AI

Mobile Price Range Prediction

(Machine Learning: Classification)

PROBLEM STATEMENT

- In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices.

PROBLEM DESCRIPTION

- 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?

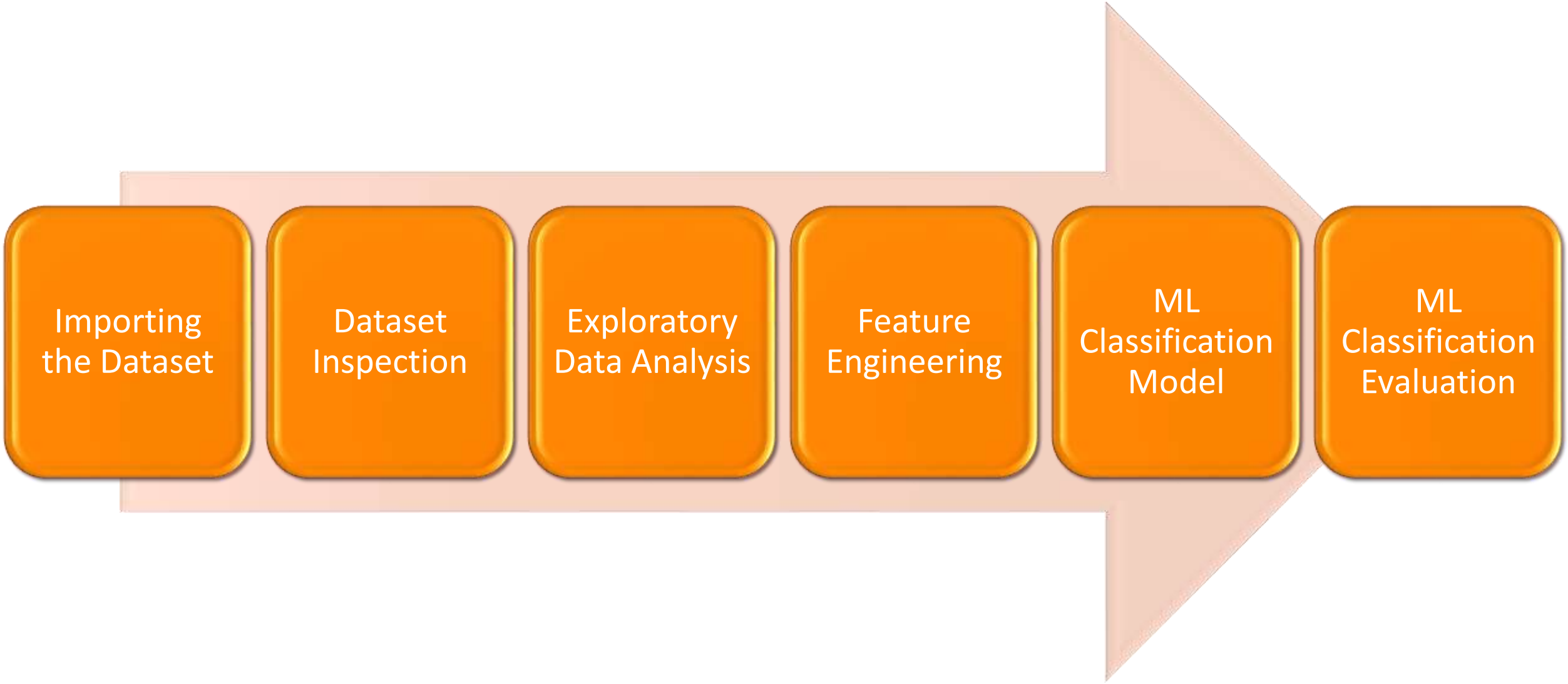
DATA DESCRIPTION

Variables 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 mega pixels
- **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 mega pixels
- **Px_height -** Pixel Resolution Height
- **Px_width -** Pixel Resolution Width
- **Ram -** Random Access Memory in Mega Bytes
- **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 Wi-Fi or not
- **Price_range -** This is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

CONTENTS

STEPS INVOLVED



DATA SUMMARY

Name of the Dataset	data_mobile_price_range.csv
Number of variables	21
Number of observations	2000
Duplicate rows	0
Total size in memory	100 kb
Missing Data (Columns)	0

VARIABLE DATA TYPE

Date Type	Column	
Numeric - int64	0) battery_power 1) blue 3) dual_sim 4) fc 5) four_g 6) int_memory 8) mobile_wt 9) n_cores 10) pc	11) px_height 12) px_width 13) ram 14) sc_h 15) sc_w 16) talk_time 17) three_g 18) touch_screen 19) wifi 20) price_range
Numeric - float64	2) clock_speed 7) m_dep	

HANDLING MISSING DATA

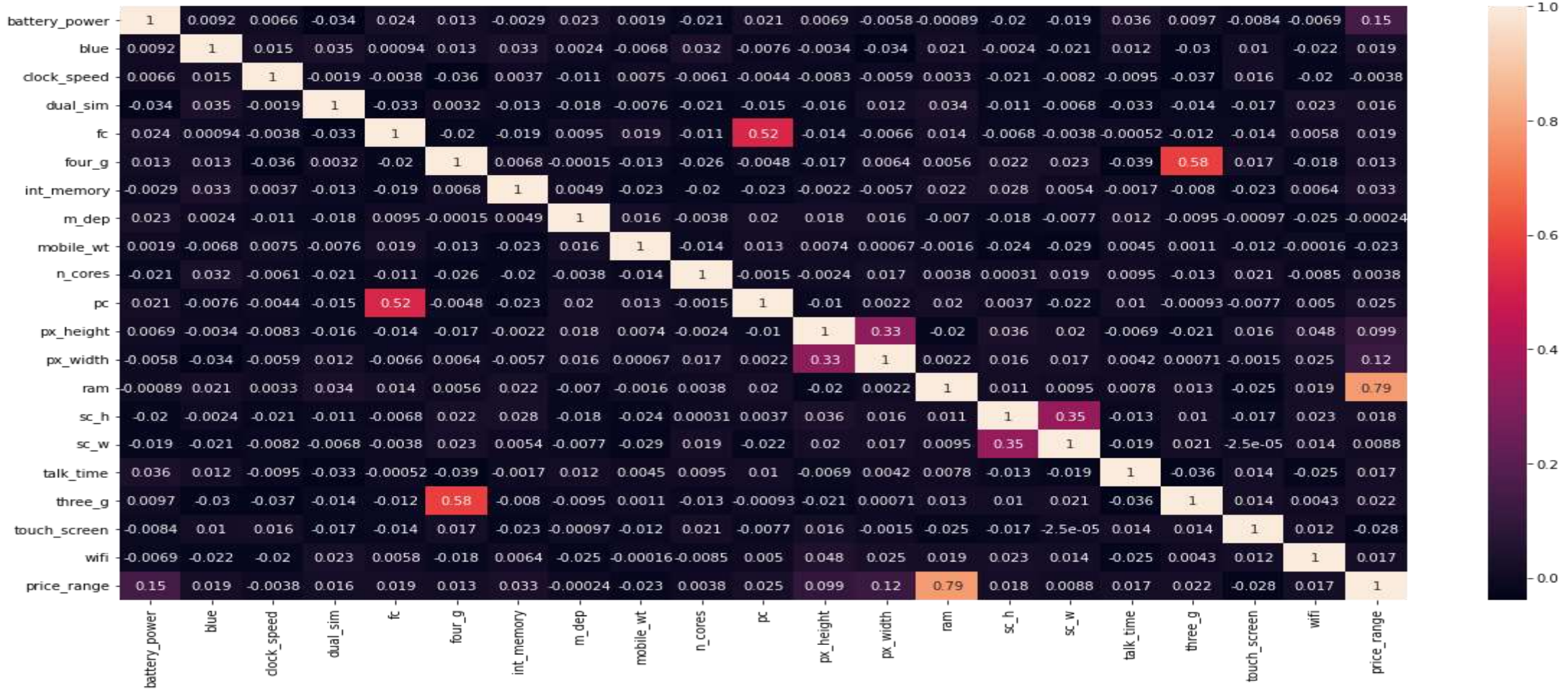
- The column px_height and sc_w contain null values .

  px_height = 180

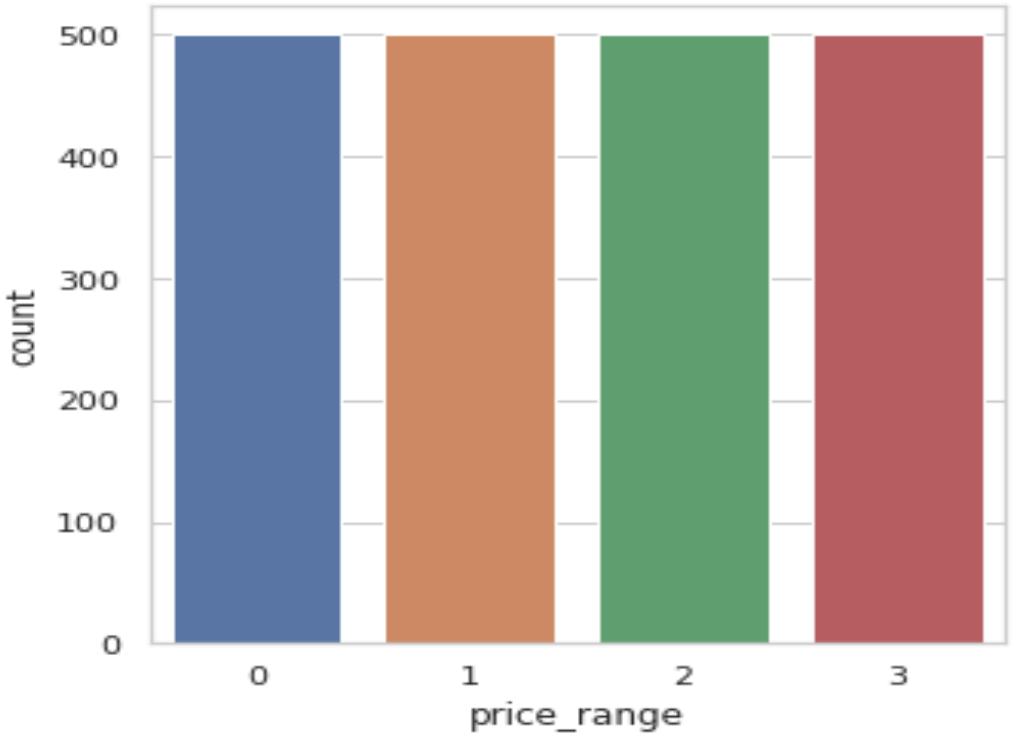
  sc_w = 2

- Minimum value of px_height and sc_w cannot be zero.
- So where there is sc_w and px_height is zero ,we have assigned mean values.

