JUMIA SMARTPHONE PRICE PREDICTION

Optimizing Retail Strategies through Data Analytics

OUTLINE

- Business Understanding
- Data Understanding & preparation
- Modelling
- Model Evaluation: Model of Choice
- Deployment
- Conclusion & Recommendations

OVERVIEW

The African e-commerce market led by platforms like Jumia has experienced rapid growth with over 100,000 retailers competing for consumer attention.

Retailers are struggling to develop effective pricing strategies due to the complexity of dynamic market conditions and limited tools for real time pricing evaluations.

The proposed solution is a predictive pricing model that leverages sales data, competitor prices and market trends to automate and optimize pricing decisions.

BUSINESS PROBLEM

Retailers on Jumia are struggling with pricing strategies due to intense competition and changing market conditions, with slow and inefficient pricing evaluation methods. To address this, we aim to develop a predictive pricing model using sales data, competitor prices, and market trends. This will automate pricing evaluations, helping retailers set competitive prices and improve performance during peak periods like Black Friday, ultimately giving them the tools to succeed in the competitive e-commerce landscape

PROJECT OBJECTIVES

Predictive Model

Develop a predictive model to forecast optimal smartphone pricing based on key product features.

Analyze the dataset

Analyze the dataset to identify key features that impact smartphone pricing on the Jumia platform.

Smartphone Market Analysis

Analyze the smartphone market in detail to identify the most dominant smartphone brand on the Jumia platform.

DATA UNDERSTANDING

The data was scrapped from Jumia KE on 31-oct-2024 focusing on 12,000 smartphones listed by popularity. The data is saved in CSV and available for review in the data repository.

Key product features were extracted and organized from the product name column to gain insights into each phone's features. This was done using Regular expression module.

A structured data frame was created and numerical data extracted from Rating, Number of Reviews and Search Ranking column for further analysis.

DATA PREPARATION

The Dataset was cleaned for duplicates and missing values, revealing the following key insights:

Screen Size: Most phones have an average screen size of 6.6 inches (5.0–6.88 inches).

RAM: 4 GB is the most common (1 GB to 8 GB range).

Storage (ROM): 128 GB is standard, with some phones offering 32 GB to 256 GB.

Warranty: Most phones come with a 1-year warranty, some with 2 years, and a few with no warranty.

Price: The average price is KES 13,735, with most phones priced between KES 10,000–15,000.

Rating: Average rating of 4.3, indicating positive feedback.

Reviews: Average number of reviews is 79, with some phones receiving hundreds.

Page Listings: Phones span 1–300 pages, showing a wide range of options.

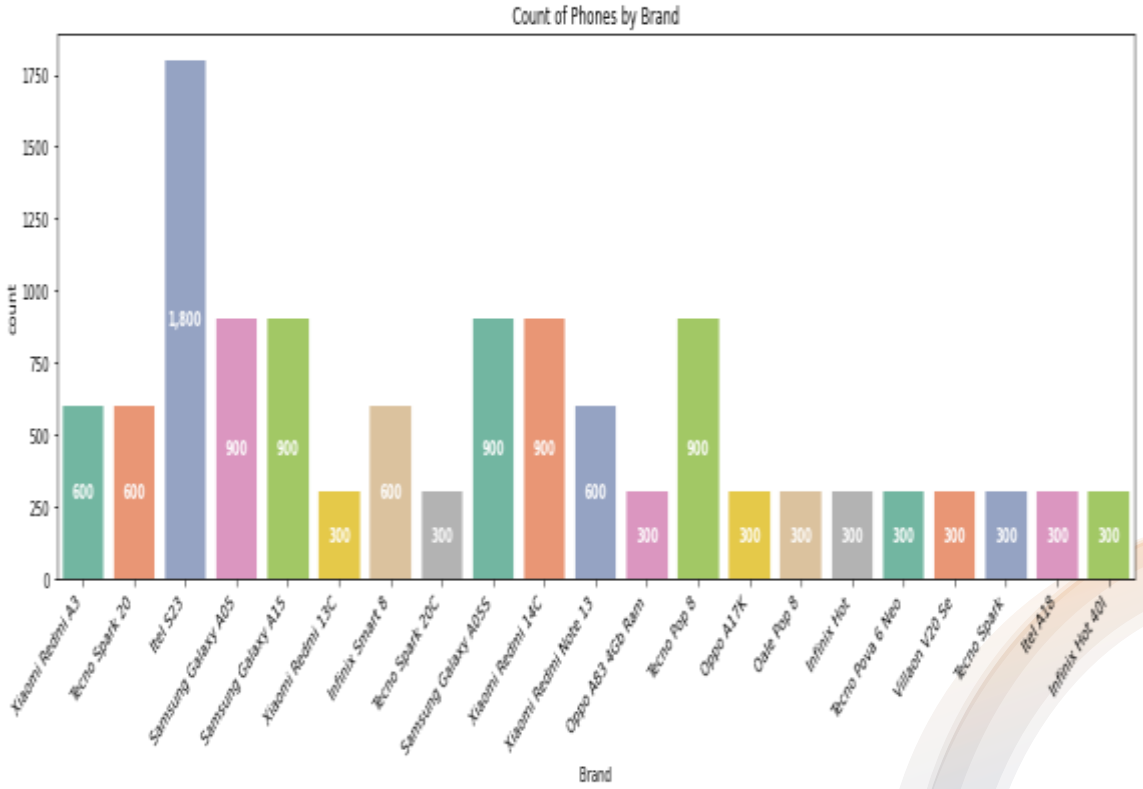
Rank: Phones typically rank between 1 and 30, indicating competitive listings.

