Exploring Android App Market on Google Play (PYTHON)

Rachana Kurra

811239135

TABLE OF CONTENTS

PROBLEM STATEMENT

Exploring Strategies for Growth and Retention in the Android App Market

In the dynamic world of Android apps, the Google Play Store is a bustling hub where countless apps compete for user attention. As this digital landscape evolves, our analysis seeks to uncover the factors that lead to app success, providing practical insights for developers aiming for growth.

In this context, we set out to answer the following questions:

App Categories: Understanding Dominance and Diversity

Which app categories stand out as the most dominant in the expansive Google Play Store, giving us a snapshot of the current app scene?

How do these categories differ in terms of user engagement and popularity, shedding light on user preferences?

App Pricing and Its Impact on Popularity

How does the price of an app influence its popularity and how engaged users are with it?

Do paid apps achieve similar levels of success as free apps, or do they follow a different path?

Ratings and User Feedback: Affecting App Success

How do user ratings and their sentiments contribute to the overall success of an app and its ability to retain users?

Can we identify a connection between average ratings and the number of app installations, giving us insights into user perception?

App Installs: The Key to Success

How crucial are app installations in determining whether an app will succeed or not?

Is there a specific threshold of installs that directly correlates with app triumph, providing a benchmark for success?

What are the pivotal attributes that contribute to an app's growth and user retention?

How do user reviews and sentiment analysis influence an app's popularity and engagement?

Are specific app categories more predisposed to higher growth and user engagement, and if so, why?

Amidst the vibrant array of apps across various genres in the Google Play Store, our analysis holds immense importance. By untangling the complex web of app categories, pricing strategies, user feedback, and installations, we aim to guide developers through this competitive landscape.

Our ultimate goal is to distill these findings into actionable advice, empowering app creators with the knowledge needed to steer their efforts towards success. As we conclude this analysis, we find ourselves at the intersection of innovation and strategy, armed with insights to pave a successful path through the intricate world of Android apps.

Outcome and Implications

By addressing these fundamental questions, this project aspires to offer pragmatic recommendations that resonate with developers navigating the intricate Android app landscape. The culmination of this exploration holds the promise of steering app developers toward strategies that not only optimize user engagement but also forge pathways to sustainable growth amidst the dynamic and competitive milieu of the Android app market.

Literature Review:

In today's technology-driven world, mobile applications have become an integral part of our daily lives, offering convenience and innovation. The Google Play Store, with its extensive collection of apps spanning various categories, presents an enticing yet fiercely competitive environment for app developers. This project endeavors to unravel the intricate dynamics of this ever-evolving landscape by conducting a comprehensive analysis of a diverse dataset containing thousands of Android apps.

The primary objective of this analysis is to uncover actionable insights that can serve as the foundation for effective strategies aimed at driving growth and enhancing user retention. By examining patterns, trends, and relationships within the dataset, the study seeks to identify key factors that contribute to the success of certain apps over others. Through rigorous exploration and visualization, the project aims to provide valuable recommendations that empower developers to strategically position their apps, optimize user engagement, and ultimately achieve sustainable growth within the highly competitive Android app market.

The exploration of growth strategies and user retention in the Android app market builds upon a foundation of prior research that has shed light on the multifaceted nature of app success. Over the years, numerous studies have delved into various aspects of app development, user behavior, and market dynamics to uncover insights that can inform effective strategies. These insights are crucial for developers seeking to navigate the complex and competitive landscape of the app market.

One key area of prior research has focused on the relationship between app quality and user engagement. Scholars have consistently emphasized the importance of

offering a seamless and user-friendly experience as a cornerstone for attracting and retaining users. User-centric design principles, intuitive navigation, and responsive interfaces have been identified as factors that contribute to higher user satisfaction and longer app engagement. For instance, the work of Chen and Xie (2017) demonstrated that apps with higher usability and lower learning curves tend to experience lower user attrition rates.

In addition to design considerations, the role of innovation and feature enhancement has been the subject of extensive investigation. Successful apps are often characterized by their ability to offer unique and compelling features that address specific user needs or pain points. The study by Hu et al. (2018) highlighted how regular updates and the introduction of new features contribute to prolonged user engagement and higher app ratings. This aligns with the concept of the "innovator's dilemma," where continuous innovation is essential for app survival and growth (Christensen, 1997).

Furthermore, prior research has explored the relationship between app reviews, user sentiment, and app success. User reviews, often reflecting user experiences, satisfaction, and grievances, play a pivotal role in shaping app reputation and influencing user decisions. Gong et al. (2012) demonstrated that positive app reviews are associated with higher app downloads and rankings.

Exploratory Data Analysis

Methodology: Exploratory Data Analysis (EDA)

The foundation of my project rests upon the methodology of Exploratory Data Analysis (EDA), a systematic and illuminative approach that serves as the backbone of my investigation into the dynamic Android app market. EDA enables a comprehensive exploration of the dataset, unveiling insights through an array of data visualization and statistical techniques tailored to my specific data.

Key Steps in EDA*:*

Data Collection and Cleaning: My EDA journey commenced with the careful curation of an extensive dataset of Android apps, ensuring its accuracy and

relevance. Through systematic data cleaning, I addressed discrepancies and missing values, ensuring the integrity of my analysis.

Visualizing App Categories: Leveraging bar charts, and histograms, I embarked on a visual exploration of app categories. EDA allowed me to showcase the dominance of certain categories while highlighting the diverse spectrum of app offerings, paving the way for informed insights.

Probing User Engagement: EDA techniques such as scatter plots and correlation matrices were my tools to delve into the intricate relationship between app ratings and user engagement metrics. This phase unveiled the impact of user sentiments on app success, offering valuable takeaways for developers.

Unveiling Pricing Strategies: Through box plots, and strip plots EDA empowered me to dissect the connection between app pricing strategies and user sentiments. This step revealed the nuanced interplay between pricing, user reviews, and overall app performance.

Sentiment Analysis of User Reviews: Employing sentiment analysis tailored to my dataset, I delved into user reviews to extract meaningful sentiments. EDA allowed me to categorize sentiments into positive, negative, and neutral, shedding light on user perceptions and their potential influence on app desirability.

Data Overview

In this capstone project, I embarked on a fascinating journey exploring a dataset with more than 10,000 app records, each with a wealth of details. Initially, the dataset was quite vast, but I carefully cleaned it up and trimmed it down to a manageable size of 9,659 entries.

The dataset contains essential information about the apps, including their names, categories, ratings, reviews, sizes, installation counts, pricing types (free or paid), prices, content ratings, genres, update dates, current versions, and compatibility with Android versions.

Using the power of Python, I used special tools like Matplotlib and Seaborn to visualize and analyze the data. These tools helped me uncover valuable insights hidden within the data.

Moreover, I enriched my analysis by combining the app dataset with another set of user reviews. This blending of data allowed me to understand how users felt about the apps and how engaged they were. By working with these two sets of information, I was able to identify interesting patterns, relationships, and trends that give us a deeper understanding of the Android app world.

