

Capstone Project 3

Mobile Price Range Prediction

Submitted by: Karan Tiwari

Content

Introduction

Problem Statement

Data Description

Data Cleaning

Exploratory Data Analysis

Data modeling

Machine learning Models

Conclusion

Introduction

- **Nowadays, having a mobile phone is practically necessary to stay in touch with the outside world. As a result, massive amounts of mobile phones are getting manufactured and due to that lot of data is also generating**
- **The more expensive phones will have a lot more features than the less expensive ones, mobile phone forecasts can be useful for evaluating a phone's price range based on its characteristics. This information can be used to inform industry-level decisions on mobile phone specifications.**

Problem Statement

- **In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices. The objective is to find out some relation between features of a mobile phone(eg:- RAM, Internal Memory, etc) and its selling price. In this problem, we do not have to predict the actual price but a price range indicating how high the price is.**

Data Description

Battery_power -Total energy a battery can store in one time measured in mAh
Blue -Has Bluetooth or not
Clock_speed-speed at which microprocessor executes instructions
Dual_sim-Has dual sim support or not
Fc-Front Camera megapixels
Four_g-Has 4G or not
Int_memory-Internal Memory in Gigabytes
M_dep-Mobile Depth in cm
Mobile_wt-Weight of mobile phone
N_cores-Number of cores of processor
Pc -Primary Camera megapixels
Px_height-Pixel Resolution Height
Px_width-Pixel Resolution Width
Ram-Random Access Memory in Megabytes
Sc_h-Screen Height of mobile in cm
Sc_w-Screen Width of mobile in cm
Talk_time-longest time that a single battery charge will last when you are
Three_g-Has 3G or not
Touch_screen-Has touch screen or not
Wifi-Has wifi or not
Price_range-This is the target variable with value of 0(low cost), 1(medium cost), 2(high cost)

Dataset have 2000 rows and 21 columns

Data Cleaning

Dataset

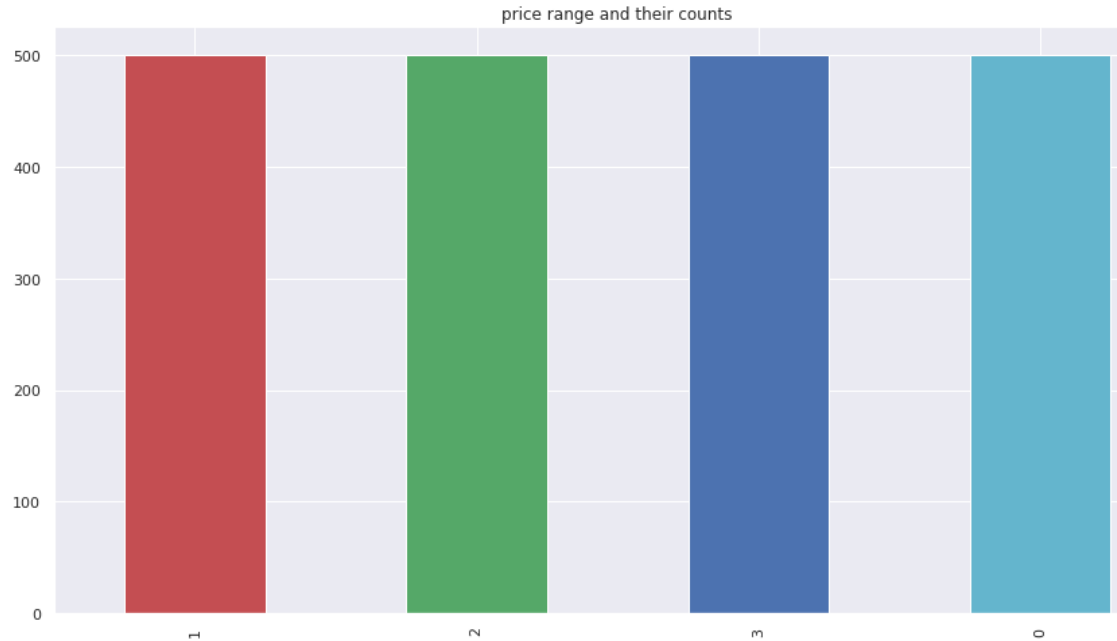
- Does not have any outlier.
- Does not have any null values.
- Does not have any duplicate rows

So we don't have to do any transformation.



