MGT 6203 Group Project Proposal

Team #: Team 48

Team Members:

1. Mohamed Yousry Hussien; mhussien6 – I have an MSc. in petroleum engineering. I worked as a reservoir engineer for 8 years and I currently work as a planning engineer. I have worked on projects including data-driven competitors' intelligence and oil and gas wells' productivity prediction models.
2. Virginia Sahagun; vsahagun3 - Bachelor's in Pharmacological Chemistry from UCSD. Currently I work as an analyst. Previously, I have worked on research projects in genetics and pharmacology.
3. Shadman Islam Chowdhury, schowdhury87- I completed my bachelor's in computer science and now working as a business analyst. My previous projects -prostate cancer survivorship platform, NSF award data search platform, Maternal Health risk assessment tool.
4. Nazmus Sakib Sumon, nsumon3- I completed my MBA in Information Systems and am currently working as a Data Analyst. Previously I worked in projects related to optimizing student transportation services and building applications to manage fuel usage.
5. Adam Hall, ahall323 – I graduated with a BS in Finance and have worked as an analyst in Capital Markets and Business Intelligence since graduation. I have previously worked on projects in wealth management risk analytics, credit and lending approvals, and optimization in workflow automation.

OBJECTIVE/PROBLEM

Project Title: What features drive smartphone popularity in Indian Smartphone markets?

Background Information on chosen project topic: The smartphone market in India is a fierce and competitive market, where new brands of smartphones try to appeal to widely varied consumer demographics. It is projected that it will grow 10 % in 2023 to reach 175 million units [1], with the country having almost 600 million smartphone users [2]. With such a competitive and fragmented market, it has lowered the barriers for new entrants. So, conceptualizing a winning strategy starts with understanding which Smartphone features attract the consumers the most. This way the manufacturers can determine the proper pricing and marketing strategy and understand which features to focus on the most to penetrate this extremely competitive market.

Problem Statement: With a market as varied with different brands and configurations of smartphones as the Indian market, we want to use Ecommerce (Flipkart) data to understand which smartphone factors (Price, camera, screen etc.) drive smartphone popularity in such a market.

State your Primary Research Question (RQ): The primary research question we are focusing on is what are the specific features that impacts the popularity/ratings of a mobile phone in an Ecommerce site the most in India.

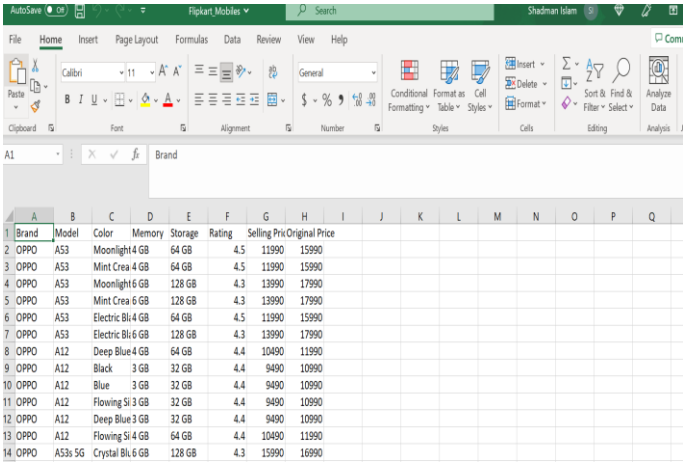
Supporting Research Questions: 1**)** Another research question we are going to focus on is how the feature choices vary among different price ranges. 2)The trend in impact of factors over time (2019 vs 2023)

Business Justification: We believe understanding the factors that impact the choices of consumers when buying a new phone, can play a big role to define the specification, the pricing strategy and the marketing mix of any new smartphones that mobile brands manufacture. In a country like India, with many brands, it is really important to understand which factors can affect the consumer's attitude's and behaviors towards smartphone purchase.