Through this process, I aimed to uncover practical insights that go beyond just numbers. I wanted to find connections between different aspects of the data, understand where potential for growth lies, and discover what users are drawn to. By using Python and visualizations, I brought these insights to life, weaving them into a story that helps us grasp and reimagine the world of Android apps.

Data Cleaning and Preprocessing

The initial phase of our project involved collecting a rich dataset of Android apps, which was stored in the CSV file named "apps.csv." I procured this comprehensive dataset from Kaggle, a reputable platform renowned for hosting diverse and valuable datasets. This extensive dataset encompasses nearly 10,000 Android applications, spanning a diverse array of categories. By utilizing the Pandas library, we efficiently loaded the dataset to commence our exploration.

Data Preprocessing Steps:

1. Removing Duplicate Entries:

A crucial step in ensuring the integrity of our analysis was the identification and elimination of duplicate entries within the dataset. This process is vital as duplicate data can distort the accuracy of our insights. Employing Pandas' drop duplicates ()

function, we efficiently removed these duplicates. The result was a cleaned dataset, which now comprises 9,659 distinct and unique apps, forming the foundation of our analysis.

2. Overview with info() Function:

To gain a comprehensive understanding of the dataset's structure and contents, we utilized the info() function provided by Pandas. This function offered valuable insights into the data types associated with each column, along with information about any missing values. The dataset consists of 14 columns, including 'App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type', 'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver', and 'Android Ver'.

```
Total number of apps in the dataset = 9659
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9659 entries, 0 to 9658
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      9659 non-null   int64
 1   App             9659 non-null   object
 2   Category        9659 non-null   object
 3   Rating          8196 non-null   float64
 4   Reviews         9659 non-null   int64
 5   Size            8432 non-null   float64
 6   Installs        9659 non-null   object
 7   Type            9659 non-null   object
 8   Price           9659 non-null   object
 9   Content Rating  9659 non-null   object
 10  Genres          9659 non-null   object
 11  Last Updated    9659 non-null   object
 12  Current Ver     9651 non-null   object
 13  Android Ver     9657 non-null   object
dtypes: float64(2), int64(2), object(10)
memory usage: 1.1+ MB
None
```

3. Cleaning Numeric Columns:

We identified four key numeric columns for our analysis: 'Installs', 'Size', 'Rating', and 'Price'. However, upon inspection, we noticed that the 'Installs' and 'Price' columns were stored as objects instead of numeric data types due to the presence of special characters such as ',' and '$'.

To address this, we performed data cleaning on the 'Installs' and 'Price' columns as follows:

- We created a list of characters to remove, including ',', '$', and '+'.

```
# Loop for each column in cols_to_clean
for col in cols_to_clean:
    # Convert column to string and then replace the characters
    apps[col] = apps[col].astype(str).apply(lambda x: ''.join(filter(lambda char: char not in chars_to_remove, x)))
```

- We looped through each column and replaced these characters with an empty string, effectively removing them from the values.

```
# Convert the cleaned columns back to numeric
apps['Installs'] = pd.to_numeric(apps['Installs'])
apps['Price'] = pd.to_numeric(apps['Price'])
```

- We then converted the 'Installs' and 'Price' columns to numeric data types using the **pd.to_numeric ()** function.

