**Business and marketing.**

* **Introduction**:

Recently years, with the spreading of using technologies especially smartphones and competition between companies that make up these smartphones, different features and prices for smartphones appeared to attract people to buy these devices. The features may be related to the components of the smartphone itself or it may be related to the services that smartphones are supported. In this project, we have tried to analyze the relationship between the features of smartphones and the unevenness of their prices that help the manufacturers of smart devices, mobile phone dealers in their trade and help users to buy smart devices at the best prices. To get the results, we apply a machine-learning approach with specific classification algorithms to classify mobile prices based on 20 features for smartphones such as battery power, ram and so on.

* **Purpose:**

The purpose of our dataset is for business and marketing.

* **Question:**

Which features affect the price of the mobile?

* **Source**:

The dataset that we are used in our project from Kaggle source that is one of the big sources for the dataset. Kaggle is one of the famous community for data scientists that involve a lot of datasets that help people to chive their objectives, explore and build helpful models and it provides challenges in data science.

* **Attributes**:

1. The power of the battery (battery\_power): This means the total of energy that battery can store in one time and is measured in mAh.
2. Blue: if the mobile has Bluetooth or not.
3. Clock\_speed: The speed of microprocessor to execute the instructions.
4. Dual\_sim: if the device has support of dual sim or not.
5. Fc: Front Camera in mega pixels.
6. Four\_g: if the mobile has and support 4G or not.
7. Int\_memory: Internal memory in GigaBytes.
8. M\_dep: Mobile depth in cm.
9. Mobile\_wt: weight of mobile phone.
10. N\_cores: number of cores of processor.
11. Pc: Primary camera in mega pixels.
12. Px\_height: Pixel Resolution Height.
13. Px\_width: Pixel Resolution Width.
14. Ram: Random Access Memory in Megabytes.
15. Sc\_h: The height of screen of mobile in cm.
16. Sc\_w: The width of screen of mobile in cm.
17. Talk\_time: The longest time that a single battery charge will last when you are.
18. Three\_g: if the mobile has and support 3G or not.
19. Touch\_screen: if the mobile has touch screen or not.
20. Wifi: if the mobile has wifi or not.

* **The objective for the dataset Analysis:**

The objective of the dataset analysis is to classify the prices of mobile into one of four categories. These categories of mobile prices are categorized from 0 to 3 in increasing order of prices i.e 0 will be for the lowest range and 3 for the highest range. These categories are based on 20 different features for mobile as described above to find if there is a high relation between the prices of mobile and these features based on the accuracy result of classification algorithms used. Also, find which one of the features is affected more than the others on mobile prices to help manufacturers of smart devices to develop these features to increase their profits and compete with other companies. On the other hand, it will help mobile phone dealers in their trade to bring the mobiles that have the best features to increase their profits and in marketing for mobiles. Also, it helps users to buy smart devices at the best prices with good features.

* **Dataset pre-processing:**

Our dataset is already pre-processed and cleaned and we don’t make any pre-process on it, we use it directly for classification in our project.

* **The data analysis step-by-step**

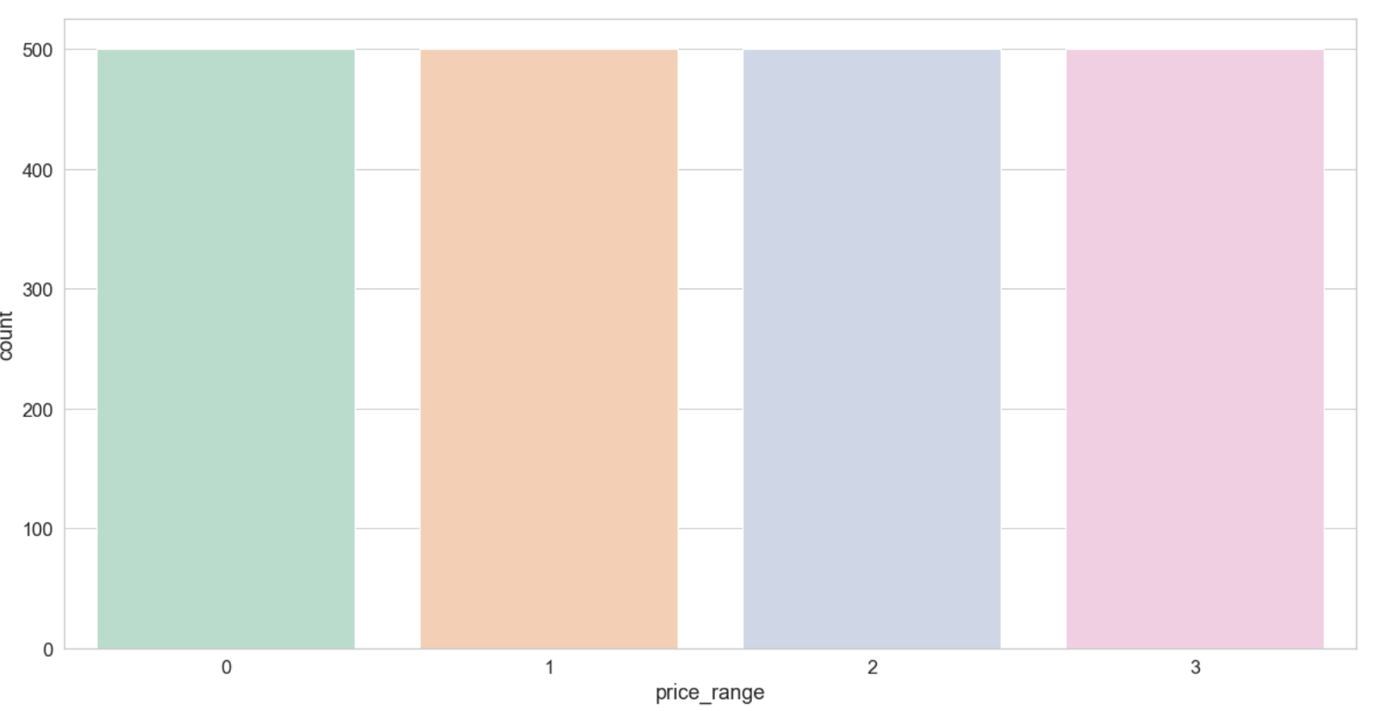
For the analysis of our data using python, first of all, I install python language on my computer and install all needed libraries and packages using pip command to be used over the project.

