**Topic: Determining success of mobile devices based on reviews, rating and sales through data visualization.**

In this project firstly data had to be collected from all the different websites which had ratings of mobile phones based on different factors of rating them. Such as camera, memory, graphics, user interface performance of each mobile phones performed by the experts and ranked them accordingly.

These data were collected from different websites which contains data about these specific ranking of each devices based on the respective factors. This was done with the help of using BeautifulSoup4 package under python3 which helped to crawl over the website and collect the necessary data out of it. All of this data is then stored in a CSV file which can be used for further work on it.

A screenshot of a cell phone

Description automatically generated

Fig 1. Screenshot of the CSV file containing the complete dataset

Fig 1. Shows the screenshot of the dataset which was collected from different websites and stored as a table for ease of normalizing and knowledge discovery of data. Here we can observe that mobile devices were chosen from different companies and which were the most popular and best-selling of that company, so as to have a better point of reference for each of the company.

A screenshot of a cell phone

Description automatically generatedFig 2. Screenshot of the Normalized dataset

Fig 2. Show that values of the same data which was in Fig 1. But now it is normalized for a better understanding and bring all the values under one specific range which will be in between 0 and 1. This makes the graphs much easier to understand than the previous approach.

Now Analysis of Variance (ANOVA) test was performed on the same data set to understand the relation of each factor in the total ranking and the price of the mobile device.

A picture containing flower

Description automatically generated

Fig 3. Results of ANOVA test

Fig.3 Is the screenshot of the results obtained out of the ANOVA test where the F value is 316.124 which is much more than the expected F value which was 2.05 to satisfy the null hypothesis which meant all the means of factors are equal and hence the null hypothesis is reject and the conclusion out of it is that some factors have much more contribution than the others.

Tukey’s Significance test was performed next which is used to find out the significance between two factors or how different they are and below are the results of our dataset.

A close up of text on a white background

Description automatically generated

Fig 4. Results of Tukey HSD test

From Fig 4. A lot of new things were discovered where we see the camera ratings and the graphics performance rating didn’t have any significance with each other resulting in not reject the hypothesis. Which might be true since cameras aren’t much related to the graphical performance of a device. Also, the memory management and the UX came out to be false because there must be some factors in memory management not affecting the UX which could be concluded in the further tests or study.

A screenshot of a cell phone

Description automatically generated

Fig 5 Total of the specs score to the price graph

Fig 5. Was plotted to observe the change in the total specs score in the devices with the price ranges. And we can see there are a few devices even with a lower price point are still providing score range as high as the expensive phones which can be important to look at when comparing them to the sales of each devices over the year.A screenshot of a cell phone

Description automatically generated

Fig 6. Price to camera rating bar graph

Fig 6. Is plotted to check for the relation of the price of the phone and how well it performs with when it comes to the price of the device which shows there are a few exceptions being performing good even though being cheap and vice versa. All of this data is going to be taken under consideration in the further analysis of this study.

Later in this study, correlation of matrix will be studied so as to come up with firm results for the relation between each factor with respect to the price of the device as a whole. And all these results would be then compared to the annual sales of each device so as to measure the success of devices as compared to what ranking they had in each factor individually.