# Google Play Store Data Analysis

---

## Objective

The objective of this project is to analyze the Google Play Store dataset to uncover insights about app ratings, installs, reviews, and sentiment distribution. The analysis uses Python libraries for data manipulation and visualization to provide actionable recommendations for app developers and marketers.

---

## Data Preparation

### 1. Data Import

- Datasets `googleplaystore.csv` and `googleplaystore_user_reviews.csv` were imported into Python using pandas.
- Key fields such as `Size`, `Installs`, `Price`, and `Rating` were cleaned and converted to appropriate data types.
- Missing values in the `Rating` column were handled by removing rows with null values to ensure accurate analysis.

### 2. Preprocessing Steps

- **Size Conversion**: Converted app sizes from MB and KB to numeric values in MB.
- **Sentiment Analysis**: Used the VADER sentiment analyzer to classify user reviews as Positive, Neutral, or Negative.
- **Installs and Price**: Removed unnecessary symbols (e.g., "+", ",", "$") and converted fields to numeric values.

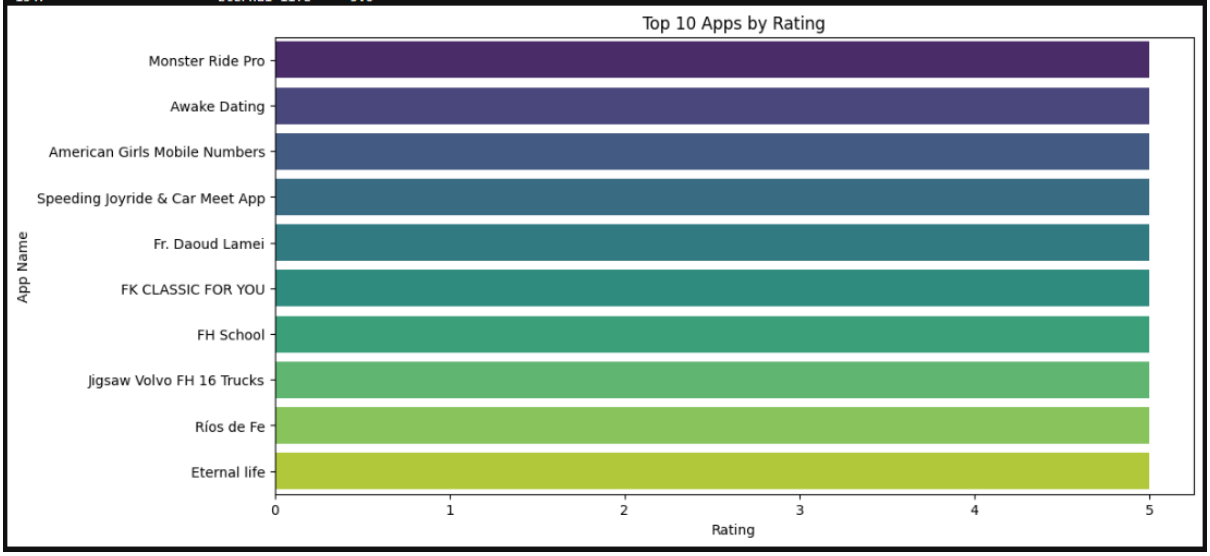
---

