

PROJECT REPORT

on

MOBILE APPS AND USER REVIEWS ANALYSIS

**Presented in partial fulfilment of the requirements for the completion
of the Data analysis and visualization course
project**

at



Submitted to

Dr Smita Agrawal

Submitted by

22BCE133
(Yasvi Kachhadiya)

22BCE122
(Jeenal Darji)

22BCE029
(Bansi Prajapati)

ABSTRACT

This project analyzes mobile app data and user reviews from the Google Play Store to derive insights into app performance, user satisfaction, and sentiment trends. Key objectives include understanding app categories, ratings, installs, and user sentiment. By leveraging advanced data preprocessing techniques, KPI derivation, and visualization, this analysis highlights significant trends and actionable insights, providing valuable knowledge for app developers, marketers, and analysts. The study employs Python and Tableau to preprocess, analyze, and visualize data, creating an engaging exploration of app usage and user engagement.

INTRODUCTION

Mobile applications are integral to the digital ecosystem, influencing various industries, from education to entertainment. Understanding user feedback and app performance is critical to app development and marketing strategies. This project explores Google Play Store app data to uncover trends in user engagement, sentiment analysis, and app performance.

The primary objectives are to:

1. Preprocess and clean app and review data.
2. Identify trends in user ratings, installs, and sentiment polarity.
3. Create KPIs that provide actionable insights.
4. Visualize results using Tableau dashboards.

This project combines structured app data with unstructured user reviews to provide a comprehensive analysis of user behavior and app success.

SCOPE

The analysis and visualization conducted in this project have applications in:

1. **App Development:** Insights into user feedback help refine app features.
2. **Marketing Strategy:** Understanding sentiment trends aids in targeted advertising.
3. **User Engagement:** Identifying highly rated categories informs app recommendations.
4. **Revenue Optimization:** Revenue-related KPIs guide pricing strategies.

Future work could include predictive modeling to forecast app success based on historical trends.

PROBLEM STATEMENT

Mobile apps' success is driven by user satisfaction, engagement, and feedback. Analyzing app ratings, installs, and reviews provides valuable insights into trends and patterns. This project aims to:

1. What are the most popular categories of apps based on the number of installs?
2. Is there a relationship between app ratings and the size of the app?
3. How do pricing and app type (free vs. paid) affect app ratings?
4. Are apps with higher installs or reviews more likely to have higher sentiment subjectivity?
5. What does the heatmap reveal about ratings and sentiments across app categories?
6. What insights are revealed by the pair plot for numerical features?
7. How do sentiment polarities vary across categories?
8. What are the trends in total installs across app categories?
9. How is sentiment polarity distributed among apps?
10. What is the proportion of apps across different categories?

OBJECTIVE

Trend Analysis: Understand how app ratings, installs, and reviews vary across categories.

Sentiment Insights: Analyse user sentiment and its relationship with app installs.

KPI Development: Create unique, advanced KPIs to evaluate app success.

Visualization: Use Tableau to present findings effectively.

DATASET DESCRIPTION

apps.csv

- Contains structured data about Google Play Store apps.
- Key Columns:
 - **App:** Name of the app.
 - **Category:** App category (e.g., "Education," "Game").
 - **Rating:** Average user rating (1-5 scale).
 - **Reviews:** Number of reviews.
 - **Installs:** Total downloads.
 - **Price:** Cost of the app (if paid).
 - **Content Rating:** Target audience (e.g., "Everyone," "Teen").
 - **Last Updated:** Last update date.

user_reviews.csv

- Contains user reviews for apps in apps.csv.
- Key Columns:
 - **Translated_Review:** Text of the user review.
 - **Sentiment:** Sentiment category ("Positive," "Negative," "Neutral").
 - **Sentiment_Polarity:** Polarity score (-1 to 1).
 - **Sentiment_Subjectivity:** Subjectivity score (0 to 1).

Attribute Name	Description	Data Type	Significance
App	Name of the mobile application	Categorical	Primary identifier for the app. Used to link datasets.
Category	App category or genre	Categorical	Helps in grouping apps for category-level analysis.
Rating	Average user rating (1-5)	Numerical (float)	Indicates user satisfaction. A higher rating reflects better app quality.
Reviews	Total number of reviews	Numerical (int)	Shows user engagement and app popularity.
Size	Size of the app (in MB or textual format)	Numerical (float)	Indicates the resource requirements for the app.
Installs	Number of downloads	Numerical (int)	Reflects app popularity and reach.
Type	Free or Paid app	Categorical	Differentiates between monetization models.
Price	Price of the app (if paid)	Numerical (float)	Provides insights into app pricing strategies.
Content Rating	Target audience of the app	Categorical	Indicates the appropriate audience for the app (e.g., Everyone, Teen).
Genres	Additional genres/subcategories of the app	Categorical	Provides more granularity within main categories.
Last Updated	Date of the last app update	Date	Reflects app maintenance and developer activity.
Current Ver	Current version of the app	Categorical	Provides version information to track updates.
Android Ver	Minimum Android version required	Categorical	Indicates device compatibility.
Translated_Review	Text of the user review	Text	Provides qualitative insights into user feedback.
Sentiment	Sentiment category of the review	Categorical	Helps in analyzing user sentiment (Positive, Negative, Neutral).
Sentiment_Polarity	Polarity score of the review sentiment	Numerical (float)	Indicates sentiment direction: -1 (negative) to +1 (positive).
Sentiment_Subjectivity	Subjectivity score of the review (0-1)	Numerical (float)	Shows whether the review is opinion-based or factual.

DATA CLEANING AND PREPROCESSING

Handling Missing Values:

- Dropped rows with missing critical information (Android Ver, Current Ver).

Text Cleaning:

- Removed non-alphanumeric characters and stopwords from Translated_Review.
- Applied stemming to reduce words to their root form.

Outlier Detection:

- Used IQR to detect and remove outliers in Reviews, Installs, and sentiment scores.

Normalization:

- Scaled Reviews, Installs, and sentiment scores to a uniform range using Min-Max scaling.

