



Capstone project

SUPERVISED ML- CLASSIFICATION

Topic- Mobile price range prediction

**By – Vivek Kumar Soni &
Ashish pandey**

Contents of the Presentation

- Problem Statement
- Data Summary
- Exploratory data analysis
- Data correlation
- Data wrangling
- Machine Learning models
- Model Explanation
- Challenges
- Conclusion

Problem Statement

AI

- Mobile phones have become the greatest necessity for all individuals now days . People want more features and best specification in a phone and that too at cheaper price. The demand for phones is so high that the there is a huge competition prevelling between mobile manufactures. To stay ahead in the race these companies try to bring in a new features and innovations so that people are lured buying their brand smartphones.
- Price of a mobile phone is influenced by various factors . Brand name, newness of the model, specification such as internal memory, camera, ram ,sizes , connections etc. are some of the important factors in determining the prices. As a business point of view it become an utmost priority to analyse these factors from time to time amd come up with best set of specification and price ranges so that people buy there mobile phones.

-

Data summary

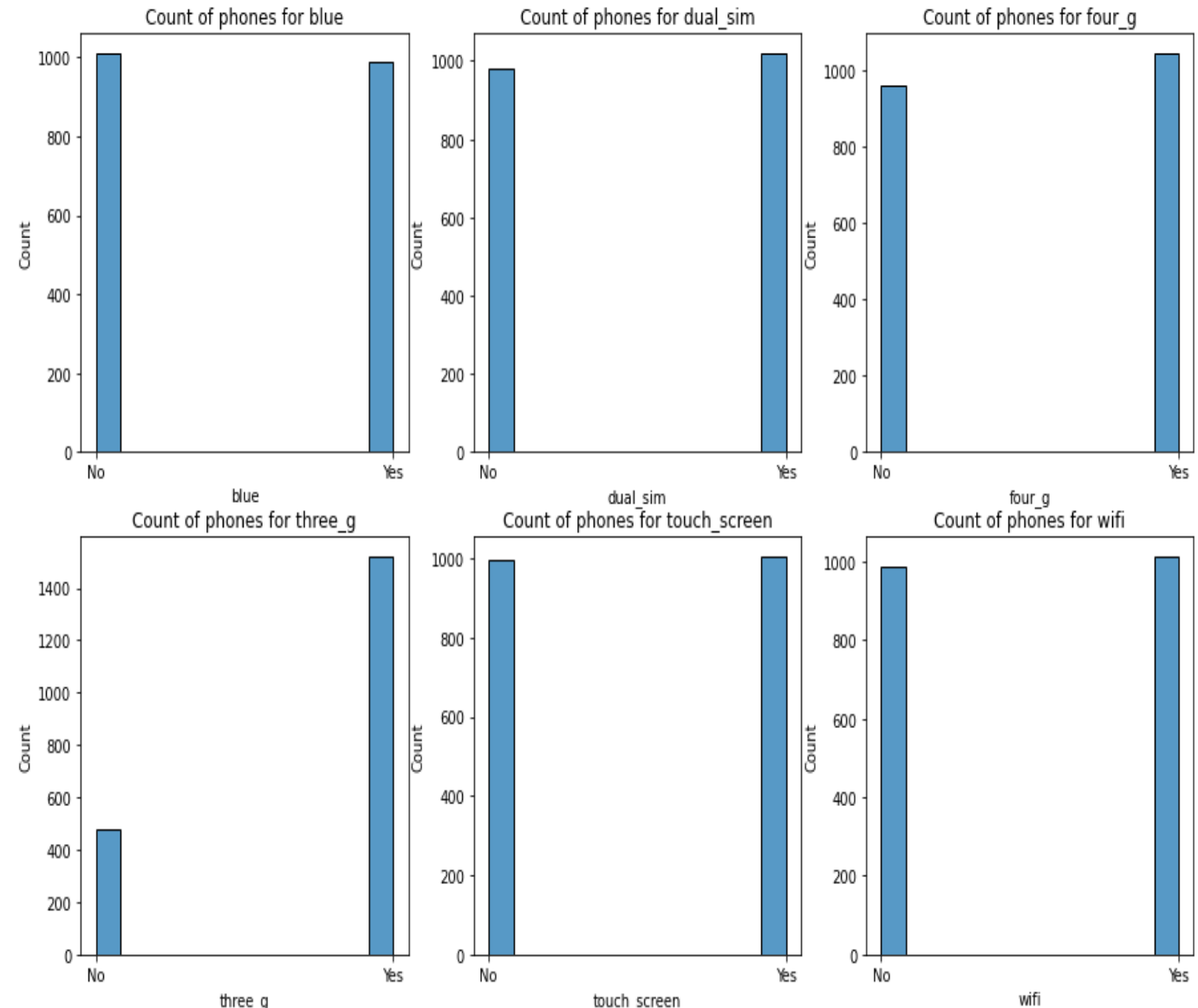
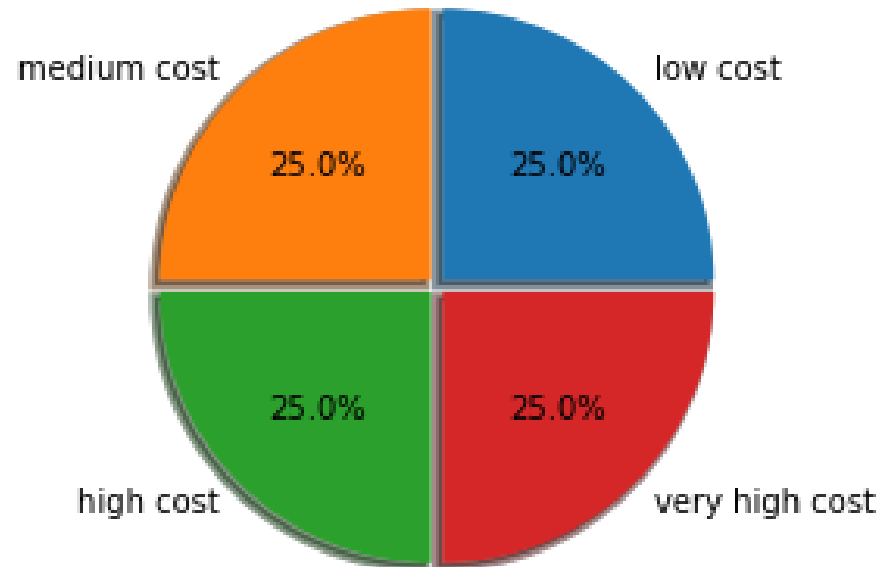
A red square logo with the white letters 'AI' inside.

