

# Analiza podataka pomoću Python Pandas alata kroz primjer

---

Hohnjec, Andrea

Undergraduate thesis / Završni rad

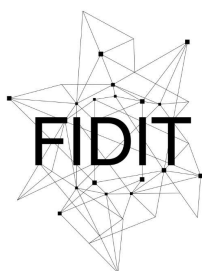
2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:195:342603>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-16**



Sveučilište u Rijeci  
**Fakultet informatike  
i digitalnih tehnologija**

Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



zir.nsk.hr

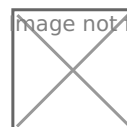


Image not found or type unknown

Sveučilište u Rijeci – Odjel za informatiku

Jednopredmetna informatika

Andrea Hohnjec

# Analiza podataka pomoću Python Pandas alata kroz primjer

Završni rad

Mentor: dr. sc. Lucia Načinović Prskalo

Rijeka, 2019.

Rijeka, 15.3.2018.

## Zadatak za završni rad

Pristupnik: Andrea Hohnjec

Naziv završnog rada: Analiza podataka pomoću Python Pandas alata kroz primjer

Naziv završnog rada na eng. jeziku: Data analysis with Python Pandas with practical examples

Sadržaj zadatka: Zadatak završnog rada je opisati i pojasniti postupak analize podataka i osnovne pojmove koje se vežu uz to područje, opisati pojedine faze u postupku analize podataka te primijeniti alat Python Pandas za implementaciju analize podataka na odabranom skupu podataka.

Mentor

Dr. sc. Lucia Načinović Prskalo

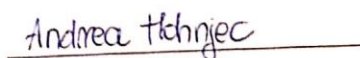


Voditelj za završne radove

Dr. sc. Miran Pobar



Zadatak preuzet: 22.3.2018.



(potpis pristupnika)

|                                                                                            |    |
|--------------------------------------------------------------------------------------------|----|
| 1.Uvod .....                                                                               | 4  |
| 2.Vrste analize podataka .....                                                             | 6  |
| 2.1.Deskriptivna analiza podataka (descriptive data analysis) .....                        | 6  |
| 2.2.Istraživačka analiza podataka (exploratory data analysis) .....                        | 6  |
| 2.3.Prediktivna analiza podataka (predictive data analysis) .....                          | 7  |
| 3.Strukture podataka.....                                                                  | 8  |
| 3.1.Niz podataka .....                                                                     | 8  |
| 3.2.Podatkovni okvir.....                                                                  | 9  |
| 3.3.Panel .....                                                                            | 10 |
| 4.Opis postupka analize podataka .....                                                     | 12 |
| 4.1.Prikupljanje podataka .....                                                            | 12 |
| 4.2.Čišćenje podataka .....                                                                | 12 |
| 4.3.Primjena različitih operacija organiziranja i upravljanja nad skupovima podataka ..... | 12 |
| 4.3.1.Sortiranje.....                                                                      | 12 |
| 4.3.2.Agregacija .....                                                                     | 15 |
| 4.3.3.Grupiranje .....                                                                     | 18 |
| 4.3.4.Filtriranje.....                                                                     | 19 |
| 4.3.5.Transformiranje.....                                                                 | 21 |
| 4.3.6.Spajanje.....                                                                        | 24 |
| 5.Primjena alata Python Pandas za analizu podataka na odabranim skupovima podataka .....   | 26 |
| 5.1.Opis skupova podataka .....                                                            | 26 |
| 5.2.Opis Python Pandas alata .....                                                         | 26 |
| 5.3.Primjena i rezultati.....                                                              | 27 |
| 5.3.1.Primjer 1 .....                                                                      | 27 |
| 5.3.2.Primjer 2 .....                                                                      | 28 |
| 6.Zaključak .....                                                                          | 31 |
| Literatura .....                                                                           | 32 |
| Popis slika .....                                                                          | 33 |

## 1.Uvod

Analiza podataka je proces uvida, čišćenja, transformiranja i modeliranja podataka s ciljem otkrivanja korisnih informacija, stvaranja zaključaka i lakšeg donošenja odluka. Analiza podataka ima višestruke aspekte i pristupe, obuhvaćajući različite tehnike pod različitim imenima, a koristi se u mnogim područjima poslovanja i znanosti. U današnjem poslovnom svijetu analiza podataka ima ulogu u donošenju znanstvenih odluka i pomaganju poduzeća da djeluje učinkovitije (1).

Još od pojave računanja, analiza podataka i računalna tehnologija razvijaju se i utječu jedno na drugo. Kako je kroz povijest količina podataka rasla tako su se uvodile nove metode za analizu podataka.

Za provedbu analize podataka potrebni su ulazni podaci prema kojima će se raditi analiza. Podatke je moguće prikupiti iz različitih izvora kao što su životno okruženje, poput prometnih kamera, satelita i intervjua. Mogu se preuzeti sa internetskih stranica ili putem čitanja određene dokumentacije.

Analiza podataka postupno se razvijala s vremenom, a danas preuzima sve veću ulogu u mnogim tvrtkama.

Analiza podataka ima korijene u statistici. Statistika je kroz povijest imala važnu ulogu za vlade diljem svijeta, najviše kod popisa stanovništva koji se koristio za vladine aktivnosti kao što je oporezivanje. Pomoću prikupljenih podataka koje je nakon toga bilo potrebno analizirati s ciljem otkrivanja korisnih informacija, razvila se analiza podataka. Analiza rasta stanovništva po određenim regijama mogla je pomoći vladama u određivanju broja bolnica koje će biti potrebne u određenom području. Izum računala i napredak u računalnoj tehnologiji doveo je do velikih poboljšanja u području provedbe analize. Primjerice, prije postojanja računala, za popis stanovništva u SAD-u 1880. godine bilo je potrebno 7 godina da se prikupljeni podaci obrade i analiziraju kako bi se došlo do konačnog izvještaja (2).

Američki statističar Herman Hollerith izumio je „tabulatorski stroj“ 1890. godine. To je bio električni stroj koji je radio na bušene kartice, te je mogao sustavno obrađivati podatke koji su na njima snimljeni. Njegovom primjenom brojanje glasova u SAD-u bilo je tri puta brže od prijašnjeg postupka ručnog prebrojavanja, te se Hermana Holleritha smatra začetnikom elektromehaničke obrade podataka (3).

Veliki preokret u analizi podataka dogodio se i 1980-ih godina s pojavom relacijskih baza podataka koje su omogućile korištenje SQL naredbi. To je olakšalo postupak prikupljanja podataka i dovelo do veće upotrebe. Zbog manjih troškova količina prikupljenih podataka se znatno povećala i konstantno rasla. Iz tog razloga William H. Inmon predložio je „skladište podataka“, sustav optimiziran za izvještaje i analizu velike količine podataka neke tvrtke. Podaci su pohranjeni u obliku pogodnom za analizu, imajući u vidu potrebe korisnika. Stoga korisnicima omogućuje jednostavno postavljanje upita i brže pregledavanje podataka (4). Howard Dresner predložio je pojam „Business Intelligence“, koji objedinjuje skup metodologija (Data Warehousing, Data Mining, OLAP) i softverskih alata kojima se omogućuje korištenje podataka iz različitih skladišta podataka (Data Warehouse) i njihovo pretvaranje u informaciju potrebnu za donošenje poslovnih odluka. Prihvatile su ga mnoge velike tvrtke u svojem poslovanju 1989. godine (5).

Početkom 1990-ih godina pojavio se pojam „Data mining“, odnosno rudarenje podataka. Označava računski proces otkrivanja podataka u velikim skupinama na drugačiji način od prijašnjih metoda. Primjenom sortiranja, organiziranja ili grupiranja velikog broja podataka

dolazi se do relevantnih informacija. Zahvaljujući tehnologijama baza podataka i skladištenja podataka omogućen je razvoj rudarenja podataka koji tvrtkama omogućuje pohranjivanje i obradu veće količine podataka za analizu. Rudarenje podataka tvrtke također mogu koristiti za predviđanje, primjerice za predviđanje potencijalnih potreba kupaca na temelju analize prijašnjih podataka o njihovoj kupovini (6).

Sljedeća značajna promjena bila je razvoj interneta. Američki informatičar Lawrence "Larry" Page zajedno sa Sergeyom Brinom razvio je Google pretraživač (7) - internetski pretraživač koji obrađuje i analizira velike količine podataka smještenih na web serveru, na temelju upita koji zadaje korisnik.

Početkom 2010-ih godina objavljen je Amazon Redshift - skladište podataka u oblaku koje nudi paralelne obrade velike količine podataka (8). Razvio se i Google BigQuery koji obrađuje upit nad tisućama Googleovih poslužitelja. Razvojem svakog dogodio se pad troškova i smanjile su se prepreke za obradu velikih količina podataka (9).

Postupak analize podataka počinje prikupljanjem podataka nakon čega je potrebno je provesti samu analizu. Prikupljeni podaci mogu biti u neadekvatnom obliku te na taj način otežavati daljnji rad. Stoga je potrebno proučiti podatke s kojima se namjerava raditi i dovesti ih u oblik prikladan za rad, to jest provesti čišćenje podataka. Ukoliko u nekom retku ili stupcu nedostaju vrijednosti ili vrijednosti nisu u odgovarajućem obliku, potrebno ih je unijeti i promijeniti u ispravan oblik. Zato se koriste se različite strukture podataka ovisno o skupu podataka. Za jednodimenzionalan skup podataka koristi se struktura niz podataka. Ako su podaci prikazani u dvodimenzionalnoj tablici koristimo podatkovni okvir, ili strukturu panel za višedimenzionalne podatke. Nakon čišćenja nad skupom podataka izvode se potrebne operacije kao što su sortiranje, spajanje, preoblikovanje, filtriranje i agregiranje podataka. Kada se skup podataka dovede u najprikladniji oblik, podaci se mogu vizualizirati te prikazati u grafičkom obliku.

U sljedećem poglavlju obrađene su vrste analize podataka, deskriptivna, istraživačka i prediktivna analiza. Nakon toga u 3. Poglavlju opisane su, te objašnjene strukture podataka: niz podataka, podatkovni okvir i panel. Zatim, u 4. poglavlju primijenjene su operacije nad skupovima podataka: sortiranje, biranje, spajanje, preoblikovanje, filtriranje, grupiranje i agregiranje. U zadnjem poglavlju opisana su dva skupa podataka, provedena je analiza te su skupovi podataka prikazani u grafičkom obliku.

## **2.Vrste analize podataka**

### **2.1.Deskriptivna analiza podataka (descriptive data analysis)**

Deskriptivna analiza podataka uključuje izradu tablica sa značenjima, mjerama i unakrsnih tablica koje se koriste za ispitivanje različitih hipoteza. Te su hipoteze često o uočenim razlikama među podskupinama. Specijalizirane deskriptivne tehnike koriste se za mjerenja segregacije, diskriminacije i nejednakosti. Diskriminacija se često mjeri revizijskim studijima ili metodama dekompozicije. Za ovu analizu bitno je točno mjerenje razina vremena i prostora (10).

Tablica značenja po podskupinama može pokazati bitne razlike među podskupinama. Često se značenja razlikuju samo zbog slučajnih varijacija, te je potreban statistički zaključak da bi se utvrdilo mogu li te razlike proizaći iz slučajnosti.

Tablice poprečnog presjeka ili dvodimenzionalne tablice prikazuju proporcije jedinica s različitim vrijednostima za svaku od dvije varijable ili proporcije ćelije. Na primjer, možemo se pitati koliki udio stanovništva ima stečenu srednjoškolsku diplomu, a ujedno prima neku vrstu pomoći (hrana, novac), što označava presjek obrazovanja i primanja pomoći. Tako je moguće ispitati proporcije u svakoj obrazovnoj skupini koja prima pomoć. U ispitanim primjerima pokazalo se da što je razina obrazovanja viša, razina pomoći se smanjuje.

Deskriptivna analiza pomaže opisati, prikazati ili sažeti podatke na smislen način. Ne dopušta donošenje zaključaka izvan podataka koje smo analizirali ili donošenje zaključaka o proizvedenim hipotezama. Predstavljanje podataka na smislen način omogućuje jednostavniju vizualizaciju podataka.

### **2.2.Istraživačka analiza podataka (exploratory data analysis)**

Istraživačka analiza podataka važan je dio procesa analize podataka. Primjena ove analize odrediti će vrste drugih analiza koje podatkovni analitičar može koristiti za ispitivanje određenog skupa podataka. Prikadna je za kvalitativne i kvantitativne podatke korištene u viševarijantnim analizama u području društvenih znanosti. Naglasak je stavljen na korištenje vizualnog prikaza kako bi se otkrile informacije o podacima koji se ispituju. Osnova ove analize je pretraživanje s naglaskom na korištenju alternativnih tehnika za pristup istom skupu podataka, a pretežno se fokusira na otkrivanje novih značajki u podacima (11). Svakako, potrebno je očistiti podatke, što uključuje modeliranje, transformiranje i vizualizaciju podataka kako bi se omogućio što bolji uvid u skup podataka.

