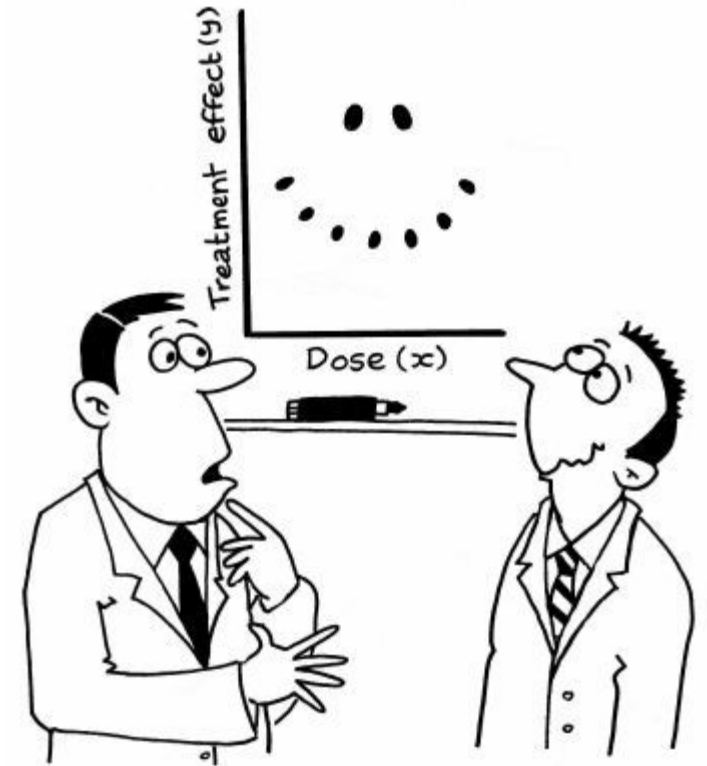

Vrednosti van granica

— Ana Vuksić 1086/2016 —

Šta su vrednosti van granica?

- Vrednosti van granica su podaci čija vrednost je značajno udaljena i odskače od drugih njemu bliskih vrednosti.

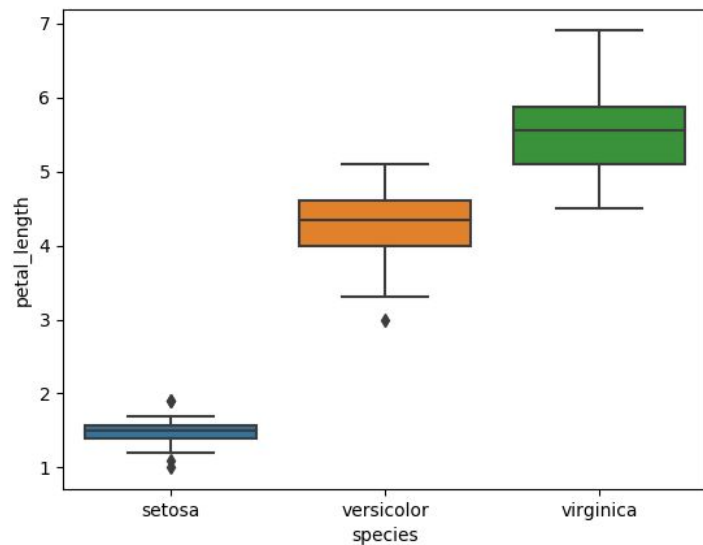


"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

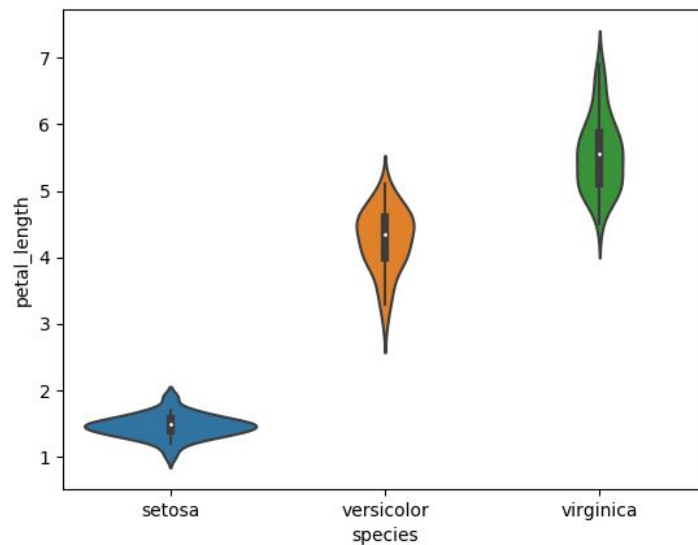
- Kako one nastaju?
 - Greska pri unosu podatak
 - Greska pri merenju (instrumentom)
 - Namerno (radi testiranja)
 - Greske nastale pri procesiranju
 - Prirodno, nisu greske
- Da li ih je lako prepoznati-detektovati?

