-Introduction

-Exploratory and Descriptive Analysis

-Research Questions

- Date Preprocessing

-Models

-results and Conclusion

**Introduction**

Laptops Dataset contains laptop specifications and it’s prices, In our project we will dive into the Laptops dataset, do analysis, get insights and finally we will create a model to Predict Laptop Prices.

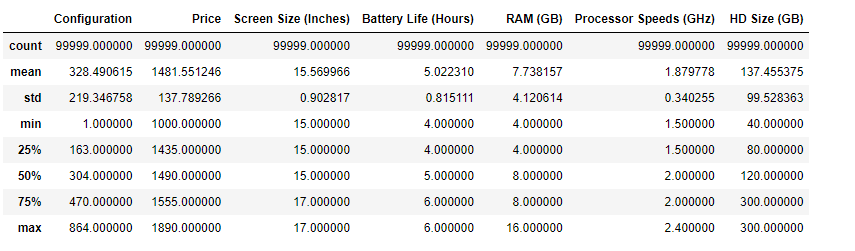
If any user wants to buy a laptop then our application should be compatible to provide a tentative price of laptop according to the user configurations.

**Exploratory and Descriptive Analysis**

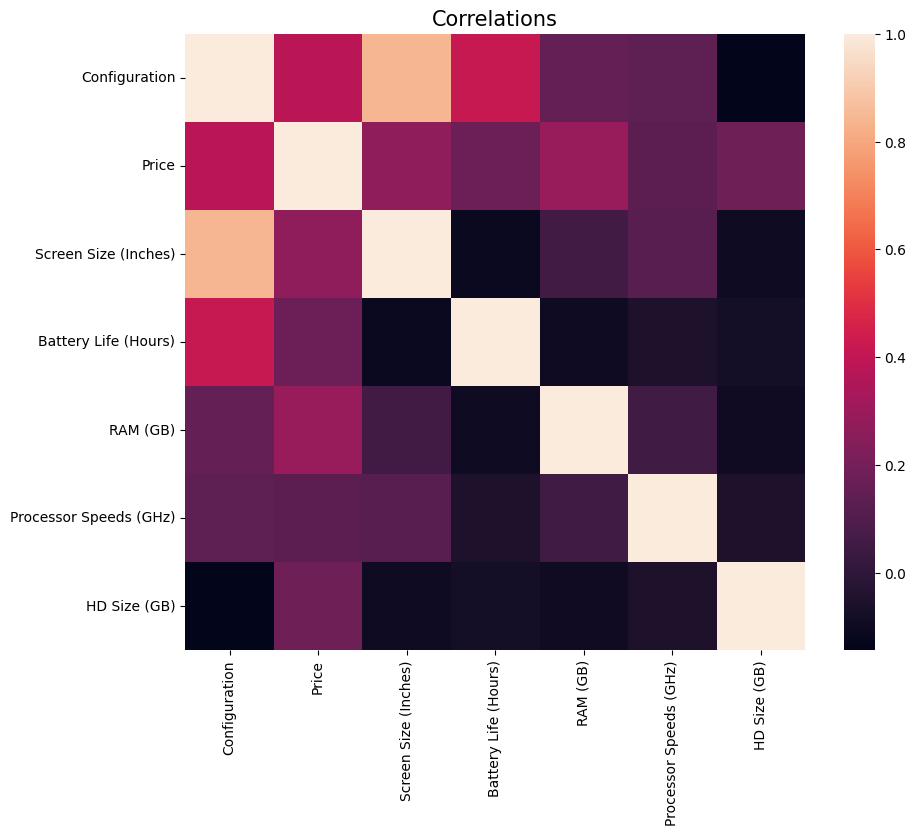
Our dataset consist of 100k rows and 9 columns (features),

We have 7 numeric features and 2 categorical Features.