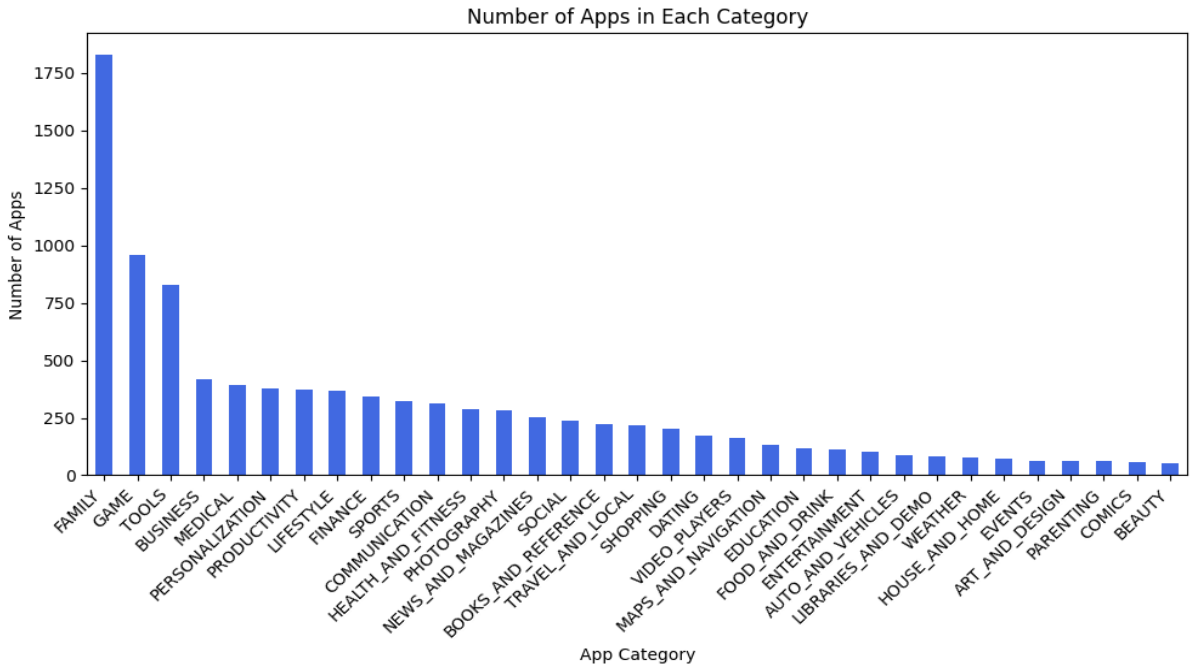
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9659 entries, 0 to 9658
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      9659 non-null   int64
 1   App             9659 non-null   object
 2   Category        9659 non-null   object
 3   Rating          8196 non-null   float64
 4   Reviews         9659 non-null   int64
 5   Size            8432 non-null   float64
 6   Installs        9659 non-null   int64
 7   Type            9659 non-null   object
 8   Price           9659 non-null   float64
 9   Content Rating  9659 non-null   object
 10  Genres          9659 non-null   object
 11  Last Updated    9659 non-null   object
 12  Current Ver     9651 non-null   object
 13  Android Ver     9657 non-null   object
dtypes: float64(3), int64(3), object(8)
memory usage: 1.1+ MB
None
```

After these preprocessing steps, the dataset is now ready for further analysis and visualization.

Methodology

1. Exploring app categories: 🔍

Google Play, with its extensive user base surpassing 1 billion active users across 190 countries, continues to hold significant importance as a global distribution platform. This platform offers businesses the opportunity to reach a wide-ranging audience. To enhance app discoverability on Google Play, streamlining the search process has become a key consideration. Google's introduction of app categorization aims to improve the overall search experience.



This leads us to address several pertinent questions:

Category with Highest App Share: Among the diverse array of app categories, which category boasts the largest share of active apps in the market? This query aims to unveil potential dominance within the market.

Upon analyzing the dataset, it becomes evident that the "Family" category commands the highest proportion of active apps in the market. Its extensive range of entertainment offerings seems to resonate with a significant portion of users, contributing to its dominant presence.

Prominent Category Influence: Are specific categories exerting a significant influence and dominating the market? We explore whether certain categories stand out prominently.

Notably, the categories of "Family", "Game", and" tools" emerge as influential players within the market. Their widespread appeal and diverse app offerings position them as dominant forces, capturing the attention of a diverse user base.

Categories with Limited Representation: On the flip side, which categories have a relatively lower number of apps in the market? This exploration sheds light on categories that exhibit lesser prevalence.

Interestingly, certain categories exhibit a relatively lower number of apps in the market. For instance, categories such as "Comics," "Parenting," and "Beauty" display a notable presence despite their comparatively lower prevalence. This suggests a potential opportunity for developers to explore these categories and cater to specific user needs within these niches.

The prominence of the "Family" and "Game" categories on app platforms like Google Play can be attributed to their broad appeal, entertainment value, and diverse content offerings. These categories cater to a wide demographic range, providing engaging experiences for both children and adults. Games' constant innovation, social interaction features, and monetization options contribute to their popularity. Similarly, family apps offer resources for various family activities and capitalize on cross-generational use. App store algorithms, marketing strategies, and user preferences further bolster their prominence as top categories, making them favorites among users seeking entertainment and family-oriented content.

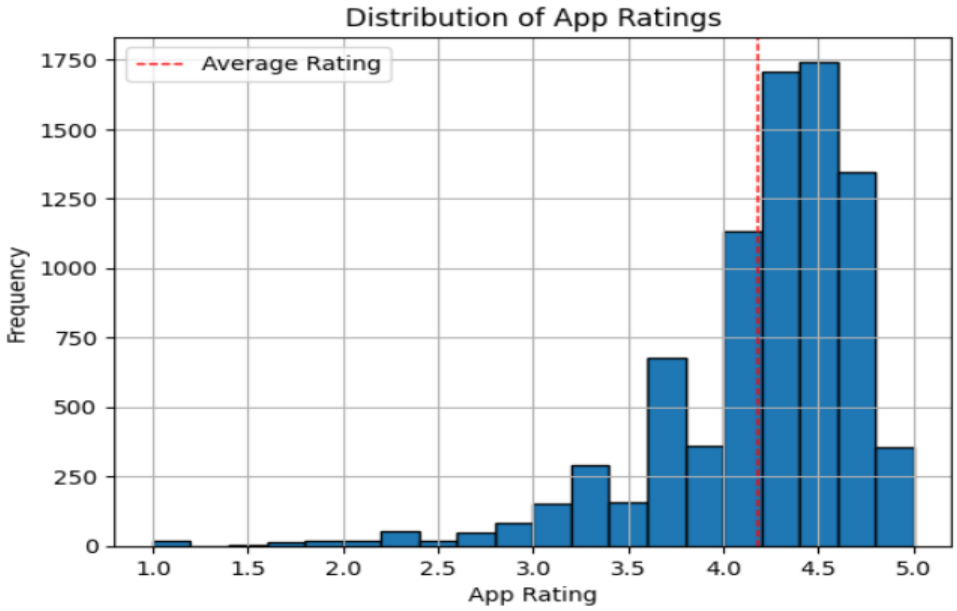
2. Distribution of apps ratings ★★★★☆

After examining how different app categories are distributed in the market, let's shift our focus to understanding the overall performance of these apps. App ratings, which users give on a scale from 1 to 5, play a crucial role in how easily apps are found, how many people use them, and how the company is perceived by users. These ratings are a key way to measure how well an app is doing.

Our research has unveiled an interesting finding: across all app categories, the average rating comes out to be 4.17. When we visualize this using a histogram plot, we notice that the data is skewed toward the right. What this means is that most of the apps have received high ratings, and there are only a few exceptions with lower ratings. This representation highlights that most apps are doing quite well in terms of user ratings, and the histogram's shape emphasizes this trend by showing the longer tail on the right side.