The contents of the data had these features:

- Battery power – total energy a battery can store in one time measured in mAh
- Clock_speed- speed at which microprocessor executes instructions Fc , Pc Front and Primary camera megapixels.
- Int memory – internal memory in gigabyte
- M_dep – mobile depth in cm
- Mobile_wt – weight of mobile phone
- N_cores – number of cores of processor
- Px_height,px_width – pixel resolution height and width
- Ram – random access memory
- Talk time longest time that a single batterycharge to will last when you are on call
- Blue , 4g, 3g, dual_sim, touchscreen, wifi, - some supportef and unsupported categories
- Price_range – this is the largest variable with value of 0,1,2

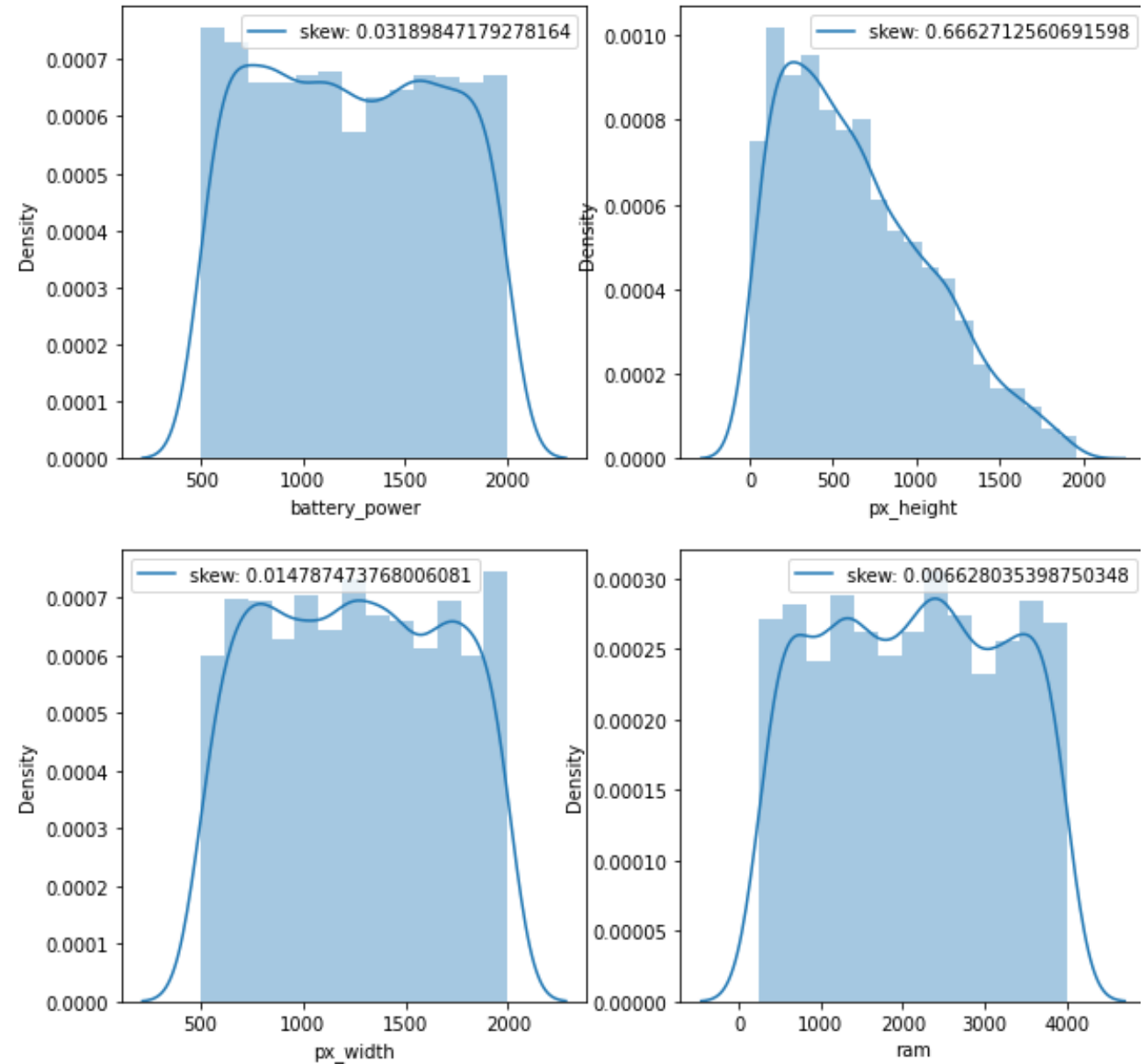
EDA – Univariate Analysis

- Our dependent variable – price is equality distributed amongst variable.
- Other dichotomous features have equal no of observation for each category, expect 3g.