Vizuelna reprezentacija

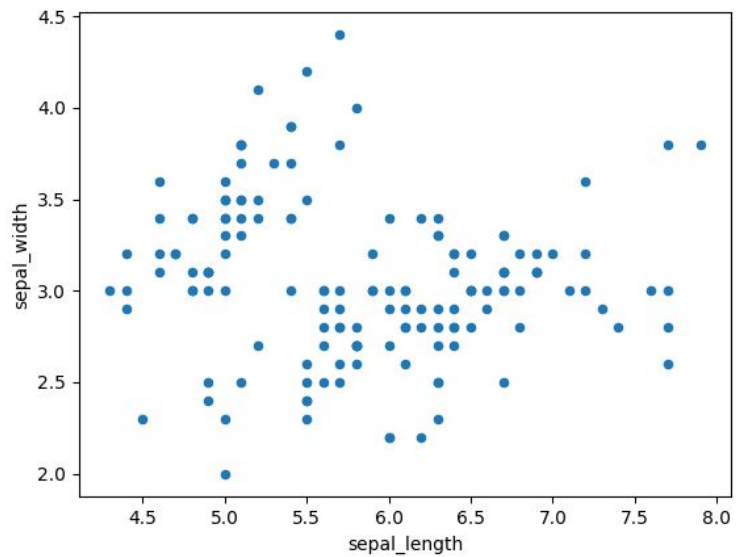
kutijasti



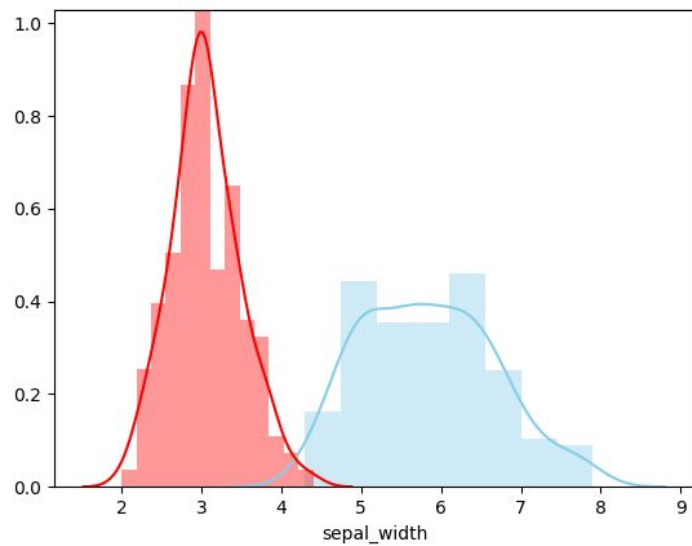
violinjski



raštrkani



histogram

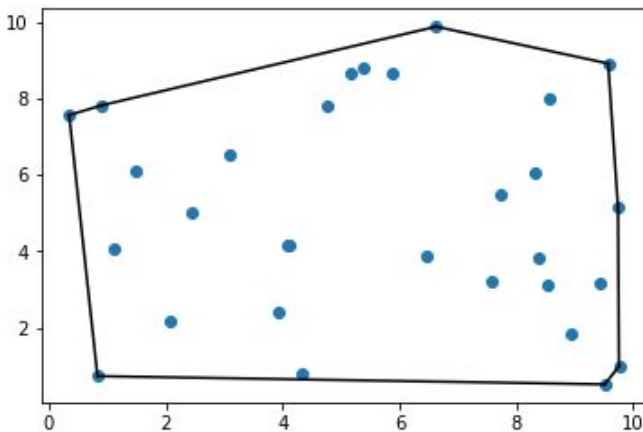
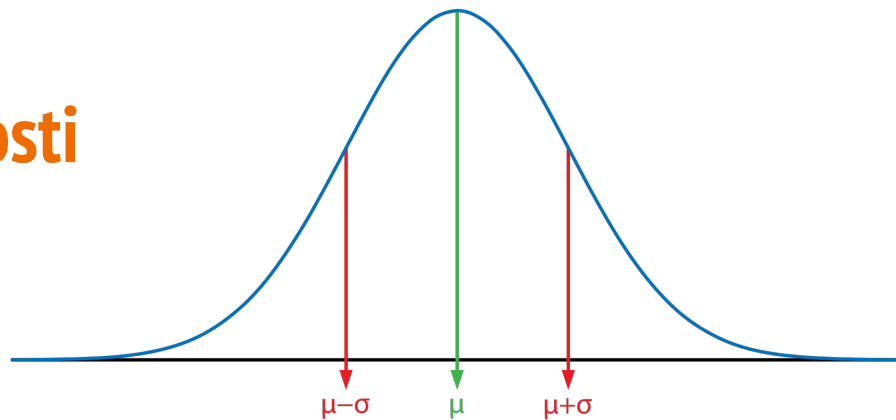


Metodi za analizu vrednosti van granica

- Ekstremne vrednosti
- Klasterovanje
- Modeli zasnovani na udaljenosti
- Modeli zasnovani na gustini
- Verovatnosni modeli
- Informaciono-teoretski model

