



Popular Phone Features in India

Adam Hall, Mohamed Hussien, Virginia Sahagun,
Nazmus Sakib Sumon, Shadman Chowdhury

Table of Contents

1. Background Information and Research Findings
2. Value Proposition of Our Research
3. Initial Hypothesis
4. Our Approach to the problem
5. Data Source Overview
6. Data Cleaning Process
7. Finding from Exploratory Data Analysis
8. Methodology
9. Results From Machine Learning Model
10. Conclusion and Future Plans

Background Info & Research



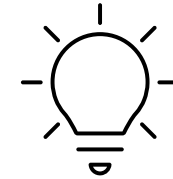
- The Indian market is ultra-competitive
- Market expected to grow 10 % in 2023 to reach 175 million units[1]
- New Competitors emerging like Realme,Xiomi etc.
- Emergence of Ecommerce site like Flipkart helps increase sale of smartphones through partnership
- New market entrants needs insights on features that are the main drivers of popularity among the buyers

Value Proposition



Overview of the problem

- What are the main drivers of ratings in Indian Smartphone market?
- No predictive method to determine rating based on smartphone configuration



Planned Solution

- A Regression based model to predict ratings
- Determining the features that impact the model the most(i.e-main drivers)

Initial Hypothesis

- Research show that price, brand, storage capacity, speed and battery life may be most useful predictors of popularity overall[2]
- Research done in Neighboring country Bangladesh showed that price and operating system are the main factors[3]
- The trend of choice has been similar throughout the years



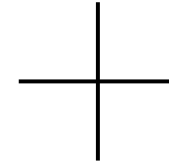
The Novelty in Our Approach



Focus in South Asian
Market

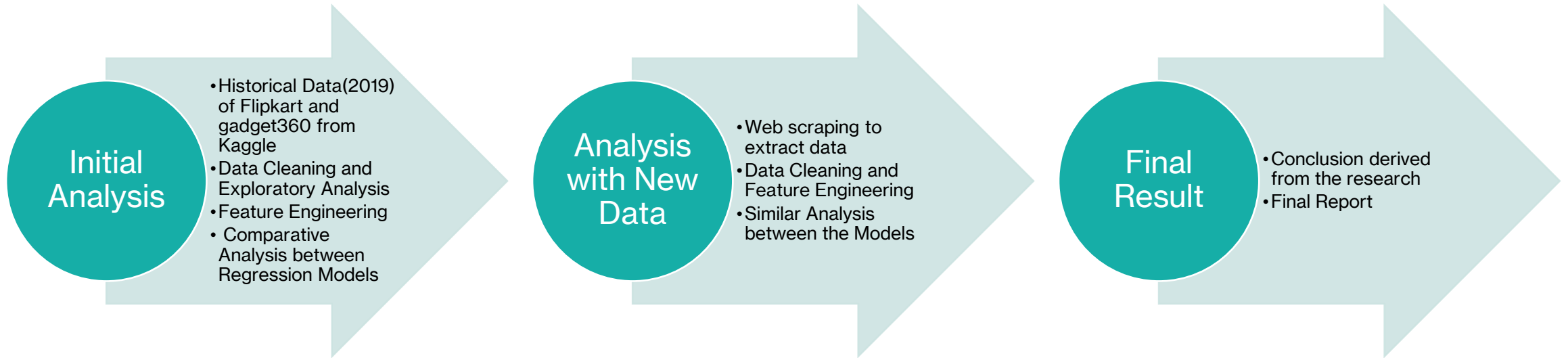


Focusing on Historical
Trend



Augmentation of the
Data Source

Approach

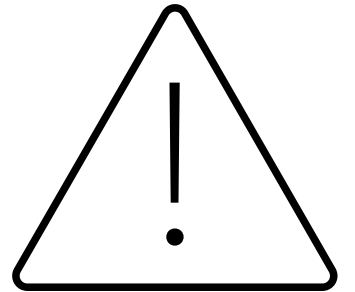


Similar projects review

- Google Play App Rating Prediction[4]
 - Regression based models used- Linear Regression, XGBoost Regressor, Random Forrest Regressor
 - Random Forrest Regressor had the lowest RMSE
 - Due to Multicollinearity Linear Regression had very high RMSE
- Phone Price Classification and Explanatory Data Analysis[5]
 - Main price driver is RAM and Battery life

Web Scrapping and Challenges

- Initial data found from 2019
- No research done on latest data(2023)
- Got the latest data from Gadget360 and Flipkart through Web Scrapping
- Challenges-
 - Different URLs for getting the Full mobile specifications
 - Removing "rare" specifications.
 - A lot of sponsored Ad that caused trouble getting appropriate data



Data Cleaning and Processing Process

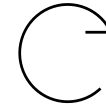
- Two complementary datasets from Flipkart.com and Gadgets360.com (includes additional features missing in Flipkart).
- Two snapshots of time (each of 2 datasets):
 - 2019 version (Kaggle.com)
 - 2023 version (web scraped)



Two different datasets
collected from Kaggle



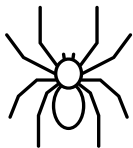
Whitespace removed
from the columns that
will be joined together



Unit conversion for
RAM, Storage size, etc.