## Visualizations and Analysis

### 1. Top 10 Apps by Rating

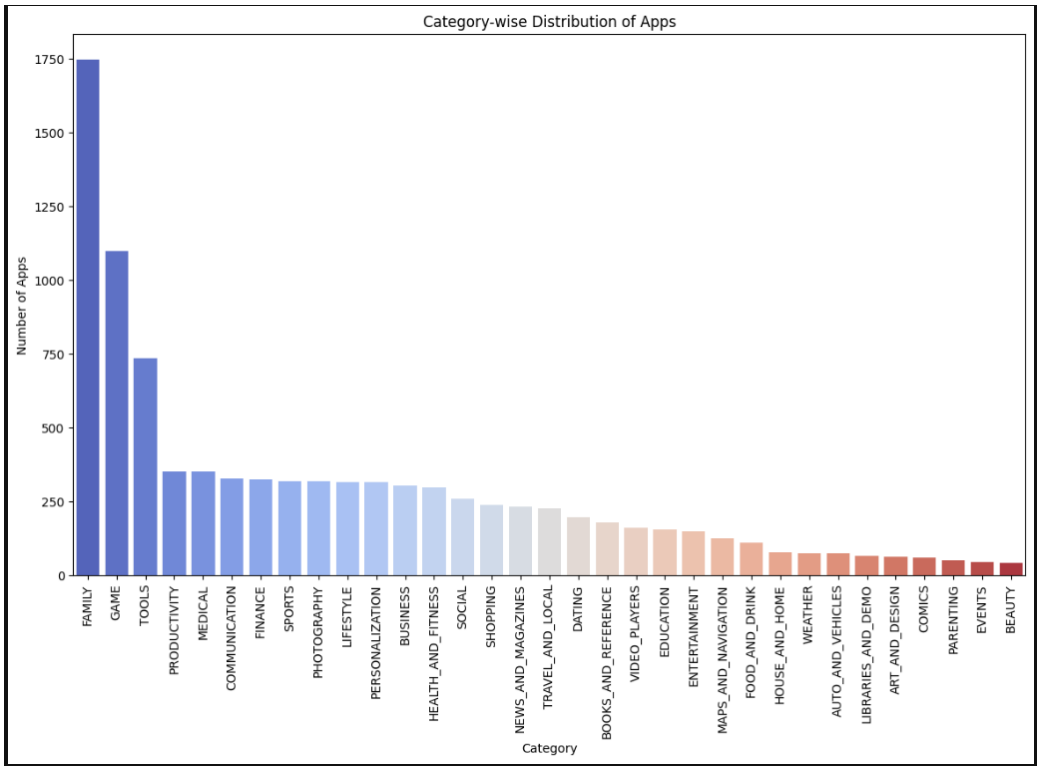
- Identified the apps with the highest user ratings.
- A bar chart displayed the ratings of the top 10 apps.
- **Key Insight**: Apps with niche functionalities tend to have higher user satisfaction.

```
Top 10 Apps by Rating:
                                App  Rating
10776                 Monster Ride Pro     5.0
615                     Awake Dating     5.0
612      American Girls Mobile Numbers     5.0
654    Speeding Joyride & Car Meet App     5.0
10820                 Fr. Daoud Lamei     5.0
10529              FK CLASSIC FOR YOU     5.0
10416                      FH School     5.0
10407        Jigsaw Volvo FH 16 Trucks     5.0
10357                      Ríos de Fe     5.0
1547                     Eternal life     5.0
```




## 2. Category-Wise Distribution of Apps

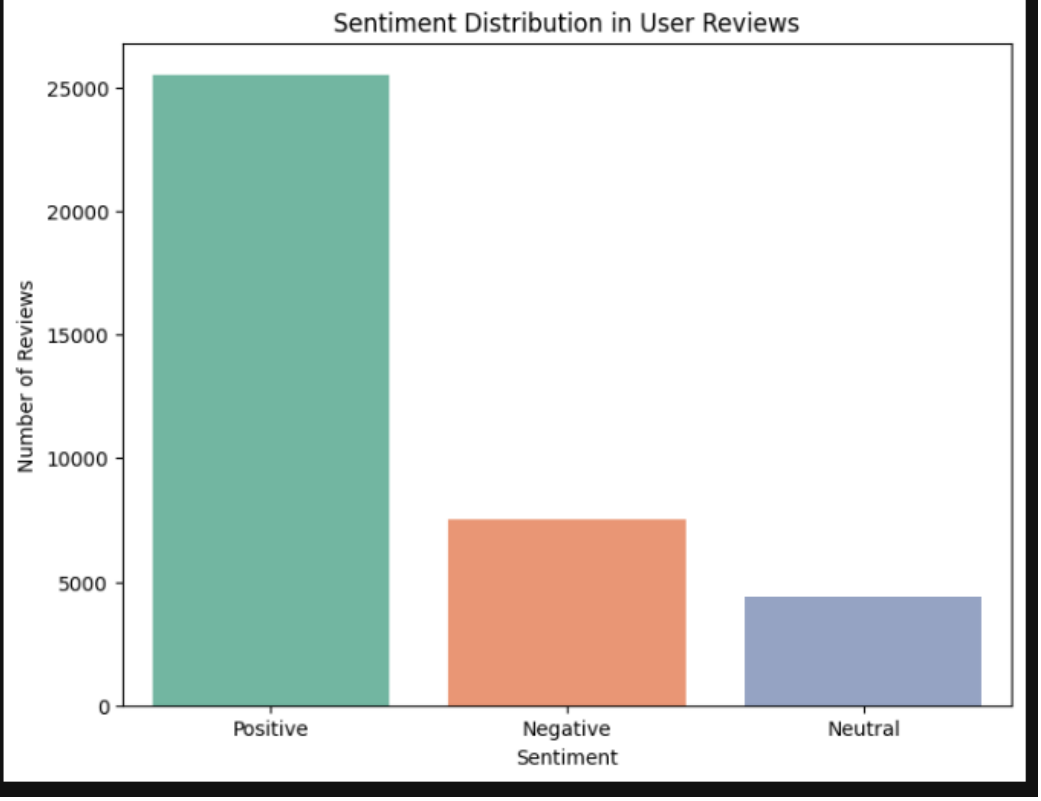
- Counted apps in each category and visualized the distribution using a bar chart.
- **Key Insight**: Categories like `Family` and `Game` had the highest number of apps, indicating market saturation in these segments.



