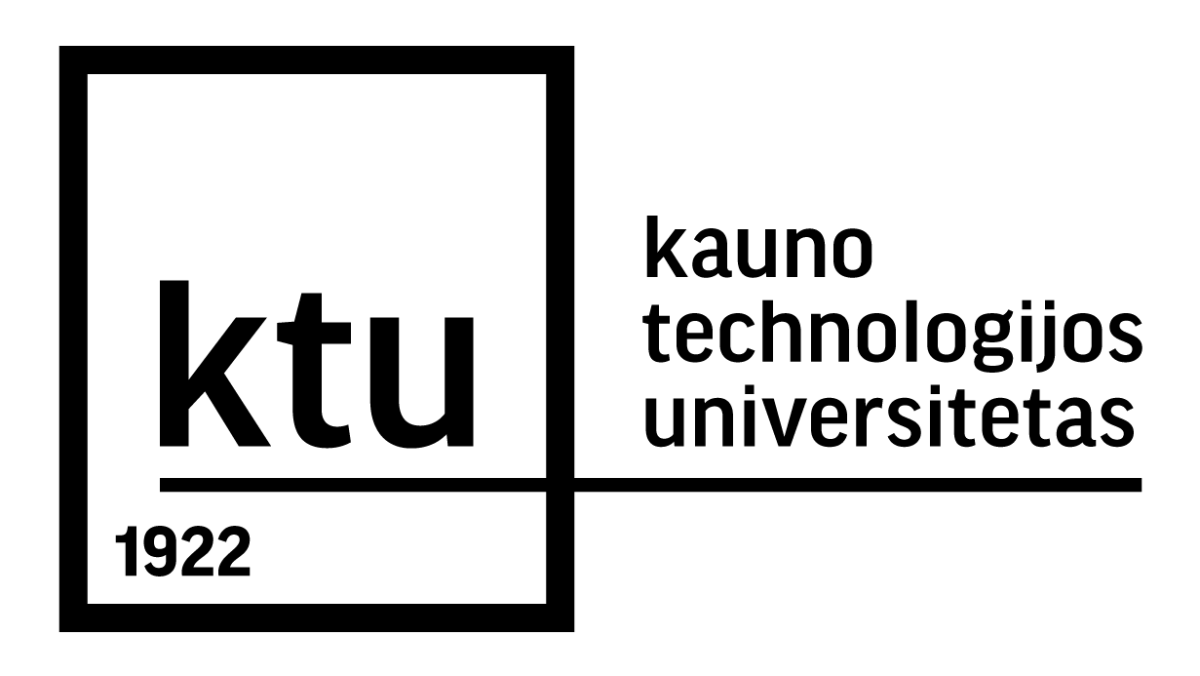
KAUNO TECHNOLOGIJOS UNIVERSITETAS

INFORMATIKOS FAKULTETAS



Intelektikos pagrindai (P176B101)

Laboratorinis darbas Nr.1

Atliko:

IFF-8/3 gr. studentas

Dovydas Zamas

2021 m. kovo 4 d.

Priėmė:

Lekt. PaulauskaitėTarasevičienė Agnė

KAUNAS 2021

Contents

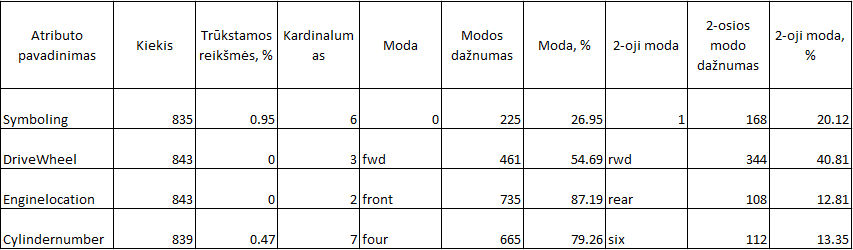
[1. Duomenų rinkinys 3](#_Toc65774127)