Overall, the Market is dominated by mid-range phones with standard features

COUNT OF PHONES BY BRAND

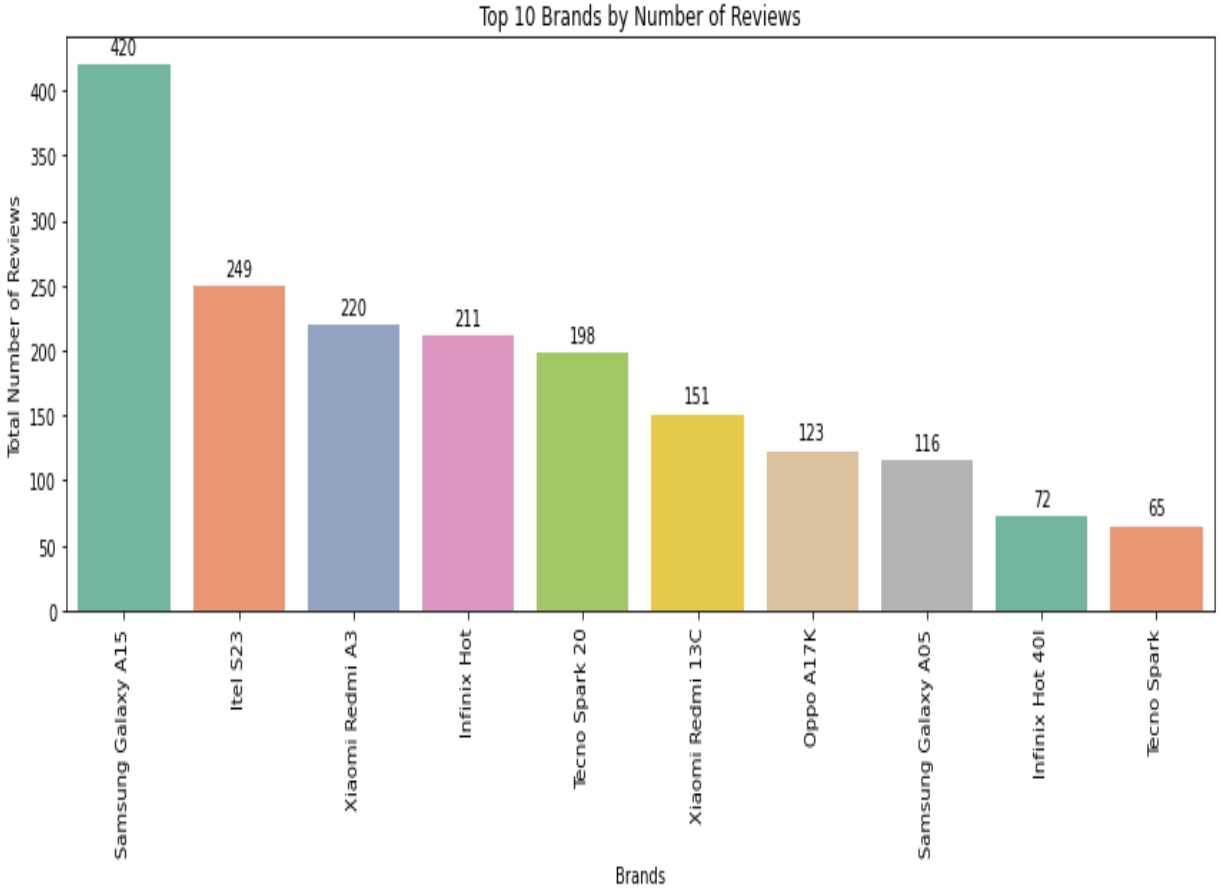
Itel S23 appears to be the most listed product which could possibly imply that this phone is a fast-moving product due to demand among customers on the platform followed by Samsung Galaxy A05, XIOMI Redmi 14C, Samsung Galaxy A15 and Tecno POP 8 at a very close range which highlights the competitive landscape in this segment.