CORRELATION BETWEEN DIFFERENT VARIABLES

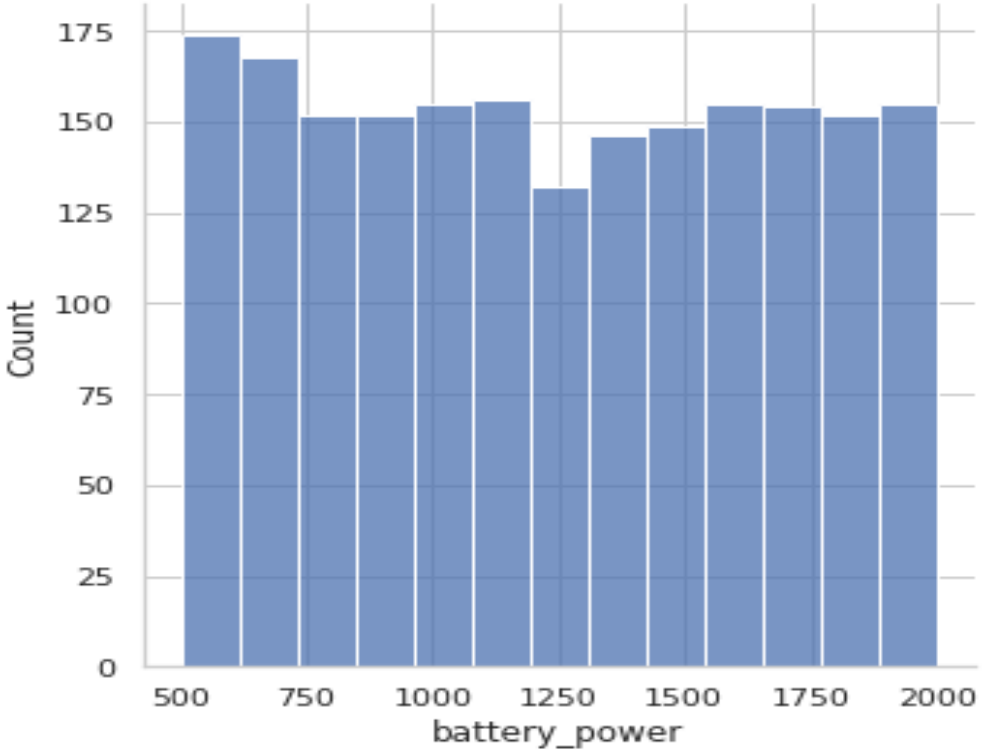


MOBILE PHONE PRICE RANGE



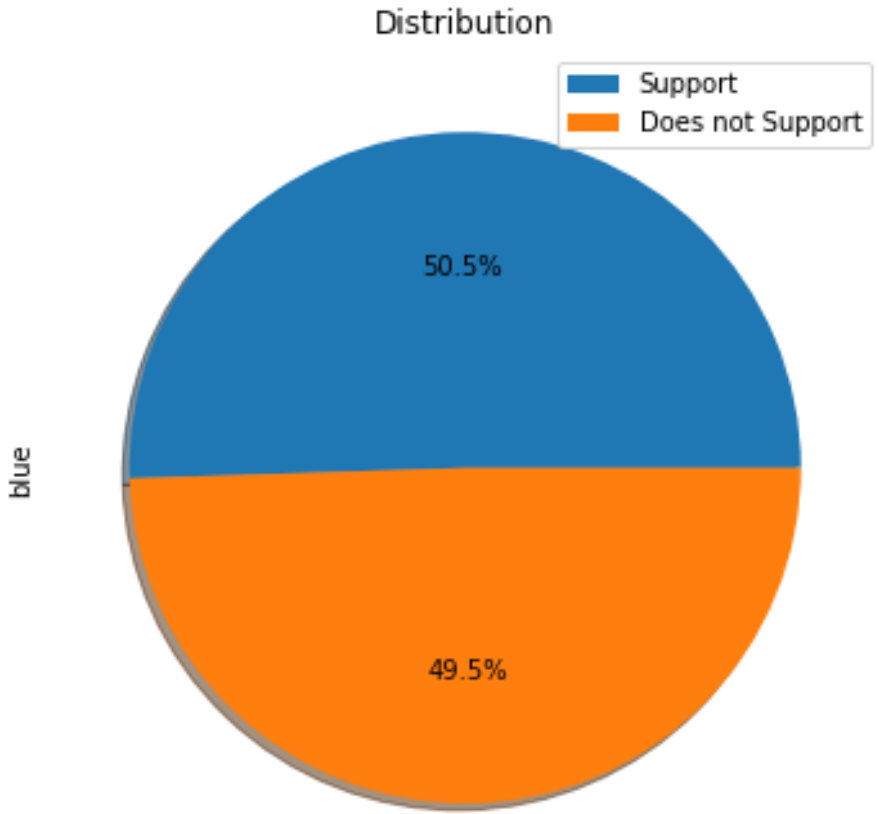
- So, there are mobile phones in 4 price ranges. The number of elements is almost similar.

DATA DISTRIBUTION – BATTERY_POWER



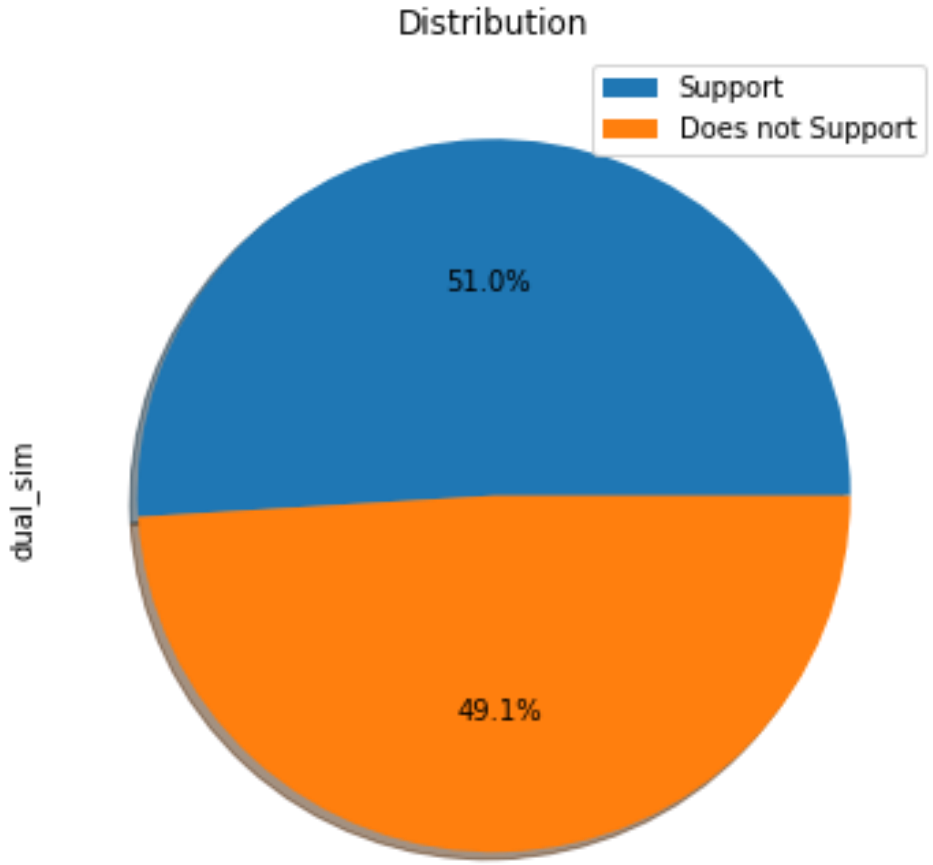
- This plot shows how the battery mAh is spread.

MOBILE PHONE - BLUETOOTH



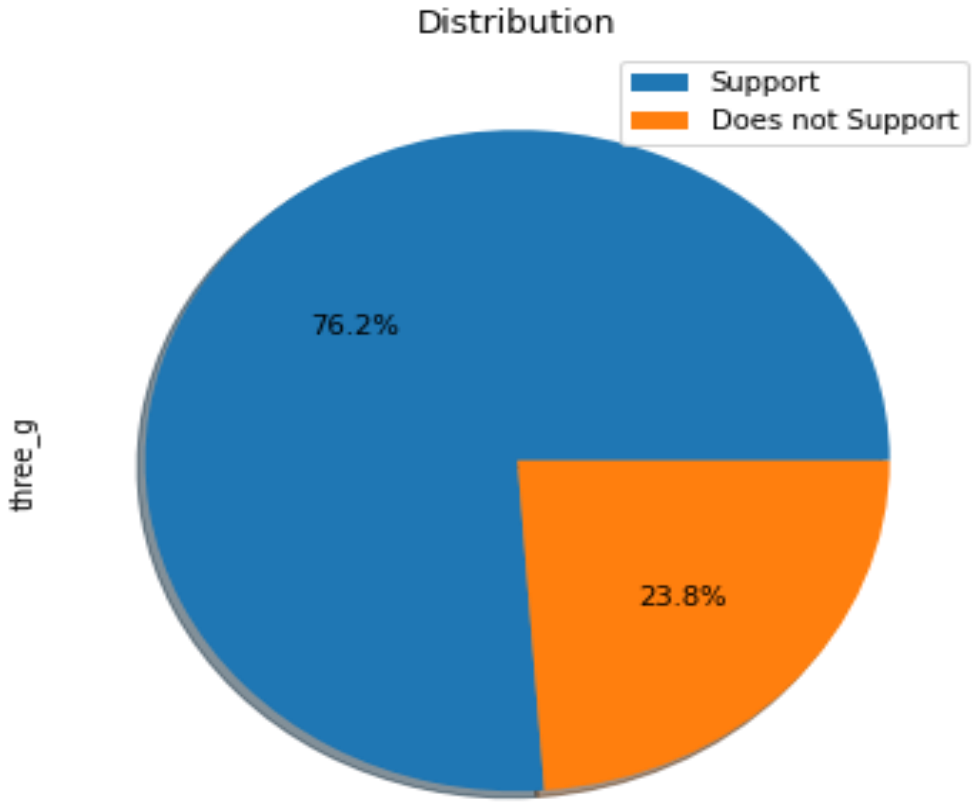
- we see the percentage count of how many devices support Bluetooth and how many does not support.