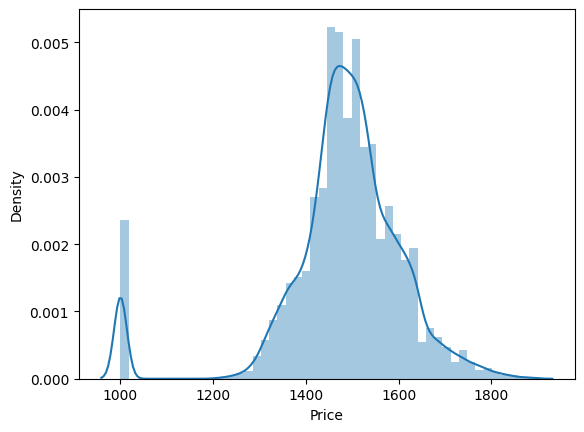
Descriptive analysis table:



Correlation between Numerical Features

****

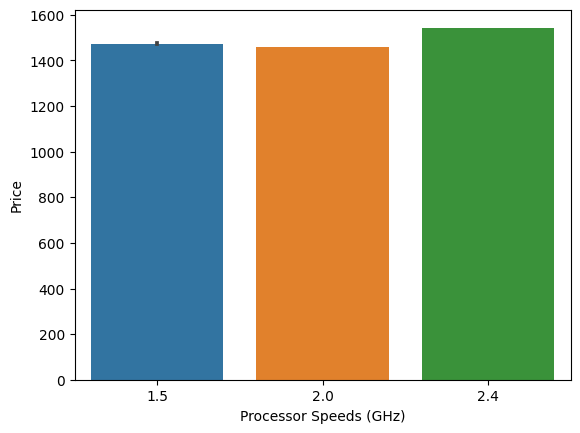
The distribution of the Price Variable



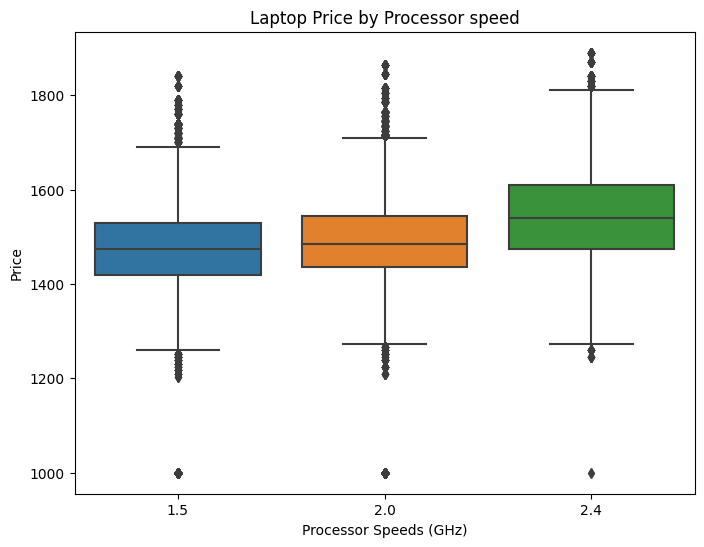
As we see most of laptop prices ranges from 1400 $ to 1600%.

Least price is 1000$ and max is around 1800$.

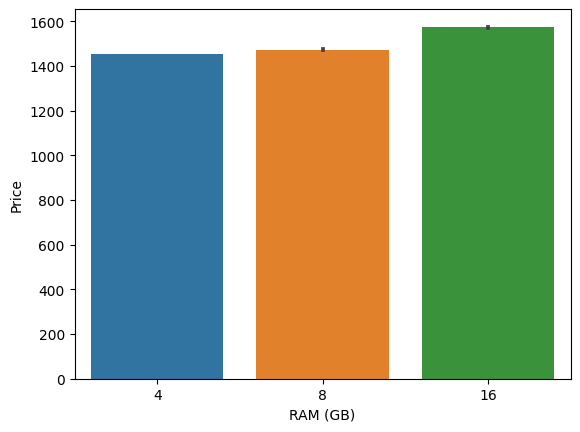
Bar plot between Price of laptop and Processor speed