These findings could inform inventory decisions and marketing strategies to enhance visibility and sales for these trending products.



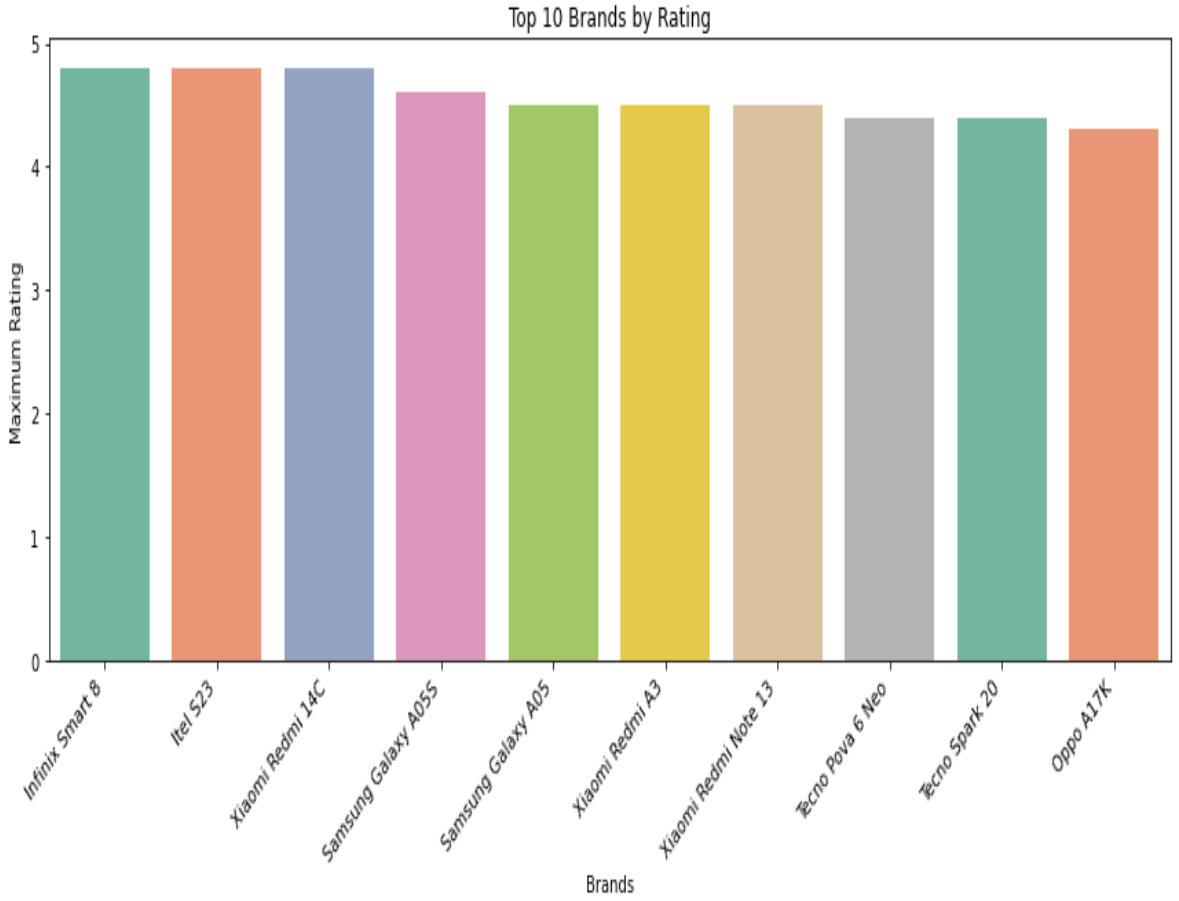
TOP 10 BRANDS BY NUMBER OF REVIEWS

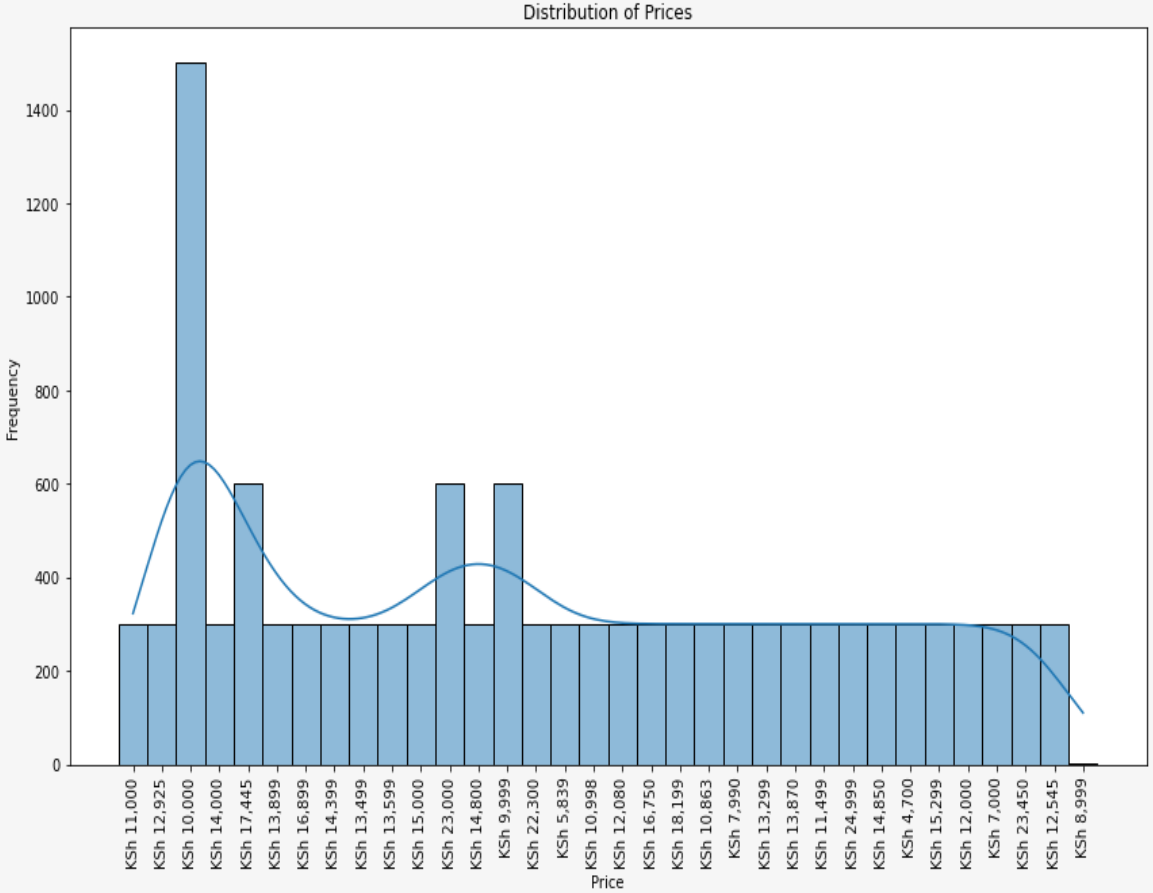
Samsung Galaxy A15 leads with over 400 reviews, followed by Itel S23 with 250. Xiaomi Redmi A3 and Infinix Hot have around 200 reviews, while other models, like Tecno Spark, have fewer. Possibly suggesting greater consumer interest in the Galaxy A15 and Itel S23.



TOP 10 BRANDS BY RATING

The bar plot shows the top 10 smartphone brands with maximum ratings near 5, indicating high customer satisfaction and performance across their top brands, highlighting a competitive market.