MOBILE PHONE – DUAL SIM



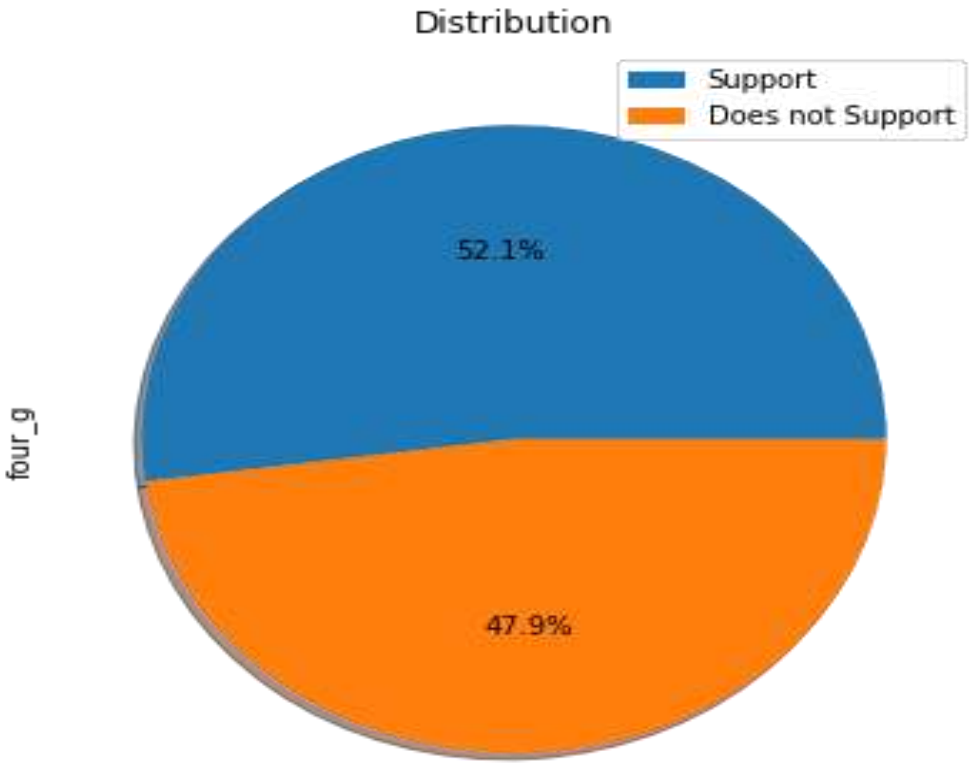
- we see the percentage count of how many devices support Dual sim and how many does not support.

MOBILE PHONE – 3G



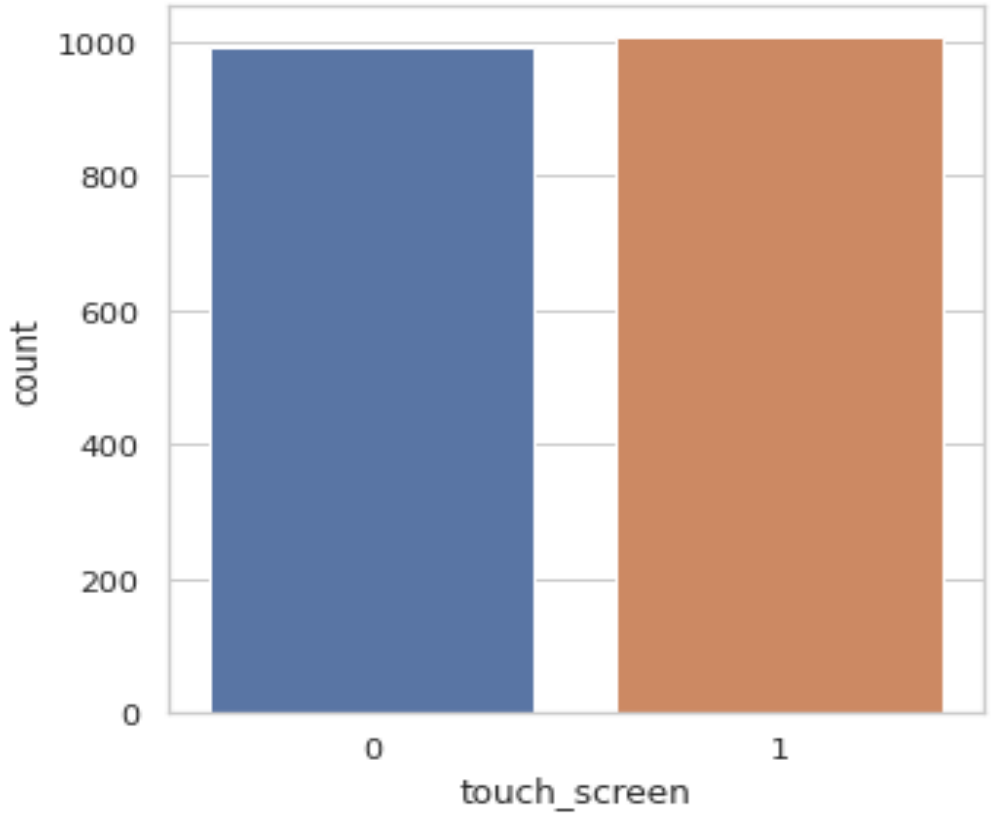
- we see the percentage count of how many devices support 3G and how many does not support.

MOBILE PHONE – 4G



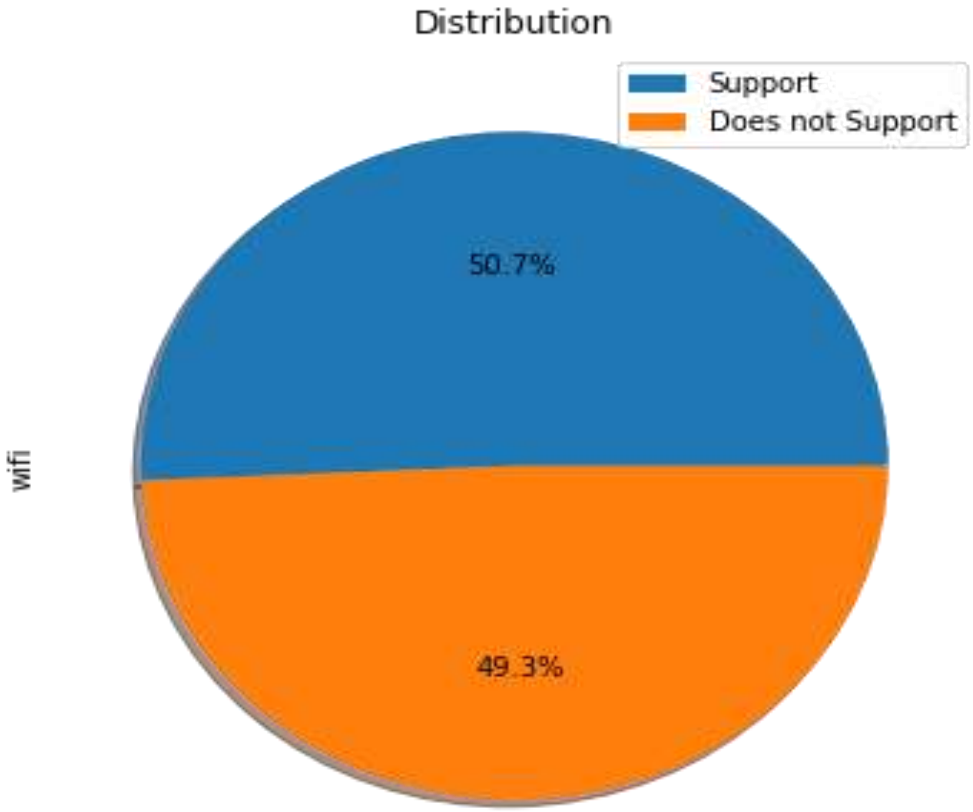
- we see the percentage count of how many devices support 4G and how many does not support.

