

Donošenje zaključaka na osnovu podataka

Pavle Pađin

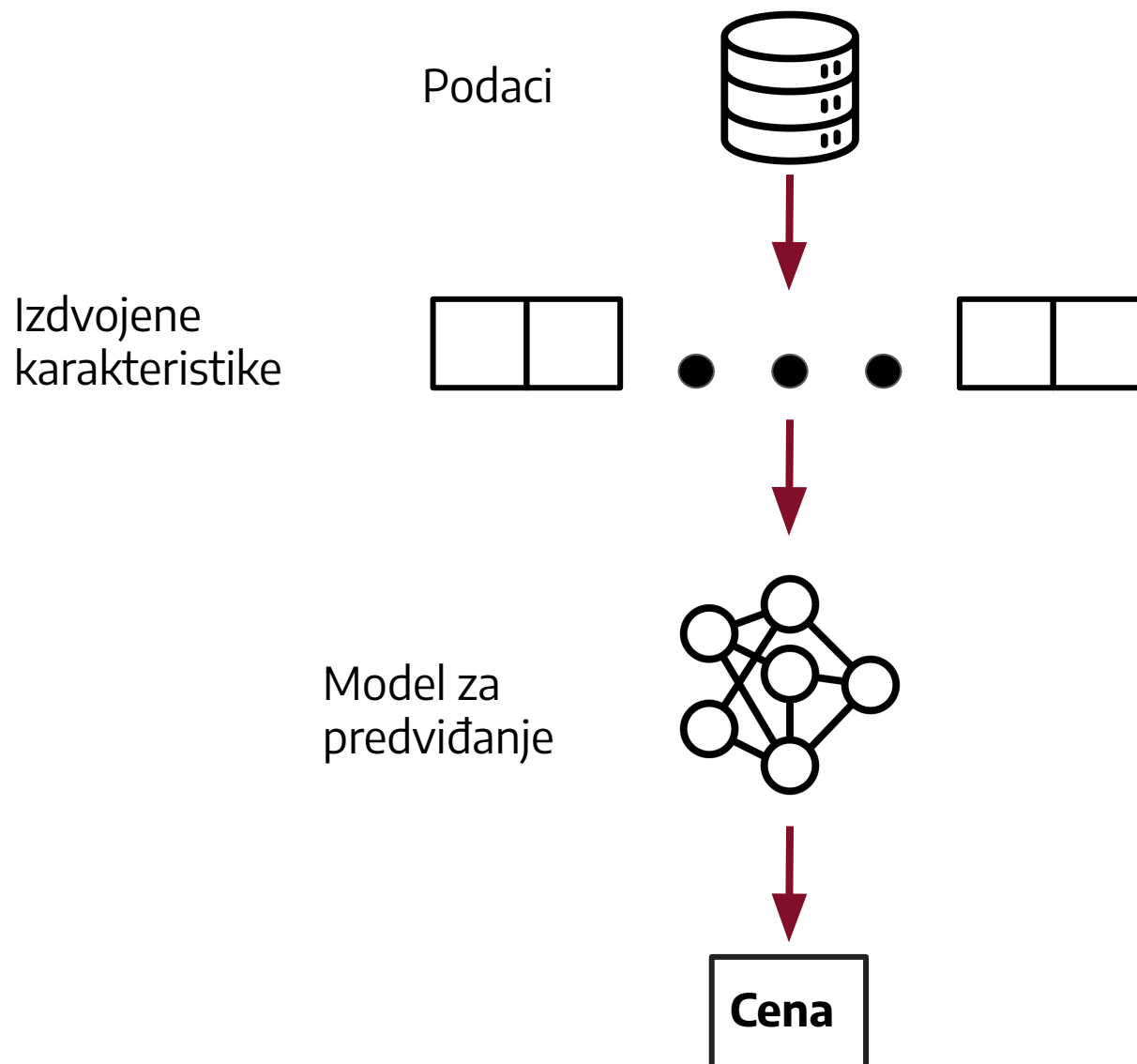
Matematička gimnazija

11. 04. 2022.

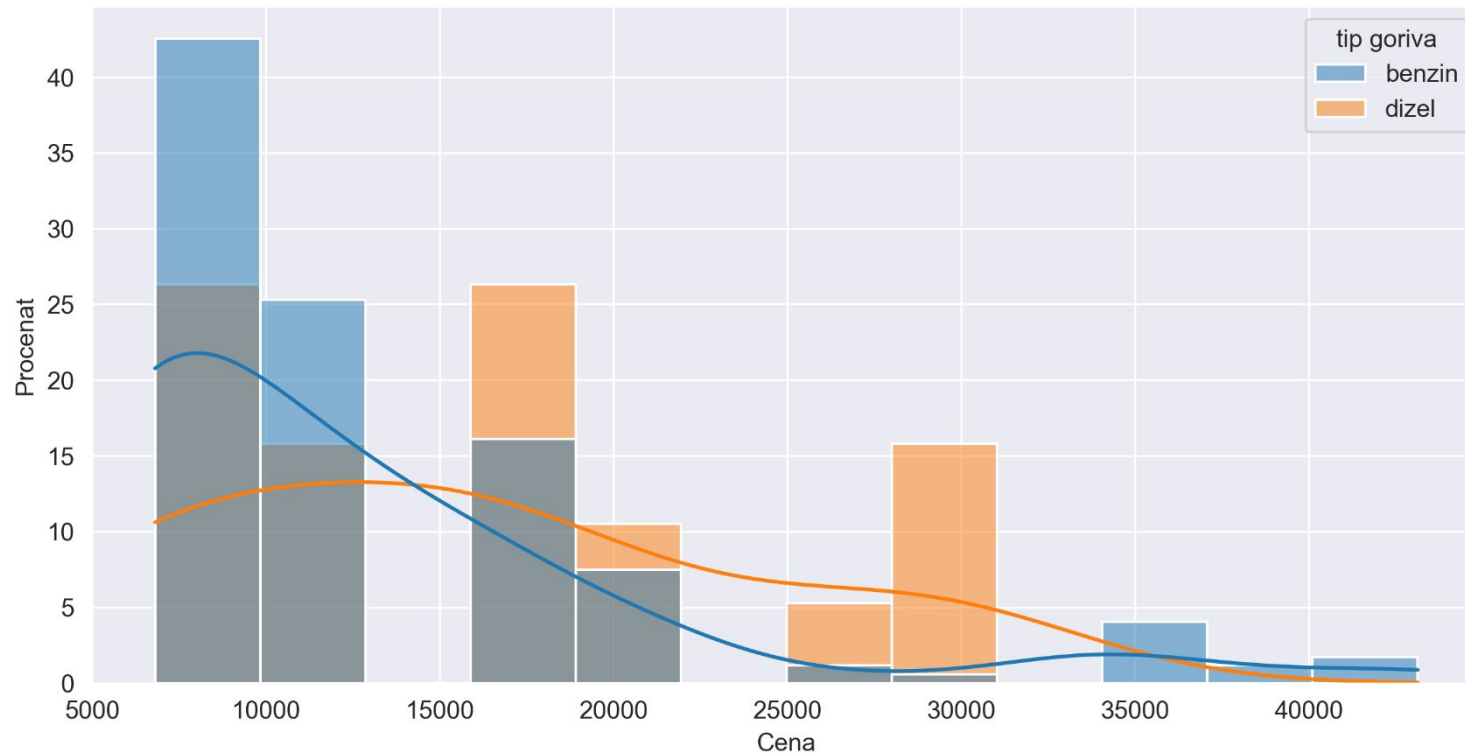
- Baza podataka automobila sa njihovim cenama

