Group Members

Md Hammad Rasheed

Md Zeeshan Rasheed

Assignment: Capstone_Project

Email Id

mdhammadr@gmail.com

mdzeeshanrasheed5@gmail.com

Phone Number

+91 8789322372

+91 8863026181

Dataset: CAR_DETAILS.csv

Objective:

Based on dataset we have to build Machine Learning model to Predict Selling Price of Car.

Dataset Preview:

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
2	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
3	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner
4	Hyundai Verna 1.6 SX	2012	600000	100000	Diesel	Individual	Manual	First Owner
5	Datsun RediGO T Option	2017	250000	46000	Petrol	Individual	Manual	First Owner
6	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
7	Maruti Alto LX BSIII	2007	140000	125000	Petrol	Individual	Manual	First Owner
8	Hyundai Xcent 1.2 Kappa S	2016	550000	25000	Petrol	Individual	Manual	First Owner
9	Tata Indigo Grand Petrol	2014	240000	60000	Petrol	Individual	Manual	Second Owner
10	Hyundai Creta 1.6 VTVT S	2015	850000	25000	Petrol	Individual	Manual	First Owner

STEP-1:

Analyzing Dataset and there are 8 columns and 4341 rows

- Important columns for predicting selling price is based on fuel, seller_type, transmission, owner and Brand.
- Extracted Brand column in the dataset using Excel

Dataset Preview:

	name	Brand	year	selling_price	km_driven	fuel	seller_type	transmission	owner
2	Maruti 800 AC	Maruti	2007	60000	70000	Petrol	Individual	Manual	First Owner
3	Maruti Wagon R LXI Minor	Maruti	2007	135000	50000	Petrol	Individual	Manual	First Owner
4	Hyundai Verna 1.6 SX	Hyundai	2012	600000	100000	Diesel	Individual	Manual	First Owner
5	Datsun RediGO T Option	Datsun	2017	250000	46000	Petrol	Individual	Manual	First Owner
6	Honda Amaze VX i-DTEC	Honda	2014	450000	141000	Diesel	Individual	Manual	Second Owner

Saved the file in .csv format.

I found that the dataset have Brand is one of the most important part for selling price of any car model with this number of columns is 9 and number of rows is 4341

Step 2:

For selling price prediction , we have 7 most important our dataset.

Brand	:	Tata, Maruti, Audi, Skoda, Fiat, Honda etc.
Driven (KM)		
Year		
Transmission	:	Automatic, Manual.
Seller type	:	Dealer, Individual, Trustmark Dealer.
Owner type	:	First Owner, Second Owner, Third Owner, Fourth and Above Owner, Test Drive Car.
Fuel type	:	Petrol, Diesel, LPG, CNG, Electric.

Step 3:

Analyzing the data types of columns of dataset.

Brand	:	Object
Driven (KM)	:	Integer
Year	:	Integer
Transmission	:	Object
Seller type	:	Object
Owner type	:	Object
Fuel type	:	Object

Step 4: Proceed with python programming.

Libraries (used to build the model):

Pandas

Pandas is a Python library used for working with data sets.
It has functions for analyzing, cleaning, exploring, and manipulating data

NumPy

NumPy is a Python library used for working with arrays.
It also has functions for working in domain of linear algebra, fourier transform, and matrices.

Scikit-learn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python.
It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Step 4: Continue…

Matplotlib

Matplotlib is a cross-platform, data visualization and graphical plotting library (histograms, scatter plots, bar charts, etc) for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB.

Seaborn

Seaborn is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on top matplotlib library and is also closely integrated with the data structures from pandas.

Step 5: Building a Streamlit web app

GitHub Repository Link

Click Here

Streamlit Web App

Click Here