Osnovni koraci u istraživačkoj analizi podataka:

- Otkrivanje temeljne strukture skupa podataka,
- Deklariranje potrebnih varijabli,
- Otkrivanje nepravilnosti u podacima.

Kod istraživačke analize izrazito su bitna početna istraživanja podataka kako bi se otkrile nepravilnosti, testirale hipoteze i provjerile pretpostavke uz pomoć sažetih statistika i grafičkih prikaza. Zahtjeva dobro razumijevanje podataka i potrebno je imati što bolji uvid u podatke. Ne postoji definirana tehnika kojom će se provesti analiza, već ona ovisi o pojedincu koji je provodi, njegovom tumačenju i individualnim pogledom na skup podataka. Istraživačka analiza prvenstveno se bavi razumijevanjem podataka prije nego se krene u daljnju analizu.

### 2.3. Prediktivna analiza podataka (predictive data analysis)

Prediktivna analiza može procijeniti buduće rezultate na osnovu arhive prošlih podataka pomoću tehnika analize. Obuhvaća razne statističke tehnike od prediktivnog modeliranja, strojnog učenja do rudarenja podacima pomoću kojih se analiziraju arhivirani podaci kako bi se predvidjeli nepoznati događaji u budućnosti. Može predvidjeti buduće ishode sa značajnom preciznošću. Korištenjem prediktivne analize tvrtka može dobiti uvid u buduće ishode korištenjem prošlih i trenutnih podataka. Budući ishodi mogu biti pouzdano procijenjeni u danima, mjesecima ili godinama. Tako tvrtke koje raspolažu velikom količinom podataka, primjenom prediktivne analize mogu pratiti svoj razvoj i biti spremne na buduće događaje.

Kako bi se ostvarilo predviđanje budućih događaja, prediktivna analiza koristi algoritme klasifikacije, te algoritme za klasteriranje.

Prediktivna analiza može pomoći tvrtkama sa otkrivanjem postojećih, te stjecanju budućih klijenata. Za to se koriste se algoritmi klasifikacije. Na primjer, podaci o klijentu pojedine tvrtke mogu se razvrstati prema kategorijama mjesečnih primanja. Svaki pročitani zapis iz skupa podataka sadrži informaciju o ciljnoj varijabli - vrijednost koja se želi predvidjeti i prediktorske varijable – skup ulaznih varijabli. Prediktorske varijable mogu biti životna dob klijenta, spol ili zanimanje. Temeljem ovih kategorija i prediktorskih varijabli moguće je procijeniti mjesečne prihode klijenata koji još ne postoje u bazi tvrtke.

Algoritmi za klasteriranje omogućuju analizu uzoraka podataka i otkrivanje njihove međusobne ovisnosti. Algoritam nastoji cijeli skup podataka formirati u grupe, to jest klastere, tako da sadrže podatke koji su međusobno slični, a maksimalno se razlikuju od drugih grupa klastera. Zatim se oblikuju predikcije, temeljem kojih se može procijeniti vjerojatnost da se neki događaj ostvari. Podjela u grupe omogućuje primjenu tehnika rudarenja podataka nad grupiranim, to jest smanjenim skupovima podataka.

## 3.Strukture podataka

### 3.1.Niz podataka

Niz podataka je jednodimenzionalni niz koji može sadržavati podatke bilo koje vrste. Može sadržavati cijeli broj, string, float i druge. Oznake osi nazivaju se indeksi. Može se napraviti od ulaznih podataka poput niza, rječnika, skalarne vrijednosti ili konstante.

Niz podataka koristi četiri parametra i kreira se pomoću naredbe:

```
pandas.Series(data, index, dtype, copy)
```

Objašnjenje parametara:

- Data- Podaci mogu biti u formama kao što je ndarray, liste ili konstante,
- Index- Vrijednost indeksa mora biti jedinstvena, iste dužine kao i niz podataka. Ako ih ne zadamo zadat će se prema postavljenoj vrijednosti ovisno o dužini niza podataka,
- Dtype- Označava vrstu podatka, ako nije navedena procijenit će se prema vrsti unesenih podataka,
- Copy- Služi za kopiranje podataka, ako nije navedeno drugačije postavljen je na vrijednost „false“.

Kreiranje niza podataka, indeksi nisu zadani:

```
import pandas as pd
import numpy as np
data = np.array(['ponedjeljak', 'utorak', 'srijeda', 'četvrtak', 'petak', 'subota', 'nedjelja'])
podatkovni_niz = pd.Series(data)
print (podatkovni_niz)
```

```
0    ponedjeljak
1         utorak
2         srijeda
3         četvrtak
4         petak
5         subota
6         nedjelja
dtype: object
```

Slika 1.Niz podataka

Kreiranje niza podataka, indeksi su zadani:

```
import pandas as pd
import numpy as np
podaci = np.array(['ponedjeljak', 'utorak', 'srijeda', 'četvrtak', 'petak', 'subota', 'nedjelja'])
podatkovni_niz = pd.Series(podaci, index=[100, 101, 102, 103, 104, 105, 106])
print (podatkovni_niz)
```

```
100    ponedjeljak
101         utorak
102         srijeda
103     četvrtak
104         petak
105         subota
106     nedjelja
dtype: object
```

Slika 2. Niz podataka sa zadanim vrijednostima

Kreiranje niza podataka, korištenjem skalarne vrijednosti:

```
import pandas as pd
import numpy as np
podatkovni_niz = pd.Series(5, index=[0, 1, 2, 3])
print (podatkovni_niz)
```

```
0    5
1    5
2    5
3    5
dtype: int64
```

Slika 3. Niz podataka, skalarne vrijednosti

### 3.2. Podatkovni okvir

Podatkovni okvir je dvodimenzionalna podatkovna struktura. Podaci su poredani u tablicama u redove i stupce. Stupci mogu sadržavati podatke različitih vrsta. Veličina im je promjenjiva, a redovi i stupci su označeni osima. Između redova i stupaca mogu se izvoditi aritmetičke operacije.

Podatkovni okvir može se kreirati koristeći sljedeću naredbu:

```
pandas.DataFrame(data, index, columns, dtype, copy)
```

Objašnjenje parametara:

- Data- Podaci mogu biti u različitim oblicima, podatkovnim nizovima, mapama, popisima, navodima, mogu biti konstante ili još jedan podatkovni okvir,
- Index- Indeksi služe za oznake redaka. Ako nisu zadani, automatski se postavljaju ovisno o dužini niza podataka,
- Columns- Služi za oznake stupaca. Kao i kod indeksa, ako nisu zadane vrijednosti, postavljaju se na dužinu ovisno o broju unesenih podataka,
- Dtype- Označava vrstu podatka pojedinog stupca,
- Copy- Ova se naredba koristi za kopiranje podataka, a zadana vrijednost je „false“.

Kreiranje podatkovnog okvira:

```
import pandas as pd
podaci = ['ponedjeljak', 'utorak', 'srijeda', 'četvrtak', 'petak', 'subota', 'nedjelja']
dataframe = pd.DataFrame(podaci)
print (dataframe)
```

```
0
0 ponedjeljak
1   utorak
2   srijeda
3   četvrtak
4     petak
5   subota
6  nedjelja
```

Slika 4.Podatkovni okvir

Kreiranje podatkovnog okvira, zadani su nazivi stupaca:

```
import pandas as pd
podaci = {'Dan':['Ponedjeljak', 'Utorak', 'Srijeda', 'Četvrtak', 'Petak', 'Subota', 'Nedjelja'], 'Broj':[1,2,3,4,5,6,7]}
dataframe = pd.DataFrame(podaci)
print (dataframe)
```

```
   Dan  Broj
0 Ponedjeljak    1
1   Utorak    2
2   Srijeda    3
3   Četvrtak    4
4     Petak    5
5   Subota    6
6  Nedjelja    7
```

Slika 5.Podatkovni okvir sa zadanim nazivima stupaca i vrijednostima

Korištenje podatkovnog niza u podatkovnom okviru:

```
import pandas as pd

podaci = {'jedan' : pd.Series([1, 2, 3], index=['a', 'b', 'c']),
          'dva' : pd.Series([1, 2, 3, 4], index=['a', 'b', 'c', 'd'])}

dataframe = pd.DataFrame(podaci)
print (dataframe)
```

```
   jedan  dva
a    1.0    1
b    2.0    2
c    3.0    3
d     NaN    4
```

Slika 6.Podatkovni niz u podatkovnom okviru

### 3.3.Panel

Panel je trodimenzionalno polje koje sadrži 3 osi:

- items – os 0 – Svaka stavka odgovara podatkovnom okviru,
- major\_axis – os 1 – Indeksi koji označavaju redak svakog podatkovnog okvira,
- minor\_axis – os 2 – Označava stupce svakog podatkovnog okvira.

Panel se može kreirati koristeći sljedeću naredbu:

pandas.Panel(data, items, major\_axis, minor\_axis, dtype, copy)

Objašnjenje parametara:

- Data - Podaci mogu biti u različitim oblicima, podatkovnim nizovima, mapama, popisima, navodima, može biti rječnik ili još jedan podatkovni okvir,
- Items- Označava os 0,
- major\_axis- Označava os 1,
- minor\_axis- Označava os 2,
- dtype- Označava vrstu podatka pojedinog stupca,
- Copy- Ova se naredba koristi za kopiranje podataka, a zadana vrijednost je „false“.

Panel može biti kreiran na dva načina, koristeći višedimenzionalno polje ili rječnik podatkovnog okvira:

```
import pandas as pd
import numpy as np

data = {'Item1' : pd.DataFrame(np.random.randn(4, 3)),
        'Item2' : pd.DataFrame(np.random.randn(4, 2))}
p = pd.Panel(data)
print p
```

```
Dimensions: 2 (items) x 4 (major_axis) x 3 (minor_axis)
Items axis: Item1 to Item2
Major_axis axis: 0 to 3
Minor_axis axis: 0 to 2
```

Slika 7.Panel

## 4.Opis postupka analize podataka

Prvi korak u analizi podataka je prikupljanje podataka s kojima se planira raditi. Nakon toga podatke je potrebno proučiti te pripremiti za analizu. Podatke je potrebno očistiti od nepotrebnih ili pogrešnih vrijednosti. Kada smo sigurni da su podaci u pogodnom obliku, nad njima se primjenjuju potrebne operacije za organizaciju i upravljanje.

### 4.1.Prikupljanje podataka

Podaci se prikupljaju iz različitih izvora. Mogu se prikupljati iz životnog okruženja, poput prometnih kamera, satelita ili uređaja za simuliranje . Također, podaci se mogu prikupiti putem intervjua, mogu se preuzeti sa internetskih stranica ili putem čitanja određene dokumentacije.

Prvobitno prikupljene podatke potrebno je obraditi ili organizirati za analizu. To može uključivati oblikovanje podataka u redove i stupce u obliku tablice za daljnju analizu, poput proračunskih tablica ili upotrebu statističkog softvera.

### 4.2.Čišćenje podataka

Kroz skupove podataka potrebno je provesti čišćenje, to jest oblikovati podatke tako da se s njima može raditi. Podaci s kojima planiramo raditi mogu biti u obliku koji nije pogodan za daljnji rad. Mogu nedostajati pojedine vrijednosti, oblikovanje može biti nedosljedno, mogu sadržavati neispravne ili čak nepotrebne zapise.

Upotrebom Pythonovih biblioteka *Pandas* (12) i *NumPy* (13), omogućuje se čišćenje podataka koje uključuje:

- Ispuštanje nepotrebnih stupaca u DataFrame,
- Promjena indeksa podatkovnog okvira,
- Korištenje `.str ()` metode za čišćenje stupaca,
- Korištenje funkcije `DataFrame.applymap()` za čišćenje cjelokupnog skupa podataka,
- Preimenovanje stupaca u prepoznatljiviji skup imena,
- Preskakanje nepotrebnih redaka u datoteci,
- Stavljanje stupaca u DataFrame.

U odabranom skupu podataka želimo se usredotočiti samo na određene informacije, stoga stupce koji su nam nepotrebni možemo isključiti, te raditi analizu nad podacima koji su nam potrebni.

### 4.3.Primjena različitih operacija organiziranja i upravljanja nad skupovima podataka

#### 4.3.1.Sortiranje

U alatu Python *Pandas* postoje dvije mogućnosti sortiranja - po vrijednosti retka i stvarnoj vrijednosti stupca.

U datoteci su stavke sortirane prema stupcu „NAZIV ULICA“ , nazivi ulica sortirani su abecednim poretom.