marka	tip goriva	broj vrata	snaga motora	...	cena
audi	benzin	4	102		13950
mercedes	dizel	4	123		28000
nissan	benzin	2	130		17200

- Želimo da nađemo karakteristike koje nam pomažu da predvidimo cenu automobila
- [Ovde](#) se može naći originalna baza

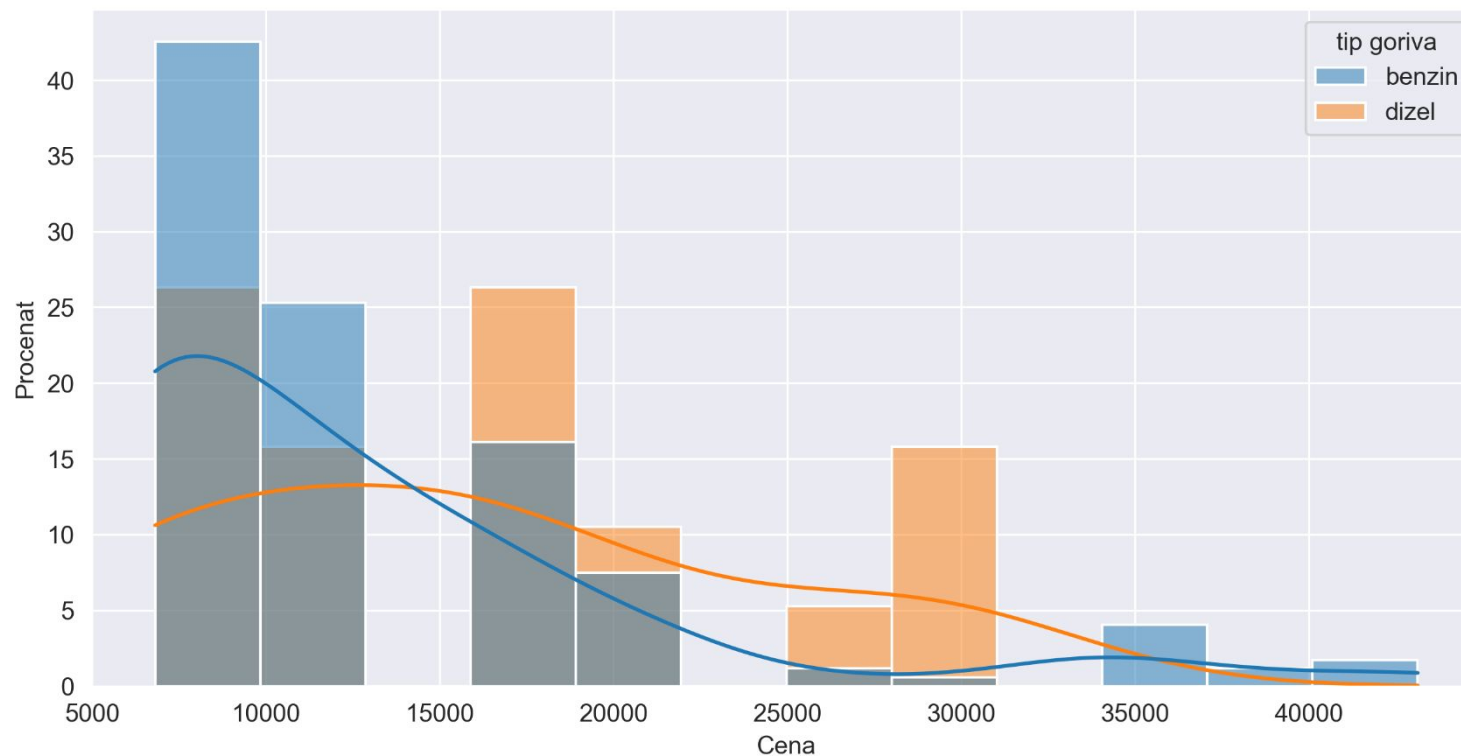


- Da li je tip goriva dobar pokazatelj cene?



