

Projekt UPA – 2. část

Výsledky

Téma: Zdravotnictví

Tím: xryban00

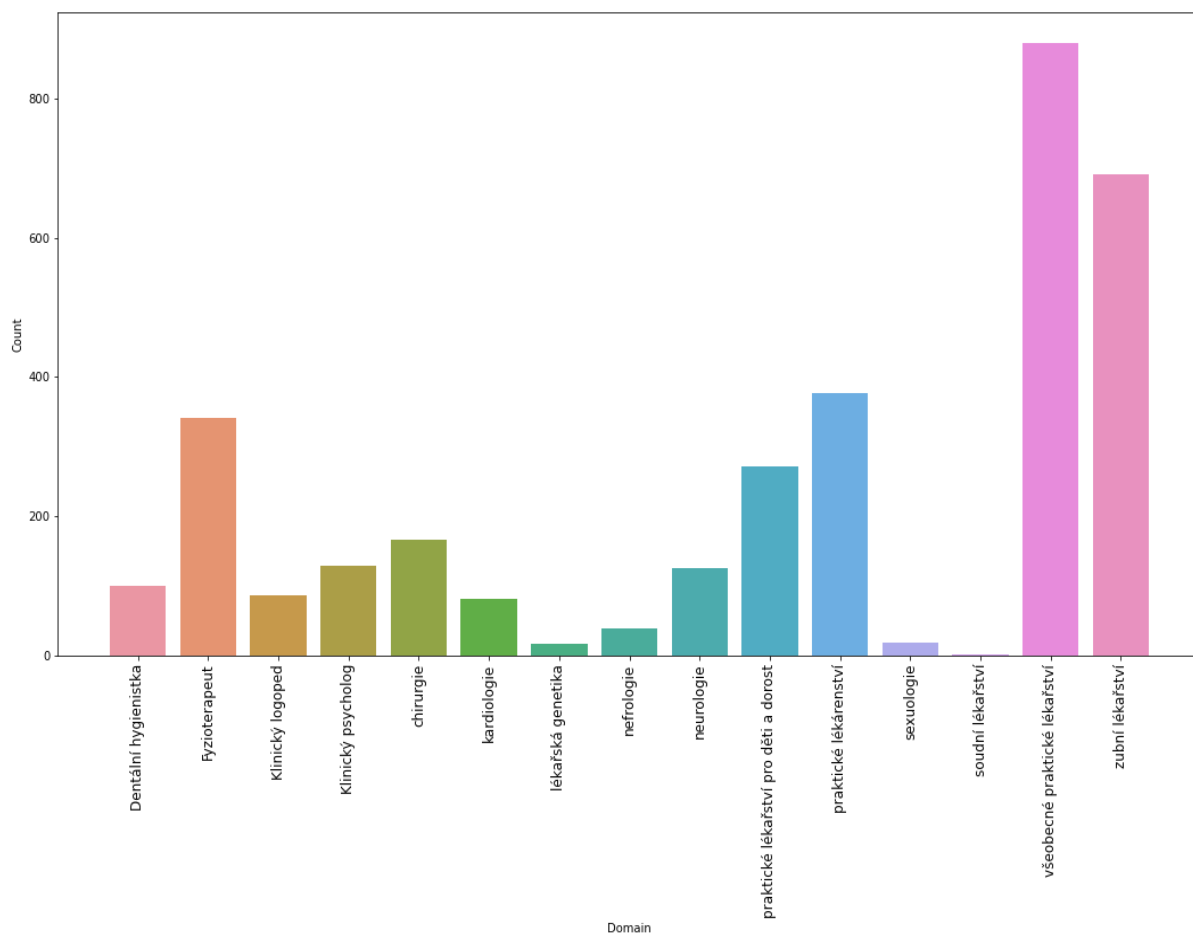
Členovia tímu: Dávid Pukanec (xpukan02)
Adam Rybanský (xryban00)
Michal Sandanus (xsanda03)

