

UPA projekt: Zdravotnictví v ČR, část 2.

Cílem druhé části tohoto projektu je získat nové znalosti zkombinováním informací z datových sad o obyvatelstvu ČR a poskytovatelů zdravotních služeb. Konkrétně počty poskytovatelů na kraj, historii poskytovatelů, poměry poskytovatelů různých oborů na počet obyvatel, a podobně.

Dalším cílem je připravit datové sady na dolování dat pro klasifikátor, který na základě počtu poskytovatelů vybraných zdravotních služeb a věkového složení obyvatel odhadne velikost okresu.

Dotazy

Dle zadání máme provést následující analýzy:

- vytvořte sloupcové grafy zobrazující počty poskytovatelů určitého oboru pro Brno a zbytek Jihomoravského kraje (15 poskytovatelů)
- vytvořte čárový (spojnicový) graf zobrazující historii počtu poskytovatelů v 5 zvolených oborech (čtvrtletní hodnoty)
- sestavte žebříček krajů dle počtu obyvatel na jednoho praktického lékaře

Dále jsme si zvolili následující dotazy:

- vytvořte grafy zobrazující počet gynekologů na ženu v určitém kraji
- vytvořte grafy zobrazující počet lékáren v krajích

Dalším úkolem je připravit trénovací sadu pro klasifikátor, který na základě počtu poskytovatelů vybraných zdravotních služeb a věkového složení obyvatel odhadne velikost okresu (jako malý/střední/velký).

Implementace a výsledky

Data jsme vhodně zpracovali, a poté výsledky vizualizovali pomocí grafů.

U několika dotazů jsme narazili na problém, že Praha oficiálně nepatří do žádného okresu.

Při analýze je ale vhodné počítat se všemi údaji, proto je v tomto výzkumu Praha brána jako 77. okres.

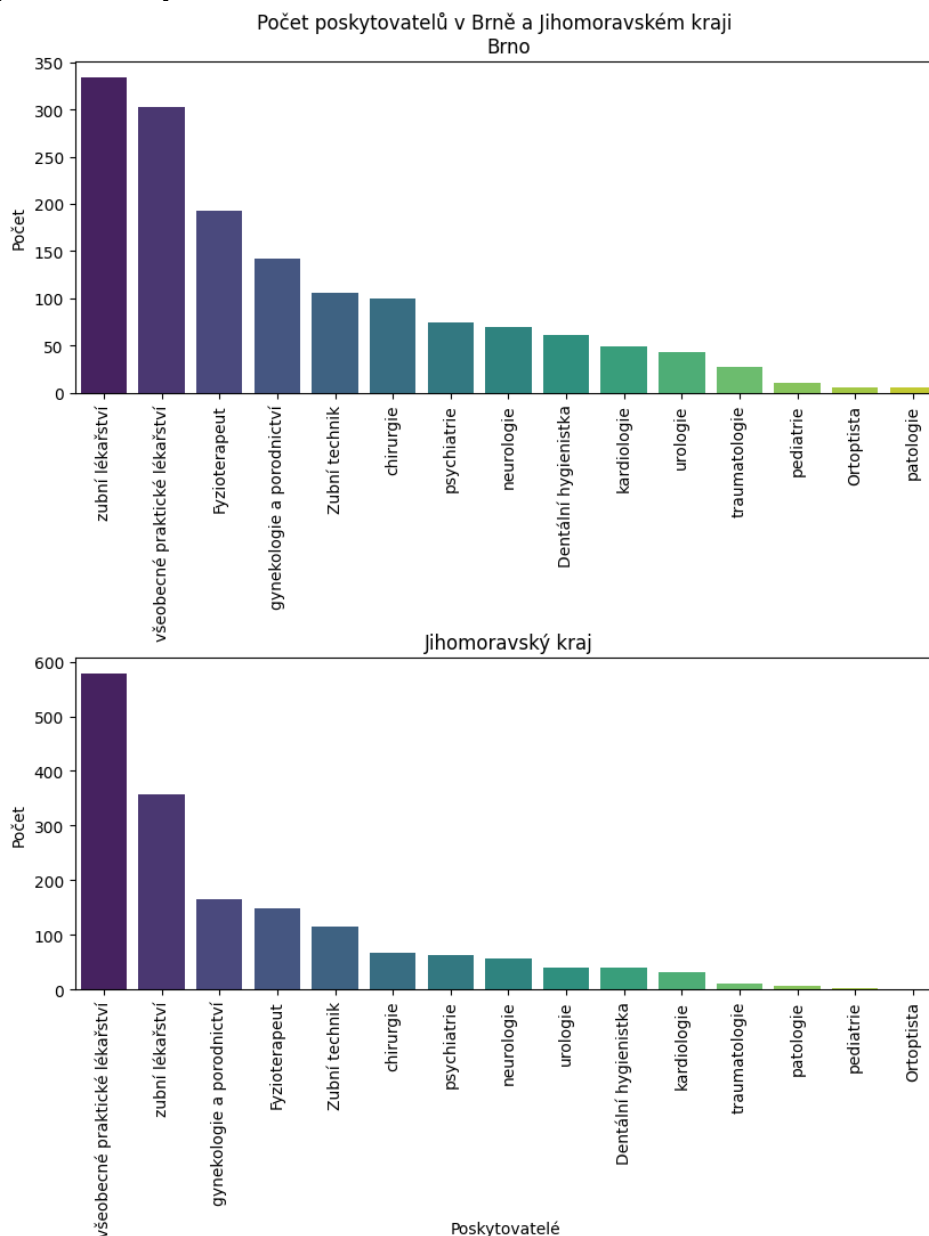
Počet poskytovatelů v oborech v Brně a Jihomoravském kraji

