

WEB MINING FINAL PROJECT REPORT

Second Hand Car Data Scraper

**Yunus Emre Cilan - 160709008
Mehmet Bozyel - 160709016**

We brought together all the vehicle advertisements published on a sample website by web crawling. We have taken the necessary data from the web pages of the vehicle advertisements we have in the parcel. These data are the city, brand, model and price of the vehicles. After passing this data in filtering and analysis processes, a csv file was output where we give the average price of the vehicles in each city according to the brand and model. The project was developed to help vehicle sellers and buyers get an idea of price information.

Technologies used in the project:

- Python3
- beautifulsoup4
- urllib3
- pandas

We encountered blocking during testing on some websites. These sites detected us as bots and blocked us. That's why the way it works on every website is the same. We developed the project based on the <https://www.arabam.com/> website.

How It Works

First, we create the links of the pages with the listings according to the url format. We get the html of these pages by making a get request with urllib. Then we parse the html pages with beautifulsoup4 and add the link of each vehicle advertisement to the list.

With Urllib, we get the html of each post with get. We parse it with BeautifulSoup4 and add the vehicle features to the list according to the class names. We use replace to bring the features to the format we want. Then we turn our list into a dataframe with pandas and print it to the cities.csv file.

With the analyzer.py file, we take the average price of the same city, brand and model vehicles and print it to the output.csv file.

To run, run in the following order:

- pip3 install beautifulsoup4
- pip3 install urllib3
- python3 scraper.py
- python3 analyzer.py

City	Marka	Seri	Average Price
ADANA	Volkswagen	Passat	99083
ADANA	Volkswagen	Polo	97290
ADANA	Volkswagen	Jetta	141333
ADANA	Volkswagen	Golf	140450
ADANA	Honda	Civic	138600
ADANA	Honda	Jazz	100000
ADANA	Ford	Focus	121833
ADANA	Ford	Fusion	67900
ADANA	Ford	Fiesta	98500
ADANA	Ford	Taunus	19750
ADANA	Renault	Clio	74787
ADANA	Renault	R 12	25250
ADANA	Renault	R 9	24625
ADANA	Renault	Fluence	92625
ADANA	Renault	Megane	133312
ADANA	Renault	Latitude	130000
ADANA	Renault	Laguna	78500
ADANA	Tofaş	Doğan	30000
ADANA	Tofaş	Şahin	30966
ADANA	Chevrolet	Aveo	115000
ADANA	Fiat	Linea	80833
ADANA	Fiat	Egea	130500
ADANA	Fiat	Brava	43750
ADANA	Fiat	Palio	54250