Navigating the Realm of Google Play Store Apps: Discovering Hidden Trends and Insights

As a junior data scientist, I find myself at the intersection of data and discovery, poised to explore datasets that hold the promise of revealing hidden gems of knowledge.

Today, I invite you to join me on a captivating journey as we navigate through a dataset carefully curated from the depths of the data-rich landscape at Kaggle.

This particular dataset, extracted from the intriguing realm of Google Play Store Apps, is an invitation to unravel the mysteries that shroud the world of mobile applications.

With over millions of apps available on the Google Play Store, this dataset provides a unique opportunity to analyze and understand the trends, preferences, and user behavior in the mobile app market. By delving into this dataset, we can gain valuable insights that can shape future app development strategies and enhance user experiences. So, let's embark on this exciting adventure together and uncover the untapped potential hidden within the world of Google Play Store Apps.

Introduction

In this scientific study, we examine and investigate a dataset called “googleplaystore.csv” that includes information on numerous apps accessible on the Google Play Store. Information such as:

Top Paid Apps in the Family Category: Using a bar diagram, we highlight the most popular paid apps in the Family category, giving a clear picture of their installation levels.

Genres in Focus: Installs Analysis: Using a pie chart, we display the distribution of premium family applications across various genres, demonstrating which genres receive the most installs.

Installations of Apps by Category: We offer an array of the number of installs for each app category, giving us a good idea of where user interest is.

A pie diagram on user prefrences: depicts the distribution of installs across various app categories, offering information on user preferences.

Price Comparison by Mean Across Categories: We compare the average pricing of applications using a bar chart. across several categories, providing insights on expense patterns.

Exploring High-Value applications: We look into the most costly applications in each category, providing a glimpse into premium features.

Assumptions

Data We Can Rely On: Our journey begins with the understanding that the dataset we're using accurately represents the different aspects of apps on the Google Play Store. This forms the foundation for everything we'll discover.

Real People, Real Opinions: As we dive into the dataset, we're considering user reviews as genuine reflections of how people feel about these apps. This helps us get a sense of whether users are happy with what they're using.

Getting the Picture from Ratings: We're looking at user ratings as clues about whether an app is good or popular. These ratings give us hints about what people like and what's working well.

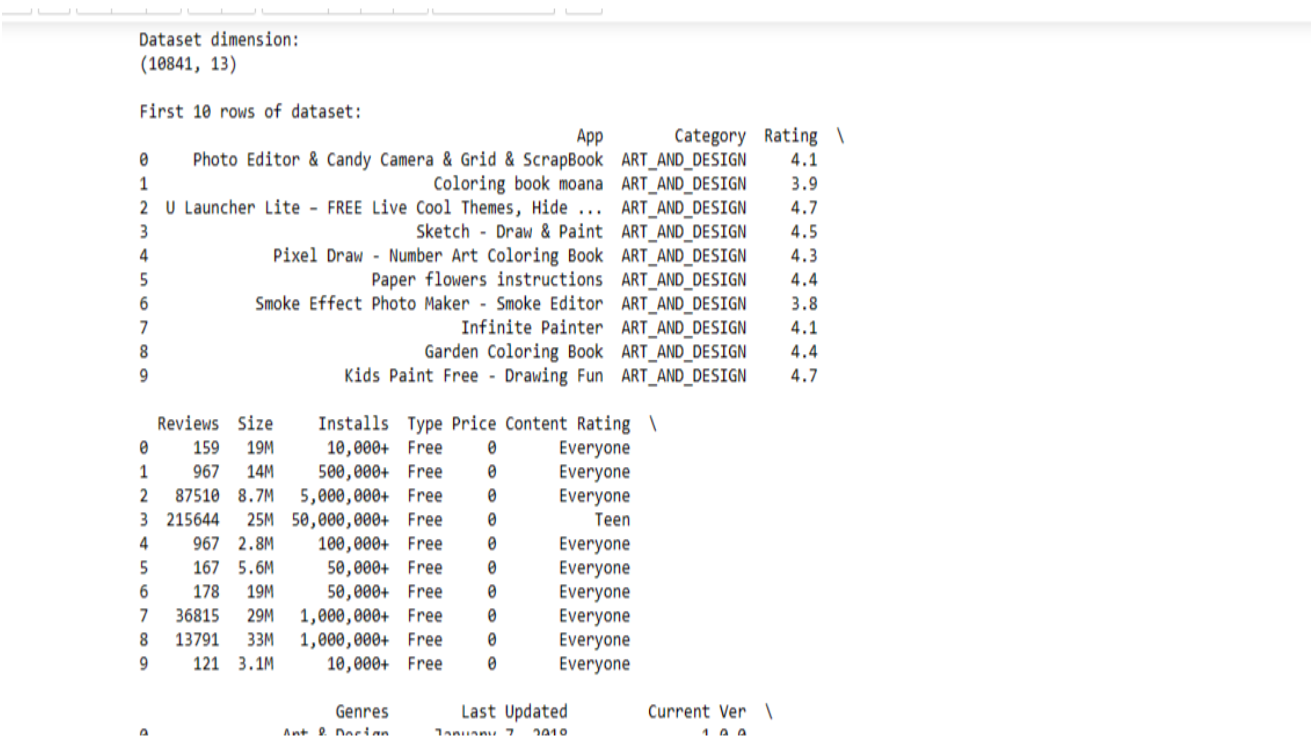
Seeing Different App Types: The way we're looking at things relies on apps being grouped into categories that make sense. This helps us understand each type of app better and see what sets them apart.