Dotazy skupiny A

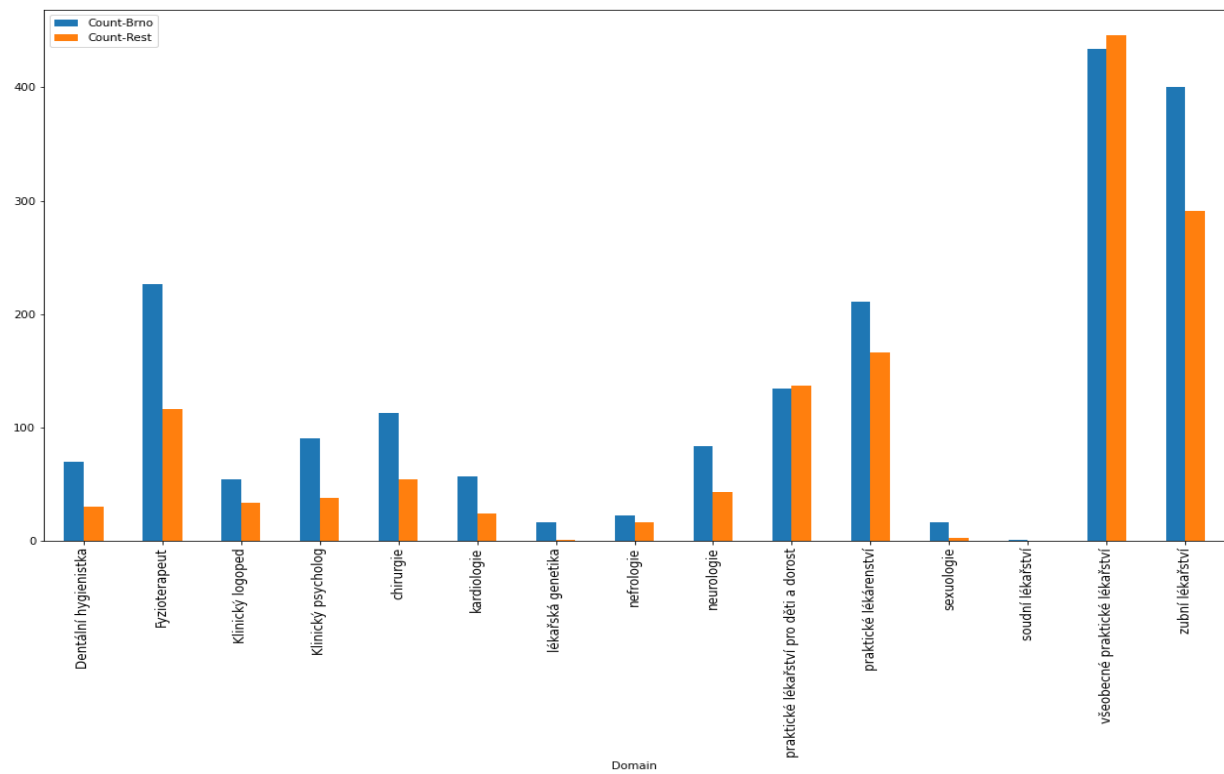
Dotaz A1

Zadanie: Vytvořte sloupcové grafy zobrazující počty poskytovatelů určitého oboru pro Brno a zbytek Jihomoravského kraje. Pro zobrazení libovolně vyberte 15 typů poskytovatelů.

Jedna interpretácia úlohy: graf pre celý Juhomoravský kraj

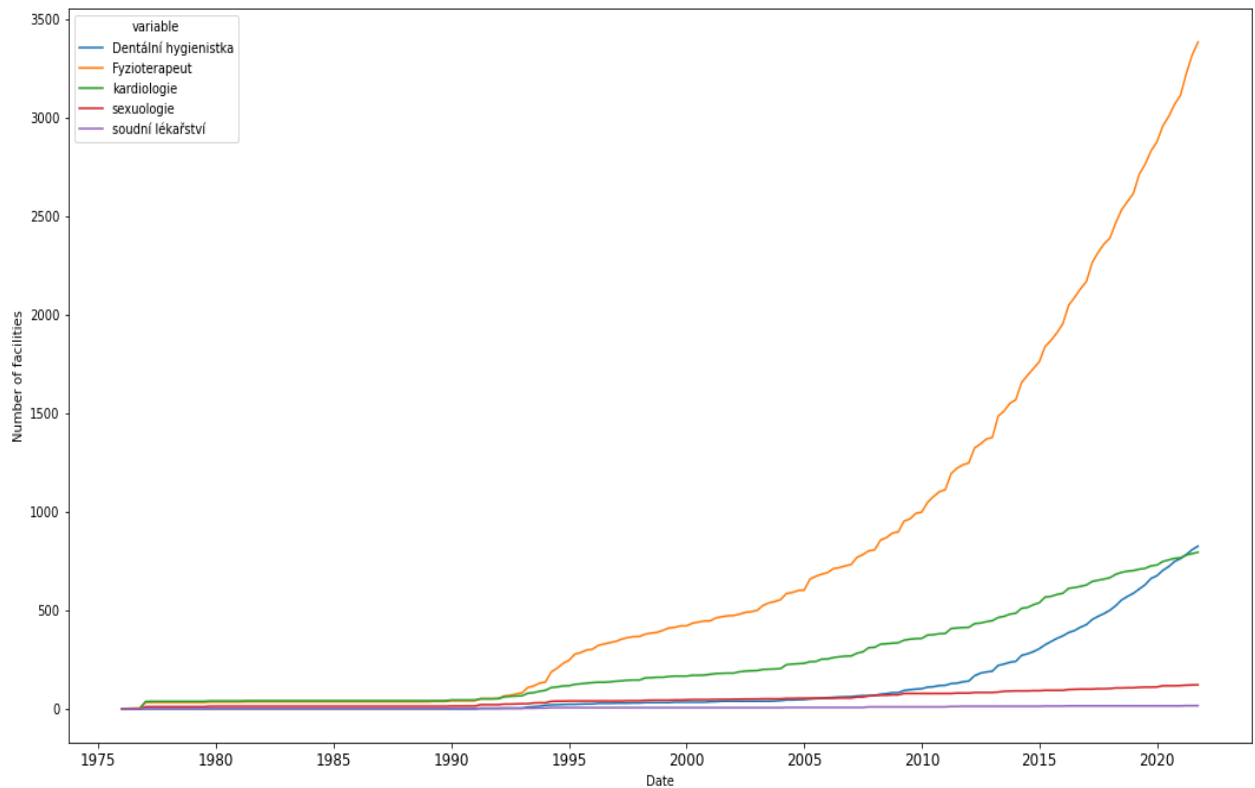


Iná interpretácia úlohy: graf porovnania Brna(mesto + venkov) a zvyšku Juhomoravského kraja



Dotaz A2

Zadanie: Vytvořte čárový (spojnicový) graf zobrazující historii počtu poskytovatelů v 5 zvolených oborech. Vyberte obory, u kterých lze očekávat výraznější změny v počtu poskytovatelů (např. dentální hygienistka). Nemusíte zobrazovat data pro každý měsíc, stačí čtvrtletní hodnoty.

