

Основне академске студије
Информациони инжењеринг

Методе и технике науке о подацима

Визуализација података у језику Python

(материјали за вежбе)

Библиотека *Matplotlib*

посвећена прављењу визуализација података

по узору на команде за *MATLAB*

ослањање на библиотеку *NumPy*

подршка за разноврсне визуализације

стубичасти графикон (енгл. *bar plot*)

графикон расејања (енгл. *scatter plot*)

тортни графикон (енгл. *pie chart*)

хистограм

кутијаста графикон (енгл. *boxplot*)

контурне визуализације

...

Библиотека *Matplotlib*

ОСНОВНИ ПОДАЦИ

дело Џона Д. Хантера и сарадника

настанак 2003. године

слободна за употребу

актуелна верзија 3.10.0

Интернет сајт

<https://matplotlib.org/>

Припреме за коришћење

инсталација библиотеке

библиотека доступна у оквиру дистрибуције *Anaconda*

могућа засебна инсталација

употребом *pip*

употребом *conda*

учитавање библиотеке и потребног садржаја

```
import matplotlib.pyplot as plt
```

коришћење псеудонима **plt**

Основни програмски систем рада при прављењу визуализације

припрема података који треба да буду визуализовани
постављање и подешавање радне површине (слике) на којој треба извести визуализацију

исцртавање главног садржаја визуализације одабраног типа
могућност комбиновања више визуализација

исцртавање пратећих садржаја
наслов, ознаке, напомене, легенда...

Основне функције у интерфејсу *matplotlib.pyplot*

постављање радне површине (слике)

функција **figure(...)**

исцртавање главног садржаја

стубичасти графикон

функција **bar(...)**

графикон расејања

функција **scatter(...)**

тортни графикон

функција **pie(...)**

хистограм

функција **hist(...)**

кутијастаи графикон

функција **boxplot(...)**

виолински графикон

функција **violinplot(...)**

Основне функције у интерфејсу *matplotlib.pyplot*

исцртавање пратећих садржаја

наслов

функција **title(...)**

ознака X осе

функција **xlabel(...)**

ознака Y осе

функција **ylabel(...)**

подеоци X осе

функција **xticks(...)**

подеоци Y осе

функција **yticks(...)**

легенда

функција **legend(...)**

Исцртавање графикана

пример – библиотеке

```
import random as rand  
import numpy as np  
import matplotlib.pyplot as plt
```

Исцртавање графикана

пример – улазни подаци

```
rand.seed(3)
```

```
labele = ["K1", "K2", "K3"]
```

```
podaci = rand.choices(labele, k=10)
```

```
brojnost = [podaci.count(labela)  
            for labela in labele]
```

Исцртавање графикана

пример – улазни подаци

```
gpsv = np.random.default_rng(3)
```

```
n = 100
```

```
a = np.linspace(0, 20, n)
```

```
b = np.sort(gpsv.triangular(2, 5, 7, n)  
            ) + gpsv.normal(0.3, 0.15, n)
```

