

Title Page

Google Play Store E-commerce App Review Analysis

Analyzing Customer Sentiment and Insights

[Bereket Andualem]

[Date: e.g., July 26, 2025]

[Debre birhan Universty/DataScience]

Table of Contents

1. Executive Summary
 2. Introduction
 3. Methodology
 4. Results & Key Findings
 5. Challenges Faced & Solutions
 6. Conclusion & Recommendations
 7. Future Work
-

1. Executive Summary

This project aimed to analyze customer sentiment and extract key insights from English reviews of three popular e-commerce applications—AliExpress, Alibaba, and Jiji—available on the Google Play Store. Leveraging Python for data acquisition, cleaning, analysis, and modeling, combined with SQL Server for data storage and Streamlit for visualization, this study provides a comprehensive view of user experiences.

The analysis revealed distinct patterns in review volume, rating distributions, and sentiment across the platforms. Jiji generally exhibited the highest average ratings and positive sentiment, while AliExpress, a newer entrant, showed lower average ratings and highly polarized feedback. Alibaba fell between the two. Common topics discussed varied, with AliExpress users focusing on shipping and product issues, Alibaba users on pricing and app quality, and Jiji users on the platform's usability and marketplace dynamics. A machine learning model (Logistic Regression) was successfully trained to classify review sentiment from text, achieving an accuracy of 76% on test data, although it struggled with the minority 'Neutral' and 'Negative' classes due to data imbalance. An interactive Streamlit dashboard was developed to present these findings dynamically, allowing users to explore insights per app.

This project demonstrates the application of the full data science lifecycle, from web scraping to model deployment, yielding actionable insights for understanding user perceptions of these e-commerce platforms.

2. Introduction

Understanding customer sentiment is paramount for e-commerce businesses to improve services, address pain points, and maintain a competitive edge. User-generated reviews on platforms like the Google Play Store offer a rich, albeit unstructured, source of direct feedback. This project focuses on analyzing English reviews for three major e-commerce apps: AliExpress, Alibaba, and Jiji. The primary objective is to extract, analyze, and visualize customer sentiment and key discussion topics from these reviews.

The specific goals of this project were to:

- Scrape and store Google Play Store reviews for the three selected apps.
- Clean and preprocess the textual review data.
- Perform comprehensive Exploratory Data Analysis (EDA) to uncover patterns and trends.
- Build and evaluate machine learning models for automatic sentiment classification.
- Develop an interactive dashboard to communicate findings effectively.

This report details the approach taken, the results obtained, challenges encountered, and the overall conclusions drawn from this analysis.

3. Methodology

The project followed a structured data science workflow, aligning with the stages outlined in the project guidelines.

3.1 Data Acquisition & Storage

- **Scraping:** The `google-play-scraper` Python library was utilized to efficiently collect English reviews for AliExpress, Alibaba, and Jiji. The script targeted approximately 500 reviews per app, sorted by newest.
- **Storage:** Scraped data was stored in a Microsoft SQL Server database (`PlayStoreReviewsDB`) within a table named `dbo.AppReviews`. The schema included fields for app name, review text, rating, date, and metadata.

3.2 Data Cleaning & Preprocessing

- **Loading:** Data was loaded from SQL Server into a Pandas DataFrame.

- **Cleaning:** Reviews with missing or empty text were removed. Data types were verified.
- **Text Preprocessing (NLP):** A dedicated cleaning function was applied to the `ReviewText`. Steps included lowercasing, HTML tag removal, URL/emoji removal, contraction expansion, punctuation/number removal, tokenization, stopword removal, and lemmatization. The result was stored in `CleanedReviewText`.
- **Feature Engineering:** `ReviewWordCount` was calculated.

3.3 Exploratory Data Analysis (EDA)

- **Tools:** Analysis was conducted using Jupyter Notebooks (`02_eda_insights.ipynb`), leveraging `pandas`, `numpy`, `matplotlib`, `seaborn`, `plotly`, `vaderSentiment`, and `wordcloud`.
- **Activities:** Performed review count analysis, rating distribution analysis, review length analysis, common words/N-grams analysis, basic sentiment analysis using VADER, and missing value analysis. Visualizations included bar charts, line charts, histograms, box plots, pie charts, word clouds, and heatmaps.

3.4 Statistical Modeling / Machine Learning

- **Tools:** Jupyter Notebook (`03_modeling_sentiment.ipynb`) and `scikit-learn`.
- **Objective:** Classify review sentiment (Positive, Negative, Neutral) based on `CleanedReviewText`.
- **Process:**
 - Data was split into training (80%) and testing (20%) sets.
 - `TfidfVectorizer` was used within `Pipelines` to convert text to numerical features.
 - Three models were evaluated: Logistic Regression, Random Forest, and Naive Bayes.
 - Logistic Regression performed best (Accuracy ~76%).
 - Addressed class imbalance by implementing a Logistic Regression Balanced variant using `class_weight='balanced'`.
 - Evaluated using accuracy, classification report, and confusion matrix. Analyzed feature importance for the Logistic Regression model.