Time Matters: Our insights are shaped by considering that the dataset represents a certain period in time. This helps us understand the dataset; by analyzing it, we can gain valuable insights into the past and make informed decisions for the future. trends and behaviors that were happening during that specific time.

Data Summary

Before we dive into the details, we took a step back to get a sense of what our dataset is all about. We did this by summarizing it in two ways:

Dimensions of the Dataset: Imagine the dataset as a big table with rows and columns. We counted how many rows and columns there are. This helps us understand the size of our data, like the size of a puzzle. We concluded that our data had a dimension of 10841 rows and 13 columns. "Please refer to the image provided below for a visual representation."



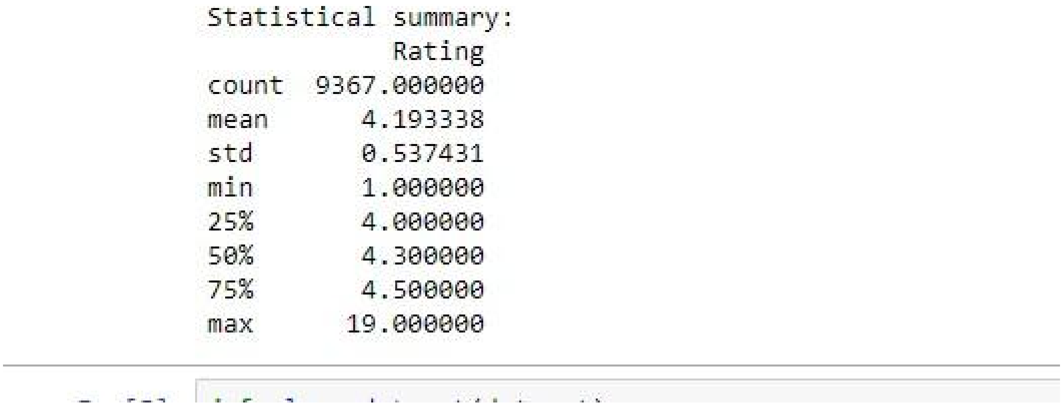
```
Dataset dimension:
(10841, 13)

First 10 rows of dataset:
                                                 App        Category  Rating  \
0     Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                                Coloring book moana  ART_AND_DESIGN     3.9
2  U Launcher Lite – FREE Live Cool Themes, Hide ...  ART_AND_DESIGN     4.7
3                              Sketch - Draw & Paint  ART_AND_DESIGN     4.5
4              Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3
5                         Paper flowers instructions  ART_AND_DESIGN     4.4
6            Smoke Effect Photo Maker - Smoke Editor  ART_AND_DESIGN     3.8
7                                   Infinite Painter  ART_AND_DESIGN     4.1
8                               Garden Coloring Book  ART_AND_DESIGN     4.4
9                      Kids Paint Free - Drawing Fun  ART_AND_DESIGN     4.7

  Reviews  Size     Installs  Type Price Content Rating  \
0     159   19M      10,000+  Free     0       Everyone
1     967   14M     500,000+  Free     0       Everyone
2   87510  8.7M   5,000,000+  Free     0       Everyone
3  215644   25M  50,000,000+  Free     0           Teen
4     967  2.8M     100,000+  Free     0       Everyone
5     167  5.6M      50,000+  Free     0       Everyone
6     178   19M      50,000+  Free     0       Everyone
7   36815   29M   1,000,000+  Free     0       Everyone
8   13791   33M   1,000,000+  Free     0       Everyone
9     121  3.1M      10,000+  Free     0       Everyone

                       Genres        Last Updated         Current Ver  \
0                Art & Design     January 7, 2018               1.0.0
```

