

Second Project (Proposal)

SDAIA T5 Bootcamp

Used Cars

Done by:
Aljoharah Alaqeel
Nasser Alquraini
Raghad Alnasser

Introduction:

Cars are one of the most important transportation in our life. Using your own car and not being dependent on public transportation can be convenient, economic, and saves you time.

In this project our dataset contains information about used cars listed on www.cardekho.com. This data can be used for a lot of purposes such as price prediction to exemplify the use of linear regression in Machine Learning.

Resource:

I got the dataset from Kaggle website

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

About this data:

This Data contains 13 columns and 8129 rows.

Description of Dataset:

Feature	Description
Name	Name of the car and its model.
Year	The year of manufacture.
Selling_price	How much the car was sold out.

Km_driven	Indicates the number of kilometers the car traveled.
Fuel	The type of fuel, diesel, or petrol.
Seller_type	Who is the seller of the car, individual or a dealer.
Transmission	Type of transmission manual or automatic.
Owner	Number of people owned the car.
Mileage	Indicates the number of miles the car traveled.
Engine	Type of engine of a car.
Max_power	Maximum power of a vehicle.
Torque	Torque of a vehicle.
Seats	Number of seats the vehicle has.

Tools:

- Jupyter
- Pandas for data manipulation.
- Matplotlib for visualization.
- Sklearn
- Seaborn
- Linear Regression
- Decision Tree
- Word
- Numpy
- Github
- Prezi

Questions this project will answer:

1. What is the highest car preferred by fuel type?
2. What is the most selling car by price?
3. Number of seats by type of car?
4. What is the highest transmission type?
5. What is the highest km driven by type of car?

Question for modeling using linear regression and decision tree:

- 1- Does year, selling price, kilometer driven, fuel, seller type, transmission, owner, engine, seats affect the selling price?