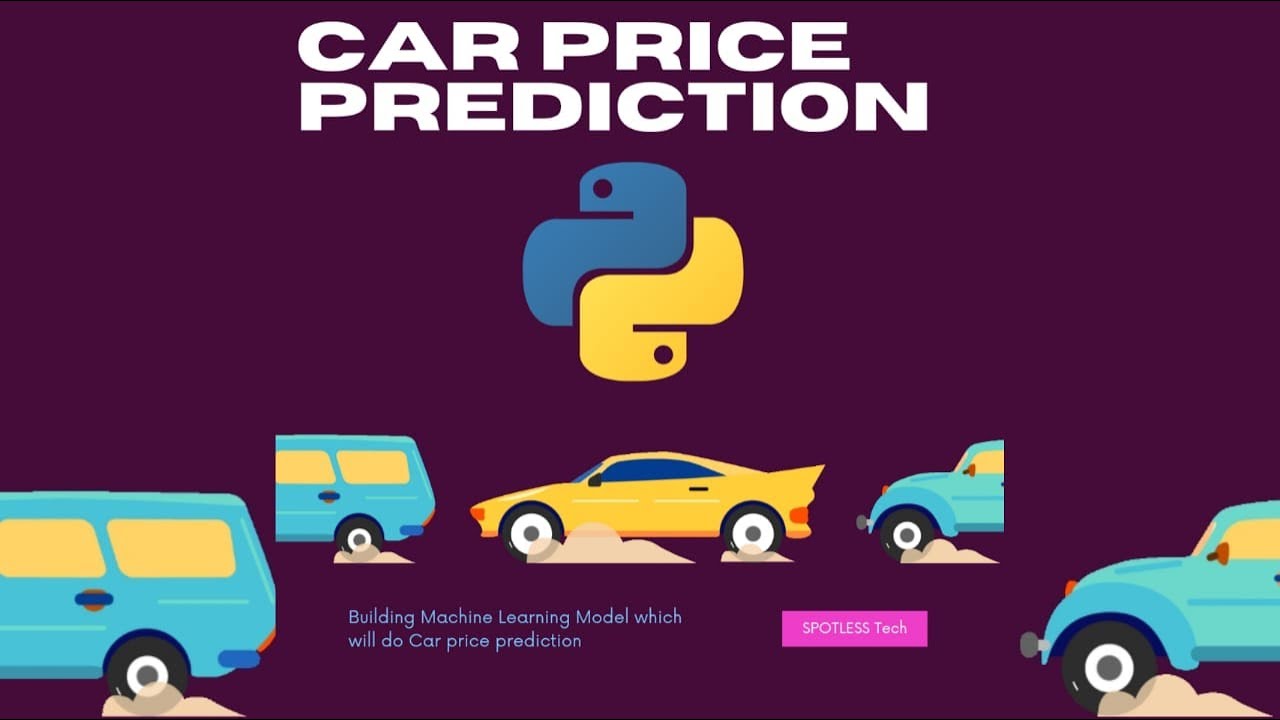
**Car price prediction………..**



~~~~

Contents:

[Project Description:](about:blank)

[Project Technical Details:](about:blank)

[Exploratory Data Analysis](about:blank)

[Data Visualization:](about:blank)

[Model Building:](about:blank)

[Deployment using Flask:](about:blank)

Prediction of Car price

**Project description:**

1. The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely
   1. Brand or Type of the car one prefers like Ford, Hyundai
   2. Model of the car namely Ford Figo, Hyundai Creta
   3. Year of manufacturing like 2020, 2021
   4. Type of fuel namely Petrol, Diesel
   5. Price range or Budget
   6. Type of transmission which the customer prefers like Automatic or Manual
   7. Mileageto name a few characteristic features required by the customer. The project Car Price Prediction deals with providing the solution to these problems. Through this project, we will get to know which of the factors are significant and tell us how they affect the car’s worth in the market
2. Project Technical Details:

The following diagram shows the various steps that we have followed in our project.