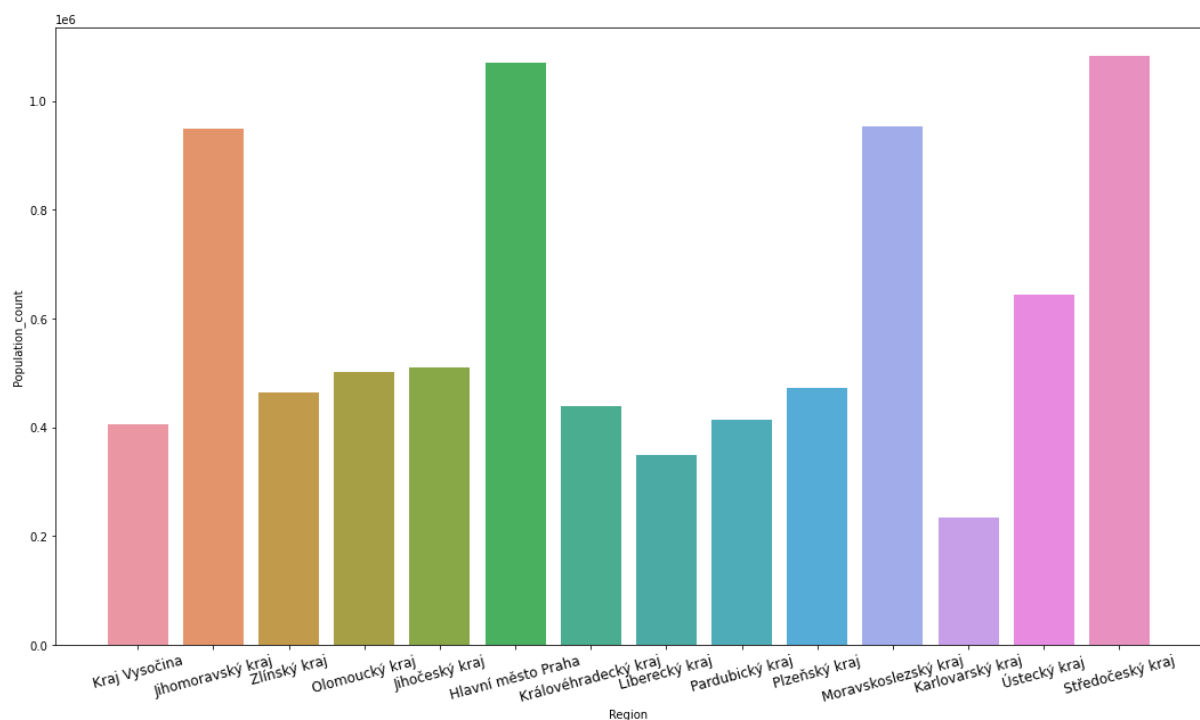


Dotazy skupiny B

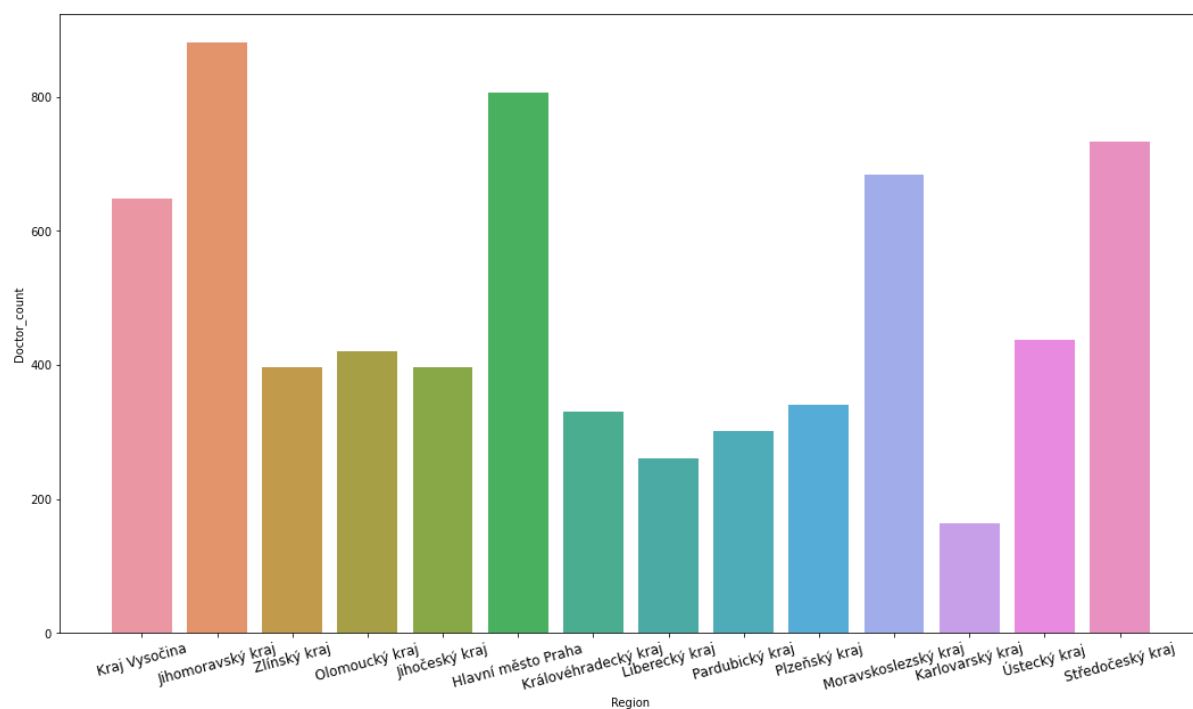
Dotaz B1

Zadanie: Sestavte žebříček krajů dle počtu obyvatel na jednoho praktického lékaře (obor všeobecné praktické lékařství), výsledky zobrazte graficky. Graf bude pro každý kraj zobrazovat počet praktických lékařů v kraji, celkový počet obyvatel a počet obyvatel na jednoho lékaře. Graf můžete zhotovit kombinací dvou grafů do jednoho (jeden sloupcový graf zobrazí první dvě hodnoty a druhý, čárový graf, hodnotu třetí). Pro přesnější výsledky použijte počet obyvatel kraje nad 20 let.

