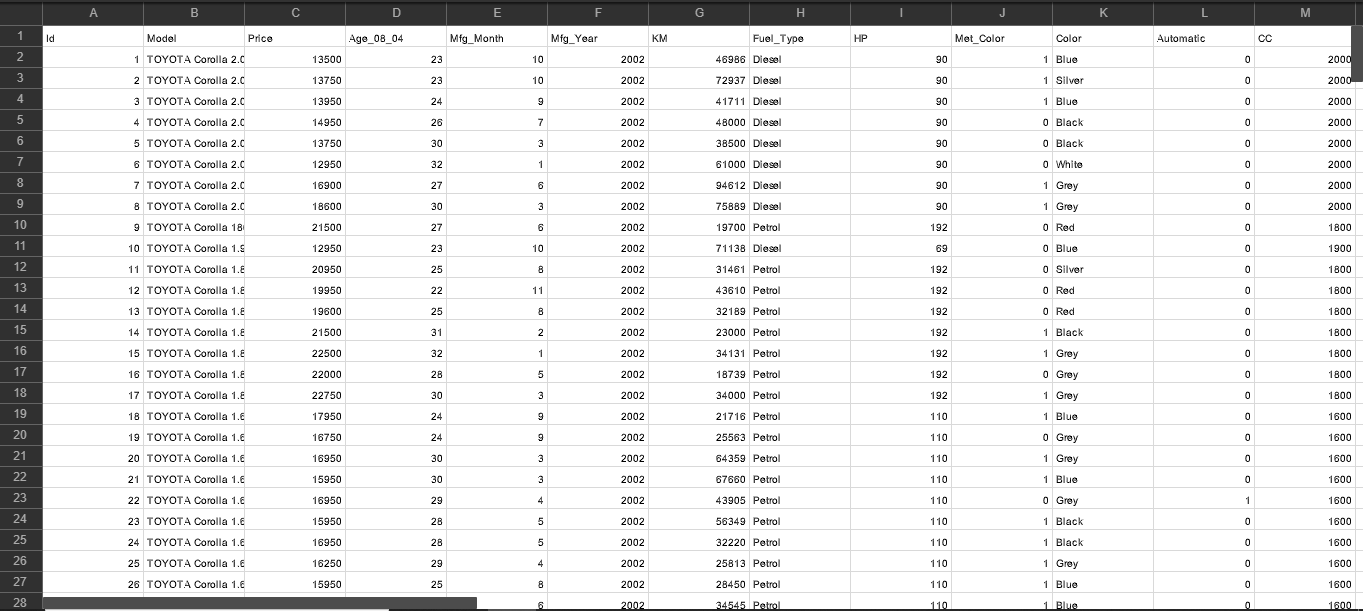
Data Analytics in Automotive Industry

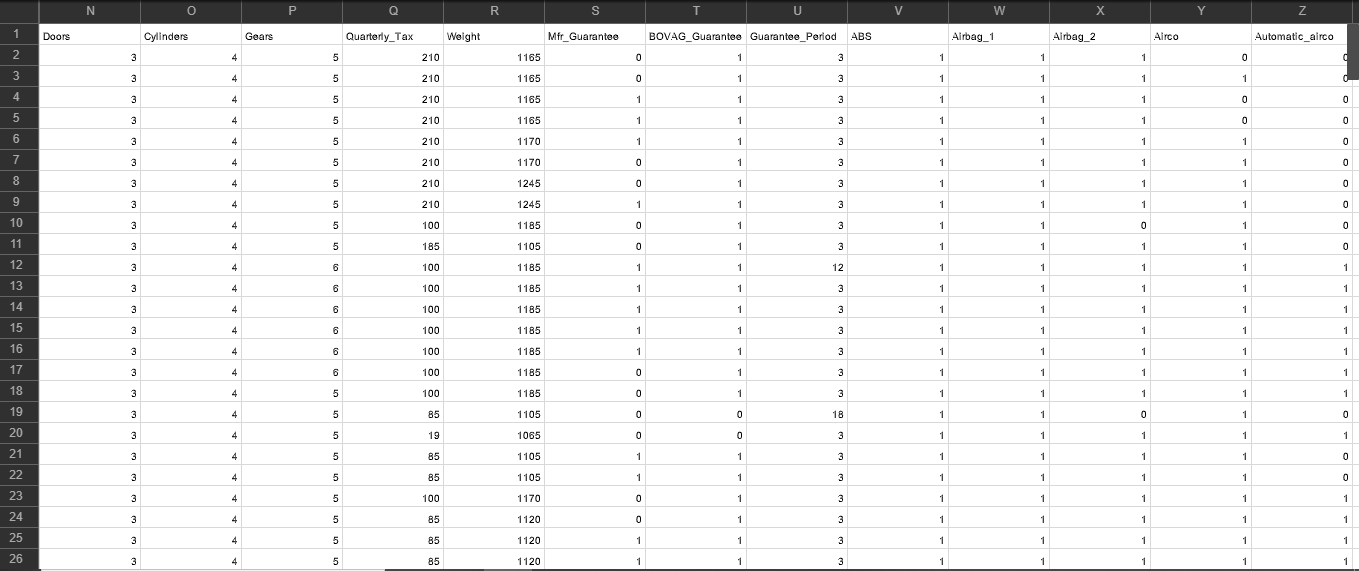
Udit Manish Parekh

**Data Analytics In Automotive Industry**

**Exploring the data with the tool**

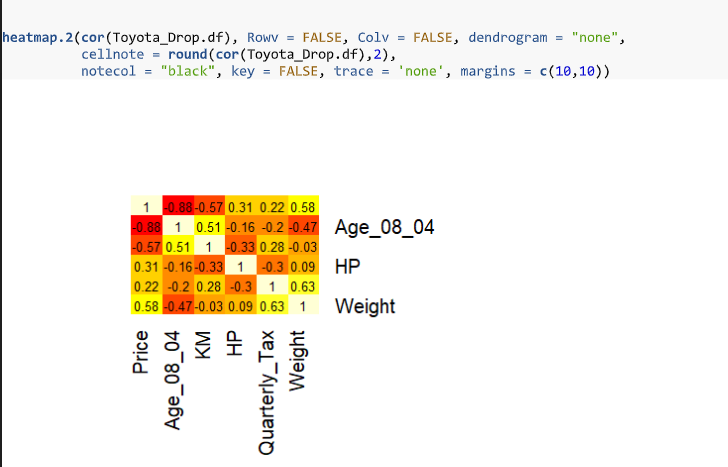
In this project, I reviewed a dataset that includes various variables such as the model of the vehicle, the price, the age, time of manufacturing, fuel type, color, CC, gear, guarantee period, airbag, and other numerous variables. The data is corrected by automotive retailers and analyzed to generate interesting details about the relationship between the various variables. For instance, automotive retailers evaluate the relationship between the age of the car and its price. Such insights help in the pricing of the car as well as determining other factors such as market conditions and the ideal time to initiate the sales. The data is also used to generate insights such as the relationship between the guarantee period of the car and its age, as well as the age of the car and its KM. The figure below is a screenshot of the data used in this study.



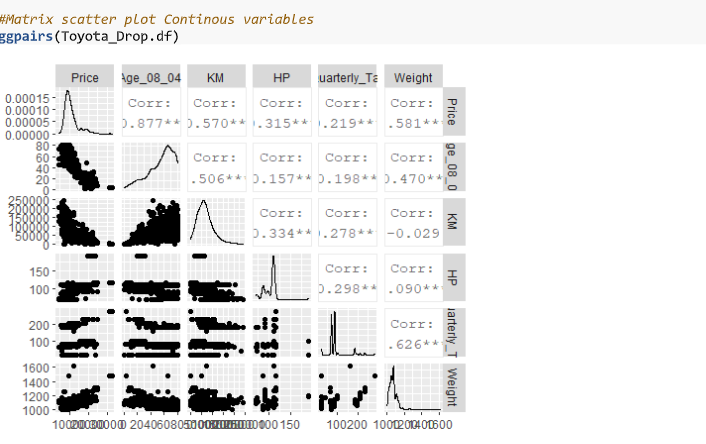


**Classifications Basic Concepts and Decision Trees**

The main technique use to process data is correlation analysis. The analysts focused on identifying relationships between the different variables in the dataset. The programs used to analyze the variables depending on a set of parameters. For instance, the figure below is an example of a code used to analyze relevant continuous variables. The variables processed in this program are price, age, weight, HP, and quarterly tax.



The figure below is a scatter plot for continuous variables.



**Classifications Alternative Techniques**

The dataset above could also be classified using other classification techniques. Rule-based classifiers are some of the typical classification techniques, and they process data by a collection of "it…then…" rules. Rule-based classifiers can be characterized into mutually exclusive rules and exhaustive rules depending on the possible combination of attribute values. Another typical classification alternative technique is the direct method that directly extracts rules from data. Some of the techniques in the direct method are RIPPER, CN2, and 1R.

**Summary of Results**

Analyzing the variables generated some interesting correlations. For instance, age is negatively correlated with price. This means that the older the car, the lower its price. Similarly, the price of the car has a negative correlation with the distance traveled in KM. Other interesting insights generated from the analytical project indicate that weight and height have a positive correlation, KM and age also have a positive correlation, and quartile tax also has a positive correlation with weight.