Izgled datoteke:

|    | NAZIV ULICA                               | DUŽINA U m | STAZA               |
|----|-------------------------------------------|------------|---------------------|
| 0  | Antuna Ćoljana                            | 1435       | biciklistička staza |
| 1  | Avenija Dubrava                           | 600        | biciklistička staza |
| 2  | Avenija Dubrovnik                         | 7400       | biciklistička staza |
| 3  | AVH (od Antalove do Horvatove)            | 4150       | biciklistička staza |
| 4  | AVH (od sredine mosta do Antalove)        | 1300       | biciklistička staza |
| 5  | AVH (od sredine mosta do Av. Vukovar)     | 2400       | biciklistička staza |
| 6  | Bani ulica                                | 1800       | biciklistička staza |
| 7  | Baruna Filipovića                         | 3000       | biciklistička staza |
| 8  | Bistrička ulica                           | 1500       | biciklistička staza |
| 9  | Borovje                                   | 3000       | biciklistička staza |
| 10 | Bundek                                    | 2500       | biciklistička staza |
| 11 | Domovinski most                           | 3000       | biciklistička staza |
| 12 | Držićeva avenija                          | 5500       | biciklistička staza |
| 13 | Dugave                                    | 2900       | biciklistička staza |
| 14 | Gavranova                                 | 3250       | biciklistička staza |
| 15 | Gojka Ćurka                               | 3800       | biciklistička staza |
| 16 | Grada Vukovara(od Držićeve do Koledovine) | 5450       | biciklistička staza |
| 17 | Grada Vukovara(od Savske do Držićeve)     | 5360       | biciklistička staza |
| 18 | Gundulićeva                               | 1600       | biciklistička staza |
| 19 | Heinzelova                                | 5000       | biciklistička staza |
| 20 | Horvaćanska                               | 9500       | biciklistička staza |
| 21 | I.B.Maćuranić                             | 3000       | biciklistička staza |
| 22 | Islandska-Ukrajinska                      | 3600       | biciklistička staza |

Slika 8.Datoteka biciklisticestazei.csv<sup>1</sup>

Sortiranje po vrijednosti stupca korištenjem metode `sort_values()`. Sortira vrijednosti stupca i prihvaća argument „by“ koji će označiti vrijednost stupca prema kojemu će se izvršiti sortiranje. U sljedećem primjeru sortiranje se izvodi po duljini staze, od najduže prema najkraćoj:

<sup>1</sup> Preuzeto s <https://data.gov.hr/dataset/biciklisticke-staze>

```
import pandas as pd
import numpy as np
podaci=pd.read_csv("biciklisticestazei.csv", sep=';',encoding = 'ISO-8859-1' )
nesortirano=pd.DataFrame(podaci)
nesortirano.rename(columns={'DUŽINA U m':'DUŽINA U m'},inplace=True )
nesortirano['STAZA'] = nesortirano['STAZA'].map({'biciklistička staza':
'biciklistička staza', 'Biciklističke staze sportsko-rekreativnog karaktera -
Medvednica':'Biciklističke staze sportsko-rekreativnog karaktera - Medvednica'})
sortirano=nesortirano.sort_values(by='DUŽINA U m',ascending=False)
display(sortirano)
```

|    | NAZIV ULICA                                       | DUŽINA U m | STAZA                                             |
|----|---------------------------------------------------|------------|---------------------------------------------------|
| 35 | Nasip (Staza zdravlja) od Jankomira do Dom. Mosta | 40000      | biciklistička staza                               |
| 73 | Puntarska staza                                   | 29000      | Biciklističke staze sportsko-rekreativnog kara... |
| 74 | Imunska staza                                     | 26000      | Biciklističke staze sportsko-rekreativnog kara... |
| 64 | Zagrebačka avenija i Slavonska av.                | 23700      | biciklistička staza                               |
| 75 | Vrabečka staza                                    | 20000      | Biciklističke staze sportsko-rekreativnog kara... |
| 70 | Etno staza                                        | 17000      | Biciklističke staze sportsko-rekreativnog kara... |
| 71 | Zrcalna staza                                     | 15340      | Biciklističke staze sportsko-rekreativnog kara... |
| 72 | Staza bez daha                                    | 13000      | Biciklističke staze sportsko-rekreativnog kara... |
| 25 | Jarun                                             | 12500      | biciklistička staza                               |
| 20 | Horvaćanska                                       | 9500       | biciklistička staza                               |
| 77 | Spojna staza sa sedlo na sedlo                    | 9500       | Biciklističke staze sportsko-rekreativnog kara... |
| 76 | Vrhunska spojna staza                             | 8500       | Biciklističke staze sportsko-rekreativnog kara... |
| 44 | Radnička                                          | 8200       | biciklistička staza                               |
| 2  | Avenija Dubrovnik                                 | 7400       | biciklistička staza                               |
| 50 | Sopnica                                           | 6000       | biciklistička staza                               |
| 12 | Dr. Ćirilovec avenija                             | 5500       | biciklistička staza                               |

Slika 9.Sortiranje

Sljedeći primjer prikazuje sortiranje po stupcu, zamjena mjesta stupaca „DUŽINA U m“ i „NAZIV ULICA“:

```
import pandas as pd
import numpy as np
podaci=pd.read_csv("biciklistickestazei.csv", sep=';',encoding = 'ISO-8859-1' )
nesortirano=pd.DataFrame(podaci)
nesortirano.rename(columns={'DUŽINA U m':'DUŽINA U m'},inplace=True )
nesortirano['STAZA'] = nesortirano['STAZA'].map({'biciklistička staza':
'biciklistička staza', 'Biciklističke staze sportsko-rekreativnog karaktera -
Medvednica':'Biciklističke staze sportsko-rekreativnog karaktera - Medvednica'})
sortirano=nesortirano.sort_index(axis=1)
display(sortirano)
```

|    | DUŽINA U m | NAZIV ULICA                           | STAZA               |
|----|------------|---------------------------------------|---------------------|
| 0  | 1435       | Antuna Đoljana                        | biciklistička staza |
| 1  | 600        | Avenija Dubrava                       | biciklistička staza |
| 2  | 7400       | Avenija Dubrovnik                     | biciklistička staza |
| 3  | 4150       | AVH (od Antalove do Horvatove)        | biciklistička staza |
| 4  | 1300       | AVH (od sredine mosta do Antalove)    | biciklistička staza |
| 5  | 2400       | AVH (od sredine mosta do Av. Vukovar) | biciklistička staza |
| 6  | 1800       | Bani ulica                            | biciklistička staza |
| 7  | 3000       | Baruna Filipovića                     | biciklistička staza |
| 8  | 1500       | Bistrička ulica                       | biciklistička staza |
| 9  | 3000       | Borovje                               | biciklistička staza |
| 10 | 2500       | Bundek                                | biciklistička staza |
| 11 | 3000       | Domovinski most                       | biciklistička staza |
| 12 | 5500       | Držićeva avenija                      | biciklistička staza |
| 13 | 2900       | Dugave                                | biciklistička staza |
| 14 | 3250       | Gavranova                             | biciklistička staza |
| 15 | 3800       | Gojka Ruka                            | biciklistička staza |

Slika 10.Sortiranje stupaca

#### 4.3.2.Agregacija

Agregacija omogućuje pretvaranje vrijednosti skupa podataka u novu jedinstvenu vrijednost primjenom određene funkcije.

Izgled datoteke:

|    | animal   | uniq_id | water_need |
|----|----------|---------|------------|
| 0  | elephant | 1001    | 500        |
| 1  | elephant | 1002    | 600        |
| 2  | elephant | 1003    | 550        |
| 3  | tiger    | 1004    | 300        |
| 4  | tiger    | 1005    | 320        |
| 5  | tiger    | 1006    | 330        |
| 6  | tiger    | 1007    | 290        |
| 7  | tiger    | 1008    | 310        |
| 8  | zebra    | 1009    | 200        |
| 9  | zebra    | 1010    | 220        |
| 10 | zebra    | 1011    | 240        |
| 11 | zebra    | 1012    | 230        |
| 12 | zebra    | 1013    | 220        |
| 13 | zebra    | 1014    | 100        |
| 14 | zebra    | 1015    | 80         |
| 15 | lion     | 1016    | 420        |
| 16 | lion     | 1017    | 600        |
| 17 | lion     | 1018    | 500        |
| 18 | lion     | 1019    | 390        |
| 19 | kangaroo | 1020    | 410        |
| 20 | kangaroo | 1021    | 430        |
| 21 | kangaroo | 1022    | 410        |

Slika 11.Datoteka zoo.csv<sup>2</sup>

Primjena agregacije na stupcu „water\_need“ - sumirane su sve vrijednosti stupca, koje daju rezultat koliko je ukupno vode potrebno za sve životinje:

<sup>2</sup> Preuzeto s [http://www.hzzo.hr/wp-content/uploads/2018/04/web\\_opca\\_032018.xls?b32def](http://www.hzzo.hr/wp-content/uploads/2018/04/web_opca_032018.xls?b32def)

```
import pandas as pd
import numpy as np
podaci=pd.read_csv('zoo.csv', delimiter = ',')
df = pd.DataFrame(podaci)
r = df.rolling(window=3,min_periods=1)
print (r['water_need'].aggregate(np.sum))
```

```
0      500.0
1     1100.0
2     1650.0
3     1450.0
4     1170.0
5      950.0
6      940.0
7      930.0
8      800.0
9      730.0
10     660.0
11     690.0
12     690.0
13     550.0
14     400.0
15     600.0
16     1100.0
17     1520.0
18     1490.0
19     1300.0
20     1230.0
21     1250.0
Name: water_need, dtype: float64
```

Slika 12.Agregacija

### 4.3.3. Grupiranje

Svako grupiranje uključuje jednu od sljedećih operacija:

- Podjela objekta,
- Primjena funkcije,
- Kombiniranje rezultata.

Primjena funkcije uključuje neku od operacija kao što je agregacija, transformacija ili filtriranje.

Izgled datoteke:

| 1  | 2   | 3                       | 4           | 5         | 6           | 7          | 8      | 9                 | 10          | 11                | 12         | 13                              | 14                                     | 15 | 16 | 17 |
|----|-----|-------------------------|-------------|-----------|-------------|------------|--------|-------------------|-------------|-------------------|------------|---------------------------------|----------------------------------------|----|----|----|
| 1  | mbu | Naziv Ust               | Nadredje    | Vrsta Ust | Upisnik     | Adresa     | Mjesto | Telefon           | Faks        | e Posta           | URL        | Celnik                          |                                        |    |    |    |
| 2  | 13  | Edukacijsk              | Sveučilišt  | Fakultet  | Znanstven   | Borongajsk | ZAGREB | (+385 1) 24       | (+385 1) 24 | dekanat@ei        | www.erf    | dekanica                        | prof. dr. sc. Snježana Sekušak Galešev |    |    |    |
| 3  | 195 | Rudarsko-Sveučilišt     | Fakultet    | Znanstven | Pierottijev | ZAGREB     |        | (+385 1) 51       | (+385 1) 41 | dekanat@www.rgn   | dekan      | prof. dr. sc. Zoran Nakić       |                                        |    |    |    |
| 4  | 196 | Institut za antropolo   | Javni insti | Znanstven | Ljudevita   | ZAGREB     |        | (+385 1) 51       | (+385 1) 51 | ured@ina          | www.inar   | ravnatelj                       | dr. sc. Saša Missoni                   |    |    |    |
| 5  | 197 | Institut za arheologij  | Javni insti | Znanstven | Ljudevita   | ZAGREB     |        | (+385 1) 61       | (+385 1) 61 | ured.ravn         | www.iarh   | ravnatelj                       | dr. sc. Marko Dizdar                   |    |    |    |
| 6  | 198 | Klinička bolnica "Dol   | Ostale znc  | Znanstven | Avenija 38  | ZAGREB     |        | (+385 1) 21       | (+385 1) 21 | kbd@kbd.hr        |            | ravnatelj                       | prof.dr.sc. Milan Kujundžić            |    |    |    |
| 7  | 199 | Fakultet f Sveučilišt   | Fakultet    | Znanstven | Jordanovsk  | ZAGREB     |        | (+385 1) 21       | (+385 1) 21 | tajništvo         | (http://ww | dekan                           | prof. dr. sc. Ivan Koprek              |    |    |    |
| 8  | 201 | Energetski institut     | Znanstven   | Znanstven | Savska ce   | ZAGREB     |        | (+385 1) 61       | (+385 1) 61 | eihp@eih          | www.eihp   | ravnatelj                       | dr. sc. Goran Granić                   |    |    |    |
| 9  | 203 | Katolički b Sveučilišt  | Fakultet    | Znanstven | Vlaška 38   | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | kbf@theo          | www.kbf    | dekan                           | prof. dr. sc. Mario Cifrak             |    |    |    |
| 10 | 204 | MUP Policijska akad     | Visoka šk   | Znanstven | Avenija 61  | ZAGREB     |        | (+385 1) 21       | (+385 1) 21 | vps@fkz.h         | www.pa.r   | dekan                           | dr. sc. Krunoslav Borovec              |    |    |    |
| 11 | 205 | Hrvatski državni arhi   | Ostale znc  | Znanstven | Marulićev   | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | hda@arhiv.hr      |            | ravnatelj                       | dr. sc. Stjepan Čosić                  |    |    |    |
| 12 | 206 | Klinika za ortopediju   | Ostale znc  | Znanstven | Šetalište 1 | LOVRAN     |        | (+385 51) 1       | (+385 51) 1 | klinika.lov       | www.ortc   | ravnatelj                       | prof.dr.sc. Branko Šestan              |    |    |    |
| 13 | 245 | Fakultet z Sveučilišt   | Fakultet    | Znanstven | Ulica cara  | OSIJEK     |        | (+385 31) 1       | (+385 31) 1 | helpdesk          | (http://ww | dekan                           | izv. prof. dr. sc. Damir Matanović     |    |    |    |
| 14 | 258 | Sveučilište u Splitu    | Sveučilišt  | Znanstven | Poljička 4  | SPLIT      |        | (+385 21) 1       | (+385 21) 1 | rektorat.o        | www.unis   | rektor                          | prof. dr. sc. Dragan Ljutić            |    |    |    |
| 15 | 260 | Akademij Sveučilišt     | Umjetničk   | Znanstven | Ilica 85    | ZAGREB     |        | (+385 1) 31       | (+385 1) 31 | alu@alu.h         | www.alu    | dekan                           | izv. prof. art. Tomislav Buntak        |    |    |    |
| 16 | 264 | Akademij Sveučilišt     | Umjetničk   | Znanstven | Trg Repub   | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | dekanat@www.adu   | dekanica   | dr. sc. Franka Perković Gamulin |                                        |    |    |    |
| 17 | 269 | Sveučilište u Zadru     | Sveučilišt  | Znanstven | Ul. Mihovi  | ZADAR      |        | (+385 23) 1       | (+385 23) 1 | rektorat@www.uniz | rektorica  | prof. dr. sc. Dijana Vican      |                                        |    |    |    |
| 18 | 285 | Sveučilišt Sveučilišt   | Sveučilišt  | Znanstven | Cara Hadri  | OSIJEK     |        | (+385 31) 1       | (+385 31) 1 | info@biol         | www.biol   | pročelnica                      | doc. dr. sc. Ljiljana Krstin           |    |    |    |
| 19 | 286 | Sveučilišt Sveučilišt   | Sveučilišt  | Znanstven | Trg Ljudev  | OSIJEK     |        | (+385 31) 1       | (+385 31) 1 | ured@fizi         | www.fizo   | pročelnik                       | izv. prof. dr. sc. Vanja Radolić       |    |    |    |
| 20 | 288 | Opća županijska boln    | Ostale znc  | Znanstven | Osječka 11  | POŽEGA     |        | (+385 34) 1       | (+385 34) 1 | info@poz          | www.poz    | ravnatelj                       | prof.dr.sc. Željko Glavić              |    |    |    |
| 21 | 289 | PLIVA HRVATSKA d.o.o    | Ostale znc  | Znanstven | Prilaz bar  | ZAGREB     |        | (+385 1) 31       | (+385 1) 31 | tatjana.m         | www.pliv   | predsjednik                     | uprave Tihomir Orešković               |    |    |    |
| 22 | 290 | Fidelta d.o.o.          | Znanstven   | Znanstven | Prilaz bar  | ZAGREB     |        | (+385 1) 81       | (+385 1) 81 | fidelta@g         | www.fide   | predsjednik                     | uprave Dr. sc. Philip John Dudfield    |    |    |    |
| 23 | 291 | Sveučilišt Sveučilišt   | Sveučilišt  | Znanstven | Cara Hadri  | OSIJEK     |        | (+385 31) 1       | (+385 31) 1 | info@ken          | www.ken    | pročelnik                       | doc. dr. sc. Berislav Marković         |    |    |    |
| 24 | 292 | Institut za istraživanj | Ostale znc  | Znanstven | Jagodno 1   | VELIKA G   |        | (+385 1) 6168-522 |             | ires@ires         | www.ires   | ravnatelj                       | dr. sc. Zoran Pišl                     |    |    |    |
| 25 | 293 | Arheološki muzej u      | Ostale znc  | Znanstven | Trg Nikole  | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | amz@amz           | www.amz    | ravnateljica                    | dr. sc. Jacqueline Balen               |    |    |    |
| 26 | 55  | Ekonomsk Sveučilišt     | Fakultet    | Znanstven | Cvite Fisk  | SPLIT      |        | (+385 21) 1       | (+385 21) 1 | dekanat@www.efst  | dekanica   | prof. dr. sc. Maja Fredotović   |                                        |    |    |    |
| 27 | 58  | Prehramb Sveučilišt     | Fakultet    | Znanstven | Pierottijev | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | dekan@pl          | www.pbf    | dekan                           | prof. dr. sc. Damir Ježek              |    |    |    |
| 28 | 61  | INA-Industrija nafte    | Ostale znc  | Znanstven | Avenija 61  | ZAGREB     |        | (+385 1) 61       | (+385 1) 61 | ina-bespl         | http://ww  | direktor                        | Gabriel Račka                          |    |    |    |
| 29 | 62  | Medicinsk Sveučilišt    | Fakultet    | Znanstven | Braće Brar  | RIJEKA     |        | (+385 51) 1       | (+385 51) 1 | dekanat@www.mec   | rektor     | prof. dr. sc. Tomislav Rukavina |                                        |    |    |    |
| 30 | 65  | Stomatolc Sveučilišt    | Fakultet    | Znanstven | Gundulić    | ZAGREB     |        | (+385 1) 41       | (+385 1) 41 | dekanat@www.sfz   | dekanica   | prof. dr. sc. Zrinka Tarle      |                                        |    |    |    |
| 31 | 79  | Fakultet a Sveučilišt   | Fakultet    | Znanstven | Vladimira   | OSIJEK     |        | (+385 31) 1       | (+385 31) 1 | dekanat@www.fazc  | dekan      | prof. dr. sc. Krunoslav Zmaić   |                                        |    |    |    |
| 32 | 81  | Ekonomsk Sveučilišt     | Fakultet    | Znanstven | Ivana Filip | RIJEKA     |        | (+385 51) 1       | (+385 51) 1 | efri@efri         | www.efri   | dekan                           | prof. dr. sc. Alen Host                |    |    |    |
| 33 | 82  | Građevinsk Sveučilišt   | Fakultet    | Znanstven | Ulica fra A | ZAGREB     |        | (+385 01) 1       | (+385 01) 1 | ured_d            | www.prar   | dekan                           | prof. dr. sc. Stjepan Iakušić          |    |    |    |

Slika 13. Datoteka ustanove\_z.csv<sup>3</sup>

U sljedećem primjeru grupirane su vrijednosti po stupcu 'Naziv Ustanove'

<sup>3</sup> Preuzeto s <https://data.gov.hr/dataset/ustanove-iz-sustava-znanosti>

```
import pandas as pd
import numpy as np
podaci=pd.read_csv('ustanove_z.csv', sep=';')
df = pd.DataFrame(podaci)
display (df.groupby('Vrsta Ustanove').groups)

{'Bolnica': Int64Index([78, 149, 153, 154], dtype='int64'),
 'Državna ustanova': Int64Index([81], dtype='int64'),
 'Fakultet': Int64Index([ 0, 1, 5, 7, 11, 24, 25, 27, 28, 29, 30, 31, 32,
 33, 35, 38, 44, 46, 47, 50, 52, 59, 64, 65, 83, 84,
 86, 87, 88, 90, 100, 101, 102, 113, 115, 117, 118, 119, 120,
 121, 122, 123, 124, 125, 127, 128, 129, 132, 135, 145, 146, 155,
 158, 159, 164, 167, 168, 173, 174, 176, 181],
 dtype='int64'),
 'Javni institut': Int64Index([ 2, 3, 39, 40, 48, 62, 66, 69, 71, 79, 80, 104, 106,
 107, 108, 112, 139, 163, 166, 170, 175, 177, 178, 180, 182],
 dtype='int64'),
 'Knjižnica': Int64Index([53], dtype='int64'),
 'Ostale znanstvenoistraživačke pravne osobe': Int64Index([ 4, 9, 10, 18, 19, 22, 23, 26, 37, 42, 43, 45, 49,
 51, 54, 55, 56, 57, 60, 61, 63, 67, 68, 70, 72, 73,
 76, 82, 85, 103, 109, 110, 111, 114, 116, 126, 130, 131, 136,
 138, 141, 143, 144, 150, 151, 152, 156, 165, 169, 172, 179],
 dtype='int64'),
 'Privatni institut': Int64Index([74, 140], dtype='int64'),
 'Sveučilišni centar': Int64Index([58], dtype='int64'),
 'Sveučilišni odjel': Int64Index([16, 17, 21, 91, 93, 94, 147, 157], dtype='int64'),
 'Sveučilište': Int64Index([12, 15, 75, 77, 92, 99, 142, 148, 160, 161], dtype='int64'),
 'Umjetnička akademija': Int64Index([13, 14, 34, 171], dtype='int64'),
 'Ustanova od posebnog značaja za Republiku Hrvatsku': Int64Index([41, 105], dtype='int64'),
 'Veleučilište': Int64Index([36, 133, 134], dtype='int64'),
 'Visoka škola': Int64Index([8, 89, 95, 137], dtype='int64'),
 'Znanstveni institut': Int64Index([6, 20], dtype='int64'),
 'Znanstveno-nastavna jedinica': Int64Index([97], dtype='int64')}
```

Slika 14. Grupiranje

Korištenje metode `get_group()` omogućuje odabir određene grupe podataka.

U sljedećem primjeru iz skupa podataka izvršeno je grupiranje po vrsti ustanove. Zatim su iz grupe izdvojene samo ustanove kod kojih je zadovoljen uvjet da je vrsta ustanove bolnica.

```
import pandas as pd
import numpy as np
podaci=pd.read_csv('ustanove_z.csv', sep=';')
df = pd.DataFrame(podaci)
grupirano=df.groupby('Vrsta Ustanove')
display(grupirano.get_group('Bolnica'))
```

| mbu | Naziv Ustanove | Nadredjena Ustanova                               | Vrsta Ustanove | Upisnik | Adresa              | Mjesto                               | Telefon | Faks              | e Posta           | URL                                  | Celnik                          |                                              |
|-----|----------------|---------------------------------------------------|----------------|---------|---------------------|--------------------------------------|---------|-------------------|-------------------|--------------------------------------|---------------------------------|----------------------------------------------|
| 78  | 352            | Klinika za dječje bolesti                         | NaN            | Bolnica | Znanstvena ustanova | Klaićeva 16                          | ZAGREB  | 01/4600 100       | 01/4826 053       | kdb@kdb.hr                           | www.kdb.hr                      | sanacijski upravitelj dipl. oec. Osman Kadić |
| 149 | 337            | Specijalna bolnica za ortopediju i opću kirurg... | NaN            | Bolnica | Znanstvena ustanova | Ul. Dalmatinskih brigada bb, Matulji | MATULJI | (+385 51) 277-350 | (+385 51) 273-901 | NaN                                  | NaN                             | ravnatelj prof. dr. sc. Boris Nemec          |
| 153 | 348            | Specijalna bolnica Sv. Katarina                   | NaN            | Bolnica | Znanstvena ustanova | Bračak 8                             | ZABOK   | 049/204 888       | 049/204 887       | info@svkatarina.hr                   | www.svkatarina.hr               | ravnateljica mr. sc. Nevenka Kovač           |
| 154 | 349            | Thalassoterapija Opatija                          | NaN            | Bolnica | Znanstvena ustanova | Maršala Tita 188/1                   | OPATIJA | +385 51 202 600   | +385 51 271 424   | thalassoterapija-opatija@ri.t-com.hr | www.thalassoterapija-opatija.hr | ravnatelj dipl. iur. Emil Bratović           |

Slika 15. Grupiranje uz uvjet

#### 4.3.4. Filtriranje

Filtriranje koristimo kada iz skupa podataka želimo izdvojiti određene podatke koji nas zanimaju. Možemo izdvojiti pojedini stupac ili redak. Izdvajamo ih prema definiranim kriterijima, a kao rezultat dobijemo podskup skupa podataka koji želimo izdvojiti. Podaci koje dobijemo kao rezultat istog su tipa kao i ulazni podaci.

Podatke možemo filtrirati koristeći funkciju `filter()`.

```
DataFrame.filter(items=None, like=None, regex=None,axis=None)
```

Objašnjenje parametara:

- **Items**- sadržava oznake osi,
- **Like**- traži podatke za koje je zadani argument istina,
- **Regex** – traži podatke koji su zadani regularnim izrazom,
- **Axis**- os na kojoj želimo provesti filtriranje, za niz podataka koristi se „index“, a za podatkovni okvir „columns“.