# Duomenų rinkinys

Laboratoriniui darbui pasirinktas automobilių specifikacijų rinkinys. Iš automobilio duomenų rinkinio buvo pašalinti unikalūs atributai, pvz., „car\_ID“, „CarName“. Modifikuotą automobilio duomenų rinkinį sudaro šie atributai: „symboling“[[[1]](#footnote-1)], „drivewheel“[[[2]](#footnote-2)], „enginelocation“[[[3]](#footnote-3)], „cylindernumber“[[[4]](#footnote-4)], „enginesize“[[[5]](#footnote-5)], „compressionratio“[[[6]](#footnote-6)], „horsepower“[[[7]](#footnote-7)], „citympg“[[[8]](#footnote-8)], „highwaympg“[[[9]](#footnote-9)].

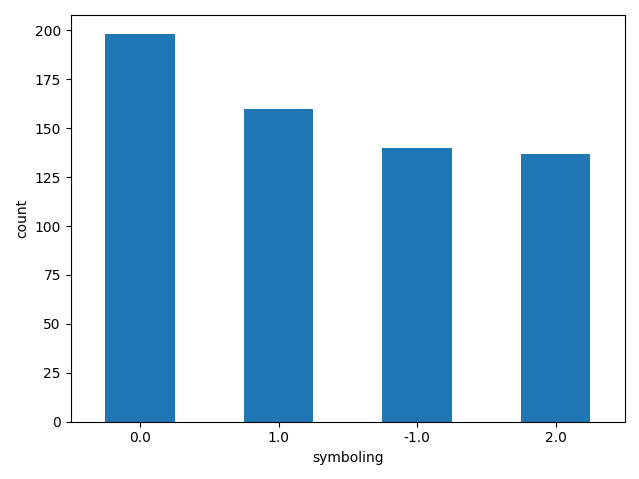
# Duomenų rinkinio kokybės analizė

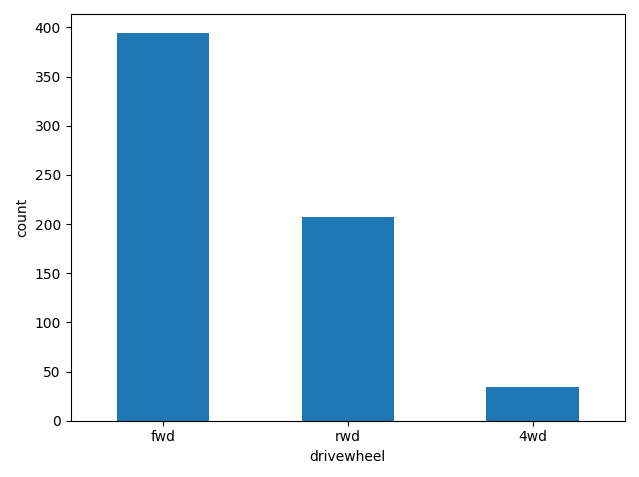
|  |
| --- |
| pav. 1 Tolydinio tipo atributų kokybės analizės lentelė |
|  |