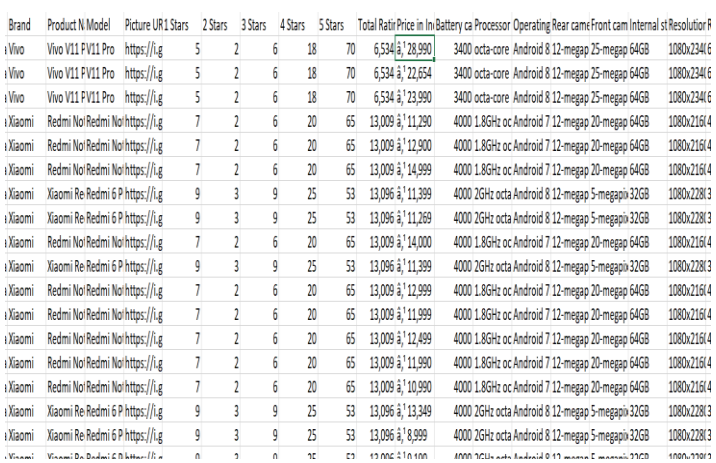
DATASET/PLAN FOR DATA

Data Sources: The initial data sets for preliminary EDA and forming the research questions were collected from Kaggle. They are- 1) Gadget360- https://www.kaggle.com/datasets/deepann/4000-laptops-data-from-gadgets360?select=mobiles.csv 2) Flipkart- https://www.kaggle.com/datasets/devsubhash/flipkart-mobiles-dataset 3)Flipkart Dataset-2023 version (Scrapped) 4) Processor Data- Also will be scrapped.

Data Description: One of the datasets were the NDTV/gadgets 360 dataset which contained smartphone specifications, the second dataset contained mobile phone price, rating and specifications from an Indian Ecommerce site called Flipkart. For a separate research question we will also use a new version of the Flipkart dataset which we will scrape and also use processor data. We plan on combining just the specifications dataset with the ecommerce dataset to see how specification impact the ratings and popularity in the ecommerce site. We will use another version of the Flipkart dataset (number 3) to do the historical trend analysis of feature choices. Below are snippets of the data from NDTV and Flipkart (2019 version)-



Brand	Model	Color	Memory	Storage	Rating	Selling Pric	Original Price
OPPO	A53	Moonlight	4 GB	64 GB	4.5	11990	15990
OPPO	A53	Mint Crea	4 GB	64 GB	4.5	11990	15990
OPPO	A53	Moonlight	6 GB	128 GB	4.3	13990	17990
OPPO	A53	Mint Crea	6 GB	128 GB	4.3	13990	17990
OPPO	A53	Electric Bl	4 GB	64 GB	4.5	11990	15990
OPPO	A53	Electric Bl	6 GB	128 GB	4.3	13990	17990
OPPO	A12	Deep Blue	4 GB	64 GB	4.4	10490	11990
OPPO	A12	Black	3 GB	32 GB	4.4	9490	10990
OPPO	A12	Blue	3 GB	32 GB	4.4	9490	10990
OPPO	A12	Flowing Si	3 GB	32 GB	4.4	9490	10990
OPPO	A12	Deep Blue	3 GB	32 GB	4.4	9490	10990
OPPO	A12	Flowing Si	4 GB	64 GB	4.4	10490	11990
OPPO	A53s 5G	Crystal Bl	6 GB	128 GB	4.3	15990	16990



Brand	Product N	Model	Picture UR	1 Stars	2 Stars	3 Stars	4 Stars	5 Stars	Total Ratir	Price in In	Battery ca	Processor	Operating	Rear came	Front cam	Internal st	Resolution F
Vivo	Vivo V11	FV11 Pro	https://i.g	5	2	6	18	70	6,534	â¹28,990	3400	octa-core	Android 8	12-megap	25-megap	64GB	1080x2340
Vivo	Vivo V11	FV11 Pro	https://i.g	5	2	6	18	70	6,534	â¹22,654	3400	octa-core	Android 8	12-megap	25-megap	64GB	1080x2340
Vivo	Vivo V11	FV11 Pro	https://i.g	5	2	6	18	70	6,534	â¹23,990	3400	octa-core	Android 8	12-megap	25-megap	64GB	1080x2340
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹11,290	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹12,900	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹14,999	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Xiaomi Re	Redmi 6 P	https://i.g	9	3	9	25	53	13,096	â¹11,399	4000	2GHz octa	Android 8	12-megap	5-megapix	32GB	1080x2280
Xiaomi	Xiaomi Re	Redmi 6 P	https://i.g	9	3	9	25	53	13,096	â¹11,269	4000	2GHz octa	Android 8	12-megap	5-megapix	32GB	1080x2280
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹14,000	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Xiaomi Re	Redmi 6 P	https://i.g	9	3	9	25	53	13,096	â¹11,399	4000	2GHz octa	Android 8	12-megap	5-megapix	32GB	1080x2280
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹12,999	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹11,999	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹12,499	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹11,990	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Redmi No	Redmi No	https://i.g	7	2	6	20	65	13,009	â¹10,990	4000	1.8GHz oc	Android 7	12-megap	20-megap	64GB	1080x2160
Xiaomi	Xiaomi Re	Redmi 6 P	https://i.g	9	3	9	25	53	13,096	â¹13,349	4000	2GHz octa	Android 8	12-megap	5-megapix	32GB	1080x2280
Xiaomi	Xiaomi Re	Redmi 6 P	https://i.g	9	3	9	25	53	13,096	â¹8,999	4000	2GHz octa	Android 8	12-megap	5-megapix	32GB	1080x2280

