

# *iPhone Customer*

## *Reviews*

### Data Analysis



*Supervised by:*

Dr. Sara El-Mowafy

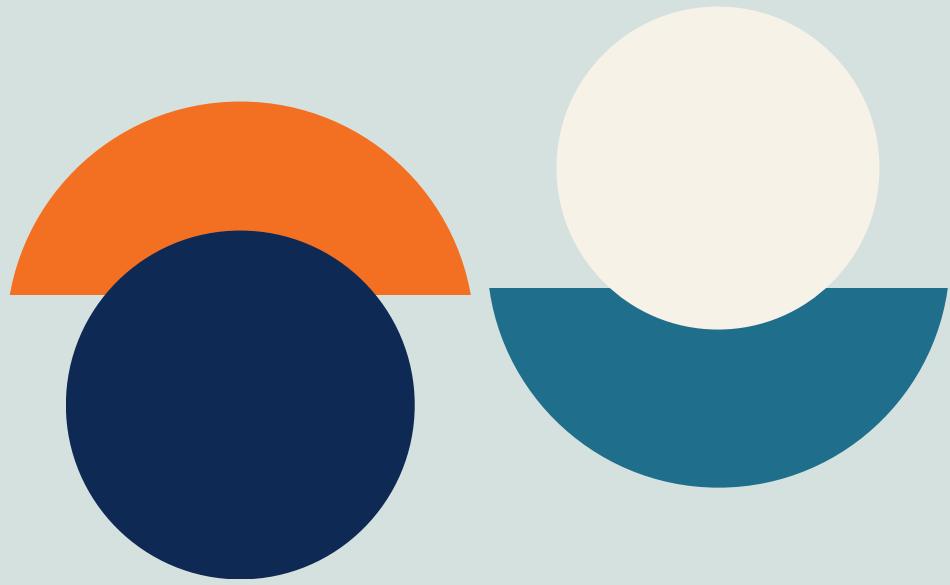
*Prepared by:*

Code Crushers Team

# *Introduction*

This report analyzes customer reviews for various products, focusing on factors such as ratings, review titles, descriptions, and verification status. The data collected spans multiple countries and covers a wide range of product variants. By understanding customer feedback, the goal is to provide actionable insights that can help improve product quality, customer satisfaction, and ultimately drive sales growth.





# *Key Questions*

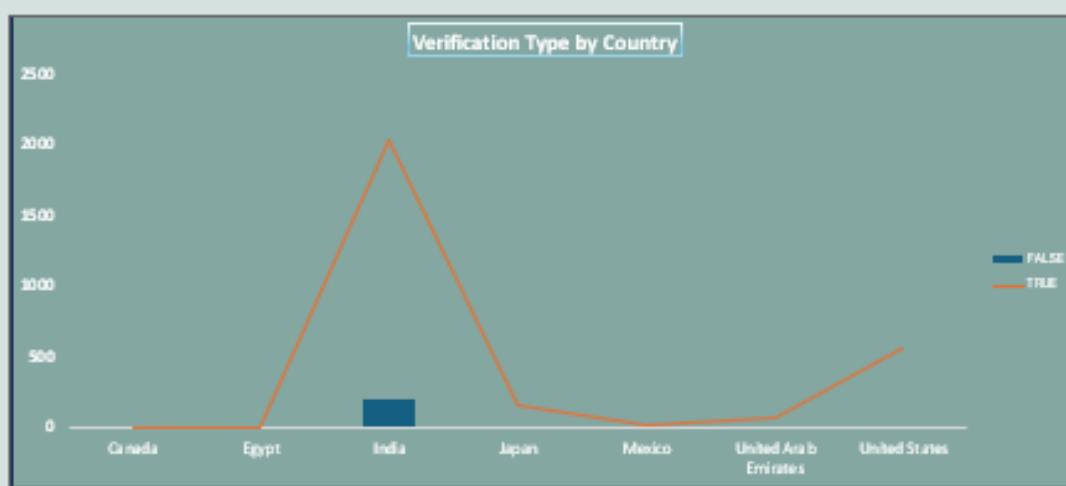
- How do customer ratings vary across different countries?
- Which countries have the highest and lowest product ratings?
- Are verified reviews generally more positive than unverified reviews?
- What are the most common themes or topics in the review titles and descriptions?
- Use text analysis to identify the most repeated words or common phrases.
- Which product variants receive the best reviews?
- Are certain variants consistently rated higher than others?
- How do review trends change over time?
- Do product ratings improve or decline as more reviews are collected?

# Data Analysis

## Descriptive Statistics:



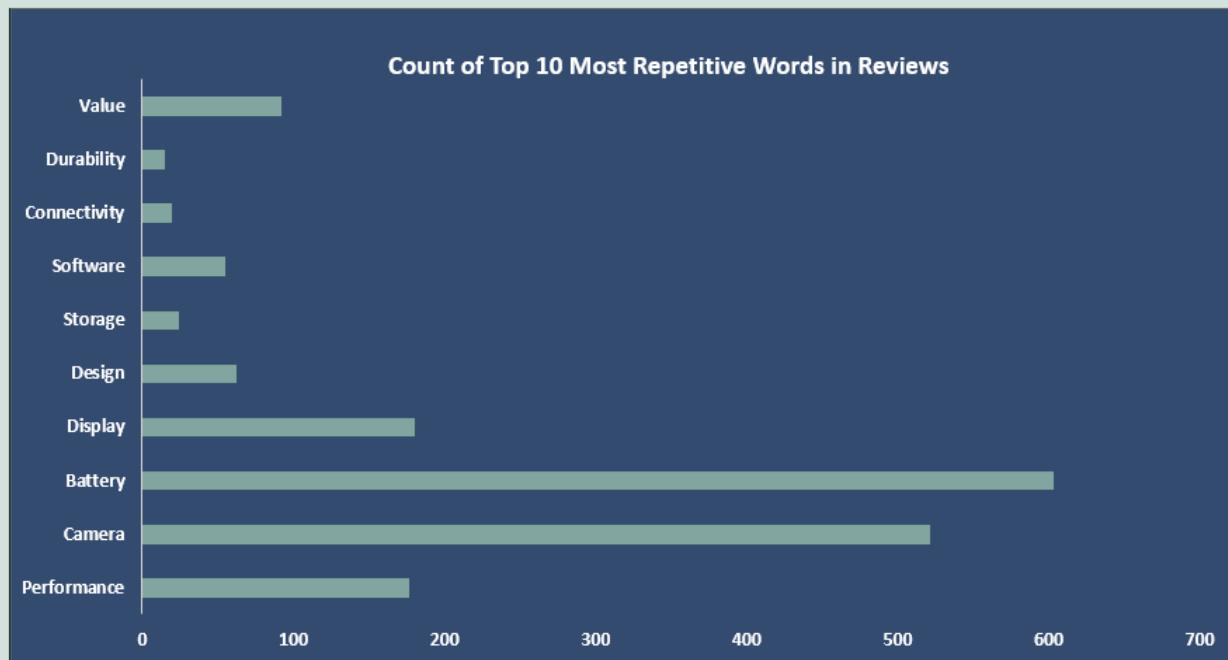
The distribution of products around the world reveals that India ranks first