MOBILE PHONE – TOUCH SCREEN



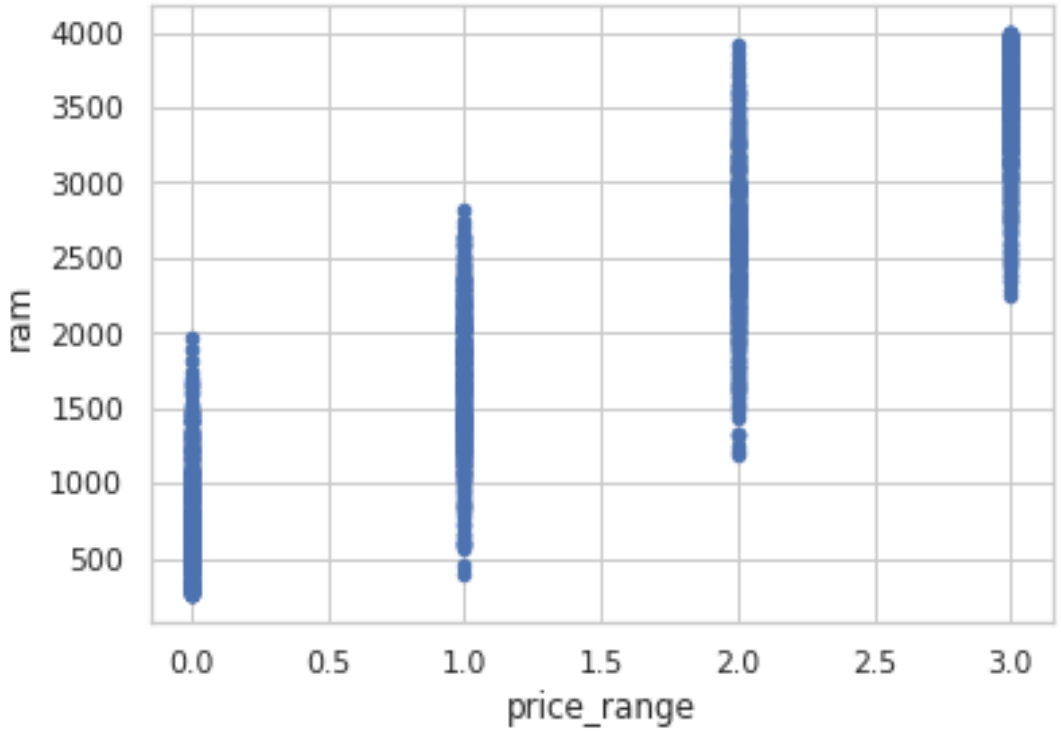
- we see the count of how many devices have Touch screen and how many does not.

MOBILE PHONE – WIFI



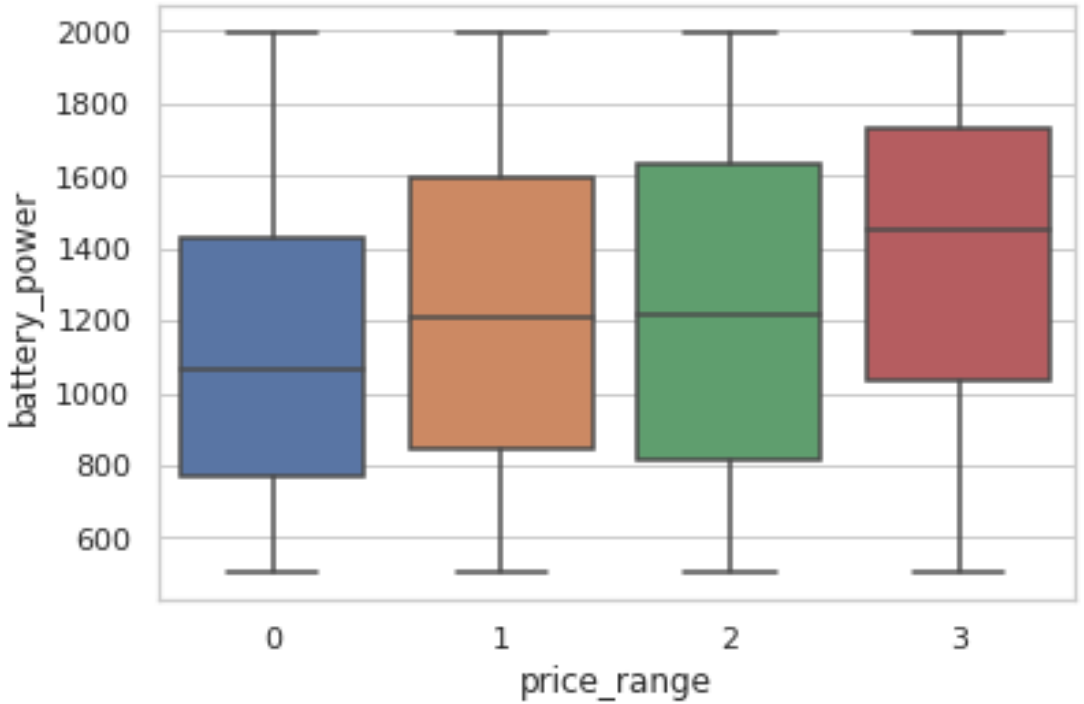
- we see the percentage count of how many devices support Wi-Fi and how many does not support.

MOBILE PHONE – PRICE RANGE VS RAM



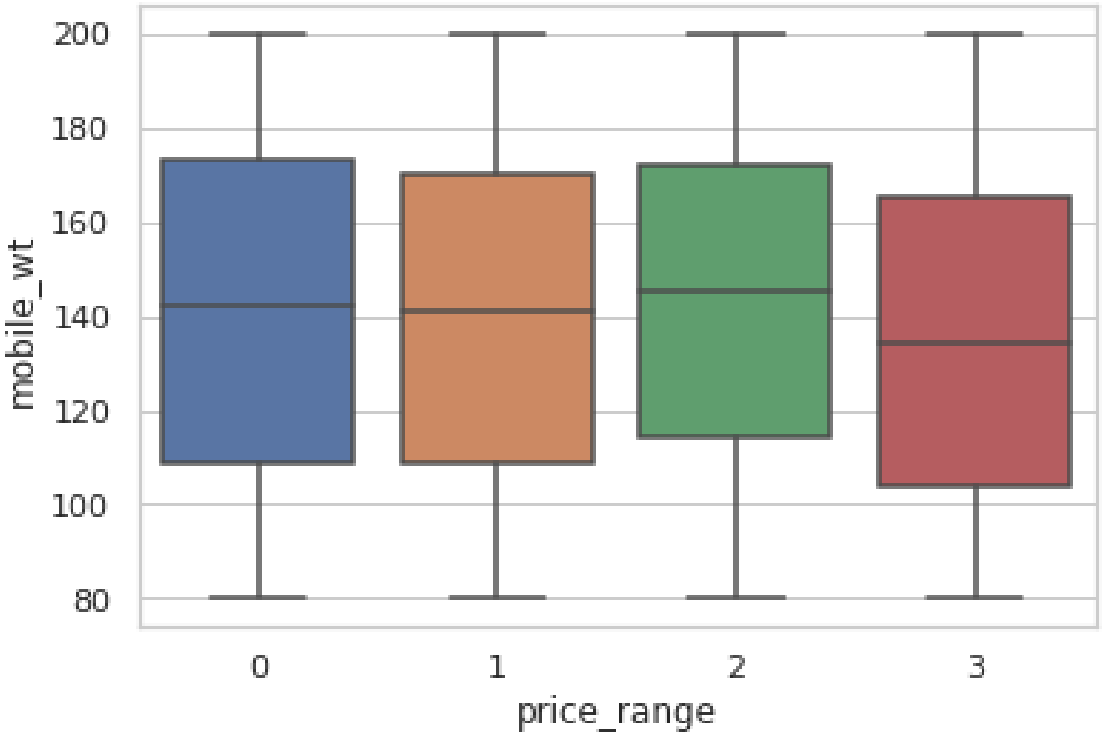
- Ram has continuous increase with price range while moving from Low cost to Very high cost.

MOBILE PHONE – PRICE RANGE VS BATTERY POWER



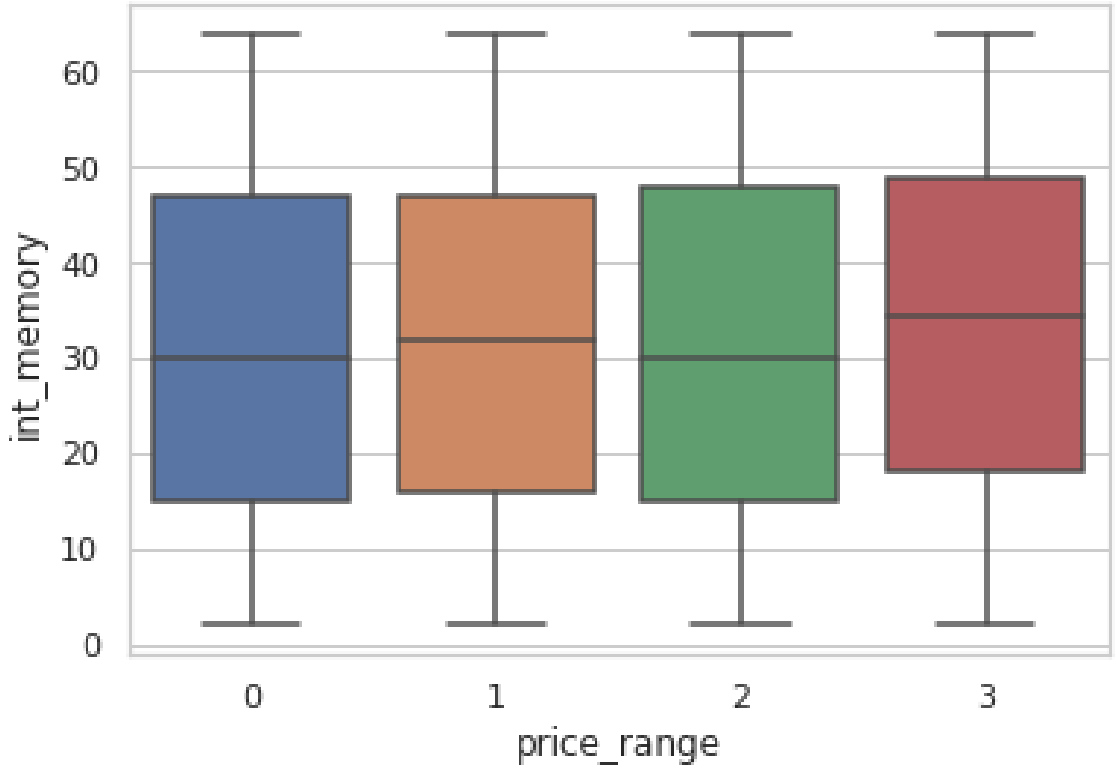
- Battery power has continuous increase with price range while moving from Low cost to Very high cost.