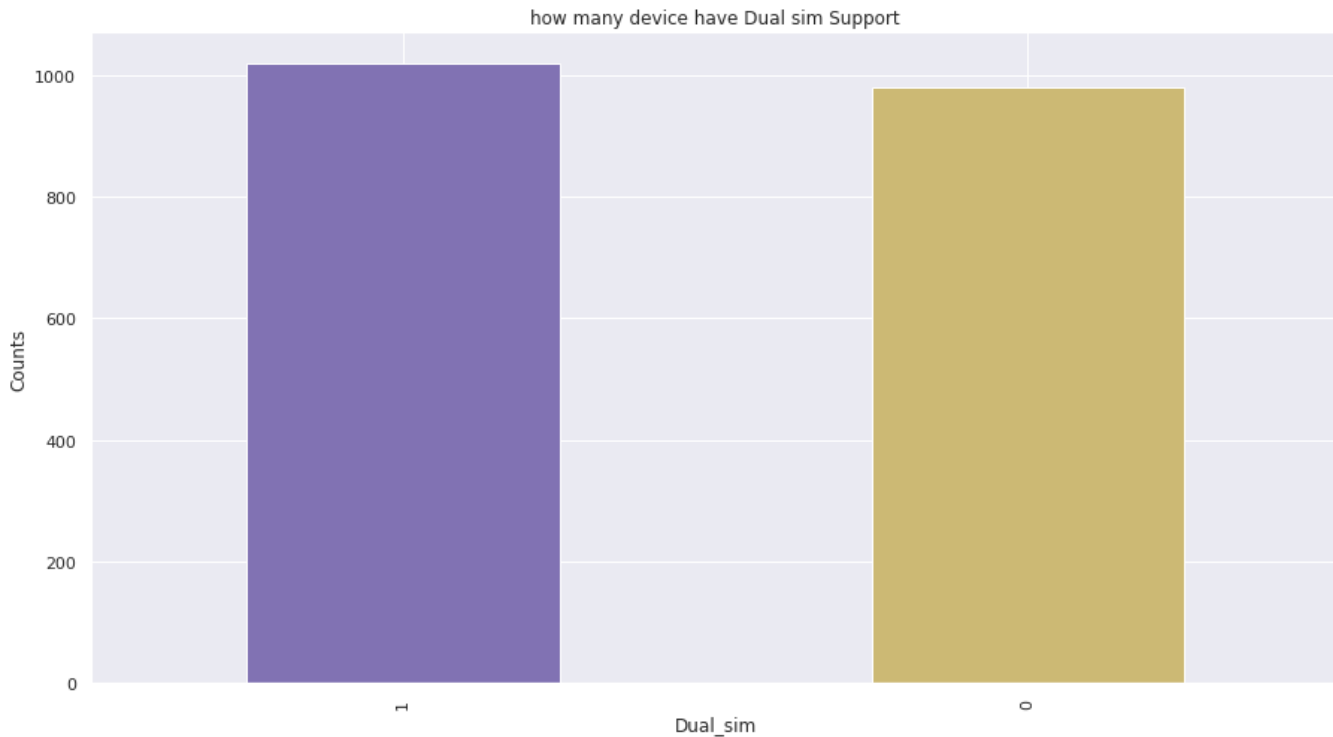
Exploratory Data Analysis

Price range and their counts.