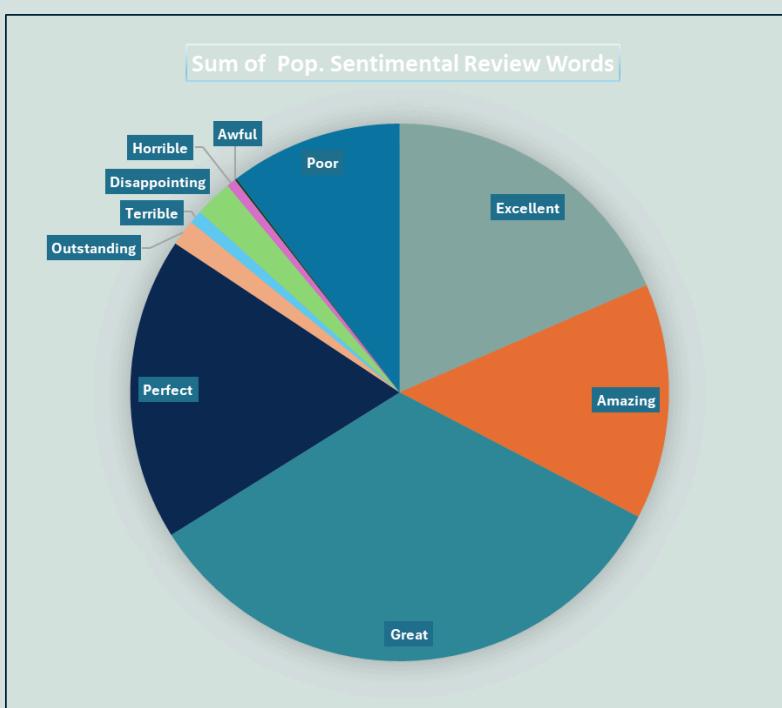
This chart reveals that India has the highest number of both verified and non-verified products.

# Data Analysis

## Text Analysis:



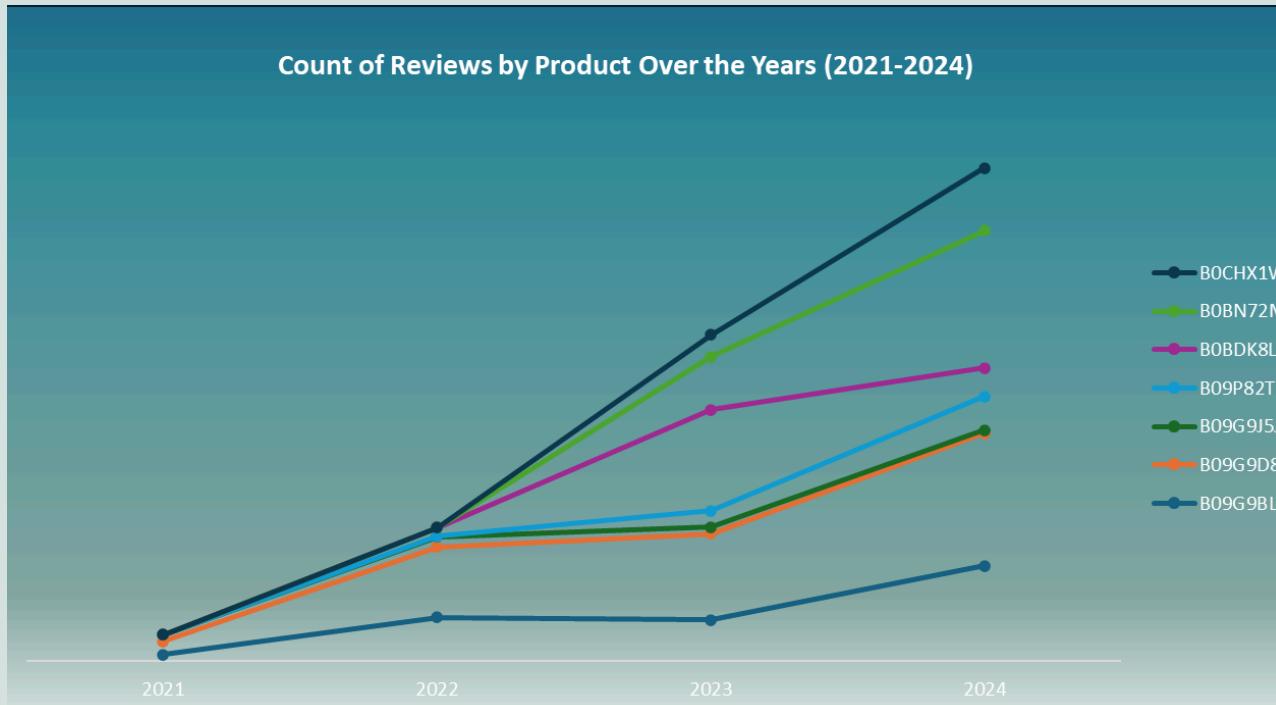
The analysis of the 10 most common words used in customer reviews reveals that customers prioritize battery life and camera quality when making their choices