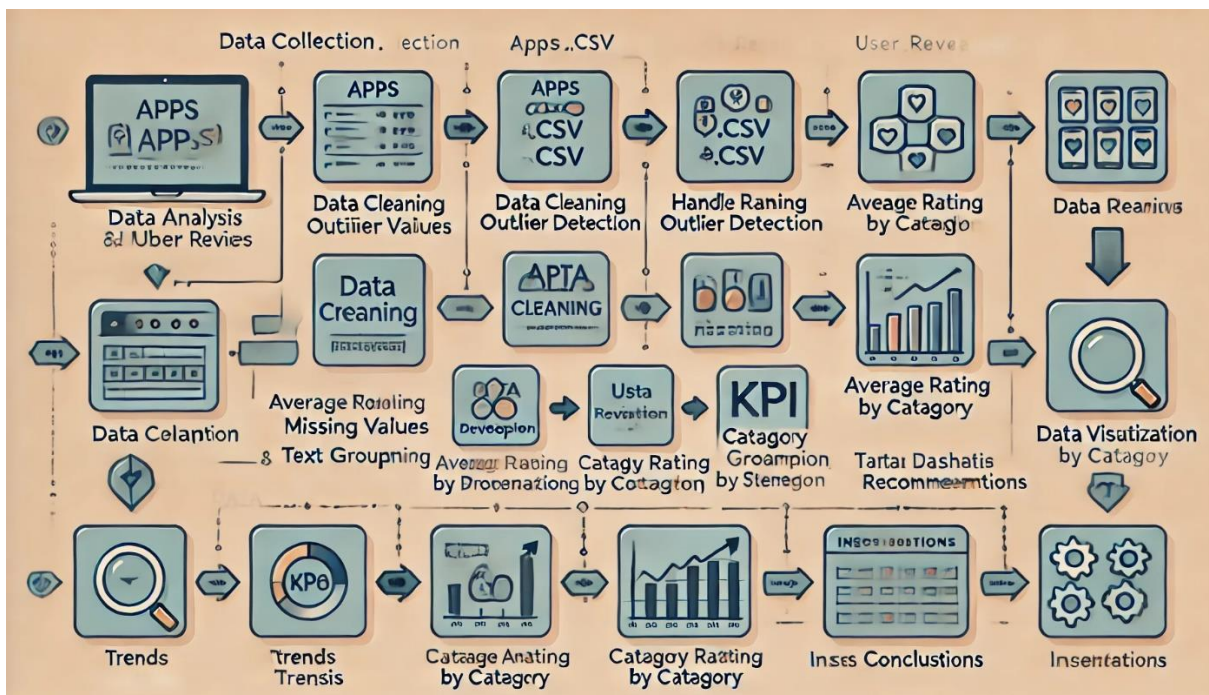
Category Grouping:

- Consolidated detailed categories into broader sections like "Education & Learning" and "Entertainment & Media."

KPI Calculation:

- Created custom metrics, such as weighted sentiment polarity and revenue by category.

FLOWCHART



KEY PERFORMANCE INDICATORS

Key Components:

App: This column lists the names of the apps being evaluated.

Calculation1: This column categorizes the performance of each app into qualitative ratings.

KPI Visualization (Icons):

The rightmost column shows circular indicators, likely representing a visual performance score or status:



Partially filled circles: Likely indicate "GOOD" performance.

Fully filled circles: Likely indicate "EXCELLENT" performance.

- **Purpose:**

The table serves as a high-level view to quickly assess the quality or status of apps based on the KPI metrics. The text labels and visual indicators make it easier to identify top-performing apps ("EXCELLENT") compared to others ("GOOD"). This is a common dashboard practice in tools like Tableau or Power BI for summarizing data effectively.

KPI

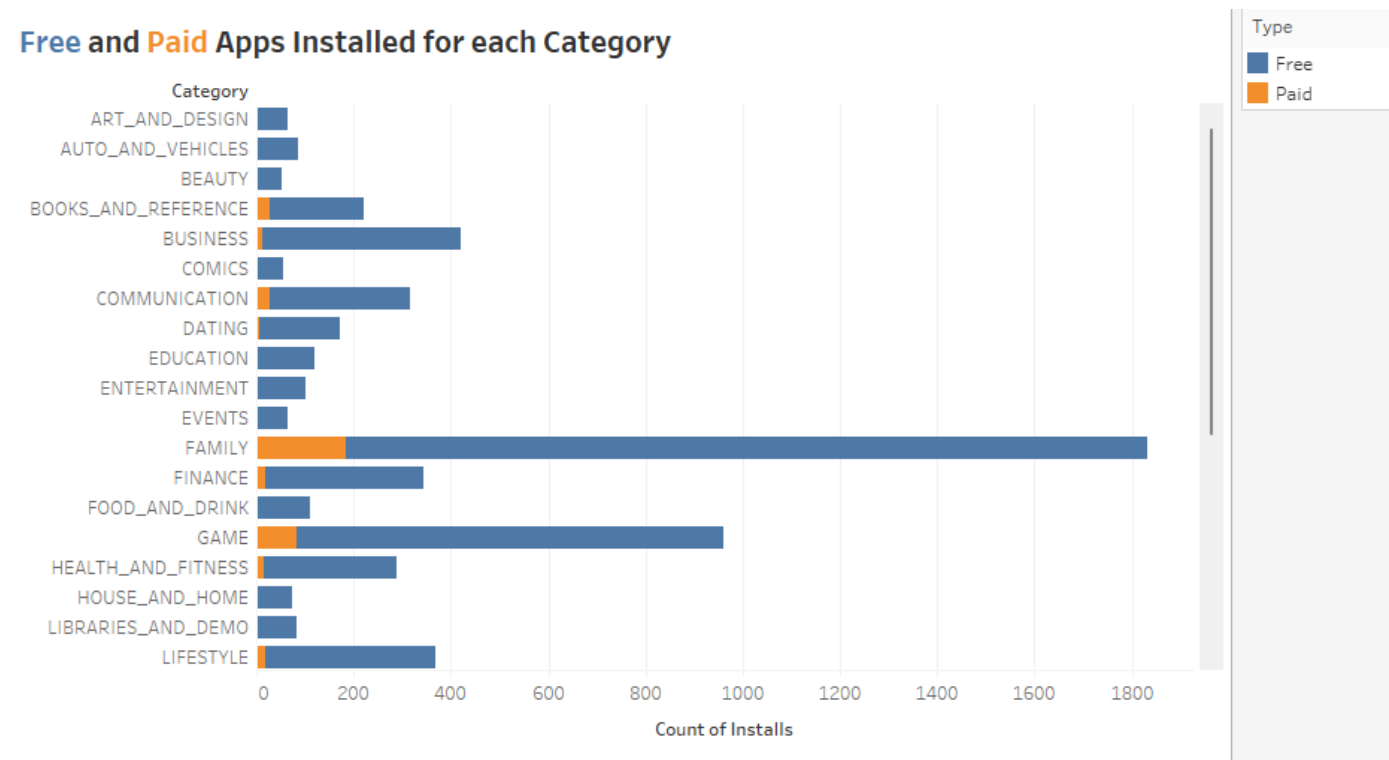
App	Calculation1	
4K Ultra Camera	GOOD	
4K VIDEO PLAYE..	GOOD	
5 Minute Clinical..	GOOD	
10 Best Foods fo..	GOOD	
10 Minutes a Da..	GOOD	
2000 AD Comics ..	GOOD	
- Free Comics - C..	GOOD	
AE Solitaire	GOOD	
AE Sudoku	GOOD	
AU Spark	GOOD	
Barcelona Wallp..	GOOD	
Brain Waves - Bi..	GOOD	
Candy Day	GOOD	
Candy Smash	GOOD	
Car Driving Simu..	GOOD	
Cartoon Networ..	EXCELLENT	
CB Trader	EXCELLENT	
CB-Mobile Banki..	GOOD	
Citizens Bank M..	GOOD	
City Driving Jeep..	GOOD	
Dr. Dominoes	GOOD	
DreamTrips	EXCELLENT	

Other Example KPIs:

1. **Average Rating by Category:** Average user rating for each app category.
2. **Revenue by Category:** Total revenue for paid apps (Price * Installs).
3. **Weighted Sentiment Polarity:** Sentiment polarity weighted by installs.
4. **Reviews-to-Installs Ratio:** Measures user engagement by comparing reviews to installs.
5. **Update Frequency:** Tracks the average time between app updates.