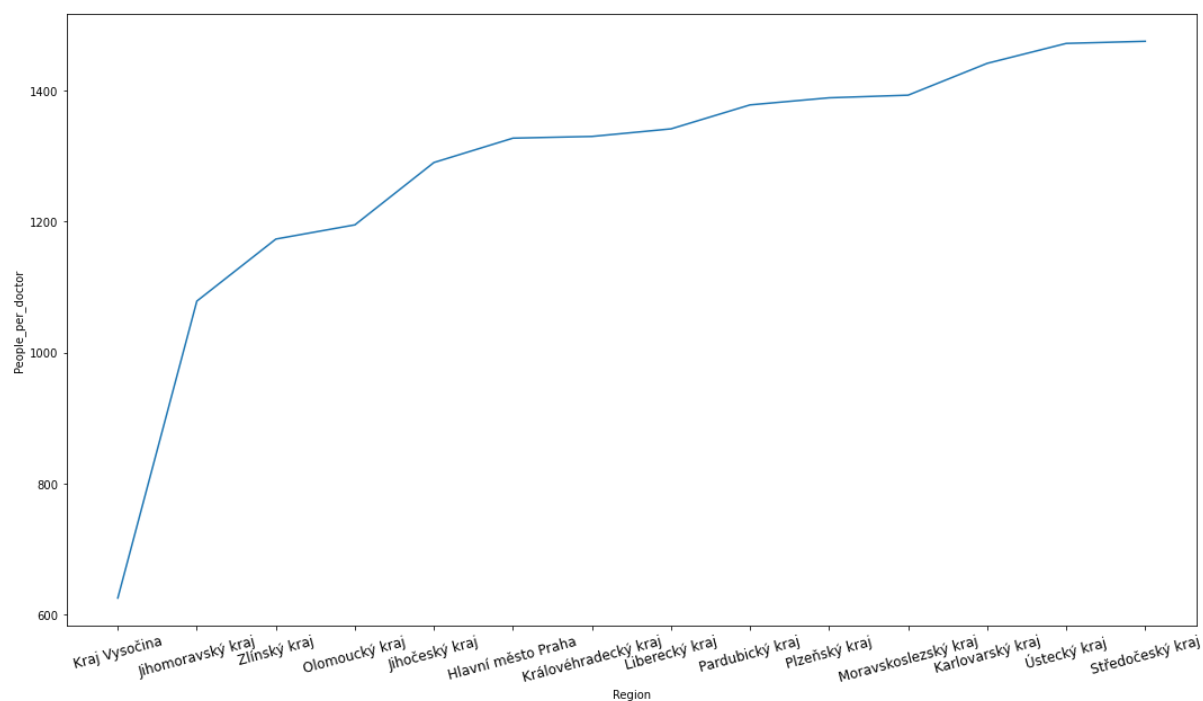
Jednotlivé grafy: graf počtu obyvatelů



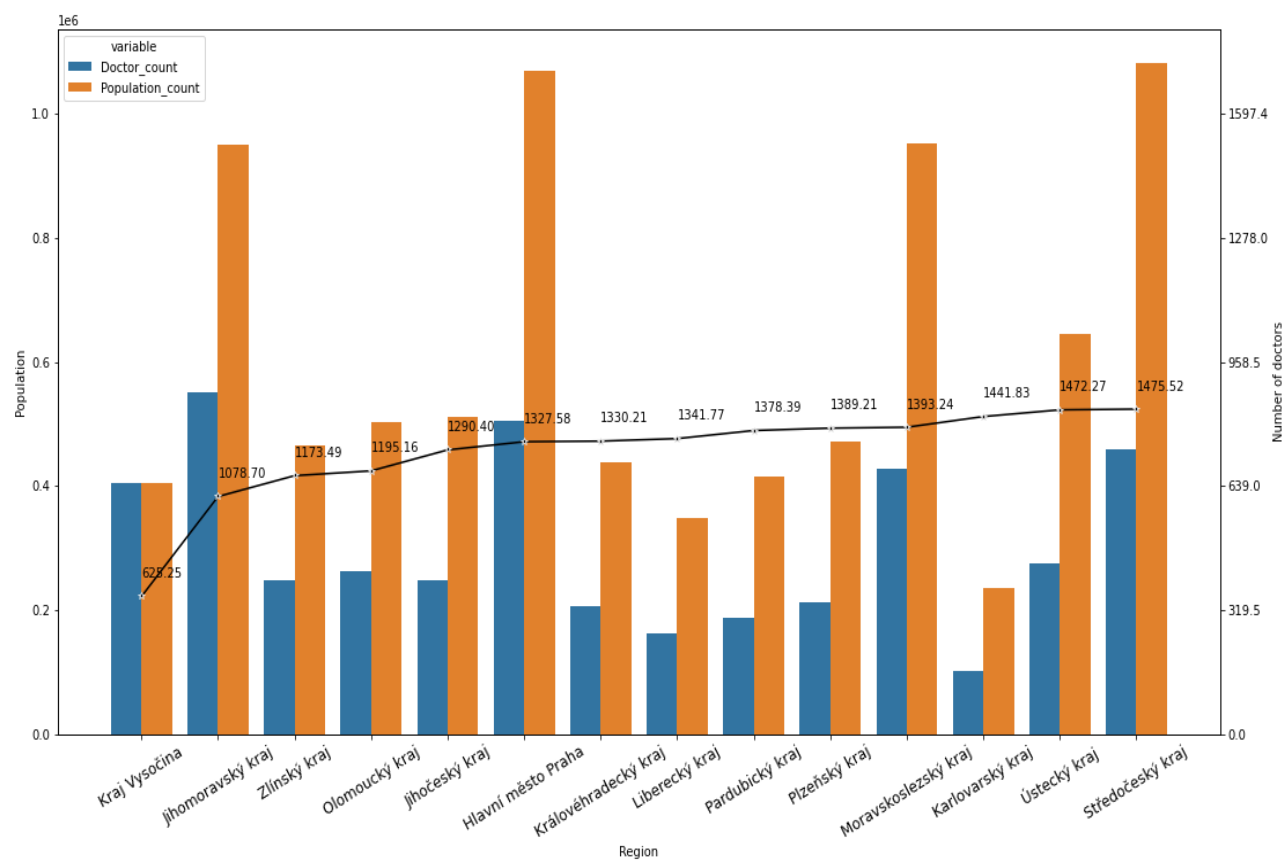
Jednotlivé grafy: graf počtu poskytovatelů