## 3. Sentiment Analysis on User Reviews

- Analyzed user sentiments and classified reviews into Positive, Neutral, or Negative.
- A bar chart displayed the distribution of sentiments.
- **Key Insight**: Positive sentiments dominated, highlighting overall user satisfaction, with significant improvement opportunities from Neutral and Negative feedback.

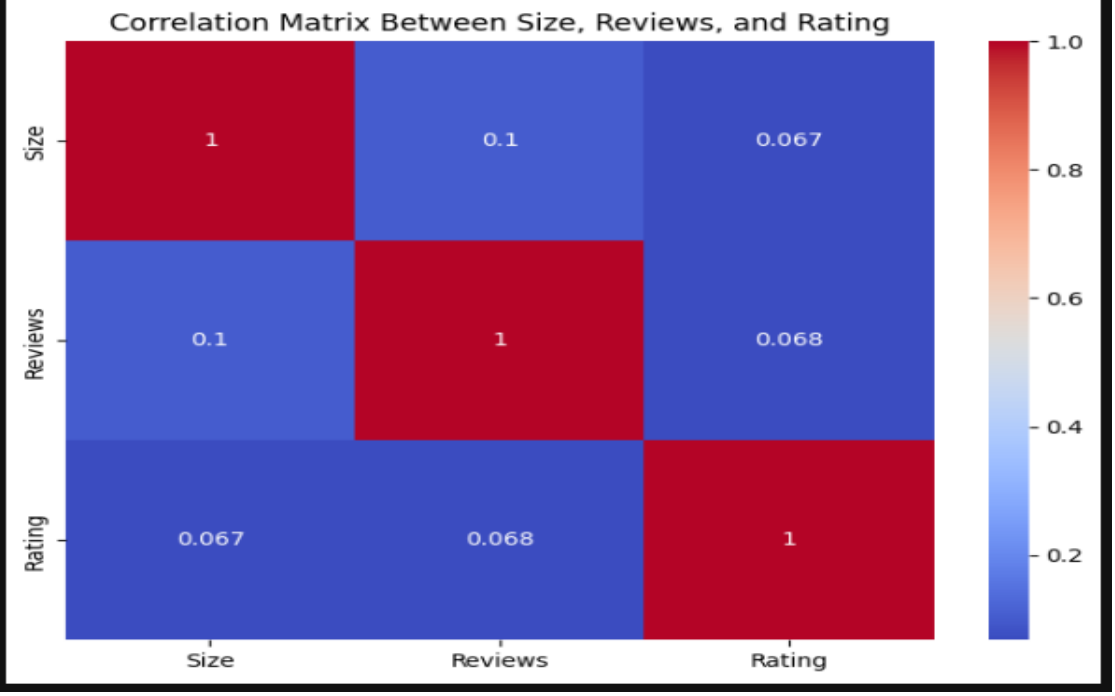
```
Sentiment Analysis Results:
Sentiment
Positive    25526
Negative     7505
Neutral      4396
Name: count, dtype: int64
```




## 4. Correlation Analysis

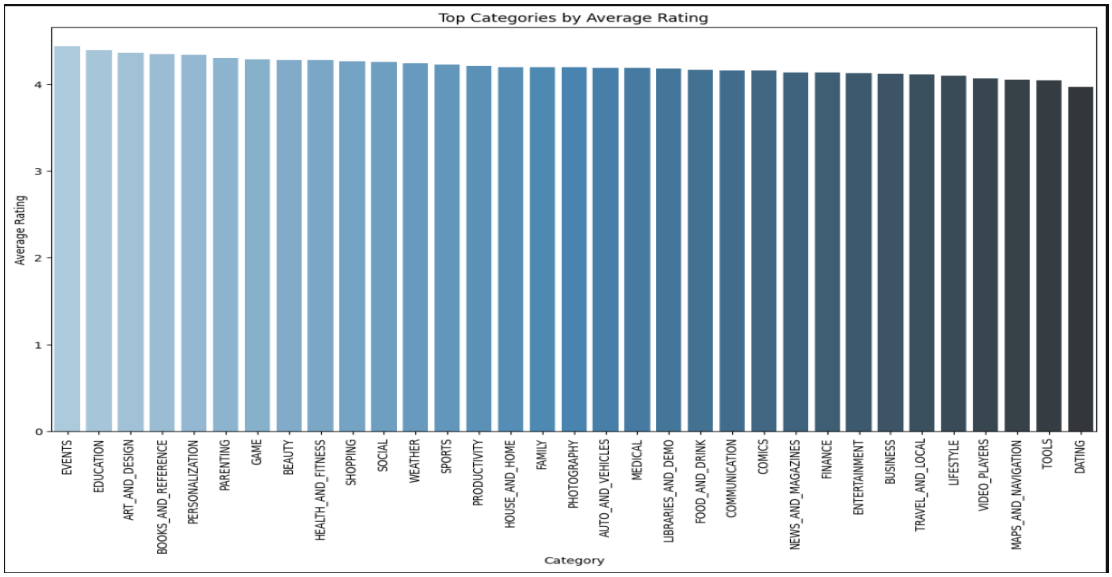
- Investigated the relationship between app size, number of reviews, and ratings.
- A heatmap was created to visualize correlations.
- **Key Insight**: Higher review counts correlated moderately with higher ratings, suggesting user engagement as a factor.

```
Correlation Between App Size, Reviews, and Rating:
              Size   Reviews     Rating
Size      1.000000  0.103888  0.067478
Reviews   0.103888  1.000000  0.068141
Rating    0.067478  0.068141  1.000000
```