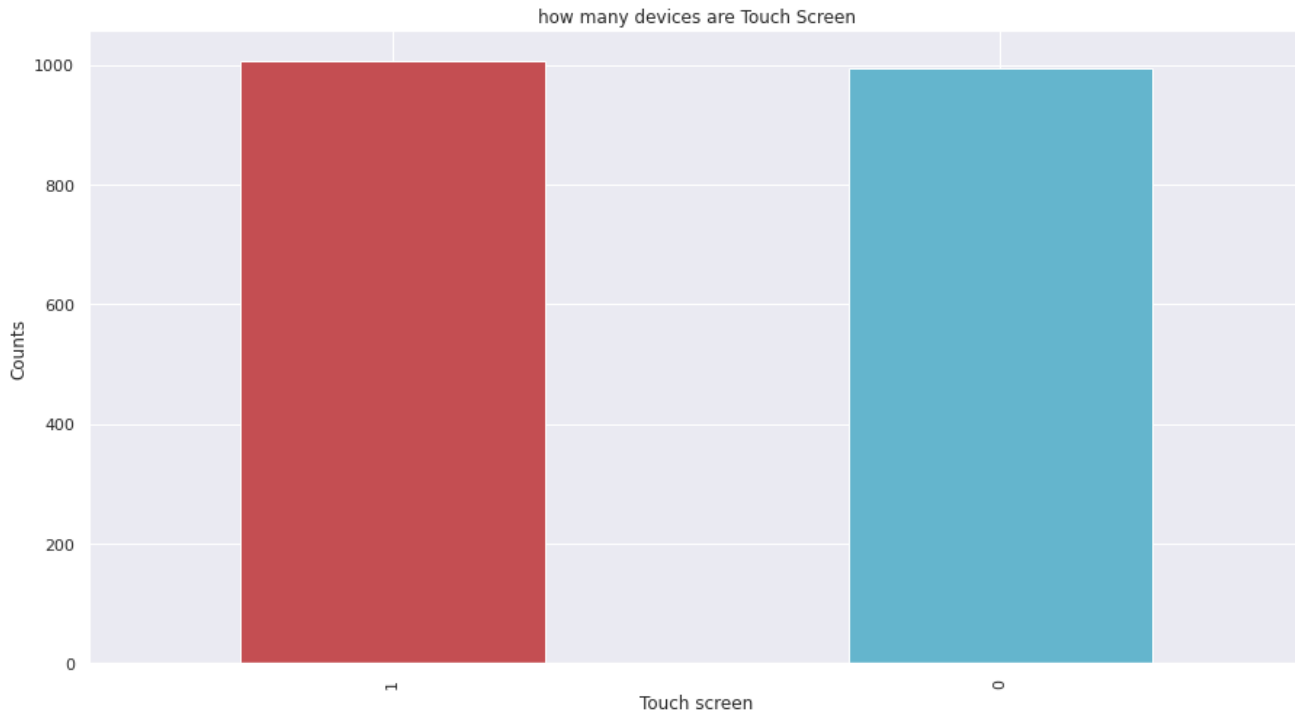
There are 4 types of mobile phones ,their price range labeled as 0,1,2,3 and all are present in equal counts.

How many devices have Dual Sim Support?



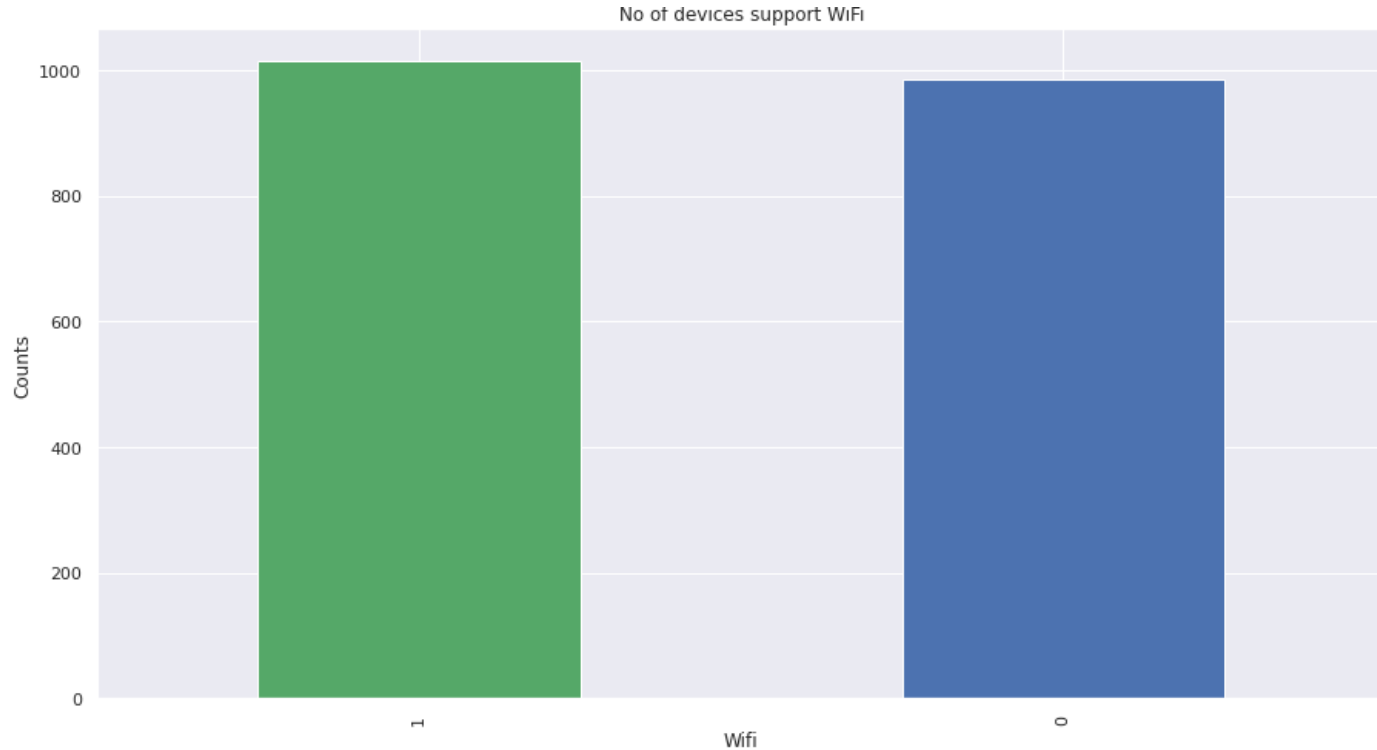
Almost 50% of the mobile phones have dual sim.