Jednotlivé grafy: graf počtu obyvatelů na lékaře



Spojený graf:



Vlastné dotazy

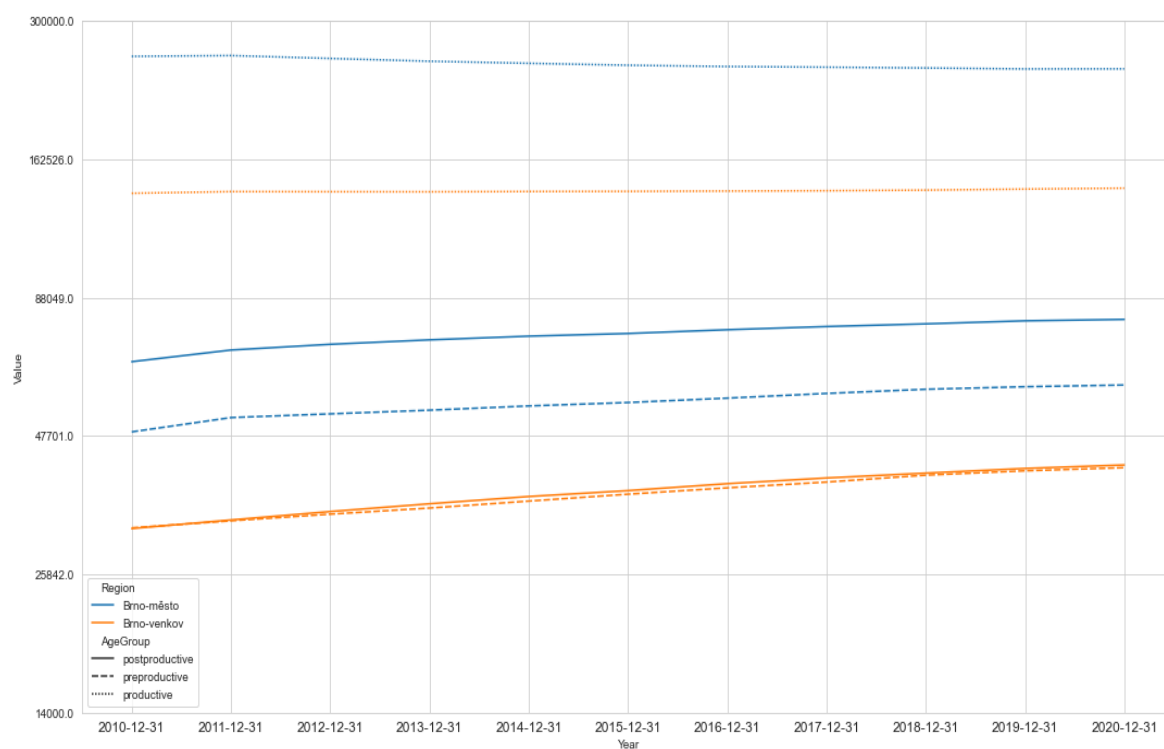
Vlastný dotaz 1

Zadanie: Vykreslite vývoj počtu obyvateľstva rozdeleného do ekonomických vekových skupín za uplynulé roky v okresoch Brno-město a Brno-venkov.