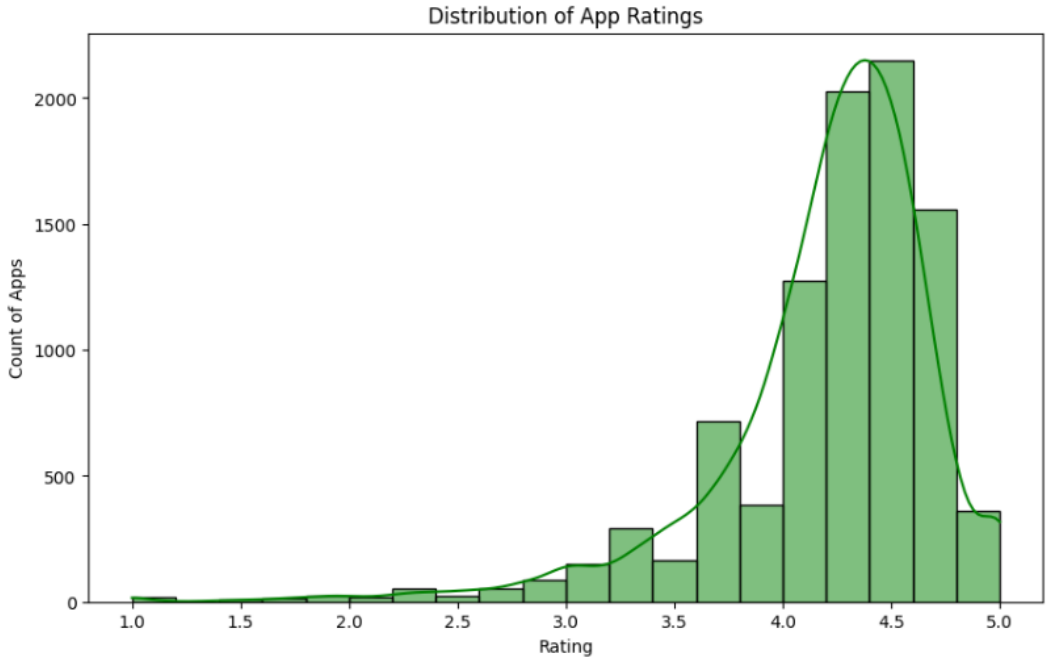

## 5. Top Categories by Average Rating

- Calculated the average rating for each category and visualized it using a bar chart.
- **Key Insight**: Categories like `Health & Fitness` and `Education` consistently received high user ratings.