Z uložených informací o poskytovatelích v jednotlivých krajích jsme vyfiltrovali ty z Jihomoravského kraje.

Jako Brno byl považován jen okres Brno-město, všechny ostatní okresy v Jihomoravském kraji jsou započítány do údajů pro Jihomoravský kraj.

Byly zkoumány obory všeobecné praktické lékařství, zubní lékařství, dentální hygienistka, zubní technik, chirurgie, psychiatrie, neurologie, urologie, pediatrie, kardiologie, psychiatrie, traumatologie, oroptistika, gynekologie a porodnictví.

Výsledky jsou následující:



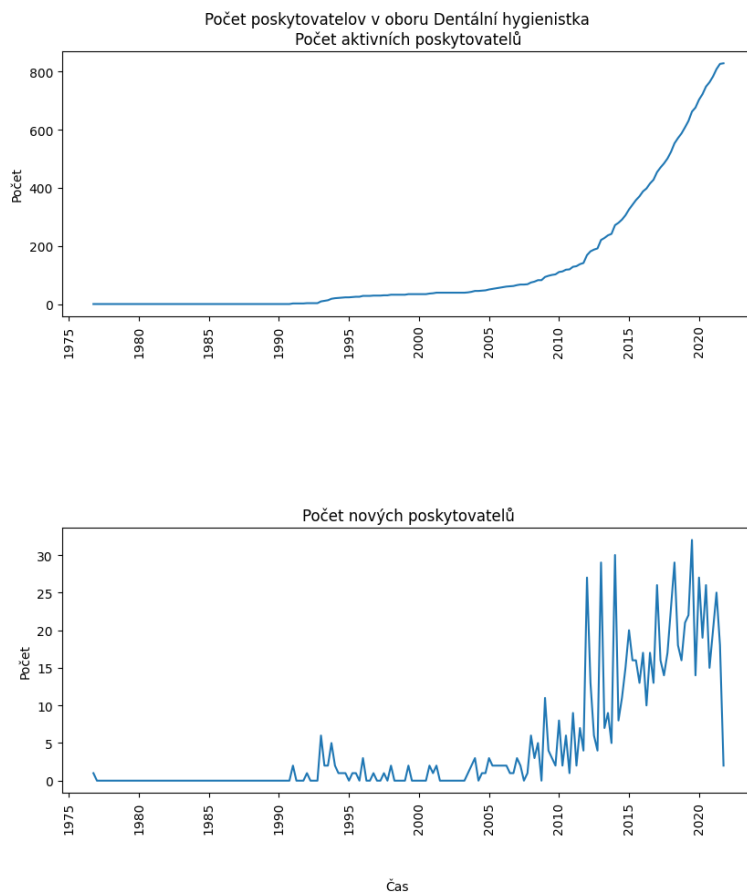
Historie počtu poskytovatelů

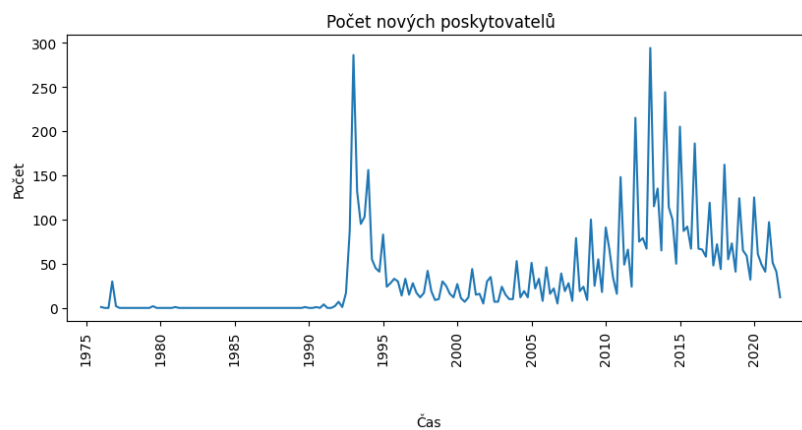
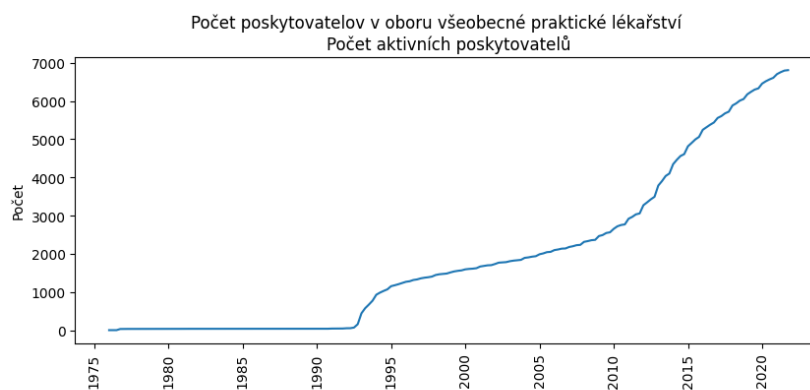
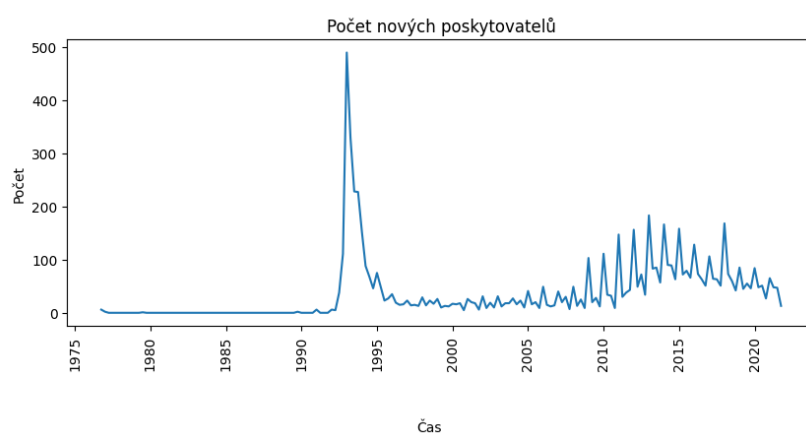
Byla zkoumána historie počtu poskytovatelů v následujících oborech: dentální hygiena, zubní lékařství, pediatrie, praktické lékařství a psychiatrie.