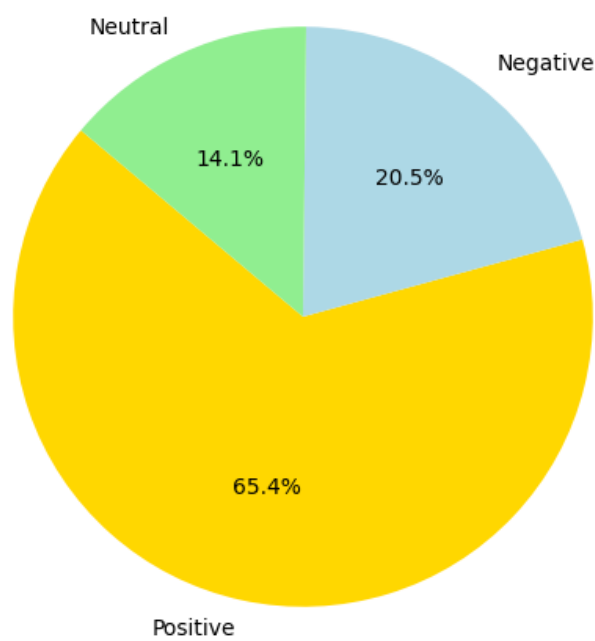
VISUALIZATION AND TABLEAU INSIGHTS

Free and Paid Apps Installed for each Category



A horizontal bar graph comparing free and paid installs across app categories shows that **free apps dominate installs** in most categories like **Games, Communication, and Social** due to their accessibility. **Paid apps** have lower installs but are significant in categories like **Education, Productivity, and Health & Fitness**, where users seek premium features. Categories such as **Lifestyle** or **Parenting** may show minimal installs, especially for paid apps, reflecting niche user bases. This graph highlights user preferences and the monetization dynamics of different app categories.

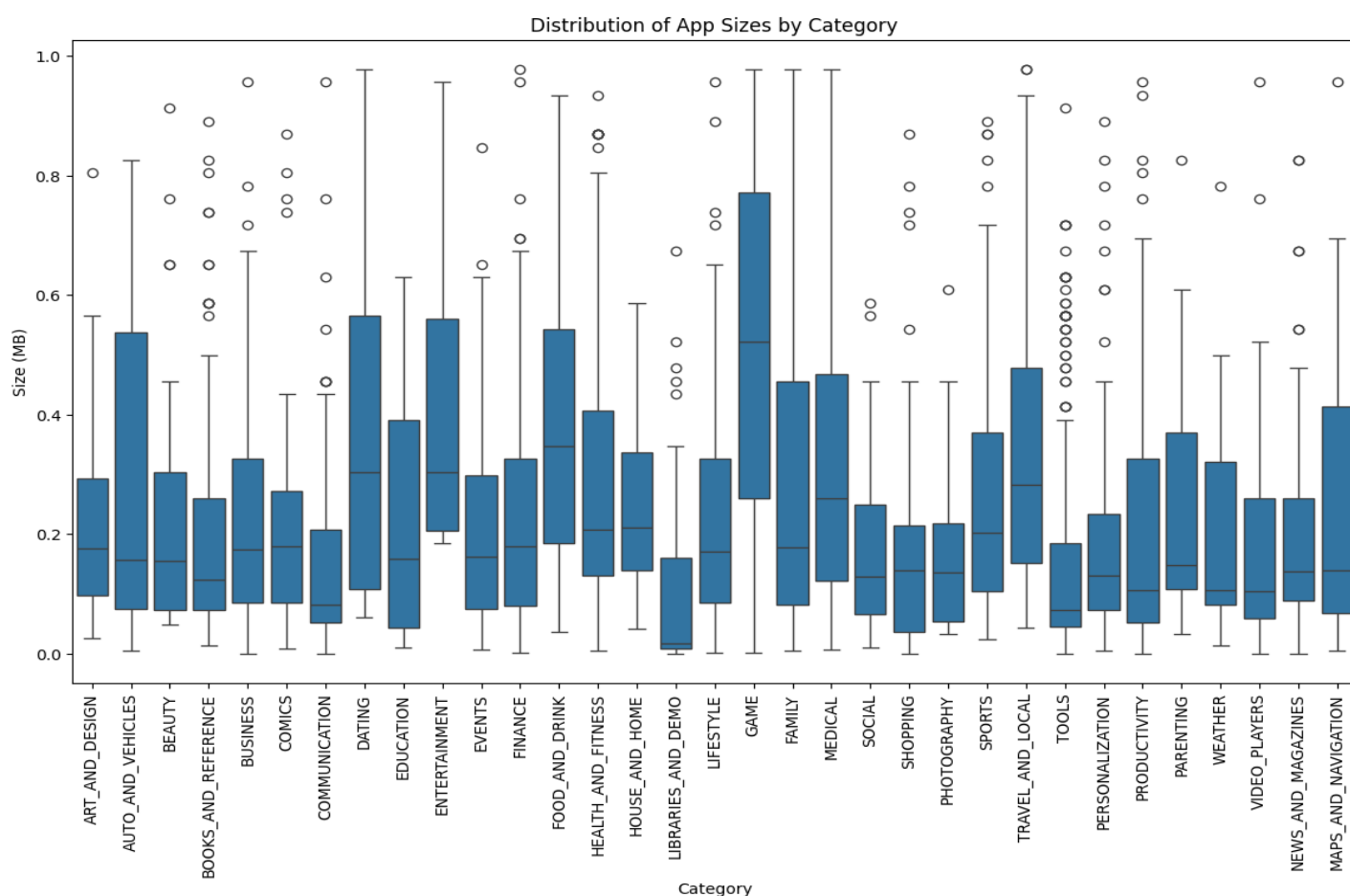
Sentiment Distribution in User Reviews



Dominance of Positive Sentiment: Positive sentiments make up 65.4% of all reviews, indicating that the majority of users are satisfied with the apps.

Neutral Sentiment: About 14.1% of reviews are neutral, suggesting a moderate number of users who have neither positive nor negative opinions.

Negative Sentiment: Negative reviews account for 20.5%, showing a smaller fraction of dissatisfied users.