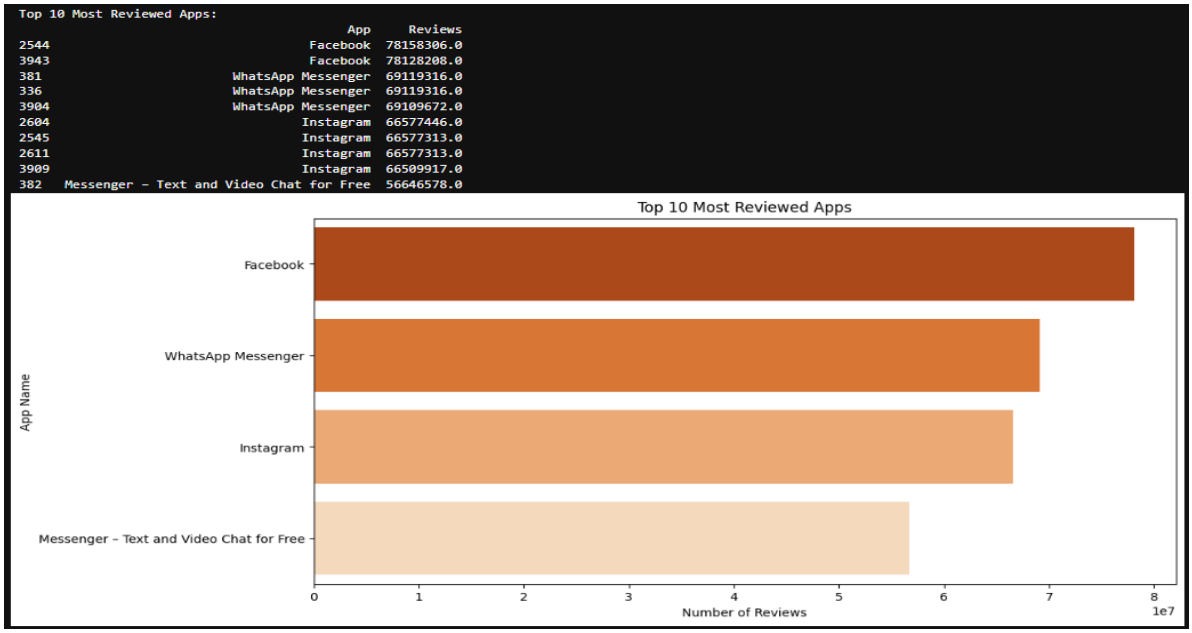
## 6. Distribution of App Ratings

- Visualized the distribution of ratings across all apps using a histogram.
- **Key Insight**: Most apps were rated between 4.0 and 4.5, indicating a generally favorable user perception.



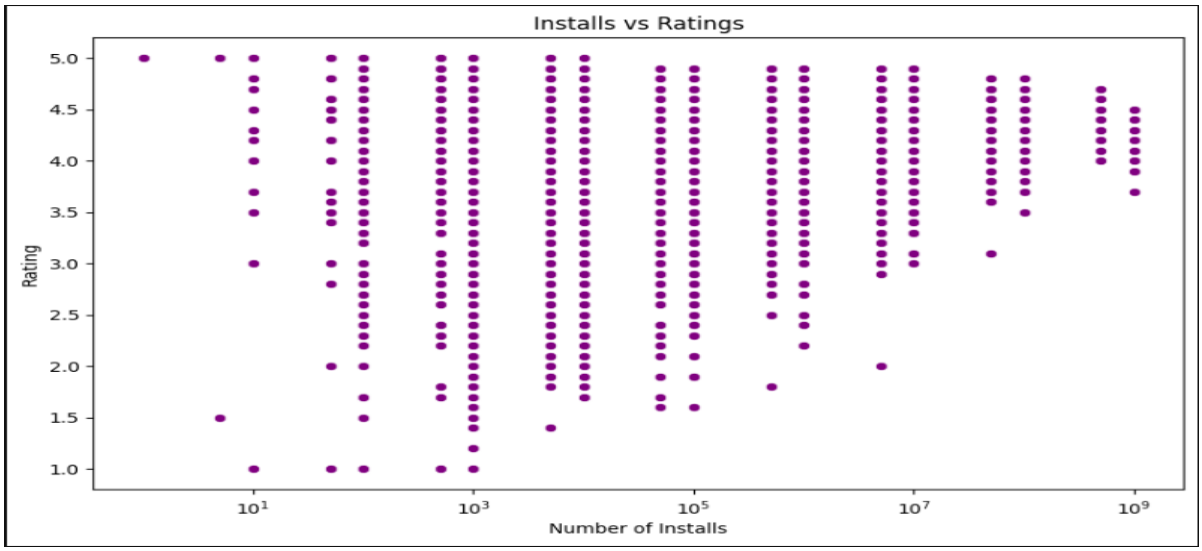

## 7. Top 10 Most Reviewed Apps

- Identified apps with the highest number of user reviews and visualized them using a bar chart.
- **Key Insight**: Popular apps like social media platforms and utility apps had significantly higher engagement.