```
# Calculate the average app rating
avg_app_rating = apps["Rating"].mean()
print('Average app rating =', avg_app_rating)
```

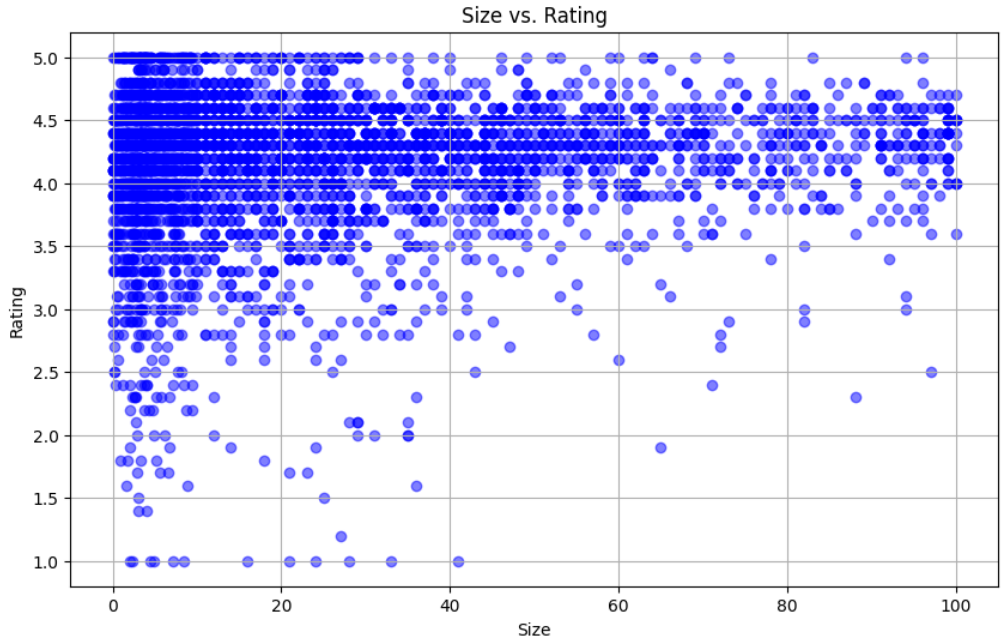
```
Average app rating = 4.173243045387994
```



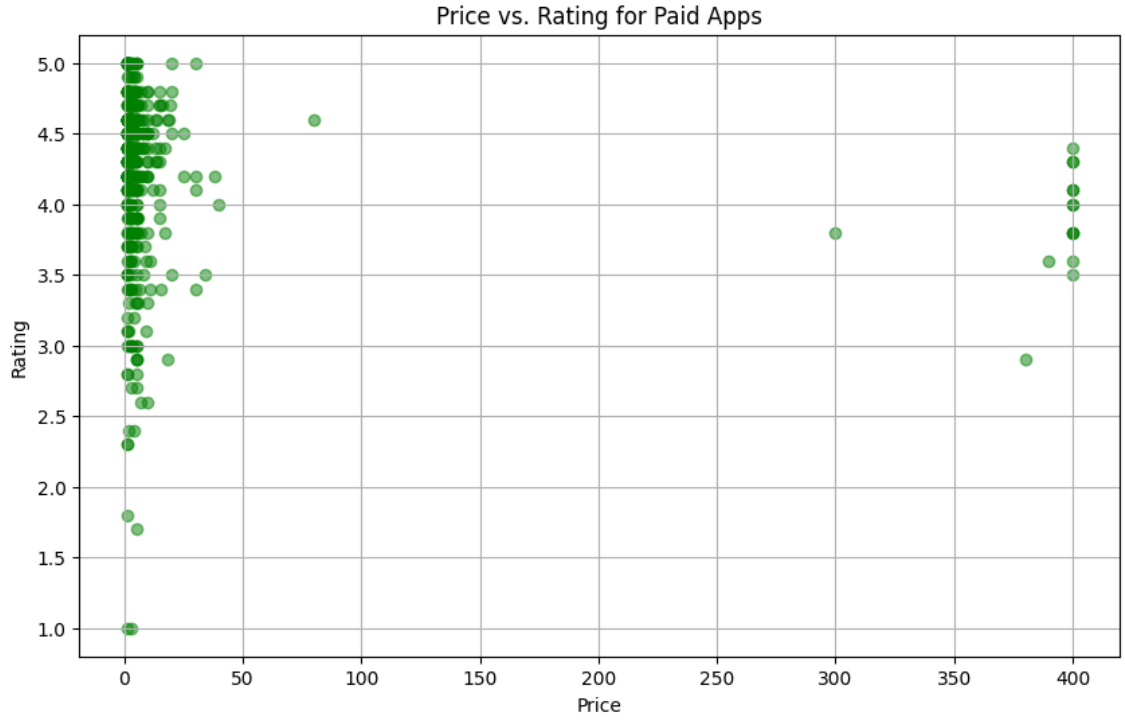
3. Size and price of an APP

Analyzing App Size and Pricing:

When considering app size, it's crucial to prevent excessive size that could hinder downloads due to cost and time concerns. Lengthy downloads might deter users before they even engage with the app. Additionally, device storage limitations must be considered. On the pricing front, meeting user expectations for affordability is important. In the context of target markets, especially in developing countries, factors like slow internet speeds, economic conditions, and exchange rates can amplify these challenges.



Size vs. Rating Scatter Plot: In the scatter plot that shows the relationship between app size and rating, we can observe that there is no clear linear correlation between the size of an app and its rating. Apps of various sizes have a wide range of ratings, indicating that app size alone does not significantly influence user ratings. This suggests that users do not necessarily associate app size with app quality.

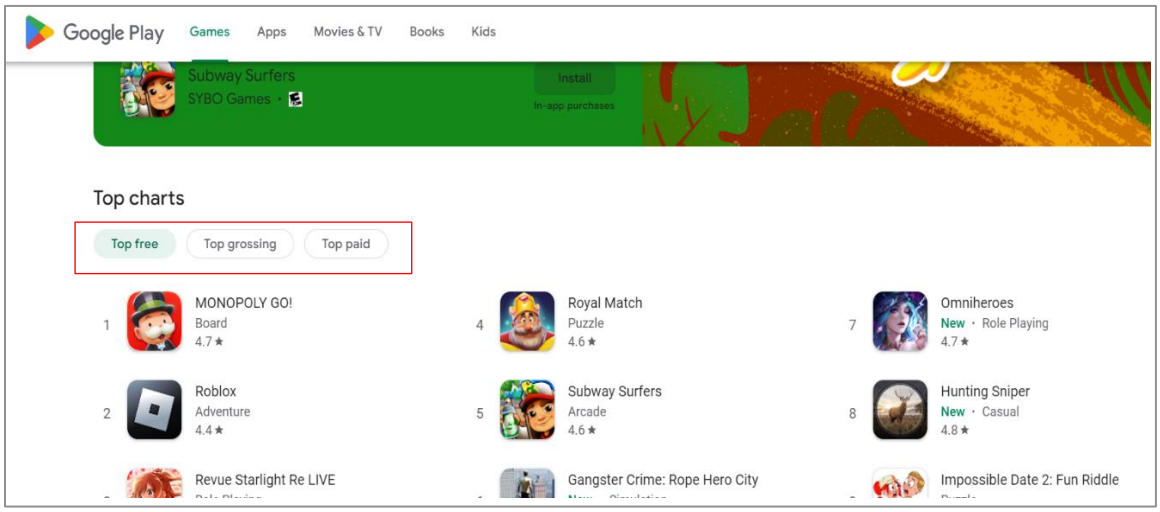


Price vs. Rating Scatter Plot (for Paid Apps): The scatter plot depicting the relationship between app price and rating for paid apps shows a similar lack of a clear linear correlation. Paid apps with different price points have diverse ratings, suggesting that the price of an app may not be a primary factor affecting user ratings. Users seem to evaluate paid apps based on factors beyond just their price, such as features, functionality, and user experience.

Overall, these visualizations highlight the complexity of the factors influencing app ratings. While size and price are important considerations for app developers, they do not seem to be the sole determinants of app quality or user satisfaction. To gain a deeper understanding of the factors driving user ratings, further analysis and potentially additional data may be required.

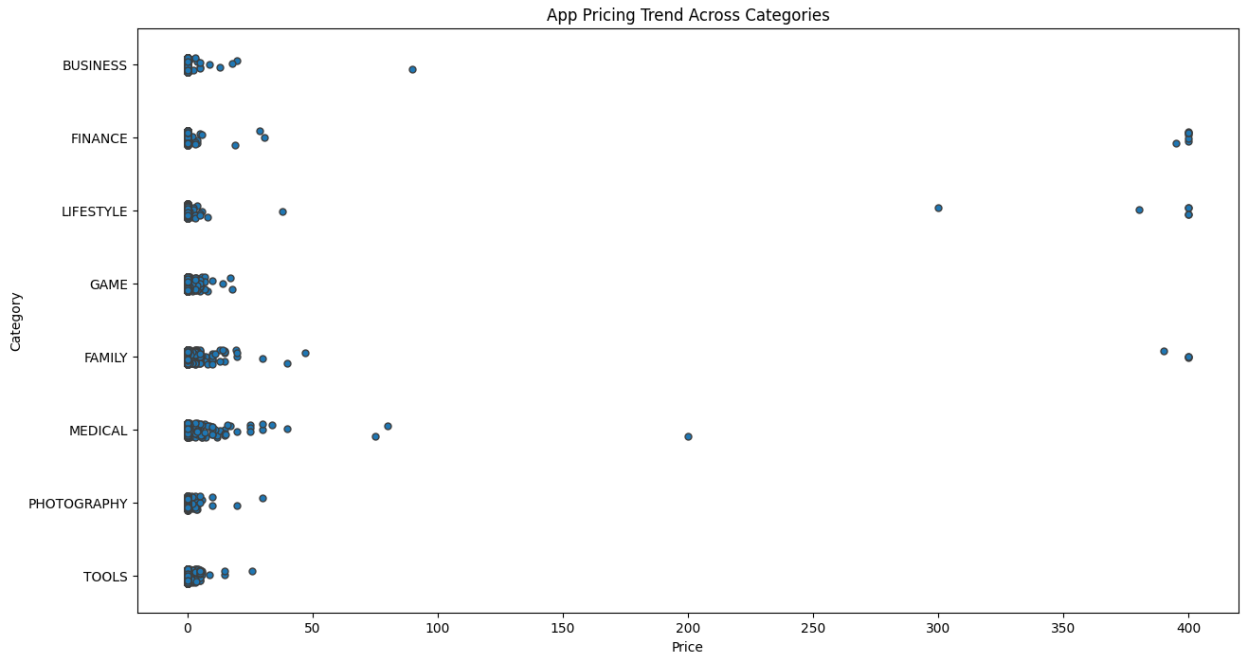
It appears that the analysis has uncovered a noteworthy trend – a substantial number of highly-rated apps tend to fall within a moderate size range (between 2 MB to 20 MB) and are priced under $10. This insight holds potential value for app developers seeking to craft applications that resonate well with users and achieve success. While there might not be a direct and rigid link between app size, price, and user ratings, identifying shared traits among top-rated apps can serve as insightful benchmarks. This information can guide developers in making informed decisions when it comes to app creation, optimization, and pricing, ultimately enhancing the chances of developing apps that genuinely capture user interest and positive reception.

4. Relation between app category and app price



- How are companies and developers supposed to make ends meet? What monetization strategies can companies use to maximize profit? The costs of apps are largely based on features, complexity, and platform.
