**Business and marketing.**

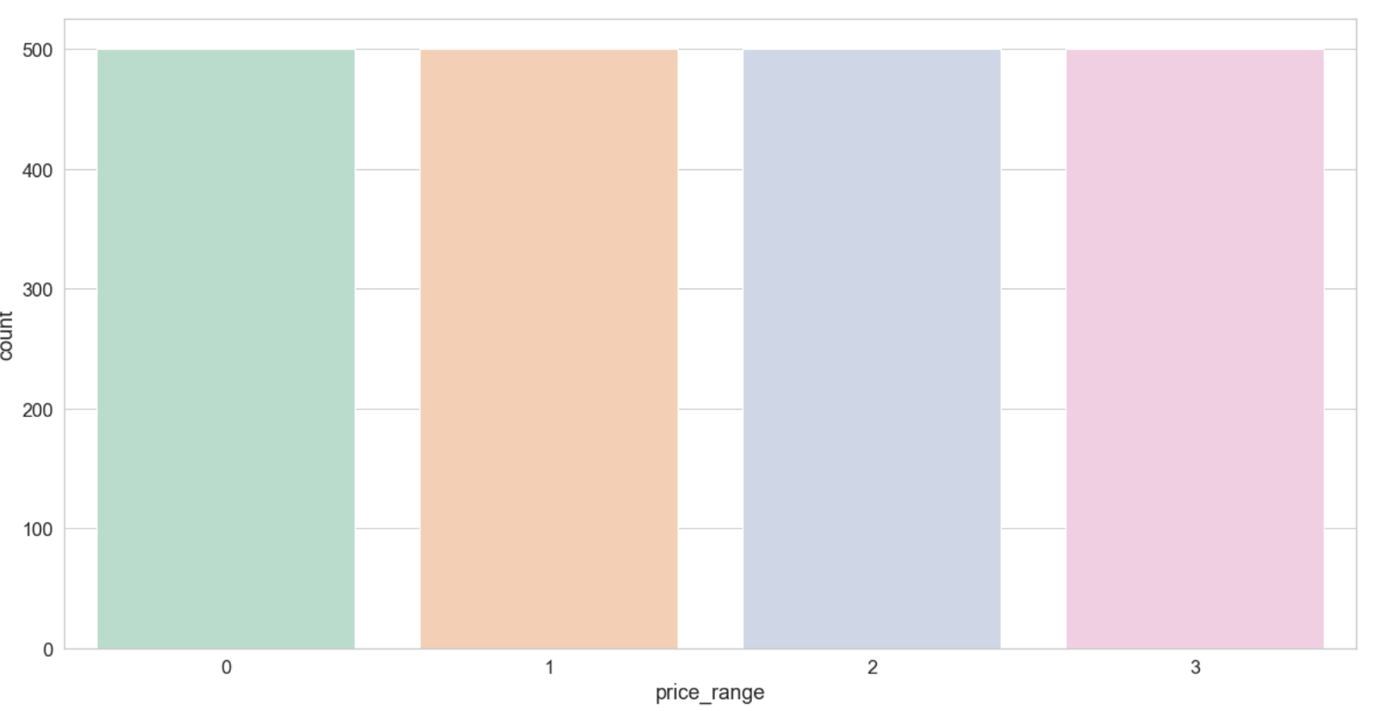
For the analysis of our data using python, first of all, I install python language on my computer and install all needed libraries and packages using pip command to be used over the project.

The steps for data analysis in python was as follows:  
1. Download the dataset that we choose from the Kaggle web site.

