# LAPTOP SELECTION AND PRICE PREDICTION

Data Science With Python Lab Project Report

Bachelor

in

Computer Science

By

**Team Name**

S190005

S190361

Rajiv Gandhi University Of Knowledge And Technologies

S.M. Puram , Srikakulam -532410

Andhra Pradesh, India

# Abstract

Introducing a new project "laptop selection and price prediction". The main aim of this project selecting of the best laptops based on financial position with the best features. After evaluation of technology, we all use laptops to complete our tasks very fastly, effectively and accurately. Nowadays laptop plays a crucial role in every field like education, manufacturing, the service sector etc..to achieve this, we took the dataset from the Kaggle website the dataset mainly contains information about the 1000 laptops on India's E-Commerence platform like Flipkart. The dataset contains technical features like Ram, Preprocessor type, Hard disk, display resolution, storage info, number of reviewers, number of ratings, etc. It also provides the image link related to the particular laptop. This project is built using Data science and machine learning models. we use Python libraries like Numpy, Scipy, pandas, Scikit-Learn and Matplotlib for the anal- ysis of the data to get desired output. In this project, we mainly compare all features of different laptops and select the best laptop among all and also predict the price of the laptop based on features and compare it with the given price. Nowadays most of the people suffer in decision-making in the selection of the best laptop for their needs. This project will help them and it also helps businessmen in decision-making in which laptop production is to be increased to get more profits. Based on reviews, this project leads to the manufacturing of new laptop based on the requirements of consumers.

# Contents

# Chapter 1

# Introduction

## 1.1 Introduction to Your Project

In today's world After evaluation of technology Laptop plays crucial part in daily life. Laptops have become essential tools in various purposes ,includ- ing work,education,business,entertainment,j and marketing.These Lap- tops have wide range of Applications in different fields.We aim to develop laptop system to accommodate users in making informed decisions when we purchasing laptop.By considering indiviual preferences,budget and specific requirements,our system will recommend suitable one among all based on the user's need.The first objective of the project is gather the information from the user budget or financial positions and desired specifications like specified ram,processor etc.. and helps them in decision making in purchusing laptop. As technology advances many laptop brands have sprung up and from every single one launches laptop with their various advantages. Of the var- ious types of laptops, specifications, and functions often cannot be used by consumers who do not meet their needs..this analzying project process takes account of prices, brands, and laptop specifications such as processor,ram, and memory. Because of this, a laptop selecting system is needed. This analzying project process takes account of prices, brands, and laptop specifications such as processor, ram, and memory. There are more number of laptops are available in market with different features,among all choosing the right one for our requirements is difficult task for every one.This challenging task can be achieved by Datascience technique.These techniques can simplify the problem by helping us in decision making of

selection of the laptop. The first objective of the project is gather the information from the user budget or financial positions and desired specifications like specified ram,processor etc.. and helps them in decision making in purchusing laptop. As technology advances many laptop brands have sprung up and from every single one launches laptop with their various advantages. Of the var- ious types of laptops, specifications, and functions often cannot be used by consumers who do not meet their needs..this analzying project process takes account of prices, brands, and laptop specifications such as processor, ram, and memory. Because of this, a laptop selecting system is needed. This analzying project process takes account of prices, brands, and laptop specifications such as processor, ram, and memory. There are more number of laptops are available in market with different features,among all choosing the right one for our requirements is difficult task for every one.This challenging task can be achieved by Datascience technique. These techniques can simplify the problem by helping us in decision making of selection of the laptop. The main aim of this project is to use the datascience methodologies to cre- ate one model that can helps users in selecting most suitable laptop for their needs.By analyzing the features of the laptop and user preferences develop best system that can provides more accurate values and recommendations. To acheive this project,we will collect the information from online websites like kaggle. The dataset which provides complete view of laptop in market in India.It can helps for analysis of the data,research and visualizing of the data.This dataset is helps to students,professionals and business persons in selecting of best laptop among all available laptop models.

## 1.2 Application

1.E-commerce Platforms: Online marketplaces that sell laptops can utilize laptop selection and price predic- tion models to assist customers in finding the most suitable laptops based on their requirements and budget. By providing personalized recommendations, e-commerce platforms can enhance the shopping experience and improve customer satisfaction. 2.Tech Review Websites: Websites or platforms that provide laptop reviews and recommendations can integrate laptop selection and price prediction capabilities. This integration would enable users to input their desired specifications and budget, and receive

personalized recommendations for laptops that meet their criteria. 3.Price Comparison Engines: Price comparison websites can incorporate laptop price prediction models to estimate the future prices of laptops. This information can be valuable for users looking to make a purchase decision, as they can anticipate price fluctuations and make informed choices. 4.Financial Analysis: Laptop manufacturers, retailers, and investors can leverage laptop price prediction models to analyze market trends and make strategic business decisions. Predicting future laptop prices can assist in inventory management, pricing strategies, and investment planning. 5.Consumer Insights and Market Research: Laptop selection and price prediction models can provide valu- able insights into consumer preferences and market demand. By analyzing patterns and trends in laptop features and prices, companies can identify emerging trends, develop targeted marketing strategies, and launch new products tailored to consumer needs. 6.Personal Budgeting and Planning: Individuals who are in the market for a laptop can use laptop selection and price prediction tools to assess various options and plan their budget accordingly. By considering factors such as desired features and predicted prices, individuals can make informed decisions that align with their financial goals. 7.Educational Institutions: Laptop selection and price prediction models can be useful for educational institutions when recommending laptops to students. By taking into account the requirements of specific academic programs and student budgets, institutions can guide students in selecting suitable laptops for their educational needs.

## 1.3 Motivation Towards Your Project

As technology advances many laptop brands have sprung up and from every single one launches laptop with their various advantages. Of the var- ious types of laptops, specifications, and functions often cannot be used by consumers who do not meet their needs..this analzying project process takes account of prices, brands, and laptop specifications such as processor, ram, and memory. Because of this, a laptop selecting system is needed. This analzying project process takes account of prices, brands, and laptop specifications such as processor, ram, and memory. There are more number of laptops are available in market with different features,among all choosing the right one for our requirements is difficult task for every one.This challenging task can be achieved by Datascience technique.These

techniques can simplify the problem by helping us in decision making of selection of the laptop. The main aim of this project is to use the datascience methodologies to cre- ate one model that can helps users in selecting most suitable laptop for their needs.By analyzing the features of the laptop and user preferences develop best system that can provides more accurate values and recommendations. To ensure the accuracy and effectiveness of the recommendation system, we will evaluate its performance using metrics such as precision, recall, and user feedback. We will continuously refine and improve the system by incor- 2po- rating user feedback, updating the laptop database, and fine-tuning the recommendation algorithms. The laptop selection project aims to develop a personalized recommendation system that assists users in selecting the most suitable laptop based on their preferences, budget, and specific require- ments. By leveraging user profiling, data analysis, and advanced algorithms, this project aims to simplify the laptop selection process and empower users to make well-informed decisions. The successful completion of this project will provide a valuable tool for individuals seeking the perfect laptop that meets their unique needs and preferences. Based on data analysis ,we will assign weights to different features and properties of laptops which requires the importance in the selection process. if we consider gaming performances the weights assigned to graphics card specifcations will be higher. Using required needs of users we will rank the laptops. Now we design user friendly interface that allows users to input their preferences, view recommended laptops, and compare different models. The interface may includes interactive features, filtering of the options which are available for us based on requirements, and visualizations to helps the user experience,make user comfortable in selection of laptop and facilitate decision-making in choosing of laptop. laptop selection project aims to uti- lize datascience techniques to simplify and optimize the process of choosing the best laptop based on their needs.This project provides personlized recom- mendations and valuables enable users to make decisions in selecting laptop for their specific needs and requirements.