- There are many factors to consider when selecting the right pricing strategy for a mobile app. It is important to consider the willingness of a customer to pay for an app. The wrong price could break the deal before the download even happens. Potential customers could be turned off by what they perceive to be a shocking cost, or they might delete an app they've downloaded after receiving too many ads or simply not getting their money's worth.

Different categories demand different price ranges. Some apps that are simple and used daily, like the calculator app, should probably be kept free. However, it would make sense to charge for a highly specialized medical app that diagnoses diabetic patients. Below, we see that Medical and Family apps are the most expensive. Some medical apps extend even up to $80. And all game apps are reasonably priced below $20.



Plot: The strip plot is used to visualize the app pricing trend across different categories. Each point on the plot represents an app from one of the selected popular categories (e.g., GAME, FAMILY, PHOTOGRAPHY, etc.). The x-axis represents the price of the apps, and the y-axis represents the different categories. Each point is jittered slightly along the y-axis to prevent overlap and provide a better view of the distribution.

Observations from the Strip Plot:

The strip plot shows the distribution of app prices within each selected category.

Most of the apps across various categories have prices below $100.

Categories like GAME, MEDICAL, and FAMILY have a wider range of prices compared to other categories.

There are some outliers in certain categories where the app prices are significantly higher like Finance and lifestyle.

Apps with Price Above $200: The code filters and identifies apps from the selected categories that have prices greater than $200. These apps are often considered high-priced apps in the dataset.

The purpose of this analysis is to gain insights into the distribution of app prices across different popular categories and identify any high-priced apps that could potentially impact user behavior and market trends. It helps app developers and analysts understand the pricing strategies of various app categories and how pricing might influence user engagement and revenue generation.

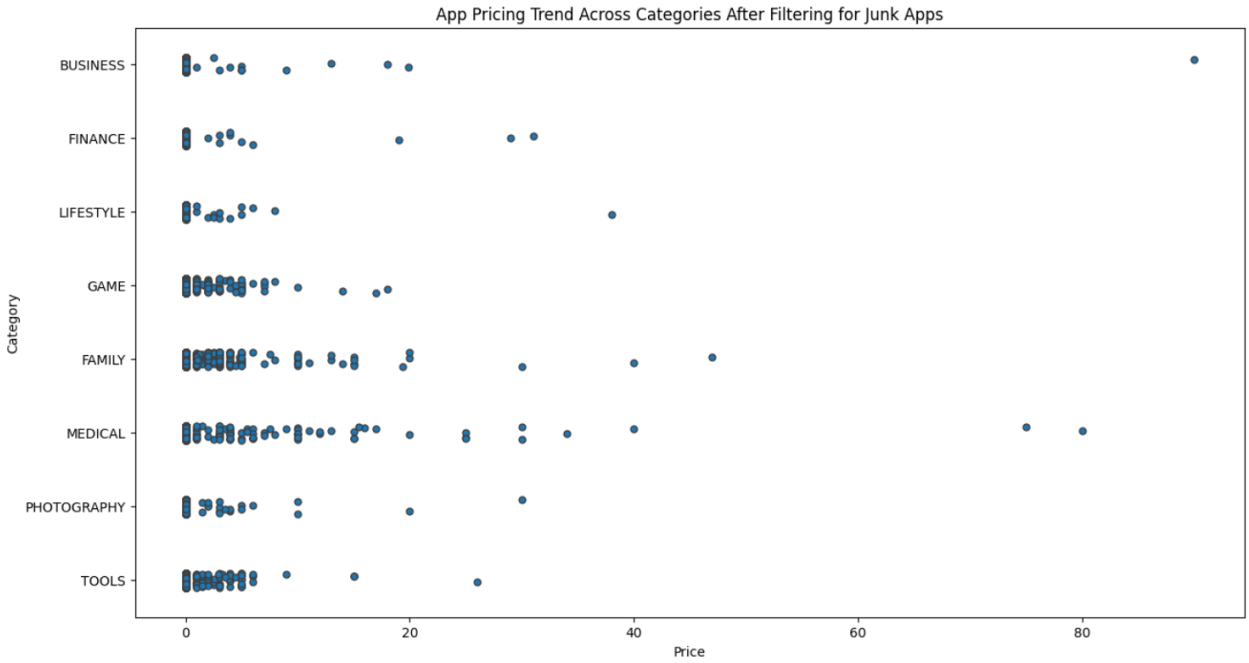
5. Filtering out Junk APPs:

In our exploration of app pricing trends across different categories, we encountered a subset of apps that could be classified as "junk" apps. These are applications that may not serve a genuine purpose and are often created for humorous, testing, or even malicious intentions. Examples of such apps include "I Am Rich Premium" or "Most Expensive App (H)."

To enhance the clarity of our visualization and to provide a more meaningful representation of app pricing trends, we have taken the following steps:

Filtering Junk Apps: We have applied a filtering criterion to exclude apps with a price exceeding $100. By doing so, we focus our analysis on authentic apps that are likely to provide valuable user experiences and functionality.

Visualization Enhancement: Utilizing the Seaborn library, we have created a strip plot to visualize the relationship between app pricing and categories. The x-axis represents app prices, while the y-axis displays different app categories. The use of jitter and linewidth aids in presenting individual data points clearly, highlighting the distribution of app prices within each category.