How many devices are Touch Screen?



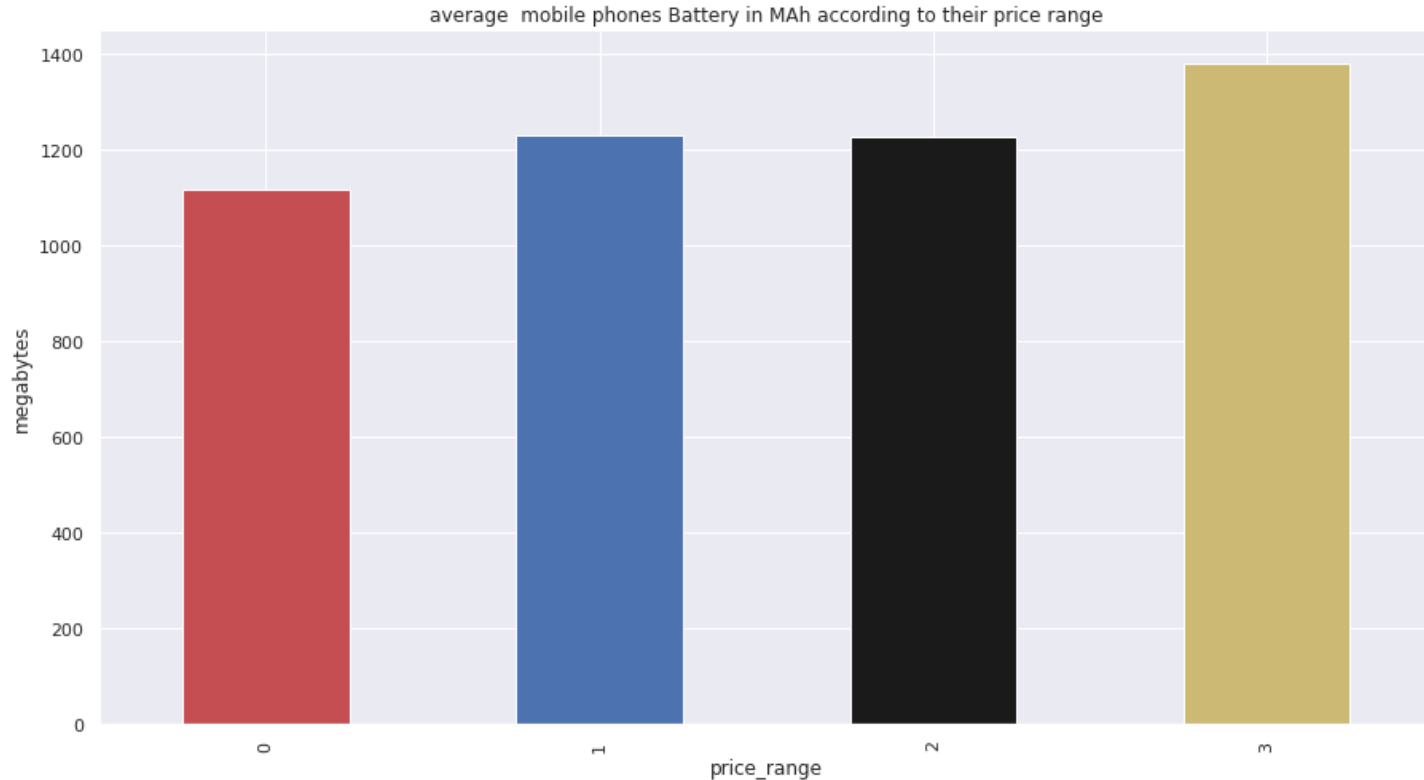
Almost 50% of the mobile phones are Touch screen.