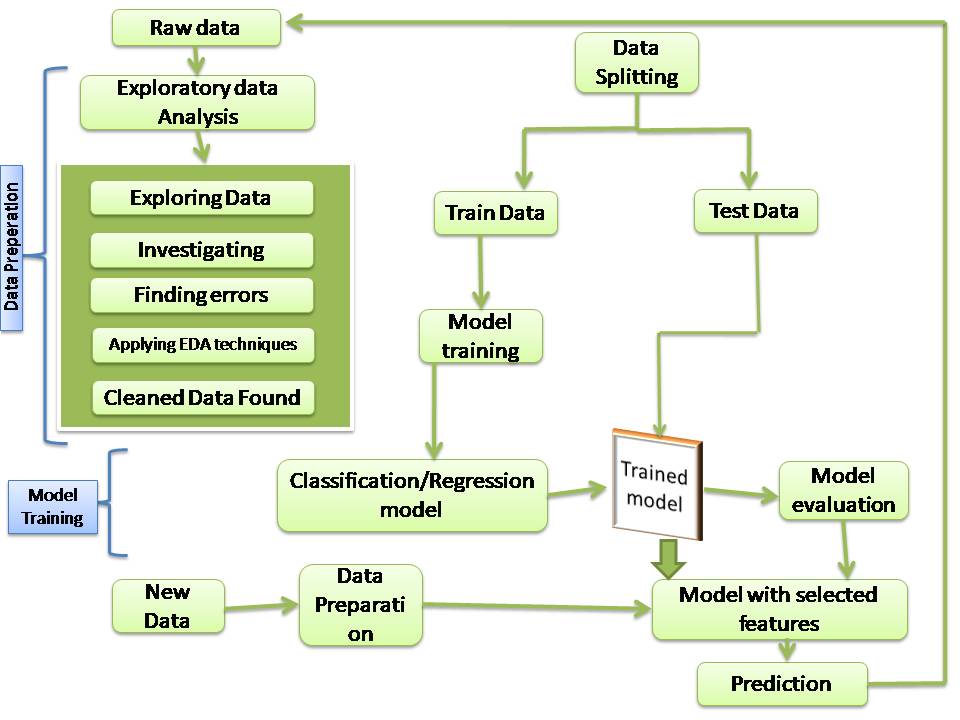
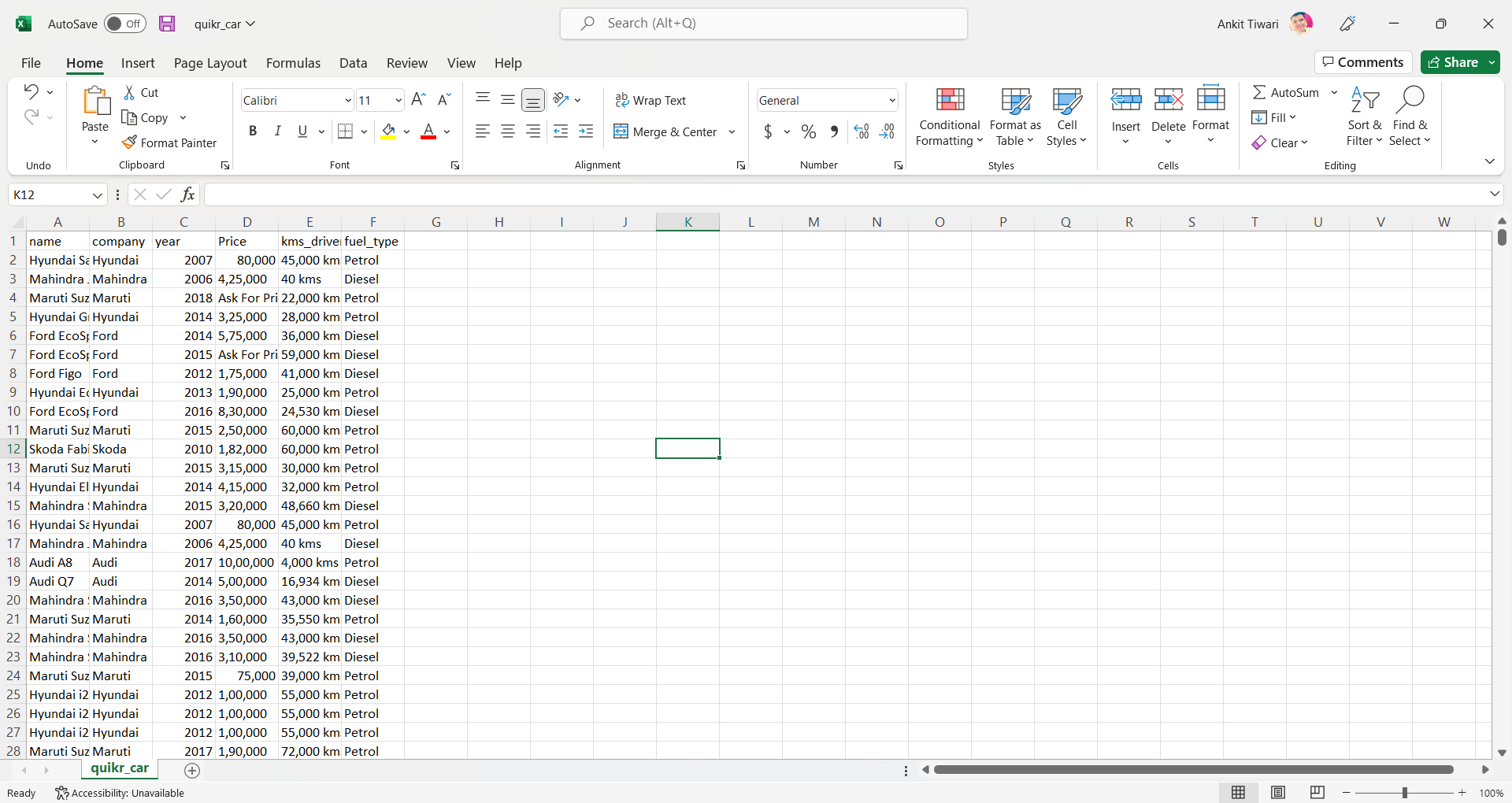


fig 1: General steps of CRISP DM process

1. **Data collection:**

Collected the below details related to the factors influencing the account dormancy through conducting various surveys and documented the details. The dataset included both categorical and numerical data.



1. There are 5 features available in the dataset and 914 Observations.
2. Since we have an imbalanced dataset, we used data cleansing create a balanced dataset.
3. Checked the null values or missing values in the dataset. Performed data imputation, filled null values with median and mode.
4. The information in the 5 features ‘Car name’ and ‘Car COMPANY’,’Car price’,’purchase year’,’kms driven’ ,’fuel type’.
5. There are 13 Features which have >30% NA values and based on characteristics and values, the following features have been dropped.
6. Problem with dataset we have and we resolve
7. ## Quality
8. - names are pretty inconsistent