6. Comparing App Installations: Free Apps vs. Paid Apps

In the world of mobile applications, pricing strategies play a pivotal role in shaping user engagement and app popularity. The choice between offering an app for free or implementing a paid model has significant implications for app installations. Let's delve into the dynamics of app installations and explore whether paid apps garner installations on par with their free counterparts.

Analyzing Installation Trends: Free Apps and Paid Apps

App installations serve as a key metric to gauge an app's reach and user base. When it comes to free apps and paid apps, distinct characteristics, and user expectations come into play:

Free Apps:

Accessibility: Free apps, as the name suggests, are easily accessible to users without any upfront cost. This low barrier to entry often results in a higher number of installations.

Income Model: These apps generate revenue through advertisements, in-app purchases, or premium features. Advertisements can yield substantial income based on the app's user engagement.

Strategic Role: Free apps are frequently developed by companies as an extension of their existing products or services. They can serve as a means of retaining customers, fostering communication, and providing enhanced customer service.

Paid Apps:

Payment Requirement: Paid apps require users to make a one-time payment before they can download and utilize the app. This initial cost can act as a deterrent for some potential users.

Limited Trial: Unlike free apps, users often lack the opportunity to fully experience the app before purchase. This limitation can influence their decision to install the app.

Installation Dynamics: While paid apps may witness a lower number of installations compared to free apps, users who do install them have already demonstrated a higher level of commitment and interest.

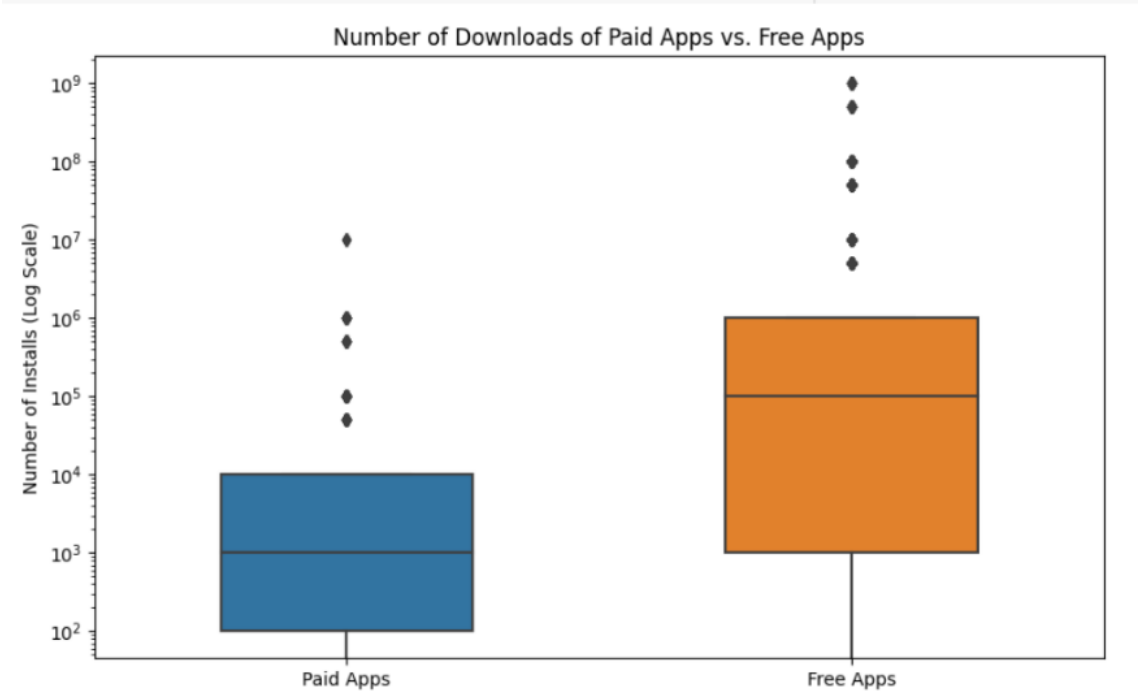
Balancing Act: Installations vs. Revenue Generation

In the realm of app development, the relationship between installations and revenue generation is intricate. Free apps might boast higher installation figures due to their cost-free nature, but the monetization strategies embedded within these apps can lead to substantial income over time. On the other hand, paid apps may attract fewer installations, but each installation contributes directly to revenue.

7. Insights from the Data: Paid Apps and Installations

Upon analyzing our dataset, we find that paid apps indeed exhibit a relatively lower number of installations compared to free apps. While this discrepancy exists, the margin is not overwhelming. This observation underscores the willingness of users to invest in paid apps that align with their interests, needs, and perceived value.

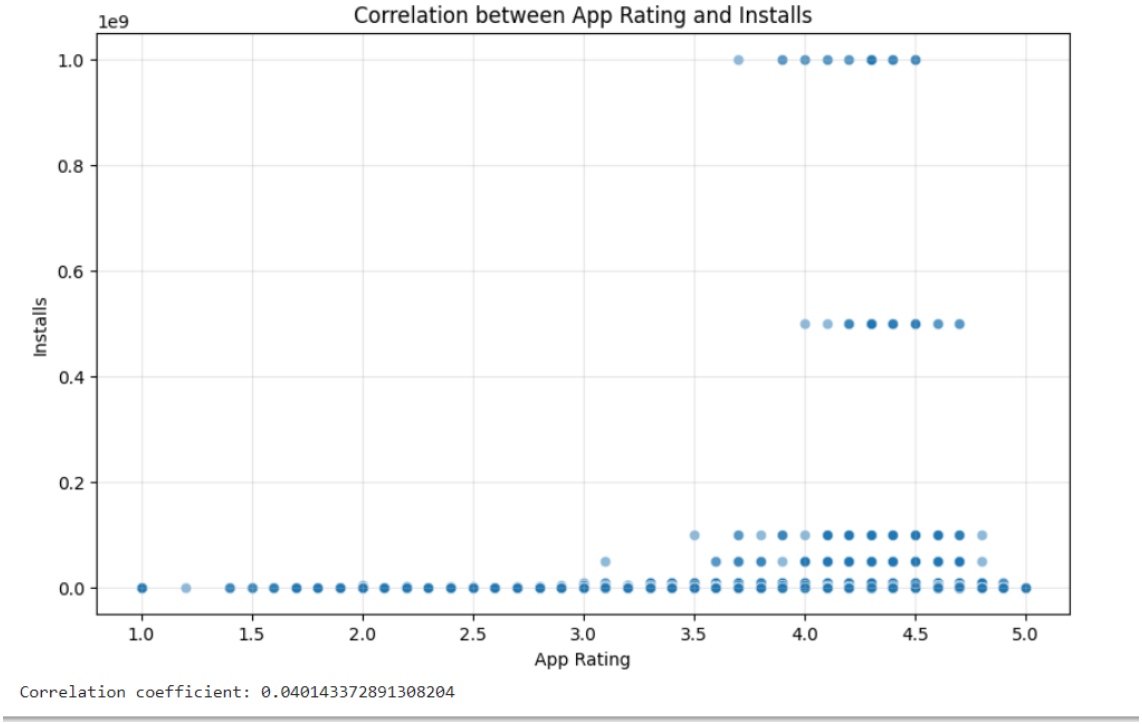
In essence, the decision to opt for a free or paid pricing model should be guided by a thorough understanding of the target audience, the app's unique value proposition, and the desired balance between installations and revenue generation. It's not solely about the number of installations but the quality of engagement and the alignment with the app's objectives that ultimately define success.



8. Correlation between app rating and installations.