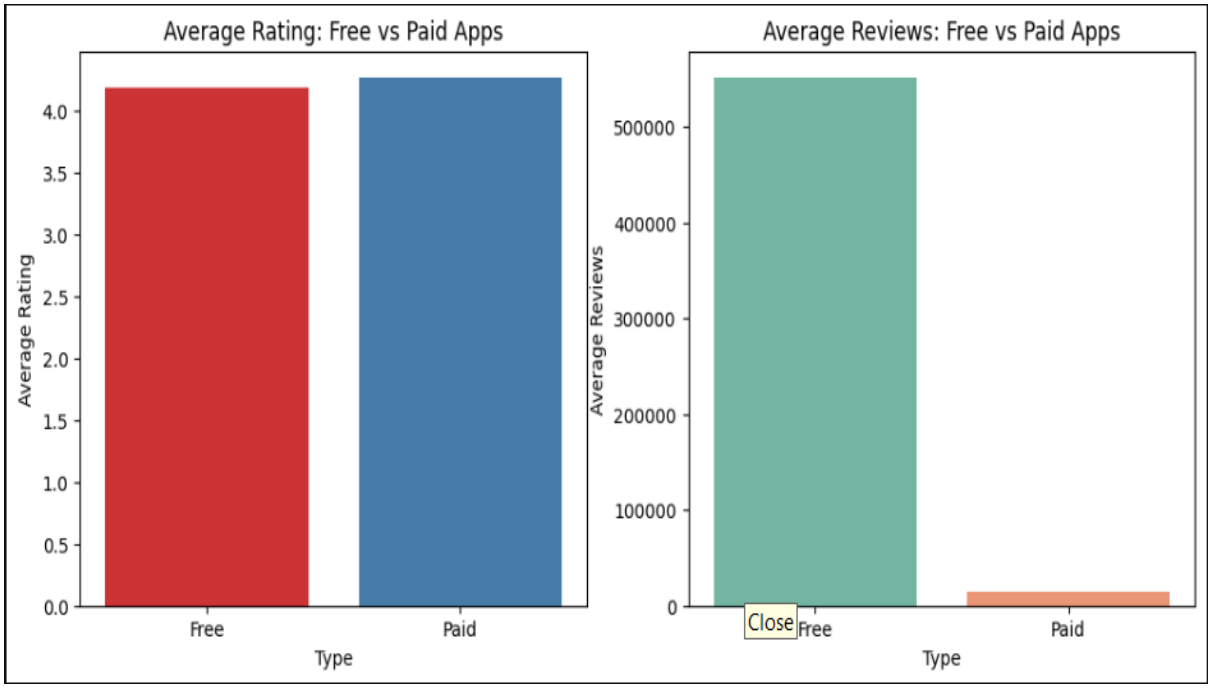
## 8. Installs vs. Ratings

- A scatter plot visualized the relationship between the number of installs and ratings.
- **Key Insight**: High installs often corresponded to average ratings, suggesting that mass adoption doesn't always equate to user satisfaction.



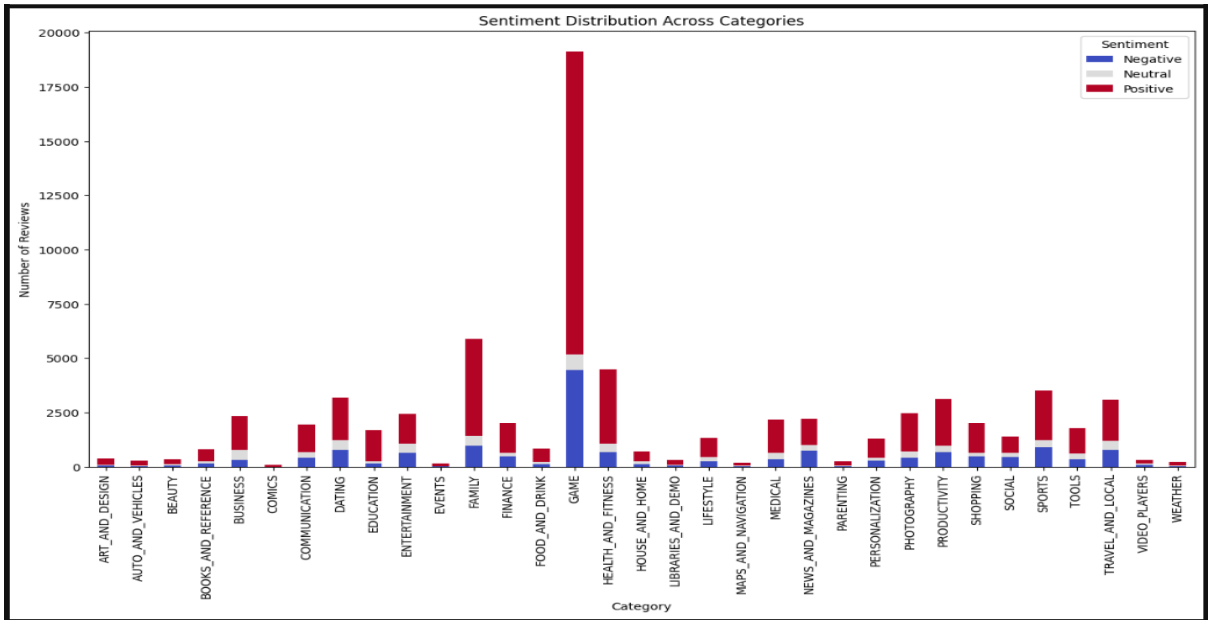

## 9. Free vs. Paid Apps Analysis

- Compared average ratings and reviews for free and paid apps using bar charts.
- **Key Insight**: Free apps had more installs and reviews, but paid apps often received higher average ratings.