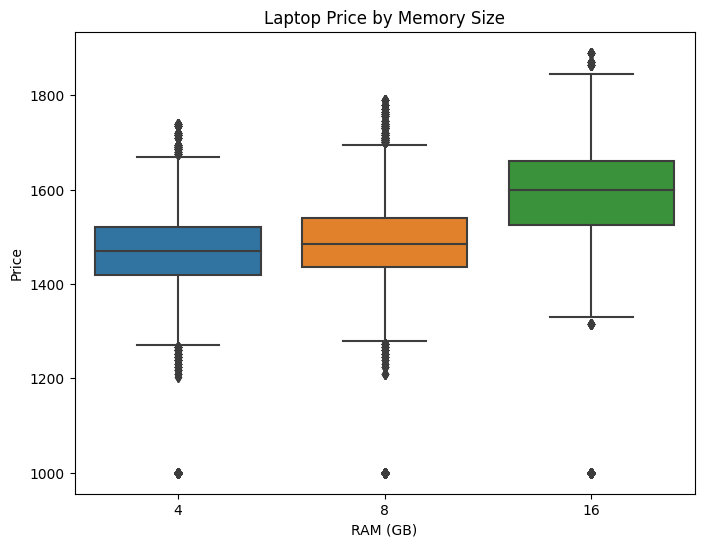
Processor speed has influence on the price but not that much that indicates that the other features has influence too.

Box plot between Price of laptop and Processor speedObviously, as the speed of the processor increases the outliers of the prices decreases.

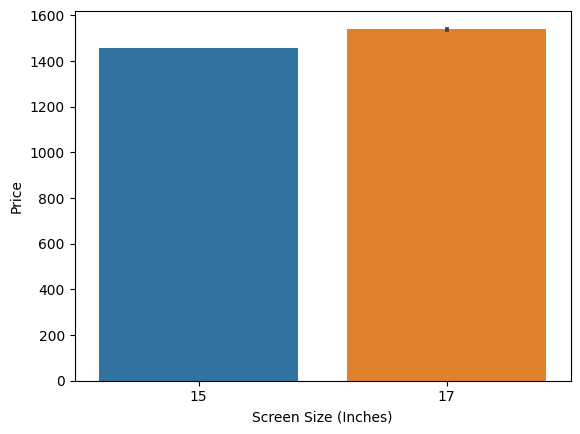
Bar plot between Price of laptop and Ram size



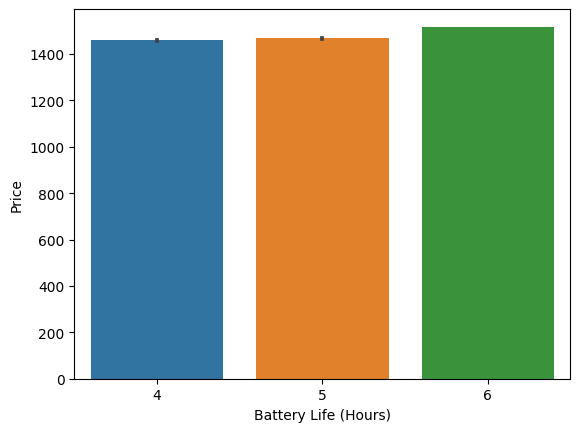
There is a significant influence of the 16g ram on the price of the laptop



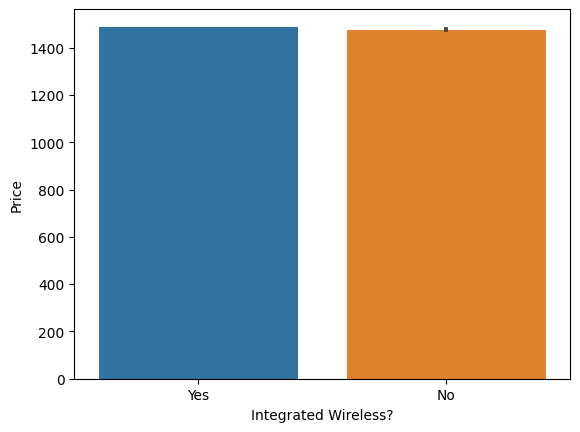
Bar plot between Price of laptop and Screen Size (Inches)