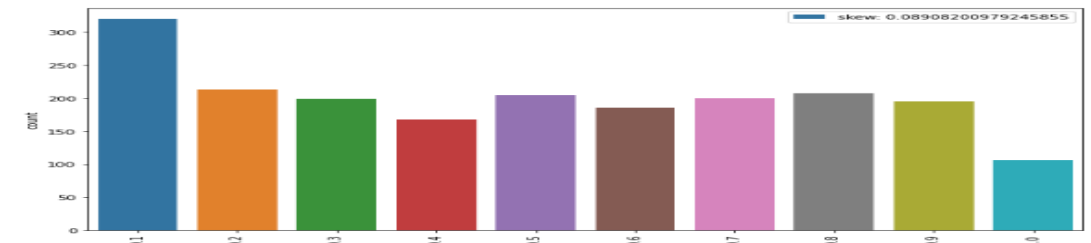
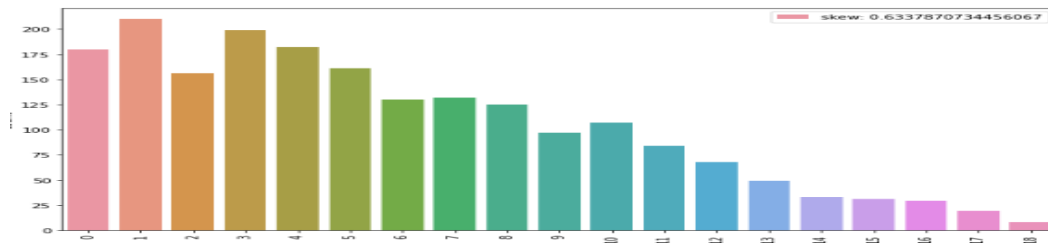
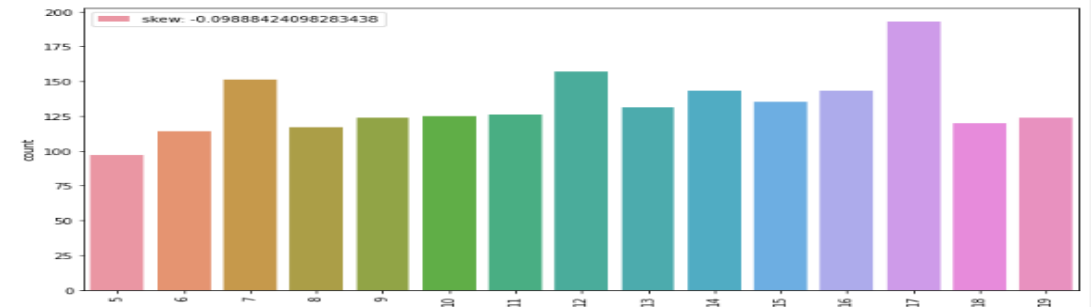
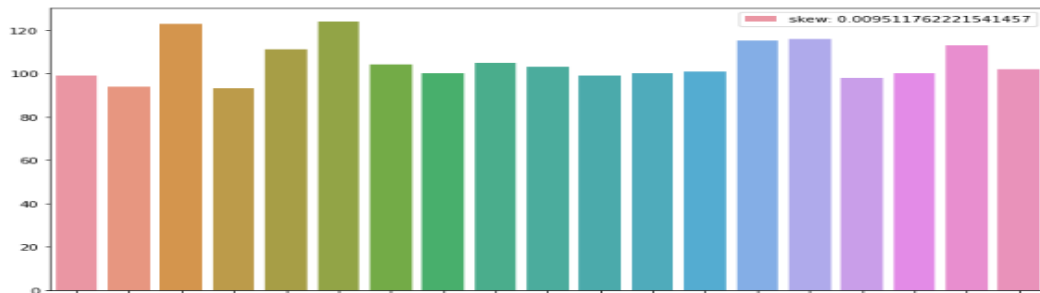
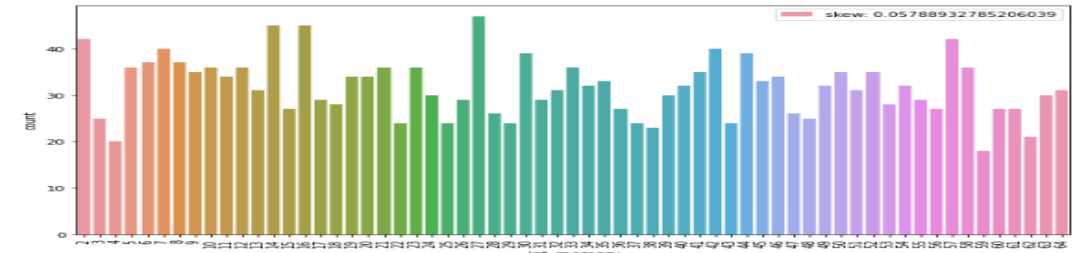
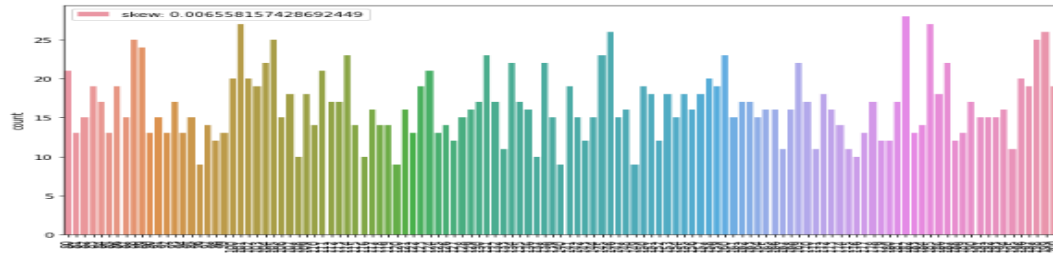
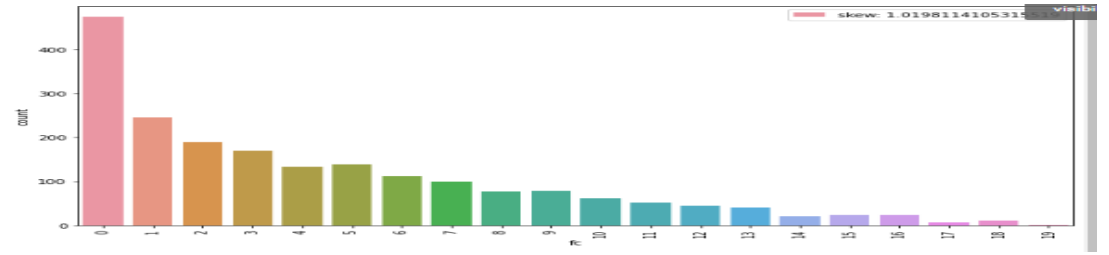
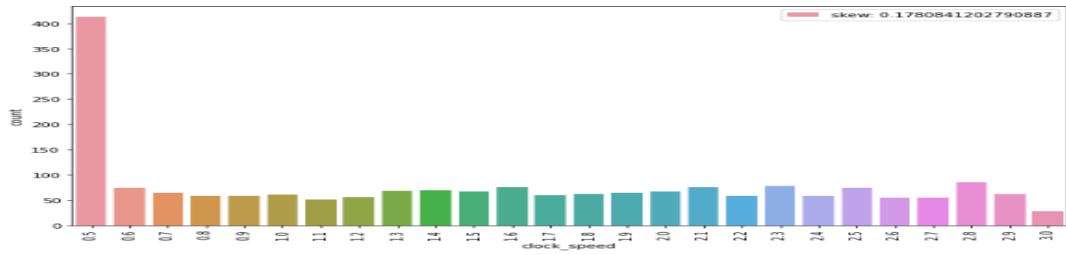
EDA – Univariate Analysis

- The count of phones is equally distributed among all the parameters in every variable except pixel height.
- The number phones tends to decrease as we go higher in pixel height i.e we can say that the graph is positively skewed.



Univariate Analysis – Numerical Variables

AI



Univariate Analysis – Numerical Variables

The logo consists of the letters 'AI' in white, bold, sans-serif font, centered within a solid red square.