9. - names have company names attached to it
10. - some names are spam like 'Maruti Ertiga showroom condition with' and 'Well mentained Tata Sumo'
11. - company: many of the names are not of any company like 'Used', 'URJENT', and so on.
12. - year has many non-year values
13. - year is in object. Change to integer
14. - Price has Ask for Price
15. - Price has commas in its prices and is in object
16. - kms\_driven has object values with kms at last.
17. - It has nan values and two rows have 'Petrol' in them
18. - fuel\_type has nan values

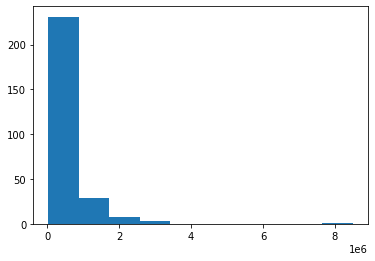
# 2. Exploratory Data Analysis:

Analysis were done on the basis of several charts which are shown below:

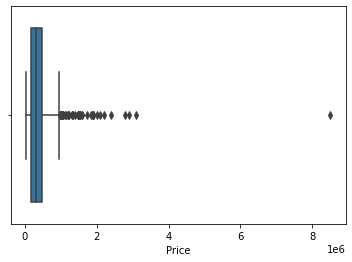
We can see here that these 4 features (Transmission, Seller\_type, Fuel\_type, and Owner) are categorical data and impacting the selling price of the car. Present price of a car directly influences Selling Price prediction. Both are highly correlated and here directly proportional to each other.