## 1.4 Problem Statement

Improving the accuracy finding price of the Laptops and selecting the best laptop The problem at hand is to select the best laptop among all given laptops with their features.Based

on the data analysis of the user recocommend accurate and valuable recommmendations. Based on the preferences of the user and specifications related laptop which user wants to purchuse.the main aim to overcome the challenging task to select best laptop among several laptops available in market. The main aim of this project is to use the datascience methodologies to cre- ate one model that can helps users in selecting most suitable laptop for their needs.By analyzing the features of the laptop and user preferences develop best system that can provides more accurate values and recommendations. To acheive this project,we will collect the information from online websites like kaggle. The dataset which provides complete view of laptop in market in India.It can helps for analysis of the data,research and visualizing of the data.

# Chapter 2

# Approach To Your Project

## 2.1 Explain About Your Project

A laptop selection and price prediction project mainly involves using machine learning techniques to analyze various features of laptops and predict their prices based on those features and also selecting the best laptop among several laptops which are available in market. The goal is to build a model that can accurately predict the price of a laptop and also choose one best laptop based on preferences and its specifications, such as the processor, RAM, storage capacity, display size, and other required information.

Gather the dataset from the online websites like githud,kaggle etc.dataset contains the laptop specifications like Ram,processor,os,storage and display size etc..the selected dataset can includes wide range of brands in market like Hp,lenovo,Dell etc..The dataset should cover a wide range of laptops from different brands, models, and price ranges.clean and preprocess the gathered data to increase its quality for training data to get accurate values of recommendations.This step involves handling missing values, removing duplicates and unneccesary data standardizing numerical features, and encoding categorical variables. It's important to procesthat can made the dataset is consistent and ready for further analysis.

After performing of data analysis on gathered dataset it helps in understanding of relationships between the different features of Laptop and Laptop prices.choosing appropriate machine learning model among random forest,logistic regression,multiplelinear regression,knn algorithm etc.Based on the nature of gathered dataset we will choose best algorithm

which gives more accurate values..

Once the model is trained using collected dataset, it can be used to predict the price of a new laptop based on its specifications and also gives best laptop based on required specifications. Users can input the relevant features of a laptop into the trained model, and it will generate an estimated price as output .If we give required features to this model it will gives recommended laptop to them The predicted price can provide valuable insights for users when making decisions about laptop purchases.

## 2.2 Data Set

The dataset collected from kaggle website contains information about various laptop specifications and their corresponding prices.THe dataset gives comprehensive view of different brands in market. Name: this gives the information about different types of laptops in market. such as Dell, HP, Lenovo, Apple, etc. these features of the laptop influences prices of the laptop. Ram: The amount of memory available for the laptop's operations. This feature represents the laptop's multitasking capability and influences its performance and price. Storage: The storage capacity of the laptop, typically measured in gigabytes (GB) or terabytes (TB). This includes information about hard disk drives (HDD) or solid-state drives (SSD). The storage capacity affects the laptop's price and usability. Display Size: The size of the laptop's screen, usually measured diagonally in inches. This feature helps users assess the laptop's visual experience. Processor: The type and specifications of the laptop's central processing unit (CPU). This feature includes information like the brand (Intel, AMD), the model (i5, Ryzen 7), the number of cores, and clock speed. The processor's performance significantly impacts the laptop's price. Price: The actual price of the laptop. This is the target variable that the machine learning model aims to predict based on the other features. os: The operating system is the software that manages computer hardware and software resources and provides a user interface for interacting with the computer.

## 2.3 Prediction technique

For laptop selection and price prediction, various machine learning techniques can be used. One common technique is regression analysis, which aims to predict a continuous numerical value, such as the price of a laptop. Regression models can capture the relationships between input features (such as processor, RAM, storage, etc.) and the output variable (price).

When it comes to the laptop selection process, the focus is typically on finding the most suitable laptop based on the user's specific needs and preferences. While price prediction is an important aspect of laptop selection, it is not the only factor considered. Therefore, the laptop selection process may involve various techniques beyond price prediction output variable is (name of the laptop). , classification techniques may not be directly used for predicting the actual price of a laptop since price is a continuous numerical value. However, classification techniques can still be valuable for various related tasks and aspects of the project.

Classification algorithms can be used to identify the importance of different features in determining the price category of a laptop. By training a classifier on price categories and input features, you can obtain insights into which features have the most significant impact on the price classification. This information can help users understand the key factors driving laptop prices.

Decision Trees: Decision trees recursively split the dataset based on different features to create a tree-like model. Each leaf node represents a predicted value. Decision trees can handle both numerical and categorical features and capture non-linear relationships between features and price. However, they may suffer from overfitting if not properly regularized.

Random Forest: Random forest is an ensemble learning method that combines multiple decision trees. It averages the predictions of individual trees to produce a more robust and accurate prediction. Random forests can handle complex relationships and reduce overfitting by averaging predictions from different trees.

## 2.4 Graphs

```
import numpy as np
import pandas as pd
import seaborn as sns
sns.distplot(d["price(in Rs.)"])
plt.show()
```

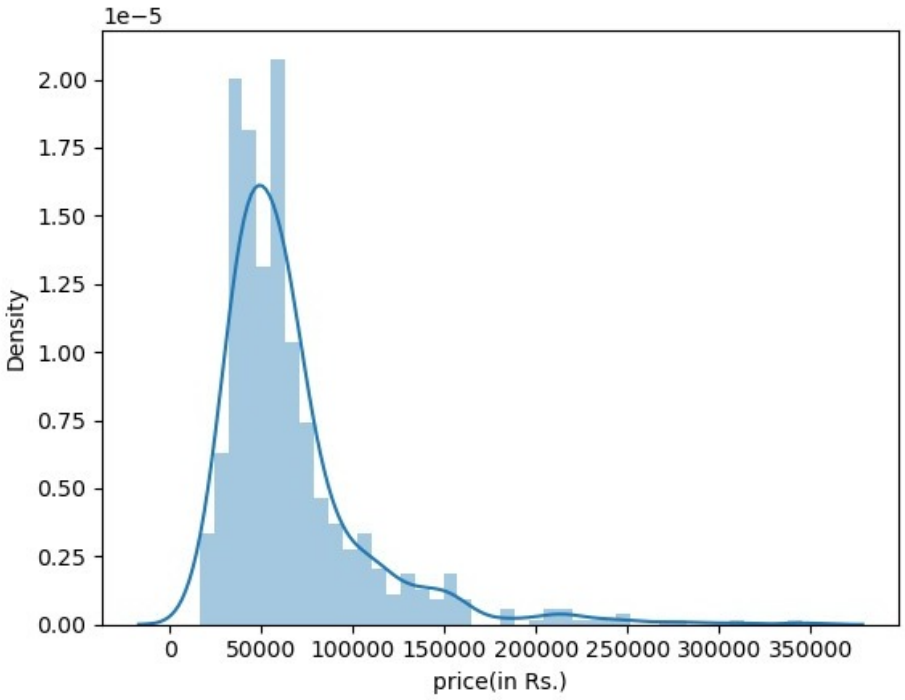



Figure 2.1: Analysis of price ranges for different Laptops

```
import seaborn as sns
import pandas as pd
import  matplotlib.pyplot as plt
k=pd.read_csv("file2.csv")
plt.scatter(k['name'],k['ram'])
```