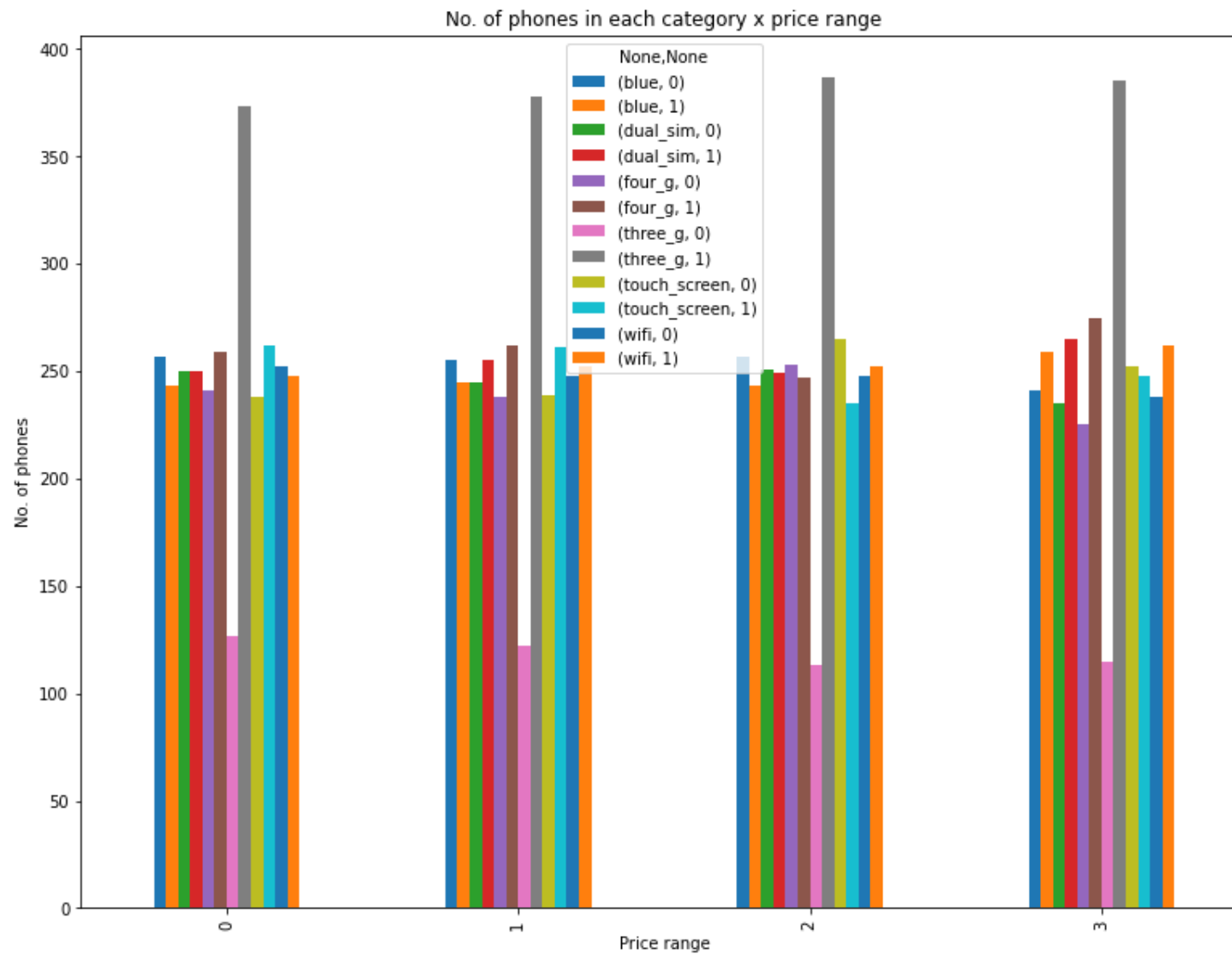
- The clock speed with 0.5GHz stands out in popularity among other phones. Rest of them have even distributed over the sales.
- The phone with higher megapixels camera doesn't show a great relevance amongst the customers.
- All the rest of features has shown a symmetric skewness in their sales numbers , expect for screen width.

Multivariate Analysis

Categorical variables

Almost equal no. of observation for each price range

For each category.