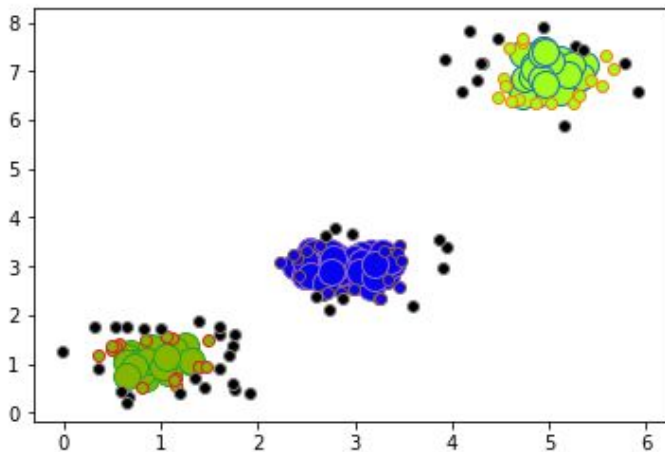
Analiza ekstremnih vrednosti

- Jednodimenzionalnih podataka
- Multidimenzionalnih podataka
- Metode zasnovane na dubini

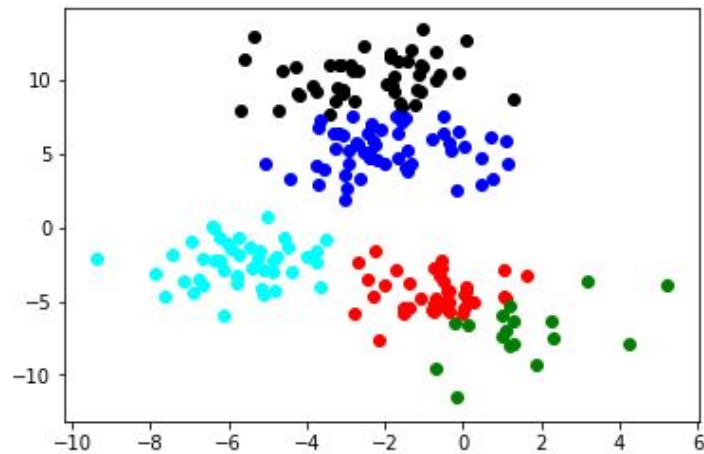


Analiza pomoću klasterovanja

DBSCAN algoritam

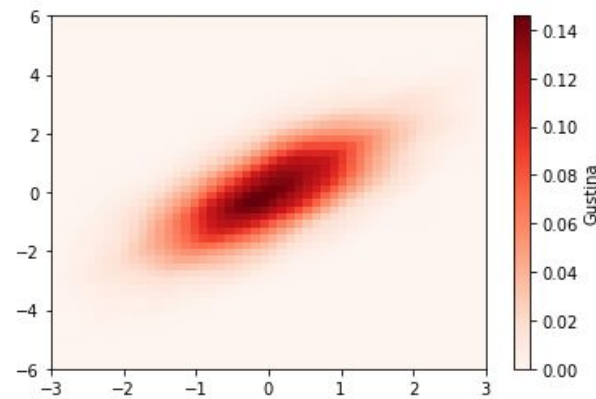
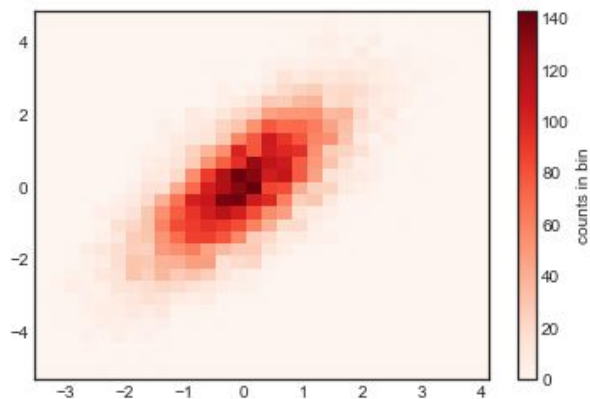


Algoritam K srednjih vrednosti



Modeli bazirani na gustini

- Histogrami (korpe)
- Rešetke
- Kerneli



Moguća rešenja

- Odstranjivanje
- Izbacivanje i čuvanje pojavljivanja
- Smena (zamena za maksimum/minimum)
- Transformacija promenljivih
- Korišćenje primerenih metoda

Kada čuvati vrednost, a kada ne?

Izbaciti:

- Znate da je pogrešan (godine života)
- Imate puno podataka (ne pravi razliku)
- Možete ih ponovo napraviti/proveriti

Ne izbacivati:

- Kritični podaci, i mala izmena utiče (temperatura kvara aviona)
- Puno vrednosti van granica

Primena

- Kontrola kvaliteta i detekcija prevare
- Finansijska prevara
- Analiza WEB poseta
- Neodobreni upadi
- Primena u zemljanim naukama

GitHub

Literatura

- Data Mining, Aggarwal, Charu C.
- Python Documentation
- Engineering Statistics Book