Key Variables: Dependent- Ratings **Independent**- Camera, Battery, Screen, Price , brand name, Processors etc.(All other smartphone features and related information)

APPROACH/METHODOLOGY

Planned Approach and model comparison: First, we will use linear regression to see if there is a linear relationship between each phone features (independent variables) on the phone ratings (dependent variable). The linear model can also be used to determine the effect of each feature on phone ratings. Secondly, we can build a decision tree model. The decision tree model can be used to create a model that predicts the impact of different phone features on ratings allowing us to identify which features have the greatest impact on phone ratings. The smartphone specifications from the Gadget360 dataset will be merged with the Flipkart dataset. For the regression analysis we can use k-fold cross validation to select the best model. Additionally, since our datasets have many features, we can use principal component analysis which will allow us to focus our analysis on the most important features. We will use R square values, predictions accuracy, mean square errors to compare the models.

Anticipated Conclusions/Hypothesis: The anticipated conclusion of our analysis will be to find the specific factors that impact the “rating” value the most. We will be looking into particular specifications that garner higher ratings. Based on our market research, the factors we believe will be the most important are price, brand, storage capacity and battery. [3] We want to see if our hypothesis on the top drivers of rating hold in the conclusion of our analysis. We also want to see if there has been any historical change in the top drivers of rating. Lastly, we hypothesize that depending on the price, the drivers start to vary along with rating, creating clusters, which we also want to find through our research as a secondary research outcome.

What Tools we will be using?: For Analysis and Modelling- R Language, RStudio; For scrapping – Python; For visualization- Tableau/R; For Code Maintenance- GitHub

What business decisions will be impacted by the results of the analysis? What could be some benefits?

The findings of our work can be used by any new smartphone manufacturers to identify the most influential factors affecting sales and devise different marketing strategies accordingly. For example. If the study shows that camera is one of the key factors in their target demography, they can focus on giving a good camera in the smartphone and shift their R&D budget from other aspects of the smartphone. The market potential of conducting such research is huge as it will lead to better product development, improved marketing strategies, creating a competitive advantage with other brands and lastly the targeted product development will lead to higher customer satisfaction.

PROJECT TIMELINE/PLANNING

Project Timeline/Key dates to achieve certain milestones by: With the project submission deadline being April 16th, we plan on completing the project in approximately 1 month 10 days. We plan on dividing our tasks into the following dates- 1. Data Extraction- By March 16th, 2. Data Processing- By March 24th, 3. Project Proposal Presentation Video- By March 25th,4. Preliminary EDA- By March 27th , 5. Feature Engineering- By March 31st ,6. Project Progress Report- By April 1st ,7. Modelling- By April 5th, 8. Model Comparison- by April 12th ,9. Final Report- by April 15th ,10. Final Presentation Video and slides- by April 17th

How will progress towards these goals and outcomes be measured: We will measure the goals of the project through the completion of each step of analysis before the given due dates and having them consolidated in the progress reports as deliverables. The progress towards the project goals will also be represented in the project proposal presentation video and the project progress report. Our final progress will be measured by detailing all our findings in the final report. The impact of our results on the market can be gauged through further analysis where we ingest our model with new data to further solidify our findings.

Appendix

References

[1] B. Standard, "Business Standard," [Online]. Available: https://www.business-standard.com/article/technology/india-smartphone-market-to-grow-10-to-reach-175-million-units-in-2023-122122300733_1.html.

[2] T. E. Times, "The Economic Times|News," 26 October 2021. [Online]. Available: https://economictimes.indiatimes.com/news/india/indias-growing-data-usage-smartphone-adoption-to-boost-digital-india-initiatives-top-bureaucrat/articleshow/87275402.cms?from=mdr.

[3] C. Buckle, "GWI," 9 January 2019. [Online]. Available: https://blog.gwi.com/chart-of-the-week/smartphone-features-consumers/.