## 10. Sentiment Distribution by Category

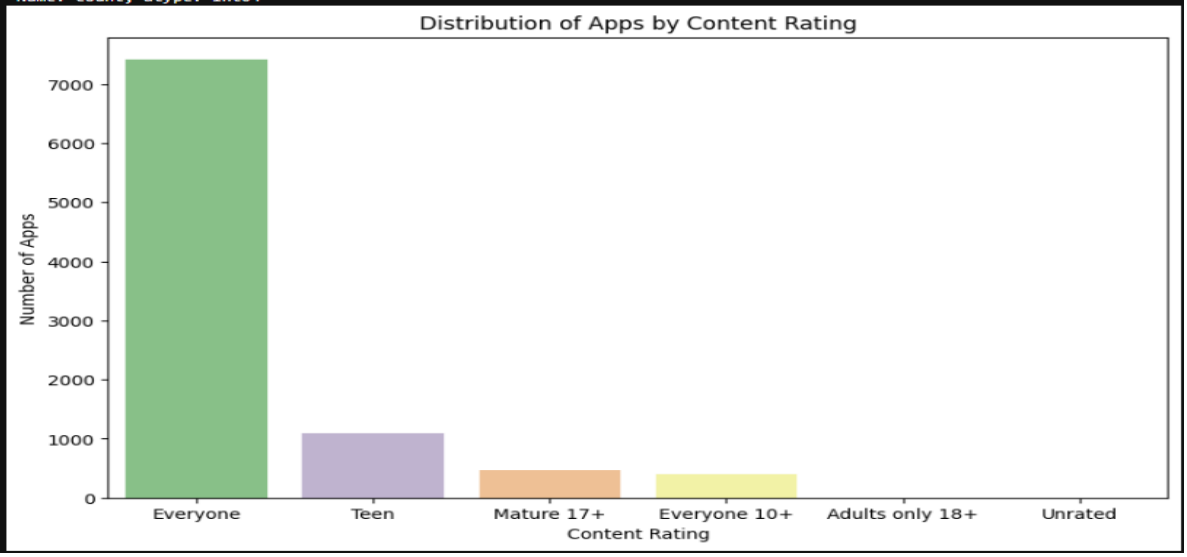
- Grouped sentiments by category and visualized them using a stacked bar chart.
- **Key Insight**: Categories like `Education` and `Entertainment` had the highest percentage of positive reviews.




## 11. Content Rating Distribution

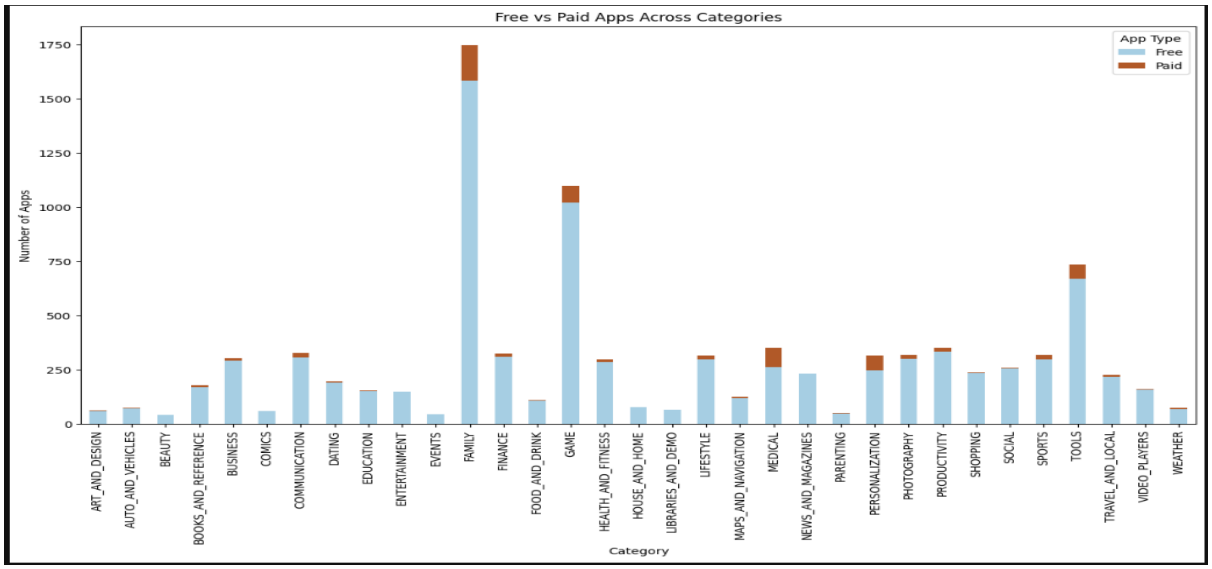
- Analyzed the distribution of apps by content rating (e.g., Everyone, Teen, Mature 17+).
- **Key Insight**: Most apps were rated for `Everyone`, emphasizing broad market targeting.

```
Content Rating Distribution:
Content Rating
Everyone           7420
Teen               1084
Mature 17+          461
Everyone 10+        397
Adults only 18+       3
Unrated               1
Name: count, dtype: int64
```