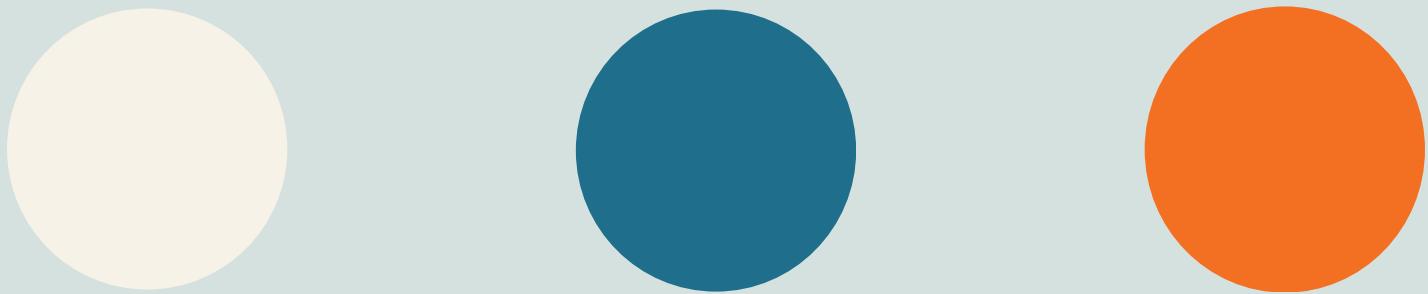
*Sentiment analysis reveals that customer reviews show a positive indication, as positive words outnumber negative ones.*

# Data Analysis

## Temporal Analysis:



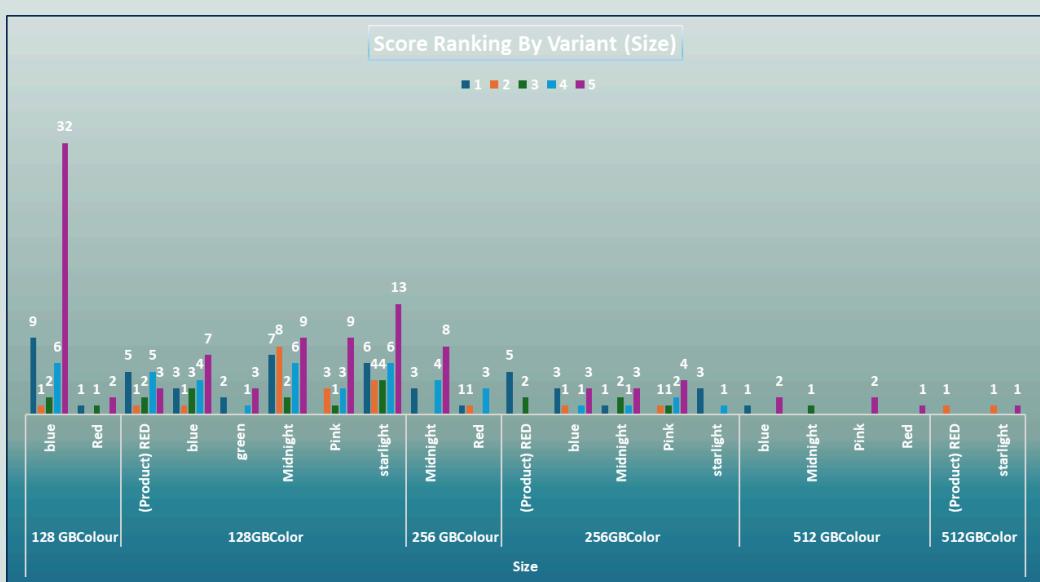
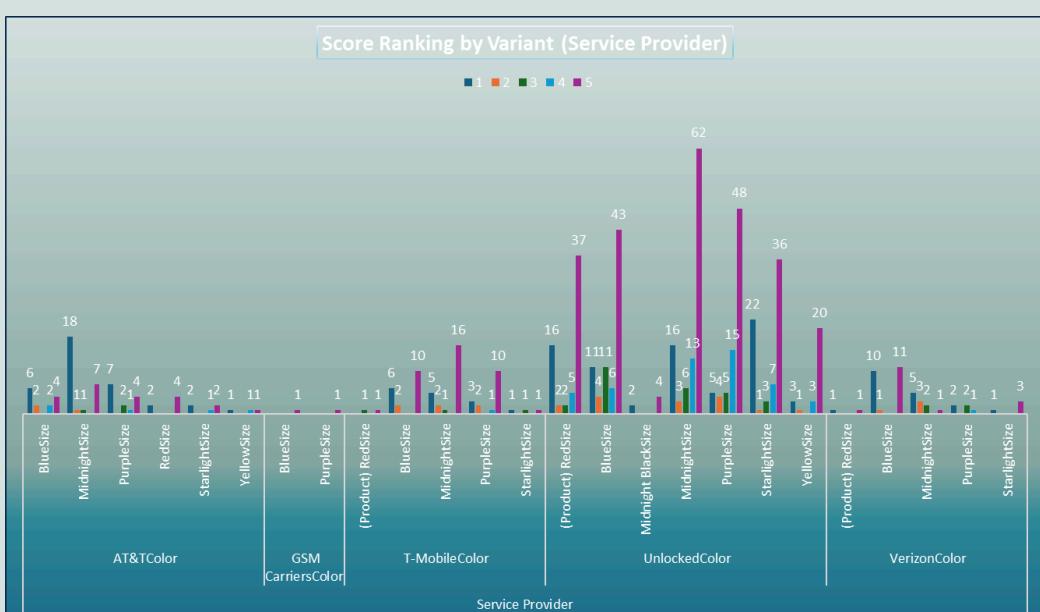
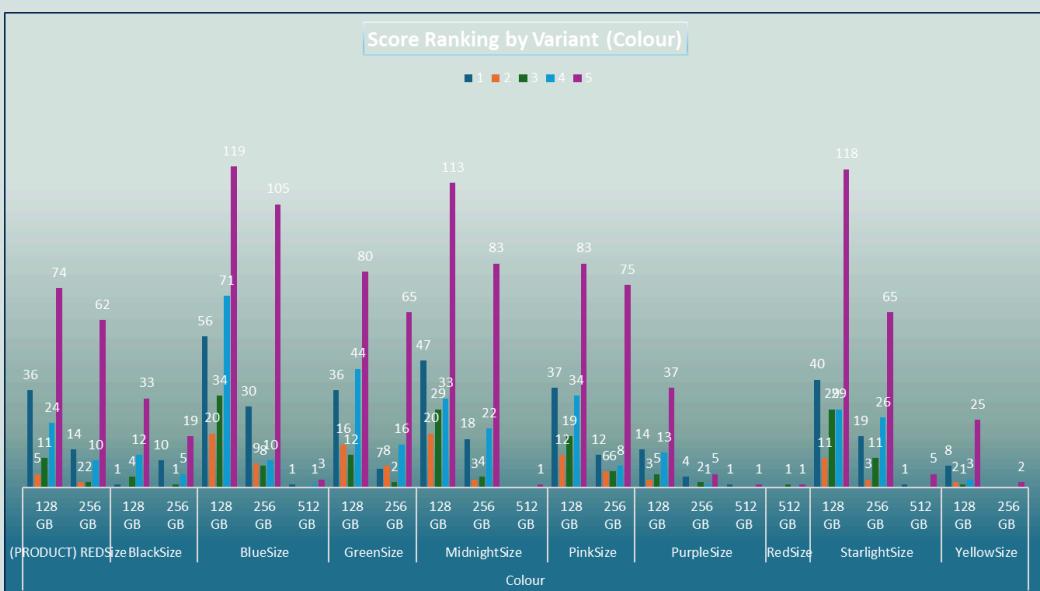
This chart reveals that the product model (BOCHX1W1XY) has consistently ranked at the top from 2021 to 2024