Statistical Snapshot: Think of our dataset like a group of friends. We took a peek at their average height, their tallest and shortest members, and how much they vary in height. In the same way, we calculated the average, minimum, maximum, and spread of some numbers in our dataset. This snapshot gives us a quick idea of what we're working with.



```
Statistical summary:
             Rating
count  9367.000000
mean      4.193338
std       0.537431
min       1.000000
25%       4.000000
50%       4.300000
75%       4.500000
max      19.000000
```

These summaries are like a sneak peek before we dig deeper into our analysis. Just like looking at a movie trailer to get a taste of what's to come!

Data Cleaning

To make sure the information we're using is reliable, we went through a process called data cleaning. Here's what we did:

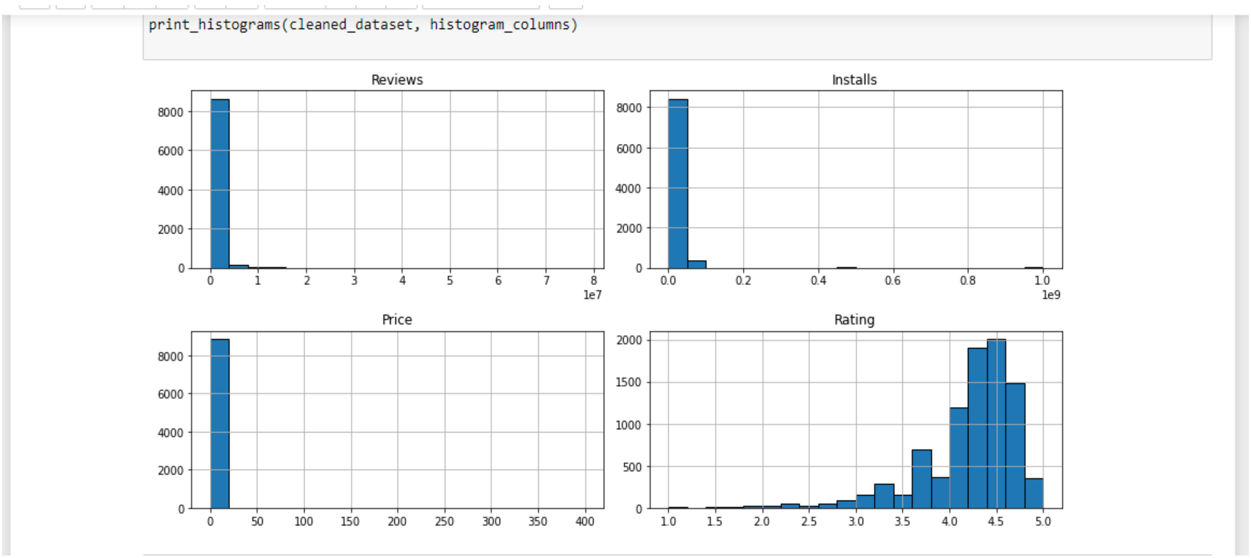
We made sure there were no repeated rows in the dataset.

We got rid of any rows that were missing important information.
We worked on the 'Installs' column. We removed extra symbols like '+' and ',' and made the numbers easier to understand.
The 'Price' column got a similar treatment. We took out the '$' sign and made sure the numbers made sense as decimals.
The numbers in the 'Reviews' column were changed to whole numbers so they were easier to work with.

By doing these steps, we're getting the data in good shape for our analysis. It's like cleaning up a room before we start decorating!

Analysis of Exploratory Data

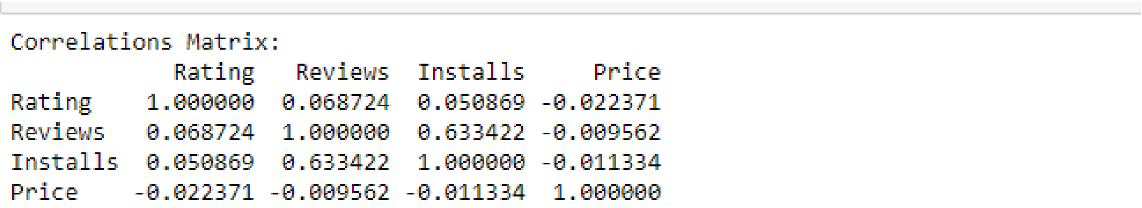
To present the range of numerical parameters (reviews,' installations,' price,' and rating),' we create histograms. These histograms enable us to determine the range of values and variations of each characteristic.



```
print_histograms(cleaned_dataset, histogram_columns)
```