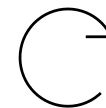
Duplicate and null
values removed



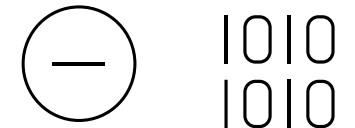
Two different datasets
collected by Web scraping



Whitespace removed
from the columns that
will be joined together



Unit conversion for
RAM, Storage size, etc.



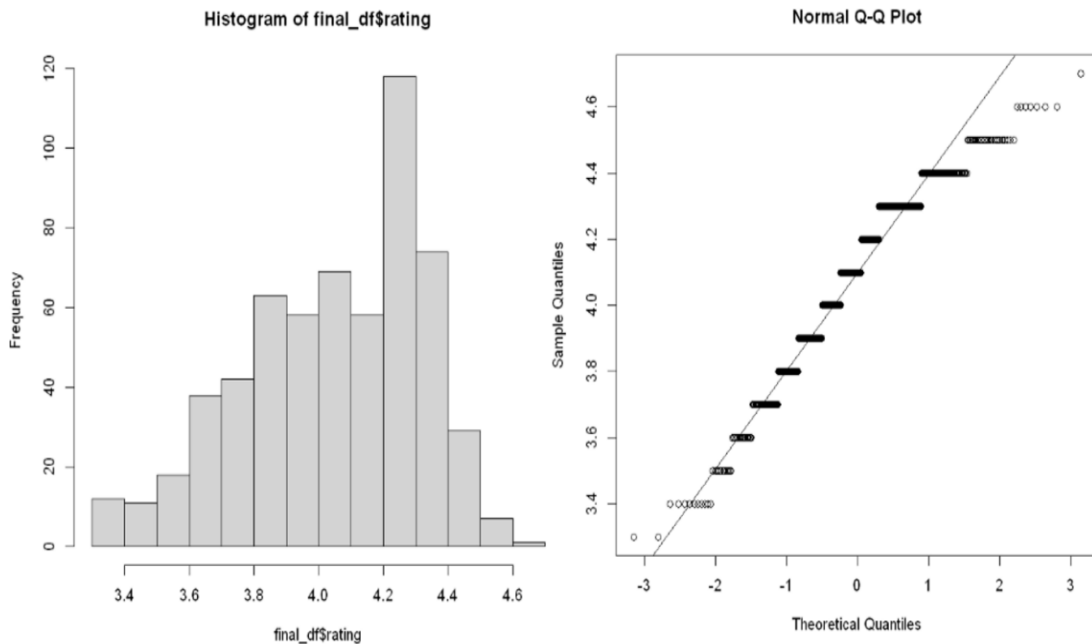
Duplicate and null
values removed

Summary of Datasets

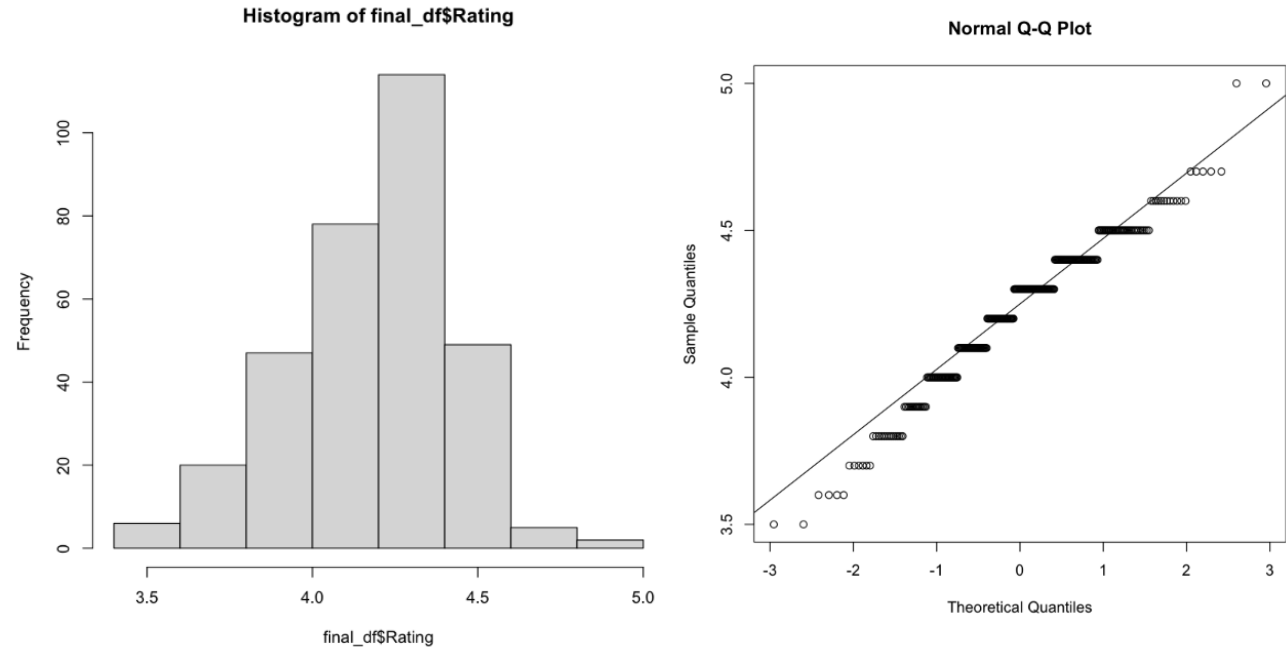
	2023 Dataset		2019 Dataset	
Step	Gadgets360	Flipkart	Gadgets360	Flipkart
Initial dataset size	8,695 rows x 50 columns	1,998 rows x 17 columns	1,359 rows x 21 Columns	3,114 rows x 8 Columns