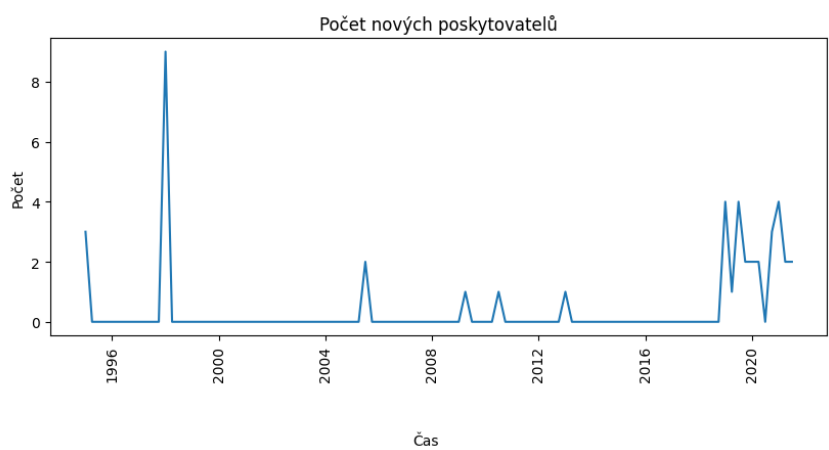
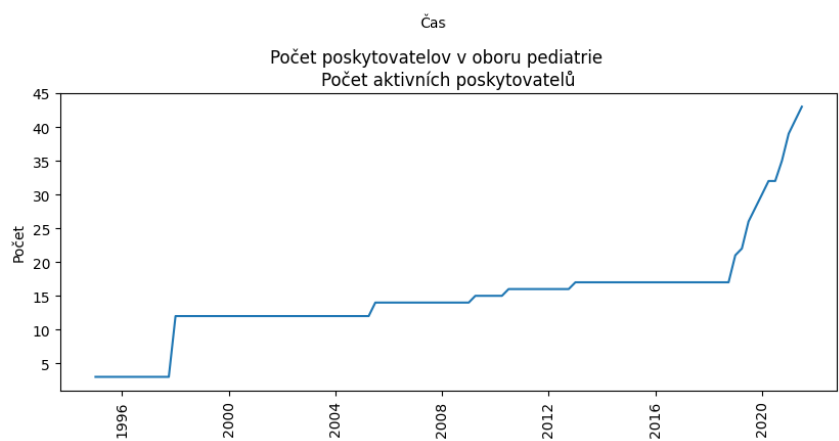
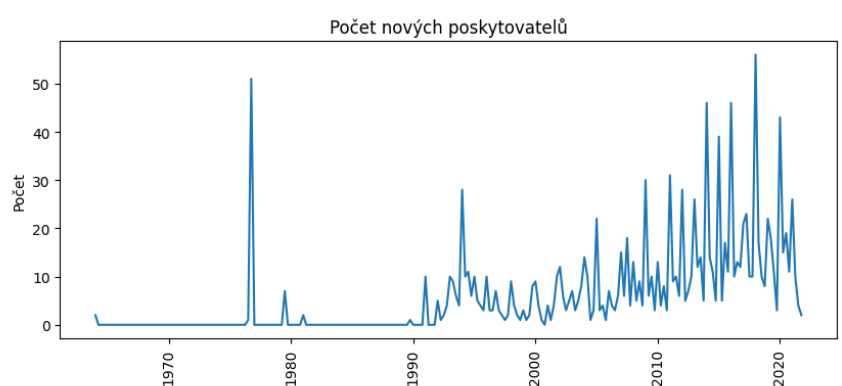
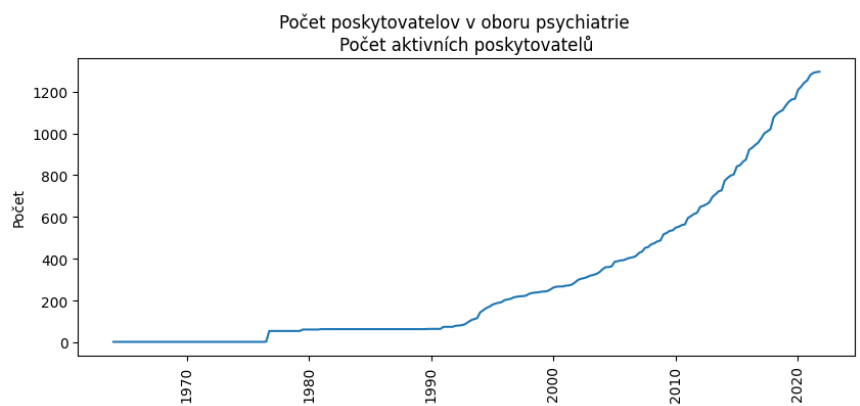
Veškeré údaje jsou počítány pro tříměsíční období (čtvrtletí).

První graf vždy zobrazuje počet aktivních poskytovatelů v oboru, druhý graf zobrazuje přírůstek poskytovatelů v oboru, a to pomocí data zahájení činnosti.