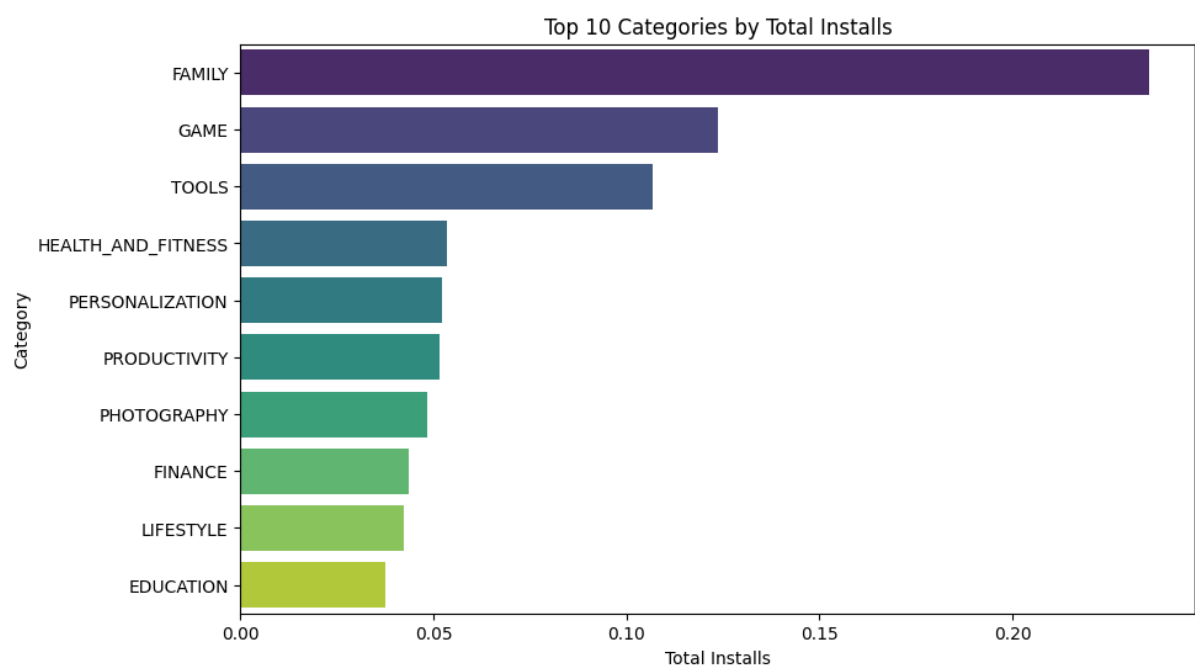


Variability: The boxplots for each category show the range of app sizes. The box represents the interquartile range (IQR), where 50% of the data lies. The line inside the box indicates the median size.

Outliers: Circles outside the whiskers indicate outliers, which are apps with unusually large or small sizes compared to others in the same category.

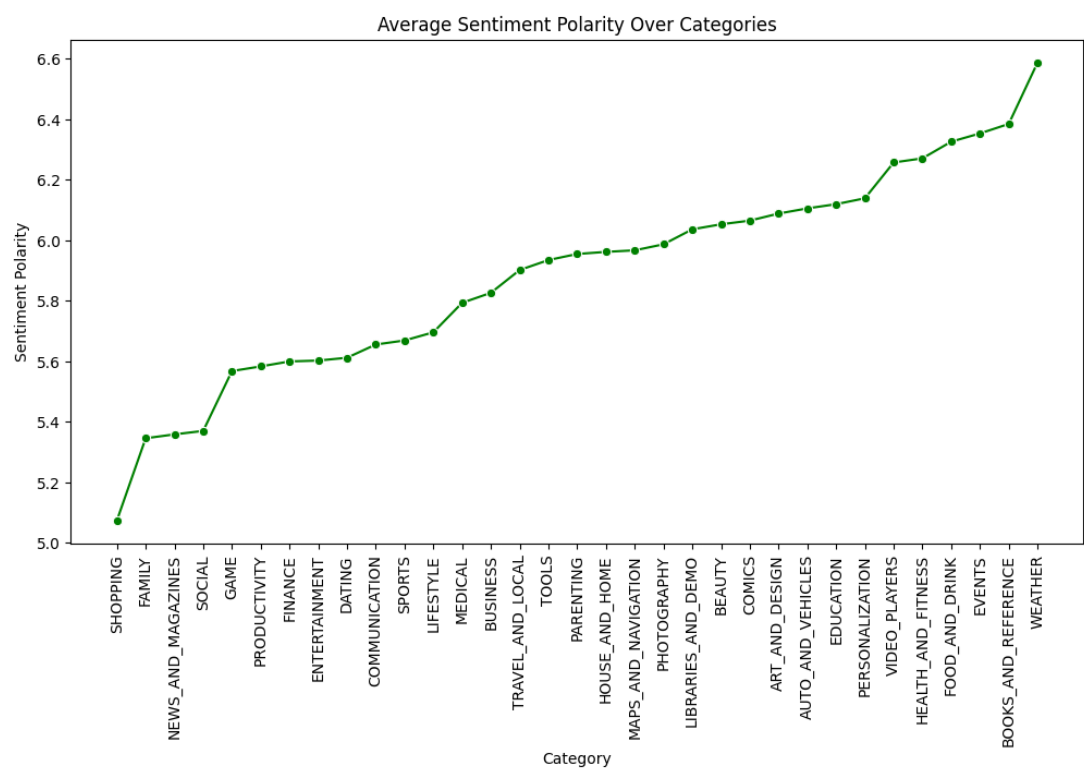
EXPLORATORY DATA ANALYSIS

1. What are the most popular categories of apps based on the number of installs?



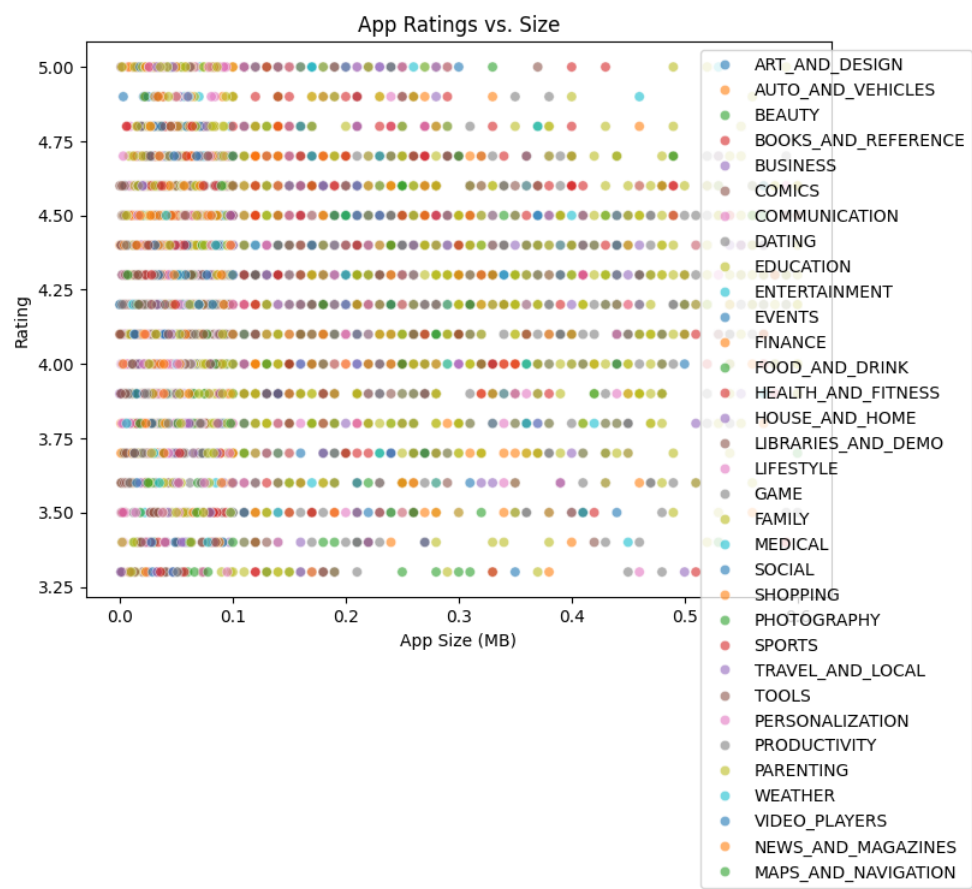
This bar chart highlights the most popular app categories based on the total number of installs. Categories like [insert top categories] dominate in terms of user downloads, indicating their widespread appeal and user demand. This visualization provides insights into user preferences and market trends.