Izgled datoteke:

|    | 1             | 2            | 3            | 4         | 5          | 6         | 7         | 8                      | 9                                                                                 | 10  | 11  | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 1 |
|----|---------------|--------------|--------------|-----------|------------|-----------|-----------|------------------------|-----------------------------------------------------------------------------------|-----|-----|----|----|----|----|----|----|----|---|
| 1  | ŽUPANIJA      | Naziv zavoda | Ulica i kbr. | Pošt. br. | Mjesto     | Tel.      | Fax.      | Ime i prez. Titula     | E-mail                                                                            | adr | web |    |    |    |    |    |    |    |   |
| 2  | Bjelovarski   | Zavod za     | Trg Eugen    | 43000     | Bjelovar   | (043) 242 | (043) 246 | Saša Križ dipl. ing.   | sasa.kriz@http://zpubbz.hr/                                                       |     |     |    |    |    |    |    |    |    |   |
| 3  | Brodsko-č     | Zavod za     | Trg pobje    | 35000     | Slavonski  | (035) 445 | (035) 445 | Matej Bad mag. ing.    | matej.bac@http://www.bpzzpu.hr/                                                   |     |     |    |    |    |    |    |    |    |   |
| 4  | Dubrovački    | Zavod za     | Petilovrije  | 20000     | Dubrovnik  | (020) 322 | (020) 321 | mr. sc. Ma dipl. ing.  | marina.or@http://www.zzpudnz.hr/                                                  |     |     |    |    |    |    |    |    |    |   |
| 5  | Grad Zagreb   | Zavod za     | Ulica Repu   | 10000     | Zagreb     | (01) 6101 | (01) 6101 | Ivica Rovi dipl. iur.  | zavod.pro@http://www.zzpugz.hr/                                                   |     |     |    |    |    |    |    |    |    |   |
| 6  | Istarska ž.   | Zavod za     | Riva 8       | 52100     | Pula       | (052) 351 | (052) 351 | Ingrid Palj dipl. ing. | ingrid.palj@http://www.zpuiz.hr/                                                  |     |     |    |    |    |    |    |    |    |   |
| 7  | Karlovački    | Zavod za     | Haulikova    | 47000     | Karlovac   | (047) 609 | (047) 609 | Mario Keč dipl. ing.   | ravnatelj@http://www.zavod-kazup.hr                                               |     |     |    |    |    |    |    |    |    |   |
| 8  | Koprivnički   | Zavod za     | Florijansk   | 48000     | Koprivnica | (048) 624 | (048) 624 | Zlatko Fili dipl. ing. | zlatko@pi@http://www.prostorno-kkz.hr                                             |     |     |    |    |    |    |    |    |    |   |
| 9  | Krapinski     | Zavod za     | Magistrats   | 49000     | Krapina    | (049) 382 | (049) 382 | Snježana i dipl. ing.  | snjezana.i@http://www.prostor-kkz.hr/                                             |     |     |    |    |    |    |    |    |    |   |
| 10 | Ličko-senjski | Zavod za     | Dr. Franje   | 53000     | Gospić     | (053) 588 | (053) 588 | Stipe Mud dipl. ing.   | stipe.mud-                                                                        |     |     |    |    |    |    |    |    |    |   |
| 11 | Međimurski    | Zavod za     | Ruđera Bc    | 40000     | Čakovec    | (040) 374 | (040) 374 | Mirjana Pi dipl. ing.  | mpintar@http://www.zavod.hr                                                       |     |     |    |    |    |    |    |    |    |   |
| 12 | Osječko-b     | Zavod za     | Ribarska 1   | 31000     | Osijek     | (031) 213 | (031) 213 | Oliver Gri dipl. ing.  | oliver.gri@http://www.prostorobz.hr                                               |     |     |    |    |    |    |    |    |    |   |
| 13 | Požeško-s     | Zavod za     | Županijski   | 34000     | Požega     | (034) 290 | (034) 290 | Mladenko dipl. ing.    | mladenko@http://www.zpu-psz.hr                                                    |     |     |    |    |    |    |    |    |    |   |
| 14 | Primorsko     | Zavod za     | Splitska 2   | 51000     | Rijeka     | (051) 351 | (051) 212 | Adam But. mag. ing.    | zavod@pg@http://www.zavod.pg.hr/                                                  |     |     |    |    |    |    |    |    |    |   |
| 15 | Sisačko-m     | Zavod za     | Trg Bana J   | 44000     | Sisak      | (044) 521 | (044) 523 | v.d. Margi dipl. ing.  | margita.m@http://www.zpusmz.hr                                                    |     |     |    |    |    |    |    |    |    |   |
| 16 | Splitsko-d    | Zavod za     | Domovins     | 21000     | Split      | (021) 400 | (021) 552 | Niko Mrčić dipl. ing.  | niko.mrcic@http://zpu-sdz.hr/                                                     |     |     |    |    |    |    |    |    |    |   |
| 17 | Šibensko-     | Zavod za     | Vladimira    | 22000     | Šibenik    | (022) 217 | (022) 217 | Damir Luč dipl. ing.   | damir.luci@http://zpu-skz.hr/                                                     |     |     |    |    |    |    |    |    |    |   |
| 18 | Varaždinski   | Zavod za     | Mali plac    | 42000     | Varaždin   | (042) 211 | (042) 211 | Davorin G dipl. ing.   | zavod@vz@http://www.varazdinska-zupanija.hr/zupanija/zavod-za-prostorno-uredjenje |     |     |    |    |    |    |    |    |    |   |
| 19 | Virovitički   | Zavod za     | Trg Ljudev   | 33000     | Virovitica | (033) 721 | (033) 721 | Jasna Barž dipl. ing.  | prostor-ol@http://juzpupvz.hr/                                                    |     |     |    |    |    |    |    |    |    |   |
| 20 | Vukovarski    | Zavod za     | Glagoljški   | 32100     | Vinkovci   | (032) 344 | (032) 344 | Pavica Fili dipl. iur. | zpuvsz@g@http://zpuvsz.hr/web/                                                    |     |     |    |    |    |    |    |    |    |   |
| 21 | Zadarska ž    | Zavod za     | Ulica braći  | 23000     | Zadar      | (023) 254 | (023) 251 | Stjepan G prof. geog   | zuz@zd.t-@http://www.zpu-zadup.hr/                                                |     |     |    |    |    |    |    |    |    |   |
| 22 | Zagrebački    | Zavod za     | Ulica grad   | 10000     | Zagreb     | (01) 6312 | (01) 6312 | Željka Kuč dipl. ing.  | z.z.kucinic@http://www.zpuz.hr/                                                   |     |     |    |    |    |    |    |    |    |   |
| 23 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |
| 24 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |
| 25 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |
| 26 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |
| 27 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |
| 28 |               |              |              |           |            |           |           |                        |                                                                                   |     |     |    |    |    |    |    |    |    |   |

Slika 16.Datoteka Zavodi\_adresar.csv<sup>4</sup>

Primjer filtriranja izdvajanjem stupaca 'Naziv zavoda', 'Ulica i kbr.' i 'Mjesto':

<sup>4</sup> Preuzeto s [https://mgipu.gov.hr/UserDocsImages//dokumenti/OpenData//Zavodi\\_adresar.csv](https://mgipu.gov.hr/UserDocsImages//dokumenti/OpenData//Zavodi_adresar.csv)

```

import pandas as pd
import numpy as np
podaci=pd.read_csv('Zavodi_adresar.csv', sep=';', encoding = 'ISO-8859-1' )
df = pd.DataFrame(podaci)
print (df.filter(["Naziv zavoda", "Ulica i kbr.", "Mjesto"]) )

```

|    | Naziv zavoda \                                    |
|----|---------------------------------------------------|
| 0  | Zavod za prostorno uređenje Bjelovarsko - bilo... |
| 1  | Zavod za prostorno uređenje Brodsko - posavske... |
| 2  | Zavod za prostorno uređenje Dubrovačko-neretva... |
| 3  | Zavod za prostorno uređenje Grada Zagreba         |
| 4  | Zavod za prostorno uređenje Istarske županije     |
| 5  | Zavod za prostorno uređenje Karlovačke županije   |
| 6  | Zavod za prostorno uređenje Koprivničko - križ... |
| 7  | Zavod za prostorno uređenje Krapinsko - zagors... |
| 8  | Zavod za prostorno uređenje Ličko-senjske župa... |
| 9  | Zavod za prostorno uređenje Međimurske županije   |
| 10 | Zavod za prostorno uređenje Osječko - baranjsk... |
| 11 | Zavod za prostorno uređenje Požeško - slavonsk... |
| 12 | Zavod za prostorno uređenje Primorsko - gorans... |
| 13 | Zavod za prostorno uređenje Sisačko - moslavač... |
| 14 | Zavod za prostorno uređenje Splitsko - dalmati... |
| 15 | Zavod za prostorno uređenje Šibensko - kninske... |
| 16 | Zavod za prostorno uređenje Varaždinske županije  |
| 17 | Zavod za prostorno uređenje Virovitičko - podr... |
| 18 | Zavod za prostorno uređenje Vukovarsko - srije... |
| 19 | Zavod za prostorno uređenje Zadarske županije     |
| 20 | Zavod za prostorno uređenje Zagrebačke županije   |

|    | Ulica i kbr.                | Mjesto         |
|----|-----------------------------|----------------|
| 0  | Trg Eugena Kvaternika 13    | Bjelovar       |
| 1  | Trg pobjede 26a             | Slavonski Brod |
| 2  | Petilovrijenci 2            | Dubrovnik      |
| 3  | Ulica Republike Austrije 18 | Zagreb         |
| 4  | Riva 8                      | Pula           |
| 5  | Haulikova 1                 | Karlovac       |
| 6  | Florijanski trg 4/1         | Koprivnica     |
| 7  | Magistratska 1              | Krapina        |
| 8  | Dr. Franje Tuđmana 4        | Gospić         |
| 9  | Ružera Boškovića 2          | Čakovec        |
| 10 | Ribarska 1/II               | Osijek         |
| 11 | Županijska 7                | Požeška        |
| 12 | Splitska 2/II               | Rijeka         |
| 13 | Trg Rana Jocića Jelačića 6  | Sisak          |

Slika 17.Filtriranje

#### 4.3.5.Transformiranje

Transformiranje je operacija koja se koristi u kombinaciji sa grupiranjem. Za razliku od agregacije koja vraća izmijenjen skup podataka, transformacija može vratiti transformirane podatke za daljnje kombiniranje. Izlazni skup podataka je istog oblika kao i ulazni.

Izgled datoteke:

|    | 1       | 2             | 3     | 4        | 5        | 6          | 7         |
|----|---------|---------------|-------|----------|----------|------------|-----------|
| 1  | account | name          | order | sku      | quantity | unit price | ext price |
| 2  | 383080  | Will LLC      | 10001 | B1-20000 | 7        | 33,69      | 235,83    |
| 3  | 383080  | Will LLC      | 10001 | S1-27722 | 11       | 21,12      | 232,32    |
| 4  | 383080  | Will LLC      | 10001 | B1-86481 | 3        | 35,99      | 107,97    |
| 5  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 48       | 55,82      | 2679,36   |
| 6  | 412290  | Jerde-Hilpert | 10005 | S1-82801 | 21       | 13,62      | 286,02    |
| 7  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 9        | 92,55      | 832,95    |
| 8  | 412290  | Jerde-Hilpert | 10005 | S1-47412 | 44       | 78,91      | 3472,04   |
| 9  | 412290  | Jerde-Hilpert | 10005 | S1-27722 | 36       | 25,42      | 915,12    |
| 10 | 218895  | Kulas Inc     | 10006 | S1-27722 | 32       | 95,66      | 3061,12   |
| 11 | 218895  | Kulas Inc     | 10006 | B1-33087 | 23       | 22,55      | 518,65    |
| 12 | 218895  | Kulas Inc     | 10006 | B1-33364 | 3        | 72,3       | 216,9     |
| 13 | 218895  | Kulas Inc     | 10006 | B1-20000 | -1       | 72,18      | -72,18    |
| 14 |         |               |       |          |          |            |           |

Slika 18.Datoteka sales\_transactions.xlsx<sup>5</sup>

U nastavku je prikazan primjer transformiranja s korištenjem grupiranja. Grupiranje je izvršeno po broju narudžbe pojedinog kupca, ukupnim zbrojem svih njegovih narudžbi.

