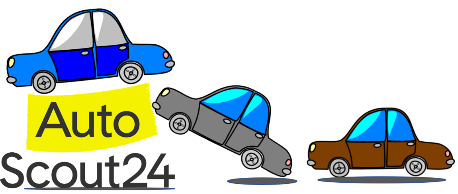
**Exploratory Data Analysis (EDA) Project**

[](https://camo.githubusercontent.com/d7751b01ccfca108cfe104168c9eedd9be63eb0c9ee867cfe01827f60b244daa/68747470733a2f2f646f63732e676f6f676c652e636f6d2f75633f69643d31464f63754539416b417163304a553664556c792d63756f556733624f70393248)

**EDA for Car Price Prediction Model**

**Descriptions :**

* A .json file containing a dataset consisting of 15919 rows and 54 columns is provided.
* This dataset, scraped from the on-line car trading company in 2019, contains many features of 9 different car models.
* The features (variables) of this dataset are too messy and distored.

**What is expected of you?**

* Read the .json file and assign the dataset into a DataFrame using pandas.
* Implement all aspects of the **EDA process** to the dataset.
  + Fix corrupted data formats,
  + Handle with outliers and missing values,
    - Domain (automobiles) knowledge is important.
    - Always use the internet to do the research that you need.
    - Think carefully to decide whether a data is outliers or not. Examples :
      * There is no conventional car model with an average fuel consumption of 1 - 1.5 liters per 100 km.
      * Or you need to know that it cannot be a 300 euro car.
      * Or if there is only one car with 3 doors out of the 15919 cars, this is what you should pay attention to and examine.
  + Drop the columns / rows you determined unnecessary as a result of your analysis,
  + Use visualization tools while doing all these processes.
* As a result, get the dataset ready to provide an appropriate input to the ML models.
* Save cleaned dataset into a .csv file.

**Need to Study :**

* .str.method,
* .contains(),
* .extract(),
* .to\_datetime(),
* .get\_dummies(),
* .add\_prefix(),
* .sample(),
* regex,
* .to\_numeric(),
* .isin(),
* .corr().

## Introduction

Welcome to "**AutoScout Data Analysis Project**". This is the capstone project of **Data Analysis** Module. **Auto Scout** data which using for this project, scraped from the on-line car trading company in 2019, contains many features of 9 different car models. In this project, you will have the opportunity to apply many commonly used algorithms for Data Cleaning and Exploratory Data Analysis by using many Python libraries such as Numpy, Pandas, Matplotlib, Seaborn, Scipy you will analyze clean dataset.

The project consists of 3 parts:

* First part is related with 'data cleaning'. It deals with Incorrect Headers, Incorrect Format, Anomalies, Dropping useless columns.
* Second part is related with 'filling data'. It deals with Missing Values. Categorical to numeric transformation is done.
* Third part is related with 'handling outliers of data' via Visualisation libraries. Some insights are extracted.