## Comparative Analysis:



This chart reveals that a rating score of 5 dominates across all products, except for B09G9J5JZX and B09P82T3PZ, which perform poorly in other rating categories.

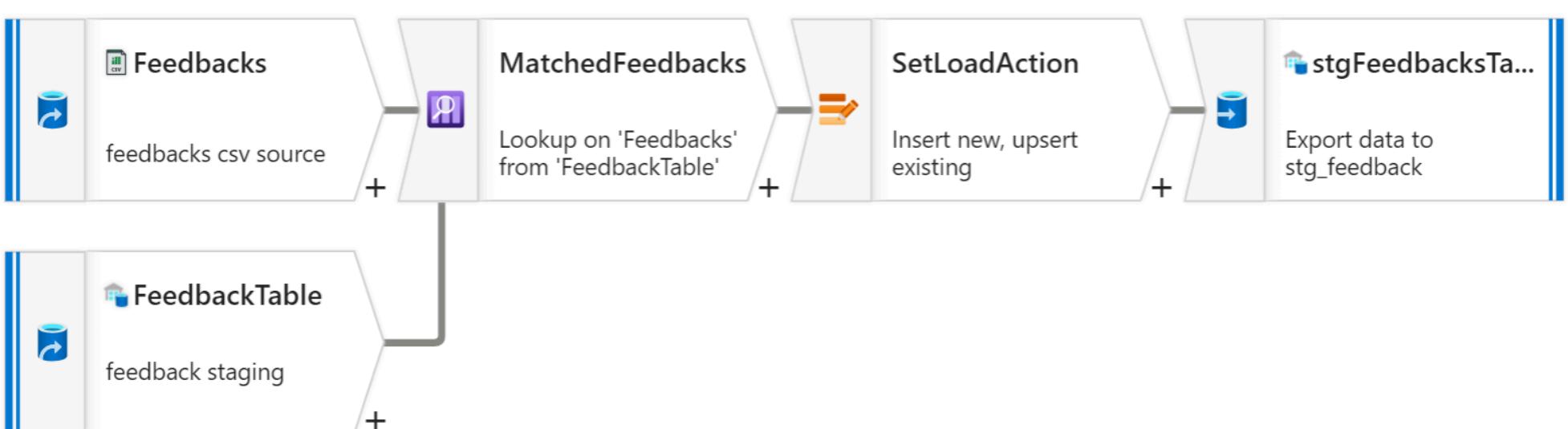


The following three charts analyze the variants specified in the dataset (color, service provider, size) based on their rating scores and reveal that the 'Blue 128 GB - Midnight' variant ranks first.