Ekonomické vekové skupiny: Predproduktívny vek: 0 - 15

Produktívny vek: 15 - 65

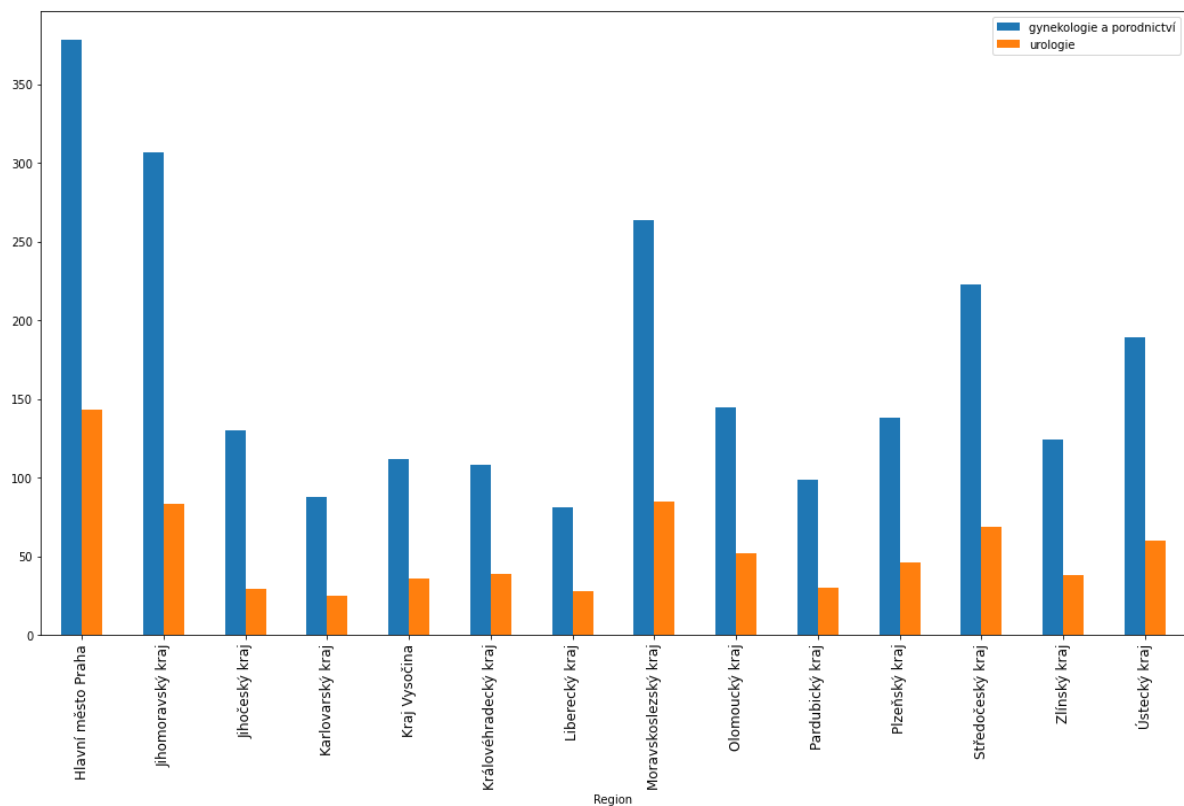
Postproduktívny vek: 65 a viac



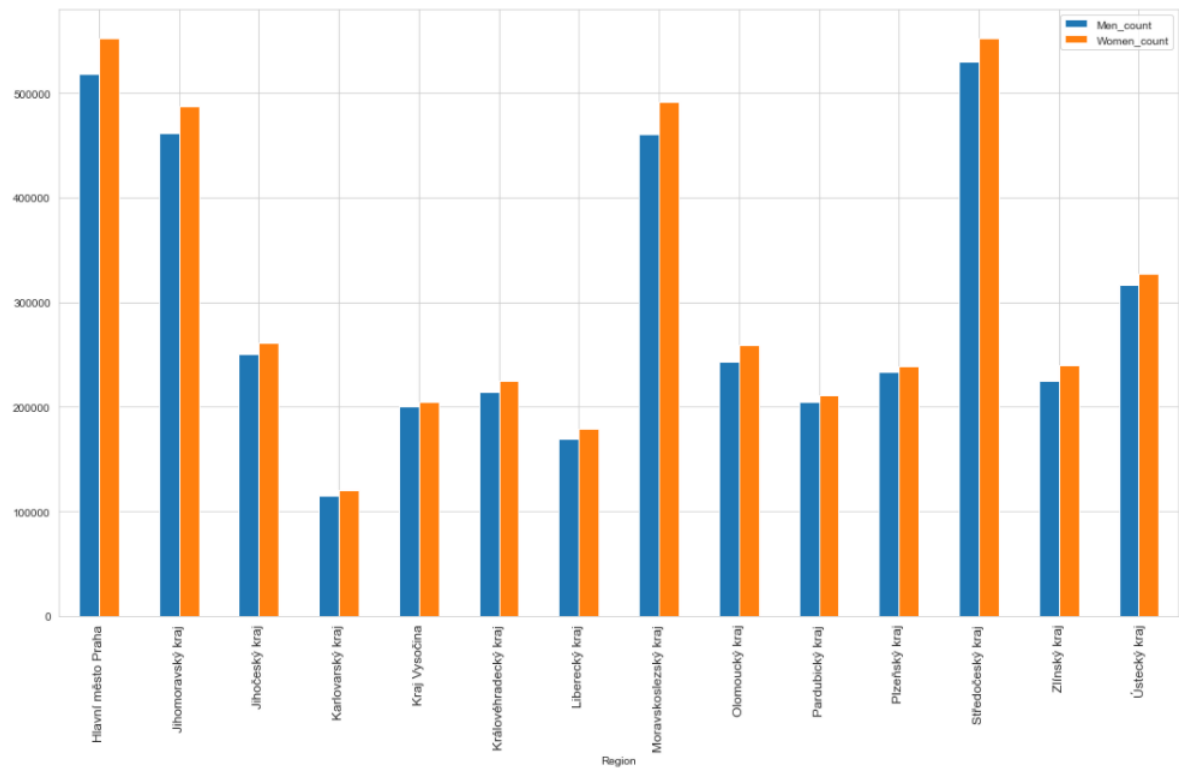
Vlastný dotaz 2

Zadanie: Porovnanie počtu urologických ambulancií na počet mužských obyvateľov kraja (nad 20 rokov) a gynekologických ambulancií na počet ženských obyvateľov kraja (nad 20 rokov) v rámci krajov Českej republiky.

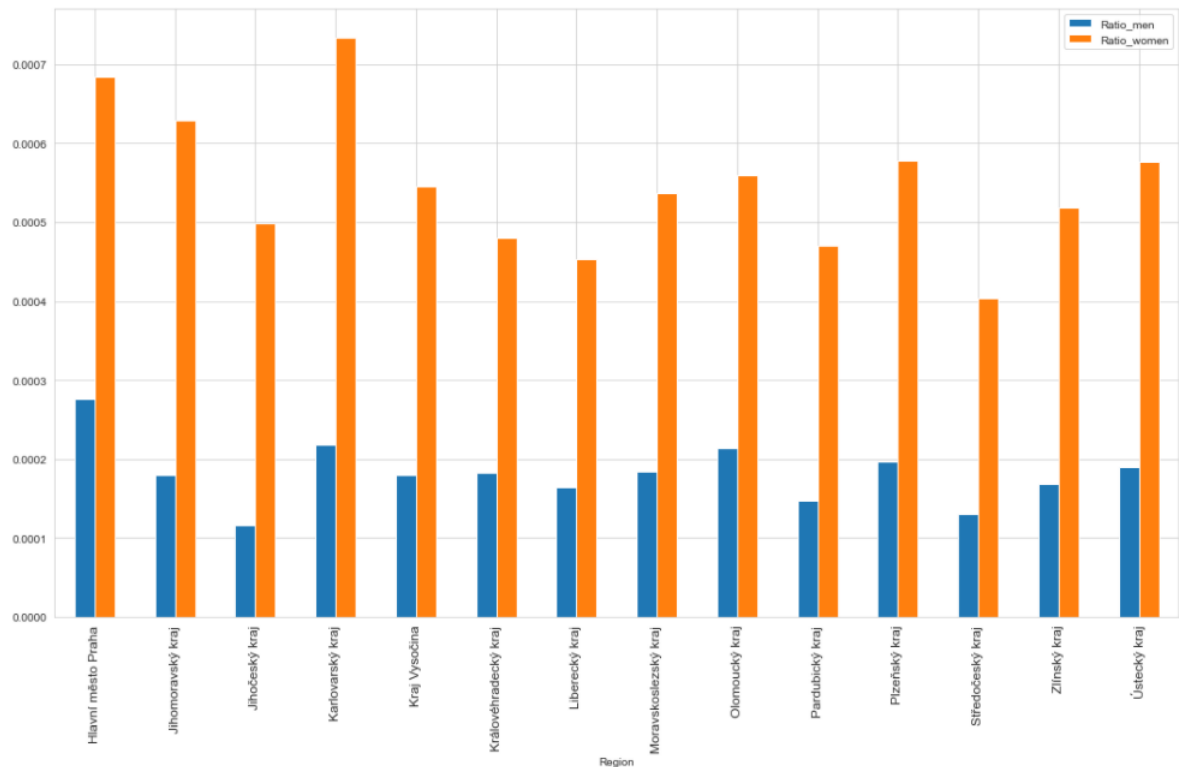
Graf 1: porovnanie počtu gynekologických a urologických ambulancií v rámci krajov