How many devices Support Wi-Fi?



Almost 50% of the mobile phones have Wi-Fi connectivity.

Average mobile phones Battery in MAh according to their price range.



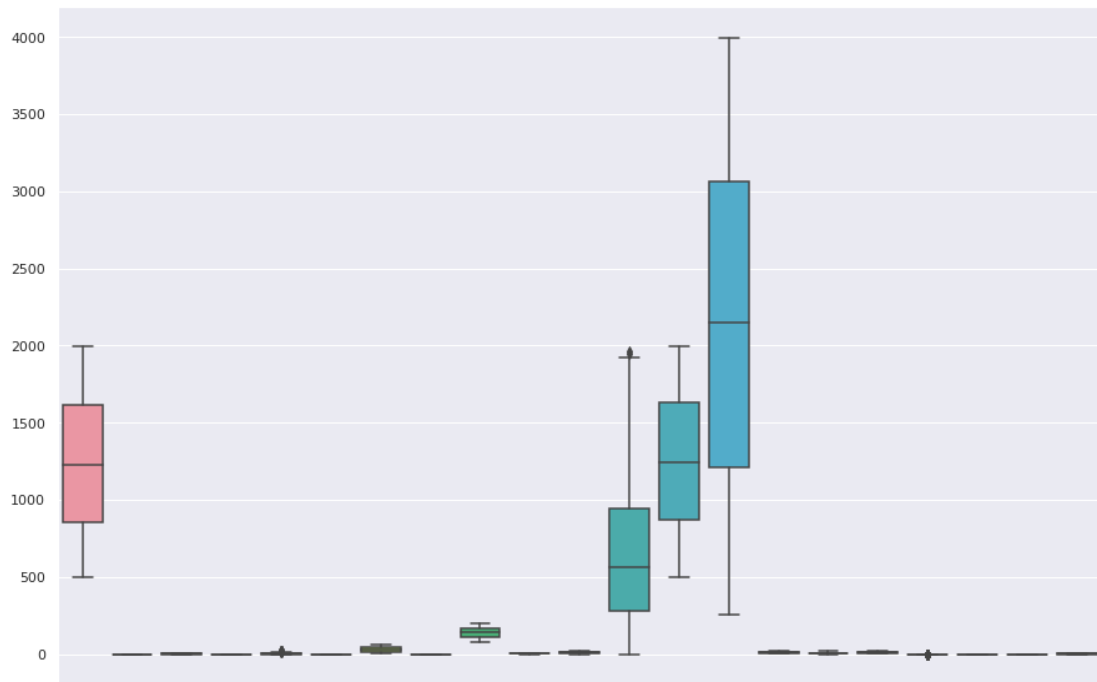
For price range 0 the average battery is around 1100 mAh, for price range 1 and 2 it's around 1230 mAh and for price range 3 it's around 1370 mAh.