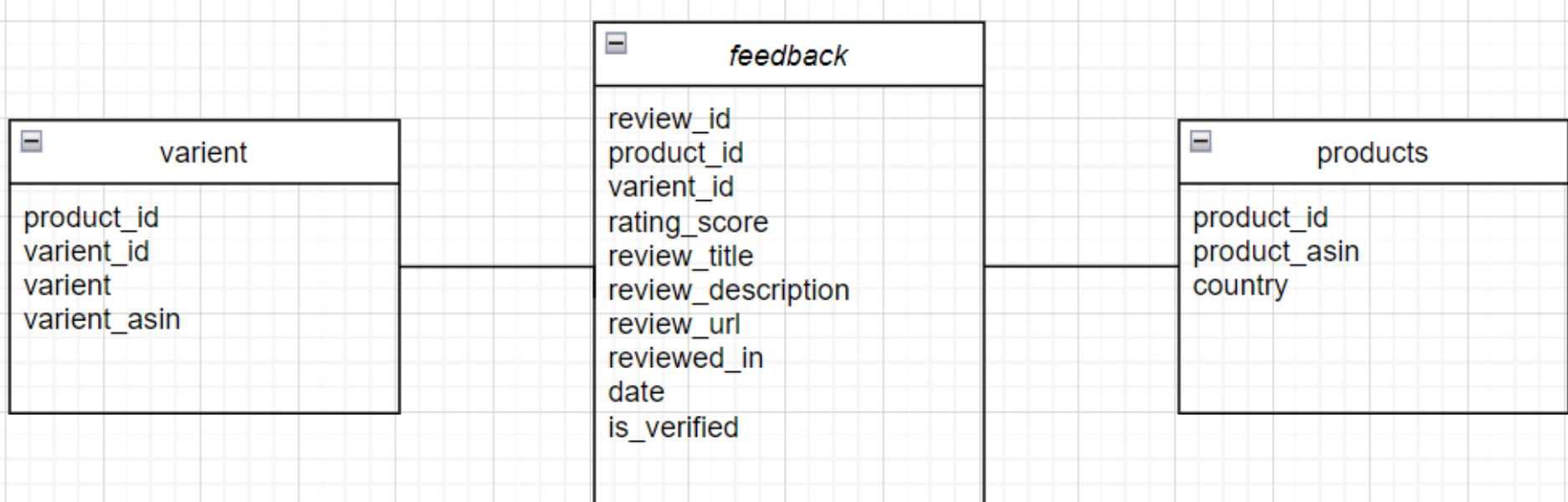
# Insights

- Encourage customers to make verified purchases by offering incentives such as discounts, loyalty points, or exclusive deals. Verified purchases could also be highlighted more prominently in the product listings to build trust and increase sales.
- Tailor marketing strategies and product offerings to different regions. For Europe, investigate the lower ratings by collecting more specific feedback and addressing any regional concerns, such as delivery times or product expectations.
- Highlight the features that customers consistently praise (e.g., "Great," "Excellent," and "Amazing") in marketing materials. At the same time, address the issues frequently mentioned in negative reviews (e.g., "Poor" or "Disappointing") by improving product quality or providing clearer expectations.
- Focus marketing efforts on the most popular variants (e.g., color or size) and consider expanding inventory of those variants. Additionally, use the high-performing variants as benchmarks to improve the lower-rated ones.
- Maintain or improve customer service and product quality to sustain the positive trend in review scores. Regularly update customers on product improvements and engage with those who leave feedback to show that their input is valued.

# Data warehouse, model, and data pipeline



- We loaded our feedback data into azure data lake storage to be our source data that will load into our data warehouse.
- We created a scheduled data pipeline to extract the data from the data lake storage and load it into a staging table so the data can be transformed later on based on our data warehouse model.
- The data pipeline is configured to load the data if it's new and update it if the data already exists.

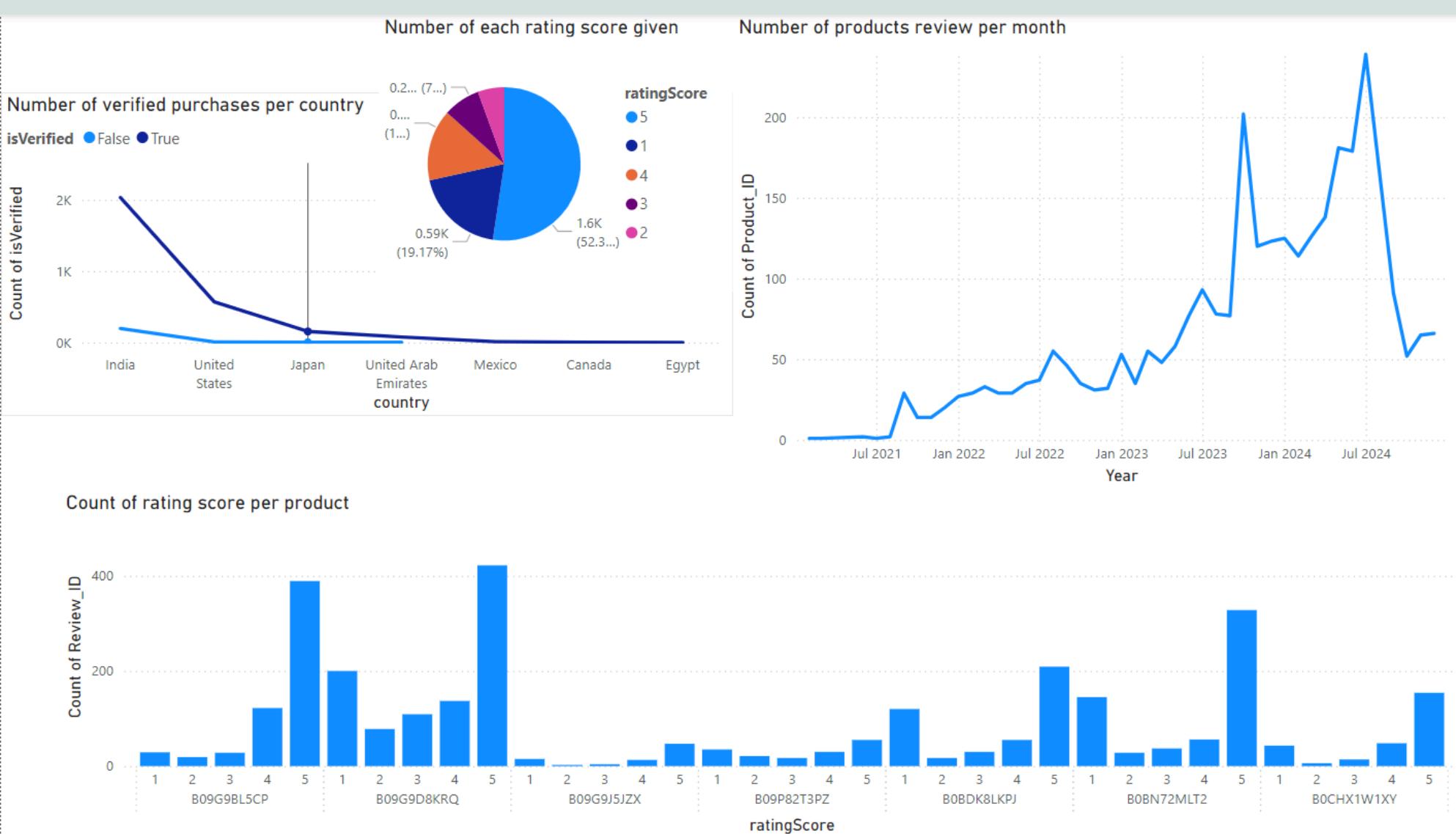


We created our data warehouse model so the feedback table is our fact table and products and varient tables are both dimension tables.

