umírněnost	82	57	17	33	49	3	14	6	9	1	0	0
Oblíbenost, popularita	181	121	41	77	104	2	34	23	23	8	1	0
Spravedlnost	351	205	102	189	162	7	60	46	56	20	5	0
Sexualita	568	381	131	234	334	9	114	71	74	26	6	0
Sebeúcta	549	343	150	314	235	10	83	56	70	48	11	0
Vzrušení, zábava	368	252	79	172	196	13	91	48	44	13	1	0
Užívat si života	721	493	155	397	324	14	141	75	105	43	11	1
Tolerance	318	193	95	163	155	9	56	34	48	20	5	0
Tradice	121	75	30	40	81	1	13	14	27	5	3	0
Vzhled, přitažlivost	596	414	124	338	258	10	116	80	82	35	6	0
Duchovní rozvoj	423	265	109	236	187	7	72	44	55	34	7	0
Poctivost	514	330	134	254	260	11	85	61	74	25	5	0
Přátelství	877	562	223	488	389	22	172	96	124	52	7	1
Pokora	370	223	103	183	186	7	54	41	66	27	6	0

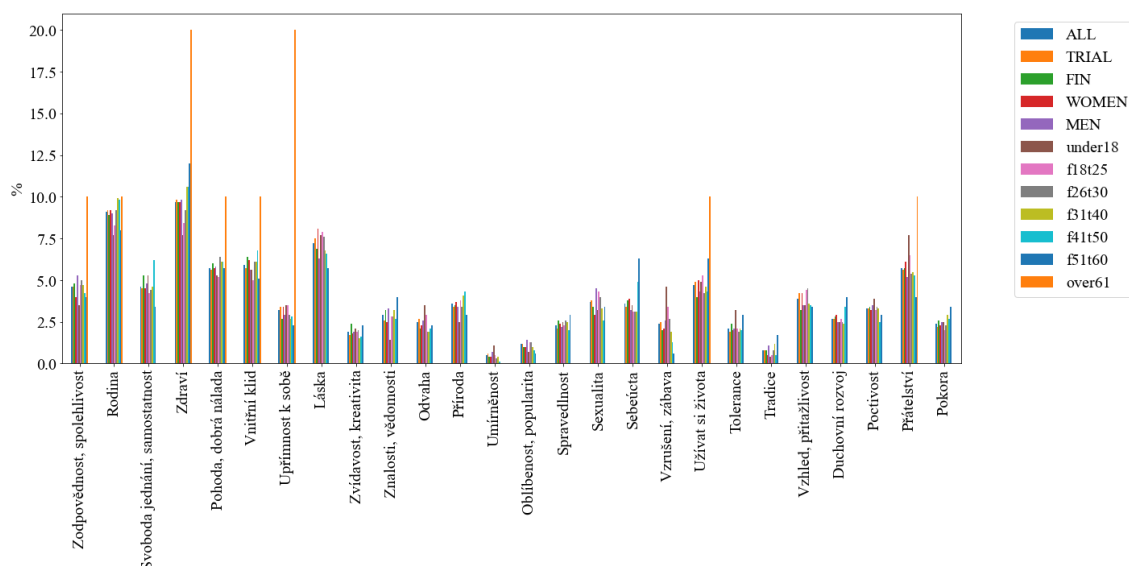
Tabulka 0.150 Procentuální rozdělení odpovědí podle kategorií

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Zodpovědnost, spolehlivost	4.6	4.6	4.8	4	5.3	3.5	4.7	5	4.7	4.2	4	10
Rodina	9.1	9.2	8.9	9.2	9	7.7	8.3	9.2	9.9	9.8	8	10
Svoboda jednání, samostatnost	4.6	4.5	5.3	4.5	4.8	5.3	4.2	4.4	4.6	6.2	3.4	0
Zdraví	9.7	9.8	9.7	9.7	9.8	7.7	8.4	9.2	10.6	10.6	12	20
Pohoda, dobrá nálada	5.7	5.6	6	5.7	5.8	5.3	5.2	6.4	6.1	6.1	5.7	10
Vnitřní klid	5.9	5.7	6.4	6.2	5.6	5.6	5	6.1	6.1	6.8	5.1	10
Upřímnost k sobě	3.2	3.4	2.7	3.4	2.9	3.5	3.5	2.9	2.7	2.8	2.3	20
Láska	7.2	7.5	6.9	8.1	6.3	7.7	7.9	7.6	6.8	6.6	5.7	0
Zvídavost, kreativita	1.9	1.7	2.4	1.8	1.9	2.1	1.9	2	1.5	1.6	2.3	0
Znalosti, vědomosti	2.9	2.6	3.2	2.5	3.3	1.4	2.8	2.8	3.2	2.7	4	0
Odvaha	2.5	2.7	2.1	2.3	2.6	3.5	2.9	1.9	1.9	2.1	2.3	0