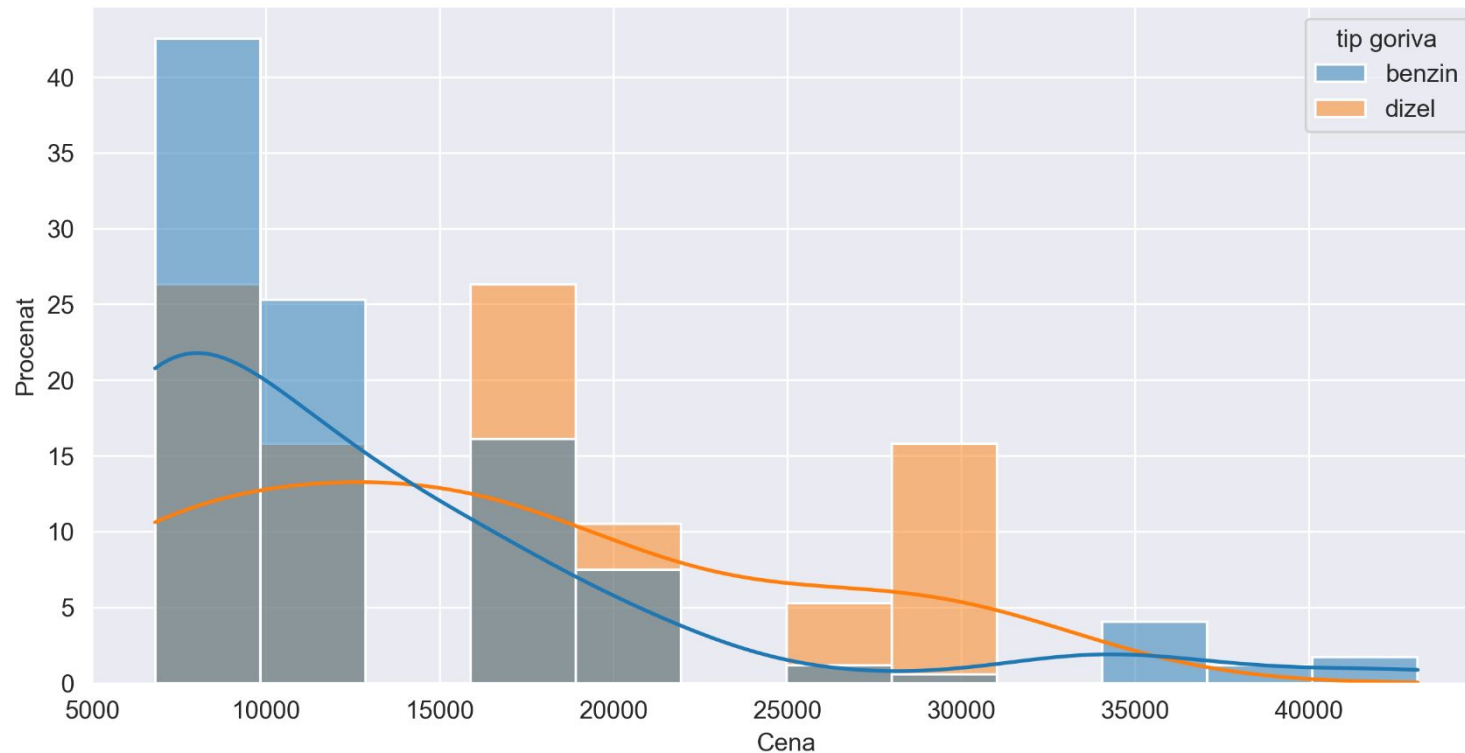
- Da li je tip goriva dobar pokazatelj cene?

