

## Termination Project

### 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.

FinalList

Name	DXO	CPU	GPU	MEM	UX	Total	Price
Xiaomi Redmi K20 Pro	102	140217	183068	60120	68349	451761	400
Xiaomi Mi 9	110	136311	180695	63725	64828	445567	445
Honor 20 pro	113	127930	125978	74776	62718	391410	500
Samsung Galaxy S10	116	121378	166060	59383	64365	411193	590
OnePlus 7T Pro	114	144011	197877	78107	74703	494705	600
OnePlus 7 Pro	113	142000	178641	78724	74619	473991	670
Apple iPhone 11	109	155235	211480	75370	82044	524137	700
Huawei p30	116	129216	131850	67960	65166	394199	750
LG G7 ThinQ	85	97120	103076	44170	53175	297546	760
Apple iPhone 11 Pro	115	167055	213503	83000	83365	546930	1000
Samsung Galaxy Note 10+ 5G	117	128071	171608	89136	65118	453939	1100
Apple iPhone 11 Pro Max	118	166509	215130	82035	83267	546948	1200

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.

```

NORMALIZED DATA
['Xiaomi Redmi K20 Pro', 'Xiaomi Mi 9', 'Honor 20 pro', 'Samsung Galaxy S10', 'OnePlus 7T Pro', 'OnePlus 7 Pro', 'Apple iPhone 11', 'Huawei p30', 'LG G7 ThinQ', 'Apple iPhone 11 Pro', 'Samsung Galaxy Note 10+ 5G', 'Apple iPhone 11 Pro Max']

[0.5151515151515151, 0.7575757575757576, 0.8484848484848485, 0.9393939393939394, 0.8787878787878788, 0.8484848484848485, 0.7272727272727273, 0.9393939393939394, 0.0, 0.9090909090909091, 0.9696969696969697, 1.0]

[0.616243654822335, 0.5683917923786373, 0.44055194108815326, 0.3468649460213855, 0.670494030170873, 0.6417387574176021, 0.8309859154929577, 0.45894044469864875, 0.0, 1.0, 0.4425680989490241, 0.992192750411096]

[0.7138700983454406, 0.6926928088243168, 0.2043836007639174, 0.5620861370410695, 0.8460295928748639, 0.6743623610045156, 0.9674264194049298, 0.25678690631302764, 0.0, 0.9854802148963893, 0.6115979795455763, 1.0]

[0.5026167605167274, 0.3859887379927128, 0.31609804571050015, 0.3706525339516396, 0.713083802583637, 0.7103814243126863, 0.9562437893342166, 0.3971844981782047, 0.0, 1.0, 0.3955945677376615, 0.9967538920172242]

[0.35471244940621804, 0.4348841346795356, 0.680647600409198, 0.33832228795089625, 0.7547257928212427, 0.7684472712716275, 0.6938575812836365, 0.5290664057287728, 0.0, 0.8635413423475515, 1.0, 0.8420806831828492]

[0.6183390670483797, 0.5935036607565296, 0.37635624413597324, 0.45567798173230367, 0.7905269404415362, 0.7074722736786393, 0.9085372210327103, 0.3875389932719064, 0.0, 0.9999278273630524, 0.6270719561198368, 1.0]

[0.0, 0.05625, 0.125, 0.2375, 0.25, 0.3375, 0.375, 0.4375, 0.45, 0.75, 0.875, 1.0]

```

Fig 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.

FVALUE AND PVALUE					
316.12405382717657 3.0603169653409916e-37					
SUMMARY					
	sum_sq	df	F	PR(>F)	
C(treatments)	1.608840e+11	4.0	316.124054	3.060317e-37	
Residual	6.997745e+09	55.0	NaN	NaN	

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.

Multiple Comparison of Means – Tukey HSD, FWER=0.10						
group1	group2	meandiff	p-adj	lower	upper	reject
CPU	DXO	-137810.4167	0.001	-149434.1187	-126186.7146	True
CPU	GPU	-137920.4571	0.001	-149544.1591	-126296.7551	True
CPU	MEM	-66545.5833	0.001	-78169.2854	-54921.8813	True
CPU	UX	-67778.0	0.001	-79401.702	-56154.298	True
DXO	GPU	-110.0404	0.9	-11733.7425	11513.6616	False
DXO	MEM	71264.8333	0.001	59641.1313	82888.5354	True
DXO	UX	70032.4167	0.001	58408.7146	81656.1187	True
GPU	MEM	71374.8738	0.001	59751.1718	82998.5758	True
GPU	UX	70142.4571	0.001	58518.7551	81766.1591	True
MEM	UX	-1232.4167	0.9	-12856.1187	10391.2854	False

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.