After executing the data pipeline our data warehouse would have the following tables:

- Staging table that the source data got loaded into initially.
- From that staging table, we follow our data warehouse model and start by creating the two dimensions tables then create our main fact table.

*After loading the data into our data warehouse we created a power-bi dashboard to give us some insights into our data*

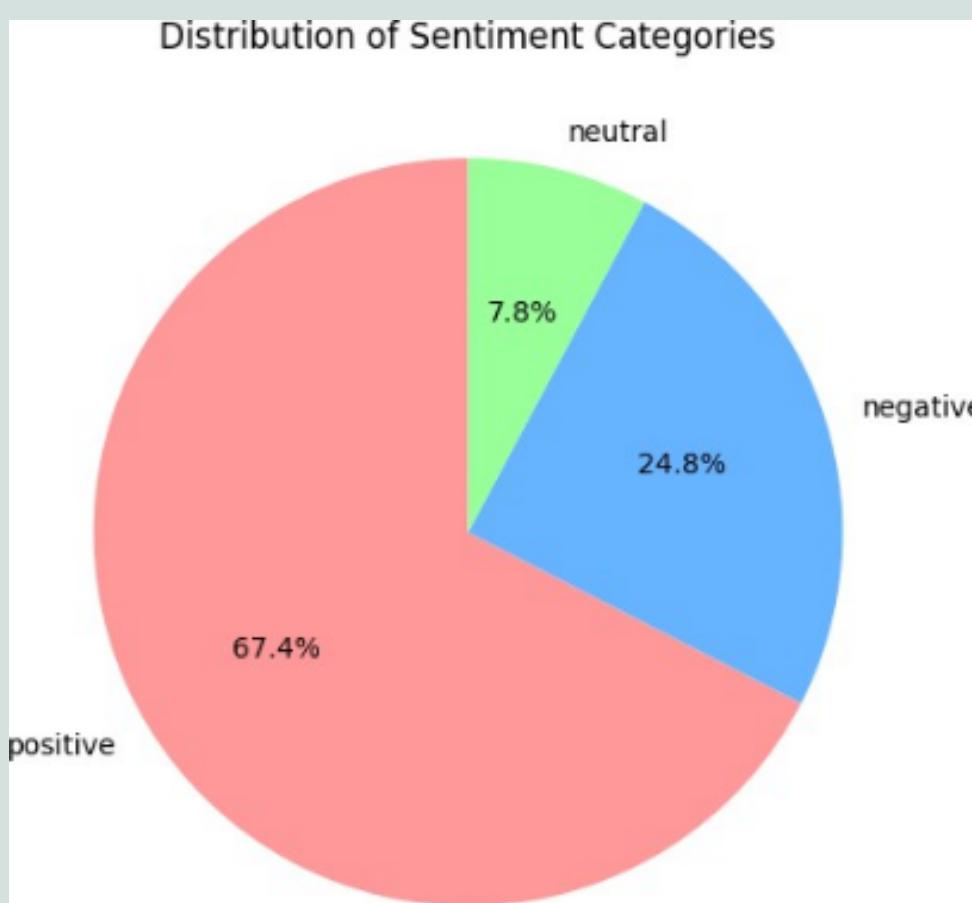


# Sentiment Analysis

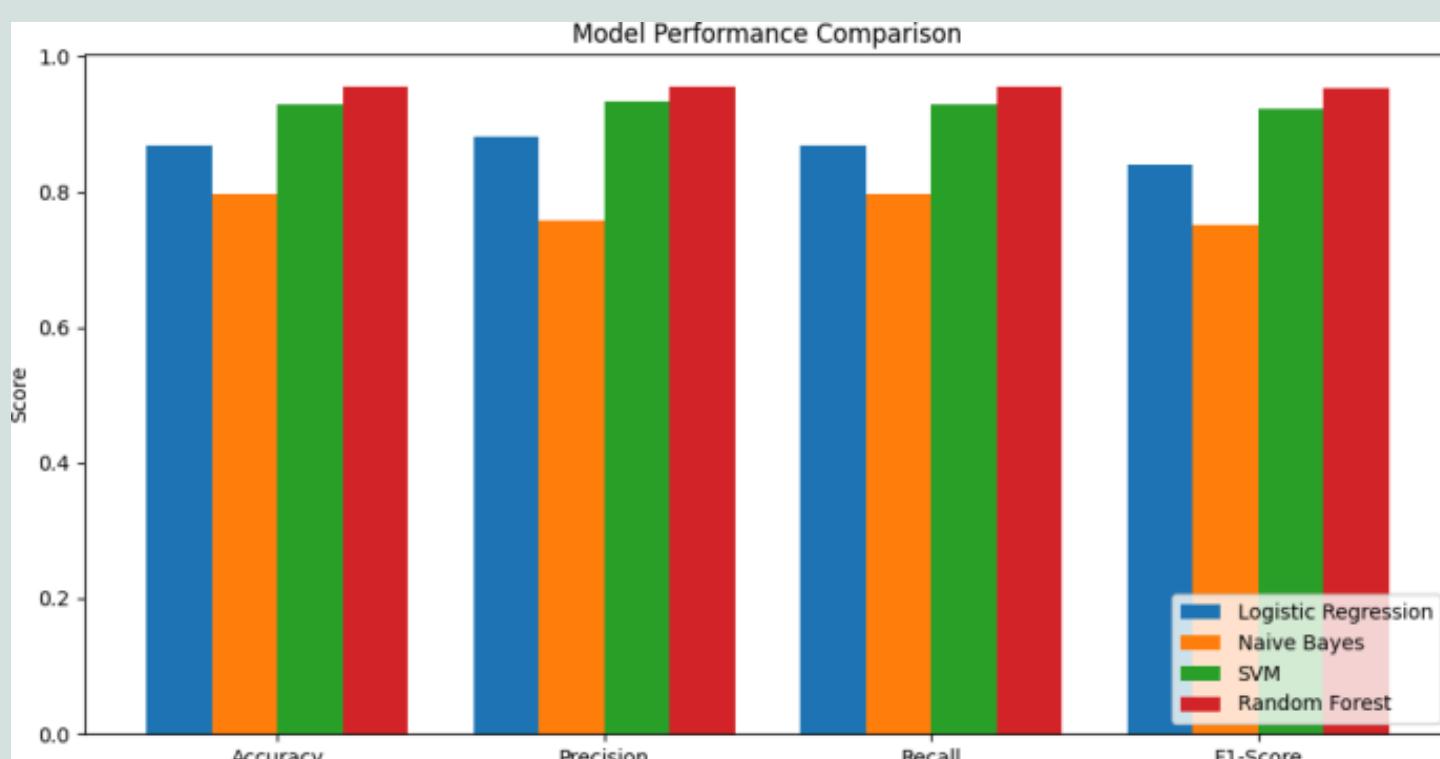
*This report presents the results of our sentiment analysis project on customer feedback. We analyzed X customer reviews and classified them into positive, neutral, and negative sentiments using various machine learning models.*

