- Names and ID's of all participants:

Bartłomiej Umiński 397506, Dariusz Kęsicki 384759

- Short description of the topic and the web page:

We decide to scrape informations about offers of cars for sale from webpage www.gumtree.pl

- Short description of your scraper mechanics:

Using beautifulsoup, scrapy and selenium packages, we download data from gumtree.pl about the parameters of cars offered for sale. A detailed description of the code mechanics what each line does is in the code comments.

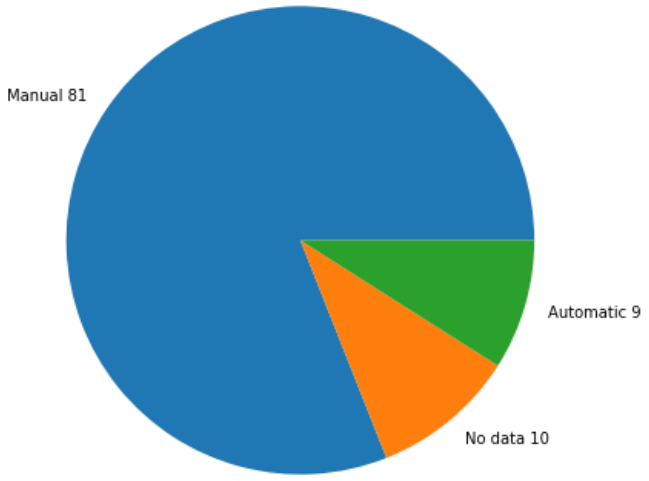
- Short technical description of the output you get:

Output of our code is a pandas table with 10 columns: name of offer, price, date when the offer was added, location, year of production, kilomiters, type of gearbox, brand of car, model and type of car.

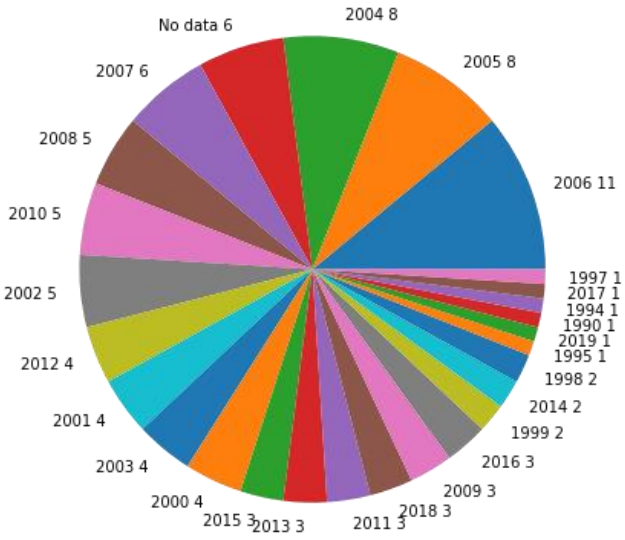
- Extremely elementary data analysis - you need to prove, that collected data can be used for further analysis, but nothing more (hard limit of data analysis: one page):

A separate sript was used for data analysis. Fisrt we count types of gearbox. From 100 first offers at page 81 had manual type of gearbox, 9 automatic and 10 offers had no data about gearbox. A pie chart was created to visualize the results. Then the years of car production from the offers were counted. The results were visualized on the chart.

Type of gearbox

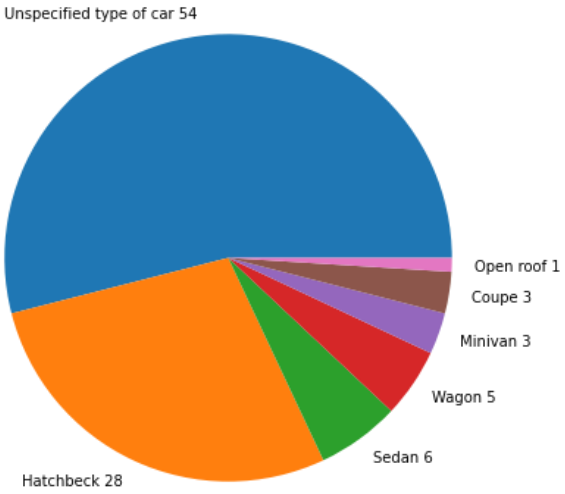


Year of production



Type of car was the last variable that was counted and the results are visualized in the graph.

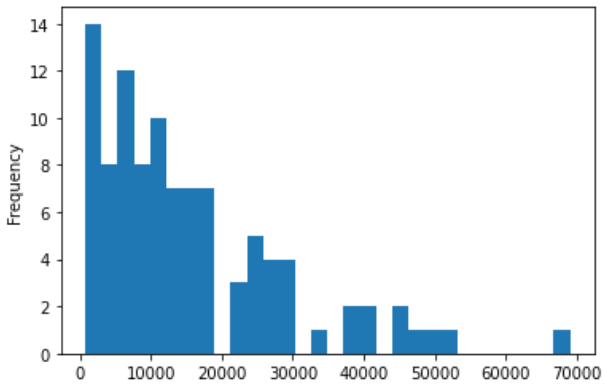
Type of car



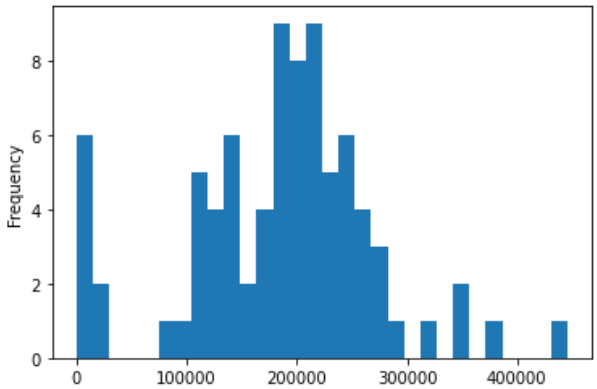
Brands were also counted. We obtain: Opel 11, No data 10, Mercedes-Benz 9, Renault 9,Audi 7 Volkswagen 5, Fiat 5, BMW 5, Citroen 5, Ford 4, Peugeot 4, Skoda 4,Honda 3, Hyundai 3, Others 3, Volvo 3, Daewoo 2, Kia 2, Seat 2, Subaru 1, Toyota 1, Nissan 1, Alfa Romeo 1.

Next the numerical data was analyzied. Minimal price of offer was 800 zł and maximum 69000 zł. Mean price was 15621 zł and standard deviation 13407 zł. First quartile value was 5875 zł, second 11100 zł and third 23500 zł. Minimal kilometers 1 and maximum 446000. Mean value of kilometers traveled was 184774 and standard deviation 87305. First quartile value was 135000, second 198000 and third 230000. Additionally, a histograms was drawn.

Price



Kilometers



- Detailed description which participant wrote which part of the project.

Bartosz Umiński: Beatiful Soup script, Selenium script

Dariusz Kęsicki: Scrapy script, description.pdf, Data analysis script