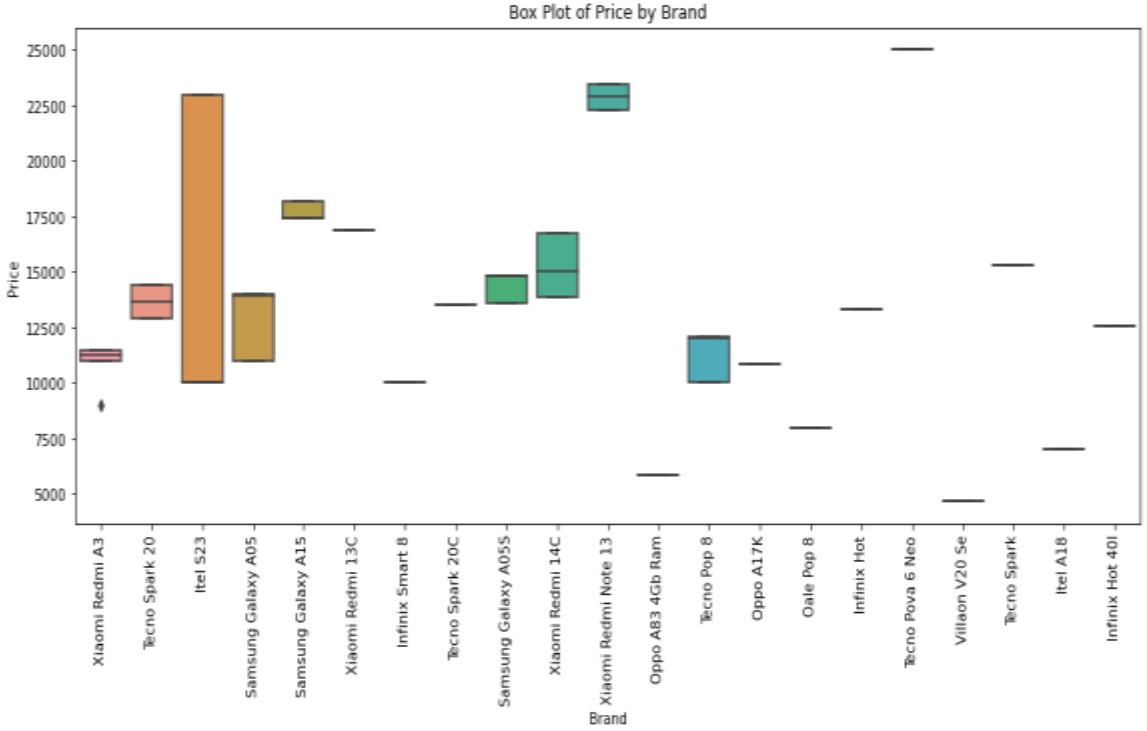
DISTRIBUTION OF PRICES

The histogram shows that most prices are concentrated around KES 11,000, with fewer phones priced higher, creating a right-skewed distribution.