Car age is affecting negatively as the Selling Price decreases for an older car.

plt.hist(data.Price.unique()) #histogram



sns.boxplot(data.Price)

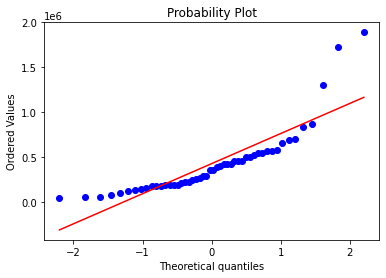


from scipy import stats

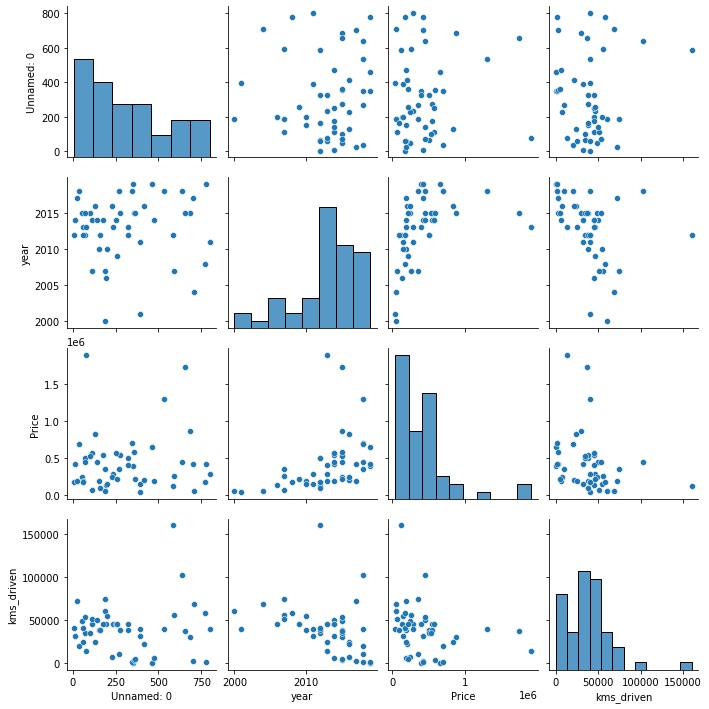
import pylab

stats.probplot(temp.Price, dist = "norm", plot = pylab)

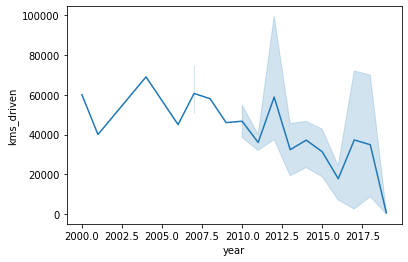
plt.show()



sns.pairplot(temp.iloc[:, :])



sns.lineplot(data = temp, x = "year", y = "kms\_driven")



Correlation ########################################

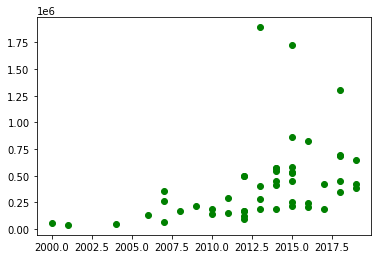
# Correlation measures the relationship between two variables. We mentioned that

