

Car Price Prediction

Submitted by:

Bharathi B V

**ACKNOWLEDGMENT**

With great pleasure, I am sincerely express our deep sense everlasting profound gratitude and heartfelt thanks to several individuals from whom I received impetus motivation and invaluable work during the internship project work. I am very grateful to Sajid and Keshavinternship 17th batch guide, for the help and I do express our deep gratitude to you both for able to guidance, keep interest and constant encouragement thrs.

**INTRODUCTION**

* Business Problem Framing

To analyze the the details of used cars based on different categories.

* Conceptual Background of the Domain Problem

Here, we are train the model based on the data of various cities and we will find out To analyze the details of used cars based on different categories.

* Review of Literature

All possible information from all the available data tables more the information, more than for EDA and feature Engineering. its take more important to take the average during the aggregation of data from tans the table rather than taking the counts before the loan was applies no future information steps into the data to be used for modelling.

* Motivation for the Problem Undertaken

This is useful to number of peoples who are buying to used cars gives a clarity on each

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

1 Included in ‘Data-cleaning ipynb’.

2 Selecting relevant features.

3 Exploratory the data analysis and data cleaning.

4 Null value imputation

5 Handling outliers

6 Training a machine learning model

7 Hyperparameter tunning

8 Evaluate the model

9 prediction of the model

* Data Sources and their formats

Creating index on all the table to be joined this will speed up and processing here we are using EDA feature Engineering, Data visualization and statics approach perform data cleaning, outlier handling, missing values build model etc.

* Data Preprocessing Done

Partitioning and splitting that dataset account when credit loan default was applied as that dataset is skewed, stratification is used allocate the samples evenly based on sample classes so that training set and test set have similar ration of classes

* Data Inputs- Logic- Output Relationships

1. Dataset is highly imbalanced we can use our models are majority and minority logic input and outputs.
2. Majority classes-the dataset is too small. down sampling the majority class will not help, so we will up sample the minority class.
3. Minority – we can balance the dataset either by up sampling the minority class or down sampling the majority class.

* State the set of assumptions (if any) related to the problem under consideration

I am not taken any presumptions of this problem.

* Hardware and Software Requirements and Tools Used

1 processor=AMD Ryzen 3 3200U Radeon Vega Mobile Gfx

2 Motherhood-85EA

3 RAM – 4 GB

4 keyboard version – 51.24

5 AMD Ryzen to keep temperature under control

6 smart Trooper cabinet

Packages

1. # Install python2 libraries
2. sudo apt-get install python
3. sudo apt-get install python-pip python-numpy python-scipy python-matplotlib
4. sudo pip install pandas
5. # install python 3 libraries
6. Pip3 install jupyterhub
7. import bs4
8. from bs4 import BeautifulSoup
9. import requests
10. import re
11. import time
12. from datetime import datetime
13. import pandas as pd
14. import numpy as np
15. import seaborn as sns
16. import matplotlib.pyplot as plt
17. %matplotlib inline
18. from selenium import webdriver
19. from mpl\_toolkits.mplot3d import Axes3D

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

1 examining the data

2 check the basic details (null value, D type, Shape etc)

3 identify the target and independent features and perform EDA using Data visualization and Statistical approach accordingly

4 perform data cleaning, outliers handling, missing value imputation

5 feature engineering

6 perform hyperparameter tunning

7 evaluate the model again

* 8 make prediction
* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms used for the training and testing.

Lambda

* Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

We have used seaborn, heatmap and barplot

* Interpretation of the Results

Hyderabad City

Most available brands - Maruti(248), Hyundai(112), Honda(17), Toyota(11), Volkswagon(9)

Cars availability - Petrol(319), Diesel(85), Petrol+CNG(9), Petrol+LPG(4)

Recent year models availability - 2016(21), 2017(47), 2018(43), 2019(30), 2020(3)

Availability based on Gear - Automatic(25), Manual(392)

High Budget car - Toyota Innova Crysta2.4 VX 8 STR (1801049/-), 2018 Model

Low Budget car - Hyundai AccentGLE (166099/-), 2008 Model

Delhi City

Most available brands - Maruti(1069), Hyundai(404), Honda(173), Toyota(150), Renault(82)

Cars availability - Petrol(1122), Diesel(993), Petrol+CNG(81)

Recent year models availability - 2016(216), 2017(217), 2018(210), 2019(122), 2020(61), 2021(2)

Availability based on Gear - Automatic(182), Manual(2014)

High Budget car - Toyota Fortuner2.8 4x2 AT (2918399/-), 2018 Model

Low Budget car - Cheverlet SparkLS 1.0 (112000/-), 2010 Model

Mumbai City

Most available brands - Maruti(802), Hyundai(331), Honda(176), Toyota(93), Volkswagon(84)

Cars availability - Petrol(1083), Diesel(579), Petrol+CNG(71), Petrol+LPG(1)

Recent year models availability - 2016(224), 2017(219), 2018(148), 2019(95), 2020(37)

Availability based on Gear - Automatic(258), Manual(1476)

High Budget car - Toyota Land CruiserLC200 VX 2 PREMIUM (3495000/-), 2010 Model

Low Budget car - Tata NanoXT TWIST (125000/-), 2014 Model

Bangalore City

Most available brands - Maruti(378), Hyundai(228), Honda(41), Renault(29), Volkswagon(9)

Cars availability - Petrol(658), Diesel(116), Petrol+CNG(1), Petrol+LPG(5), Electric(1)

Recent year models availability - 2016(77), 2017(71), 2018(54), 2019(37), 2020(7), 2021(1)

Availability based on Gear - Automatic(70), Manual(711)

High Budget car - MG HECTORSHARP 2.0 DIESEL (1964099/-), 2019 Model

Low Budget car - Chevorlet SparkLS 1.0 (189499/-), 2009 Model

Chennai City

Most available brands - Maruti(300), Hyundai(167), Honda(42), Volkswagon(23), Ford(16)

Cars availability - Petrol(481), Diesel(124), Petrol+LPG(4)

Recent year models availability - 2016(81), 2017(79), 2018(68), 2019(41), 2020(6)

Availability based on Gear - Automatic(60), Manual(549)

High Budget car - Tata HexaVaricor 400 XT (1801499/-), 2019 Model

Low Budget car - ChevorletSparkLS 1.0 (171799/-), 2011 Model

**CONCLUSION**

The availability of cars in 'Delhi'(2200+) is the highest among other 4 cities..

Compared to other 4 cities 'Hyderabad'(415+) has less available cars.

'Maruti' brand cars are widely available with a count of around 2800 cars in all the cities, followed by Hyundai(1240+), Honda(449), Toyota(280+)

Most of the cars runs with 'Petrol' with a count of 3663.

High budget car among all cities:

Toyota Land CruiserLC200 VX 2 PREMIUM (3495000/-), 2010 Model, available in Mumbai

Low budget car among all cities:

Maruti AltoLX (91000/-), 2008 Model, available in Delhi