Business Report: Analysis of Google Play Store and Apple App Store Data

Executive Summary

This report presents an analysis of mobile application data from both the Google Play Store and Apple App Store. The analysis was conducted using Python with pandas for data manipulation and exploration. The dataset contains comprehensive information about apps across both platforms, including categories, ratings, prices, and other relevant metrics. The analysis focused on data cleaning, feature engineering, and preliminary exploration of app distributions across categories and genres.

Methodology

The analysis followed these key steps:

1. Data Loading: Imported datasets from both Google Play Store and Apple App Store

2. Data Cleaning:

- Handled missing values by imputing medians for ratings and filling other missing values with "Unknown"

- Removed duplicate entries

- Processed price information by removing currency symbols and converting to float

- Filtered out non-English apps

- Dropped rows with missing critical values

3. Data Exploration:

- Examined value counts for categories/genres in both platforms

- Compared the distribution of apps across different categories

- Analyzed price distributions

Key Findings

1. Data Composition

The datasets contain:

- Google Play Store: 10,841 apps across 33 categories

- Apple App Store: 7,197 apps across 23 genres

2. Category Distribution

Google Play Store Top Categories:

1. Family: 1,814 apps (16.7%)

2. Game: 1,055 apps (9.7%)

3. Tools: 812 apps (7.5%)

Apple App Store Top Genres:

1. Games: 3,165 apps (44.0%)

2. Entertainment: 402 apps (5.6%)

3. Education: 385 apps (5.3%)

3. Price Analysis

- Google Play Store apps predominantly free (most have $0 price)

- Apple App Store shows a mix of free and paid apps

- Price conversion was successfully implemented for Google Play Store data

4. Data Quality Issues Addressed

- Handled 1,474 missing ratings in Google Play Store data (13.6% of total)

- Removed non-English apps from both platforms

- Processed price information consistently

- Addressed missing values in critical columns

Business Insights

1. Market Composition Differences

The distribution of apps differs significantly between platforms:

- Google Play has a more balanced distribution across categories

- Apple App Store is heavily dominated by Games (44% of all apps)

Recommendation: Developers targeting Apple should consider the competitive game market, while Google Play offers more diverse opportunities.

2. Pricing Strategies

The prevalence of free apps on Google Play suggests:

- Freemium models dominate on Android

- Potential for monetization through ads and in-app purchases

- Apple users may be more willing to pay upfront for apps

Recommendation: Consider platform-specific monetization strategies when developing cross-platform apps.

3. Category Opportunities

The analysis reveals underserved categories:

- Google Play: "Beauty" and "Comics" have relatively few apps

- Apple App Store: "Catalogs" and "Medical" have low representation

Recommendation: Developers could explore these niche categories for less competitive opportunities.

Technical Challenges and Solutions

1. Data Cleaning:

- Addressed significant missing values in ratings and other fields

- Standardized price formats across currencies

- Handled special cases in price data (e.g., "Everyone" appearing in price column)

2. Data Integrity:

- Verified no duplicate entries in either dataset

- Ensured consistent data types across columns

3. Feature Engineering:

- Successfully converted price strings to numerical values

- Created clean category/genre distributions for analysis

Limitations

1. The analysis doesn't yet explore relationships between variables (e.g., ratings vs. price)

2. Temporal aspects (last update dates) haven't been analyzed

3. No sentiment analysis of reviews has been performed

4. Installation numbers are in string format and haven't been converted to numerical values

Next Steps

1. Deeper Analysis:

- Explore correlations between ratings, price, and other metrics

- Analyze temporal trends in app updates and releases

- Compare performance metrics between platforms

2. Advanced Modeling:

- Predict app ratings based on other features

- Cluster apps to identify successful patterns

3. Visualization:

- Create visual representations of category distributions

- Develop dashboards for interactive exploration

Conclusion

This analysis provides a solid foundation for understanding the mobile app ecosystem across both major platforms. The cleaned datasets are now prepared for more sophisticated analysis, and the initial findings reveal important differences in market composition between Google Play and Apple App Store. These insights can inform business decisions around app development, marketing strategies, and platform prioritization.

The technical work has addressed key data quality issues, ensuring reliable results for future analyses. Further investigation into the relationships between app characteristics and success metrics could yield valuable predictive insights for app developers and publishers.