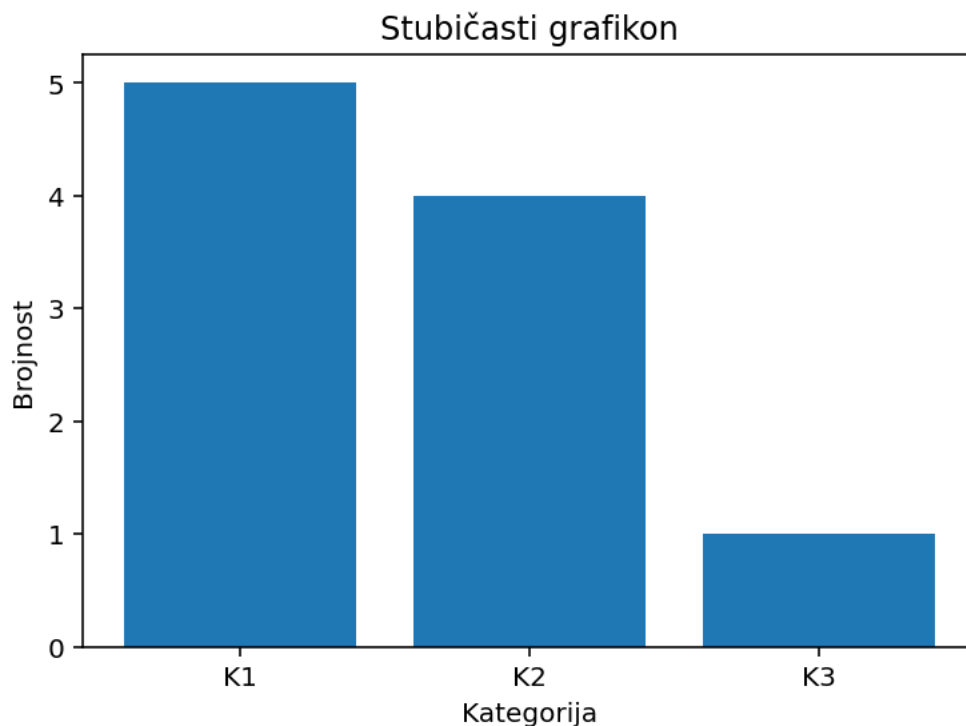
Исцртавање графикана

пример – стубичасти графикон

```
plt.figure()  
plt.bar(range(1, len(labele) + 1),  
        brojnost, tick_label=labele)
```

```
plt.title("Stubičasti grafikon")  
plt.xlabel("Kategorija")  
plt.ylabel("Brojnost")
```

Исцртавање графика пример – стубичасти графикон



Исцртавање графикана

пример – графикон расејања

```
plt.figure()
```

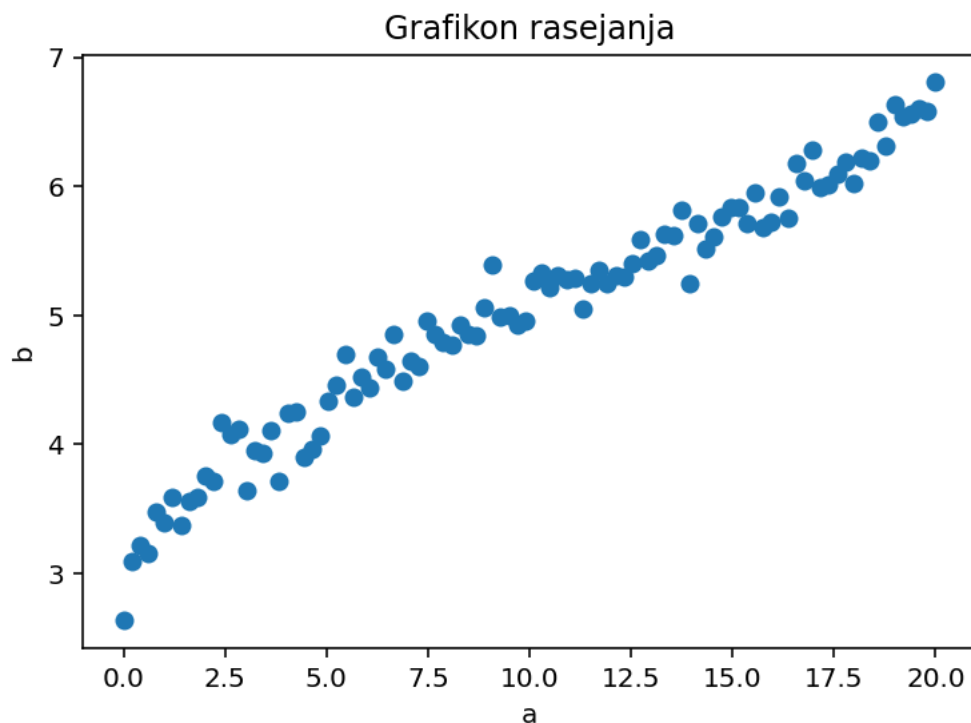
```
plt.scatter(a, b)
```

```
plt.title("Grafikon rasejanja")
```

```
plt.xlabel("a")
```

```
plt.ylabel("b")
```

Исцртавање графика пример – графикон расејања



Исцртавање графика

пример – хистограм

```
plt.figure()
```

```
plt.hist(b)
```

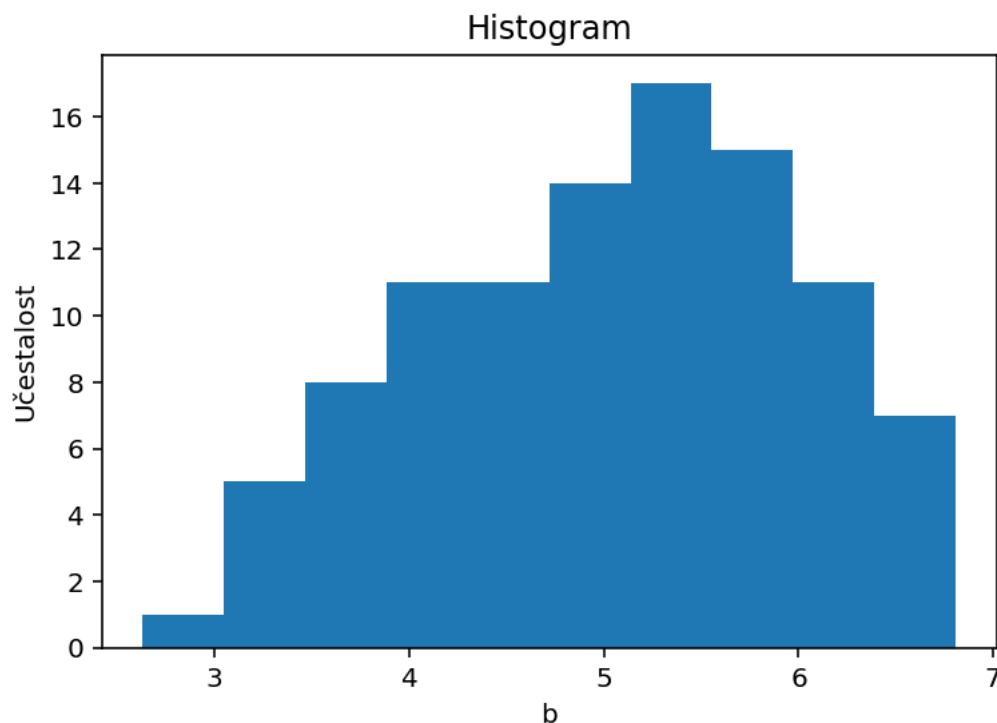
```
plt.title("Histogram")
```

```
plt.xlabel("b")
```

```
plt.ylabel("Učestalost")
```