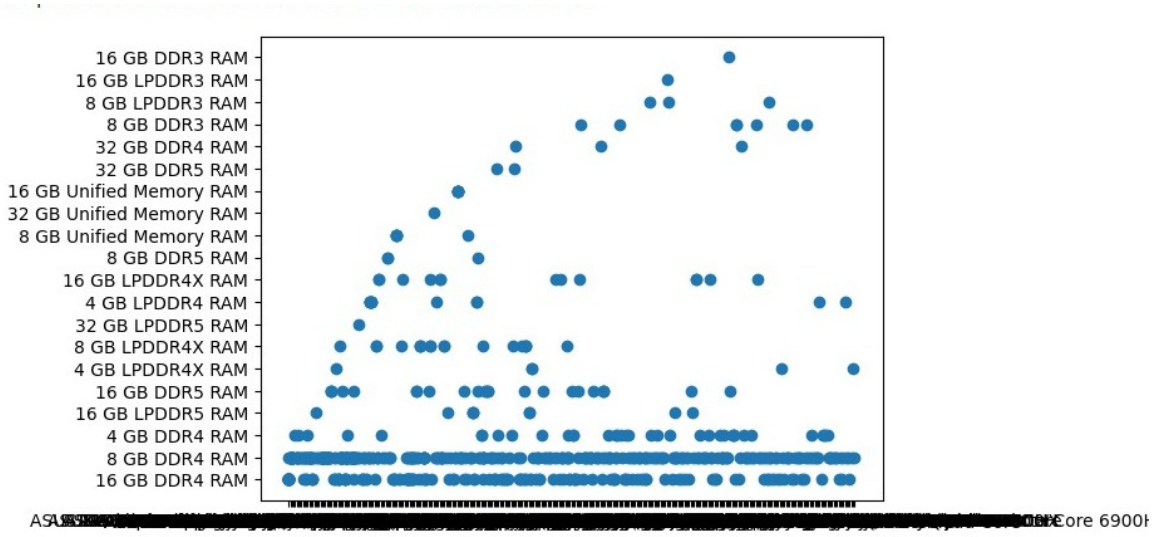



Figure 2.2: RAM Distribution

```
import numpy as np
import pandas as pd
import  matplotlib.pyplot as plt
import seaborn as sns
p=sns.barplot(k['processor'],k['price(in Rs.)'])
```

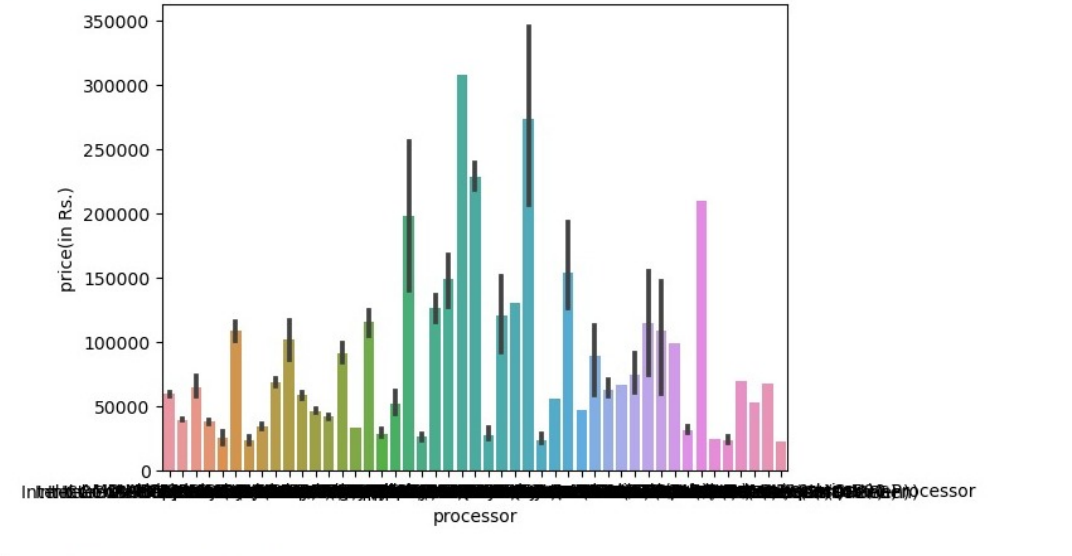



Figure 2.3: visual representation of the relationship between different processors and their prices.

```
import pandas as pd
```

```
import seaborn as sns
import  matplotlib.pyplot as plt
g=sns.JointGrid(x="rating",y="display(in inch)",data=d)
g=g.plot(sns.regplot,sns.distplot)
```

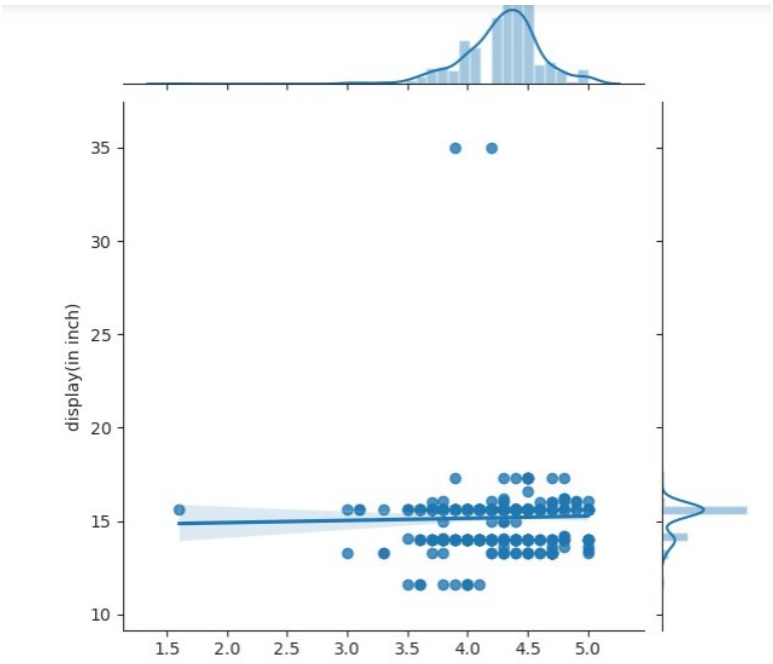



Figure 2.4: Distribution of ratings and display sizes

```
import seaborn as sns
import pandas as pd
import  matplotlib.pyplot as plt
plt.figure(figsize=(12,6))
d.groupby('name').size().sort_values(ascending=False).head(5).plot(kind = '
plt.xlabel('name of the laptop')
plt.ylabel('Number of Laptops')
plt.title('Top 5 most popular laptops brand')
plt.show()
```

