

T5 – Data Science Bootcamp



Project Proposal

Predict the price of a car – Linear regression



Razan Aljuhani

Problem statement

Companies try to produce cars with special features to take attention in the market and compete with other competitors. They need to understand the dynamic changes of pricing in the market through the years, and the factors affecting the pricing of cars in the market.

Proposed solution

The project will build an effective model using Linear regression that prediction the price of a used car depending on its features and attributes. So, the companies can control the design of the cars, and make professional plan of business to meet specific price levels.

Aim and goals:

The project aims to apply the learned skills from Data Science Bootcamp provided by SDAIA. It will build a predictive model of the used car prices by achieving the following goals:

- Identify significant attributes in the car that increase its price.
- Prediction of the price of a particular car based on its attributes.

Dataset

Vehicle dataset that consists of data about the used cars listed on this Indian website www.cardekho.com. The third version of dataset was updated in 2020 and contain 8111 data points and 13 attributes.

The given attributes in the dataset:

1. **name:** Car full name.
2. **year:** Year of selling car
3. **selling_price:** The cost of car. (Target in the prediction model)
4. **km_driven:** Number of kilometers the car is driven
5. **fuel:** Diesel or petrol.
6. **seller_type:** Sold by individual or dealer.
7. **transmission:** Automatic or Manual
8. **owner:** Number of previous owners.
9. **mileage:** Mileage of the car.
10. **engine:** Engine capacity.

11.max_power: Max power of engine

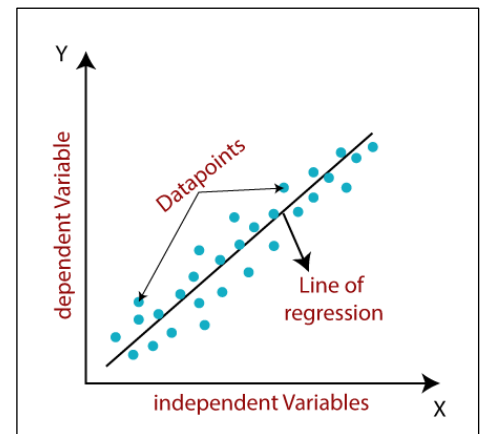
12.torque: Torque of the car

13.#seats: Number of seats in the car

Reference: <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho?select=Car+details+v3.csv>

Algorithm

Linear Regression is an ML algorithms and Familiar in the data science field for analyzing and manipulating data. It defined a linear relationship between a dependent variable and the other given independent variables. In real life, linear regressions can be used in developing businesses to predict, evaluate trends and determine estimates.



Tools

The needed tools to satisfy the project requirements:

- **Environment:** Jupyter Notebook.
- **Programming Language:** Python.
- **Data processing:** Pandas, NumPy.
- **Modeling:** Sickit-Learn.
- **Visualization:** Seaborn, Matplotlib.

MVP Goal:

The project mainly goes through different steps including:

- 1- Preprocessing.
- 2- Modeling.
- 3- Visualization.
- 4- Testing.