Исцртавање графика

пример – хистограм



Подаци

скуп података о притужбама на авионску буку за подручје Сан Франциска (САД)

постоји 8778 појава и 8 обележја

свака појава обухвата збирне податке о броју притужби и броју позивалаца за један месец једне године у једној заједници
обележја за годину, месец, почетак месеца, заједницу, број притужби...

извор скупа података

скуп података *Aircraft Noise Report Summary*

матична Интернет страна

https://data.sfgov.org/Transportation/Aircraft-Noise-Report-Summary/q3xd-hfi8/about_data

у оквиру портала отворених података Сан Франциска (САД)

подаци доступни у више формата

може се користити формат CSV (CSV датотека доступна кроз опцију *Export*)

скуп података креиран 9. 4. 2016. а најновије повезано ажурирање 24. 3. 2026.

назначено одељење које је објавило скуп података је *Airport (SFO)*

назначени власник скупа података је *NoiseAbatementOffice@flysfo.com*

лиценца и могућности коришћења скупа података

Open Data Commons Public Domain Dedication and License

<https://opendatacommons.org/licenses/pddl/1-0/>

Задатак 1.

Учитати скуп података из одговарајуће *CSV* датотеке.

Проверити садржај учитаних података и типове колона и, у случају да недостајуће вредности постоје, уклонити из скупа података све редове у којима су такве вредности присутне.

Из учитаног скупа података задржати само колоне о години, месецу, заједници, броју притужби и броју позивалаца.

Задатак 2.

Нацртати стубичасти графикон који приказује број притужби по месецима 2016. године за заједницу Санта Круз.

Подесити ознаке стубова на графикону.

Подесити наслов графикона и називе оса.

Задатак 3.

Нацртати стубичасти графикон који приказује број притужби по месецима 2010. године узимајући у обзир све заједнице.

Стубичасти графикон поставити хоризонтално.

Подесити ознаке стубова на графикону.

Подесити висину и боју стубова.

Подесити боју и дебљину ивице код стубова.

Подесити наслов графикона и називе оса.

Задатак 4.

Нацртати тортни графикон који приказује расподелу броја притужби у новембру 2012. године по појединачним заједницама.

Подесити ознаке и њихов положај на графикону.

Подесити наслов графикона.

Подесити позицију наслова тако да буде избегнуто преклапање с осталим садржајем.

Задатак 5.

Нацртати хистограм за месечни број позивалаца у периоду од 2010. до 2012. године.

Подесити наслов хистограма и називе оса.

Задатак 6.

Нацртати хистограм за месечни број позивалаца у периоду од 2010. до 2012. године тако да број интервала буде 20 а скала за учесталост логаритамска.

Подесити приказ на хистограму тако да буде контурног типа.

Подесити наслов хистограма и називе оса.

Задатак 7.

На заједничкој површини нацртати два хистограма за месечни број притужби тако да буду један преко другог.

Први хистограм формирати на основу података за 2009. годину а други на основу података за 2012. годину.

Подесити оба хистограма тако да буду искоришћене исте границе интервала.

Подесити за оба хистограма скалу за учесталост тако да буде логаритамска.

Подесити боје за оба хистограма тако да садржај оба хистограма буде истовремено видљив.

Подесити заједнички наслов и називе оса.

Поставити легенду.

Основна литература

Matplotlib. Matplotlib — Visualization with Python. Internet: <https://matplotlib.org/>

Matplotlib. Using Matplotlib — Matplotlib 3.10.3 documentation. Internet: <https://matplotlib.org/stable/users/index>

Rougier NP. Scientific Visualization: Python + Matplotlib. 2021. Internet: <https://www.labri.fr/perso/nrougier/scientific-visualization.html>