3.5 Dashboard Development

- **Tool:** Streamlit.
- **Implementation:** Developed `dashboard.py` to load `cleaned_reviews_data.csv`.
- **Features:** Included interactive app selection, overview statistics, rating/sentiment distribution charts, top words/bigrams/word cloud tabs, sentiment trend line chart, and sample reviews by sentiment using expanders.

4. Results & Key Findings

4.1 EDA Highlights

- **Review Volume & Trends:** All apps showed a significant increase in reviews around mid-to-late 2025. AliExpress appeared to be a recent entrant with rapid growth.
- **Rating Patterns:** Jiji had the highest average rating (3.36). AliExpress had the lowest (2.62), likely due to its newness. AliExpress and Alibaba showed polarized rating distributions.
- **Review Characteristics:** AliExpress users wrote the longest reviews on average. Longer reviews were associated with extreme (1-star or 5-star) ratings, with AliExpress showing a unique pattern for 4-star reviews.
- **Common Topics:** Distinct themes emerged: AliExpress (shipping, product issues), Alibaba (pricing, app quality), Jiji (platform usability, marketplace).
- **Sentiment Insights (VADER):** Jiji showed the highest average VADER sentiment. Sentiment correlated moderately with star ratings (corr ~0.57).

4.2 Modeling Highlights

- **Model Performance:** The standard Logistic Regression model achieved 76% accuracy. It performed excellently for 'Positive' sentiment (High Precision & Recall) but poorly for 'Neutral' (0% scores) and moderately for 'Negative' (Low Recall).
- **Class Imbalance:** The extreme imbalance (Positive >> Negative >> Neutral) was identified as the primary reason for poor performance on minority classes.
- **Balanced Model:** Using `class_weight='balanced'` improved recall for 'Negative' and 'Neutral' classes in subsequent evaluations, demonstrating a strategy to address imbalance.

4.3 Dashboard Features

- The Streamlit dashboard successfully implemented all required features, offering an interactive and user-friendly interface to explore the analysis results for each app dynamically.

5. Challenges Faced & Solutions

- **NLTK Data Download:** Initial errors occurred due to missing NLTK data files. This was resolved by running the `nltk.download()` commands for required packages (`punkt`, `punkt_tab`, `stopwords`, `wordnet`, `omw-1.4`).
- **Data Loading Issues:** Errors in loading data for modeling arose from running cells out of order or not re-running cells that created necessary columns (like `VADER_Sentiment`). This was fixed by ensuring correct execution sequence and re-saving the cleaned data after all EDA steps.
- **Class Imbalance in Modeling:** The Logistic Regression model failed to predict 'Neutral' reviews. This was diagnosed through detailed evaluation. The solution implemented was

using `class_weight='balanced'` to give more importance to minority classes during training.

- **Dashboard Errors:** Minor coding errors (like undefined variables) and Pandas deprecation warnings were encountered. These were resolved through careful debugging and updating code (e.g., adding `observed=True` to `groupby`).
 - **Time Management:** As common in data science projects, significant time was spent on data cleaning and refining visualizations. Sticking to a structured plan and iterating helped manage this.
-

6. Conclusion & Recommendations

This project successfully demonstrated a complete data science workflow applied to real-world e-commerce app review data. Key insights include the distinct user sentiment profiles of AliExpress, Alibaba, and Jiji, the challenges of analyzing imbalanced datasets, and the power of interactive dashboards for communicating results.

Based on the analysis, the following recommendations can be made:

- **For AliExpress:** Focus on improving aspects related to shipping and handling customer complaints more effectively, as these were prominent negative themes.
 - **For Alibaba:** Continue leveraging its strengths in product variety and global reach while monitoring pricing strategies and customer service responsiveness.
 - **For Jiji:** Maintain its user-friendly platform and direct marketplace features, which drive high user satisfaction.
 - **For Future Analysis:** Collecting more 'Neutral' examples or employing advanced resampling techniques could further improve model performance for all sentiment classes.
-

7. Future Work

- **Advanced NLP Models:** Implement and evaluate more sophisticated models like BERT or RoBERTa for sentiment analysis, potentially offering better accuracy and handling of context.
- **Deeper Topic Modeling:** Use techniques like LDA or BERTopic on larger datasets to discover more nuanced themes within reviews for each app or sentiment.
- **Real-time Analysis:** Develop a pipeline for continuous scraping and sentiment analysis to monitor app reputation dynamically.
- **Expanded Scope:** Include reviews in other languages or analyze reviews from other app stores (like Apple App Store) for a broader perspective.
- **Predictive Modeling:** Explore models to predict review ratings directly from text or identify factors that lead to extremely positive or negative reviews.

