

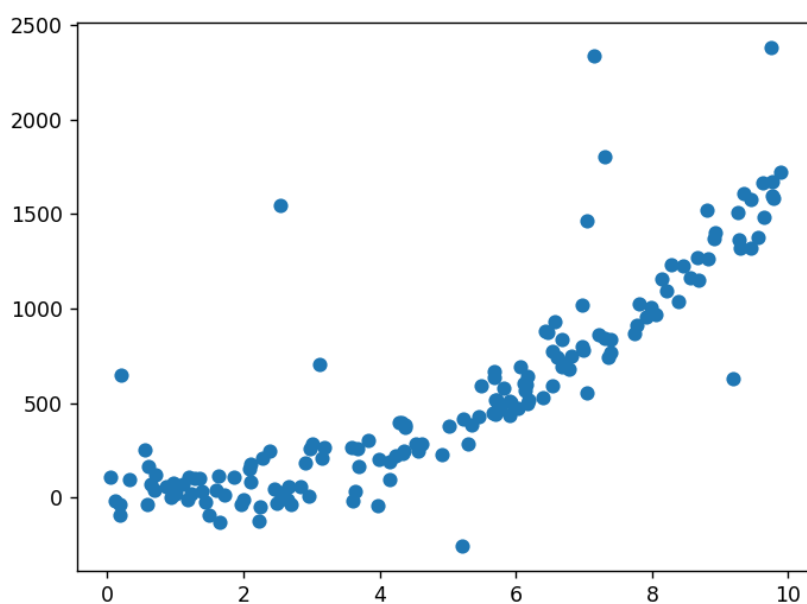
Jednostruka linearna regresija

- **Članovi tima:**

Mitrović Andrej SV81-2021

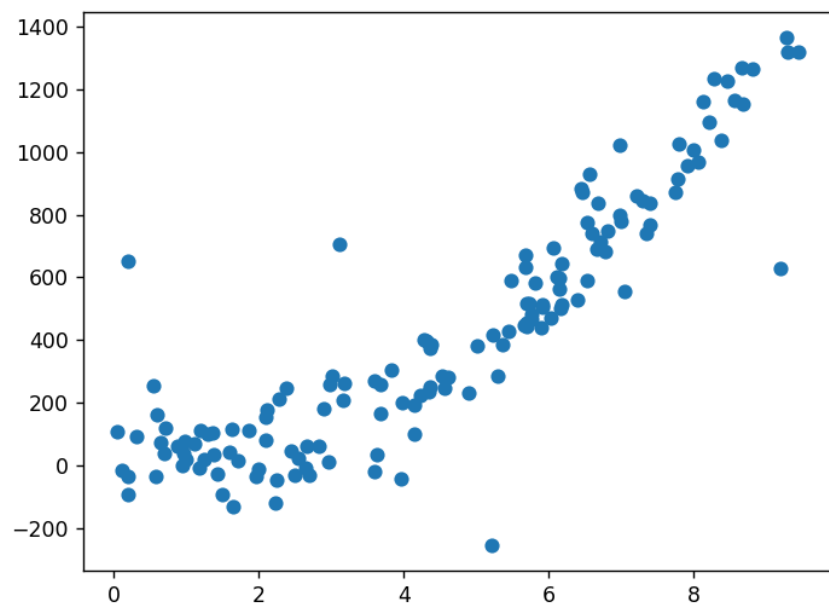
- **Pristup problemu:**

Nakon učitavanja podataka u pandas DataFrame, iscrtavamo testne podatke pomoću pyplot scatter funkcije:



Pomoću grafika možemo da primjetimo postojanje outlier-a. Potrebno je da prilagodimo podatke prije treniranja modela.

Uklanjanje outlier-e pomoću Z-score metode koja nam govori koliko je udaljena tačka od srednje vrijednosti u standardnim devijacijama. Za početak uzimamo $Z=3$ i analizom grafika pratimo promjene outlier-a u odnosu na različite Z vrijednosti.



- **Algoritmi optimizacije:**
 - Batch GD
 - Stohaistic GD
 - Mini-Batch GD

Testiranjem sva 3 algoritma na trening skupu podataka računamo RMSE rezultate.

Mini Batch GD je imao znatno manje RMSE i pokazao se kao najoptimalniji kombinujući pozitivne karakteristike BGD i SGD algoritama.

- **Zaključak:**
 - Test skup podataka nije imao nepostojeće vrijednosti pa je bilo potrebno samo srediti outlier-e.
 - Standardizacija podataka nije dovela do promjena u rezultatu RMSE.
 - Odabran je Mini Batch GD kao algoritam optimizacije nakon testiranja RMSE rezultata sva 3 Gradient Descent algoritma.
 - Krajnji RMSE rezultat nije bio <180
 - Moguće je dalje optimizovati funkciju za uklanjanje outlier-a