Výsledky byly následující:



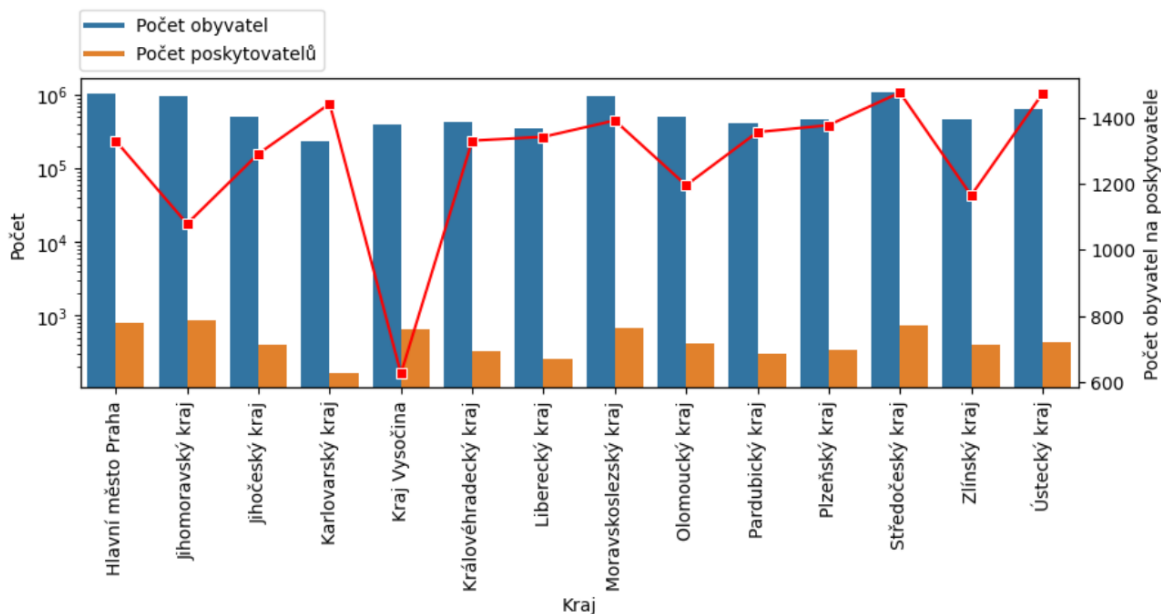




Žebříček krajů dle počtu obyvatel na jednoho praktického lékaře

Dle výsledků mají nejlepší poměr počtu obyvatel a praktických lékařů kraje Vysočina, Jihomoravský, Zlínský a Olomoucký. Nejhůře jsou na tom kraje Karlovarský, Středočeský a Ústecký.

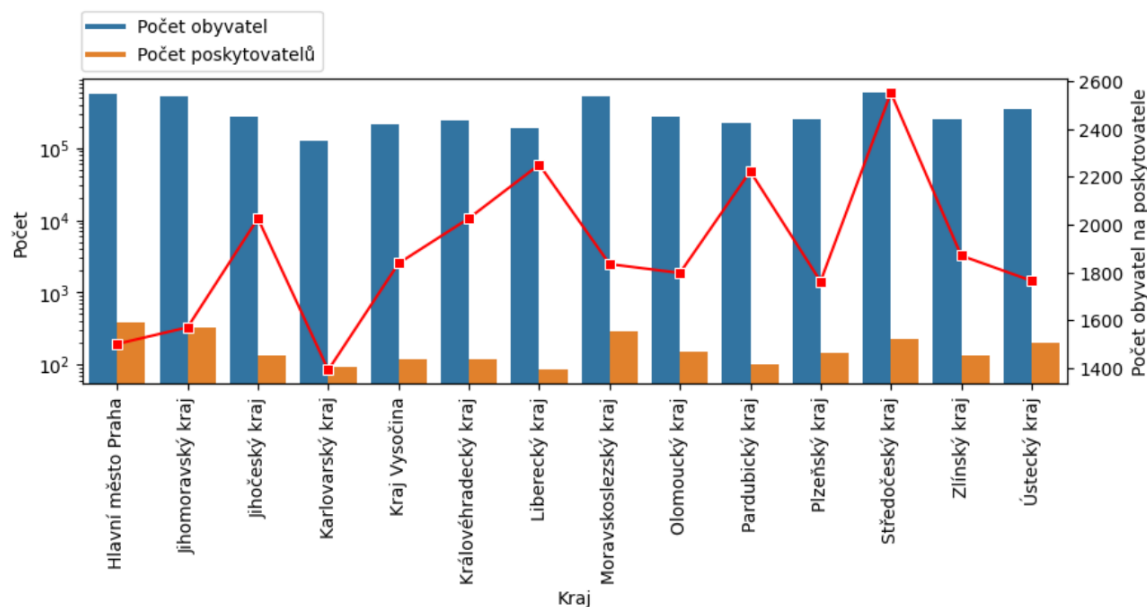
Graf zobrazující všechny kraje je následující (je v logaritmickém měřítku):