<sup>5</sup> Preuzeto s [https://pbpython.com/pandas\\_transform.html](https://pbpython.com/pandas_transform.html)

```
import pandas as pd
df = pd.read_excel("sales_transactions.xlsx")
df.groupby('order')['ext price'].sum()
```

```
order
10001    576.12
10005    8185.49
10006    3724.49
Name: ext price, dtype: float64
```

```
import pandas as pd
df = pd.read_excel("sales_transactions.xlsx")
print(df)
df.groupby('order')['ext price'].sum()
```

|    | account | name          | order | sku      | quantity | unit price | ext price |
|----|---------|---------------|-------|----------|----------|------------|-----------|
| 0  | 383080  | Will LLC      | 10001 | B1-20000 | 7        | 33.69      | 235.83    |
| 1  | 383080  | Will LLC      | 10001 | S1-27722 | 11       | 21.12      | 232.32    |
| 2  | 383080  | Will LLC      | 10001 | B1-86481 | 3        | 35.99      | 107.97    |
| 3  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 48       | 55.82      | 2679.36   |
| 4  | 412290  | Jerde-Hilpert | 10005 | S1-82801 | 21       | 13.62      | 286.02    |
| 5  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 9        | 92.55      | 832.95    |
| 6  | 412290  | Jerde-Hilpert | 10005 | S1-47412 | 44       | 78.91      | 3472.04   |
| 7  | 412290  | Jerde-Hilpert | 10005 | S1-27722 | 36       | 25.42      | 915.12    |
| 8  | 218895  | Kulas Inc     | 10006 | S1-27722 | 32       | 95.66      | 3061.12   |
| 9  | 218895  | Kulas Inc     | 10006 | B1-33087 | 23       | 22.55      | 518.65    |
| 10 | 218895  | Kulas Inc     | 10006 | B1-33364 | 3        | 72.30      | 216.90    |
| 11 | 218895  | Kulas Inc     | 10006 | B1-20000 | -1       | 72.18      | -72.18    |

```
order
10001    576.12
10005    8185.49
10006    3724.49
Name: ext price, dtype: float64
```

Slika 19. Grupiranje i korištenje metode sum()

Operacijom transform stvara se novi stupac „Order\_Total“ gdje je svakom retku pridružen ukupan zbroj narudžbi pojedinog kupca prema grupiranim narudžbama i njihovom zbroju. Dodaje se stupac „Percent\_of\_Order“ koji sadrži prosjek narudžbe, gdje se iznos pojedine narudžbe dijeli sa ukupnim iznosom u stupcu „Order\_Total“ koji je prethodno kreiran.

```

import pandas as pd
df = pd.read_excel("sales_transactions.xlsx")
df["Order_Total"] = df.groupby('order')['ext price'].transform('sum')
df["Percent_of_Order"] = df["ext price"] / df["Order_Total"]
print(df)

```

|    | account | name          | order | sku      | quantity | unit price | ext price \ |
|----|---------|---------------|-------|----------|----------|------------|-------------|
| 0  | 383080  | Will LLC      | 10001 | B1-20000 | 7        | 33.69      | 235.83      |
| 1  | 383080  | Will LLC      | 10001 | S1-27722 | 11       | 21.12      | 232.32      |
| 2  | 383080  | Will LLC      | 10001 | B1-86481 | 3        | 35.99      | 107.97      |
| 3  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 48       | 55.82      | 2679.36     |
| 4  | 412290  | Jerde-Hilpert | 10005 | S1-82801 | 21       | 13.62      | 286.02      |
| 5  | 412290  | Jerde-Hilpert | 10005 | S1-06532 | 9        | 92.55      | 832.95      |
| 6  | 412290  | Jerde-Hilpert | 10005 | S1-47412 | 44       | 78.91      | 3472.04     |
| 7  | 412290  | Jerde-Hilpert | 10005 | S1-27722 | 36       | 25.42      | 915.12      |
| 8  | 218895  | Kulas Inc     | 10006 | S1-27722 | 32       | 95.66      | 3061.12     |
| 9  | 218895  | Kulas Inc     | 10006 | B1-33087 | 23       | 22.55      | 518.65      |
| 10 | 218895  | Kulas Inc     | 10006 | B1-33364 | 3        | 72.30      | 216.90      |
| 11 | 218895  | Kulas Inc     | 10006 | B1-20000 | -1       | 72.18      | -72.18      |

|    | Order_Total | Percent_of_Order |
|----|-------------|------------------|
| 0  | 576.12      | 0.409342         |
| 1  | 576.12      | 0.403249         |
| 2  | 576.12      | 0.187409         |
| 3  | 8185.49     | 0.327330         |
| 4  | 8185.49     | 0.034942         |
| 5  | 8185.49     | 0.101759         |
| 6  | 8185.49     | 0.424170         |
| 7  | 8185.49     | 0.111798         |
| 8  | 3724.49     | 0.821890         |
| 9  | 3724.49     | 0.139254         |
| 10 | 3724.49     | 0.058236         |
| 11 | 3724.49     | -0.019380        |

Slika 20.Transformiranje

#### 4.3.6.Spajanje

Spajanje omogućuje kombiniranje podataka različitih podatkovnih okvira primjenom metode `merge()`

`pd.merge(left, right, how='inner', on=None, left_on=None, right_on=None, left_index=False, right_index=False, sort=True)`

Objašnjenje parametara:

- Left- podatkovni okvir;
- Right- podatkovni okvir;
- How- može se koristiti jedno od:
  - Left- koristi ključeve lijevog objekta,
  - Right- koristi ključeve desnog objekta,
  - Outer- koristi uniju ključeva,
  - Inner- koristi presjek ključeva;
- On- naziv stupca po kojem će se izvršiti spajanje;
- Left\_on- stupci sa lijeve strane koji će se koristiti kao ključ. Može se zadati ime stupca ili niz dužine jednake podatkovnom okviru;
- Right\_on- stupci s desne strane koji će se koristiti kao ključ;
- Left\_index- ako je 'True', koristit će se indeksi retka od lijevog podatkovnog okvira kao ključevi;
- Right\_index- ako je 'True', koristit će se indeksi retka od desnog podatkovnog okvira kao ključevi;
- Sort- omogućuje sortiranje podatkovnog okvira prema pridruženim ključevima u leksikografskom poretku.

Primjer spajanja dva različita podatkovna okvira prema nazivu retka:

```
import pandas as pd
left = pd.DataFrame({
    'id':[1,2,3,4,5],
    'Name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],
    'subject_id':['sub1','sub2','sub4','sub6','sub5']})
right = pd.DataFrame({
    'id':[1,2,3,4,5],
    'Name': ['Billy', 'Brian', 'Bran', 'Bryce', 'Betty'],
    'subject_id':['sub2','sub4','sub3','sub6','sub5']})
print (pd.merge(left,right,on='id'))
```

|   | id | Name_x | subject_id_x | Name_y | subject_id_y |
|---|----|--------|--------------|--------|--------------|
| 0 | 1  | Alex   | sub1         | Billy  | sub2         |
| 1 | 2  | Amy    | sub2         | Brian  | sub4         |
| 2 | 3  | Allen  | sub4         | Bran   | sub3         |
| 3 | 4  | Alice  | sub6         | Bryce  | sub6         |
| 4 | 5  | Ayoung | sub5         | Betty  | sub5         |

Slika 21.Spajanje prema nazivu

Primjer spajanja dva različita podatkovna okvira prema određenom nazivu stupca, spajaju se stupci lijevog podatkovnog okvira u kojem su indeksi jednaki indeksima desnog podatkovnog okvira:

```
import pandas as pd
left = pd.DataFrame({
    'id':[1,2,3,4,5],
    'Name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],
    'subject_id':['sub1','sub2','sub4','sub6','sub5']})
right = pd.DataFrame({
    'id':[1,2,3,4,5],
    'Name': ['Billy', 'Brian', 'Bran', 'Bryce', 'Betty'],
    'subject_id':['sub2','sub4','sub3','sub6','sub5']})
print (pd.merge(left,right,on=['id','subject_id']))
```

|   | id | Name_x | subject_id | Name_y |
|---|----|--------|------------|--------|
| 0 | 4  | Alice  | sub6       | Bryce  |
| 1 | 5  | Ayoung | sub5       | Betty  |

Slika 22.Spajanje prema indeksima

## **5.Primjena alata Python Pandas za analizu podataka na odabranim skupovima podataka**

### **5.1.Opis skupova podataka**

Odabrani skup podataka prvog primjera je Excel datoteka 'opca\_medicina.xls'. Sastoji se od 14 stupaca i 2339 redaka. Sastoji se od popisa medicinskih ustanova. Svaki redak predstavlja jednu ustanovu, grad gdje se ustanova nalazi, šifru, naziv djelatnosti, podatke o doktoru, status ordinacije, broj osiguranja i podatke o adresi ordinacije.

Odabrani skup podataka drugog primjera je datoteka 'vrticio.csv'. Sastoji se od 9 stupaca i 307 redaka. Prikazuje popis dječjih vrtića na području grada Zagreba. Svaki redak prikazuje tip vrtića, njegov naziv, adresu, vrstu vrtića, program, četvrt u kojoj se nalazi, napomenu, te koordinatu.

### **5.2.Opis Python Pandas alata**

Python pandas<sup>6</sup> je biblioteka otvorenog koda stvorena za programski jezik Python. Omogućuje jednostavno korištenje, a služi za analizu i vizualizaciju podataka. Nudi rad sa podatkovnim strukturama i operacijama za manipuliranje numeričkim tablicama i vremenskim nizovima. Alat se koristi u širokom rasponu, uključujući akademske i komercijalne domene, financije, ekonomiju, statistiku i analitiku. Naziv Pandas izveden je iz „panel data“ ekonometrijskog pojma za skupove podataka koji uključuju promatranja kroz više vremenskih razdoblja (14).

---

<sup>6</sup> Preuzeto s <https://pandas.pydata.org/getpandas.html>

## 5.3. Primjena i rezultati

### 5.3.1. Primjer 1

Izgled datoteke:

| 1  | 2                 | 3         | 4                                                   | 5             | 6       | 7                | 8                                   | 9          | 10                | 11        | 12             | 13                      | 14                                        |
|----|-------------------|-----------|-----------------------------------------------------|---------------|---------|------------------|-------------------------------------|------------|-------------------|-----------|----------------|-------------------------|-------------------------------------------|
| 1  | DZ / Koncesionari |           |                                                     | DOKTOR        |         |                  | Status ordinacije (da, koncesionar) | Broj osig. | ADRESA ORDINACIJE |           |                |                         |                                           |
|    | Naziv PU          | Šifra     | Naziv                                               | Naziv djelat. | Šifra   | Prezime          | Ime                                 |            | HPT broj          | HPT Naziv | Ulica          | Br.                     | Grad/općina/ gradska četvrt Grada Zagreba |
| 2  |                   |           |                                                     |               |         |                  |                                     |            |                   |           |                |                         |                                           |
| 3  | PS Bjelovar       | 900000201 | O.DZ Bjel.-Bilog.zup. lok. Čazma-om.                | opca          | 0159859 | KOLAR            | ANDREA                              | DZ         | 1162              | 43240     | ČAZMA          | KRALJA TOMISLAVA        | 16 ČAZMA                                  |
| 4  | PS Bjelovar       | 900019204 | O.DZ Bjel.-Bilog.zup. lok. Čazma-om.                | opca          | 7950829 | SIROVATKA        | ANTONIA                             | DZ         | 1295              | 43240     | ČAZMA          | KRALJA TOMISLAVA        | 16 ČAZMA                                  |
| 5  | PS Bjelovar       | 900014040 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0204765 | KOŽUL            | ANTONIJA                            | DZ         | 1679              | 43000     | BJELOVAR       | SLAVONSKA CESTA         | 17 BJELOVAR                               |
| 6  | PS Bjelovar       | 900000031 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0148717 | ŠURJAK           | BELITA                              | DZ         | 1388              | 43000     | BJELOVAR       | SLAVONSKA CESTA         | BB BJELOVAR                               |
| 7  | PS Bjelovar       | 900009888 | O.DZ Bjel.-Bilog.zup. lok. Veliki Zdenci-om.        | opca          | 0189227 | HIPKIC           | BISERKA                             | DZ         | 588               | 43290     | GRUBIŠNO POLJE | VELIKI ZDENCI, TRG KRAJ | BB GRUBIŠNO POLJE                         |
| 8  | PS Bjelovar       | 852185219 | P.O. Ust. dr. Bilanka - opća med.                   | opca          | 0094501 | ULRIČIĆ          | BLANKA                              | konces     | 1458              | 43500     | DARUVAR        | DEŽANOVAC, DEŽANOV      | 1 DEŽANOVAC                               |
| 9  | PS Bjelovar       | 900000236 | O.DZ Bjel.-Bilog.zup. lok. Čazma-om.                | opca          | 0166075 | NIKOLIĆ          | BOJANA                              | DZ         | 1159              | 43240     | ČAZMA          | KRALJA TOMISLAVA        | 16 ČAZMA                                  |
| 10 | PS Bjelovar       | 489548954 | P.O. Palatinuš dr. Božica - opća med.               | opca          | 0065986 | PALATINUŠ        | BOŽICA                              | konces     | 1584              | 43500     | DARUVAR        | PETRA PRERADOVIĆA       | BB DARUVAR                                |
| 11 | PS Bjelovar       | 876287623 | P.O. Heged dr. Branko - opća med.                   | opca          | 0030490 | HEGED            | BRANKO                              | konces     | 1668              | 43000     | BJELOVAR       | KAPELA, ULICA FAZIN     | 4 KAPELA                                  |
| 12 | PS Bjelovar       | 900012269 | O.DZ Bjel.-Bilog.zup. lok. Garešnica-om.            | opca          | 0032557 | ILIĆ             | ĐURĐICA                             | DZ         | 1704              | 43280     | GAREŠNICA      | VLADIMIRA NAZORA        | 18 GAREŠNICA                              |
| 13 | PS Bjelovar       | 459245929 | P.O. Fawzi dr. Samara - opća med.                   | opca          | 0141798 | SAMARA           | FAWZI                               | konces     | 1442              | 43000     | BJELOVAR       | IVANSKA, PETRA PRERA    | BB IVANSKA                                |
| 14 | PS Bjelovar       | 815881584 | P.O. Knežević-Miličić dr. Gordana - opća med.       | opca          | 0041424 | KNEŽEVIĆ-MILIČIĆ | GORDANA                             | konces     | 1479              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 15 | PS Bjelovar       | 485848589 | P.O. Emić dr. Gordana - opća med.                   | opca          | 0022047 | EMIĆ             | GORDANA                             | konces     | 1324              | 43240     | ČAZMA          | ŠTEFANJE, ŠTEFANJE      | 1 ŠTEFANJE                                |
| 16 | PS Bjelovar       | 900012250 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-opće med.       | opca          | 7992785 | KUKAL GJERGJAJ   | IVA                                 | DZ         | 1584              | 43000     | BJELOVAR       | SEVERIN, SEVERIN        | BB SEVERIN                                |
| 17 | PS Bjelovar       | 700870083 | P.O. Aržek dr. Ivan - opća med.                     | opca          | 0001605 | ARŽEK            | IVAN                                | konces     | 1629              | 43280     | GAREŠNICA      | VELIKA TRNOVITICA, VE   | BB VELIKA TRNOVITICA                      |
| 18 | PS Bjelovar       | 688068804 | P.O. Ladović-Vučinić dr. Jadranka - opća med.       | opca          | 0122793 | LADOVIĆ-VUČINIĆ  | JADRANKA                            | konces     | 1725              | 43000     | BJELOVAR       | IVANSKA, PETRA PRERA    | BB IVANSKA                                |
| 19 | PS Bjelovar       | 532133214 | P.O. Šveda-Breškic dr. Jasna - opća med.            | opca          | 0090506 | ŠVEDA-BREŠKIC    | JASNA                               | konces     | 1996              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 20 | PS Bjelovar       | 900023538 | O. u ust. za zdrav. sl. FBENIKS - om.               | opca          | 0131334 | SLOVAČEK-ČESARE  | JASNA                               | konces     | 2108              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 23 BJELOVAR                               |
| 21 | PS Bjelovar       | 880988096 | P.O. Šigir-Lovrić dr. Jasna - opća med.             | opca          | 0087025 | ŠIGIR-LOVRIĆ     | JASNA                               | konces     | 1408              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 22 | PS Bjelovar       | 875287522 | P.O. Šigir-Majcan dr. Jelena - opća med.            | opca          | 0087033 | ŠIGIR-MAJCAN     | JELENA                              | konces     | 2072              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 23 | PS Bjelovar       | 100910092 | P.O. Lončar dr. Josip - opća med.                   | opca          | 0050482 | LONČAR           | JOSIP                               | konces     | 1877              | 43000     | BJELOVAR       | SRIJEMSKA ULICA         | 1 BJELOVAR                                |
| 24 | PS Bjelovar       | 900000040 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0053970 | HUNJET           | JUDITA                              | DZ         | 1599              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 25 | PS Bjelovar       | 172517257 | P.O. Kepčija dr. Lovorka - opća med.                | opca          | 0040100 | KEPČIJA          | LOVORKA                             | konces     | 1799              | 43500     | DARUVAR        | PETRA PRERADOVIĆA       | 17 DARUVAR                                |
| 26 | PS Bjelovar       | 900026286 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0105368 | VIDOVIĆ          | LILIJANA                            | DZ         | 1331              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 27 | PS Bjelovar       | 900000210 | O.DZ Bjel.-Bilog.zup. lok. Čazma-om.                | opca          | 0139149 | BERMANOVIĆ       | LILIJANA                            | DZ         | 1001              | 43280     | GAREŠNICA      | VLADIMIRA NAZORA        | 18 GAREŠNICA                              |
| 28 | PS Bjelovar       | 256725675 | P.O. Runjić dr. Ljiljana - opća med.                | opca          | 0078972 | RUNJIĆ           | LILIJANA                            | konces     | 1942              | 43000     | BJELOVAR       | ANTUN MIHANOVIĆ         | 8 BJELOVAR                                |
| 29 | PS Bjelovar       | 332733270 | P.O. Pleskati dr. Ljiljana - opća med.              | opca          | 0071528 | PLESKATI         | LILIJANA                            | konces     | 1570              | 43000     | BJELOVAR       | UL. TOMAŠE G. MASARY    | 9 BJELOVAR                                |
| 30 | PS Bjelovar       | 852485247 | P.O. Zunić dr. Ljiljana - opća med.                 | opca          | 0122530 | ZUNIĆ            | LILIJANA                            | konces     | 1570              | 43500     | DARUVAR        | SIRAČ, STEJANA RADIĆ    | 120 SIRAČ                                 |
| 31 | PS Bjelovar       | 900000090 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0162388 | MATIJAŠEVIĆ      | MAJA                                | DZ         | 691               | 43000     | BJELOVAR       | SANDROVAC, ULICA BRE    | BB SANDROVAC                              |
| 32 | PS Bjelovar       | 852385234 | P.O. Kalitrođa-Horvat dr. Marija - opća med.        | opca          | 0038355 | KALITROĐA-HORVA  | MARIJA                              | konces     | 1981              | 43500     | DARUVAR        | PETRA PRERADOVIĆA       | BB DARUVAR                                |
| 33 | PS Bjelovar       | 900023406 | O.DZ Bjel.-Bilog.zup. lok. Rovinje-om.              | opca          | 0186058 | BREZNIK          | MARTINA                             | DZ         | 1342              | 43000     | BJELOVAR       | ROVINJE, TRG HRVATSK    | 11 ROVINJE                                |
| 34 | PS Bjelovar       | 900019778 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 7888848 | ČUKMAN           | MARTINA                             | DZ         | 1525              | 43000     | BJELOVAR       | ROVINJE, TRG HRVATSK    | 11 ROVINJE                                |
| 35 | PS Bjelovar       | 900018844 | O.DZ Bjel.-Bilog.zup. lok. Nova Rača-om.            | opca          | 0185671 | BURIĆ            | MILAN                               | DZ         | 1207              | 43270     | VELIKI GRĐEVAČ | NOVA RAČA, TRG STJEP    | BB NOVA RAČA                              |
| 36 | PS Bjelovar       | 880888083 | P.O. Bogdanović-Prolić dr. Milena - opća med.       | opca          | 0007875 | BOGDANOVIĆ-PROI  | MILENA                              | konces     | 1028              | 43000     | BJELOVAR       | ULICA VELIKE SREDICE    | 11 BJELOVAR                               |
| 37 | PS Bjelovar       | 172417244 | P.O. Danjek-Radić dr. Milena - opća med.            | opca          | 0017329 | DANJEK-RADIĆ     | MILENA                              | konces     | 1646              | 43500     | DARUVAR        | STJEPANA RADIĆA         | 13 DARUVAR                                |
| 38 | PS Bjelovar       | 493049304 | P.O. Begić-Konljenović dr. Milka - opća med.        | opca          | 0005207 | BEGIĆ-KOMLIJENOV | MILKA                               | konces     | 1990              | 43000     | BJELOVAR       | UL. TOMAŠE G. MASARY    | 9 BJELOVAR                                |
| 39 | PS Bjelovar       | 798479841 | P.O. Tuček dr. Mira - opća med.                     | opca          | 0104965 | TUČEK            | MIRA                                | konces     | 1601              | 43290     | GRUBIŠNO POLJE | BRACE RADIĆA            | 1 GRUBIŠNO POLJE                          |
| 40 | PS Bjelovar       | 493149317 | P.O. Grobotek dr. Mirjana - opća med.               | opca          | 0028576 | GROBOTEK         | MIRJANA                             | konces     | 1814              | 43000     | BJELOVAR       | UL. TOMAŠE G. MASARY    | 9 BJELOVAR                                |
| 41 | PS Bjelovar       | 493249320 | P.O. Drobac dr. Mirjana - opća med.                 | opca          | 0020087 | DROBAC           | MIRJANA                             | konces     | 2102              | 43000     | BJELOVAR       | UL. TOMAŠE G. MASARY    | 9 BJELOVAR                                |
| 42 | PS Bjelovar       | 463946391 | P.O. Petrović dr. Mirjana - opća med.               | opca          | 0072494 | PETROVIĆ         | MIRJANA                             | konces     | 1326              | 43280     | GAREŠNICA      | TRG HRVATSKIH BRANI     | 11 GAREŠNICA                              |
| 43 | PS Bjelovar       | 770077005 | P.O. Husein dr. Nataša - opća med.                  | opca          | 0106505 | HUSEIN           | MUSTAFA                             | konces     | 2133              | 43290     | GRUBIŠNO POLJE | BRACE RADIĆA            | 1 GRUBIŠNO POLJE                          |
| 44 | PS Bjelovar       | 900000066 | O.DZ Bjel.-Bilog.zup. lok. Bjelovar-om.             | opca          | 0142301 | OROS-DELIĆ       | NATAŠA                              | DZ         | 960               | 43000     | BJELOVAR       | VELIKO TROJSTVO, UL. B  | 71 VELIKO TROJSTVO                        |
| 45 | PS Bjelovar       | 804580456 | P.O. Lončar dr. Nataša - opća med.                  | opca          | 0050525 | LONČAR           | NATAŠA                              | konces     | 1412              | 43500     | DARUVAR        | KONČANICA, KONČANIC     | BB KONČANICA                              |
| 46 | PS Bjelovar       | 900009551 | O.DZ Bjel.-Bilog.zup. lok. Garešnički Brestovac-om. | opca          | 0150037 | DŽONLIĆ          | NEVENKA                             | DZ         | 1241              | 43280     | GAREŠNICA      | GAREŠNIČKI BRESTOVAC    | 33A GAREŠNICA                             |

Slika 23. Datoteka opca\_medicina.xls<sup>7</sup>

1. Učitavanje datoteke 'opca\_medicina.xls'. Parametar header postavljen na 1 omogućuje korištenje prvog retka kao zaglavlje datoteke.

```
podaci=pd.read_excel("opca_medicina.xls",header=[1])
```

2. Skup podataka potrebno je staviti u podatkovni okvir.

```
df = pd.DataFrame(podaci)
```

3. Neki stupci skupa podataka nisu imenovani, stoga im je potrebno dodijeliti nazive.

```
df.rename(      columns={'Unnamed: 3':'Naziv      djelatnosti','Unnamed: 7':'Status  
ordinacije','Unnamed: 8':'Broj osiguranja'}, inplace=True )
```

<sup>7</sup> Preuzeto s [http://www.hzzo.hr/wp-content/uploads/2018/04/web\\_opca\\_032018.xls?b32def](http://www.hzzo.hr/wp-content/uploads/2018/04/web_opca_032018.xls?b32def)

4. Grupiranje stupca 'Naziv PU' po stupcu 'Grad/ općina/ gradska četvrt Grada Zagreba'. Metoda count() koristi se za računanje broja uprava u pojedinoj županiji.

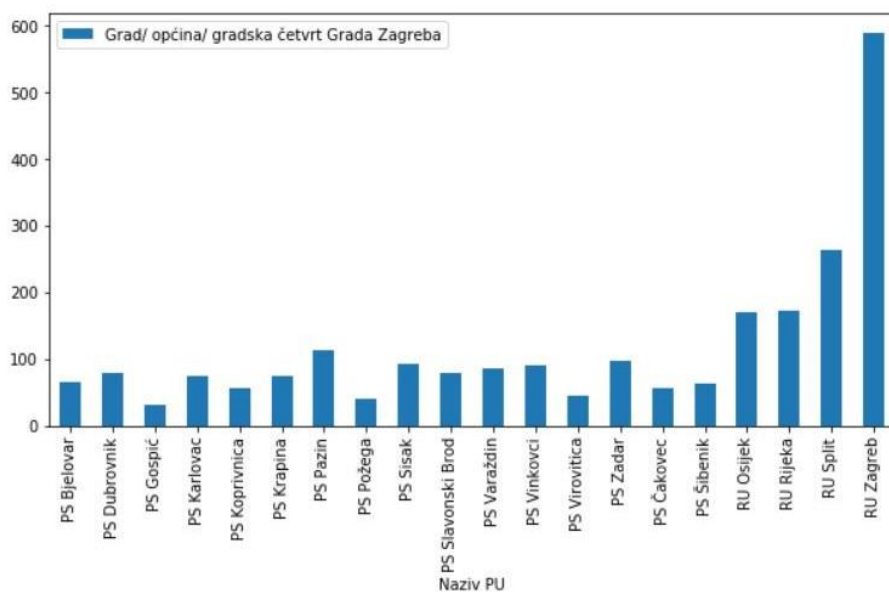
```
grupirano=df.groupby('Naziv PU')['Grad/ općina/ gradska četvrt Grada Zagreba'].count()
```