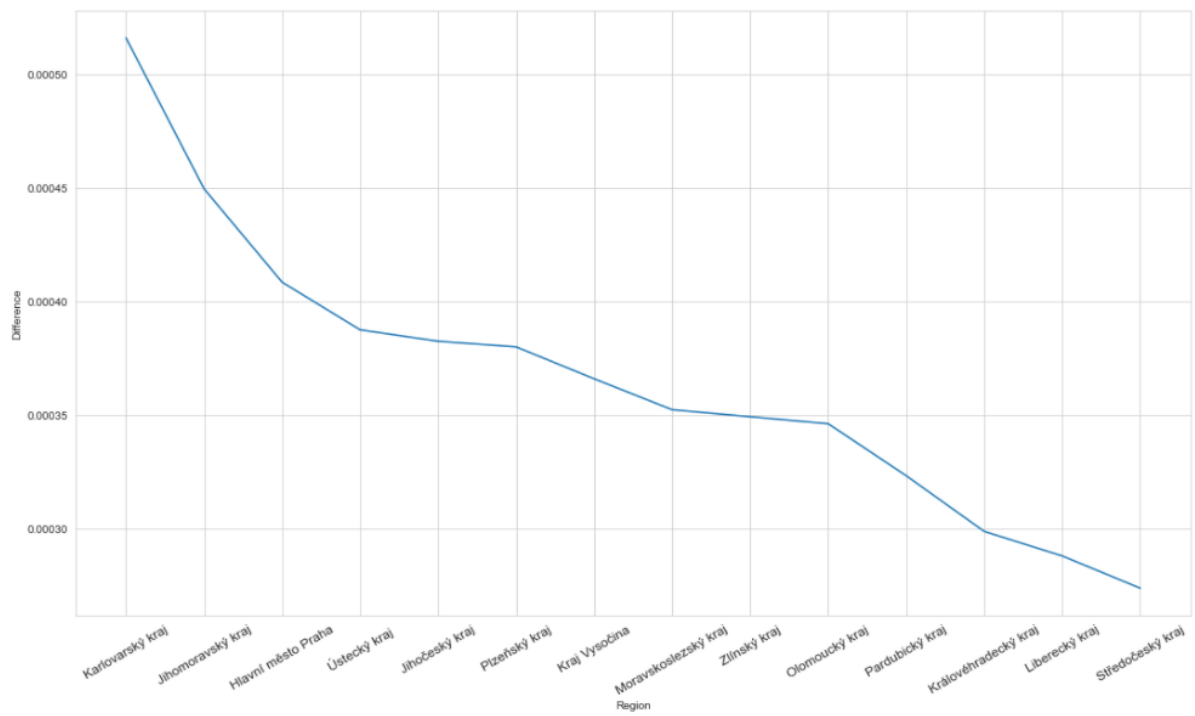
Graf 2: stĺpcový diagram pre porovnanie počtu dospelých mužov a dospelých žien v rámci krajov



Graf 3: porovnanie pomerov v rámci jednotlivých krajov



Graf 4: zoradenie krajov podľa najväčšieho rozdielu pomerov (zostupne) a vykreslené čiarovým grafom



Časť 2: predspracovanie dát

Úloha:

- Pripravte trénovací datovou sadu pro klasifikátor, který na základě počtu poskytovatelů vybraných zdravotních služeb a věkového složení obyvatel odhadne velikost města (jako kategorický atribut - malé/střední/velké podle počtu obyvatel).
 - Uvažujte následující kategorie měst: malá - do 8 tis. obyvatel, střední 8 - 30 tis. obyvatel a velká nad 30 tis. obyvatel.
 - Pro potřeby projektu vyberte libovolně 60 měst (20 z každé kategorie), pro která najdete potřebné hodnoty.
 - Atributy: počty poskytovatelů pro následující obory péče: dentální hygienistka/dětské lékařství/fyzioterapeut/chirurgie/kardiologie/Optometrista/ortopedie a traumatologie pohybového ústrojí/praktické lékařství pro děti a dorost/všeobecné praktické lékařství/zubní lékařství, procento obyvatel ve věkových skupinách 0-14 let a 15-64 roků.