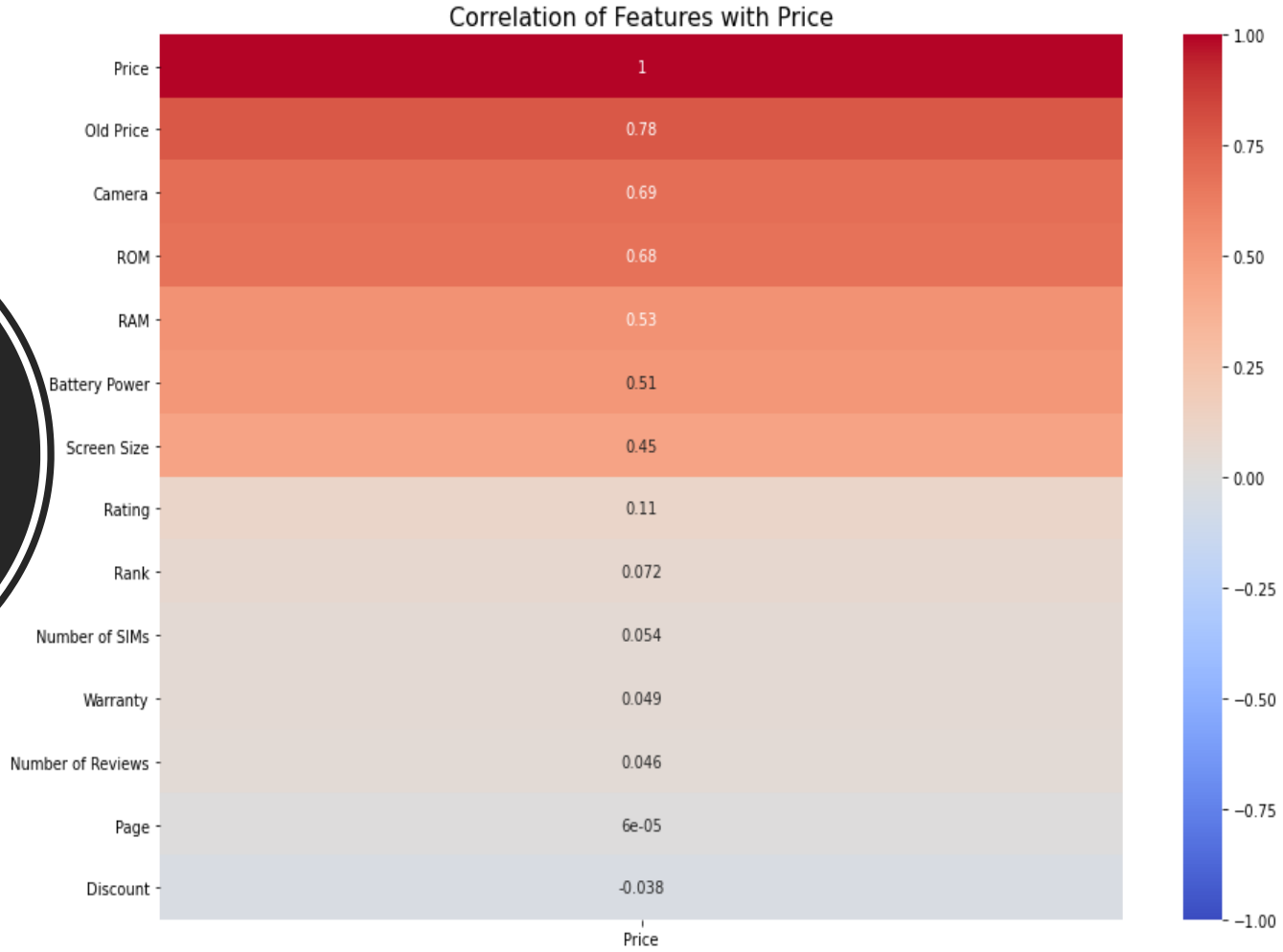
JUMIA

BOX PLOT OF PRICE BY BRAND

The box plot shows that some brands, like Itel and Samsung, have wide price ranges, while others, like Oppo and Infinix, have narrower ones. Budget-friendly models like Tecno Spark and Xiaomi Redmi A3 stand out, with occasional price outliers



CORRELATION OF FEATURES WITH PRICE



HYPOTHESIS TESTING

Hypothesis 1: Buyer Reviews vs. Product Pricing

Null Hypothesis (H_0): No relationship between reviews and pricing.

Alternative Hypothesis (H_1): A significant relationship exists.

TEST RESULTS:

p-value = 0.0000
(Statistically significant)

Spearman Correlation = 0.1536 (Weak positive relationship)

KEY INSIGHTS:

Higher-priced products tend to have slightly more reviews, but the correlation is weak.

Price alone doesn't significantly drive review counts - other factors like quality and marketing play a bigger role.

HYPOTHESIS TESTING

Hypothesis 2: Product Rank vs. Number of Reviews

Null Hypothesis (H_0): No relationship between product rank/page and number of reviews.

Alternative Hypothesis (H_1): A significant relationship exists between rank/page and number of reviews.

TEST RESULTS:

p-value for Rank = 0.0000 (Statistically significant)

p-value for Page = 1.0000 (No significant relationship)

Spearman Correlation for Rank = -0.1536 (Negative relationship)

KEY INSIGHTS:

Rank Impact: Higher-ranked products (top positions) tend to have fewer reviews, likely due to saturation of reviews for popular products.

Page Impact: No significant effect; the page position of a product does not influence review counts.

Conclusion: Rank is a significant predictor of review volume, but page placement has no notable impact.

MODELING

- Linear Regression
- Polynomial Regression
- Decision Tree Regressor
- Random Forest
- Gradient Boosting Regressor
- Extreme Gradient Boost
- Neural Networks

MODEL SELECTION

Metrics:

Mean Squared Error (MSE): Calculates the average squared difference between the predicted and actual values.

Mean Absolute Error (MAE): Calculates the average absolute difference between predicted and actual values, treating all errors equally without emphasizing large or small deviations.