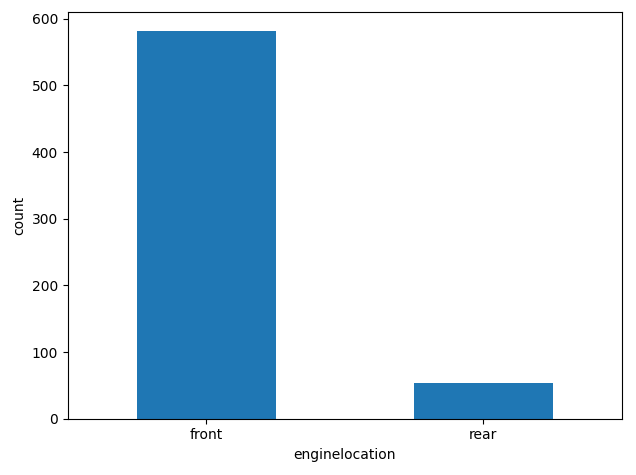


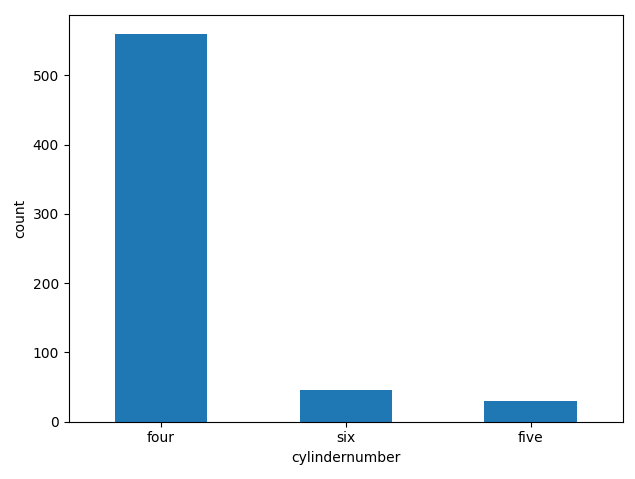
pav. 2 Kategorinio tipo atributų kokybės analizės lentelė

# Atributų histogramos









# Duomenų kokybės problemos ir sprendimai

Duomenų rinkinio atributai turėjo trūkstamų reikšmių bei išskirčių. Įrašai kurie turėjo tuščių reikšmių, bei išskirčių buvo ištrinti. Išskirčių radimui buvo pasinaudota „python“ biblioteka „pandas“, randami kvantiliai ir pagal juos atrenkami duomenys.

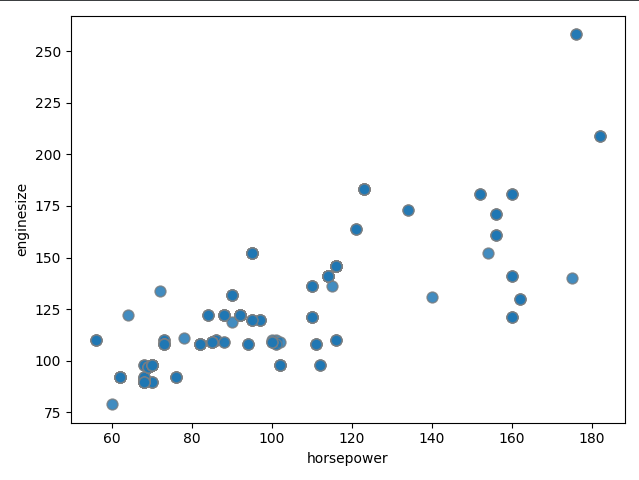
Kodo fragmentas išskirčių radimui ir šalinimui:

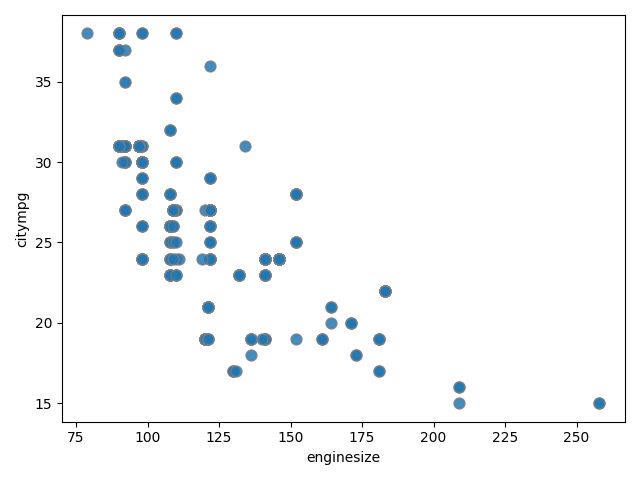
def deleteOutliers(data):  
 q\_low = data["enginesize"].quantile(0.0005)  
 q\_hi = data["enginesize"].quantile(0.9995)  
 data = data[(data["enginesize"] < q\_hi) & (data["enginesize"] > q\_low)]  
  