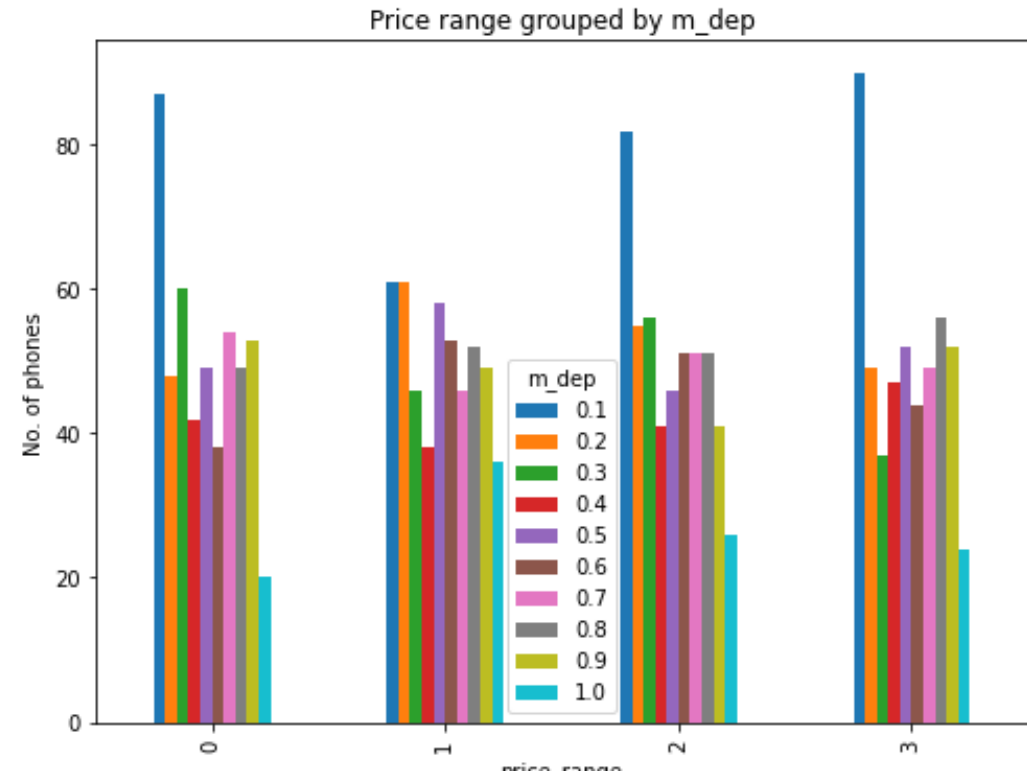
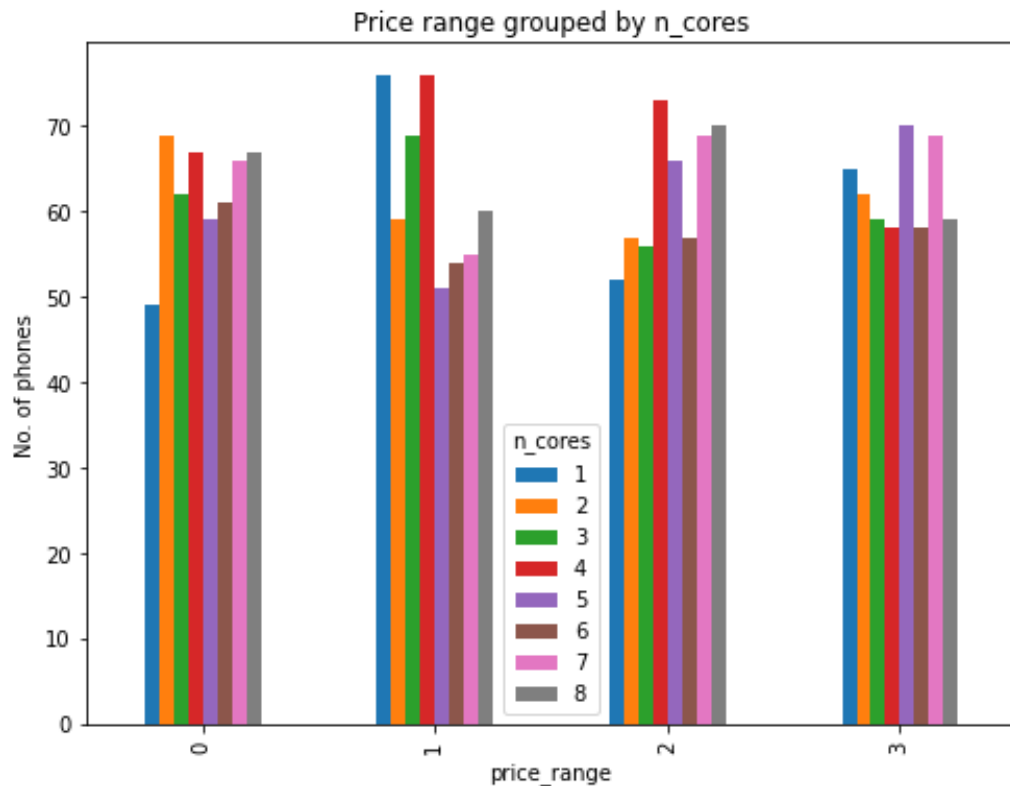


Multivariate analysis

AI

n_core and m_dep

- Count of less n_cores is high for 0 and 1 price range.
- Count of high n_cores is high for 2 and 3 price range.
- Count of phones with less thickness is high and count of phones with high thickness is low.

