

Študentska prehrana

projektna naloga

Valerij Jovanov
Fakulteta za matematiko in fiziko
Oddelek za matematiko

7. september 2022

1 Uvod

S spletne strani <https://www.studentska-prehrana.si/sl/restaurant> sem pridobil html datoteke, ter iz njih izluščil različne podatke o vseh lokalih na študentske bone, kot so:

- doplačilo pri nakupu s študentskimi boni
- mesto in lokacijo
- oceno lokala
- meni lokala
- posebnosti lokala

Te podatke sem nato spravil v CSV datoteke, ter jih s pomočjo knjižnice pandas analiziral, ter narisal nekaj grafov. Pitonski datoteki za zajem in obdelavo podatkov sta:

- glavna.pridobitev.py - datoteka za pridobitev podatkov
- glavna.obdelava.py - datoteka za luščenje podatkov in delanje CSV datotek

HTML datoteke so:

- glavna_stran.html - html datoteka glavne strani

- `podrobne_informacije.html` - html datoteke s podrobnimi informacijami o lokalih

CSV datoteke so:

- `lokali.csv` - CSV datoteka s glavnimi podatki o lokalih
- `ocene.csv` - CSV datoteka s podatki o ocenah lokalov
- `posebnosti.csv` - CSV datoteka s podatki o posebnostih lokalov
- `jedi.csv` - CSV datoteka s podatki o vseh jedeh, ter vrsti jedi

IPYNB datoteke so:

- `analiza.ipnyb` - datoteka s različnimi tabelami in grafi
- `napovedovanje.ipnyb` - datoteka s pomočjo katere poskušam iz hrane napovedati vrsto hrane

Moje hipoteze so bile:

- Ali obstaja povezava med oceno lokala ter doplačilom za obrok? - Predvidevam da bo imel lokal z manjšim doplačilom boljšo oceno.
- Imajo lokali s več pozitivnimi lastnostmi tudi višje cene? - Predvidevam, da bodo imeli lokali z več pozitivnimi lastnostmi višjo ceno
- Imajo lokali v Ljubljani višje cene? - Zaradi lokacije v centru imajo lokali verjetno višjo ceno.
- Ali obstaja optimalen lokal glede na ceno, oddaljenost od FMF, ter število menijev v primeru, da je kavarna Mafija prepolna? - Hipoteza je bila dodana zaradi lastnih interesov.

2 Pridobitev podatkov

Kot že omenjeno sem vse podatke pridobil iz spletne strani <https://www.studentska-prehrana.si/sl/restaurant> s pomočjo datoteke `glavna_pridobitev.py`. Datoteka vsebuje 4 metode. To so:

- `prenesi_iz_url(url)` - metoda kot argument sprejme url, ter se če spletna strani z url-jem obstaja, poveže na spletno stran, ter pridobi niz s html-jem spletne strani.

- `shrani_v_datoteko(besedilo, mapa, ime_datoteke)` - metoda kot argumente sprejme besedilo, mapo za shranjevanje, ter ime datoteke v katero želimo besedilo zapisati, ter besedilo zapiše v želeno datoteko in jo shrani v mapo.
- `preberi_id_lokalov(mapa, ime_datoteke)` - metoda kot argumente sprejme mapo in ime CSV datoteke s podatki o id-jih lokalov. Metoda iz datoteke prebere vse možne id-je lokalov ter jih vrne v tabeli.
- `preberi_podrobne_informacije(datoteka, num = 20)` - metoda prejme kot argument datoteko s id-ji lokalov, ter število HTML strani na datoteko. Metoda najprej pridobi informacije o id-jih lokalov, nato pa s pomočjo le teh pridobi podrobne informacije o lokalih, ter jih spravi v `num` datotek.

Na spletni strani z lokali na študentske bone ima vsak lokal svoj id, s pomočjo katerega lahko sestavimo nov url in dobimo podrobnejše informacije o lokalih kot je na primer jedilnik. S pomočjo `prenesi_iz_url(url)` sem najprej prenesel osnovno stran ter naredil CSV datoteko z id-ji. Potem sem s pomočjo te in metode `preberi_podrobne_informacije(datoteka, num = 20)`, prebral html datoteke za vsak lokal posebej ter jih spravil v 18 datotek.