 q\_low = data["symboling"].quantile(0.0005)  
 q\_hi = data["symboling"].quantile(0.9995)  
 data = data[(data["symboling"] < q\_hi) & (data["symboling"] > q\_low)]  
  
 q\_low = data["compressionratio"].quantile(0.005)  
 q\_hi = data["compressionratio"].quantile(0.9995)  
 data = data[(data["compressionratio"] < q\_hi) & (data["compressionratio"] > q\_low)]  
  
 q\_low = data["horsepower"].quantile(0.0005)  
 q\_hi = data["horsepower"].quantile(0.9995)  
 data = data[(data["horsepower"] < q\_hi) & (data["horsepower"] > q\_low)]  
  
 q\_low = data["citympg"].quantile(0.0005)  
 q\_hi = data["citympg"].quantile(0.9995)  
 data = data[(data["citympg"] < q\_hi) & (data["citympg"] > q\_low)]  
  
 q\_low = data["highwaympg"].quantile(0.0005)  
 q\_hi = data["highwaympg"].quantile(0.9995)  
 data = data[(data["highwaympg"] < q\_hi) & (data["highwaympg"] > q\_low)]  
 return data

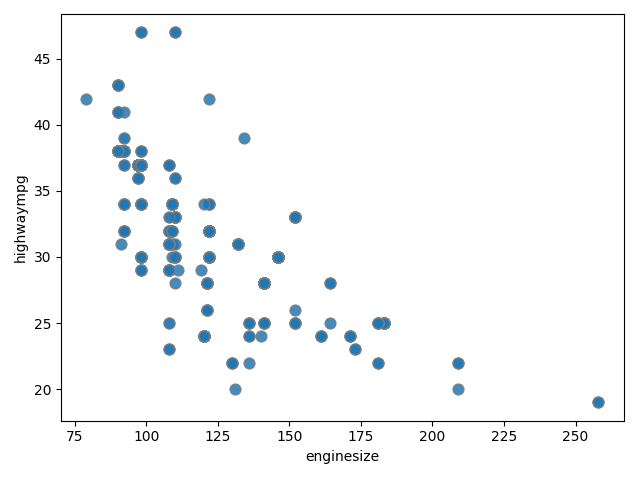
Kodo fragmentas ištrinti eilutes su trūkstamomis reikšmėmis:

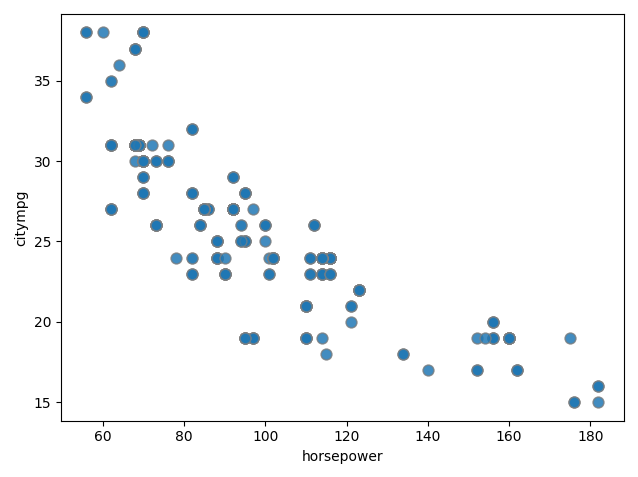
def deleteNaN(data):  
 data = data.dropna()

