Customer Sentiment and Trends Analysis for Flipkart: Uncovering Key Insights to Enhance Customer Experience and Drive Growth

Author: Prateek Sharma

Data: 15-09-2024

Company Overview: **Flipkart**, founded in 2007, is a leading Indian e-commerce platform based in Bengaluru, known for its vast product range, fast deliveries, and innovative services like cash-on-delivery. With over 300 million users, customer feedback is crucial for Flipkart's growth. By leveraging customer reviews and sentiment analysis, Flipkart can enhance its offerings and maintain a strong competitive edge in India's rapidly growing e-commerce market.

Objective: The objective of this project is to analyze customer sentiment and product review trends for Flipkart, one of India's leading e-commerce platforms. By examining customer reviews, ratings, and sentiment data, the aim is to uncover key insights into customer preferences, product performance, and factors influencing customer satisfaction. These insights will help enhance the customer experience, improve product offerings, and drive strategic growth initiatives for Flipkart.

Data Source: For this project, Flipkart's review and sentiment data is collected from Kaggle open data source accessible to everyone. This dataset contains product information and customers review, summary of review and sentiment, providing the foundation for the analysis of customer sentiment and trends.

Data Preparation:

1. Data Collection and Organization –

   After downloading the dataset from Kaggle, I organized the files systematically on my laptop by creating a structured hierarchy of folders and subfolders.

2. Data Import and Initial Organization –

   The dataset was imported into Google Sheets for initial organization and examination.

3. Data Sorting and Review –

   Data was sorted by price and ratings in both ascending and descending orders to identify and address any anomalies.

   Reviews were sorted to identify and manage entries with incorrect or missing values.

4. Handling Missing Values –

   Rows with missing or unspecified review data were identified and removed, as this data is critical for analyzing customer sentiment.

5. Final Checks and Data Typing –

   A pivot table was used to ensure that the data is accurate and complete.

   Data types were assigned appropriately:

   - Reviews: Converted to numerical values.
   - Product Price: Formatted in Indian Rupees.
   - Summary and Product Name: Kept as text

Data Cleaning:

1. Remove Duplicates –

   A total of **34,319 duplicates** were identified and removed from the dataset. Post-cleaning, **170,279 rows** remain for analysis.

2. Trim White Spaces –

White spaces were removed from approximately **28,952 cells** to ensure uniform data formatting and improve the accuracy of subsequent analyses.

3. Convert Sentiment Values –

   To facilitate easier data analysis, sentiment values were converted to numerical format. The conversion was carried out as follows:

   - **Neutral** was replaced with 0
   - **Positive** was replaced with 1
   - **Negative** was replaced with -1

   This conversion was performed by selecting the sentiment column and utilizing the **Find and Replace** function to standardize sentiment values.

4. Additional Data Validation –

   Ensured that all other data columns adhered to their respective data types and formats, and conducted a final review to verify the accuracy and consistency of the dataset.

5. Backup and Save –
   Created backup copies of both the original and cleaned datasets.
   Ensured data integrity and availability for future reference or additional analysis.

Data Analysis:

The cleaned dataset was uploaded into **RStudio** for detailed analysis using the **R programming language**. The data was loaded into the environment pane with the following code:

R code ```

```
# Data loaded in environment pane

data <- read.csv(file.choose(), header = TRUE)

# View data in Source code editor pane

View(data)
```

Summary(data)

```

The data analysis commenced with several statistical and graphical techniques to extract meaningful insights from the dataset. The analysis was performed using R libraries such as **ggplot2**, **dplyr**, **tidyr**, and **tm**.

Correlation Analysis -

We began the analysis by exploring the relationships between key variables. The following correlations were computed:

- **Correlation between Product Price and Rating**: The correlation value was **0.074**, indicating a very weak positive relationship. This suggests that price and rating are not strongly related in this dataset.
- **Correlation between Rating and Sentiment**: A strong positive correlation of **0.803** was observed, indicating a strong relationship between customer ratings and their expressed sentiment. Higher ratings tend to be associated with more positive sentiments.
- **Correlation between Product Price and Sentiment**: A weak positive correlation of **0.049** was found between product price and sentiment. This suggests that price does not play a significant role in determining the sentiment expressed by customers.

The following code was used to calculate these correlations:

R code ```

# Correlation between Product Price and Rating

cor_product_price_rate <- cor(data$product_price, data$Rate, use = "complete.obs")

print(paste("Correlation between Product Price and Rating: ", cor_product_price_rate))

# Correlation between Rating and Sentiment

cor_rate_sentiment <- cor(data$Rate, data$Sentiment, use = "complete.obs")
```

print(paste("Correlation between Rating and Sentiment: ", cor_rate_sentiment))

Correlation between Product Price and Sentiment

cor_product_price_sentiment <- cor(data$product_price, data$Sentiment, use = "complete.obs")

print(paste("Correlation between Product Price and Sentiment: ", cor_product_price_sentiment))

```

Visualization and Data Exploration -

To further explore the dataset, various visualizations were generated to provide insights into the distribution and relationships of key variables:

1. **Histogram_of_Product_Prices**
   A histogram was created to visualize the distribution of product prices in the dataset.

   R code ```

   ggplot(data, aes(x = product_price)) +

    geom_histogram(binwidth = 500, fill = "blue", color = "black") +

    labs(title = "Distribution of Product Prices", x = "Product Price", y = "Frequency") +

    theme_minimal()

   ```

2. **Bar_Plot_of_Ratings**
   A bar plot was generated to display the distribution of product ratings.