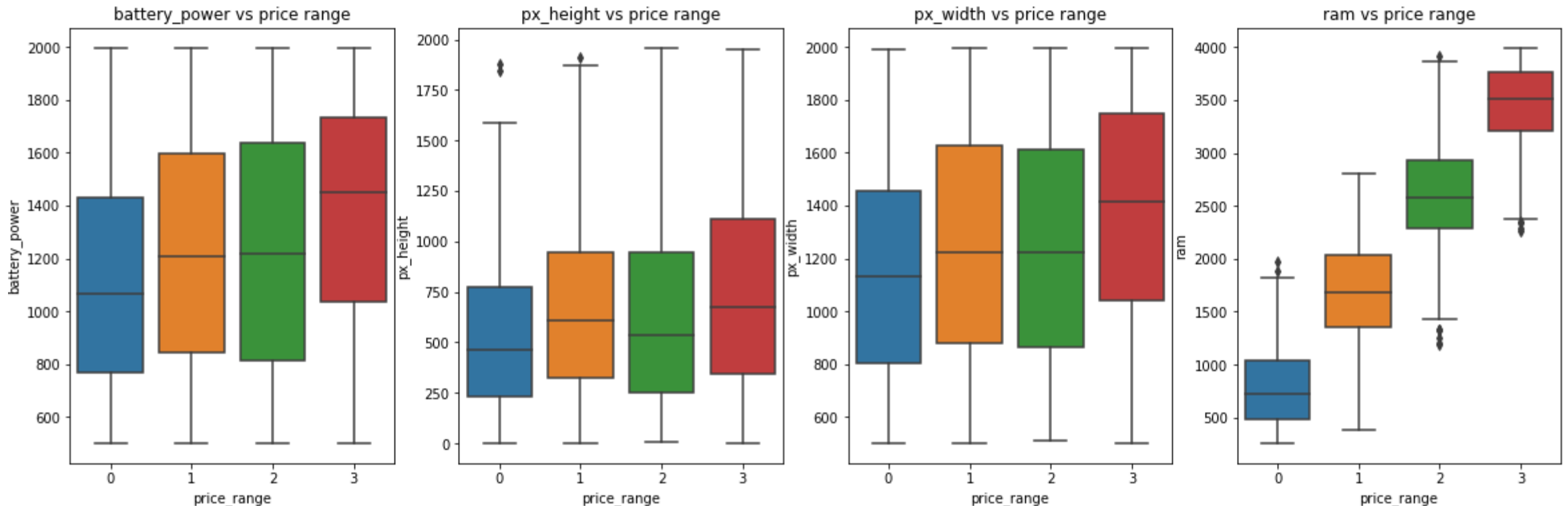


Multivariate analysis

AI

Numerical variables

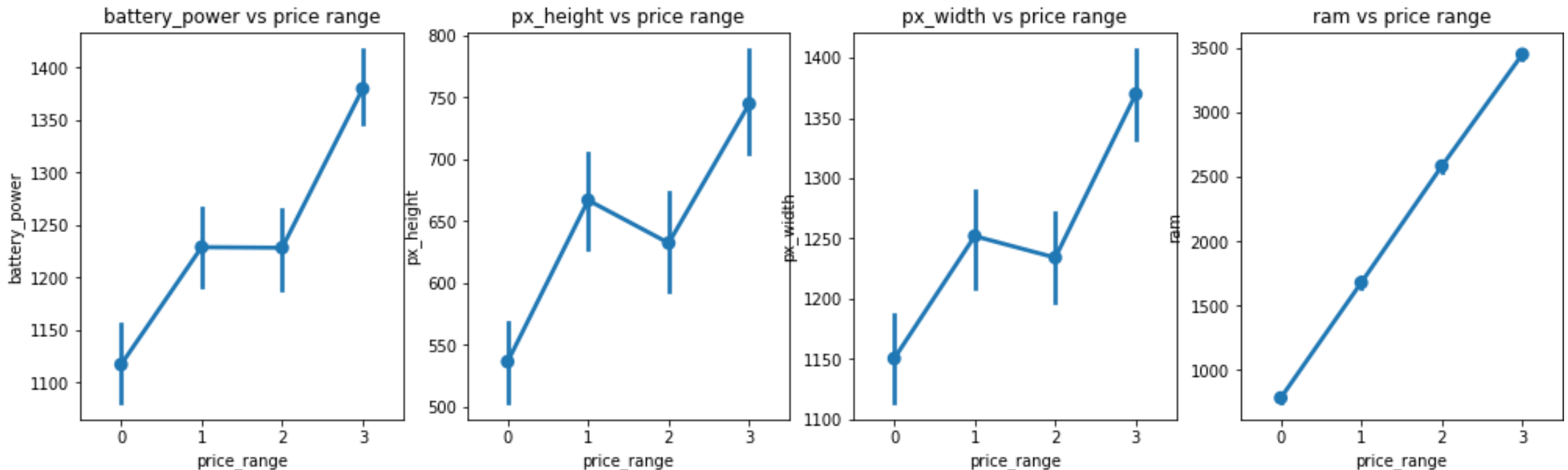
- Mean values of battery power , px_height , px_width, ram is increasing with increase in price.



Multivariate analysis- Continued

AI

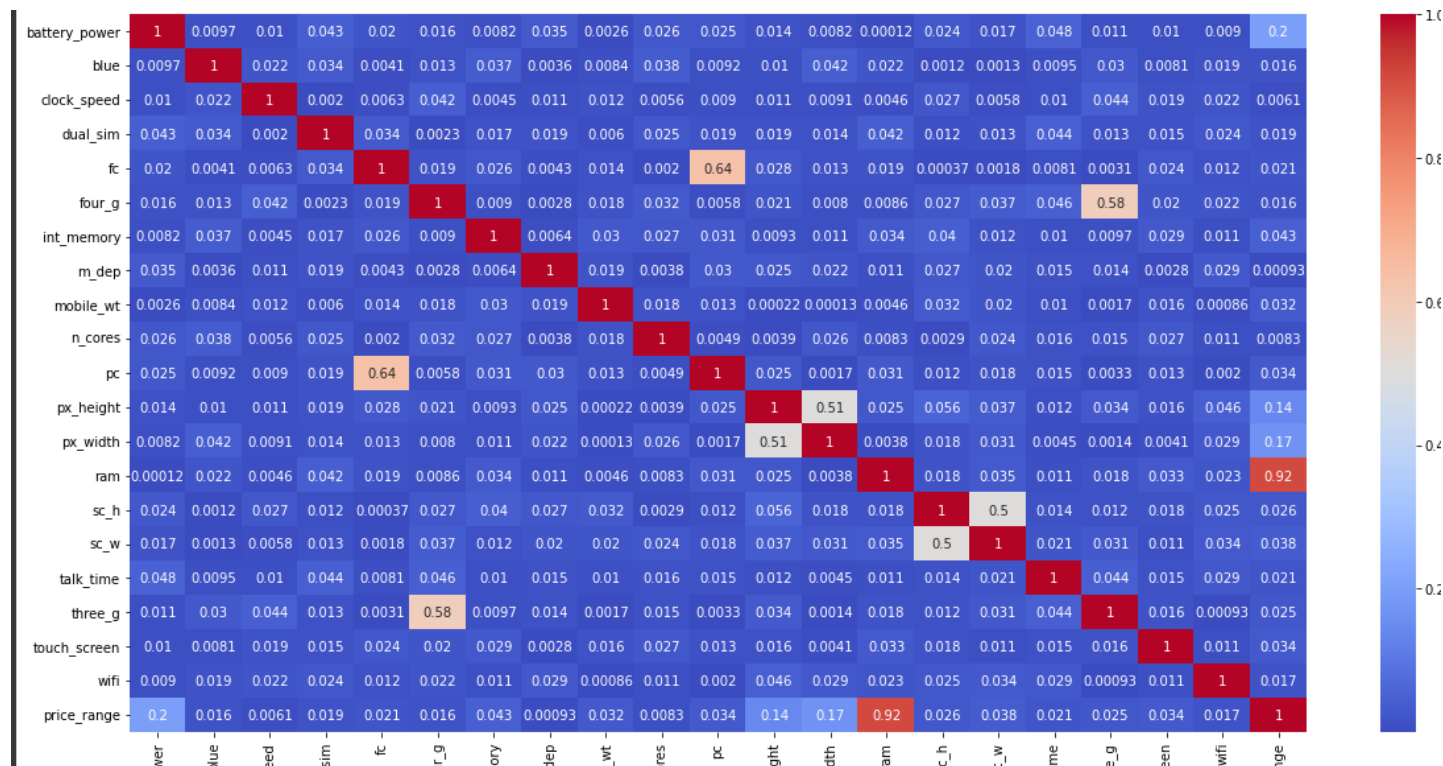
It appears that the values of these variables are rising in lockstep with the rise in prices. Ram is having RAM you have, the more money you'll spend