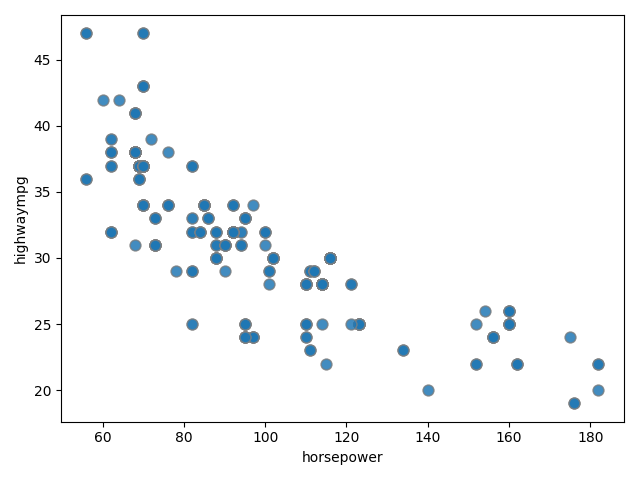
# Tolydinio tipo atributų vizualizacija



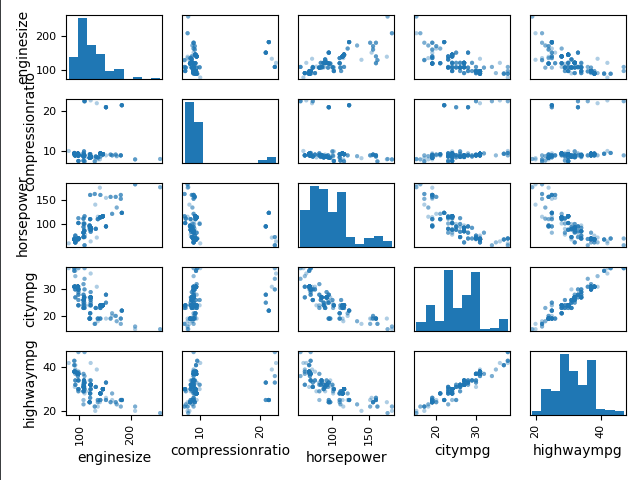




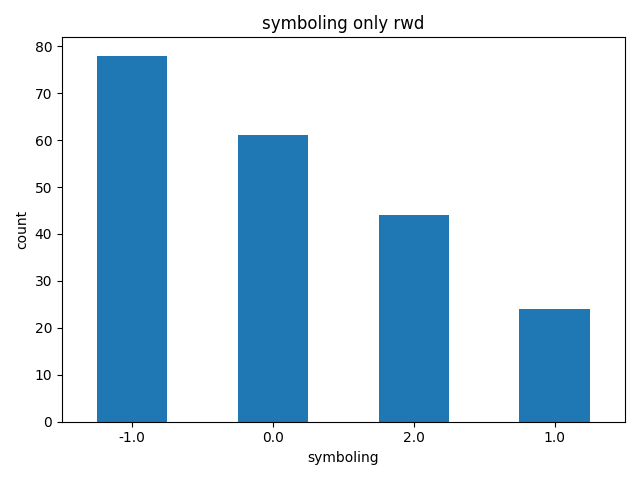


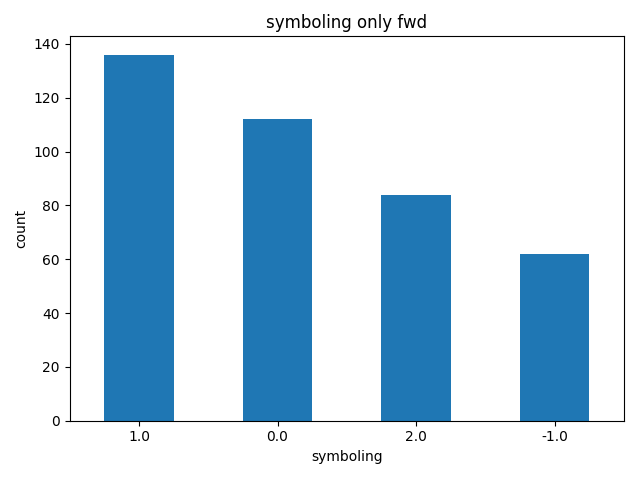