5. Iz skupa podataka dobiven je grafički prikaz broja ustanova u pojedinoj županiji.

```
df3.plot.bar(figsize=(10,5),legend=True)
```

```
import pandas as pd
podaci=pd.read_excel("opca_medicina.xls",header=[1])
df = pd.DataFrame(podaci)
df.rename( columns={'Unnamed: 3':'Naziv djelatnosti','Unnamed: 7':'Status ordinacije',
                  'Unnamed: 8':'Broj osiguranja'},inplace=True )
grupirano=df.groupby('Naziv PU')['Grad/ općina/ gradska četvrt Grada Zagreba'].count()
df3.plot.bar(figsize=(10,5),legend=True)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x4de2390>



Slika 24.Primjer1: Grafički prikaz

### 5.3.2.Primjer 2

Izgled datoteke:

|   | TIP       | NAZIV         | ADRESA                                          | VRSTA          | PROGRAMI                                         | ÈETVRT            | NAPOMENA | KOORIDNATA (E) | KOORIDNATA (N) |
|---|-----------|---------------|-------------------------------------------------|----------------|--------------------------------------------------|-------------------|----------|----------------|----------------|
| 0 | CENTRALNI | BAJKA         | Zorkovaèka ulica 8                              | Gradski vrtiæ  | smjenski program-djeca s tekoæama u razvoju s... | TREKNEJVKA-SJEVER | NaN      | 456878.03      | 5073682.79     |
| 1 | PODRUÈNI  | BAJKA         | Humska ulica 1                                  | Gradski vrtiæ  | NaN                                              | TREKNEJVKA-SJEVER | NaN      | 456324.11      | 5074495.64     |
| 2 | PODRUÈNI  | BAJKA         | Opatijski trg 9                                 | Gradski vrtiæ  | NaN                                              | TREKNEJVKA-SJEVER | NaN      | 456258.89      | 5073486.39     |
| 3 | PODRUÈNI  | BAJKA         | Selska cesta 95/1                               | Gradski vrtiæ  | NaN                                              | TREKNEJVKA-SJEVER | NaN      | 456784.22      | 5073802.17     |
| 4 | CENTRALNI | BALTAZAR      | Don Frane Buliaa 23a, Popovec                   | Privatni vrtiæ | NaN                                              | SESVETE           | NaN      | 472320.86      | 5078964.69     |
| 5 | CENTRALNI | BAMBI         | PO Mendlova 1, Zagreb (sjedite: Gorièka 17, ... | Privatni vrtiæ | NaN                                              | PODSUSED-VRAPEE   | NaN      | 449247.10      | 5075289.07     |
| 6 | CENTRALNI | BLAENA HOZANA | Trg Kardinala Franje epera 2                    | Vjerski vrtiæ  | engleski jezik-ritmika-portski program           | TRNJE             | NaN      | 459996.59      | 5073229.19     |

Slika 25.Datoteka vrticio.csv<sup>8</sup>

1. Korištenje Pythonovih biblioteka Pandas, Numpy i matplotlib:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

2. Učitavanje datoteke:

```
podaci=pd.read_csv("vrticio.csv",sep=';',encoding = 'ISO-8859-1')
```

3. Stavljanje u podatkovni okvir:

```
df = pd.DataFrame(podaci)
```

4. Promjena imena stupca:

```
df.rename(columns={'ÈETVRT':'ČETVRT'},inplace=True )
```

5. Promjena vrijednosti stupca 'TIP':

```
df['TIP']=df['TIP'].map({'PODRUÈNI':'PODRUČNI','CENTRALNI':'CENTRALNI'})
```

6. Grupiranje po tipu vrtića, korištenjem metode count() koja vraća broj centralnih i broj područnih vrtića:

```
grupirano_tip=df.groupby("TIP")["VRSTA"].count()
```

7. Računanje postotka centralne i područne vrste:

```
Postotak_centralnih=(grupirano_tip['CENTRALNI']/df['TIP'].count())*100
```

```
Postotak_podrucnih=(grupirano_tip['PODRUČNI']/df['TIP'].count())*100
```

8. Postavljanje vrijednosti izračunatog postotka koji će se prikazati na dijagramu, zaokruženi na 2 decimalno mjesta:

<sup>8</sup> Preuzeto s <http://opendataportal8502.cloudapp.net/dataset/65c23f8d-d333-4ec1-9a43-81876ad0a9fc/resource/0919ed9d-989e-449a-8457-5049b37c4a6c/download/vrticio.csv>

```
values = [Postotak_podrucnih.round(2),Postotak_centralnih.round(2)]
```

9. Određivanje boja dijagrama:

```
colors = ['r', 'g']
```

10. Naslov koji će se prikazati iznad dijagramu:

```
plt.title('Postotak vrtića po vrsti')
```

11. Odabrani oblik dijagrama:

```
grupirano_tip.plot.pie(figsize=(6,10))
```

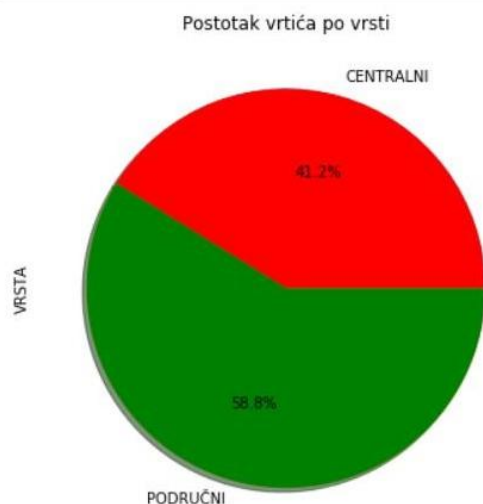
12. Prikaz zadanih parametara na dijagramu(vrijednosti, boje, izgled postotka, sjena):

```
plt.pie(values, colors=colors,autopct='%1.1f%%', shadow=True)
```

13. Grafički prikaz dijagrama:

```
plt.show()
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
podaci=pd.read_csv("vrticio.csv", sep=';', encoding = 'ISO-8859-1')
df = pd.DataFrame(podaci)
df.rename(columns={'ĚETVRT':'ČETVRT'},inplace=True )
df['TIP'] = df['TIP'].map({'PODRUĚNI':'PODRUĚNI', 'CENTRALNI':'CENTRALNI'})
grupirano_tip=df.groupby('TIP')['VRSTA'].count()
Postotak_centralnih=(grupirano_tip['CENTRALNI']/df['TIP'].count())*100
Postotak_podrucnih=(grupirano_tip['PODRUĚNI']/df['TIP'].count())*100
values = [Postotak_centralnih.round(2),Postotak_podrucnih.round(2)]
colors = ['r', 'g']
plt.title('Postotak vrtića po vrsti')
grupirano_tip.plot.pie(figsize=(6,10))
plt.pie(values, colors=colors,autopct='%1.1f%%',shadow=True)
plt.show()
```



Slika 26.Primjer2: Grafički prikaz

## 6. Zaključak

U ovome završnom radu opisan je postupak analize podataka nad skupovima podataka. Opisane su strukture podataka (niz podataka, podatkovni okvir, panel), te je objašnjena njihova primjena. Nad podacima, uz pomoć Python Pandas alata primijenjene su operacije sortiranja, agregiranja, grupiranja, filtriranja, transformiranja i spajanja, te su kao primjer prikazana dva skupa podataka nad kojima je provedena analiza.

Prvi skup podataka prikazuje popis medicinskih uprava. Učitana datoteka u Python Pandasu stavljena je u podatkovni okvir. Pridružena su imena stupcima koji nisu bili imenovani. Cilj analize bio je prikazati broj gradova pojedine uprave u grafičkom obliku. Stoga je izvršeno grupiranje naziva uprava, gdje se pojedinoj upravi pridružuju gradovi koji joj pripadaju. Kako bi se dobio broj gradova koji pripadaju upravi korištena je metoda `count()`, te je skup podataka prikazan u stupčastom dijagramu.

U drugom primjeru analize podataka korištena je datoteka koja prikazuje popis dječjih vrtića na području grada Zagreba. Datoteka ima oblik dvodimenzionalne tablice, te je stavljena u podatkovni okvir. Znakovi imena stupaca i redaka nisu se ispravno prikazivali, te su im promijenjena imena. Nakon provedbe grupiranja po tipu vrtića kao rezultat dobijemo dvije vrste vrtića. Primjenom metode `count()` dobijemo broj vrtića u pojedinoj vrsti. Izračunamo prosjeke

svake vrste čiju vrijednost prikažemo u dijagramu. Skup podataka prikazan je u kružnom dijagramu koji prikazuje postotak podjele dječjih vrtića u dvjema grupama.

Pythonova biblioteka Pandas omogućuje funkcionalnosti kao što su učitavanje datoteka različitih formata, čišćenje podataka, primjenu različitih operacija te omogućuje vizualni prikaz dobivenih podataka. Alat Python Pandas pruža visoke performanse za cijeli proces analize podataka.

## Literatura

1. [Mrežno] [Citirano: 19. 8. 2019.] [https://en.wikipedia.org/wiki/Data\\_analysis](https://en.wikipedia.org/wiki/Data_analysis).
2. [Mrežno] [Citirano: 19. 8. 2019.] <https://www.flydata.com/blog/a-brief-history-of-data-analysis/>.
3. [Mrežno] [Citirano: 20. 8. 2019.] [https://hr.wikipedia.org/wiki/Hermann\\_Hollerith](https://hr.wikipedia.org/wiki/Hermann_Hollerith).
4. [Mrežno] [Citirano: 22. 8. 2019.] [https://hr.wikipedia.org/wiki/Skladi%C5%A1tenje\\_podataka](https://hr.wikipedia.org/wiki/Skladi%C5%A1tenje_podataka).
5. [Mrežno] [Citirano: 25. 8. 2019.] <http://www.mit-software.hr/usluge/bi/bi1/>.
6. [Mrežno] [Citirano: 27. 8. 2019.] [https://hr.wikipedia.org/wiki/Rudarenje\\_podataka](https://hr.wikipedia.org/wiki/Rudarenje_podataka).
7. [Mrežno] [Citirano: 28. 8. 2019.] [https://bs.wikipedia.org/wiki/Larry\\_Page](https://bs.wikipedia.org/wiki/Larry_Page).
8. [Mrežno] [Citirano: 30. 8. 2019.] [https://en.wikipedia.org/wiki/Amazon\\_Redshift](https://en.wikipedia.org/wiki/Amazon_Redshift).
9. [Mrežno] [Citirano: 30. 8. 2019.] <https://en.wikipedia.org/wiki/BigQuery>.
10. [Mrežno] [Citirano: 30. 8. 2019.] <https://www.urban.org/research/data-methods/data-analysis/quantitative-data-analysis/descriptive-data-analysis>.
11. [Mrežno] [Citirano: 1. 9. 2019.] [https://books.google.hr/books?hl=hr&lr=&id=jF8QCBkhvQC&oi=fnd&pg=IA1&dq=exploratory+data+analysis&ots=KI5vDMExsE&sig=sTeYzEOGmwYs3JpqnkbiVU\\_SGuk&redir\\_esc=y#v=onepage&q=exploratory%20data%20analysis&f=false](https://books.google.hr/books?hl=hr&lr=&id=jF8QCBkhvQC&oi=fnd&pg=IA1&dq=exploratory+data+analysis&ots=KI5vDMExsE&sig=sTeYzEOGmwYs3JpqnkbiVU_SGuk&redir_esc=y#v=onepage&q=exploratory%20data%20analysis&f=false).
12. [Mrežno] [Citirano: 17. 9. 2019.] pandas <https://pandas.pydata.org/getpandas.html>.
13. [Mrežno] [Citirano: 17. 9. 2019.] numpy <https://pypi.org/project/numpy/>.
14. [Mrežno] [Citirano: 2. 9. 2019.] [https://en.wikipedia.org/wiki/Pandas\\_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software)).

## Popis slika

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Slika 1.Niz podataka .....                                                 | 8  |
| Slika 2.Niz podataka sa zadanim vrijednostima.....                         | 9  |
| Slika 3.Niz podataka, skalarna vrijednost .....                            | 9  |
| Slika 4.Podatkovni okvir .....                                             | 10 |
| Slika 5.Podatkovni okvir sa zadanim nazivima stupaca i vrijednostima ..... | 10 |
| Slika 6.Podatkovni niz u podatkovnom okviru .....                          | 10 |
| Slika 7.Panel.....                                                         | 11 |
| Slika 8.Datoteka biciklistickestazei.csv .....                             | 13 |
| Slika 9.Sortiranje .....                                                   | 14 |
| Slika 10.Sortiranje stupaca .....                                          | 15 |
| Slika 11.Datoteka zoo.csv .....                                            | 16 |
| Slika 12.Agregacija .....                                                  | 17 |
| Slika 13.Datoteka ustanove_z.csv .....                                     | 18 |
| Slika 14.Grupiranje .....                                                  | 19 |
| Slika 15.Grupiranje uz uvjet.....                                          | 19 |
| Slika 16.Datoteka Zavodi_adresar.csv .....                                 | 20 |
| Slika 17.Filtriranje.....                                                  | 21 |
| Slika 18.Datoteka sales_transactions.xlsx.....                             | 22 |
| Slika 19.Grupiranje i korištenje metode sum() .....                        | 23 |
| Slika 20.Transformiranje.....                                              | 24 |
| Slika 21.Spajanje prema nazivu .....                                       | 25 |
| Slika 22.Spajanje prema indeksima .....                                    | 25 |
| Slika 23. Datoteka opca_medicina.xls .....                                 | 27 |
| Slika 24.Primjer1: Grafički prikaz .....                                   | 28 |
| Slika 25.Datoteka vrticio.csv .....                                        | 29 |
| Slika 26.Primjer2: Grafički prikaz .....                                   | 30 |