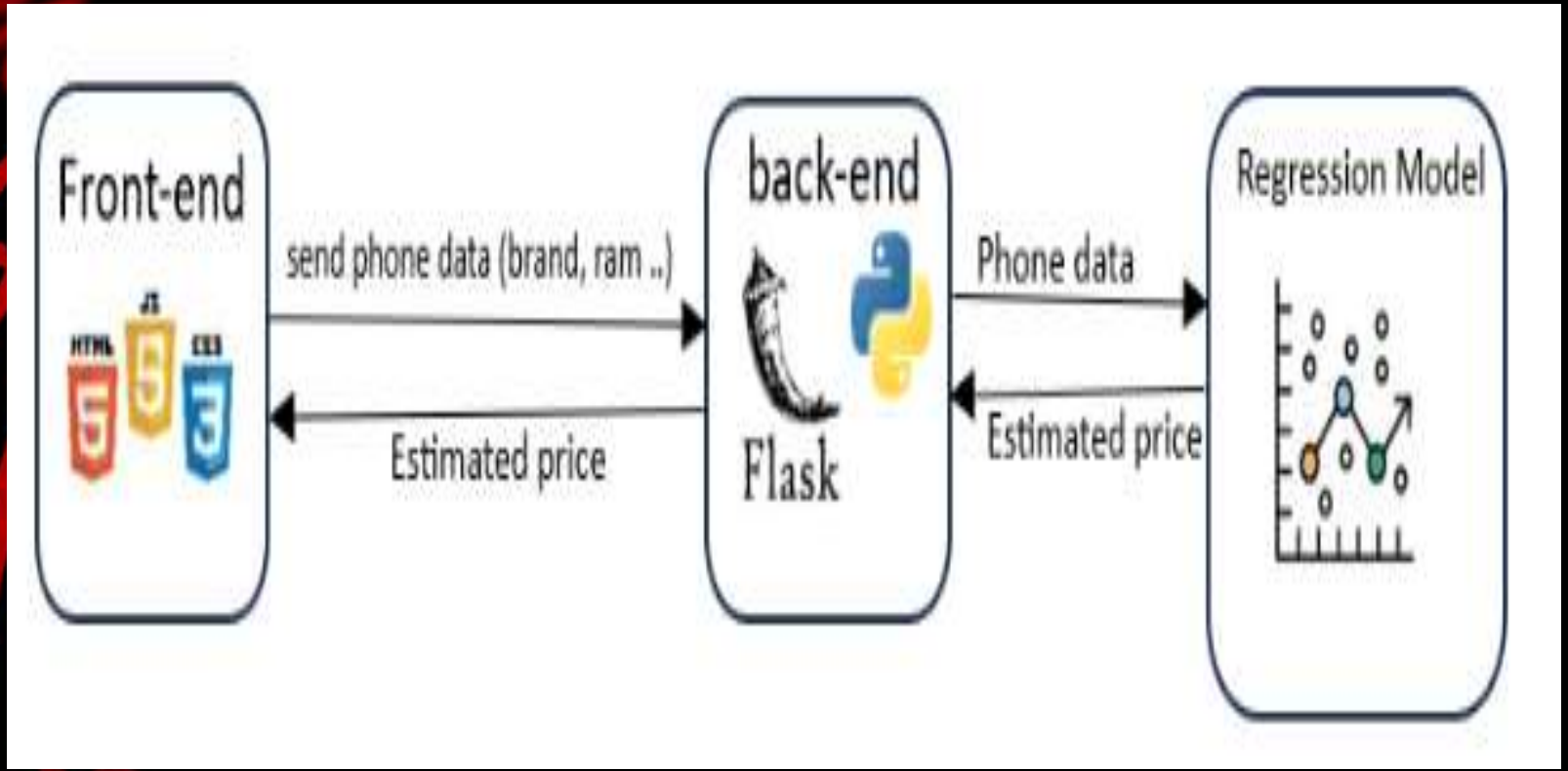
R-squared (R^2): Measures the proportion of variance in the target variable that is explained by the model. It provides a relative measure of model performance compared to a baseline

RESULTS:

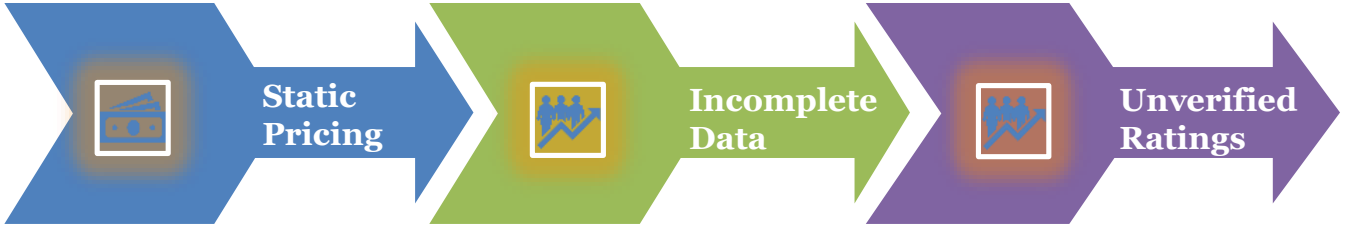
While all the models perform similarly with almost identical test metrics, Random Forest stands out slightly due to its lowest MAE (Mean Absolute Error) and MSE (Mean Squared Error). This indicates that, on average, Random Forest delivers more accurate predictions compared to the other models.