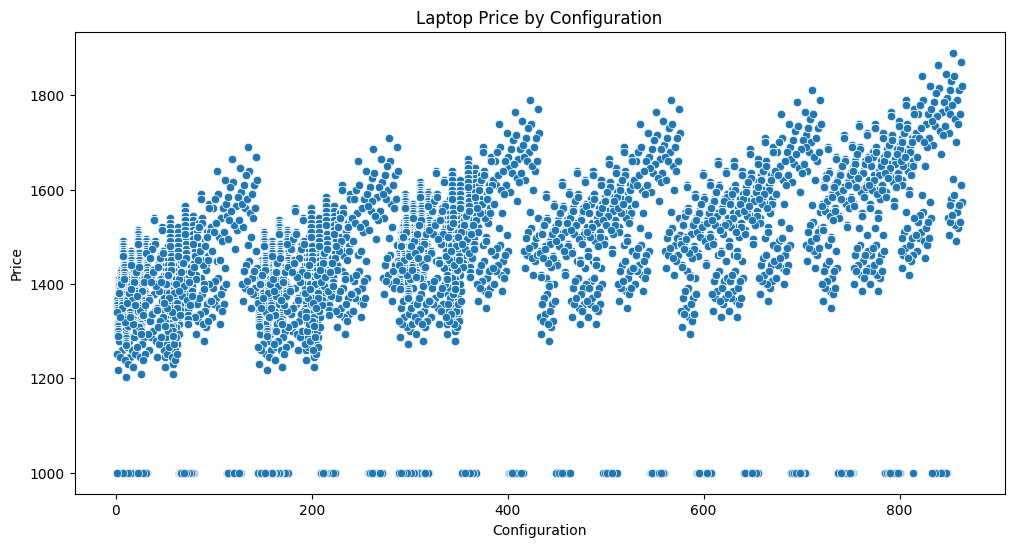
Bar plot between Price of laptop and Battery Life (Hours)



Bar plot between Price of laptop and Integrated Wireless?



Scatter plot between Price of laptop and configuration?



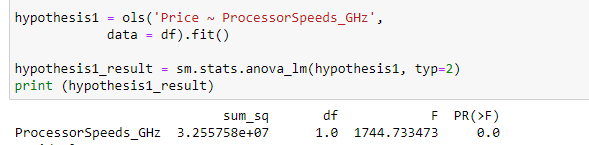
Research Questions

-**Anova Analysis questions**

1. Is there a Notable difference in the Laptops Price across Processor speed category?

H0 = The mean Laptop Price is equal for Processor Speeds Category.

H1 = At least one of the Processor Speeds category has a mean price that is not the same as the other Processor Speeds Category.



The p-value of the condition variable is < 0.05,

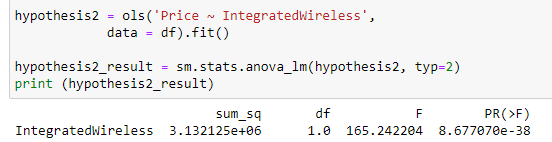
Which implies that Processor Speeds gives impact on the Laptop Price.

So the H0 is rejected.

1. Is there a Notable difference in the Laptops Price across Integrated Wireless category?

H0 = The mean Laptop Price is equal for Integrated Wireless Category.

H1 = At least one of the Integrated Wireless variable has a mean price that is not the same as the other Integrated Wireless variable Category.



The p-value of the condition variable is < 0.05,

Which implies that Integrated Wireless gives impact on the Laptop Price.

So the H0 is rejected.

-**Regression questions**

1-Can we predict the Laptop price with the given Features using Linear Regression Model ?

2-Can we predict the Laptop price with the given Features using Logistic Regression Model ?

In order to answer the above questions , we must proceed to the data preprocessing step then to the models step.

Data preprocessing

-we Encoded the categorical Variable using One Hot Encoder as machine learning model accepts only numeric values

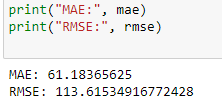
-Normalization for the numerical Data using MinMaxScaler that normalize numeric values to be ranging from 0 to 1.

- splitting the data to train and test with percentage 80/20

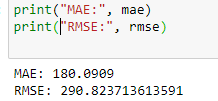
Model

We fitted a linear regression and Logistic regression models using our train data.

Linear regression result:



Logistic regression result:



Conclusion

Linear regression Model do better than Logistic regression, however the difference in results is not big .

Remembering our regression question we can say now we were able to answer those question, we could predict the Laptop Prices using Linear and logistic Regression, after doing the analysis and getting the needed insights we used Exploratory analysis and Anova analysis to understand more about our dataset then finally the project run successfully by running our models .