

Capstone Project Data Science Nanodegree March 10th , 2024

Laptop price predictions

Project overview:-

In this project i am going to develop a machine learning algorithm which will be helpful to predict the price of the laptop by using certain criteria. The data set contains the details of Manufacturer, type of the laptop, screen size and resolution, Ram, memory, GPU, weight and price. price is changing as per the specification of the laptops. My task is to develop a ML algorithm after doing required EDA, conversion of categorical variables into numbers to use them into the algorithm.

Problem Statement:-

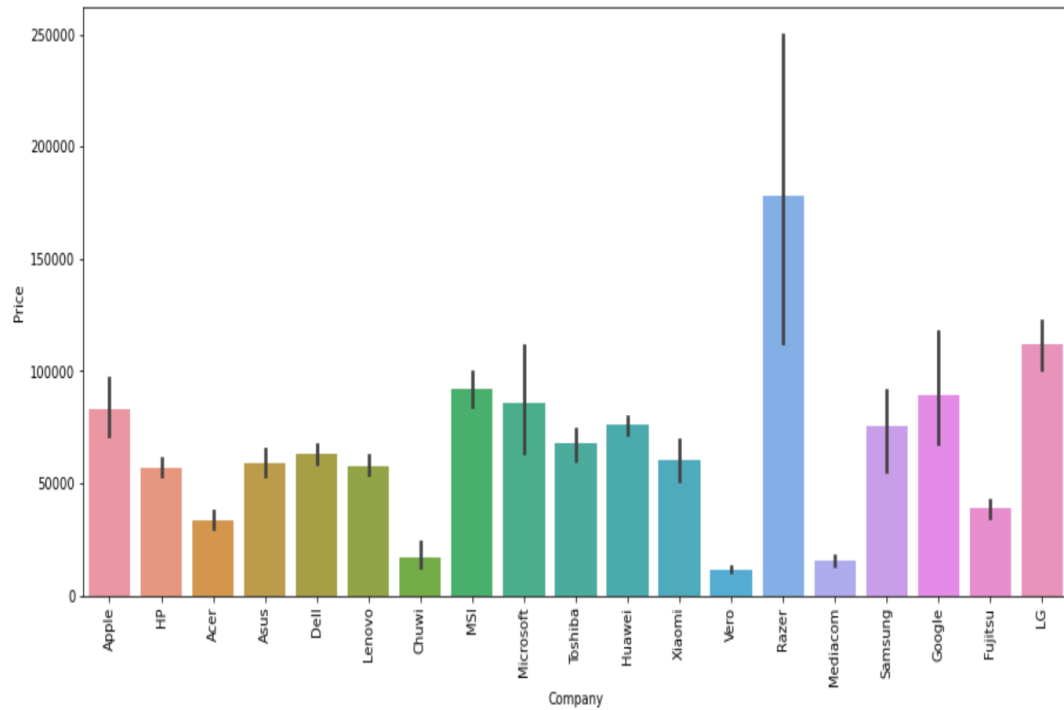
Predicting the price of the laptop which will be possible higher level of response from user based on Manufacturer, type of the laptop, screen size and resolution, Ram, memory, GPU, weight and price.

Data cleaning:-

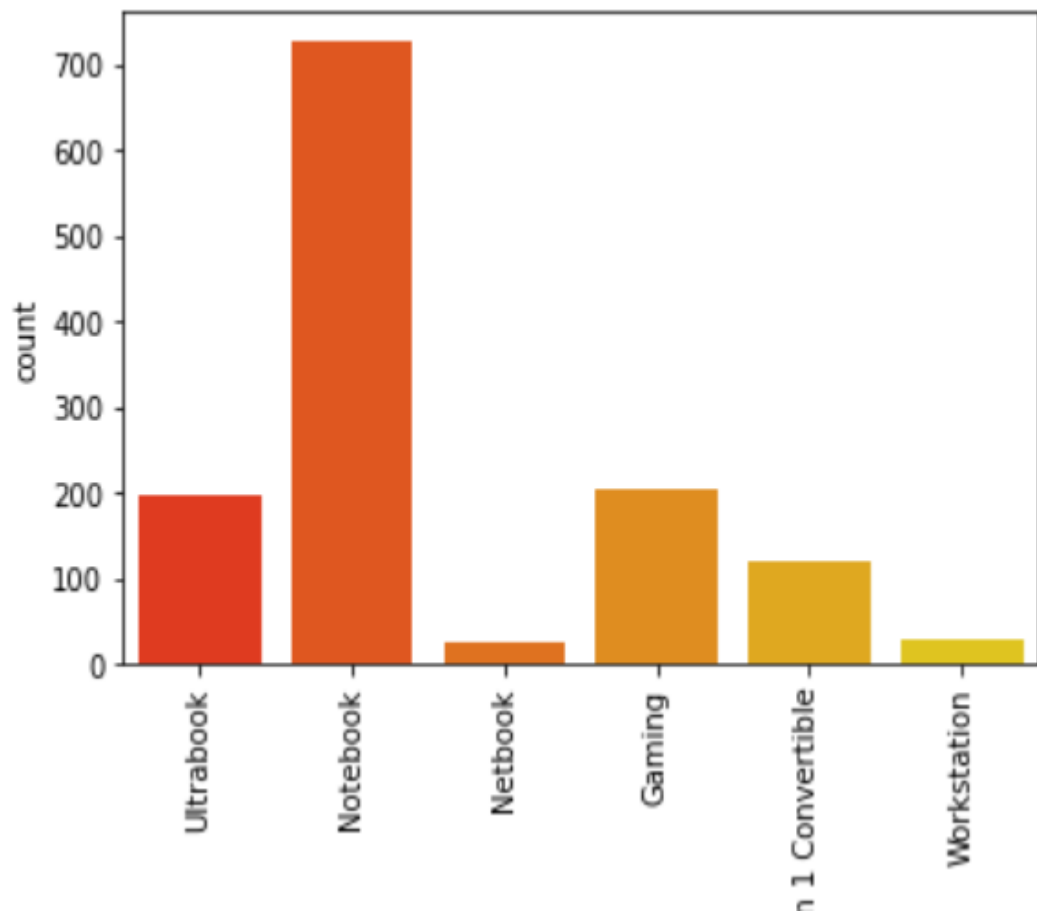
1. Removing the unnamed:0 column which was not useful.
2. Check for nulls and duplicate.
3. Check for the data types of the columns.
4. Check for the unique values in each and every column by using a function.
5. Converting strings to int by performing certain operations.(e.g. remove GB from Ram and KG from weights)