Prosečna cena po gorivu:
Benzin ~ 13k
Dizel ~ 16k

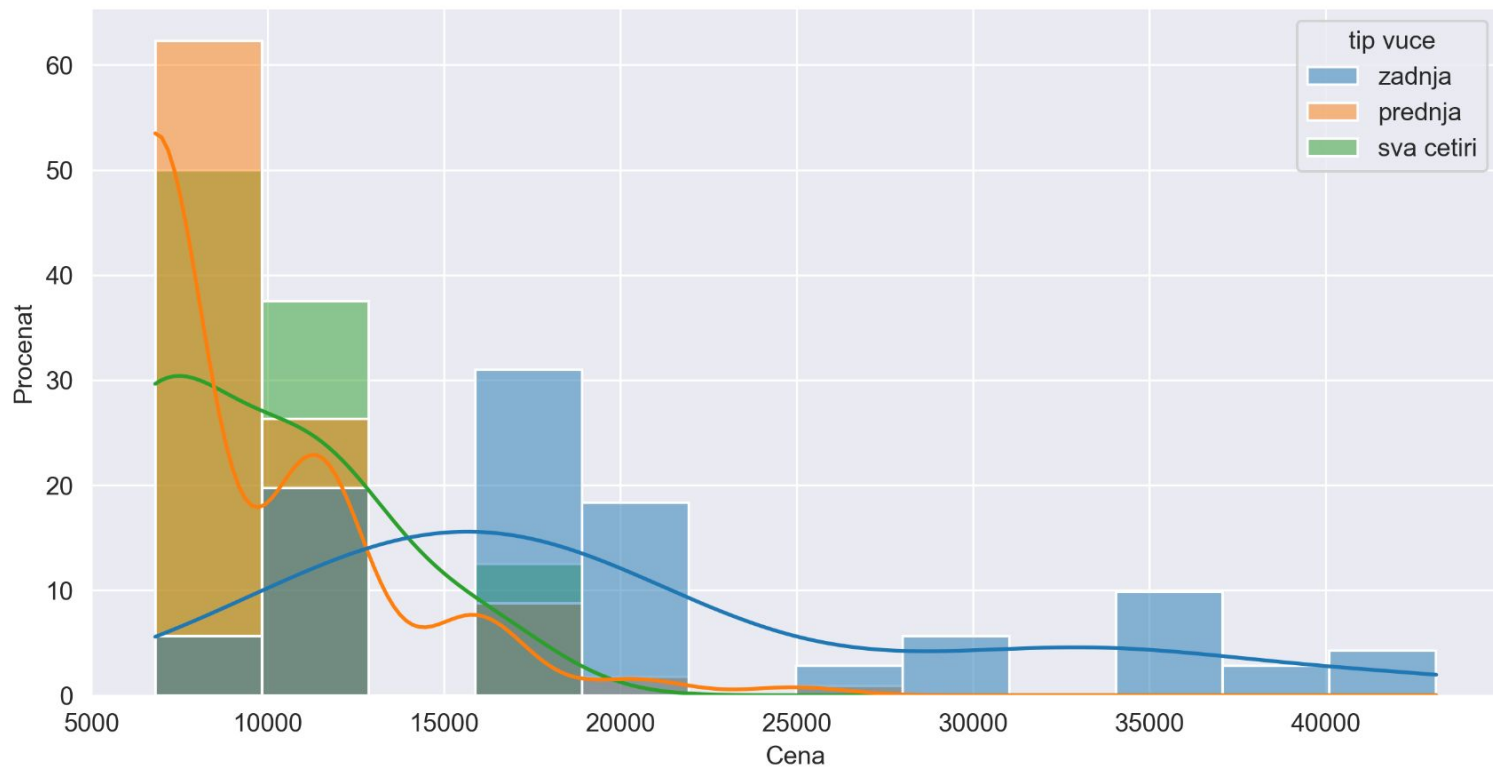


Prosečna cena po gorivu:
Benzin ~ 13k
Dizel ~ 16k

- Da li je tip goriva dobar pokazatelj cene?
- Nije! Ne donosi nam skoro nikakvo novo saznanje o ceni



- Da li je tip vuče dobar pokazatelj cene?



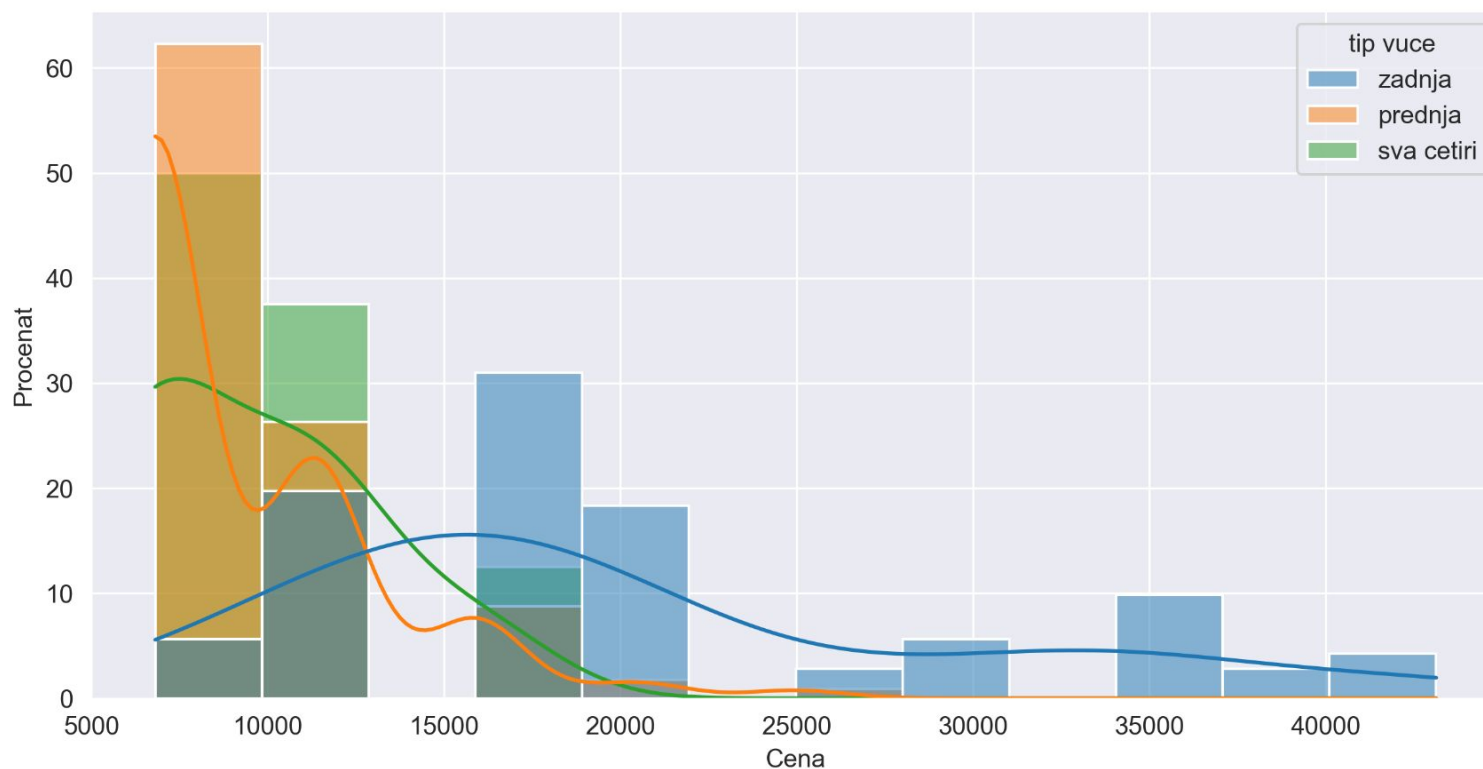
- Da li je tip vuče dobar pokazatelj cene?