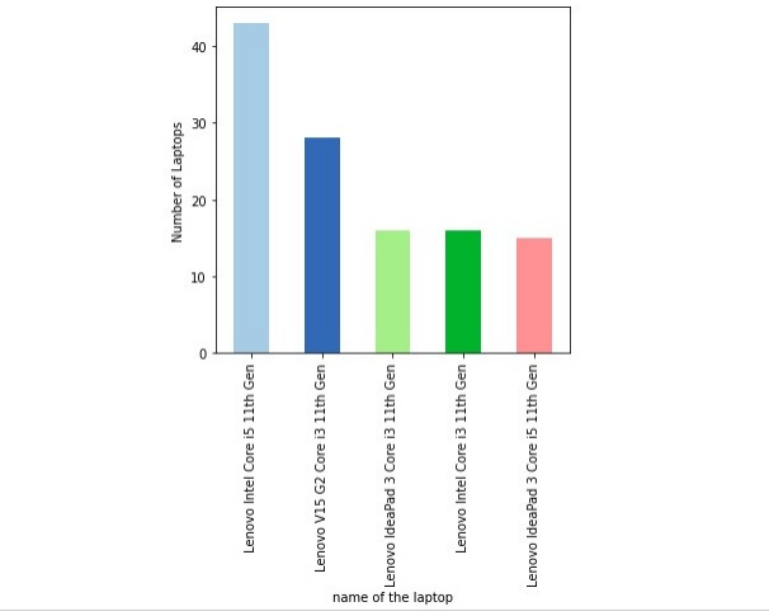



Figure 2.5: Top 5 most popular laptops brand

```
import seaborn as sns
import pandas as pd
import  matplotlib.pyplot as plt
sns.boxplot(x="display(in inch)",y="rating",data=d)
plt.show()
```

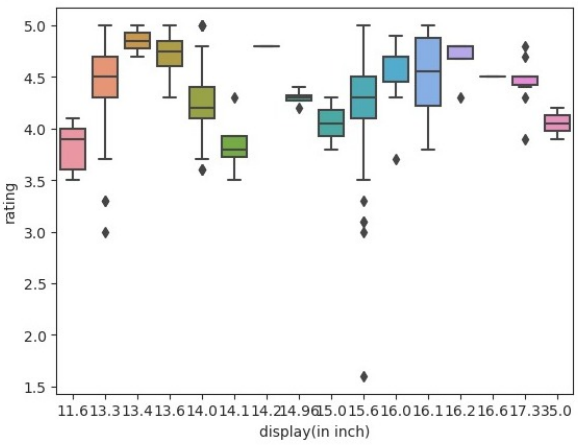



Figure 2.6: Example graph for box plot using seaborn

```
import seaborn as sns
import pandas as pd
import  matplotlib.pyplot as plt
```

```
sns.violinplot(x=d["rating"])
plt.show()
```

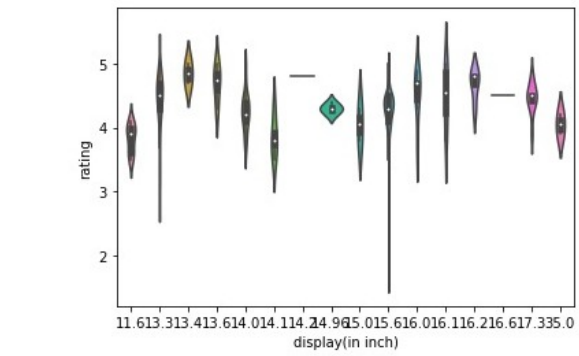



Figure 2.7: Distribution of ratings

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
plt.figure(figsize=(12,6))
d.groupby('ram').size().sort_values(ascending=False).plot(kind = 'bar',colo
plt.xlabel('Ram Size in GB')
plt.ylabel('Number of Laptops')
plt.show()
```

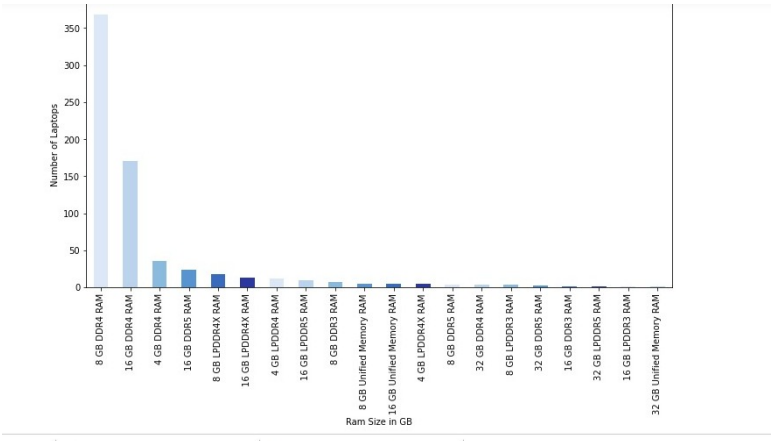



Figure 2.8: graph shows the number of laptops for each RAM category

```
import seaborn as sns
```

```
import pandas as pd
import matplotlib.pyplot as plt
sns.set_style("ticks")
sns.pairplot(d)
plt.show()
```

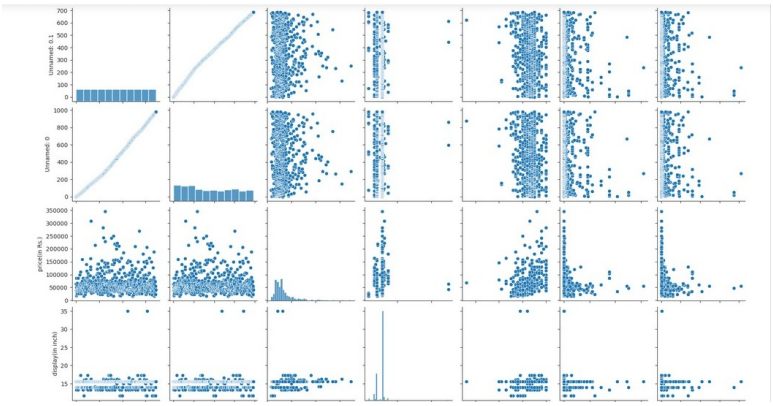


Figure 2.9: Relationships between variables

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
sns.distplot(d["price(in Rs.)"],kde=False)
plt.show()
```

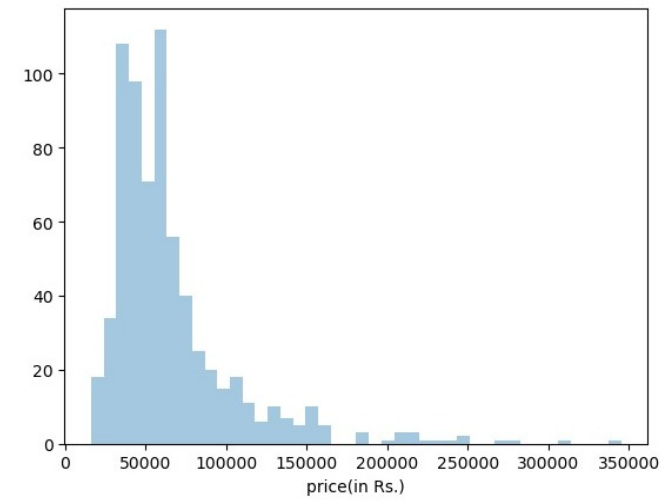



Figure 2.10: example graph of distplot

## 2.5 Visualization

```
plt.figure(figsize=(9,8))
sns.heatmap(d.corr(),square=True,annot=True,cmap="mako",center=0)
```

output:

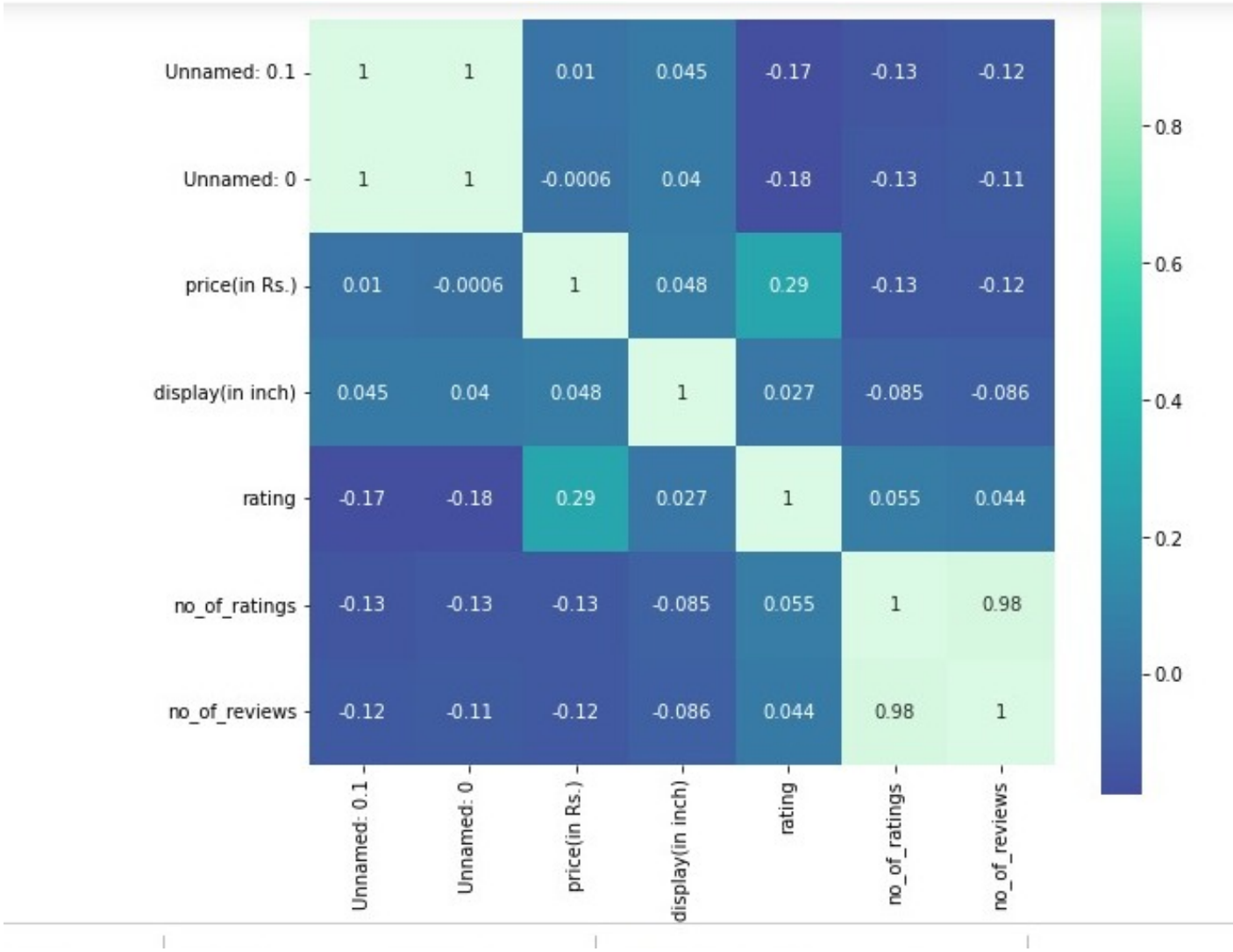



Figure 2.11: Example for Heatmap

```
sns.regplot(x="rating",y="display(in inch)",data=d)
plt.show()
```

output:

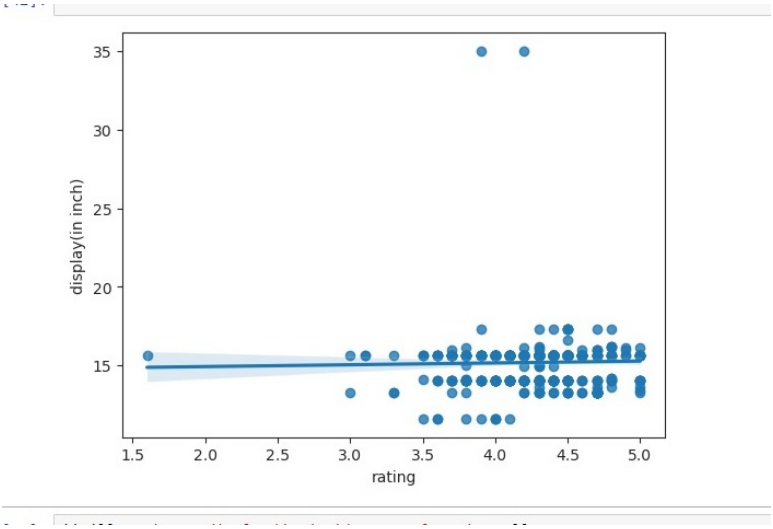



Figure 2.12: example graph for regplot

```
sns.distplot(d["rating"], hist=False)
plt.show()
```

output:

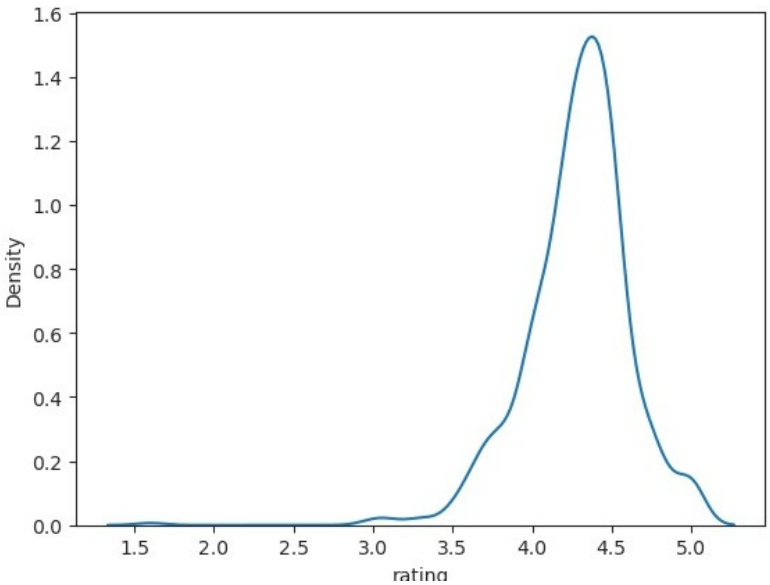



Figure 2.13: example graph for distplot

```
sns.swarmplot(x="display(in inch)",y="rating",data=d)
plt.show()
```