# a function has a purpose to predict a value, by converting input (x) to output (f(x)).

# We can say also say that a function uses the relationship between two variables for prediction

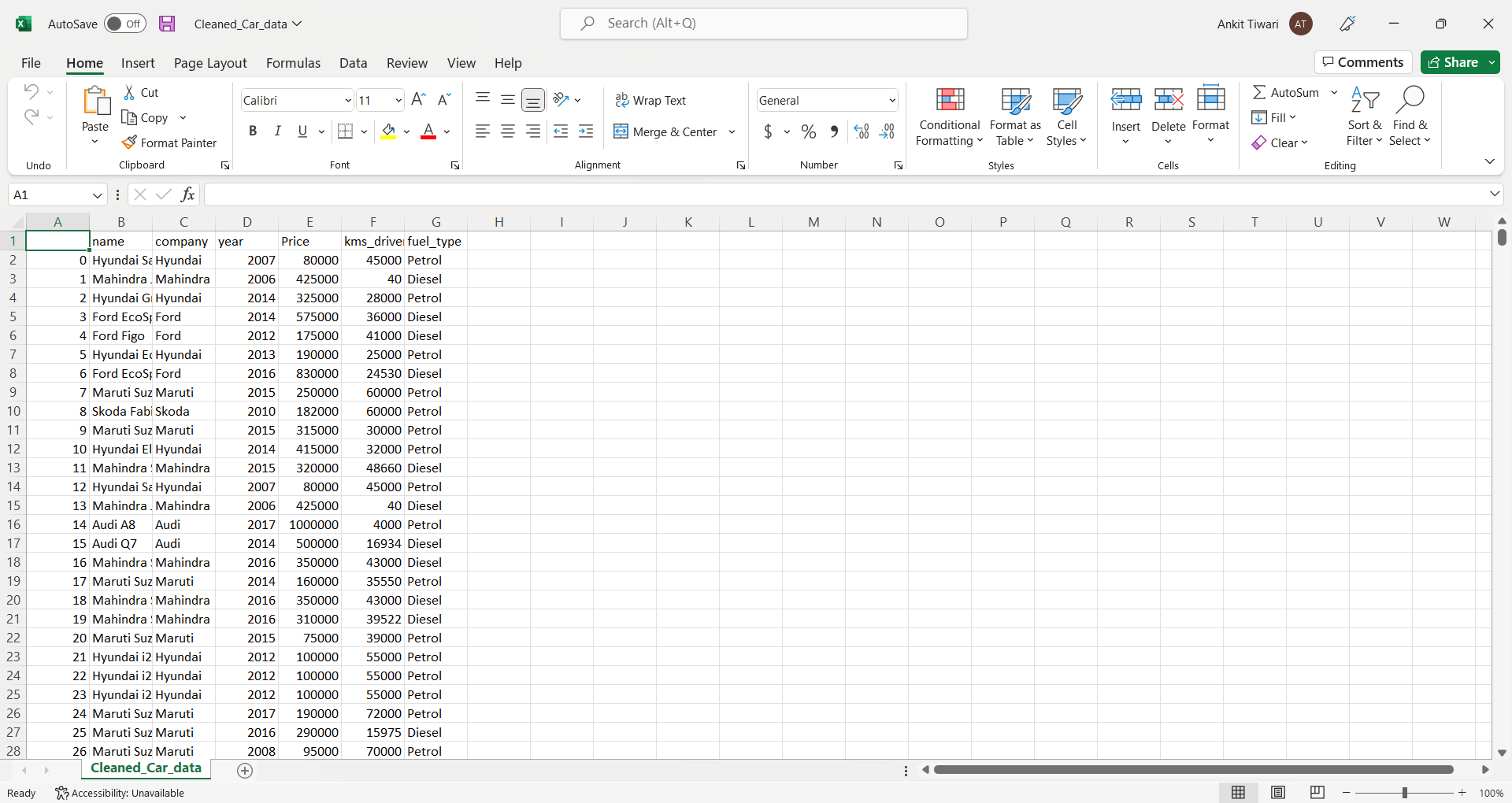
np.corrcoef(temp.Price,temp.year)

plt.scatter(y=temp.Price,x=temp.year,color='Green')



***Note: Hypothesis testing is performed on survey data. features are selected on the basis of sampled data. Domain knowledge has not been taken into account.***

**LAZY CLASSIFIER RESULTS**



***Selected model: Random Forest***

Using Lazy classifier techniques various machine learning algorithms were explored such as KNN, Ada Boost, Decision Tree, ensemble techniques , the model that gave the highest accuracy is Random Forest.