Prosečna cena po tipu vuče:

Prednja ~ 9k

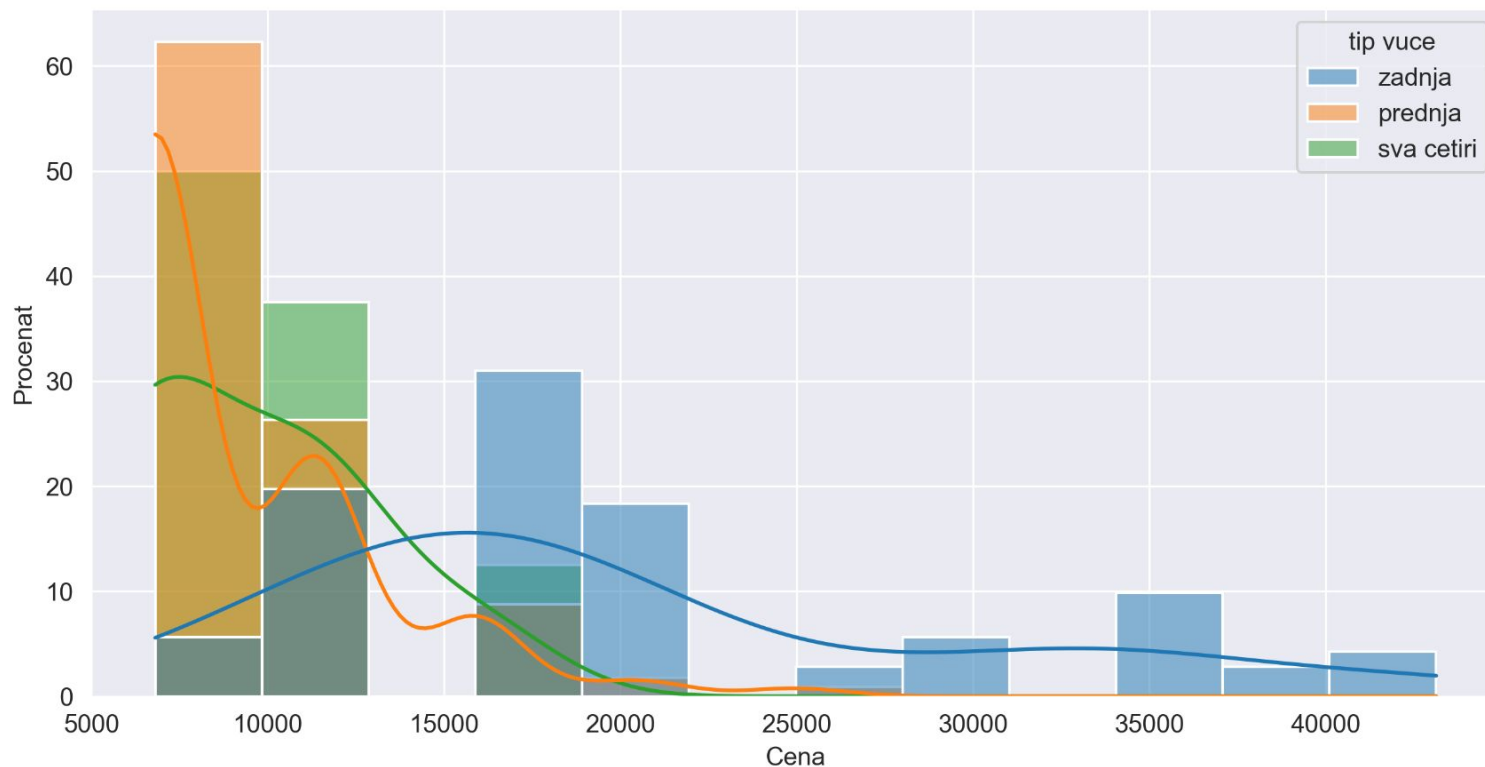
Sva 4 ~ 10k

Zadnja ~ 20k

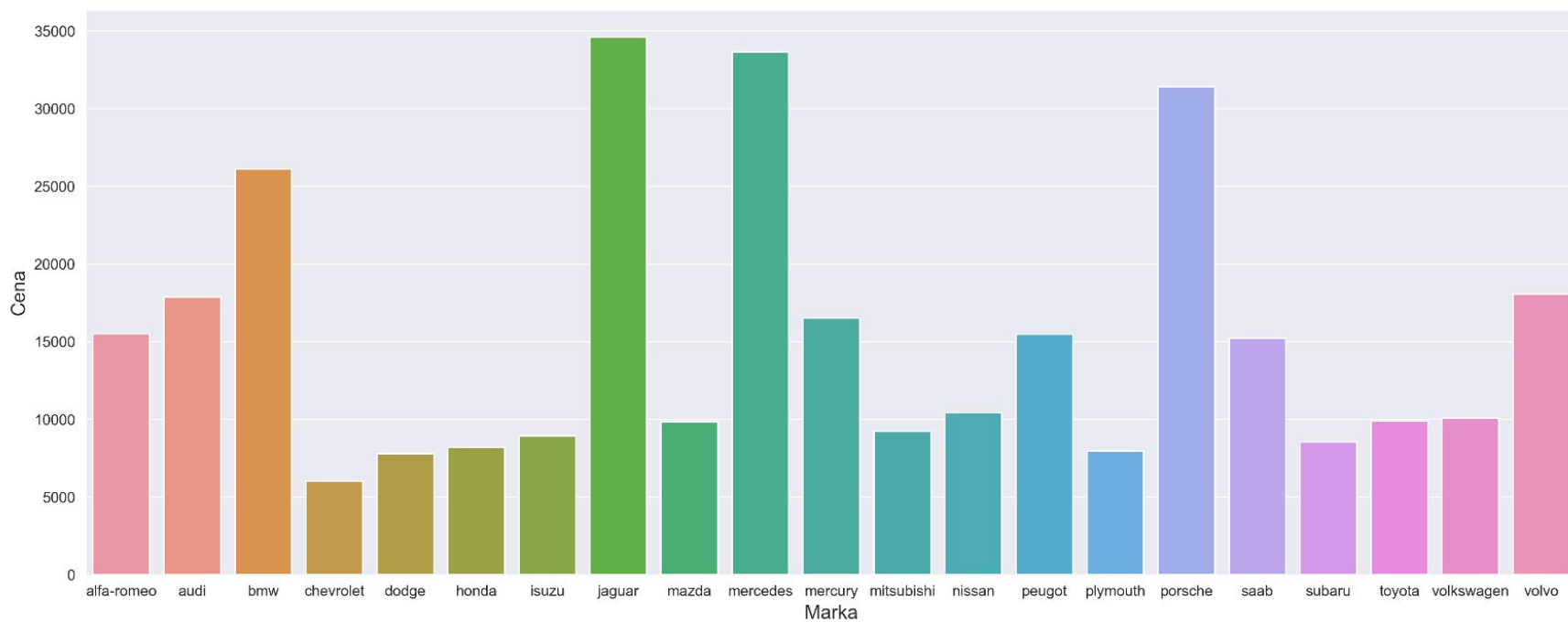


- Da li je tip vuče dobar pokazatelj cene?
- Jeste! Postoji očigledna razlika kod zadnje vuče

