DATA SCIENCE BOOTCAMP
PROJECT PROPOSAL
USED CAR PRICE PREDICTION

BAYAN JABER ALMUZAINI

QUESTION/NEED

Due to the increased price of new cars over the past years and the incapability of some customers to buy new car, there is a huge demand for used cars. Therefore, the used car price prediction system becomes important in order to grow the market, and could help to effectively determine the worthiness of the car using a variety of features.

The model aims to help the used car sellers to offer a better service by assisting them to understand what makes a car desirable, and what features are essential for a used car. Likewise, individuals who are interested in either buying or selling a used car could get benefit from the model.

Some of the questions that frame the analysis:

- Does Location in which the car being sold has any effect on the price?
- Does kilometers driven, Year of manufacturing have negative correlation with price of the car?
- How does number of seat ,Fuel type effect the pricing?

DATA DESCRIPTION

The data used in this project was downloaded from Kaggle. The dataset has almost 7000 rows and 14 different features as following:

***1. S.No.:** Serial Number*
***2. Name:** Name of the car which includes Brand name and Model name*
***3. Location:** The location in which the car is being sold or is available for purchase.*
***4.Year:** Manufacturing year of the car*
***5. Kilometers_driven:** The total kilometers driven in the car by the previous owner(s) in KM.*
***6. Fuel_Type:** The type of fuel used by the car. (Petrol, Diesel, Electric, CNG, LPG)*
***7. Transmission**: The type of transmission used by the car. (Automatic / Manual)*
***8. Owner:** Type of ownership*
***9. Mileage:** The standard mileage offered by the car company in kmpl or km/kg*
***10. Engine:** The displacement volume of the engine in CC.*
***11. Power:** The maximum power of the engine in bhp.*
***12. Seats:** The number of seats in the car.*
***13. New_Price:** The price of a new car of the same model*
***14. Price:** The price of the used car.*

The link to the dataset: https://www.kaggle.com/yogidsba/predict-used-car-prices-linearregression/notebook

TOOLS

In order to meet the requirement of the project the following tools are intended to be used:

Data Processing: *Pandas, Numpy.*

Modeling: *SciKit-Learn*

Visualization: *Matplotlib, Seaborn.*

The following algorithms are considered to be used to build the model:

- ***Linear Regression***
- ***Random Forest***

MVP GOAL

Hopefully the minimum viable product for this project would answer some of the proposed questions. And provide some useful and functional initial results which will help in confirming that the model is on the right track.