MOBILE PHONE – PRICE RANGE VS MOBILE Wt.



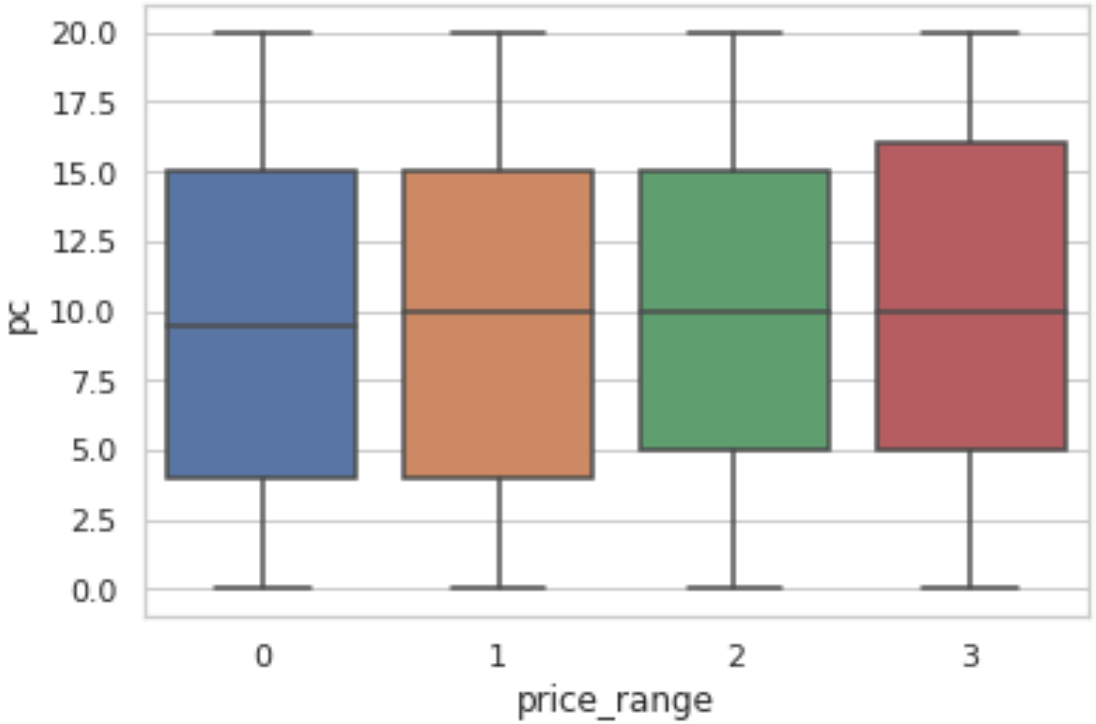
- There is not a continuous increase in Mobile weight as we move from Low cost to Very high cost.
- Mobiles with very High cost has lower mobile weight.

MOBILE PHONE – PRICE RANGE VS INTERNAL MEMORY



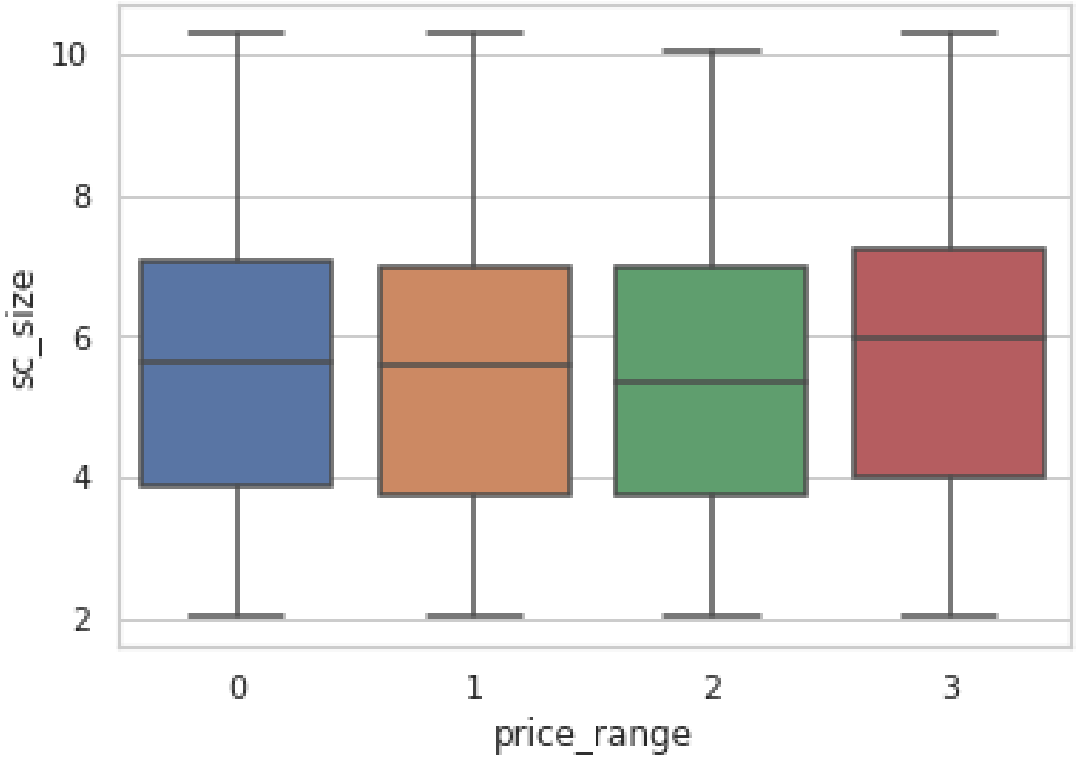
- As we move from Low cost to Very high cost, Mobiles have almost equal Internal memory.

MOBILE PHONE – PRICE RANGE VS PRIMARY CAMERA



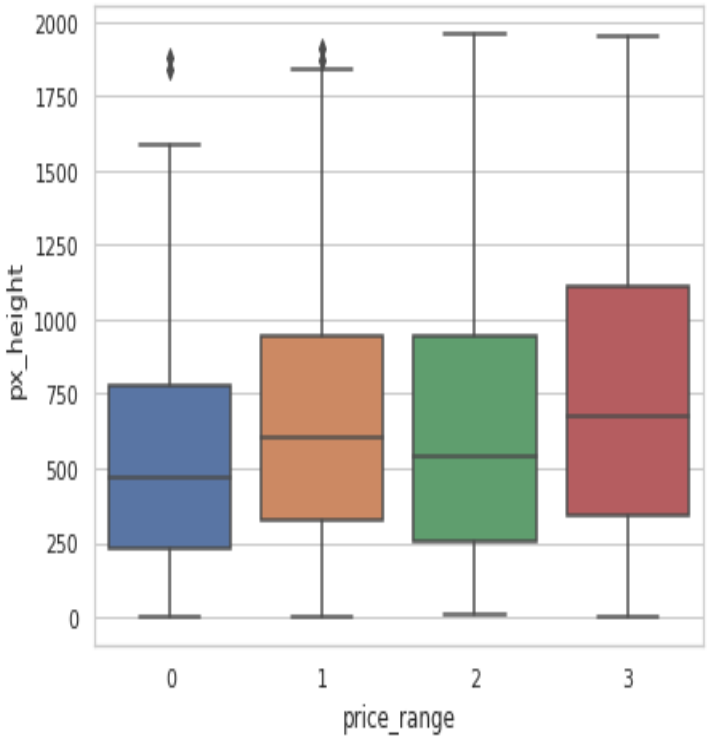
- Primary camera megapixels are showing a little variation along the target categories.

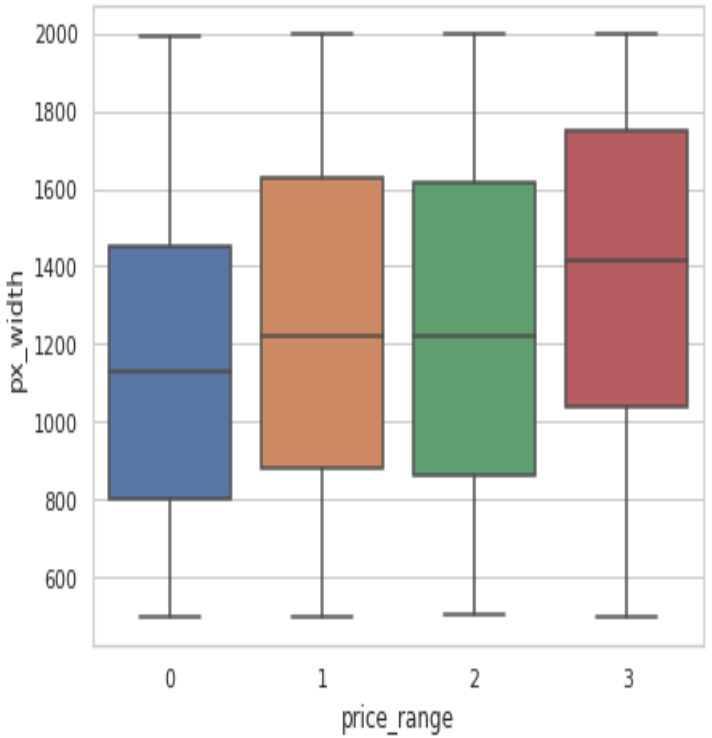
MOBILE PHONE – PRICE RANGE VS SCREEN SIZE



- In real life we use inches to tell a screen size.
- So, We have converted screen size from cm to inches.
- After conversion we can see Screen Size shows little variation along the target variables.