Prosečna cena po tipu vuče:
Prednja ~ 9k
Sva 4 ~ 10k
Zadnja ~ 20k



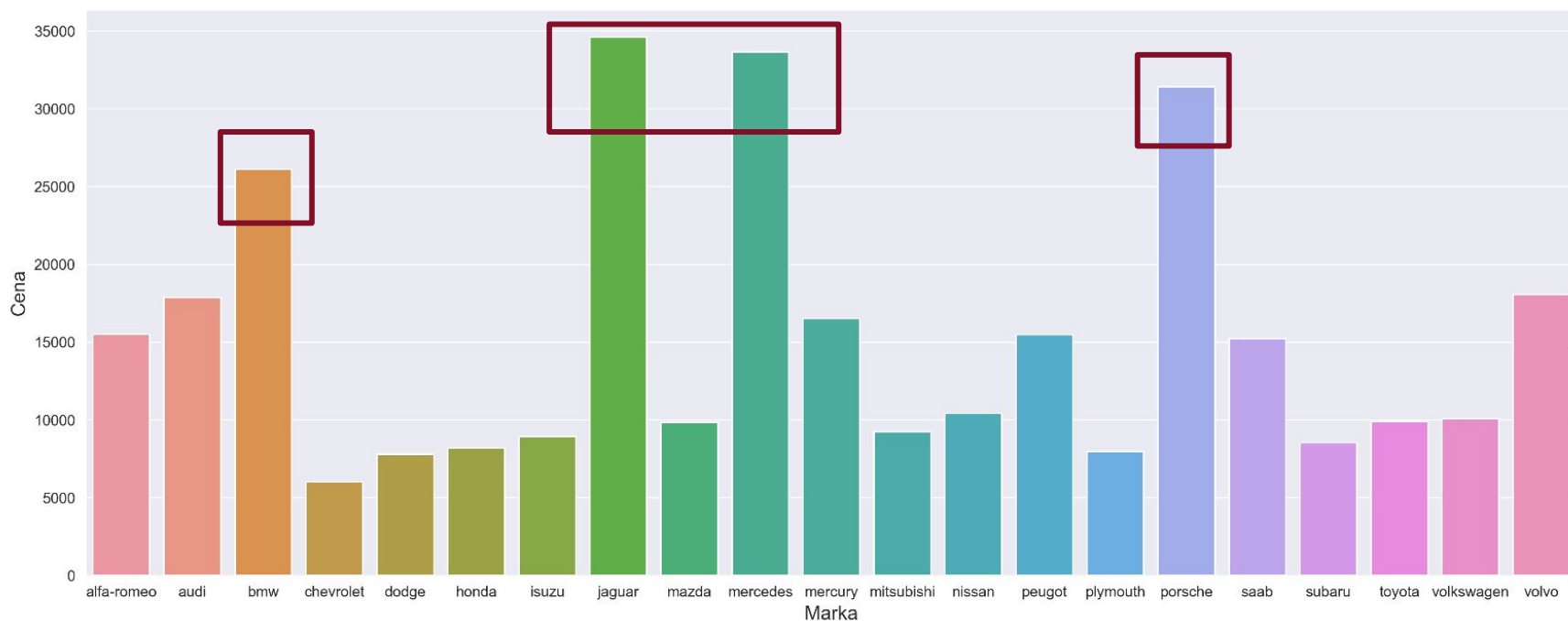
Prosečna cena po markama



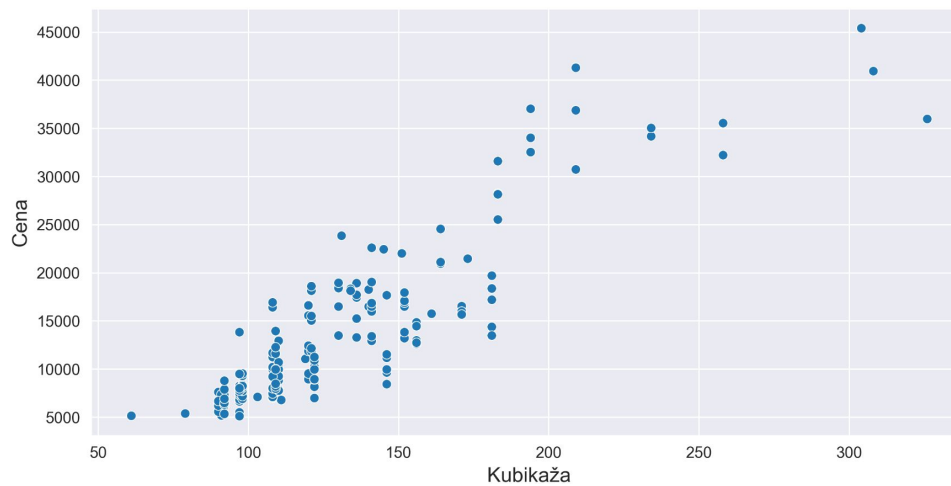
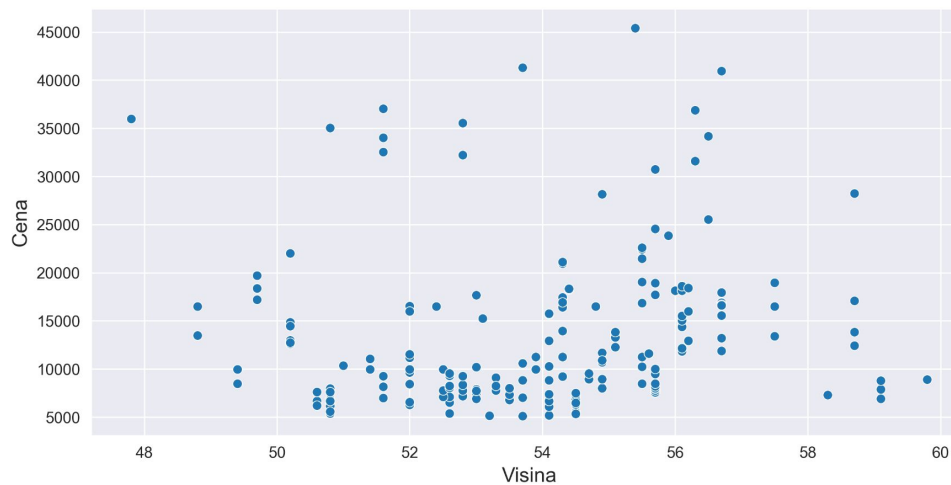
Prosečna cena po markama