Average mobile phones Ram in megabytes according to their price range



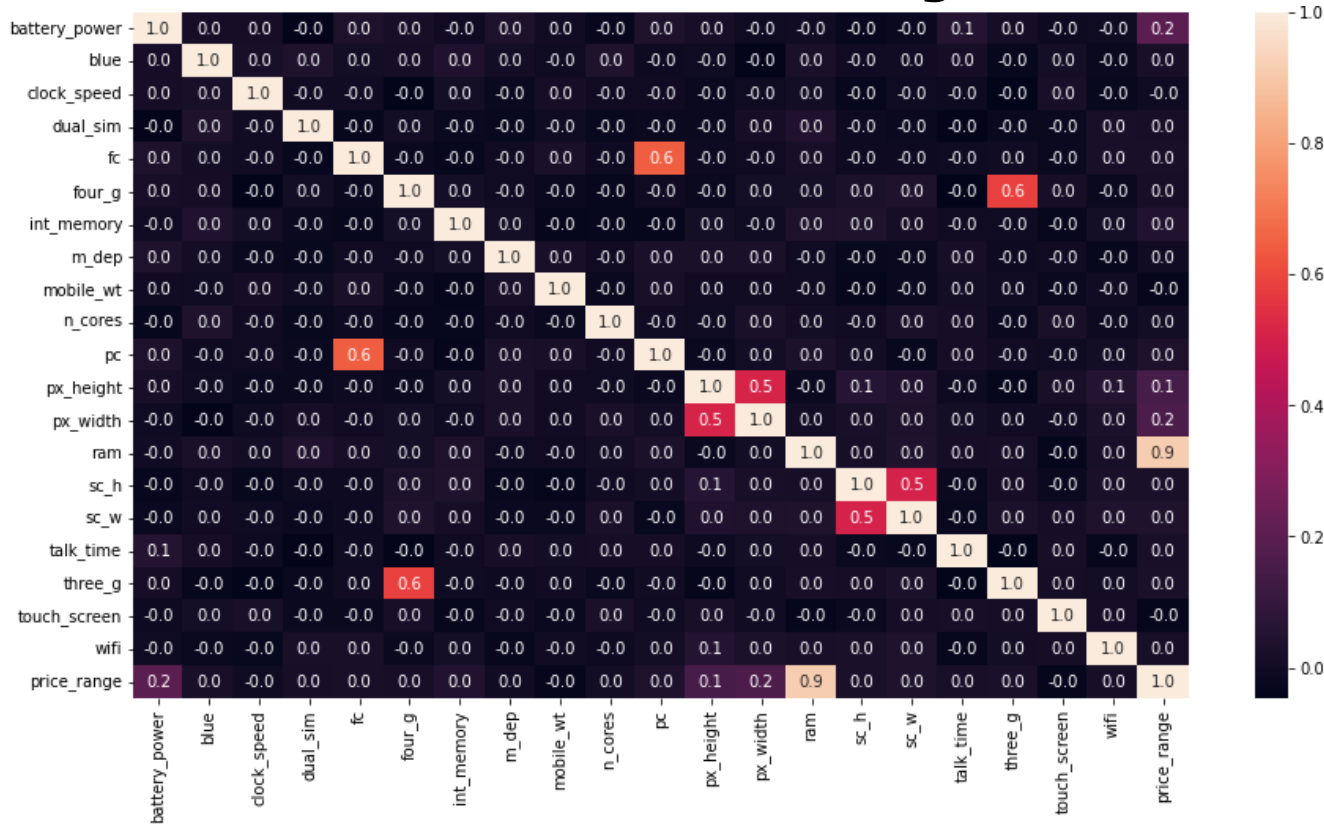
The average ram for price range 0,1,2,3 is approx 700,1650,2600 and 3400 megabytes respectively.

Outliers



No outliers found!

Multicollinearity



There is no highly correlated data inputs in our dataset, so there is no multicollinearity problem.