3 Obdelava podatkov

Glavna datoteka za obdelavo je `glavna_obdelava.py`. Ta vsebuje 7 metod:

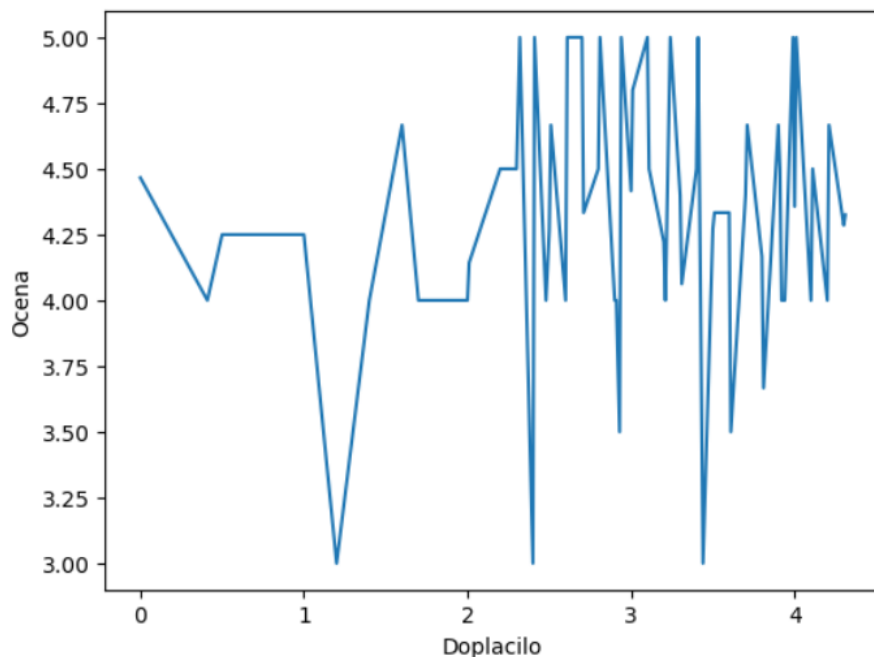
- `datoteka_v_niz(mapa, ime_datoteke)` - metoda kot argumenta sprejme mapo kjer se datoteka nahaja, ter ime datoteke. Metoda prebere podatke datoteke, ter vrne niz.
- `stran_v_lokale(stran)` - metoda kot argument sprejme html stran, ter s pomočjo regularnih izrazov iz strani izlušči podatke o id lokala, imenu lokala, naslovu lokala, doplacilu, ter mestu lokala ter jih spravi v CSV datoteko.
- `stran_v_posebnosti(stran)` - metoda kot argument sprejme html stran, ter s pomočjo regularnih izrazov prebere podatke o posebnostih lokalov, ter jih spravi v CSV datoteko.
- `stran_v_ocene(stran)` - metoda kot argument sprejme html stran, ter s pomočjo regularnih izrazov prebere podatke o ocenah lokalov, ter jih spravi v CSV datoteko.

- `pridobi_jedi(mapa)` - metoda, ki kot argument sprejme mapo z html datotekami o podrobnih informacijah o lokalih, ter s pomočjo regularnih izrazov izlušči podatke, ter naredi CSV datoteko z jedmi, vrsto jedi, ter lokalom iz katerega je ta jed.
- `pridobi_podatke_lokala(lokal, vzorec)` - pomožna metoda, ki kot argument prejme HTML lokala ter vzorec za regularne izraze in vrne slovar.
- `napisi_csv(polja, vrstice, mapa, ime_datoteke)` - metoda, ki kot argumente prejme tabelo polj za glavo, seznam slovarjev, ki jih mora zapisati v CSV datoteko, mapo in ime CSV datoteke. Metoda slovarje zapiše v dano datoteko.

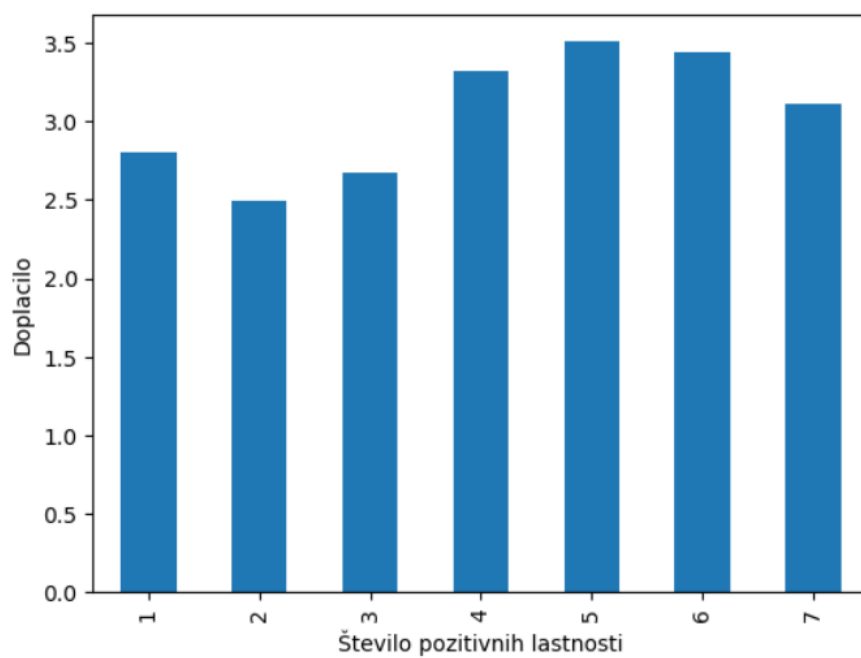
S pomočjo zgoraj opisanih metod naredim 4 CSV datoteke s podatki o lokalih.