Remove unnecessary columns	8,695 rows x 6 columns	1,998 rows x 13 columns	1,359 rows x 19 Columns	3,114 rows x 4 Columns
Drop missing values	7,800 rows	1,998 rows	1,359	2,970 rows
Drop Duplicates	4,268 rows	1,040 rows	1,359	1,079 rows
Final Merged Datasets	599 rows x 17 columns		387 rows x 23 columns	

Challenges Found through EDA

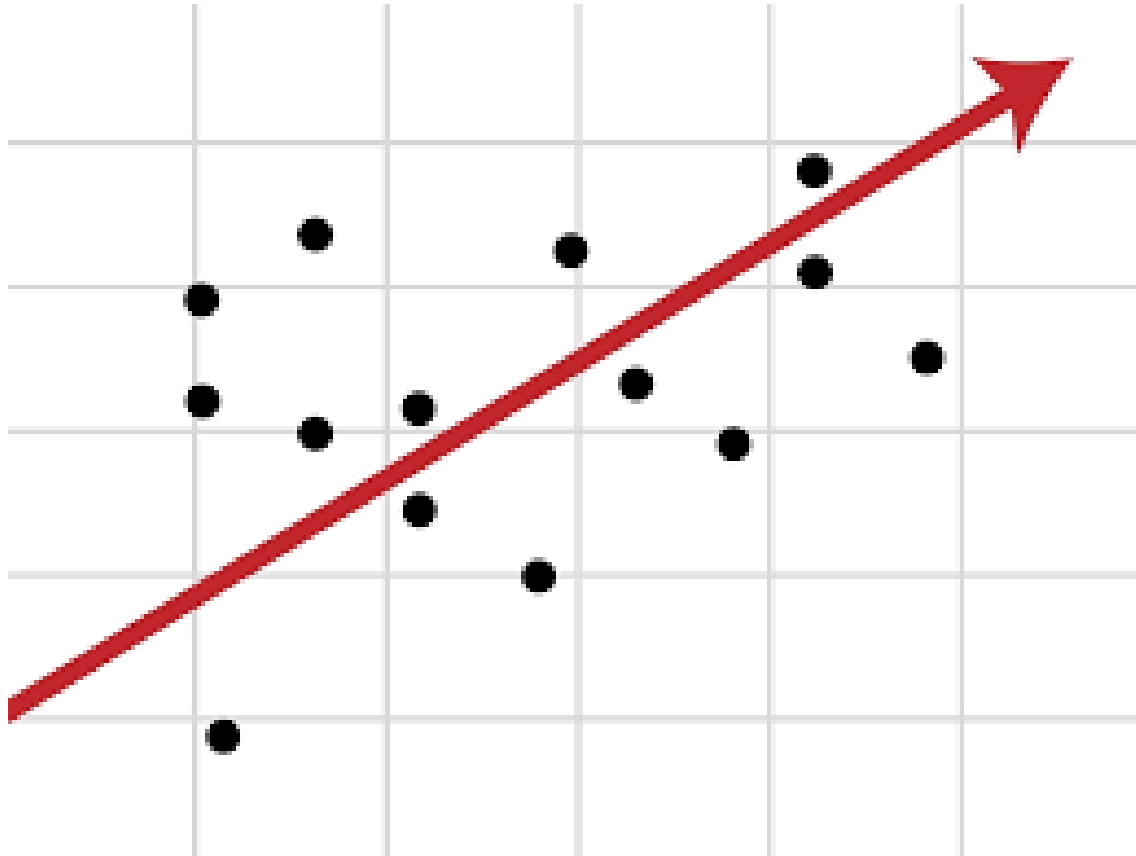
- Response variable (Rating) is not Normally distributed due to the common rating of 4.2
- Simple variable transformations were not able to normalize the data.