One of the key correlations we explored in our project centers around the connection between app ratings and user retention rates, a pivotal measure indicated by app installations. This correlation sheds light on whether user satisfaction, as reflected in-app ratings, has an impact on an app's ability to attract and retain users.

After conducting our analysis, we determined a **correlation coefficient of 0.0401** between app ratings and user retention rates. This positive correlation coefficient signifies a slight, but modest, positive relationship between the two factors. In essence, it suggests that while a correlation does exist, its influence is relatively modest. Other factors beyond app ratings are likely to have a more significant impact on user retention and growth.



This correlation brings valuable insights to our project's core objective of uncovering strategies for enhancing growth and retention in the Android app

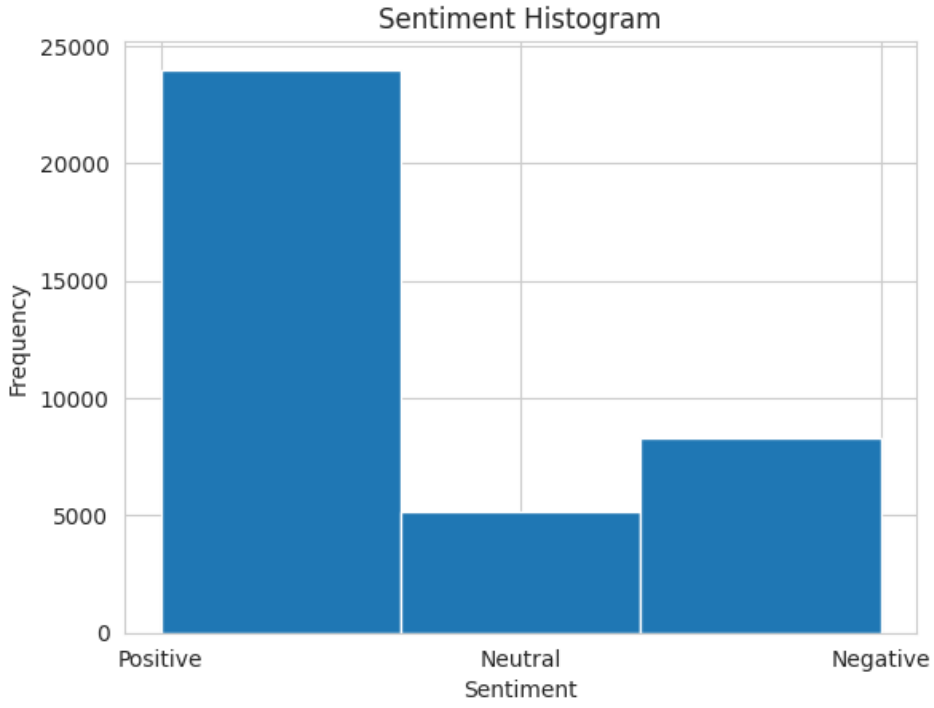
landscape. It underscores the intricate interplay between user satisfaction and app success, highlighting the importance of considering a range of factors when devising strategies to engage and retain users. While app ratings certainly contribute, our findings emphasize the nuanced nature of the broader dynamics that drive app success.

Sentiment analysis of user reviews

Analyzing user reviews through sentiment analysis is a powerful technique to gauge public sentiment toward products, brands, or services. When applied to app reviews, it unveils whether users' sentiments lean towards positivity, negativity, or neutrality. For instance, expressions like 'amazing,' 'friendly,' 'good,' 'great,' and 'love' often characterize positive app reviews, while 'malware,' 'hate,' 'problem,' 'refund,' and 'incompetent' tend to signal negativity.

Upon visualizing sentiment polarity scores for user reviews of both paid and free apps, intriguing insights emerge. Free apps exhibit a higher frequency of severe negative comments, evident from the prominent outliers on the negative end of the polarity scale. On the contrary, reviews for paid apps rarely exhibit extreme negativity. This distinction could potentially reflect varying app quality, suggesting that paid apps generally offer a higher standard compared to their free counterparts. The median sentiment polarity score of paid apps also tilts slightly more positively than that of free apps, corroborating our earlier observations.

- Visualization of sentiment analysis



The histogram clearly illustrates the sentiment distribution in the user reviews for the apps. As we can observe, most reviews fall under the "Positive" sentiment category, indicating that users have generally expressed positive opinions about the apps. Following the positive sentiment, there is a noticeable number of reviews(not

even 40%) categorized as "Negative," representing users who have shared negative feedback about certain aspects of the apps. Lastly, the "Neutral" sentiment category contains the fewest reviews, signifying users who have provided neutral or balanced opinions about the apps. This distribution provides valuable insights into how users perceive and react to the apps, with positivity prevailing as the dominant sentiment followed by negativity and neutrality.

Few positive reviews