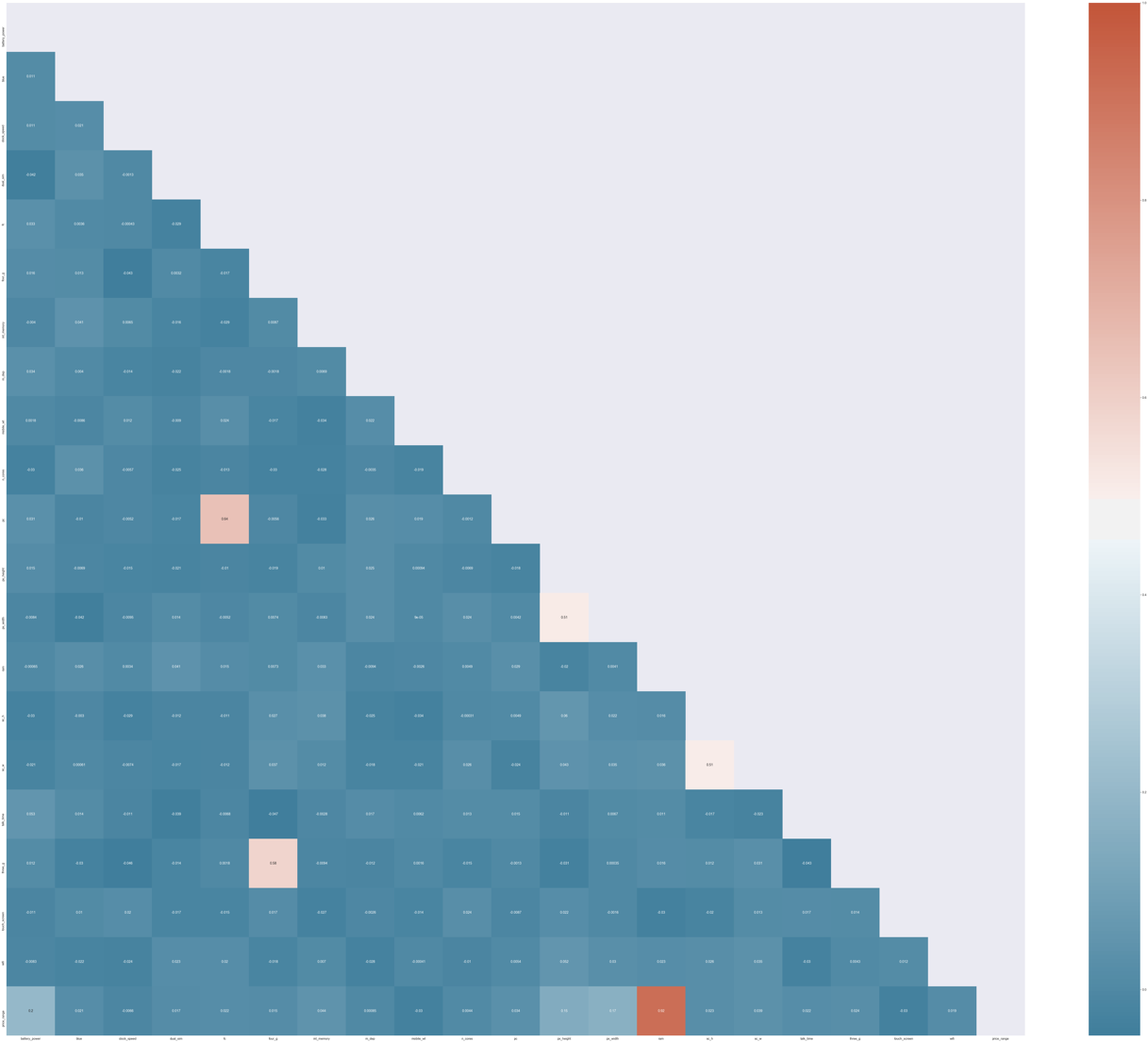
2. Read the CSV file that contains all data.

3. Specify which are the features and target of my project, where the features are the features of mobile and the target is *mobile price*.

4- My dataset is already pre-processed and cleaned and I don’t make any pre- process on it, I use it directly for classification in our project. Also, I check target and I observed that the price range is the same for all categories.



5. I apply the correlation matrix to see which feature of mobile has more effect on the mobile price to focus on it from customers or manufacturers. I note that variables *ram & mobile price* having strong positive correlation and there is a relation between *battery power & mobile price.*



5. I split the dataset for training and testing with 30% for testing and 70% for training.

6. I apply a random forest and decision tree algorithms.

|  |  |  |
| --- | --- | --- |
|  | Random forest | Decision tree |
| Accuracy score | **86.0%** | **80.0%** |
| Recall score | **85.5%** | **79.5%** |
| Precision Score | **85.5%** | **79.5%** |

7. Finally, I apply the confusion matrix to visualize the results of classification for each category of mobile price as shown in the dataset for both algorithms. I notice that category with number 1 that is for the highest range of price has the highest true results that are 138 and 13 false results in random forest and category 3 in decision tree is the highest true results is 132.

|  |  |
| --- | --- |
| Random forest | Decision tree |
|  |  |

**CONCLUSION**

I apply classification techniques using random forest and decision tree. My results show that the highest accuracy of classification using the random forest algorithm.

**REMAINING WORK**

I will apply algorithms with different train and test sizes of sizes to compare these results in order to obtain a higher accuracy rate. Also, I noticed that the results are all equal, so in the next period I will search for the reason for the equality.