	ALL	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Příroda	3.6	3.4	3.5	3.7	3.4	2.5	3.8	3.4	4.1	4.3	2.9	0
Umírněnost	0.5	0.6	0.4	0.4	0.7	1.1	0.5	0.3	0.4	0.1	0	0
Oblíbenost, popularita	1.2	1.2	1	1	1.4	0.7	1.3	1.3	1	0.8	0.6	0
Spravedlnost	2.3	2.1	2.6	2.4	2.2	2.5	2.3	2.6	2.5	2	2.9	0
Sexualita	3.7	3.8	3.4	2.9	4.5	3.2	4.3	4	3.3	2.6	3.4	0
Sebeúcta	3.6	3.4	3.8	3.9	3.2	3.5	3.1	3.1	3.1	4.9	6.3	0
Vzrušení, zábava	2.4	2.5	2	2.1	2.6	4.6	3.4	2.7	1.9	1.3	0.6	0
Užívat si života	4.7	4.9	4	5	4.3	4.9	5.3	4.2	4.6	4.3	6.3	10
Tolerance	2.1	1.9	2.4	2	2.1	3.2	2.1	1.9	2.1	2	2.9	0
Tradice	0.8	0.8	0.8	0.5	1.1	0.4	0.5	0.8	1.2	0.5	1.7	0
Vzhled, přitažlivost	3.9	4.2	3.2	4.2	3.5	3.5	4.4	4.5	3.6	3.5	3.4	0
Duchovní rozvoj	2.7	2.7	2.8	2.9	2.5	2.5	2.7	2.5	2.4	3.4	4	0
Poctivost	3.3	3.3	3.4	3.2	3.5	3.9	3.2	3.4	3.3	2.5	2.9	0
Přátelství	5.7	5.6	5.7	6.1	5.2	7.7	6.5	5.4	5.5	5.3	4	10
Pokora	2.4	2.2	2.6	2.3	2.5	2.5	2	2.3	2.9	2.7	3.4	0

Tabulka 0.151 Výsledné p-hodnoty pro χ^2 test nezávislosti mezi dvojicemi kategorií

	UNFIN	FIN	WOMEN	MEN	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
UNFIN		0.000									
FIN											
WOMEN				0.000							
MEN											
under18					0.983	0.624	0.225	0.075	0.519	0.915	
f18t25						0.522	0.000	0.000	0.358	0.888	
f26t30							0.879	0.122	0.695	0.835	
f31t40								0.532	0.918	0.824	
f41t50									0.964	0.849	
f51t60										0.000	
over60											



Obrázek 0.30 Sloupcový graf procentuálního rozdělení v kategoriích

EE 9 Zkus popsat, jak kouření ohrožuje tvoje hodnoty:

Tabulka 0.152 Průměrný počet znaků podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Průměrný počet znaků	60.3	46.6	81.1	73.0	54.2	51.9	66.3	77.1	73.9	65.1	89.4	111.0
Počet odpovědí	2521	1488	804	921	879	36	290	204	263	122	20	2

Tabulka 0.153 Průměrné sentiment skóre podle kategorií

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Pozitivní průměrné skóre	0.14	0.12	0.17	0.08	0.06	0.05	0.07	0.09	0.08	0.08	0.13	0.21
Negativní průměrné skóre	0.15	0.17	0.14	0.10	0.09	0.05	0.09	0.11	0.10	0.13	0.06	0.02

	ALL	UNFIN	FIN	women	men	under18	f18t25	f26t30	f31t40	f41t50	f51t60	over60
Neutralní												
průměrné skóre	0.68	0.67	0.69	0.82	0.85	0.90	0.83	0.80	0.82	0.79	0.81	0.77
Celkové												
průměrné skóre	0.04	-0.01	0.12	0.00	-0.02	-0.01	-0.01	0.01	0.00	-0.08	0.12	0.48

Tabulka 0.154 Nejčastější nalezená slova

	Slova	Četnost
0	health	674
1	smoking	478
2	family	379
3	want	345
4	don	284
5	smoke	221
6	money	180
7	freedom	178
8	cigarettes	154
9	love	143
10	life	138
11	peace	126
12	inner	125
13	self	122
14	good	114
15	healthy	108
16	like	101
17	threatens	96
18	time	95
19	cigarette	90

Tabulka 0.155 Nejčastější nalezená dvojsloví

	Slova	Četnost
0	(inner, peace)	113
1	(don, want)	96