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

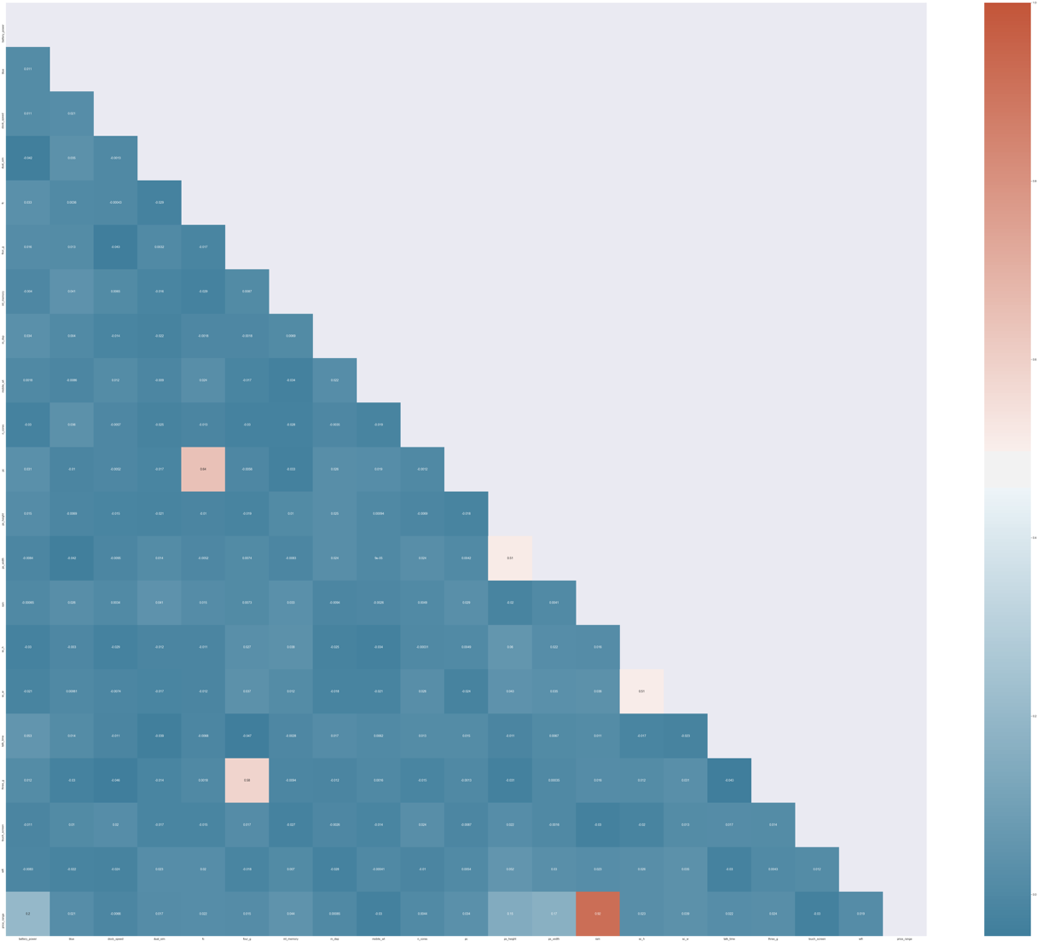
2. Read the CSV file that contains all data.

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 my 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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Random forest | Decision tree | **KNN classifier** | Logistic  Regression | Naive Bayes classifiers |
| Accuracy score | **86.0%** | **80.0%** | **93.17%** | **78.50%** | **79.83%** |

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

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 | |
|  |  | |
| KNN classifier | Logistic Regression | |
|  |  | |
| Naive Bayes classifiers | |
|  | |

* **CONCLUSION**

Todays, Machine-Learning algorithms are more important, because it helps people around the world to solve challenges, develop marketing strategies, improvement in medicine and so on. Different algorithms and techniques are used in a Machine-Learning approach. In this project, we choose a KNN classifier algorithm because it supports the classification technique and because of its high accuracy.

* **REMAINING WORK**

After we are working on this dataset, random forest algorithms and classification technique, we notice there can be some of the development for either dataset, algorithm and classification technique. So, we conclude some of the recommendations that may increase the accuracy of results or increase the benefits of the dataset. These recommendations that might be applied are as follows:

1. Increase the size of the dataset to increase the accuracy of classification.
2. Extend the idea of the dataset and applied on different things like power banks, cameras and so on.
3. Record the mobile prices as the price itself in integer format, not as category and apply regression technique on it to observe the effectiveness of mobile features on mobile prices. This recommendation to see prices clearly not as category and we don't know the domain for each category.