4 Analiza podatkov

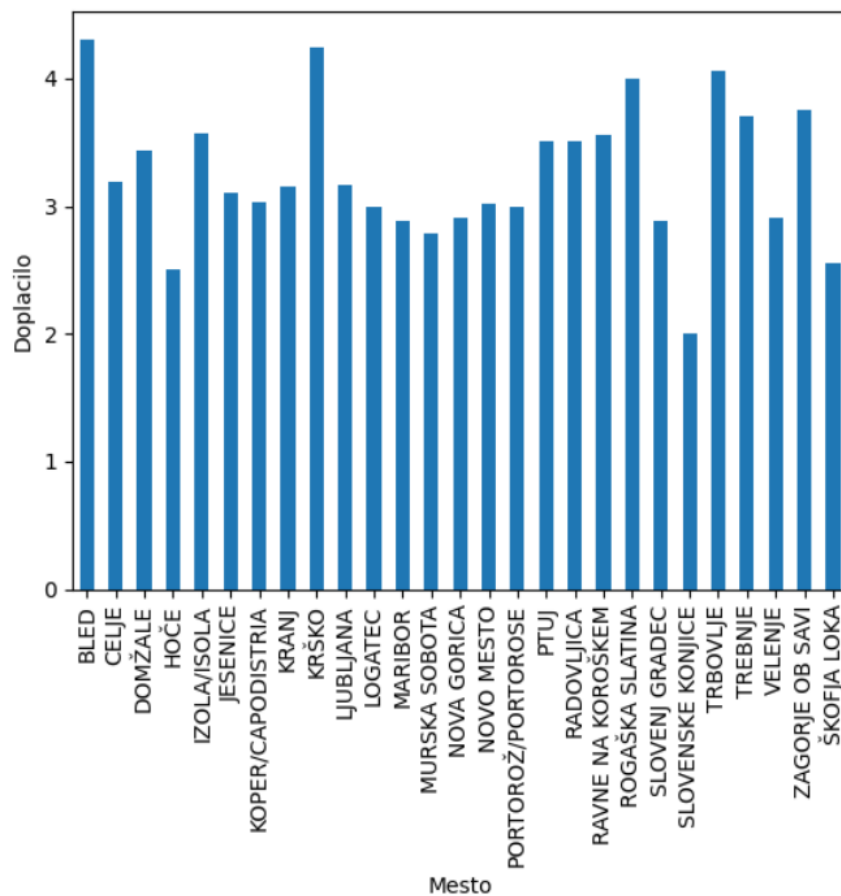
Glavna datoteka tu je `analiza.ipynb`. Najprej sem izpisal vse 4 tabele za boljšo preglednost. Oglejmo si graf ocen v odvisnosti od doplačila:



Kot vidimo iz grafa, ocene z višanjem cene vedno bolj nihajom, a žal ne vidimo bistvene relacije med doplačilom in oceno. Oglejmo si sedaj graf doplacija v odvisnosti od količine pozitivnih lastnosti oziroma posebnosti, ki jih ima lokal:



Kot lahko vidimo iz grafa imajo lokali z več pozitivnimi lastnostmi v povprečju tudi višjo ceno. Oglejmo si še graf doplačila v odvisnosti od mesta:



Nasprotno od pričakovanj Ljubljana nima najvišjega doplačila. Namesto nje to mesto zasede Bled. V datoteki je narisanih še nekaj grafov in tabel. Za konec sem s pomočjo zapiskov predavnanj naredil še klasifikator, ki iz imena jedi napove vrsto jedi. Ta je v datoteki napovedovanje.ipynb.

5 Zaključek

Od mojih hipotez je bila pravilna le druga, torej, da je doplačilo večje pri lokalih z več pozitivnimi lastnostmi. Žal nisem našel lokala, ki bi nadomestil Mafijo za kratek 15 minutni odmor, je pa menza na fakulteti za elektrotehniko dobra alternativa, če imamo malo več časa in smo malo bolj lačni.