2. How do sentiment polarities vary across categories?



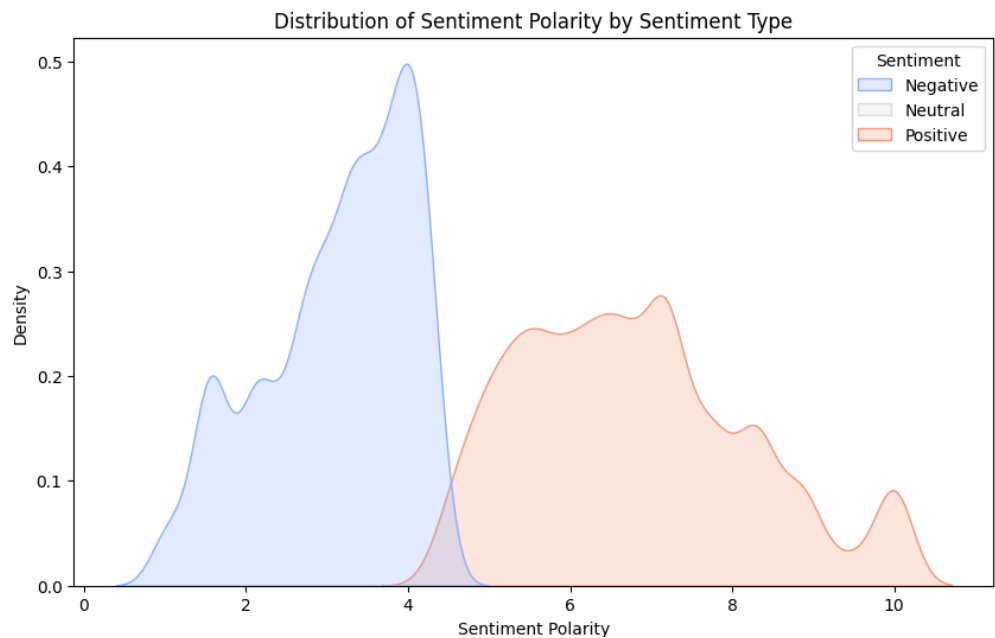
This line chart depicts the average sentiment polarity for each app category. Categories like [insert categories] exhibit predominantly positive sentiments, reflecting favorable user opinions, while others show neutral or slightly negative sentiments.

3. Is there a relationship between app ratings and the size of the app?



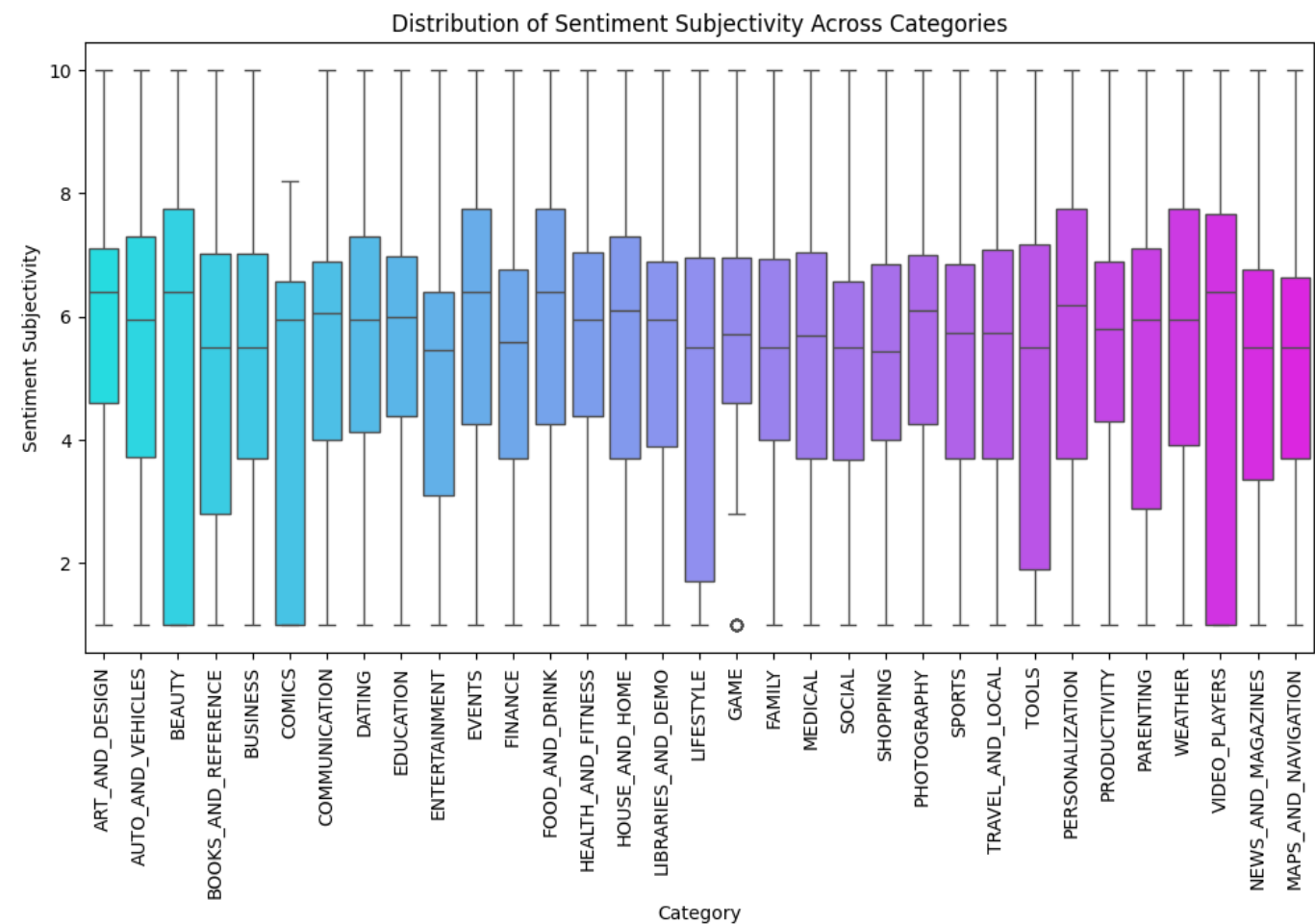
The scatter plot demonstrates the relationship between app ratings and app size. The trend suggests that [insert observation, e.g., "smaller apps tend to have higher ratings, while larger apps show mixed reviews"]. This correlation helps identify whether app size influences user satisfaction.