EDA:-

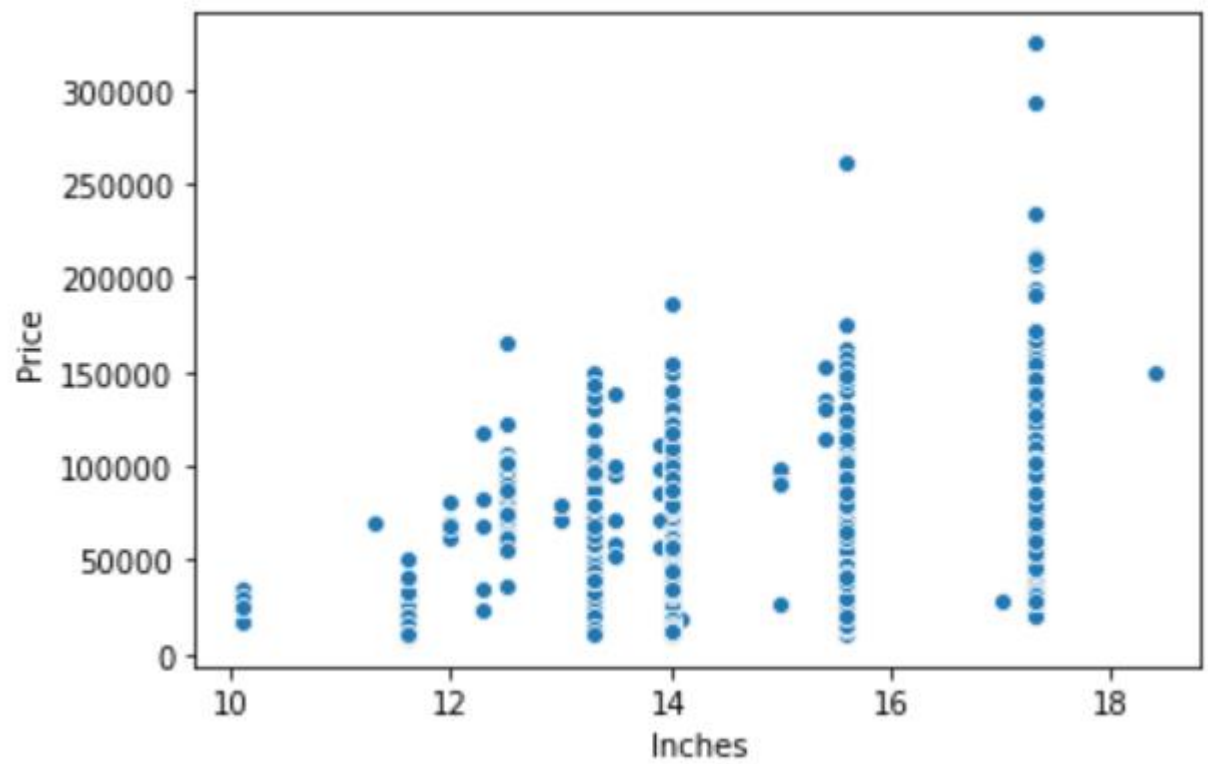
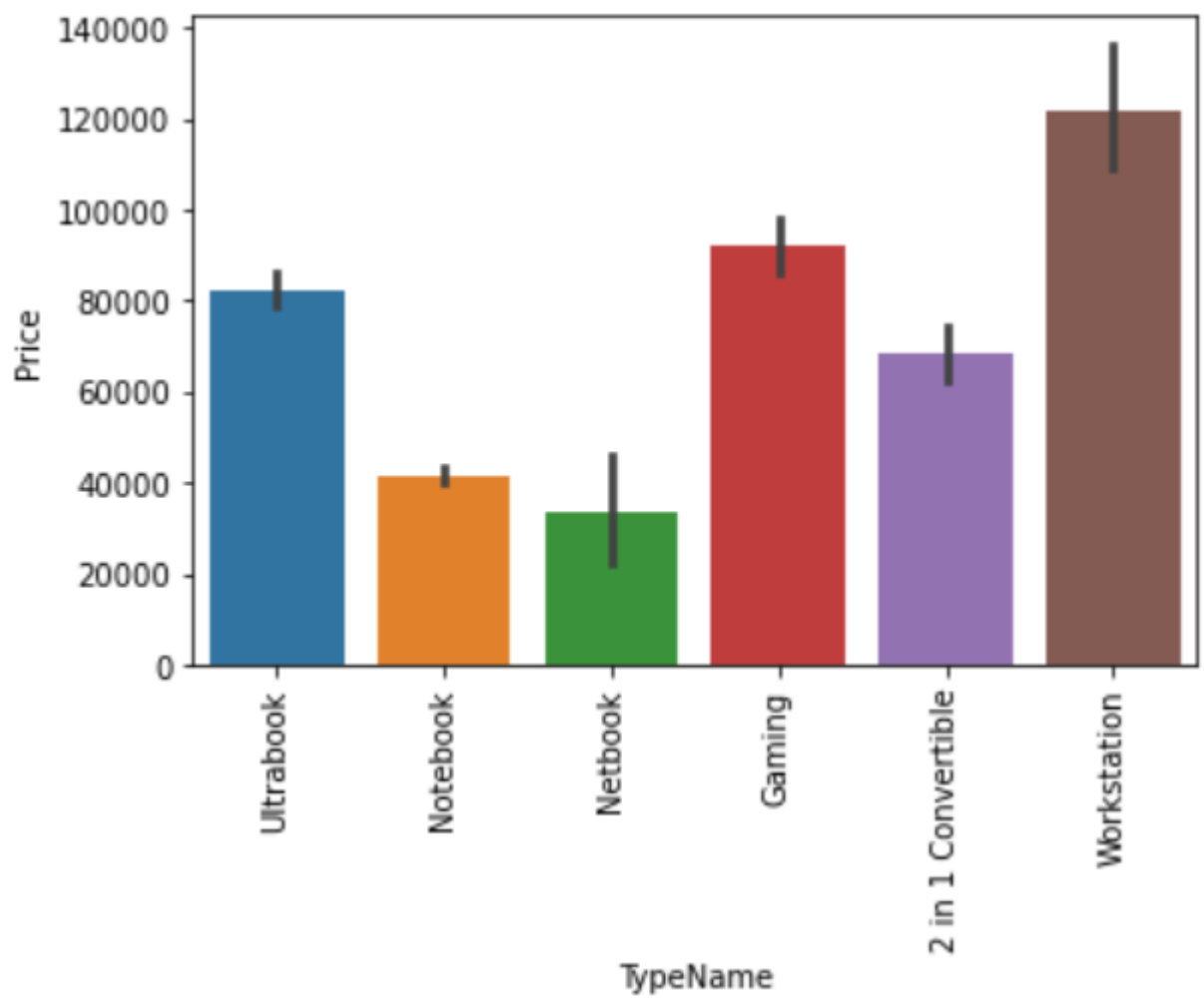
1. Checking the distribution of price so see the price ranges.
2. Checking the count of laptops against company, Typename, Ram, OpSys.
3. Check for the price ranges against each company.



4. Check for the count of various type of laptop.



5. Similarly check for the variation of price against screen, laptop type and price.



6. Convert the Screen-resolution column into X&Y and create another KPI named as PPI by using below formulae and remove screen size and resolution columns.
-

$$PPI(\text{pixels per inch}) = \frac{\sqrt{X_{\text{resolution}}^2 + Y_{\text{resolution}}^2}}{\text{inches}}$$

7. Create columns named as IPS, Touchscreen and fill binary values in it to use further.
8. Check for the correlation of each column against price.
9. Remove those columns which are not contributing much in the price.
10. Convert the memory into columns (e.g. HDD, SDD, Storage & Flash storage) to use it further in the modeling.
11. Apply the same step in GPU & OpSys columns as well.

Model Building:-

First of all I have imported all the required libraries and split the datasets into train & test part. I have tried below mentioned models.

1. Linear Regression.
2. Ridge Regression.
3. Lasso Regression.
4. Decision Tree.
5. Random forest.

After that we have performed Hyper parameter tuning followed by Prediction on the whole Dataset.