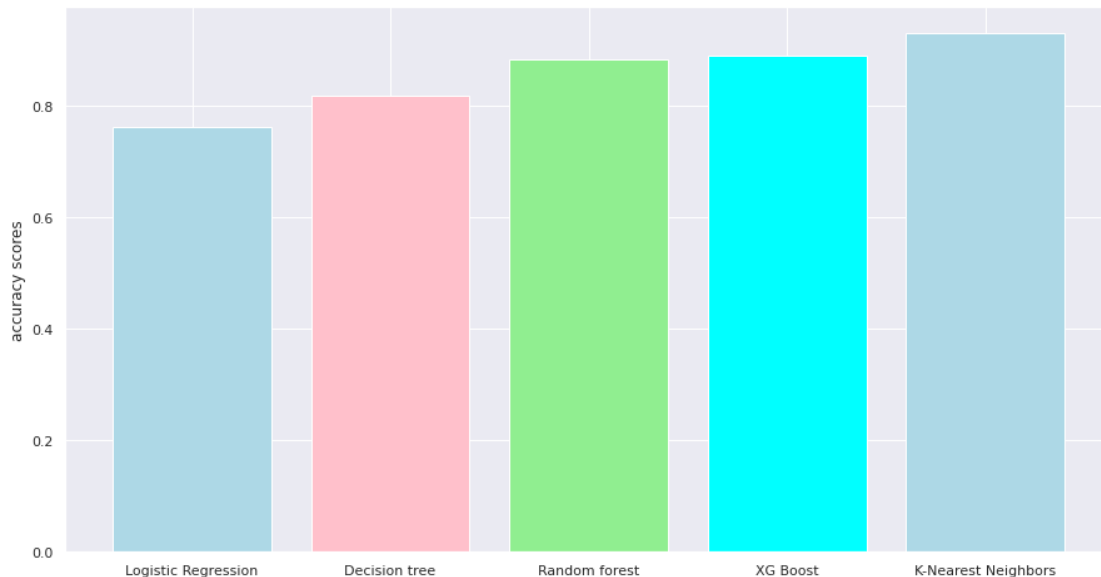
DATA MODELLING

- **First Assigning independent and dependent variable.**
- **Then do train test split in which 20% is test set and rest is for training set.**
- **x_train shape is 1600 rows and 20 columns.**
- **x_test shape is 400 rows and 20 columns.**



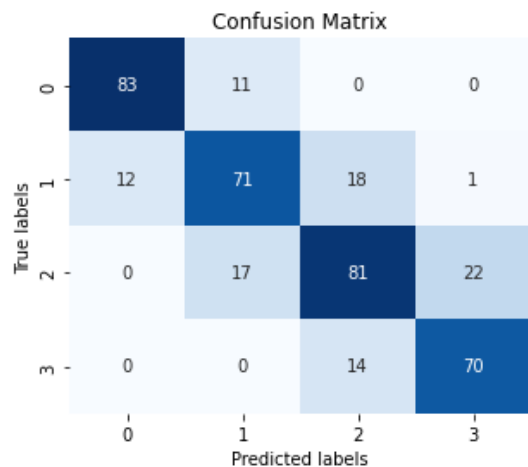
Implementing Machine learning Models

	Model_Name	Model_accuracy%
0	Logistic Regression	76
1	Decision Tree	82
2	Random forest	88
3	XG Boost	89
4	K-Nearest Neighbors	93

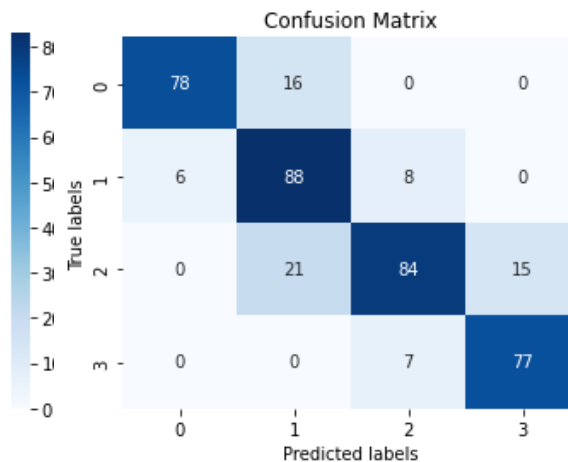


ML Models with their accuracy score .