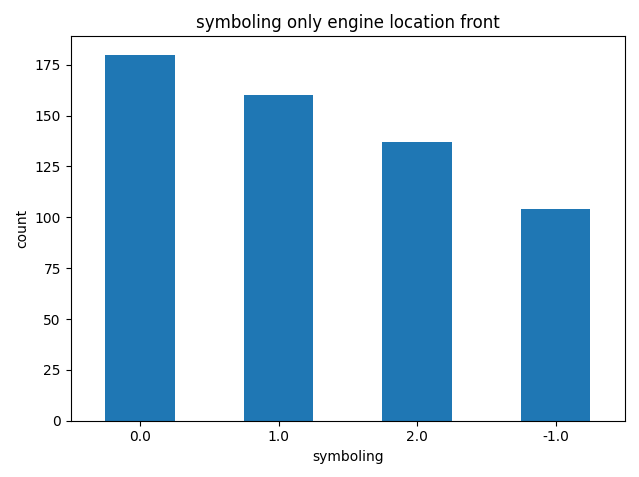
# Scatter plot Matrix diagrama

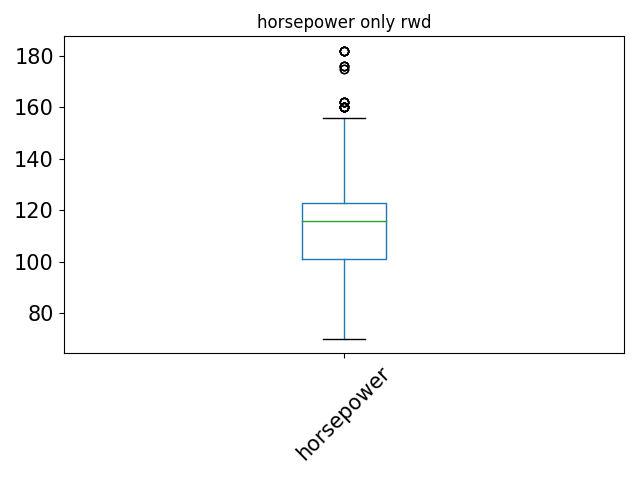


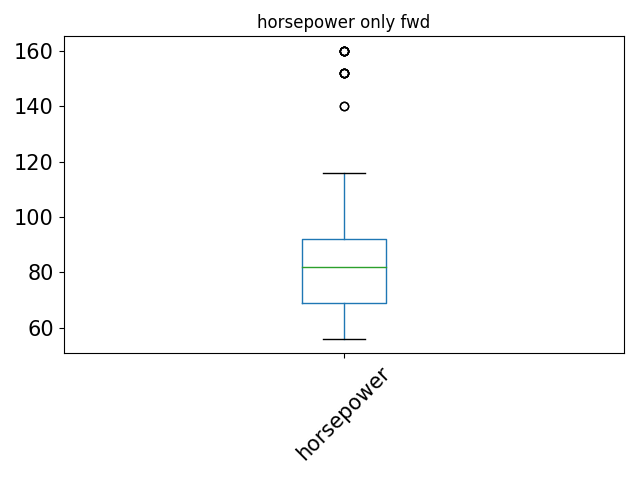
# Kategorinio tipo atributų vizualizacija