   R code ```

   ggplot(data, aes(x = factor(Rate))) +

    geom_bar(fill = "green", color = "black") +

    labs(title = "Distribution of Ratings", x = "Rating", y = "Count") +

theme_minimal()

3. **Pie_Chart_of_Sentiments**
   To visualize the distribution of sentiments, a pie chart was plotted:

   R code ```

   sentiment_counts <- table(data$Sentiment)

   sentiment_df <- as.data.frame(sentiment_counts)

   ggplot(sentiment_df, aes(x = "", y = Freq, fill = factor(Var1))) +

   geom_bar(width = 1, stat = "identity") +

   coord_polar(theta = "y") +

   labs(title = "Distribution of Sentiments", fill = "Sentiment") +

   theme_minimal()

   ```

4. **Scatter_Plot_of_Price_vs_Rating**
   A scatter plot was used to investigate the relationship between product price and rating.

   R code ```

   ggplot(data, aes(x = product_price, y = Rate)) +

   geom_point(alpha = 0.5) +

   geom_smooth(method = "lm", se = FALSE, color = "red") +

   labs(title = "Price vs. Rating", x = "Product Price", y = "Rating") +

   theme_minimal()

   ```

5. **Box_Plot_of_Product_Price_by_Sentiment**
   A box plot was created to visualize how product prices are distributed across different sentiment categories.

R code ```

ggplot(data, aes(x = factor(Sentiment), y = product_price, fill = factor(Sentiment))) +

geom_boxplot() +

labs(title = "Product Price by Sentiment", x = "Sentiment", y = "Product Price") +

theme_minimal()

```

Word Cloud for Review Analysis -

To gain insight into the textual data, a word cloud was generated from customer reviews. The **tm** and **wordcloud** packages were utilized for text processing and visualization.

R code ```

# Create a text corpus from reviews

review_corpus <- Corpus(VectorSource(data$Review))

review_corpus <- tm_map(review_corpus, content_transformer(tolower))

review_corpus <- tm_map(review_corpus, removePunctuation)

review_corpus <- tm_map(review_corpus, removeWords, stopwords("en"))

review_corpus <- tm_map(review_corpus, stripWhitespace)

# Generate word cloud

wordcloud(content(review_corpus), min.freq = 100, random.order = FALSE, colors = brewer.pal(8, "Dark2"))

```

Visualization Summary and Key Insights:

1. Distribution of Product Prices - The majority of products are priced below ₹10,000, with a sharp drop in frequency as the price increases. There are very

few products in higher price ranges, indicating that most of Flipkart's products are affordable and likely cater to budget-conscious customers.

2. Distribution of Ratings - Most customers have given high ratings, with the majority being 5-star ratings. This suggests overall customer satisfaction, although there is still a significant number of lower ratings (1-star to 3-star), indicating room for improvement in product quality or customer service.

3. Word Cloud of Sentiments - Frequently mentioned words in customer reviews include positive terms like "product," "good," "nice," "awesome," and "recommended." The word cloud suggests a positive customer sentiment toward Flipkart products, though negative terms such as "worthless" and "poor" are also visible, highlighting some dissatisfaction.

4. Distribution of Sentiments - The pie chart shows that positive sentiment (blue) dominates customer reviews, with a smaller portion reflecting negative (red) and neutral (green) sentiments. This aligns with the distribution of ratings, where positive feedback is more frequent.

5. Price vs. Rating - There appears to be a slight positive correlation between product price and rating, as indicated by the red trend line. Higher-priced products tend to receive higher ratings, suggesting that customers perceive more expensive products to be of better quality or value.

6. Product Price by Sentiment - The box plot shows that the product price distribution is similar across sentiment categories. However, products with negative sentiments (red) have a wider price range, possibly indicating dissatisfaction with both cheaper and more expensive products. Positive sentiments (blue) tend to be associated with a narrower range of product prices.

Key Takeaways:

1. Positive Customer Sentiment: Most customers leave positive feedback, with many high ratings and positive sentiment reflected in reviews.
2. Affordable Pricing: The majority of products are in the lower price range, which may appeal to budget-conscious customers. However, even higher-priced products tend to have favorable ratings.
3. Opportunity for Improvement: There is a subset of customers who provide negative feedback, often associated with both low- and high-priced

products, indicating the need for better quality control or customer service enhancements for these segments.

TOP 3 Recommendations:

1. Enhance Product Quality for High-Priced Items

- Reasoning: Although higher-priced products tend to have better ratings, there is still a considerable spread of negative sentiment for some expensive items. Flipkart should focus on improving the quality and customer satisfaction for premium products to minimize dissatisfaction among high-spending customers.
- Action: Conduct product audits or gather detailed customer feedback on high-priced items with negative reviews, and work with suppliers to ensure better quality control.

2. Focus on Customer Service for Low-Rated Products

- Reasoning: A noticeable number of customers give 1-star to 3-star ratings, indicating dissatisfaction with specific products or services. Addressing issues related to these lower-rated products can help improve overall customer satisfaction.
- Action: Implement targeted customer service outreach for products with low ratings, offering refunds, replacements, or troubleshooting assistance. Additionally, analyze these products to identify common complaints and address underlying issues.

3. Leverage Positive Reviews in Marketing Campaigns

- Reasoning: With a high volume of positive ratings and reviews, Flipkart has an opportunity to amplify this customer satisfaction through marketing. Showcasing high customer satisfaction can increase trust and attract new customers.
- Action: Use customer testimonials and high-rated products in advertisements, social media campaigns, and on the website to highlight positive user experiences. Featuring the top-rated products prominently can also encourage purchases from potential customers.