2023 dataset



2019 dataset



Methodology Overview

- Linear Regression
- Random Forest
- XGBoost
- Compare models with adjusted R^2 , MSE, and RMSE
- Training set- 70%, Test Set- 30%
- Outliers removed using Cook's distance
- Random Forest and XGBoost chosen to handle multicollinearity

Linear Regression

2019 Top Features

Brand

Battery Capacity

Internal Storage

Rear Camera

2023 Top Features

Brand

Price

Display

Rear Camera

Front Camera

Random Forest Top Features

Most Important Features			
Node Purity	2019	2023	Node Purity
4.01	Brand	Brand	10.19
2.40	Price	Screen Size	4.01
1.44	Screen Size	Price	3.00
1.17	Battery Capacity	Battery Capacity	2.33
0.95	Resolution	Days Since Release	1.71
0.56	Storage	Storage	1.62
0.48/0.44	Camera Quality	RAM	1.42

XGBoost Top Features

- Verbose=.5
- Max depth=3
- Learning Rate=0.1

2023 Variable Importance



2019 Variable Importance



Results

2023	Linear	Random Forest	XGBoost
R-Squared	66.41%	67.70%	62.98%
Mean Squared Error	0.026	0.025	0.029
Root Mean Squared Error	0.161	0.160	0.169

2019	Linear	Random Forest	XGBoost
R-Squared	57.38%	63.20%	56.10%
Mean Squared Error	0.024	0.023	0.026
Root Mean Squared Error	0.156	0.153	0.162

Conclusion and Future Impact

- Brand, Price, Display, Battery Capacity and Storage are the most consistently significant factors across different models
- Historical Trend
 - Decreased importance of the cell phone's camera quality and resolution from 2019 to 2023
 - Increase of importance of Processing speed from 2019 to 2023
- We want to focus on feature importance for multiple price points in a future iteration of the research

Sources Cited

1. B. Standard, "Business Standard," [Online]. Available: https://www.business-standard.com/article/technology/indiasmartphone-market-to-grow-10-to-reach-175-million-units-in-2023-122122300733_1.html.
2. Pekka Kekolahti, Kalevi Kilkki, Heikki Hämmäinen, Antti Riikonen, Features as predictors of phone popularity: An analysis of trends and structural breaks, *Telematics and Informatics*, Volume 33, Issue 4, 2016, Pages 973-989, ISSN 0736-5853, <https://doi.org/10.1016/j.tele.2016.03.001>.
3. Uddin, M. R., Lopa, N. Z., & Oheduzzaman, M. (2014). FACTORS AFFECTING CUSTOMERS'BUYING DECISIONS OF MOBILE PHONE: A STUDY ON KHULNA CITY, BANGLADESH. *International Journal of Managing Value and Supply Chains*, 5(2), 21.
4. <https://medium.com/analytics-vidhya/is-it-possible-to-predict-rating-of-google-play-store-apps-based-on-the-given-information-da9a44a3ac1e>
5. <https://github.com/teguharia172/Phone-Price-Classification-and-Exploratory-Data-Analysis>
6. Khan Rakib, M. R. H., Pramanik, S. A. K., Amran, M. A., Islam, M. N., & Sarkar, M. O. F. (2022). Factors affecting young customers' smartphone purchase intention during Covid-19 pandemic. *Heliyon*, 8(2), e10599. <https://doi.org/10.1016/j.heliyon.2022.e10599>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9476370/>
7. Odaymat, Rommy. (2019). Factors affecting mobile phone selection.
https://www.researchgate.net/publication/333461879_Factors_affecting_mobile_phone_selection