output: This section contain minimum of 12 pages

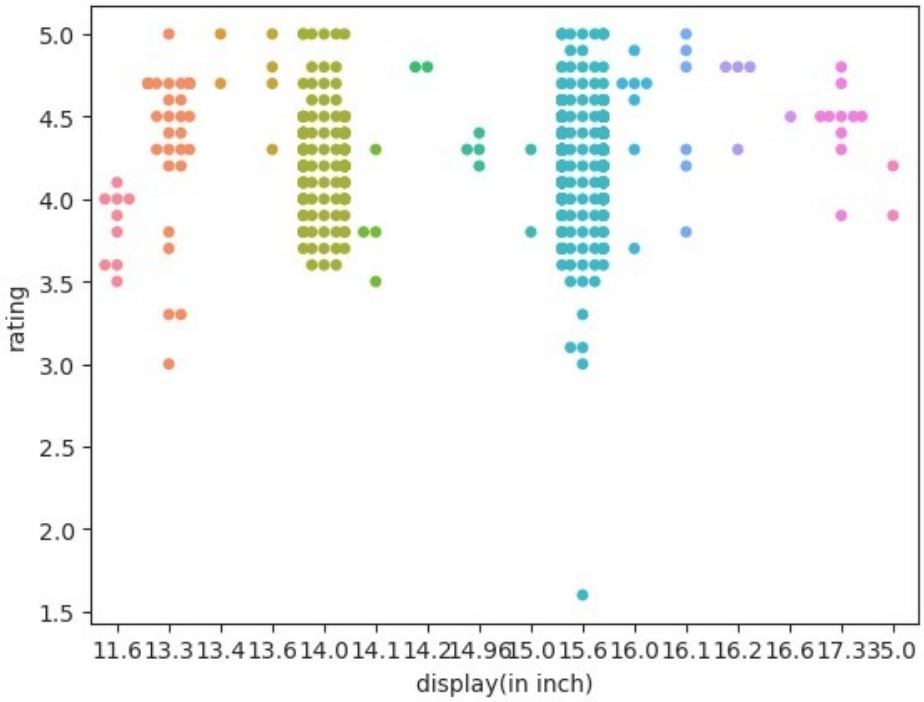



Figure 2.14: Example of swarmplot

# Chapter 3

# Code

## 3.1 Explain Your Code With Outputs

```
In [2]: import pandas as pd
        d=pd.read_csv("laptops.csv")
        d
```

Out[2]:

| | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings | no_of_reviews |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Intel Core i5 11th Gen | 62990 | Intel Core i5 Processor (11th Gen) | 16 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.5 | 14.0 | 1.0 |
| 1 | 1 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo V15 G2 Core i3 11th Gen | 37500 | Intel Core i3 Processor (11th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 1 TB HDD\|256 GB SSD | 15.6 | 4.4 | 53.0 | 3.0 |
| 2 | 2 | https://rukminim1.flixcart.com/image/312/312/l... | ASUS TUF Gaming F15 Core i5 10th Gen | 49990 | Intel Core i5 Processor (10th Gen) | 8 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.4 | 4733.0 | 463.0 |
| 3 | 3 | https://rukminim1.flixcart.com/image/312/312/x... | ASUS VivoBook 15 (2022) Core i3 10th Gen | 33990 | Intel Core i3 Processor (10th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 512 GB SSD | 15.6 | 4.3 | 10406.0 | 1040.0 |
| 4 | 4 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Athlon Dual Core | 18990 | AMD Athlon Dual Core Processor | 4 GB DDR4 RAM | DOS Operating System | 256 GB SSD | 14.0 | 3.8 | 18.0 | 3.0 |

Figure 3.1: csv file reading

```
In [13]: import pandas as pd
         d=pd.read_csv("laptops.csv")
         data=d["rating"]=d['rating'].fillna(0)
         d['no_of_ratings']=d['no_of_ratings'].fillna(0)
         d['no_of_reviews']=d['no_of_reviews'].fillna(0)
         d
```

Out[13]:

| | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Intel Core i5 11th Gen | 62990 | Intel Core i5 Processor (11th Gen) | 16 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.5 | 14.0 |
| 1 | 1 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo V15 G2 Core i3 11th Gen | 37500 | Intel Core i3 Processor (11th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 1 TB HDD\|256 GB SSD | 15.6 | 4.4 | 53.0 |
| 2 | 2 | https://rukminim1.flixcart.com/image/312/312/l... | ASUS TUF Gaming F15 Core i5 10th Gen | 49990 | Intel Core i5 Processor (10th Gen) | 8 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.4 | 4733.0 |
| 3 | 3 | https://rukminim1.flixcart.com/image/312/312/x... | ASUS VivoBook 15 (2022) Core i3 10th Gen | 33990 | Intel Core i3 Processor (10th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 512 GB SSD | 15.6 | 4.3 | 10406.0 |

Figure 3.2: filling the null values with 0

```
In [27]: h=d.describe()
         h
```

Out[27]:

| | Unnamed: 0.1 | Unnamed: 0 | price(in Rs.) | display(in inch) | rating | no_of_ratings | no_of_reviews |
|---|---|---|---|---|---|---|---|
| count | 688.000000 | 688.000000 | 688.000000 | 688.000000 | 688.000000 | 688.000000 | 688.000000 |
| mean | 343.500000 | 453.928779 | 66437.473837 | 15.184942 | 4.284884 | 718.091570 | 83.898256 |
| std | 198.752778 | 290.398881 | 40936.347731 | 1.411119 | 0.330239 | 1750.817825 | 211.596726 |
| min | 0.000000 | 0.000000 | 15990.000000 | 11.600000 | 1.600000 | 1.000000 | 0.000000 |
| 25% | 171.750000 | 193.750000 | 40355.000000 | 14.000000 | 4.100000 | 14.000000 | 2.000000 |
| 50% | 343.500000 | 432.500000 | 56049.000000 | 15.600000 | 4.300000 | 90.000000 | 11.000000 |
| 75% | 515.250000 | 711.250000 | 74990.000000 | 15.600000 | 4.500000 | 453.000000 | 53.500000 |
| max | 687.000000 | 982.000000 | 345390.000000 | 35.000000 | 5.000000 | 15492.000000 | 2054.000000 |

```
In [6]: #minimum    value present in the rating column
        l=d["rating"].min()
        l
```

Out[6]: 1.6

Figure 3.3: stastical data

```
In [57]: d.isnull()
```

Out[57]:

| | Unnamed: 0.1 | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings | no_of_reviews |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **1** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **2** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **3** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **4** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **683** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **684** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **685** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **686** | False | False | False | False | False | False | False | False | False | False | False | False | False |
| **687** | False | False | False | False | False | False | False | False | False | False | False | False | False |

688 rows × 13 columns

Figure 3.4: finding Null values

```
In [13]: #this is used for removing of the null value rows
         d= d.dropna()
         d
```

Out[13]:

| | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Intel Core i5 11th Gen | 62990 | Intel Core i5 Processor (11th Gen) | 16 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.5 | 14.0 |
| **1** | 1 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo V15 G2 Core i3 11th Gen | 37500 | Intel Core i3 Processor (11th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 1 TB HDD\|256 GB SSD | 15.6 | 4.4 | 53.0 |
| **2** | 2 | https://rukminim1.flixcart.com/image/312/312/l... | ASUS TUF Gaming F15 Core i5 10th Gen | 49990 | Intel Core i5 Processor (10th Gen) | 8 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.4 | 4733.0 |
| **3** | 3 | https://rukminim1.flixcart.com/image/312/312/x... | ASUS VivoBook 15 (2022) Core i3 10th Gen | 33990 | Intel Core i3 Processor (10th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 512 GB SSD | 15.6 | 4.3 | 10406.0 |
| **4** | 4 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Athlon Dual Core | 18990 | AMD Athlon Dual Core Processor | 4 GB DDR4 RAM | DOS Operating System | 256 GB SSD | 14.0 | 3.8 | 18.0 |