Data correlation

AI

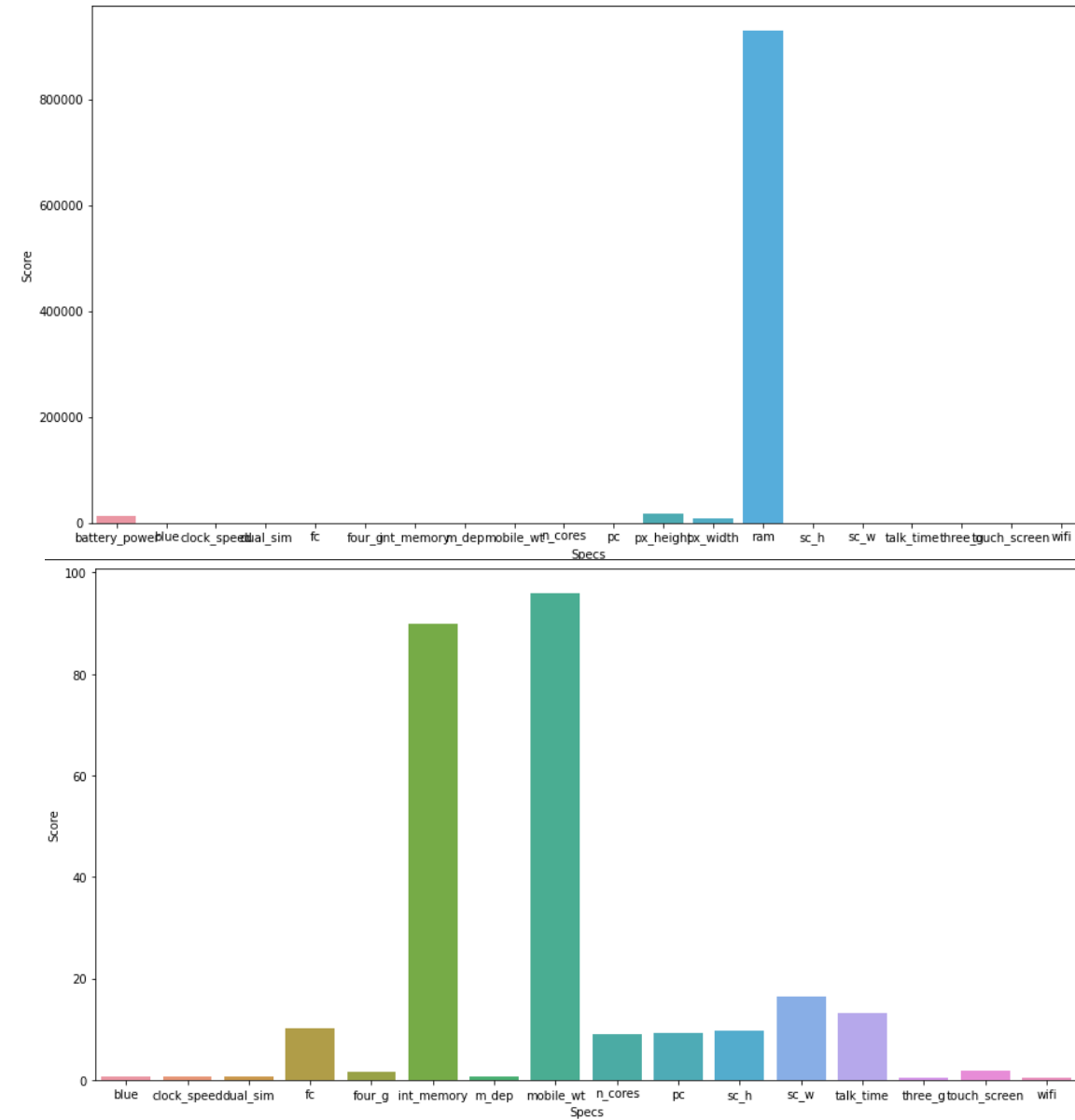
- Pc is correlated with fc.
- 39 and 4g are moderately correlated.
- Sc_h and sc_w are moderately correlated. We will try to change them into a single variable.
- Px width and px_height are moderately correlated. We will try to change them into a single variable
- Ram is highly correlated with our price range. My be one the most important factor in determining the price.



Feature Selection

The most important features ,as per both graphs are:

- Ram
- Battery power
- Pixel height pixel width
- Mobile weight internal memory
- Front camera megapixel number of cores
- Primary camera megapixels
- Screen height
- Screen width
- Talk time



Machine learning models

A red square containing the white letters 'AI' in a bold, sans-serif font.

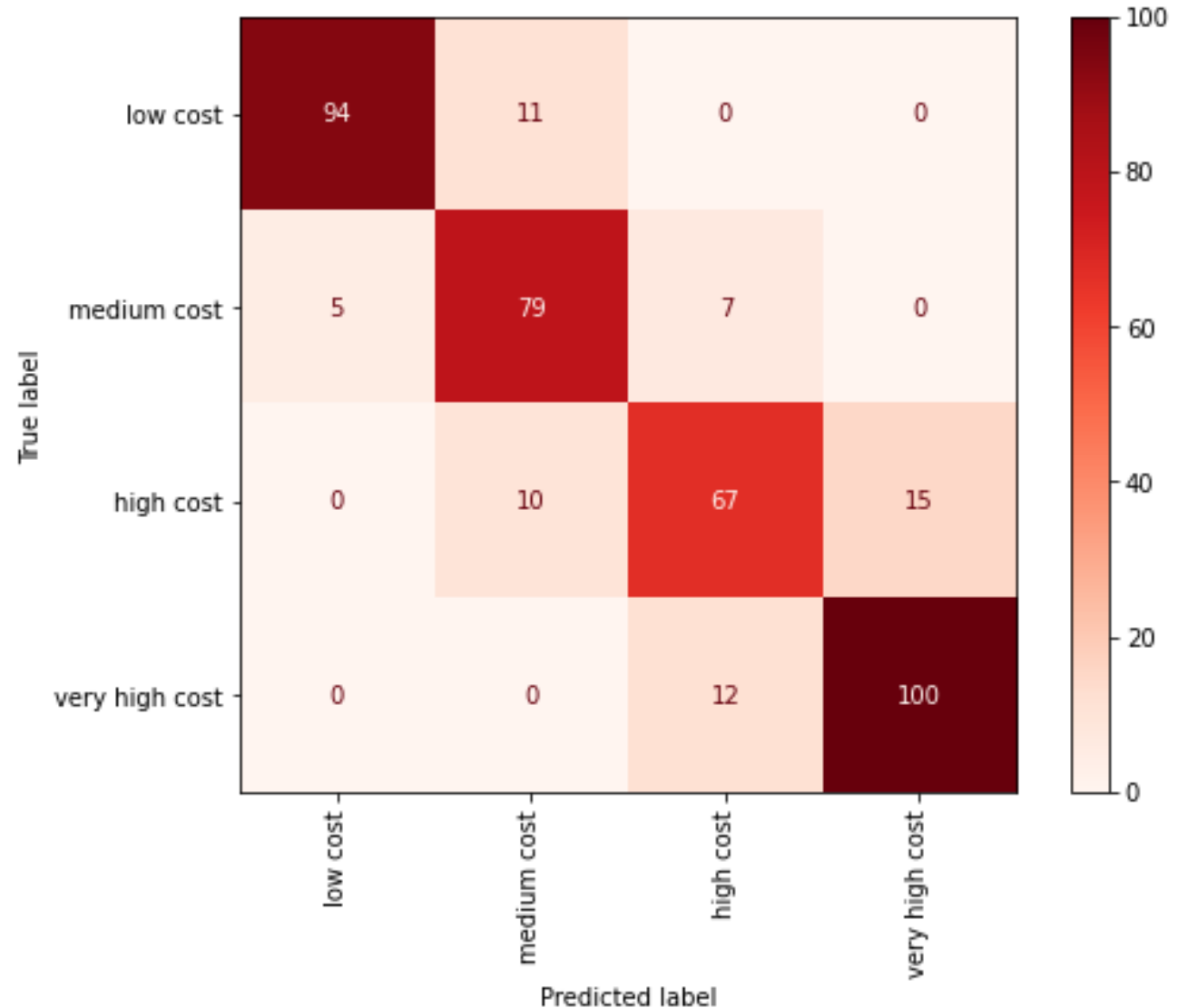
Algorithms and methods to be used to apply predictive modelling