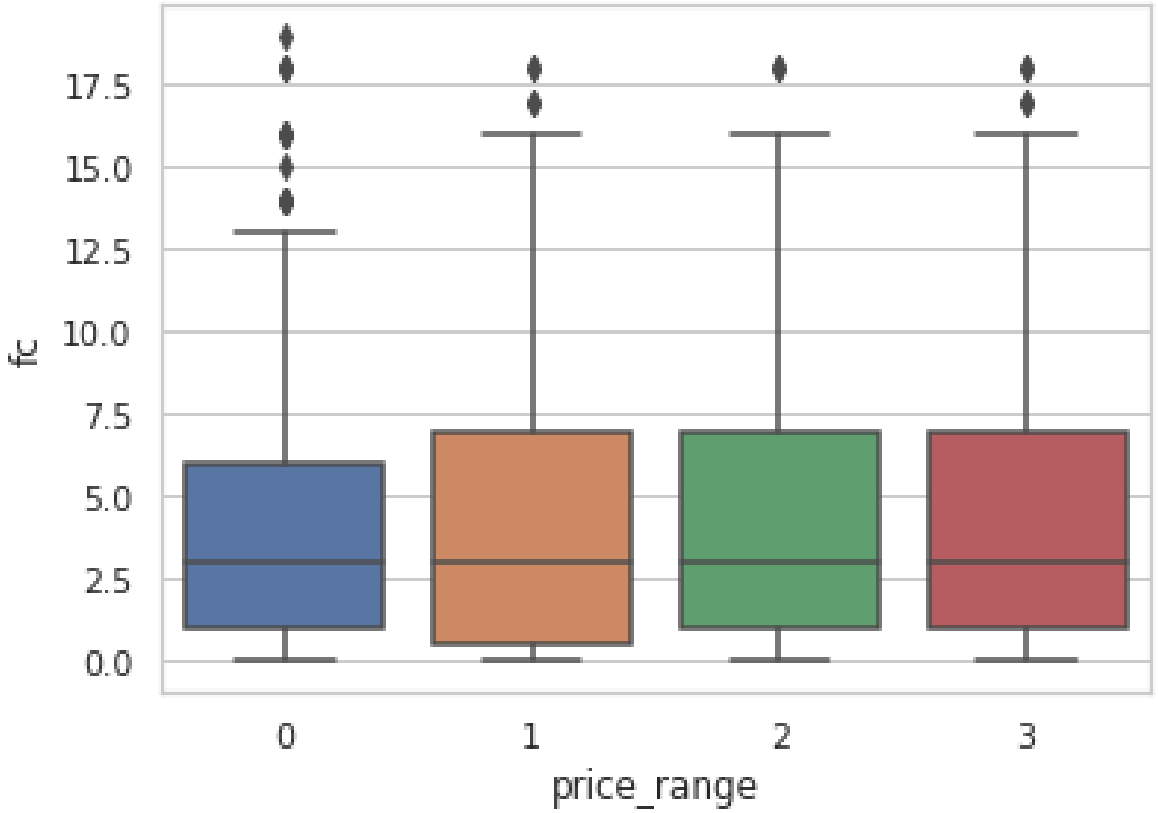
MOBILE PHONE – PRICE RANGE VS CAMERA PIXEL





- Pixel height is almost similar as we move from Low cost to Very high cost, little variation in pixel height.
- There is not a continuous increase in pixel width as we move from Low cost to Very high cost. Mobiles with 'Medium cost' and 'High cost' have almost equal pixel width.
- so we can say that it would be a driving factor in deciding price range.
- As we can see from this box plot there are Outliers!

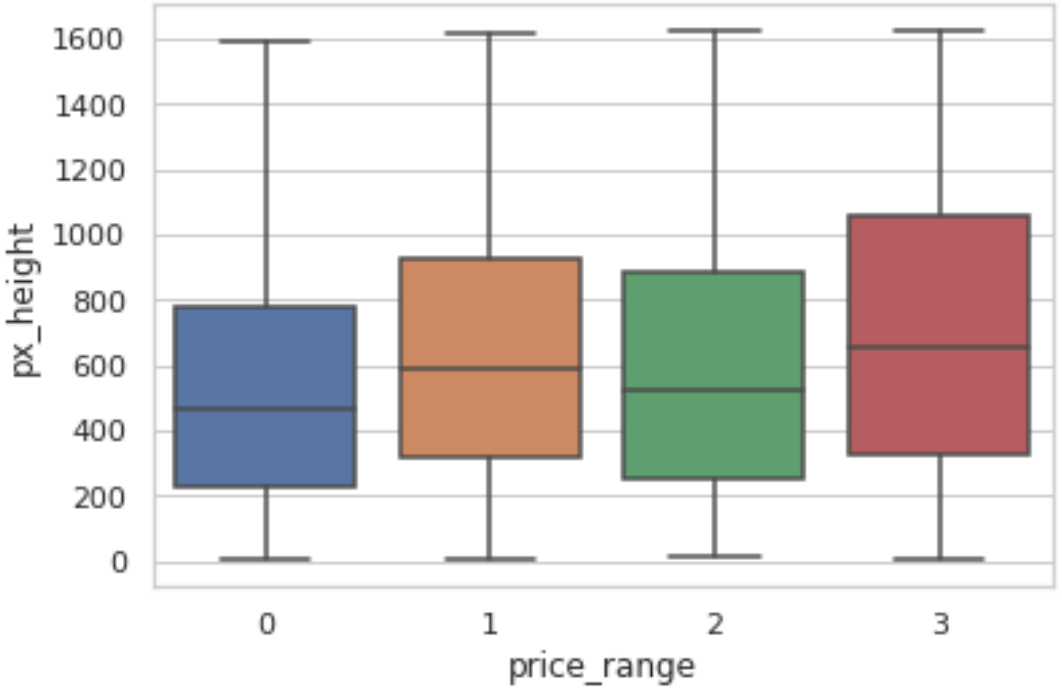
MOBILE PHONE – PRICE RANGE VS FRONT CAMERA



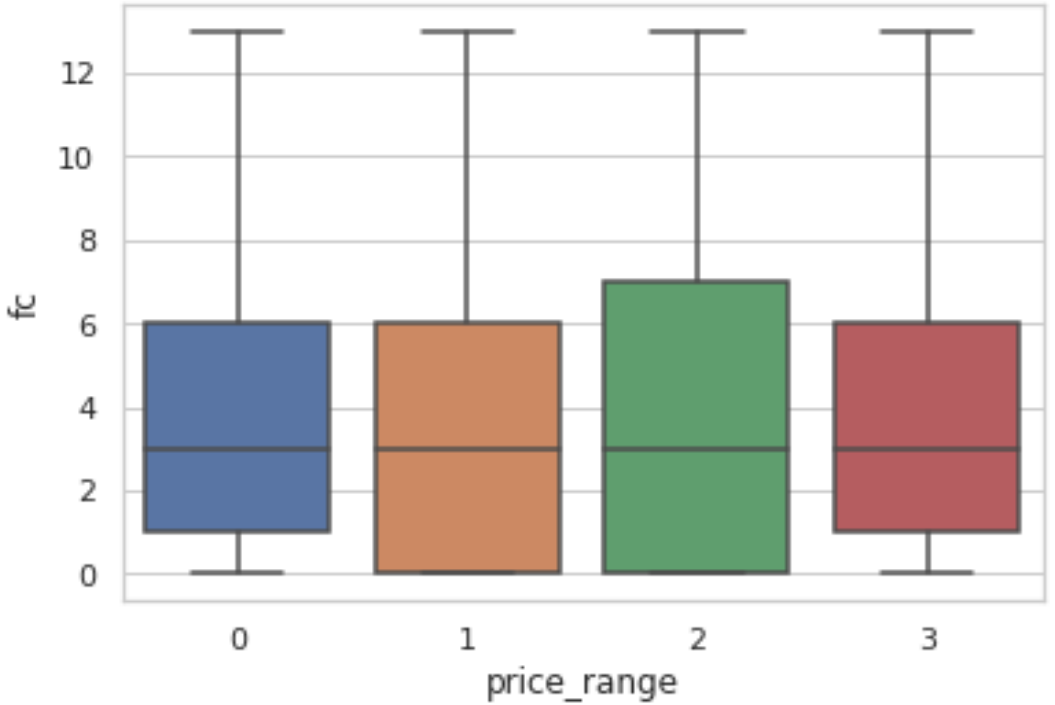
- we can see front camera shows little variation along the target variables.
- As we can see in this box plot there are Outliers! also.

OUTLIERS! TREATMENT

PRICE RANGE VS PIXEL HEIGHT



PRICE RANGE VS FRONT CAMERA



- Outliers is also something that we should be aware of, Because outliers can markedly affect our models and can be a valuable source of information, providing us insights about specific behaviors.

- After removing the outliers, the shape of the dataset is (1865, 19).

ML MODEL

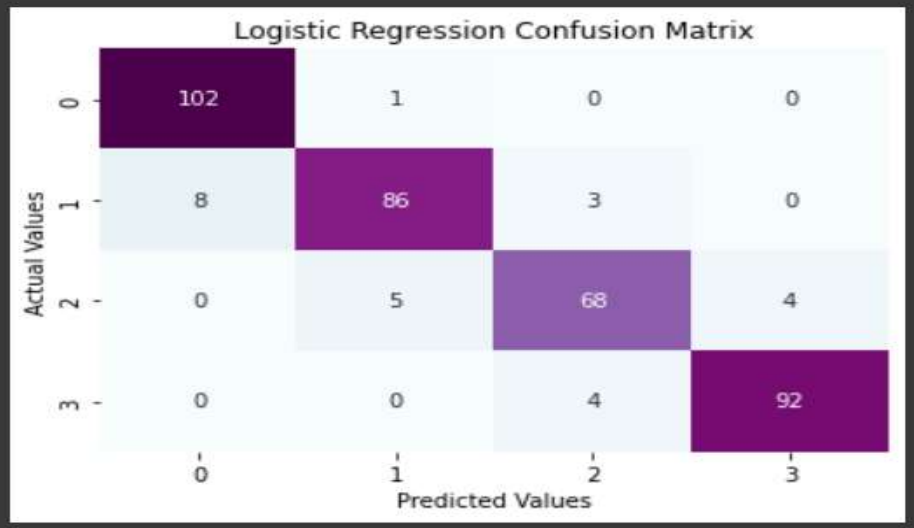
- For Train – Test data split up , we have considered 0.2 as threshold, So that train data will be around 1492 rows to train the model.

