**Driving-to-Success**

We are trying to set up a start-up car company. We are looking to buy used cars to either sell or give out for hire to clients. For this, it would be beneficial to analyse the existing used car market and figure out what type of car (brand,type,year, etc.) is the most financially sound to be invested in. For this we can web-scrape the data from a motor-selling website and analyse a Kaggle dataset of web-scraped cars. We will extract JSON and CSV files, transform with pandas, and load it into PostgreSQL database with SQLAlchemy.

**Using a CSV**

We obtained a csv 'vehicles' from kaggle https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data to be used as a dataset.

Using python we were able to transform it using the following steps:

a) Imported it into pandas and displayed the data that it contains.

b) Cut out the columns that were unnecessary and not required.

c) Dropped columns that did not have any values.

d) Renamed the columns. From that, the data frame was ready to be stored in a database

**Web Scrapping**

To scrape the web we selected 'https://www.autotrader.co.uk/'. For this, we intended to scrape and get information from it.

To do this we performed the following:

a) Accessed the page using BeautifulSoup and got the HTML in JSON format.

b) Scraped the results and stored the information in lists that contained the details that were required. These were stored in lists.

c) Using webdriver we “clicked” on the “Next Page” button and scrapped webpages until we were satisfied with the volume of data

d) From the lists, the data was input and stored in a data frame.

e) The data frame was ready to be transformed into the PostgreSQL Database.

**Connection to Local Database**

We selected PostgreSQL Database as opposed to MongoDB since the data we would be getting would be relational. This is better accessed in tables. We did the following:

a) Created the 'cars\_db' database on the PostgreSQL

b) Created the 2 tables that would hold the information from the Data Frames. c) Created a connection in pandas to the Data Base

d) Linked the Data Frames in pandas to the tables that had been created and displayed the information contained in the tables.

**Responsibilities Outline:**

CSV – Kantai Melau

Web Scrapping – Yevgeniy Denkovych

Connection to Local Database - Nnamdi Okoroji

Data Cleansing – Mihai Florea