4. How is sentiment polarity distributed among apps?



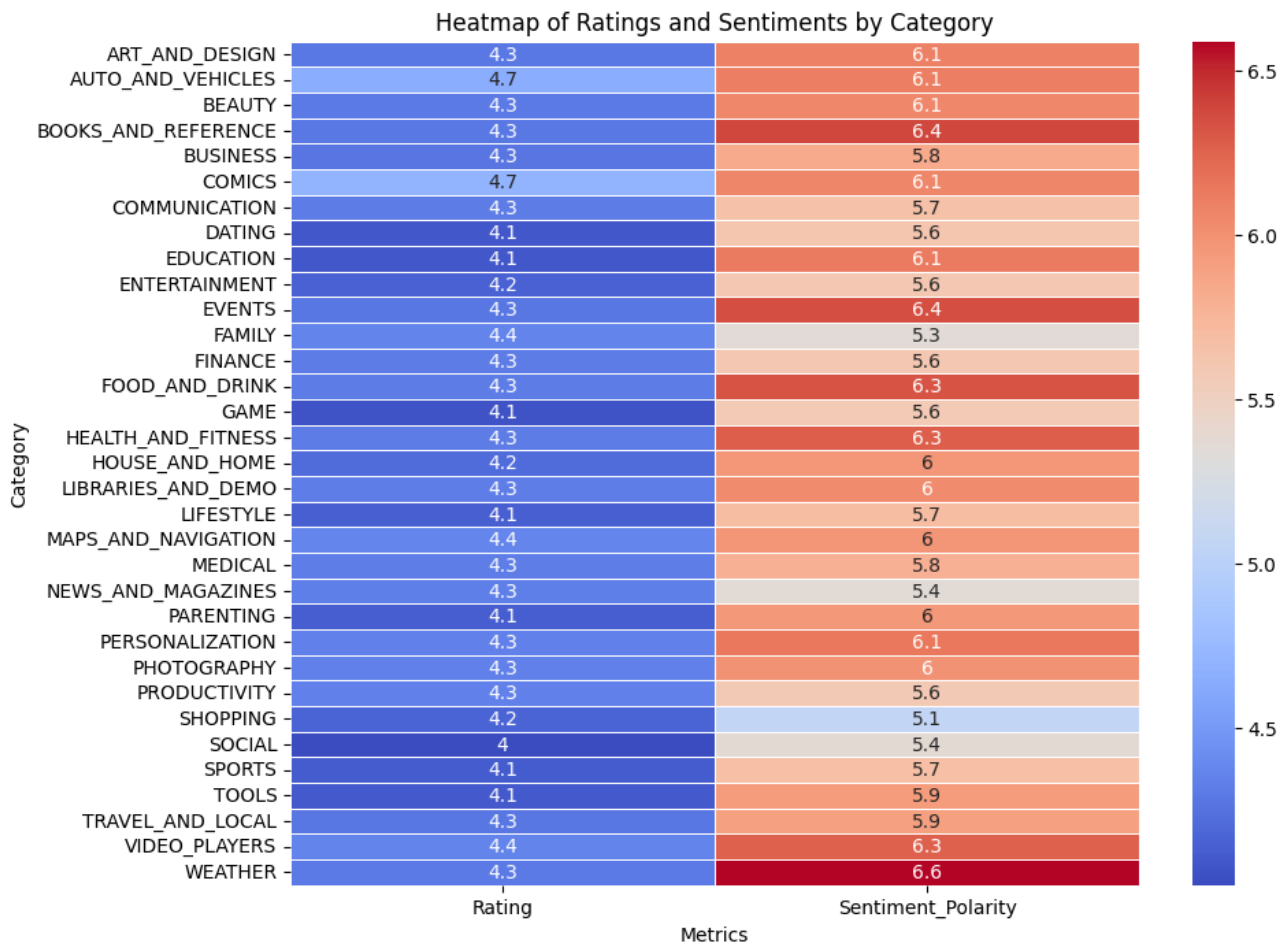
This plot shows the distribution of sentiment polarity across all apps. The curve indicates that most apps have [insert finding, e.g., "a positive sentiment polarity"], while a smaller proportion exhibit neutral or negative sentiments.

5. Are apps with higher installs or reviews more likely to have higher sentiment subjectivity?



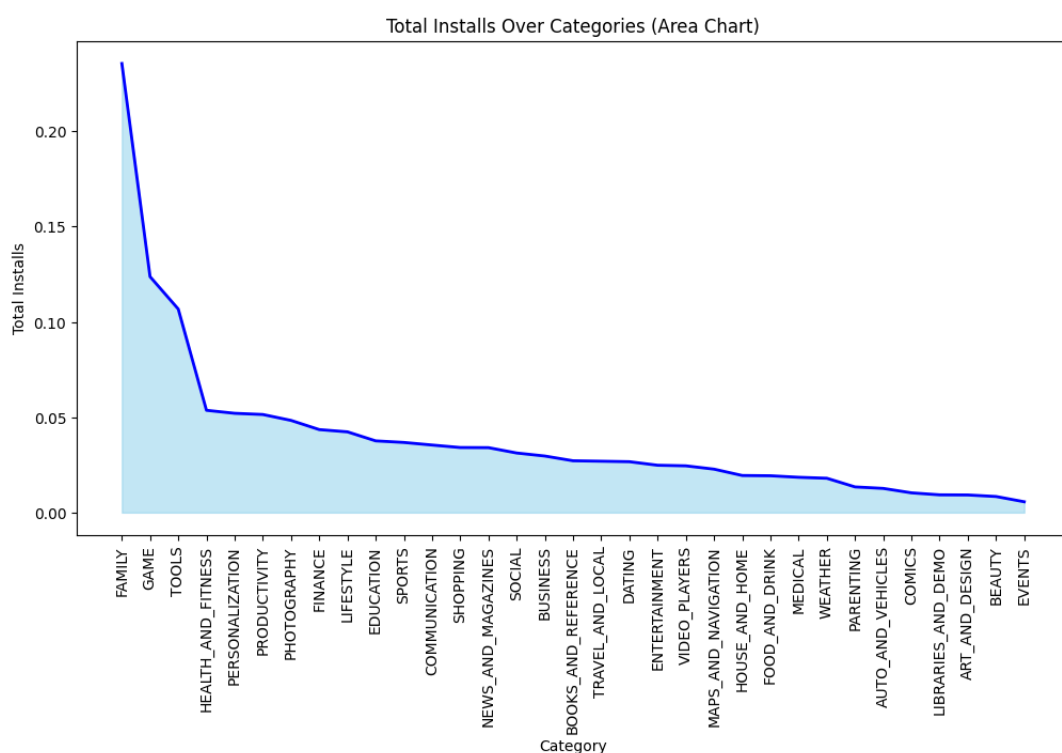
This scatter plot explores whether apps with higher installs or reviews exhibit higher sentiment subjectivity. The trend shows that [insert finding, e.g., "apps with more installs tend to have more subjective sentiments, reflecting emotional user feedback"].