Figure 3.5: Removing the null values containg rows

```
In [17]: # RESET THE INDICES
         d= d.reset_index(drop=True)
         d
```

Out[17]:

| | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings | no |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Intel Core i5 11th Gen | 62990 | Intel Core i5 Processor (11th Gen) | 16 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.5 | 14.0 | |
| 1 | 1 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo V15 G2 Core i3 11th Gen | 37500 | Intel Core i3 Processor (11th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 1 TB HDD\|256 GB SSD | 15.6 | 4.4 | 53.0 | |
| 2 | 2 | https://rukminim1.flixcart.com/image/312/312/l... | ASUS TUF Gaming F15 Core i5 10th Gen | 49990 | Intel Core i5 Processor (10th Gen) | 8 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.4 | 4733.0 | |
| 3 | 3 | https://rukminim1.flixcart.com/image/312/312/x... | ASUS VivoBook 15 (2022) Core i3 10th Gen | 33990 | Intel Core i3 Processor (10th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 512 GB SSD | 15.6 | 4.3 | 10406.0 | |
| 4 | 4 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Athlon Dual Core | 18990 | AMD Athlon Dual Core | 4 GB DDR4 RAM | DOS Operating System | 256 GB SSD | 14.0 | 3.8 | 18.0 | |

Figure 3.6: Reset the indices

```
In [18]: # saving the dataframe
         d.to_csv('file2.csv')
         d
```

Out[18]:

| | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Intel Core i5 11th Gen | 62990 | Intel Core i5 Processor (11th Gen) | 16 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.5 | 14.0 |
| 1 | 1 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo V15 G2 Core i3 11th Gen | 37500 | Intel Core i3 Processor (11th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 1 TB HDD\|256 GB SSD | 15.6 | 4.4 | 53.0 |
| 2 | 2 | https://rukminim1.flixcart.com/image/312/312/l... | ASUS TUF Gaming F15 Core i5 10th Gen | 49990 | Intel Core i5 Processor (10th Gen) | 8 GB DDR4 RAM | Windows 11 Operating System | 512 GB SSD | 15.6 | 4.4 | 4733.0 |
| 3 | 3 | https://rukminim1.flixcart.com/image/312/312/x... | ASUS VivoBook 15 (2022) Core i3 10th Gen | 33990 | Intel Core i3 Processor (10th Gen) | 8 GB DDR4 RAM | 64 bit Windows 11 Operating System | 512 GB SSD | 15.6 | 4.3 | 10406.0 |
| 4 | 4 | https://rukminim1.flixcart.com/image/312/312/x... | Lenovo Athlon Dual Core | 18990 | AMD Athlon Dual Core | 4 GB DDR4 RAM | DOS Operating System | 256 GB SSD | 14.0 | 3.8 | 18.0 |

Figure 3.7: After data processing saving dataframe into csv file

In [58]:
```
d.notnull()
```

Out[58]:

| | Unnamed: 0.1 | Unnamed: 0 | img_link | name | price(in Rs.) | processor | ram | os | storage | display(in inch) | rating | no_of_ratings | no_of_reviews |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 1 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 2 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 3 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 4 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 683 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 684 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 685 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 686 | True | True | True | True | True | True | True | True | True | True | True | True | True |
| 687 | True | True | True | True | True | True | True | True | True | True | True | True | True |

688 rows × 13 columns

d.isna().sum()

Figure 3.8: checking the null values

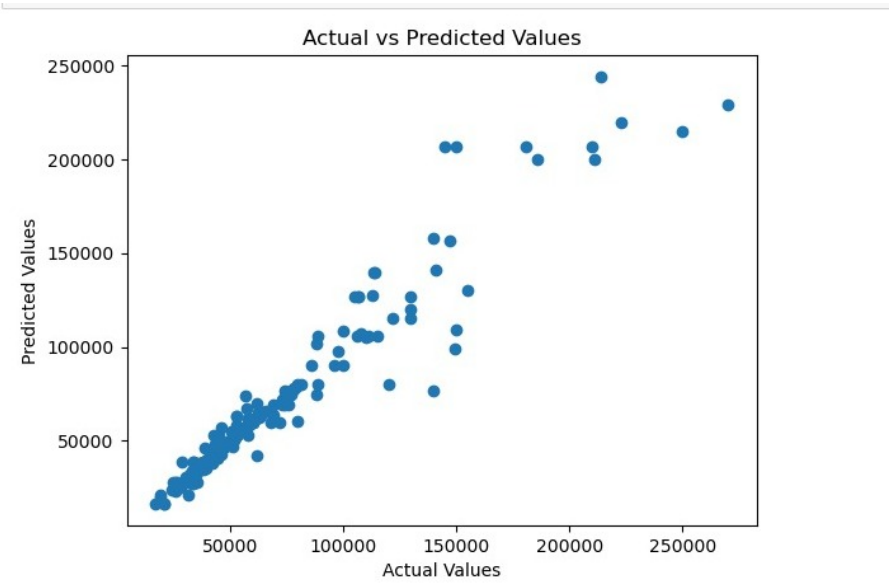
```
import pandas as pd
import numpy as np
k=pd.read_csv("file2.csv")
k=pd.DataFrame(k)
k.sort_values(['price(in_Rs.)'], inplace=True)
categorical_features =k.columns[(k.dtypes == 'object') == True].to_list()
print(categorical_features)
for feature in categorical_features:
    uniq = k[feature].unique()
    new_feature = []
    for el in k[feature]:
        new_feature.append(len(uniq) - np.where(uniq==el)[0][0])
    k[feature] = new_feature
k
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn import metrics
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_sco
import  matplotlib.pyplot as plt
```

```
import seaborn as sb
x=k.drop(['price(in Rs.)'], axis=1)
y=k["price(in Rs.)"].values
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, ra
rf_classifier = RandomForestClassifier(n_estimators=100)
rf_classifier.fit(x_train, y_train)
y_pred = rf_classifier.predict(x_test)
# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print("accuracy: ",accuracy)
MAE = mean_absolute_error(y_test, y_pred)
MSE = mean_squared_error(y_test, y_pred)
R2 = r2_score(y_test, y_pred)
print("Mean Absolute Error:", MAE)
print("Mean Squared Error:", MSE)
print("R-squared:", R2)
```

output:


```
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted Values')
plt.show()
```

output:




\

```
plt.plot(range(len(y_test)), y_test, color='blue', label='Actual')
plt.plot(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.xlabel('Data Point Index')
plt.ylabel('Price')
plt.legend()
plt.show()
```

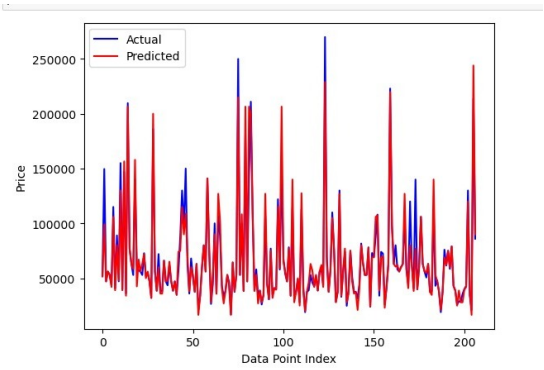



Figure 3.9: Actual vs Predicted values

```
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.plot(range(len(y_pred)), y_pred, color='green', linewidth=2, label='Bes
plt.xlabel('Data Point Index')
plt.ylabel('Price')
plt.legend()
plt.show()
```

output:

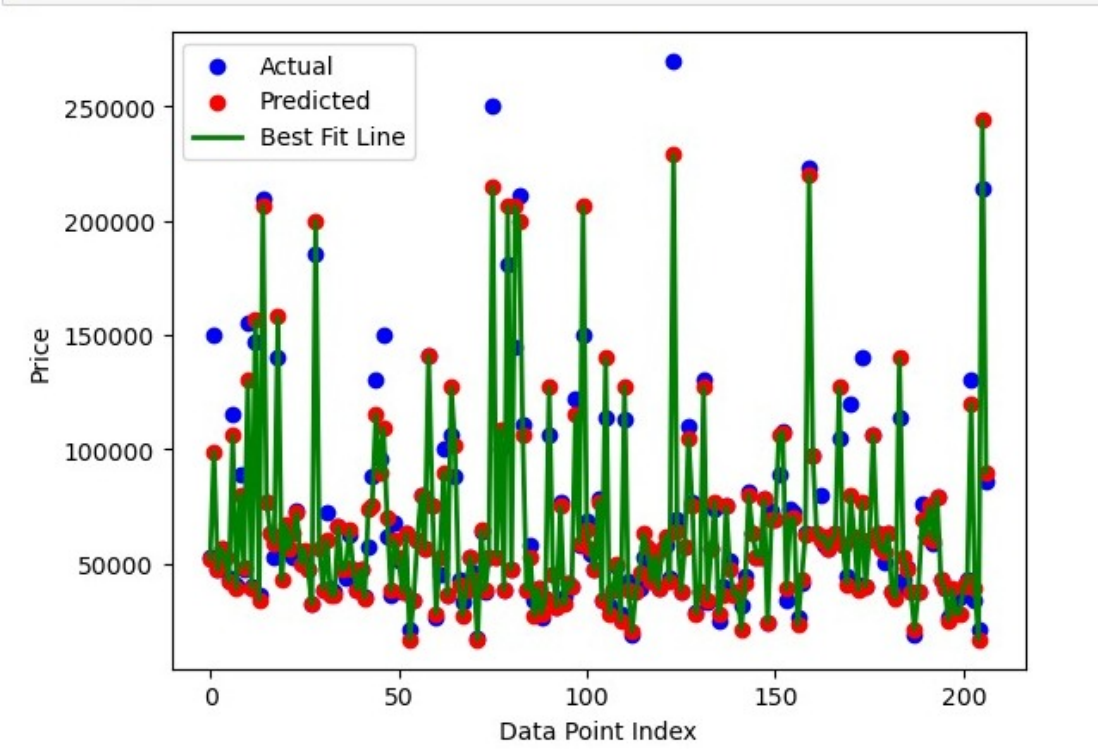



Figure 3.10: Best Fit Line For Actual vs Predicted Values

```
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.xlabel('Data Point Index')
plt.ylabel('Price')
plt.legend()
plt.show()
```

output:

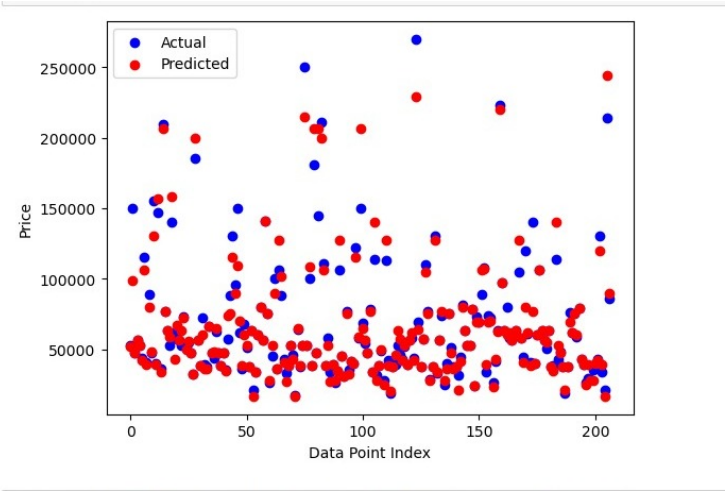



Figure 3.11: scatter plot for actual price vs predicted price

```
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
k = pd.read_csv('file2.csv')

categorical_features = k.columns[(k.dtypes == 'object') == True].to_list()
print(categorical_features)
for feature in categorical_features:
    uniq = k[feature].unique()

    new_feature = []
    for el in k[feature]:
        new_feature.append(len(uniq) - np.where(uniq == el)[0][0])

    k[feature] = new_feature

features = ['storage', 'ram', 'os']
target = 'name'
X = k[features]
y = k[target]

rf_classifier = RandomForestClassifier(n_estimators=100)
```

```
rf_classifier.fit(X, y)


new_features = ["11", "15", "20"]
new_laptop = pd.DataFrame([new_features], columns=features)
prediction = rf_classifier.predict(new_laptop)
selected_laptop = k[k[target] == prediction[0]]


print("Selected Laptop:")
print(selected_laptop)
```

```
['img_link', 'name', 'processor', 'ram', 'os', 'storage']
Selected Laptop:
     Unnamed: 0.1  Unnamed: 0  img_link  name  price(in Rs.)  processor  ram  \
31             31          32       373   330          99990         34   16
368           368         479       171   330         105990         34   16
627           627         886       373   330         107990         34   16

     os  storage  display(in inch)  rating  no_of_ratings  no_of_reviews
31   10       15             15.6     4.4          127.0           16.0
368  10       12             15.6     4.5           90.0           18.0
627  11       12             15.6     4.0           17.0            4.0
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_squared_error, r2_score
X=k.drop(['price(in Rs.)'], axis=1)
y=k["price(in Rs.)"].values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ra
rf = RandomForestRegressor()
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print('Mean Squared Error:', mse)
print('R-squared:', r2)
print("Actual values:")
print(y_test)
```

```
print("Predicted values:")
print(y_pred)
output:
R-squared value:0.9740704887906095
```

# Chapter 4

# Conclusion and Future Work

In conclusion, a laptop selection and price prediction project aims to assist users in making informed decisions when choosing a laptop by estimating its price based on its specifications. The project involves several key steps, including data collection, preprocessing, feature selection, model training, evaluation, and prediction. By analyzing features such as brand, model, processor, RAM, storage, display size, and others, machine learning models can be trained to predict the price of a laptop. Regression techniques like linear regression, decision trees, random forests, or gradient boosting are commonly used for price prediction. Classification techniques can also be valuable for categorizing laptops based on price ranges, feature importance, brand, laptop type, or sentiment analysis. Through the laptop selection and price prediction process, users can benefit from a better understanding of the factors influencing laptop prices and the ability to compare different laptops based on their preferences and budget constraints. The project provides a valuable tool to assist users in selecting the most suitable laptop for their specific needs and helps them make informed purchasing decisions. It's important to note that while machine learning models can provide price estimates, actual laptop prices may still vary due to various factors such as market trends, discounts, or economic variables. The accuracy of the price prediction models depends on the quality and representativeness of the dataset, the choice of features, and the chosen machine learning techniques. Overall, a laptop selection and price prediction project helps users navigate the wide range of laptops available in the market by providing estimates of their prices based on specifications. This empowers users to make

well-informed decisions and find laptops that best meet their requirements and budget.