- To predict the mobile price range, following Classification models are created.
  - Logistic Regression
  - Gaussian Naive Bayes
  - KNN Classifier
  - SVM Classifier
  - Random Forest

ML – EVALUATION METRICS

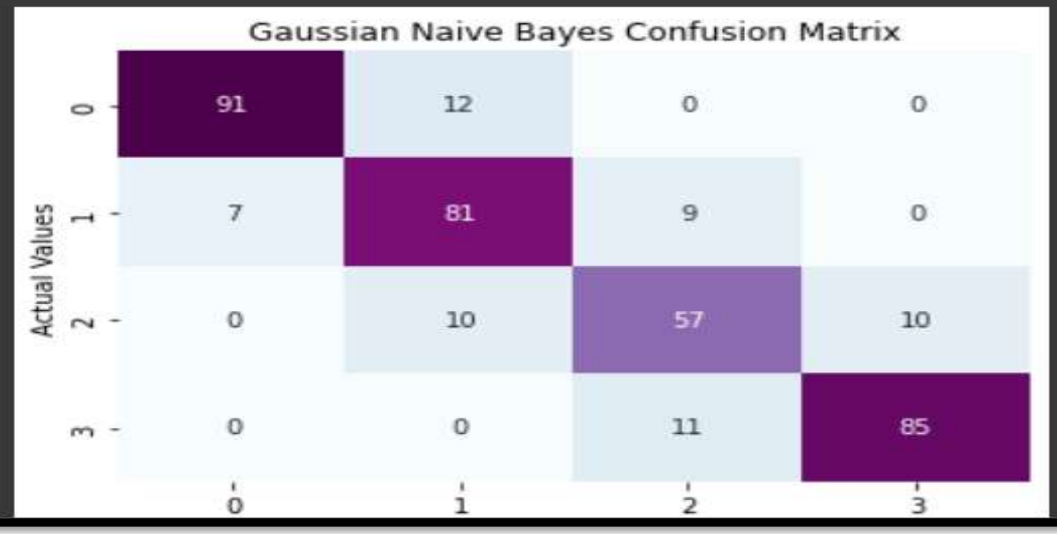
LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.93	0.99	0.96	103
1	0.93	0.89	0.91	97
2	0.91	0.88	0.89	77
3	0.96	0.96	0.96	96
accuracy			0.93	373
macro avg	0.93	0.93	0.93	373
weighted avg	0.93	0.93	0.93	373



GAUSSIAN NAIVE BAYES

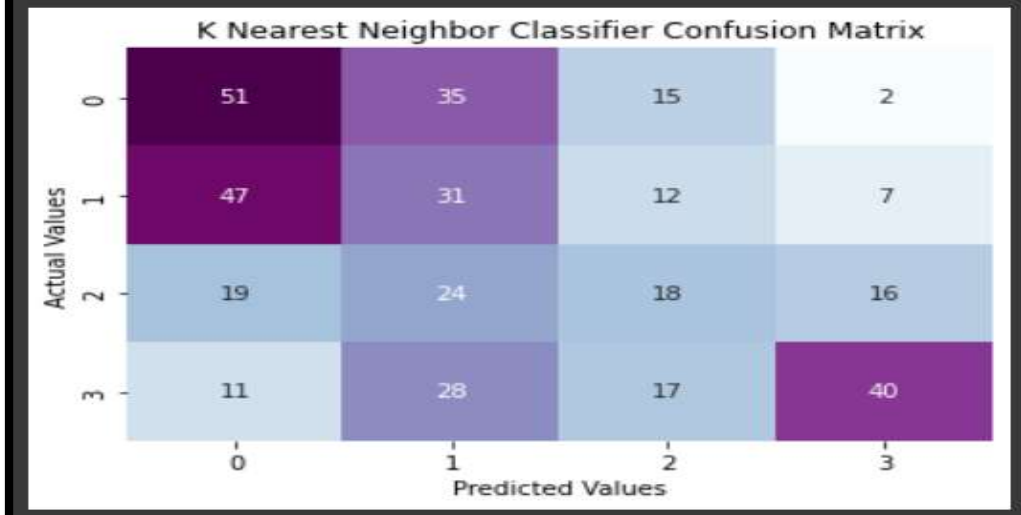
	precision	recall	f1-score	support
0	0.93	0.88	0.91	103
1	0.79	0.84	0.81	97
2	0.74	0.74	0.74	77
3	0.89	0.89	0.89	96
accuracy			0.84	373
macro avg	0.84	0.84	0.84	373
weighted avg	0.84	0.84	0.84	373



ML – EVALUATION METRICS

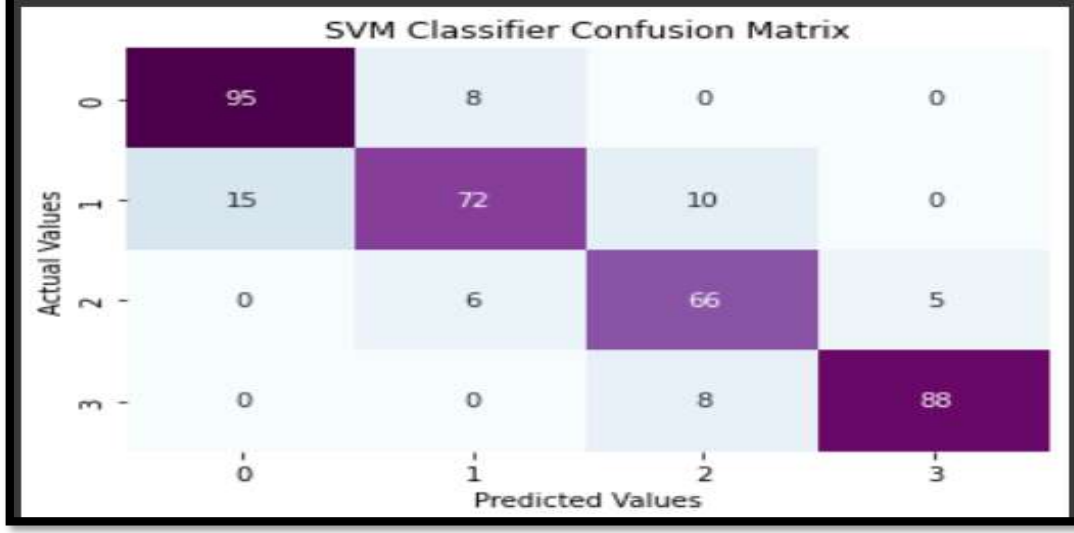
KNN CLASSIFIER

	precision	recall	f1-score	support
0	0.40	0.50	0.44	103
1	0.26	0.32	0.29	97
2	0.29	0.23	0.26	77
3	0.62	0.42	0.50	96
accuracy			0.38	373
macro avg	0.39	0.37	0.37	373
weighted avg	0.40	0.38	0.38	373



SVM CLASSIFIER

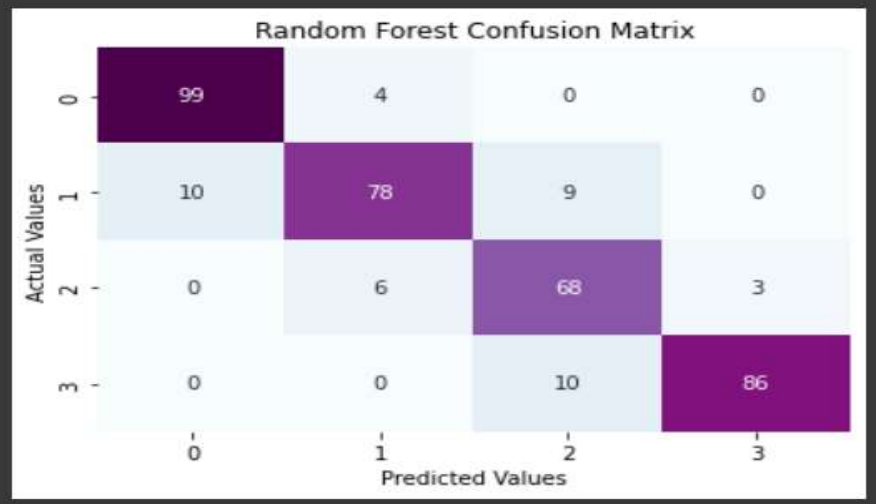
	precision	recall	f1-score	support
0	0.86	0.92	0.89	103
1	0.84	0.74	0.79	97
2	0.79	0.86	0.82	77
3	0.95	0.92	0.93	96
accuracy			0.86	373
macro avg	0.86	0.86	0.86	373
weighted avg	0.86	0.86	0.86	373