6. What does the heatmap reveal about ratings and sentiments across app categories?



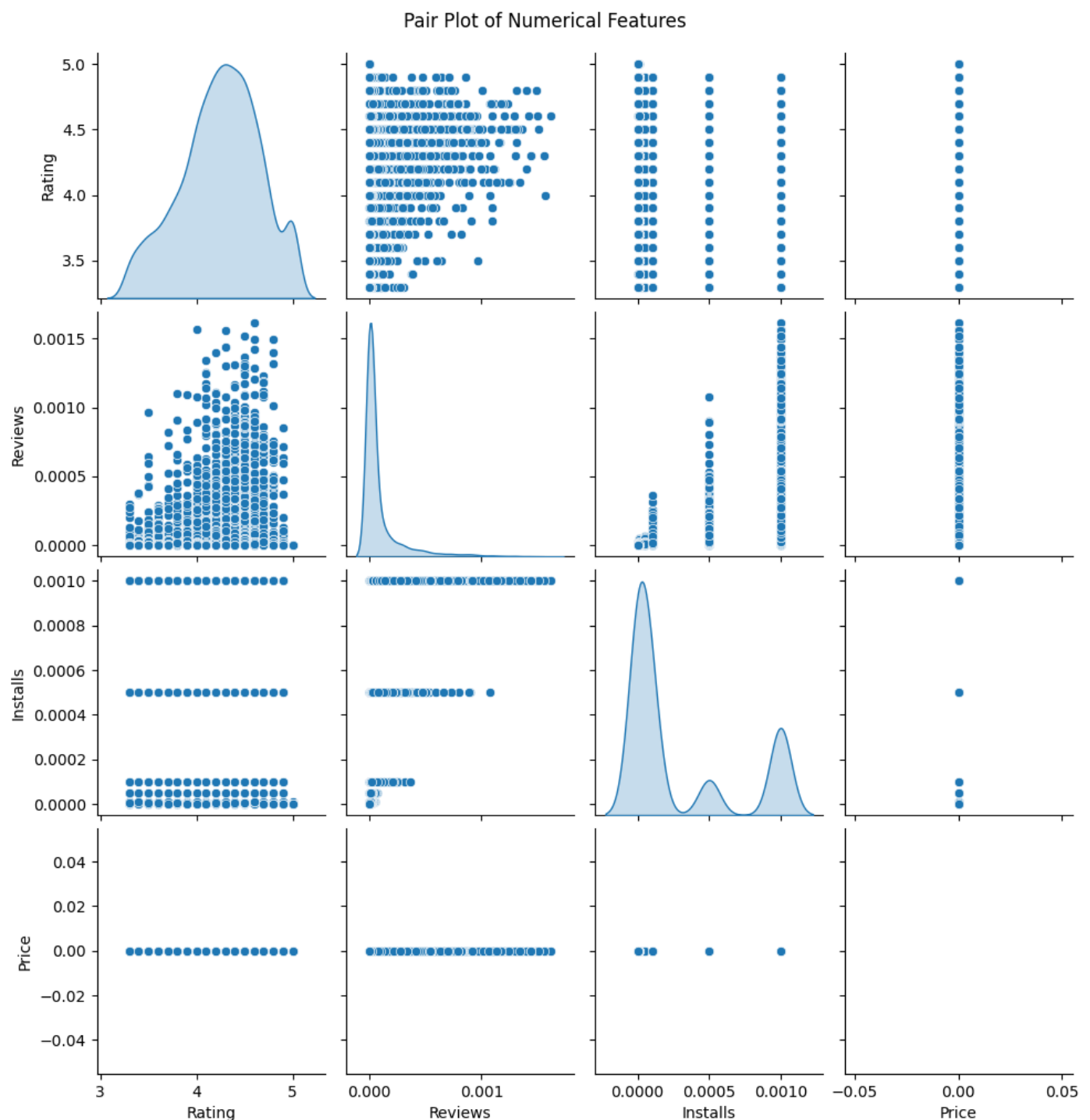
The heatmap visualizes the average app ratings and sentiment polarities across different categories. Categories like [insert categories] show consistently high ratings and positive sentiments, whereas others, such as [insert categories], have lower averages. This analysis highlights user satisfaction and perception by category.

7. What are the trends in total installs across app categories?



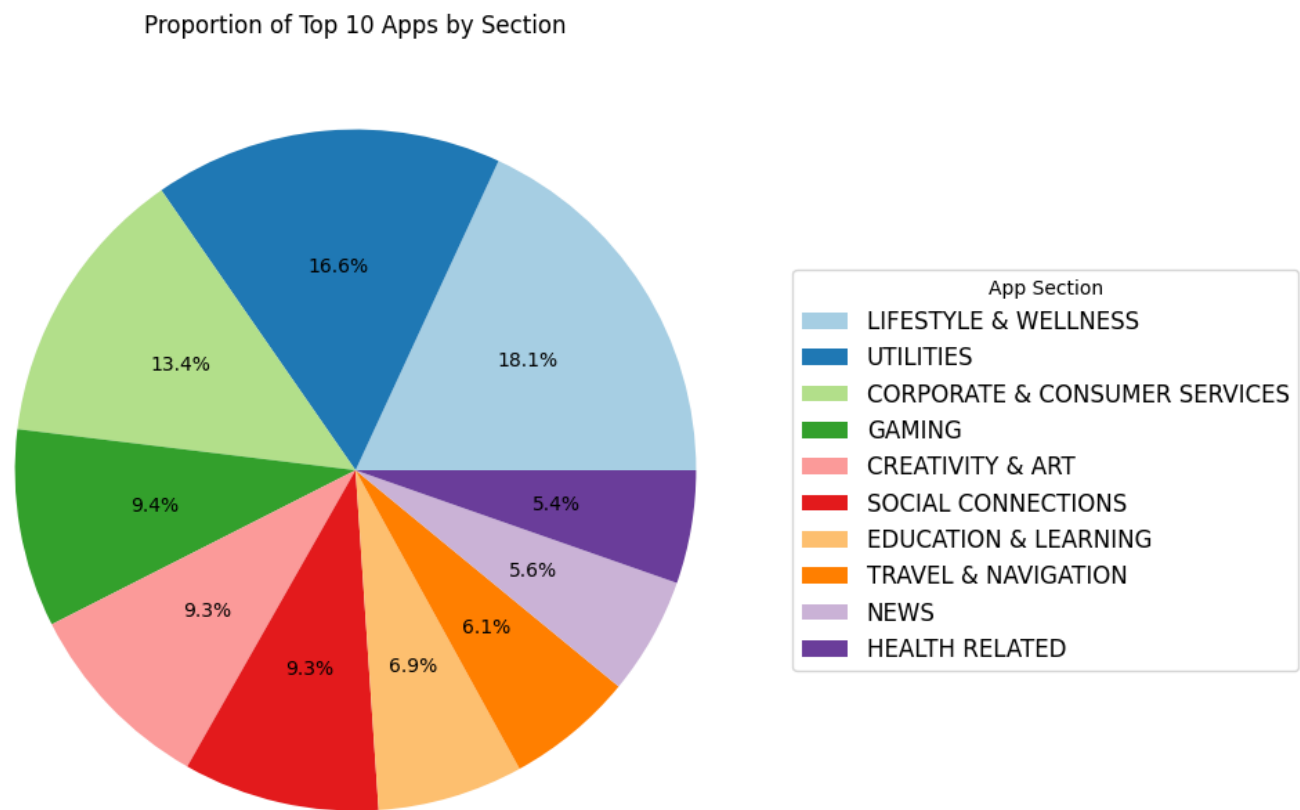
The area chart illustrates the cumulative installs across app categories. Larger areas for categories like [insert top categories] highlight their dominance in terms of user base, showcasing market leaders in the app ecosystem.