**Total number of feedback entries: 3062**

**Distribution of sentiment categories:**



*We used the following machine learning models for sentiment analysis: Logistic Regression, Naive Bayes, Support Vector Machine (SVM), and Random Forest. The text data was preprocessed and vectorized using TF-IDF.*

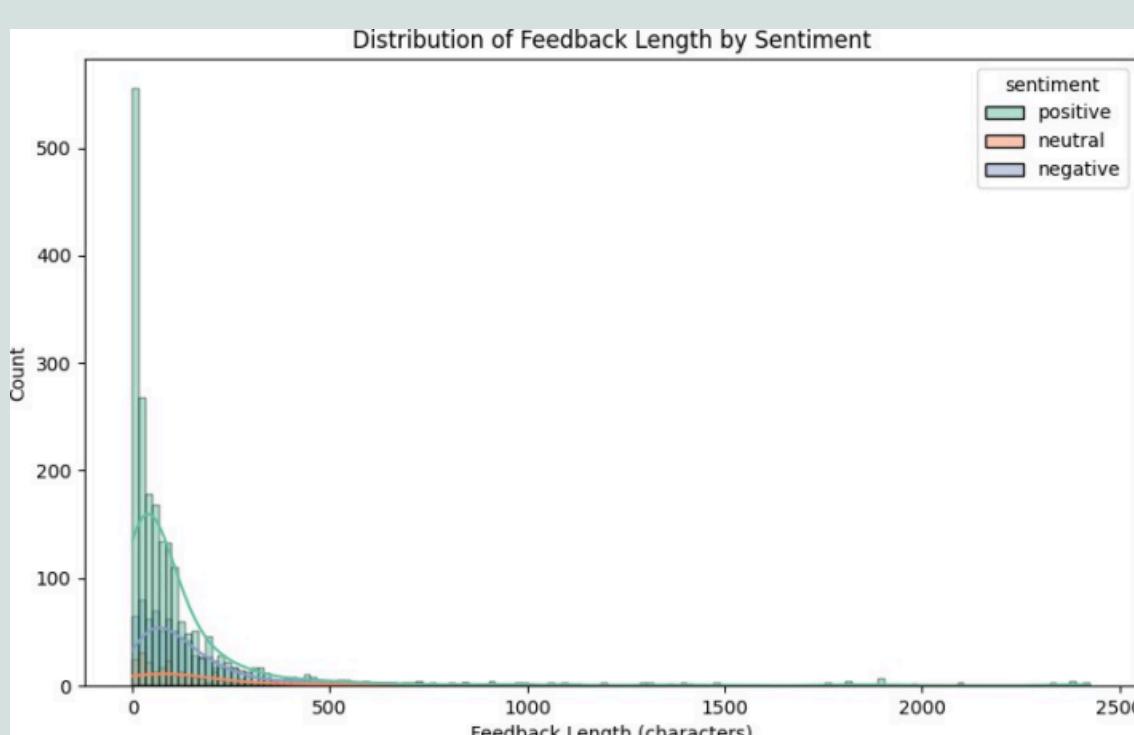


# *Model Performance Metrics*

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8681	0.8804	0.8681	0.8402
Naive Bayes	0.7975	0.7571	0.7975	0.7509
SVM	0.9288	0.9331	0.9288	0.9229
Random Forest	0.9543	0.9552	0.9543	0.9533