Žebříček krajů dle počtu žen na jednoho gynekologa

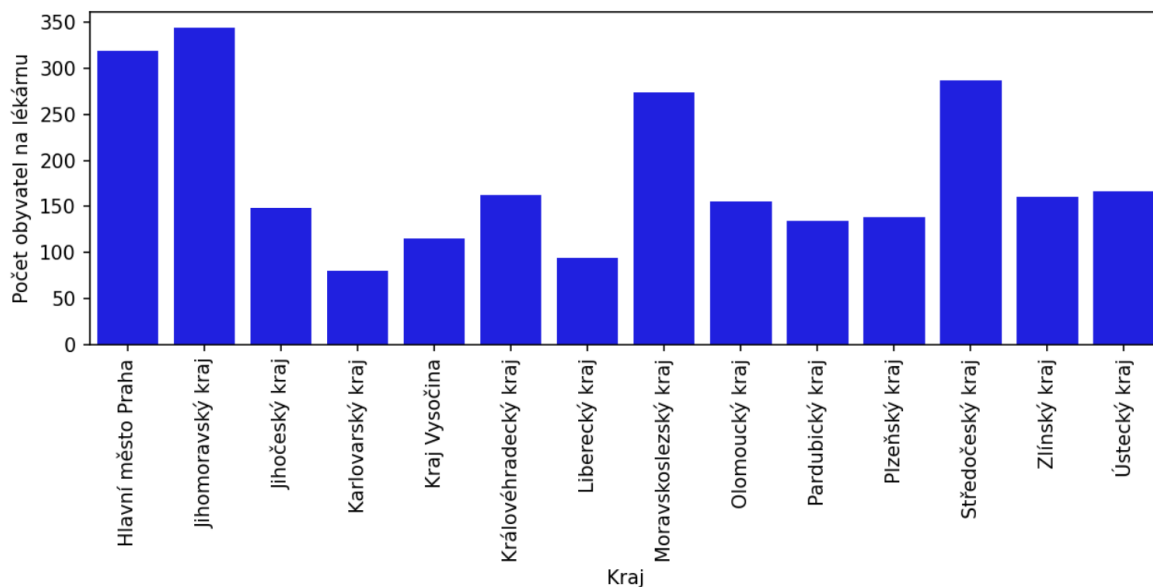
Jako zajímavá statistika se jeví i poměr žen na gynekologa v jednotlivých krajích.

Zde jsme uvažili, že gynekologa v naprosté většině případů navštěvují pouze ženy a dívky starší patnácti let.



Žebříček krajů dle počtu obyvatel na počet lékáren

Další zajímavou statistikou je poměr lékáren na počet obyvatel v jednotlivých krajích. Zde jsou na tom nejlépe Jihomoravský kraj, hlavní město Praha, Středočeský kraj a Moravskoslezský kraj. Nejhorší jsou na tom Karlovarský kraj, Liberecký kraj a kraj Vysočina.



Příprava dat pro dolovací algoritmus

Úkolem je připravit trénovací sadu pro klasifikátor, který na základě počtu poskytovatelů vybraných zdravotních služeb a věkového složení obyvatel odhadne velikost okresu (jako malý/střední/velký).¹

Okresy byly rozděleny do tří kategorií pomocí diskretizace:

- malé: 0 – 100 tisíc obyvatel
- střední: 100 – 125 tisíc obyvatel
- velké: více jak 125 tisíc obyvatel

Na rozdělení okresů byla použita technika diskretizace. Současně jsme normalizovali počet obyvatelů v okrese pomocí min-max.