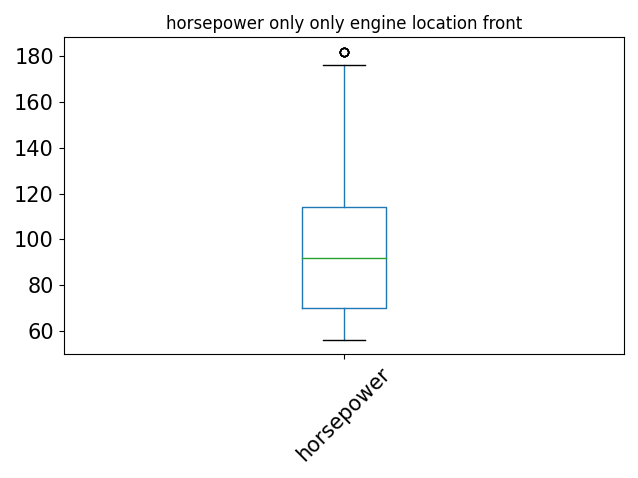


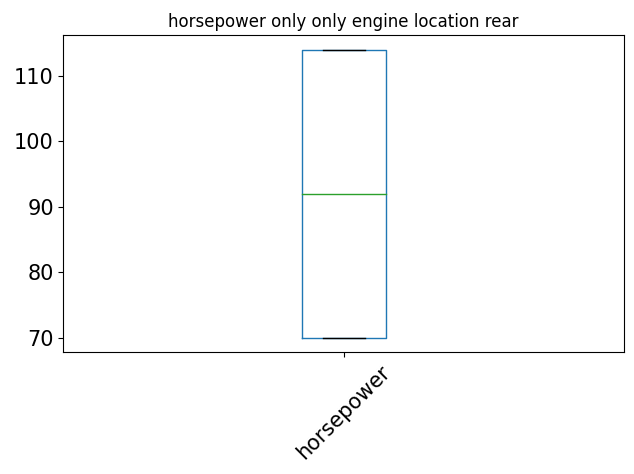




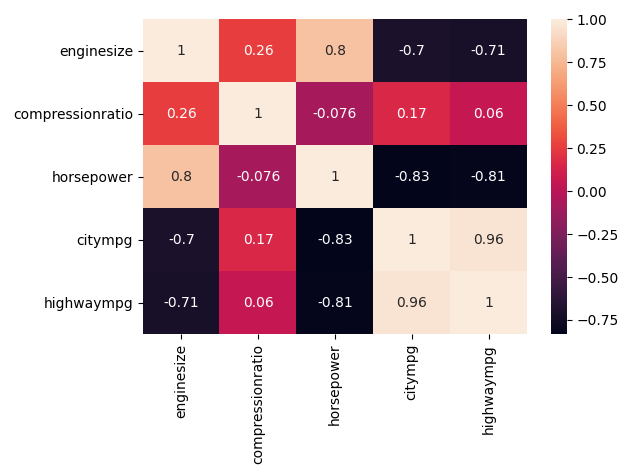








# Koreliacijos matrices diagrama

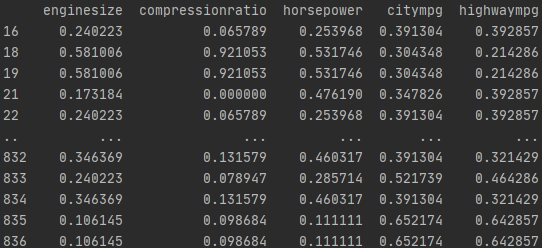


# Duomenų normalizacija

Programos kodas atlikti duomenų normalizacijai:

def normalize(data):  
 result = data.copy()  
 for feature\_name in data.columns:  
 max\_value = data[feature\_name].max()  
 min\_value = data[feature\_name].min()  
 result[feature\_name] = (data[feature\_name] - min\_value) / (max\_value - min\_value)  
 return result

Programos rezultatai:



1. Simboliai nusako mašinos rizikingumo laipsnį [↑](#footnote-ref-1)
2. Varomieji ratai [↑](#footnote-ref-2)
3. Variklio pozicija [↑](#footnote-ref-3)
4. Cilindrų skaičius esantis mašinoje [↑](#footnote-ref-4)
5. Variklio dydis [↑](#footnote-ref-5)
6. Suspaudimo laipsnis [↑](#footnote-ref-6)
7. Arklio galios [↑](#footnote-ref-7)
8. Automobilio sąnaudos mieste [↑](#footnote-ref-8)
9. Automobilio sąnaudos užmiestyje [↑](#footnote-ref-9)