- Decision tree classifier
- Random forest classifier
- Gradient boosting classifier
- Xgboost classifier
- K-nearest neighbours

Observation

AI

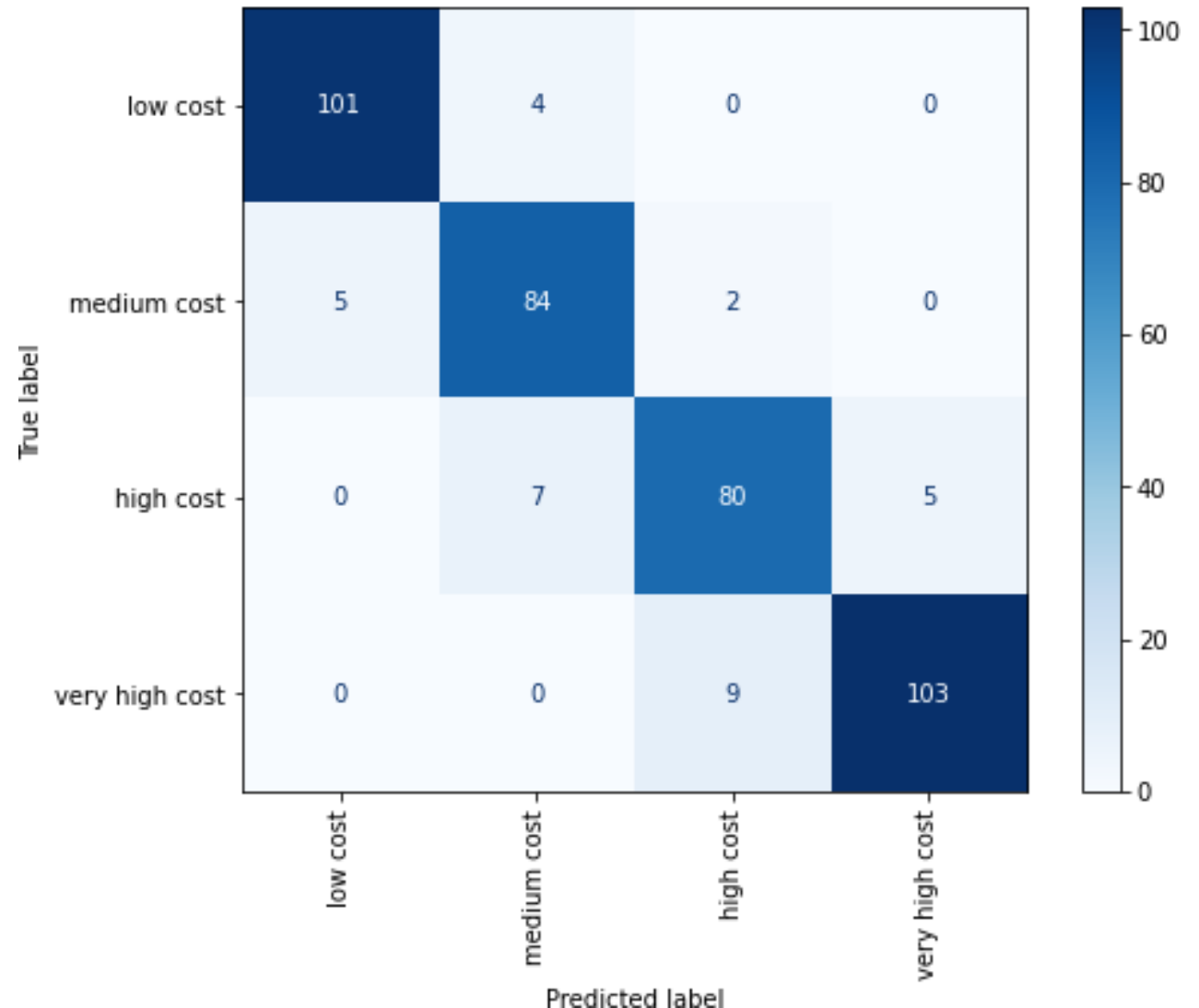
- The overall accuracy score on our test data is 84%.
- Prediction accuracy on class 1 (low cost) is 93%.
- Prediction accuracy on class 2 is 80%
- Prediction accuracy on class 3 is 63%
- Prediction accuracy on class 4 is 100%, As a result , we can see that the random forest classifier has performed good overall, and best in class 4.



Observations:

AI

- 1.The overall accuracy score on our test data is 92%.
- 2.Prediction accuracy on class 1 (Low Cost) is 100%.
- 3.Prediction accuracy on class 2 (Medium Cost) is 84%.
- 4.Prediction accuracy on class 3 (High Cost) is 80%.
- 5.Prediction accuracy on class 4 (Low Cost) is 100%. As a result, we can see that the Gradient Boosting Classifier has performed good overall, and best in class 1 & class 4.



Challenges

- The most challenging part in this exercise was to find an optimal set of parameters that could give us the best performance.
- It took hours to try every combination in cross validating and finally selecting the best value.
- The model could even perform better with even finer tunings.

Conclusion

A red square logo with the white letters 'AI' inside.

- Throughout the analysis we went through various steps to determine our predictions for the mobile price range. We started with simple eda where we analysed our dependent variable as well as other independent variables. We found out the correlation, count, relationships with the dependent variable. We looked for missing values and outliers and did some feature modifications.
- Finally we implemented 5 machine learning algorithms namely; Decision Trees, RandomForest, Gradient Boosting, K-Nearest Neighbours and XBoost. We tried hyperparameter tuning to reduce overfitting and increase model performance. The best performance was given by our Gradient Boosting model.
- We also tried to identify the important features impacting our model predictions . We saw ram, battery power, px_height and px_weight were major contributions.

THANKING YOU