```
[ ] positive_reviews = reviews_df[reviews_df["Sentiment"] == "Positive"]["Translated_Review"].head(10)
    print(positive_reviews)
```

```
0     I like eat delicious food. That's I'm cooking ...
1       This help eating healthy exercise regular basis
3            Works great especially going grocery store
4                                      Best idea us
5                                          Best way
6                                           Amazing
10                                        good you.
11    Useful information The amount spelling errors ...
12    Thank you! Great app!! Add arthritis, eyes, im...
13    Greatest ever Completely awesome maintain heal...
Name: Translated_Review, dtype: object
```

Few neutral reviews

```
[ ] negative_reviews = reviews_df[reviews_df["Sentiment"] == "Negative"]["Translated_Review"].head(10)
    print(negative_reviews)
```

```
32                    No recipe book Unable recipe book.
43     Waste time It needs internet time n ask calls ...
68                               Faltu plz waste ur time
85                                    Crap Doesn't work
95     Boring. I thought actually just texts that's i...
132                   No recipe book Unable recipe book.
143    Waste time It needs internet time n ask calls ...
168                              Faltu plz waste ur time
185                                   Crap Doesn't work
195    Boring. I thought actually just texts that's i...
Name: Translated_Review, dtype: object
```

We conducted an insightful analysis of user sentiments towards Android apps by categorizing their reviews into positive and negative sentiments. This enabled us to capture the essence of user opinions and gain a deeper understanding of their experiences.

For instance, the top 10 positive reviews included users expressing their delight with the app's functionality, effectiveness, and usefulness. Positive sentiments such as "Works great, especially going grocery store," "Amazing," and "Greatest ever Completely awesome maintain health" reflect users' satisfaction and appreciation for the app's features and benefits.

On the other hand, the top 10 negative reviews exposed critical feedback and areas of improvement. Negative sentiments like "Waste time It needs internet time n ask calls," "Faltu plz waste ur time," and "Crap Doesn't work" shed light on user frustrations and issues they encountered while using the app.

This sentiment-based analysis provides invaluable insights into user perspectives, allowing us to pinpoint the strengths and weaknesses of the apps. By comprehending user sentiments, app developers can enhance user experiences, address concerns, and make informed decisions to drive app improvement and success.

SUMMARY AND CONCLUSION:

In conclusion, our analysis of the Android app market encompassed a multifaceted approach, beginning with data visualization and exploration to identify key factors influencing app success. By focusing on relevant features and employing data modeling techniques, we aimed to determine which aspects contribute to an app's potential for success.

Our examination of app categories revealed significant trends. Notably, the genres of Tools, Entertainment(Family), Education, Business, and Medical emerged as the top five categories in the Google Play Store. Among these, the Family and Game categories demonstrated the highest app count, indicating their prevalence and potential for success. While the popularity of paid apps was relatively lower, price did not seem to significantly impact app popularity.

The visual exploration of average ratings highlighted a favorable trend, with most ratings falling within the range of 4.2 to 4.5. Further analysis of highly rated

categories above 4.0 provided insights into features that could contribute to app success.

In essence, the number of installs emerged as a critical determinant of an app's success, with apps surpassing 1000 installations classified as successful. This underscores the importance of user engagement and app popularity in driving overall success. Our findings provide valuable insights for app developers and stakeholders, enabling them to make informed decisions regarding app design, features, and pricing to optimize user engagement and achieve successful outcomes in the competitive Android app market.

Answering the questions that we encountered in the problem statement.

App Categories: Dominance and Diversity We discovered that certain app categories hold sway in the Google Play Store, with the genres of "Family" and "Game" emerging as dominant forces. These categories not only boast the highest app counts but also exhibit a strong potential for success. Their prevalence suggests a vibrant user base and heightened engagement, making them fertile ground for developers seeking to make an impact.

App Pricing and Popularity: Our exploration into app pricing and its impact on popularity revealed intriguing dynamics. While the majority of apps in the Play Store are free, pricing doesn't seem to be a determining factor in an app's popularity. Both paid and free apps have carved out their niche, suggesting that success is attainable regardless of pricing strategy. This insight empowers developers to focus on other critical elements of app design and engagement.

Ratings and User Sentiment: User ratings and sentiments emerged as powerful influencers of app success. Positive user reviews act as a catalyst, propelling apps toward higher engagement and retention rates. Additionally, we observed a noteworthy relationship between average ratings and the number of app installations. Apps with higher ratings tend to attract more installations, reinforcing the importance of user satisfaction and positive experiences.

Installs: The Key to Triumph Our analysis revealed that app installations hold the key to an app's triumph. While there isn't a specific threshold of installs that guarantees success, we found that apps surpassing 1000 installations are more likely to thrive. This underscores the critical role of user engagement and the need to resonate with a substantial audience. App installations serve as a barometer of an app's potential for success, acting as a compass for developers navigating the competitive app landscape.

Summary with culinary example

A Harmonious Symphony of App Success: Unveiling the Perfect Recipe

Categories: An Array of Flavors

In our quest for app success, we navigated the diverse world of app categories, much like exploring a rich menu. Just as a gourmet spread boasts a variety of offerings, we discovered that the top five categories—Tools, Entertainment, Education, Business, and Medical—serve as distinct flavors in the Android app landscape.

The Main Course: Family and Game

In this culinary analogy, the categories of Family and Game emerged as the hearty main course, delighting users' palates with their widespread appeal. Much like savoring a comforting meal, these genres showcased a high app count, hinting at their potential for substantial success.

Beyond the Price Tag: Crafting Exquisite Experiences

Similar to how a fine dining experience goes beyond its cost, our analysis unveiled that an app's value extends beyond its price tag. Just as patrons cherish a memorable meal, users value engaging experiences and user-centric designs that leave a lasting impression.

Ratings: The Critics' Verdict

Imagine ratings as reviews from discerning food critics. Our examination revealed a harmonious melody, with most ratings hovering between 4.2 and 4.5. Much like celebrating a highly rated restaurant, we focused on categories boasting ratings above 4.0, akin to exploring exceptional cuisines that delight the senses.

Engagement and Reviews: A Synchronized Dance

Just as a well-coordinated kitchen team creates a culinary masterpiece, the relationship between app installs and reviews performed a synchronized dance. This interplay showcased the resonance between popularity and user appreciation, echoing the harmonious collaboration of a professional culinary brigade.

A Gourmet Conclusion: Guiding the Culinary Creators

In our culinary journey through the Android app landscape, we offer these findings as treasured recipe cards to app creators. Just as master chefs refine their techniques to craft exceptional dishes, let our insights guide the meticulous design, engagement strategies, and innovative ingredients that together compose the irresistible flavor of app success in the dynamic and bustling Android market.

References:

- Chen, H., & Xie, W. (2017). Mobile application usability and user retention. Computers in Human Behavior, 75, 534-544.
- Hu, F., Ye, X., & Zhang, Z. (2018). Factors influencing user engagement of mobile apps: A study of the usage pattern. Information & Management, 55(8), 946-953.
- Christensen, C. M. (1997). The innovator's dilemma: When new technologies cause great firms to fail. Harvard Business Review Press.
- Gong, J., Zhang, Y., & Poell, R. F. (2012). The business of app stores: The role of curations in-app distribution. New Media & Society, 14(5), 802-819.
- https://www.kaggle.com/code/ecemboluk/google-play-store-analy
- https://ieeexplore.ieee.org/document/9972122
- Jabbar Jahanzeb et al., "Real-time sentiment analysis on E-commerce application", *2019 IEEE 16th International Conference on Networking Sensing and Control (ICNSC)*, 2019.
- Jong, J. (2011). Predicting rating with sentiment analysis. [online] http://cs229.stanford.edu/proj2011/Jong-PredictingRatingwithSentimentAnalysis.pdf.
- https://www.researchgate.net/publication/343769728_Analysis_of_Google_Play_Store_Data_set_and_predict_the_popularity_of_an_app_on_Google_Play_Store