Odlehle hodnoty jsme detekovali pomocí z-score. Detekovali jsme celkem 12 hodnot, například v okrese Brno-město byl velký počet všeobecné praktické lékařství nebo v Českých Budějovicích byl velký počet zubařů. Odlehle hodnoty jsme nahradili mediány.

¹ Dle zadání máme připravit data tak, aby klasifikátor na základě daných dat odhadl velikost města. Data, která máme k dispozici, tyto informace bohužel neposkytují, proto byl tento požadavek upraven na odhad velikosti okresu, a požadované hranice obyvatel byly upraveny na základě počtu obyvatel v okresech.

Popis CSV souborů

DotazA.csv

Tento soubor obsahuje data pro porovnání počtu poskytovatelů zdravotních služeb v Brně a Jihomoravském kraji a pro analýzu historie poskytovatelů vybraných služeb.

Soubor obsahuje následující sloupce (jména byla ponechána z původních CSV souborů):

- id – ID záznamu
- NazevCely – jméno poskytovatele
- DruhZarizeniKod – kód druhu zařízení
- Druhzarizeni – druh zařízení (např. samostatná ordinace všeobecného praktika)
- Kraj – kraj, ve kterém je služba poskytována (sídlo poskytovatele)
- Okres – okres, ve kterém je sídlo poskytovatele
- OkresKod – kód okresu, ve kterém je sídlo poskytovatele
- DatumZahajeniCinnosti
- OborPece – např. zubní lékař
- _id.oid – ID objektu v databázi

DotazB.csv

Tento soubor obsahuje data pro analýzu poměru počtu obyvatel kraje na praktického lékaře.

Soubor obsahuje následující sloupce:

- Id – ID záznamu
- Kraj – název kraje
- Sum_Inhabitants – počet obyvatel v daném kraji
- Sum_Providers – počet poskytovatelů zdravotnických služeb v oboru praktický lékař v kraji
- Rate – počet obyvatel na praktického lékaře

DotazCustomLekarny.csv

Tento soubor obsahuje data pro analýzu poměru počtu obyvatel na počet lékáren v kraji.

Soubor obsahuje následující sloupce:

- Id – ID záznamu
- Kraj – název kraje
- Sum – počet lékáren v kraji

DotazCustomGynekologie.csv

Tento soubor obsahuje data pro analýzu poměru počtu obyvatel na počet gynekologů v kraji.

Soubor obsahuje následující sloupce:

- Id – ID záznamu
- Kraj – název kraje
- Sum_Inhabitants – počet obyvatel v daném kraji
- Sum_Providers – počet poskytovatelů v oboru gynekologie v kraji
- Rate – počet obyvatel na gynekologa

DotazCPuvodni.csv

Tento soubor obsahuje data, která budou použita pro dolování, ještě před jejich úpravou (normalizací, diskretizací). Obsahuje následující sloupce:

- Id – ID záznamu
- Okres – název okresu
- Počet obyvatel dorost – počet obyvatel ve věku do 15
- Počet obyvatel dospělí – počet obyvatel od věku 15 do 65
- Procento dorost – procentuální zastoupení dorostu v populaci
- Procento dospělí – procentuální zastoupení dospělých v populaci
- Obor péče – zkoumaný obor péče
- Počet poskytovatelů – počet poskytovatelů v daném oboru

DotazCUpraveny.csv

Tento soubor obsahuje již upravená data pro dolování. Soubor obsahuje sloupce:

- Id – ID záznamu
- Okres
- Počet všech obyvatelů – celkový počet obyvatelů
- Počet obyvatelů dorost – počet obyvatel ve věku do 15 let
- Obor péče – zkoumaný obor péče
- Počet poskytovatelů – počet poskytovatelů v daném oboru
- Okres-kategorie – zda okres patří do malé, střední nebo velké kategorie
- Počet všech obyvatelů (norm) – normovaný počet obyvatelů v okrese

Poznámka k práci s databází a CSV soubory

V zadání je zmíněno, že veškerá práce má probíhat se soubory ve formátu CSV.

Dle domluvy po e-mailu v našem řešení pracujeme s daty z databáze. Požadované CSV soubory ale stále vytváříme, a ty obsahují jen data, která skutečně používáme v dotazech.

Při spuštění projektu jsou vytvořeny CSV s daty potřebnými k dotazům, následně jsou případně vytvořeny soubory s upravenými daty.