	Train Split			Validation		
Model	MAE	MSE	R^2 Score	MAE	MSE	R^2 Score
Linear Reg	49.70	2281 0.38	0.999	53.64	24455 .06	0.999
Polynomial Reg	2.95	567.6 6	0.999	2.69	126.1 8	0.999
Decision Tree	2.91	567.6 6	0.999	2.65	126.1 8	0.999
Random Forest	2.91	567.6 6	0.999	2.66	126.2 2	0.999
Gradient Boosting Reg	2.91	567.6 6	0.999	2.65	126.1 8	0.999
Extreme Gradient Boost	2.91	567.6 6	0.999	2.66	126.1 8	0.999
Neural Networks	78.99	1210 4.45	0.999	78.37	11375 .73	0.999

MODEL DEPLOYMENT IN WEB ENVIRONMENT



FUTURE WORK AND LIMITATIONS



01

Static pricing: The price data is static and may not reflect real-time market changes. Periodic data collection or real-time updates would provide more accurate insights.

02

Incomplete data: Variability in phone listings, such as missing details, could skew comparisons and affect value assessments. Standardizing data collection and validation would improve consistency.

03

Unverified ratings: Insights from ratings and reviews may be biased or unreliable. Using verified reviews and additional metrics, like sales data, would provide a more accurate assessment.

Benefits of Addressing these Limitations

Addressing these limitations would improve data reliability, leading to more accurate insights into price trends, brand popularity, and consumer preferences.

CONCLUSION

Features such as ROM, RAM, and screen size are key factors in determining phone pricing. To optimize pricing, sellers should consider these critical features in relation to market expectations and competition.

Product visibility within top ranks on a page has a more significant impact on customer engagement than the overall page placement.

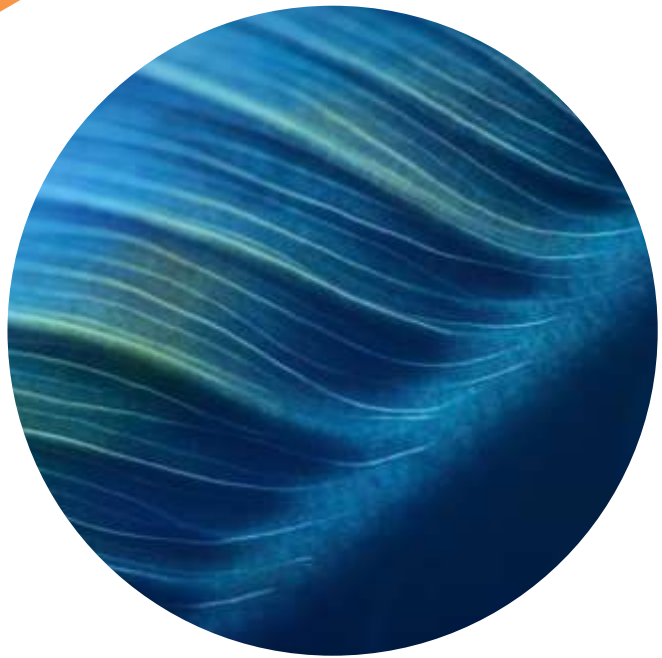
While there is a weak positive relationship between pricing and the number of reviews, pricing alone is not a major driver of review volume. Other factors like product visibility and marketing play a more substantial role.

RECOMMENDATIONS

Holistic Pricing and Marketing Approach: Focus on strategic pricing, effective marketing, and visibility through ads, while adjusting based on market trends and customer preferences.

Feature-based Pricing Optimization: Tailor pricing based on key features like ROM, RAM, and screen size to ensure competitiveness in the market while aligning with customer expectations.

A data-driven approach: Combining these elements will lead to better retail strategies and higher performance in the competitive smartphone market.



QUESTIONS & ANSWERS

APPRECIATION