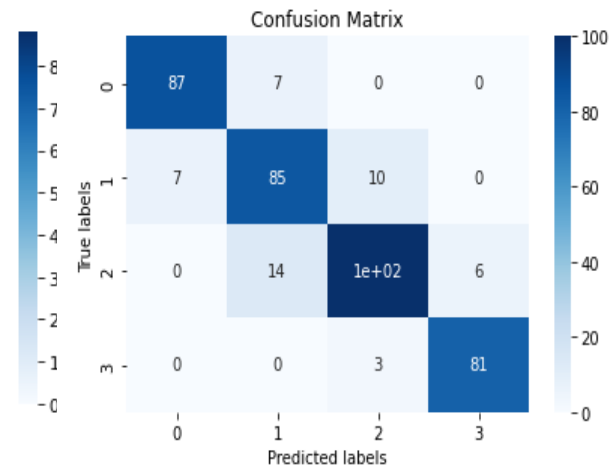
Confusion matrix of ML Models



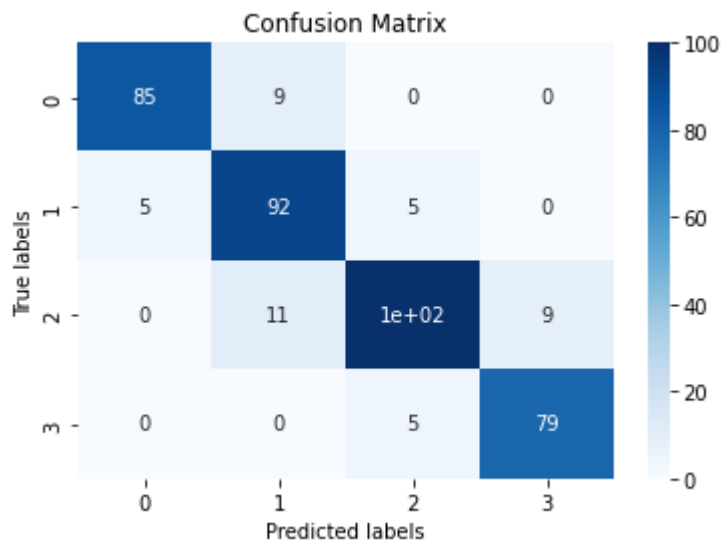
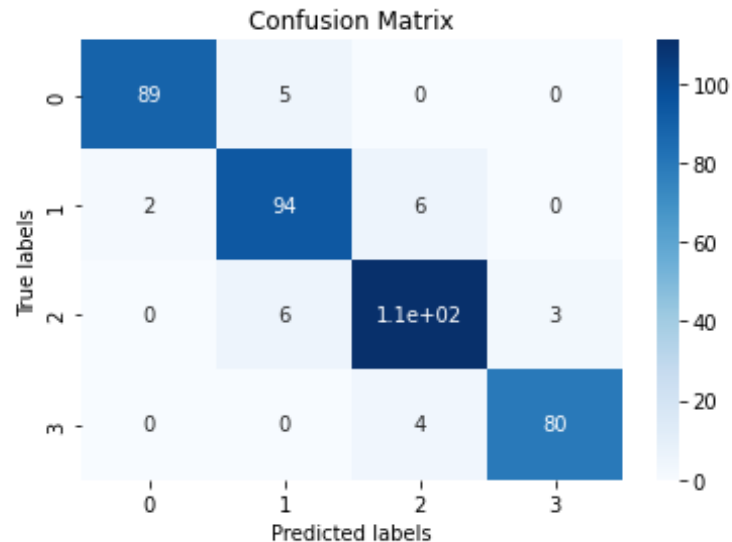
Logistic regression



Decision Tree



Random Forest

**XG Boost****K-Nearest Neighbors**

Conclusion

1: This data contains 2000 rows and 21 columns.

2: Data does not have any null value.

3: Data does not have any outlier.

4: EDA

1: There are 4 types of mobile phones, their price range labeled as 0, 1, 2, 3 and all are present in equal counts.

2: Almost 50% of the mobile phones have dual sim.

3: Almost 50% of the mobile phones are Touch screen.

4: Almost 50% of the mobile phones have Wi-Fi connectivity.

5: For price range 0 the average battery is around 1100 mAh, for price range 1 and 2 it's around 1230 mAh and for price range 3 it's around 1370 mAh.

6: The average ram for price range 0, 1, 2, 3 is approx 700, 1650, 2600 and 3400 megabytes respectively.

7: Those mobile phones which are of class 0 are the cheaper phones and those mobiles which come in class three are the expensive mobile phones.

5: There is no highly correlated data inputs in our dataset, so there is no multicollinearity problem.

6: Then we apply Logistic Regression, Decision Tree, Random Forest, XGBoost and K-Nearest Neighbors machine learning models.

7: The accuracy of logistic regression is 76% and it was the least accuracy we got among all ML models.

8: The accuracy of Decision tree model is 82%.

9: The accuracy of Random forest model is 88%.

10: The accuracy of XGBoost model is 89%.

11: The accuracy of K-Nearest Neighbors model is 93%.

12: So we conclude that K-Nearest Neighbors is our best model.