## 12. Free vs. Paid Apps per Category

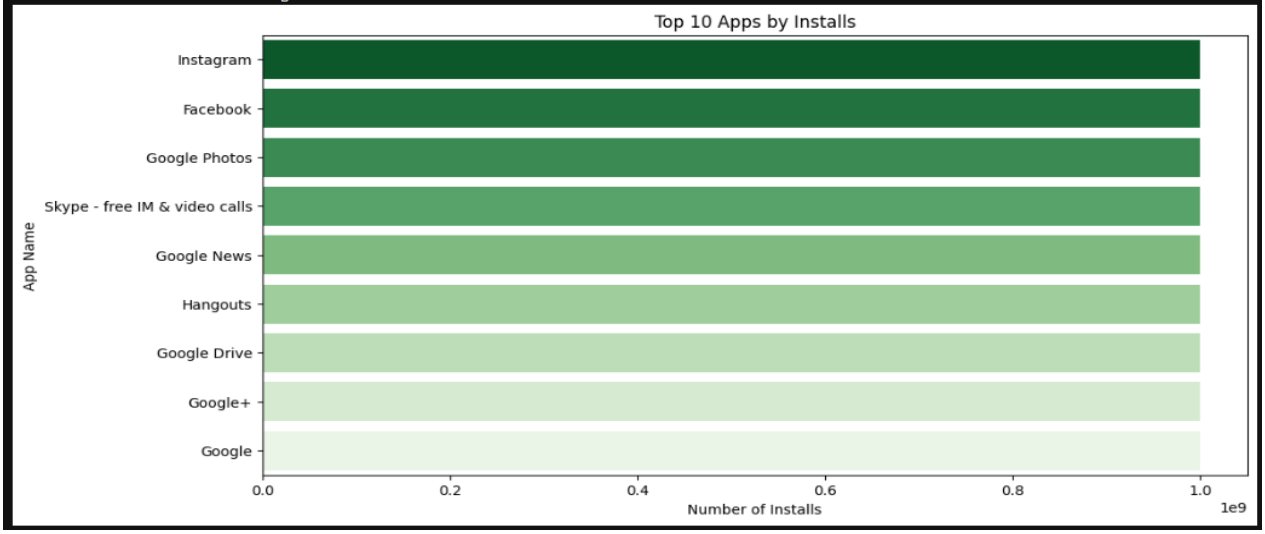
- Analyzed the proportion of free and paid apps across categories.
- **Key Insight**: Categories like `Tools` and `Business` had a higher proportion of paid apps compared to `Family` and `Game`.




## 13. Top 10 Apps by Installs

- Identified the most installed apps and visualized them using a bar chart.
- **Key Insight**: Utility and social media apps dominated the install counts.4

```
Top 10 Apps by Installs:
                              App      Installs
2545                    Instagram  1.000000e+09
2544                     Facebook  1.000000e+09
5395                Google Photos  1.000000e+09
2604                    Instagram  1.000000e+09
4234  Skype - free IM & video calls  1.000000e+09
9844                 Google News  1.000000e+09
4153                    Hangouts  1.000000e+09
4170                Google Drive  1.000000e+09
4144                     Google+  1.000000e+09
4150                      Google  1.000000e+09
```



---

# Insights and Key Findings

### Category Performance

- **High-Performing Categories**: `Health & Fitness`, `Education`, and `Entertainment` consistently scored higher in user satisfaction.
- **Saturated Categories**: `Family` and `Game` categories had the most apps, suggesting intense competition.

### Sentiment Trends

- Positive reviews dominated across categories.
- Neutral and negative reviews provided specific areas for improvement in user experience.

### Cost and Engagement

- Smaller apps generally performed better in terms of user ratings.
- Free apps drove higher engagement but lower revenue potential.

---

# Recommendations

### For Developers

1. Focus on high-performing categories like `Health & Fitness` and `Education`.
2. Optimize app size to improve performance and user satisfaction.

### For Marketers

1. Highlight positive reviews in advertising to boost installs.
2. Experiment with freemium models in saturated categories.

### For Product Teams

1. Address issues highlighted in Neutral and Negative reviews.
2. Use sentiment analysis to prioritize feature enhancements.

---

# Conclusion

This analysis of the Google Play Store dataset provided valuable insights into app performance, user sentiment, and market trends. By leveraging these findings, app developers and marketers can make data-driven decisions to enhance their products and strategies.