|  |
| --- |
| Create a new app.py file. Now, import every important module and library to deploy the model. Also load the model in the app.py file. 4. Data Preparation Here we will clean the data and prepare it for training the model  For the web application you need to create a html file for the structure of website and you can add CSS and JS for styling and other thing. |
| **Output** |

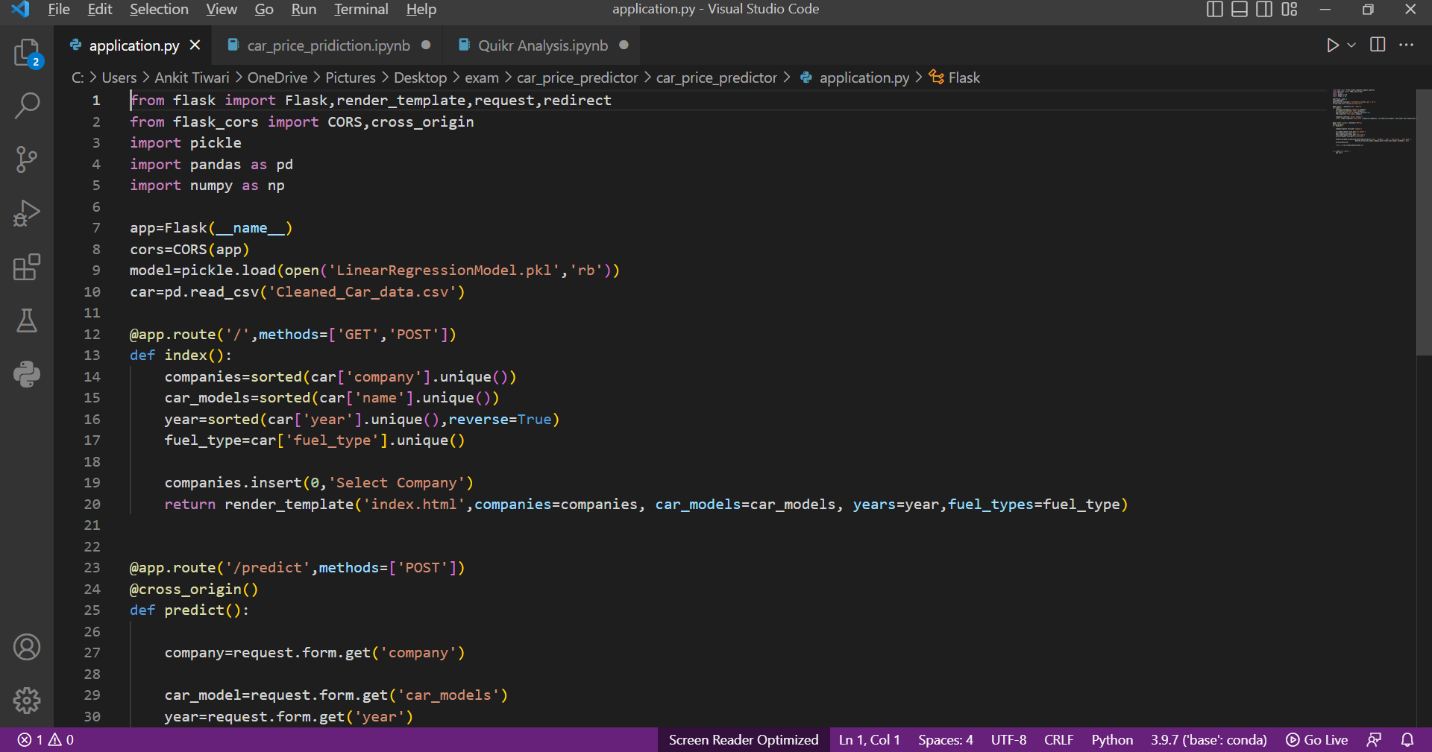
**Dormant account prediction uses the following packages and library from python:**

|  |
| --- |
| import pandas as pd  import numpy as np  import matplotlib.pyplot as plt  import matplotlib as mpl  %matplotlib inline |

# 6. Deployment using Flask:

Deployment process was done using flask technique.

**Deployment Architecture**

****

****

Created app.py file to show prediction of dormant accounts

|  |
| --- |
| After providing values    Predicted result……………… |