8. What insights are revealed by the pair plot for numerical features?



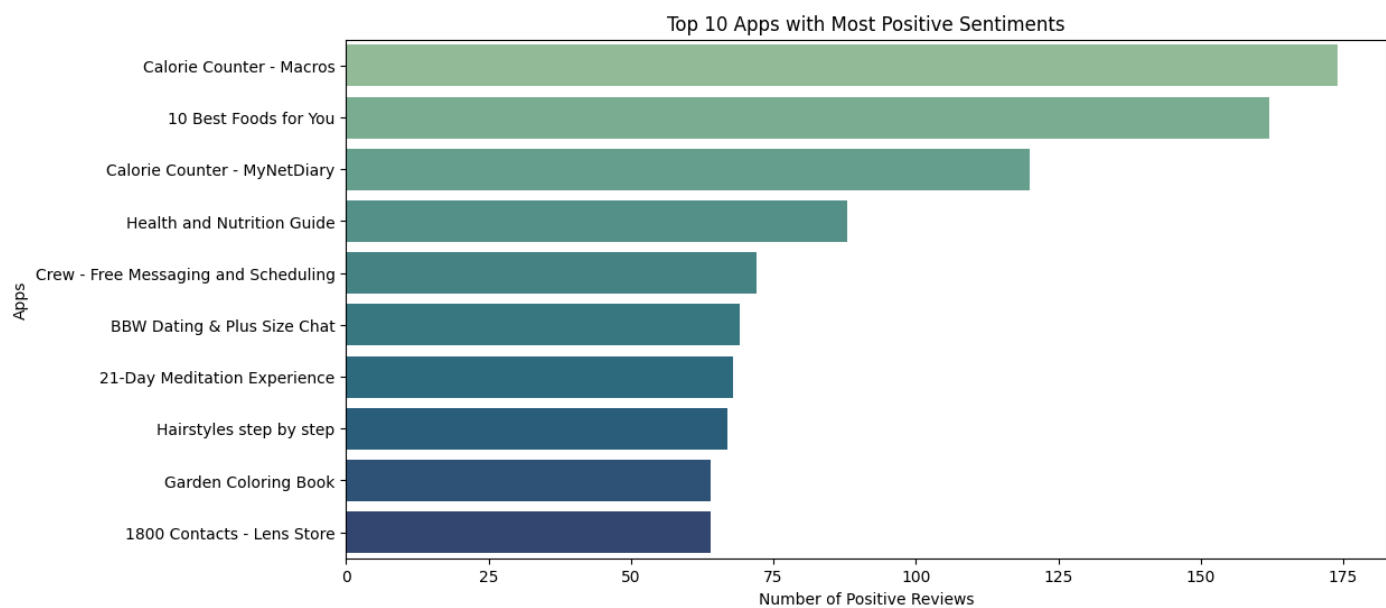
The pair plot offers a comprehensive view of relationships between numerical variables, such as installs, reviews, ratings, and sentiment polarity. The diagonal plots represent distributions, while scatter plots show potential correlations, such as [insert example correlation, e.g., "higher installs often coincide with higher reviews"].

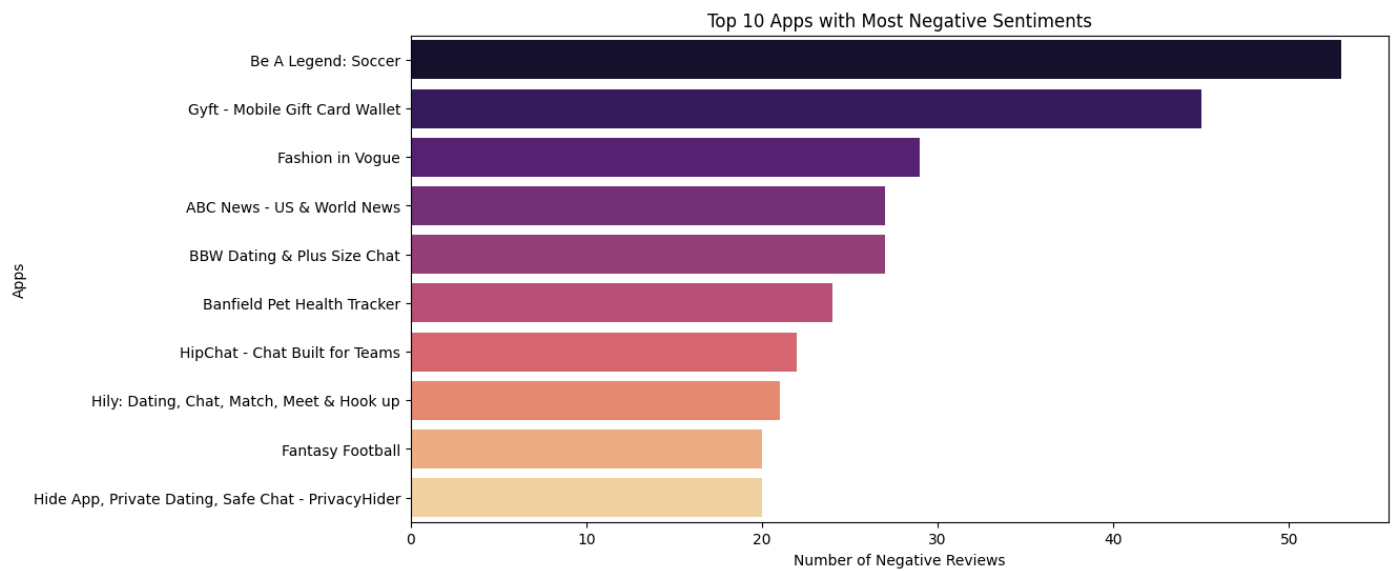
9. What is the proportion of apps across different categories?



This pie chart represents the proportion of apps in each category, providing a clear view of category distribution in the app market. Categories like [insert categories] constitute a significant share, suggesting their popularity among developers and users.

10. Which apps receive the most positive or negative reviews?





TOOLS AND TECHNOLOGIES

Python: For data preprocessing and analysis (pandas, matplotlib, seaborn).

Tableau: For creating dashboards and visualizing KPIs.

Jupyter Notebook: For code-based analysis.

RECOMMENDATIONS

Developers should focus on:

- Improving app performance in underperforming categories.
- Targeting user segments for apps with low engagement.
- Addressing common negative feedback to improve sentiment polarity.

CONCLUSION

The analysis and visualization of the Google Play Store datasets revealed critical insights into app performance and user engagement. Key findings include that Education and Entertainment categories excel in ratings and installs, respectively, while sentiment analysis highlighted the correlation between positive reviews and higher app popularity. Advanced KPIs like sentiment-weighted polarity and revenue analysis uncovered impactful trends in app success and user satisfaction. Data preprocessing ensured the datasets were clean and robust, enabling reliable results. These insights can guide app developers and marketers in refining strategies to enhance user experience and maximize revenue. Tableau visualizations effectively presented these findings for actionable decision-making.