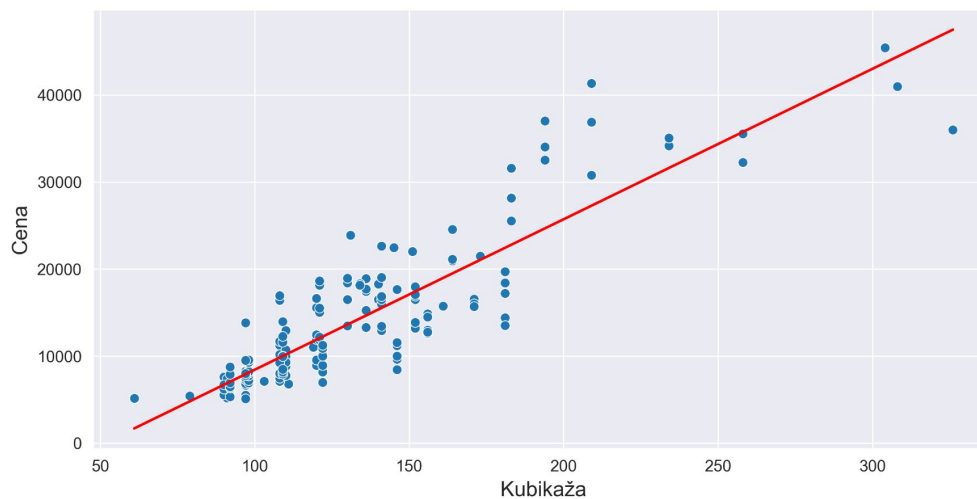
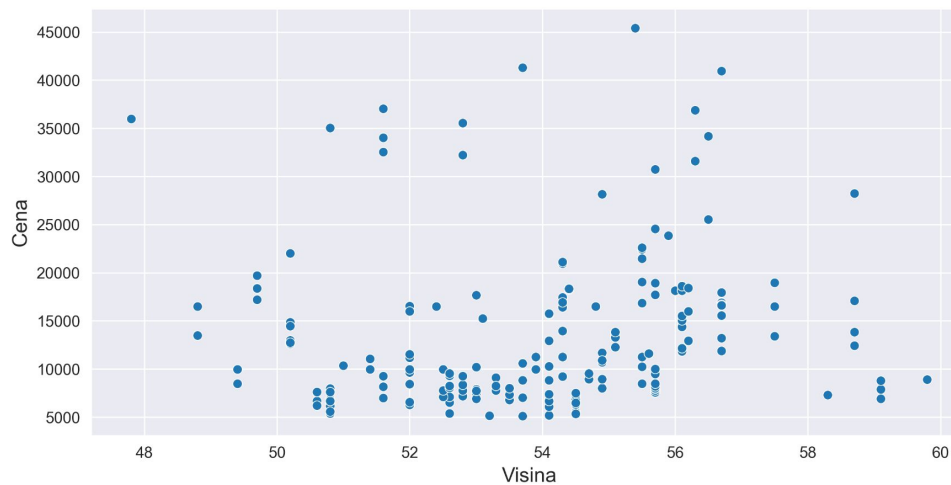
- Neke marke nam daju puno informacija o potencijalnoj ceni vozila



Gde je prisutna korelacija?



Gde je prisutna korelacija?



Kako „brže” da uočimo korelaciju?