Zadanie:

Zvolte si jednu z dolovacích úloh ze skupiny C. Pro tuto úlohu pripravte data tak, aby výsledná data mohla být použita dolovacím algoritmem. Tzn. pripravte soubor ve formátu CSV, kde každý řádek bude odpovídat jednomu objektu, každý sloupec nějakému atributu. Dále pak ve vybraných datech detekujte odlehlé hodnoty a nahraďte je jinou vhodnou hodnotou, pro jeden zvolený sloupec provedte normalizaci hodnot a pro jiný sloupec diskretizaci hodnot. Opět implementujte nástroj, který potřebná data extrahuje z úložiště vytvořeného v první části projektu a dále je požadovaným způsobem upraví. Samotné dolování z dat není vyžadováno.

Popis vytvoreného CSV súboru pred úpravami

CSV súbor sa volá "C1_OUTPUT.csv", pozostáva zo 60 riadkov. Každý riadok reprezentuje jeden okres, a v stĺpcoch sú uvedené atribúty: percento obyvateľov vo veku 0-14, percento obyvateľov vo veku 15-64, počty zvolených poskytovateľov zdravotných služieb, a veľkosť okresu (small, medium alebo big, pričom small je do 95 000 obyvateľov, medium od 95 000 do 120 000, a big od 120 000). Náhodne sme vybrali 20 okresov každej veľkosti.

Tu je ukážka .csv súboru:

	Region	15-64_percentage	0-14_percentage	Size	Dentální hygienistka	Fyzioterapeut	Optometrista	chirurgie	dětské lékařství	kardiologie	ortopedie a traumatologie pohybového ústrojí	praktické lékařství pro děti a dorost	zubní lékařství
0	Benešov	0.579097	0.161048	medium	13.0	25.0	2.0	11.0	2.0	6.0	8.0	30.0	48.0
1	Beroun	0.589125	0.171869	medium	2.0	22.0	7.0	17.0	5.0	10.0	9.0	22.0	46.0
2	Blansko	0.582136	0.158906	medium	7.0	20.0	9.0	9.0	10.0	5.0	17.0	22.0	54.0
3	Brno-město	0.592070	0.150299	big	61.0	193.0	57.0	100.0	5.0	49.0	77.0	90.0	334.0
4	Brno-venkov	0.585274	0.175605	big	9.0	33.0	20.0	13.0	0.0	8.0	13.0	44.0	66.0

Popis vytvoreného CSV súboru po úpravách

Odstránenie outlierov robíme nahradením vychýlených hodnôt minimami a maximami nachádzajúcich sa v medzikvartilovom rozpätí.

Počty nahradných hodnôt pre stĺpce

15-64_percentage	0-14_percentage	Dentální hygienistka	Fyzioterapeut	Optometrista	Chirurgie
2	3	5	6	4	4
Dětské lékařství	Kardiologie	Ortopedie a trau...	Praktické lék...	Zubní lékařství	Spolu
3	5	5	5	6	48

Normalizáciu robíme pomocou z-score cez scipy knižnicu. Normalizujeme stĺpec obsahujúci počty zariadení Dentalnej hygieny v okresoch.

```
dent_hyg = df_out_rem['Dentální hygienistka'].values
df_normalized = zscore(dent_hyg)
print(df_normalized)
```

```
[-0.33896708  1.94620422  1.94620422  0.34658431  0.57510144  0.80361857
 1.94620422  1.94620422  0.11806718  1.71768709  0.11806718  1.94620422
-1.02451847 -1.2530356  -0.11044995  1.94620422 -0.33896708  1.48916996
-0.56748421  1.26065283 -0.11044995 -0.33896708 -1.02451847 -0.79600134
-0.79600134  0.11806718 -0.79600134 -0.79600134 -1.02451847 -1.2530356
-1.2530356 -0.79600134 -1.2530356 -1.2530356  0.34658431 -0.79600134
-0.79600134 -1.02451847  0.11806718 -1.02451847  0.11806718 -0.56748421
 0.11806718 -0.33896708 -0.33896708  0.34658431 -0.79600134  0.57510144
-0.79600134 -0.11044995 -0.33896708  0.34658431  0.80361857  1.48916996
-0.79600134 -0.11044995 -1.02451847  1.71768709 -0.79600134  0.57510144]
```

Následne sme otestovali priemer na 0 a smerodajnú odchýlku na 1.

Diskretizáciu sme robili pomocou Bucketing/Plnenia kedy sme rozdelili atribút Chirurgie na 5 kategórii (0..4) reprezentujúcich veľkosti mesta (malé, stredne-malé, stredné, stredne-veľké, veľké).

	Region	Size	chirurgie	Discretization_chirurgie
0	Ústí nad Orlicí	big	31.0	4
1	Brno-město	big	31.0	4
2	Zlín	big	25.0	3
3	Karviná	big	31.0	4
4	Mladá Boleslav	big	20.0	2

Následne sme výsledky spojili do finálneho datasetu tak že sme stĺpec Dentálnej hygieny nahradili normalizovanými hodnotami a stĺpec chirurgie sme nahradili diskretizovaným stĺpcom Chirurgie.

	Region	15-64_percentage	0-14_percentage	Size	Dentální hygienistka	Fyzioterapeut	Optometrista	chirurgie
0	Ústí nad Orlicí	0.587107	0.153907	big	-0.294911	35.0	8.0	31.0
1	Brno-město	0.592070	0.150299	big	2.232897	47.0	16.0	31.0
2	Zlín	0.588118	0.146502	big	2.232897	47.0	13.0	25.0
3	Karviná	0.592256	0.139571	big	0.164691	39.0	16.0	31.0
4	Mladá Boleslav	0.600920	0.158641	big	2.232897	36.0	7.0	20.0

dětské lékařství	kardiologie	ortopedie a traumatologie pohybového ústrojí	praktické lékařství pro děti a dorost	zubní lékařství
3.0	4.0	23.0	28.0	54.0
5.0	14.0	26.0	44.0	90.0
12.0	14.0	21.0	30.0	90.0
7.0	14.0	26.0	44.0	90.0
4.0	9.0	17.0	27.0	67.0

Výsledek je uložený do súbora C1_OUTPUT_2.csv