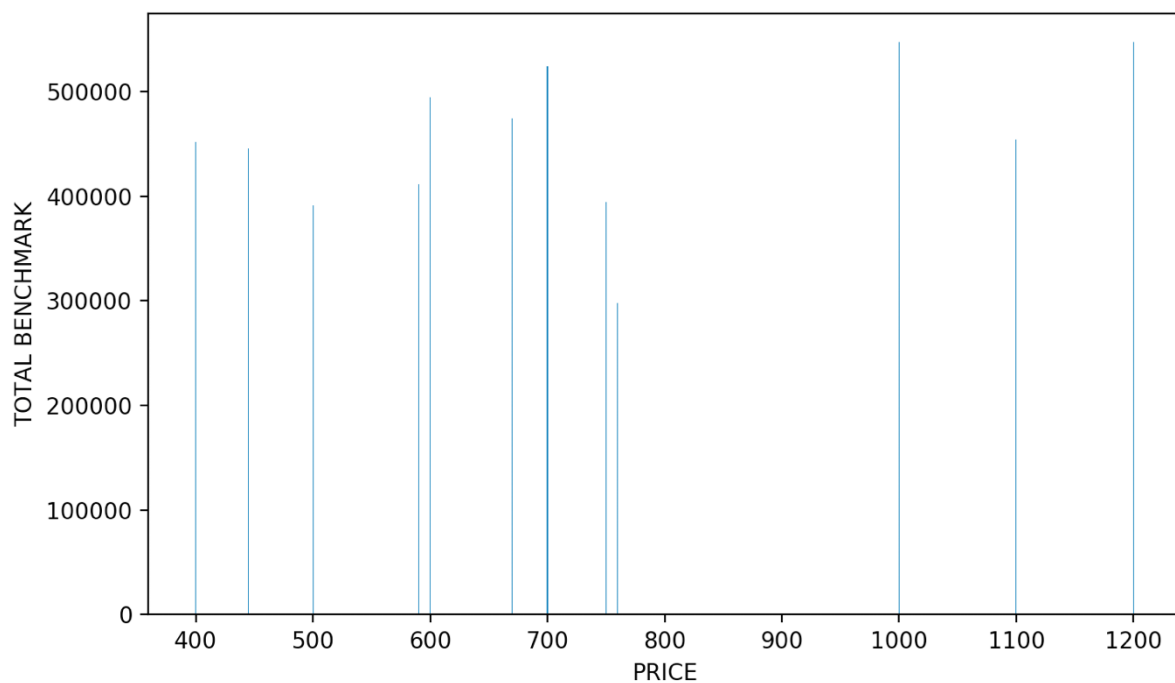


Fig 5 Total of the specs score to the price graph

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.

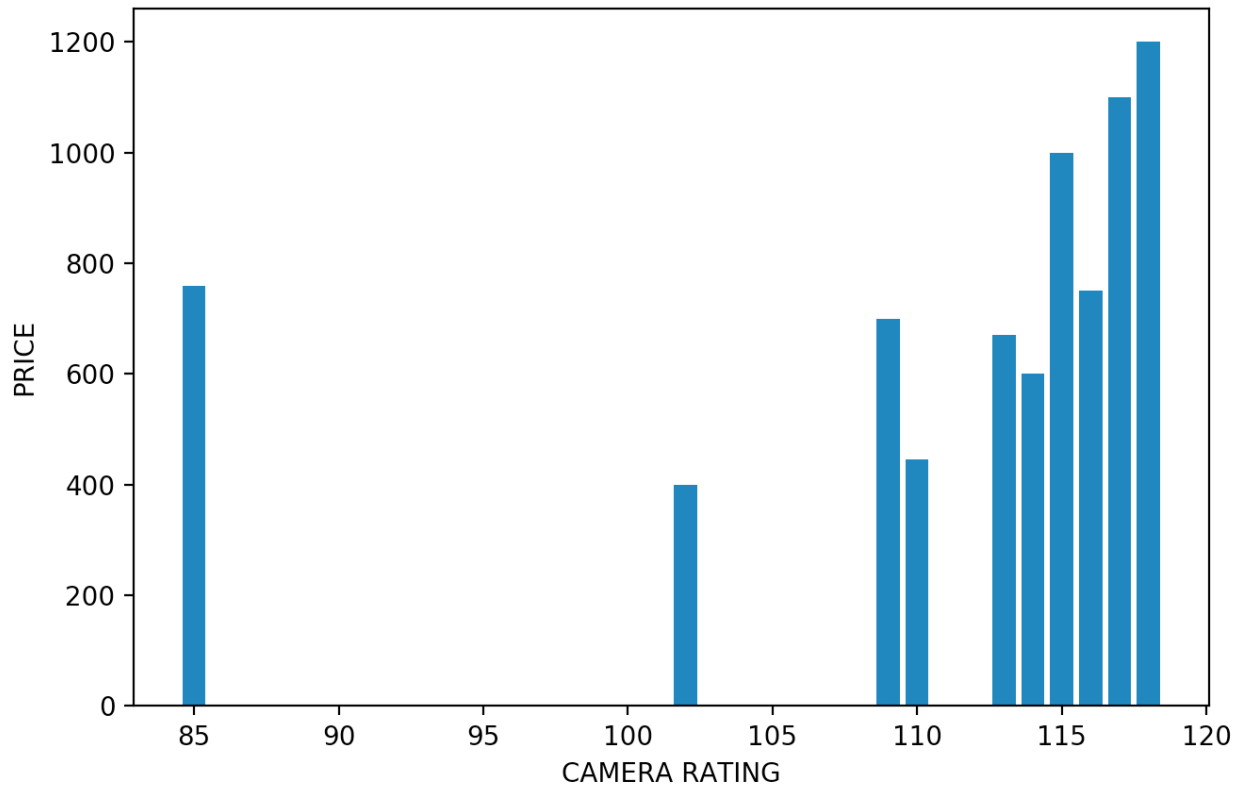


Fig 6. Price to camera rating bar graph

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.