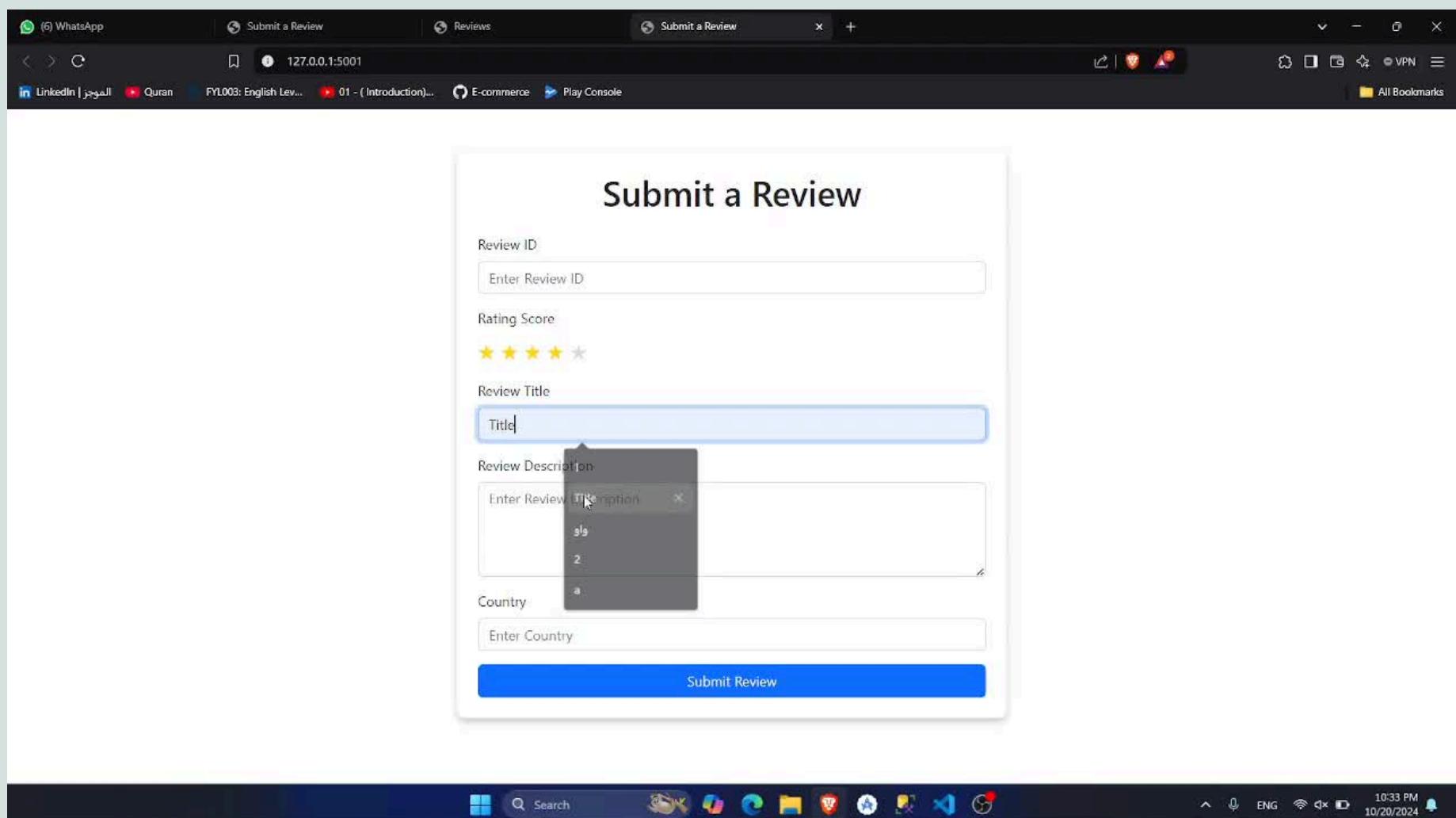
## ***Key Insights:***

- *The majority of customer feedback is positive.*
  - *Negative feedback often mentions specific product features.*
  - *Neutral feedback tends to be more factual and less emotional.*



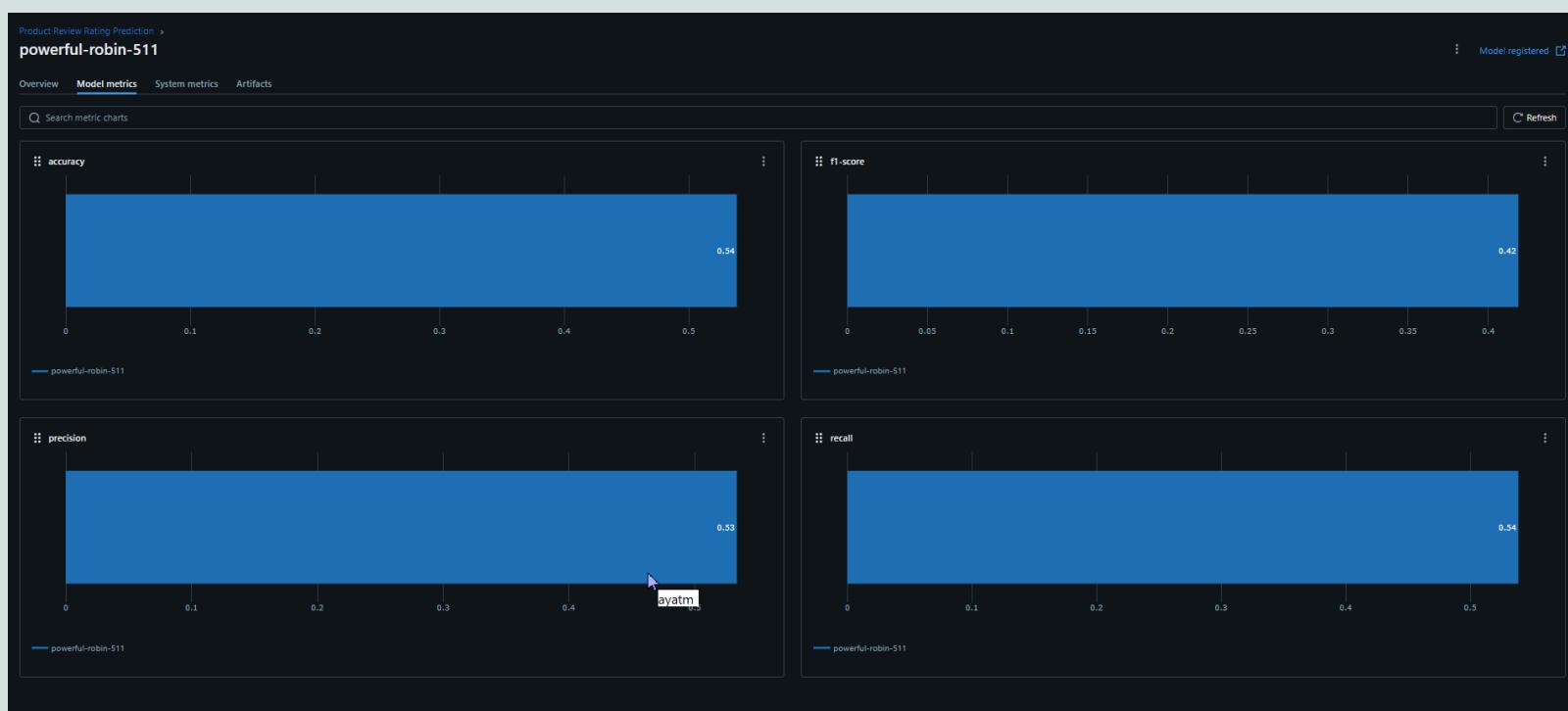
# Deployment

**Deploy a sentiment analysis model using a Flask web application to collect product reviews. The reviews are stored both in a database and a dataframe for efficient storage and further analysis**



# MLOps

**Track and manage sentiment analysis models using MLflow by creating the ML model**



 iPhone



THANK  
**YOU**