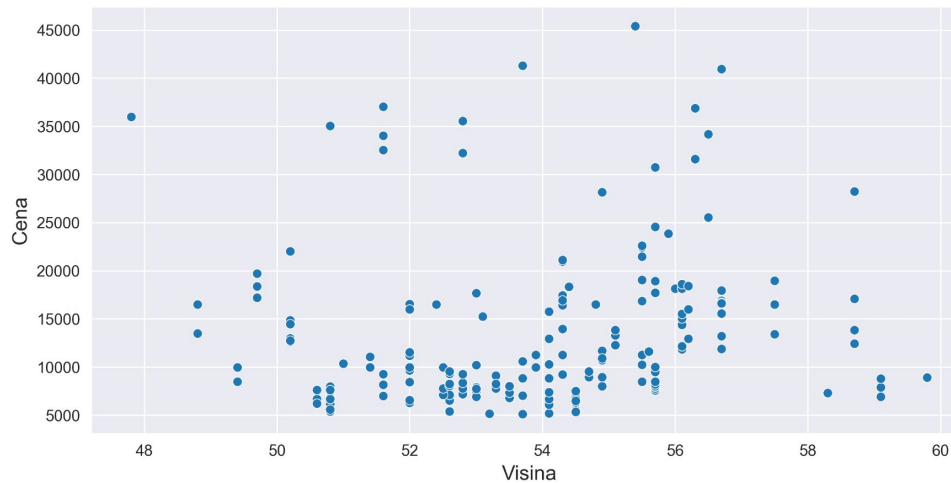


- Možemo da kvantifikujemo korelaciju - *Pearson* korelacija

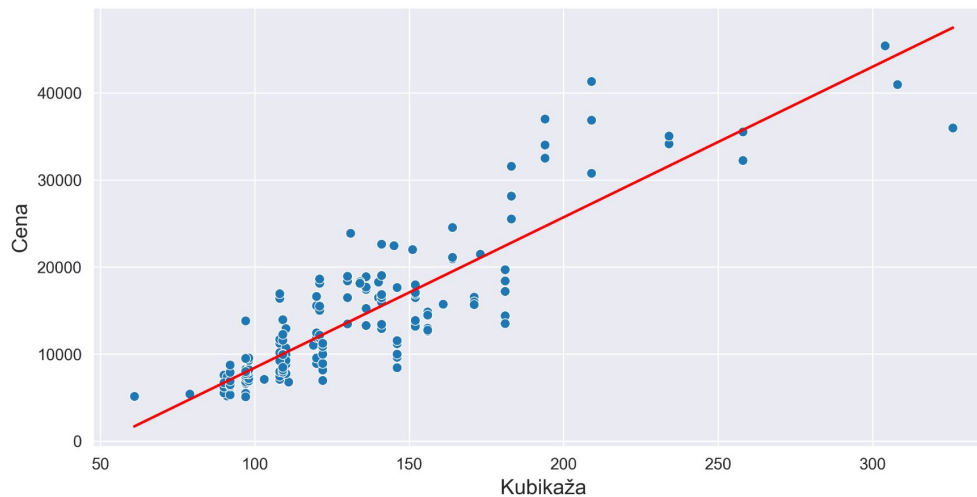
$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_x)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

- Vrednost uvek između -1 i 1
- Vrednosti blizu **1** - jaka pozitivna korelacija
- Vrednosti blizu **-1** - jaka negativna korelacija
- Vrednosti blizu **0** - nema korelacije

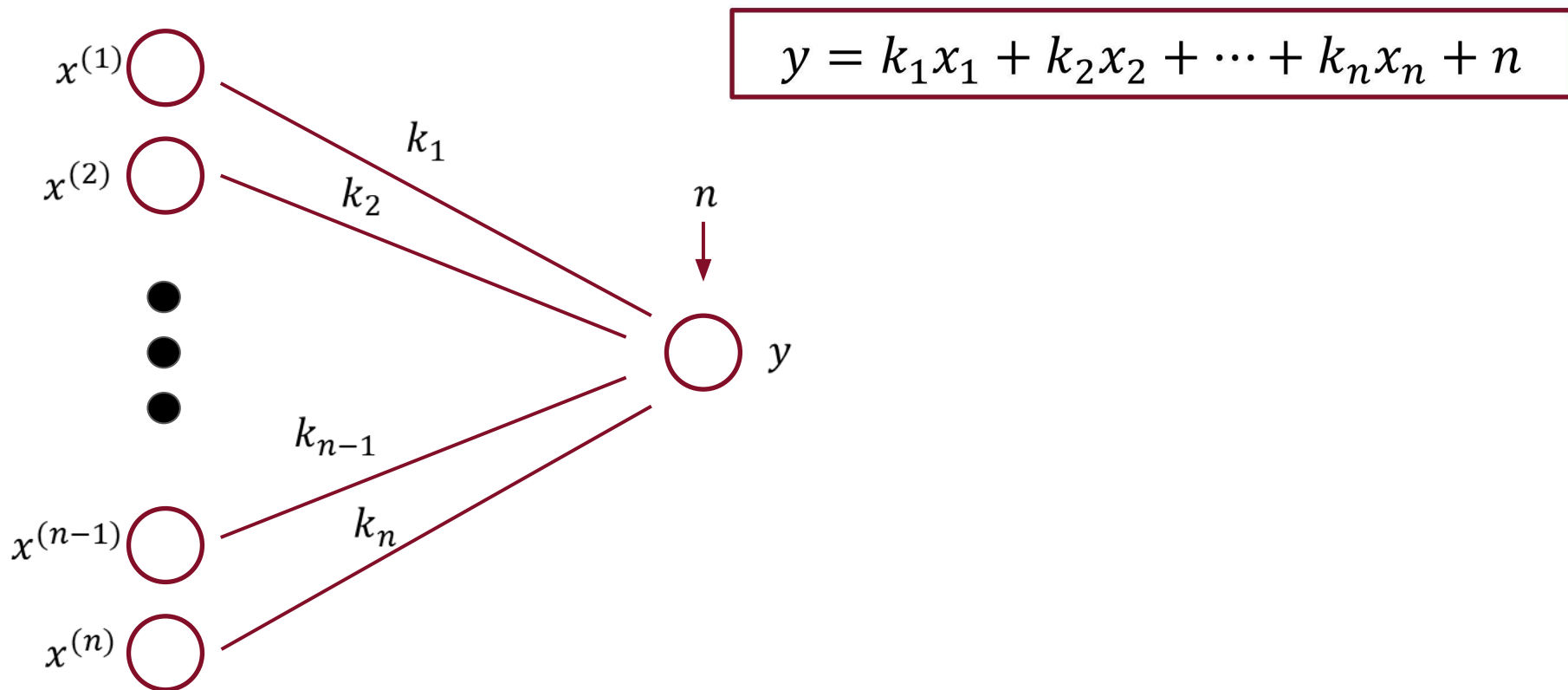
Kvantifikovanje korelacije



Korelacija između
cene i visine: **0.13**



Korelacija između
cene i kubikaže: **0.89**



Hvala na pažnji!

Pitanja?