ML – EVALUATION METRICS

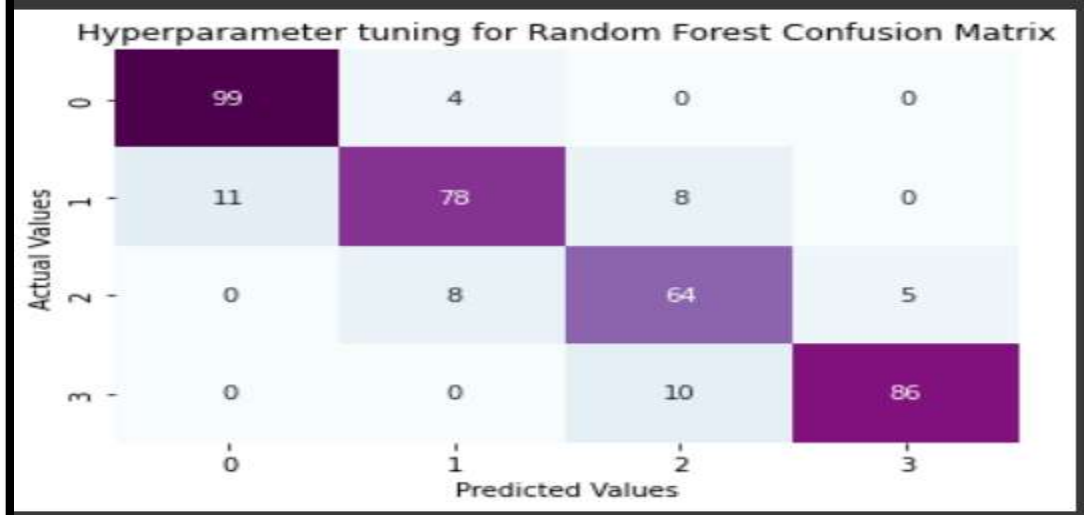
RANDOM FOREST

	precision	recall	f1-score	support
0	0.91	0.96	0.93	103
1	0.89	0.80	0.84	97
2	0.78	0.88	0.83	77
3	0.97	0.90	0.93	96
accuracy			0.89	373
macro avg	0.89	0.89	0.88	373
weighted avg	0.89	0.89	0.89	373



HYPERPARAMETER TUNING FOR RANDOM FOREST

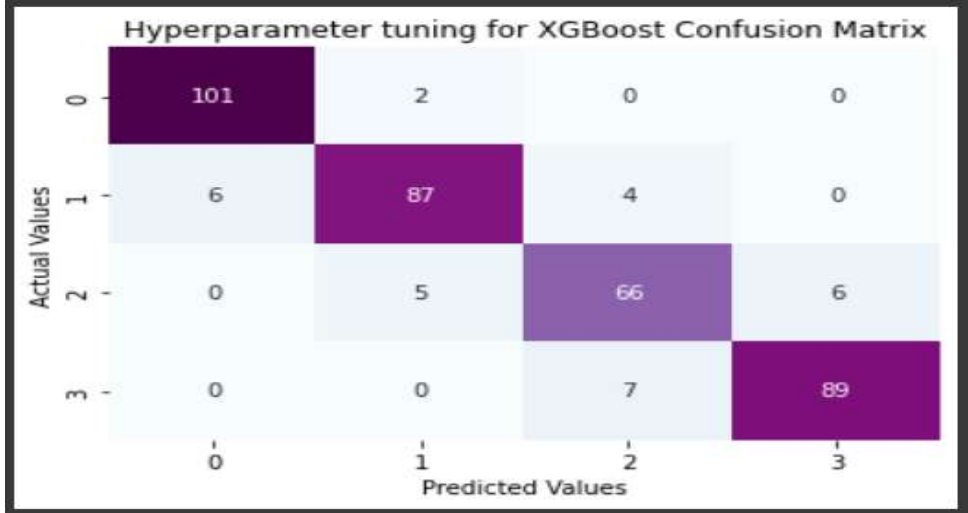
	precision	recall	f1-score	support
0	0.90	0.96	0.93	103
1	0.87	0.80	0.83	97
2	0.78	0.83	0.81	77
3	0.95	0.90	0.92	96
accuracy			0.88	373
macro avg	0.87	0.87	0.87	373
weighted avg	0.88	0.88	0.88	373



ML – EVALUATION METRICS

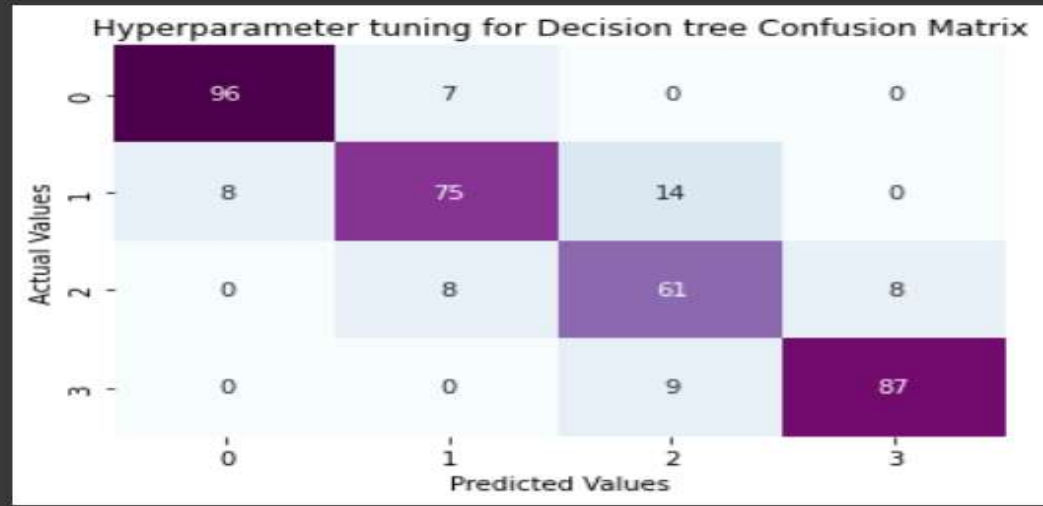
HYPERPARAMETER TUNING FOR XGBOOST

	precision	recall	f1-score	support
0	0.94	0.98	0.96	103
1	0.93	0.90	0.91	97
2	0.86	0.86	0.86	77
3	0.94	0.93	0.93	96
accuracy			0.92	373
macro avg	0.92	0.92	0.92	373
weighted avg	0.92	0.92	0.92	373

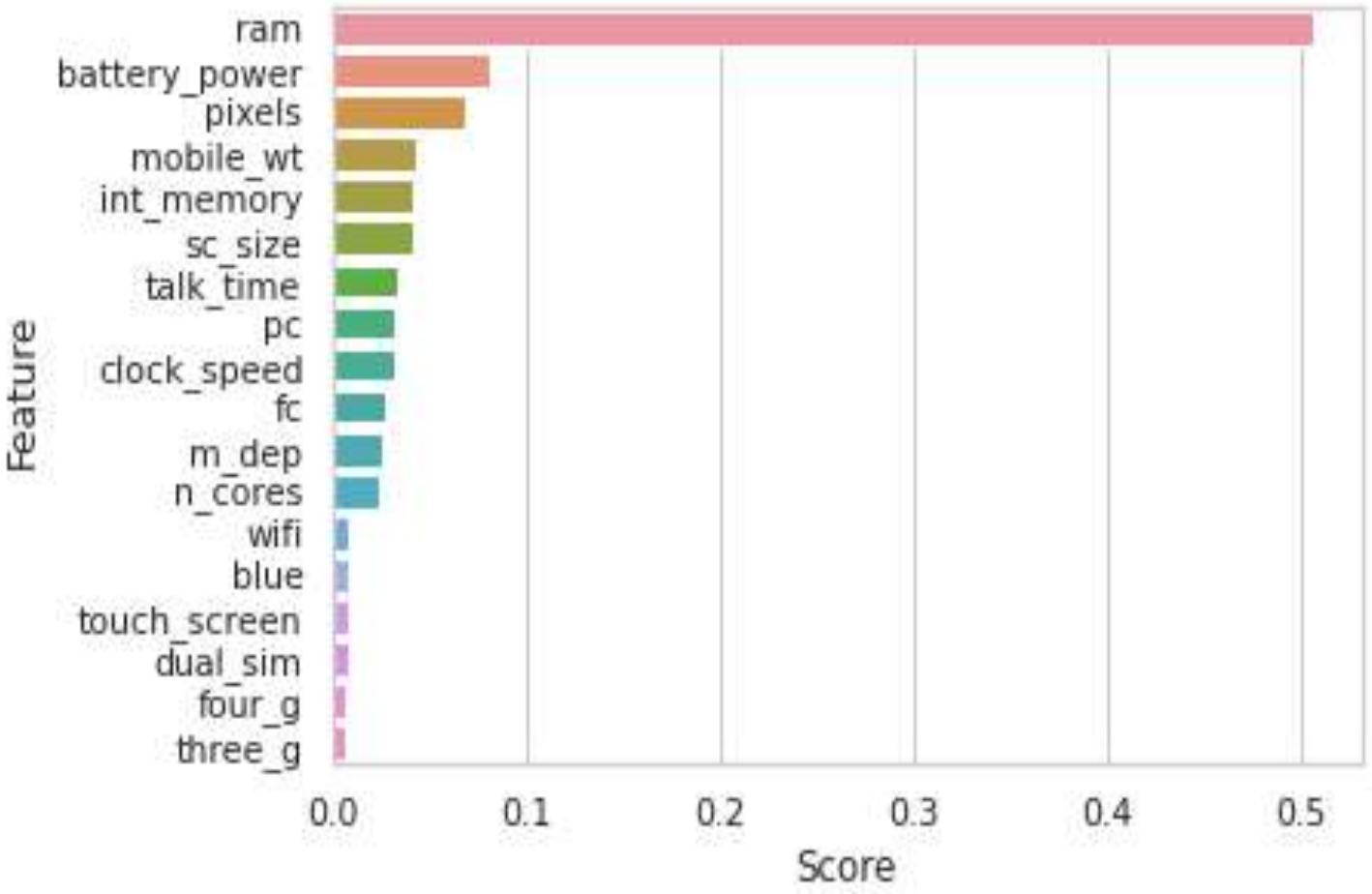


HYPERPARAMETER TUNING FOR DECSION TREE

	precision	recall	f1-score	support
0	0.92	0.93	0.93	103
1	0.83	0.77	0.80	97
2	0.73	0.79	0.76	77
3	0.92	0.91	0.91	96
accuracy			0.86	373
macro avg	0.85	0.85	0.85	373
weighted avg	0.86	0.86	0.86	373



FEATURE IMPORTANCE



- The price of the mobile majorly depends on the ram, followed by battery power and camera pixel size.

CONCLUSION

From all the above experiments we can conclude that logistic regression and XG Boosting with using hyperparameters we got the best results.

This would help organizations and consumers alike to make more educated decisions when it comes to price.