Analysis of The relationship

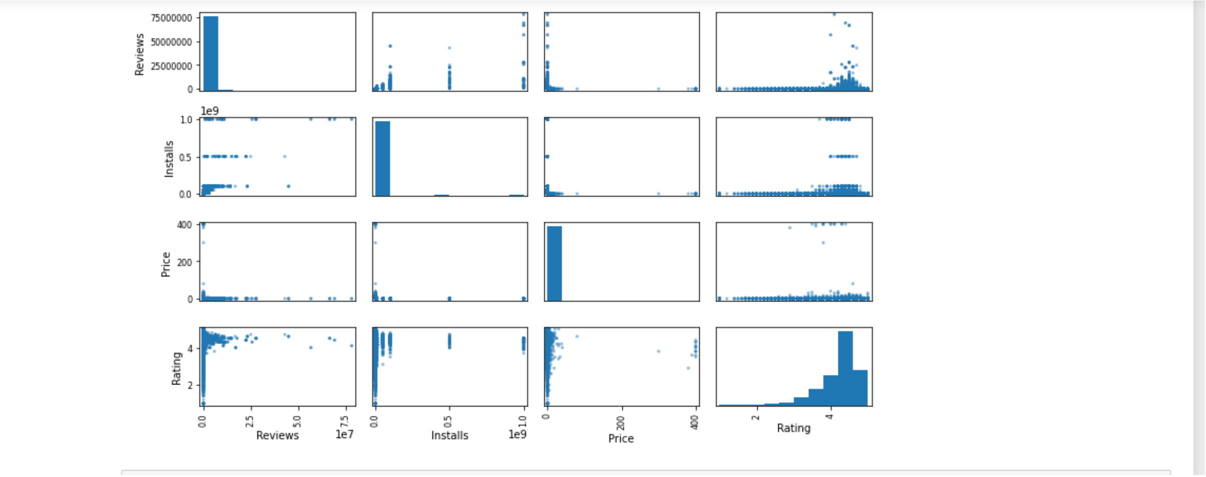
Using the compute_correlations_matrix() function, we determine the Pearson correlation factors among numerical attributes. This technique lets us look into correlations between defining features.



```
Correlations Matrix:
            Rating   Reviews  Installs     Price
Rating    1.000000  0.068724  0.050869 -0.022371
Reviews   0.068724  1.000000  0.633422 -0.009562
Installs  0.050869  0.633422  1.000000 -0.011334
Price    -0.022371 -0.009562 -0.011334  1.000000
```

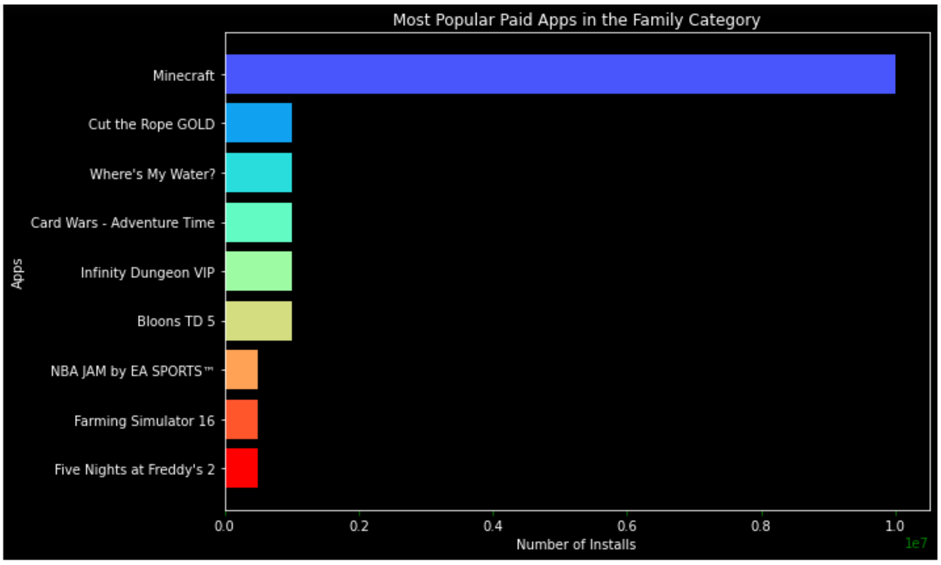
Network of scatter

The scatter matrix, which can be obtained by the print_scatter_matrix() function, is a matrix of scatter plots. These visualizations aid in seeing the interplay of numerical properties.



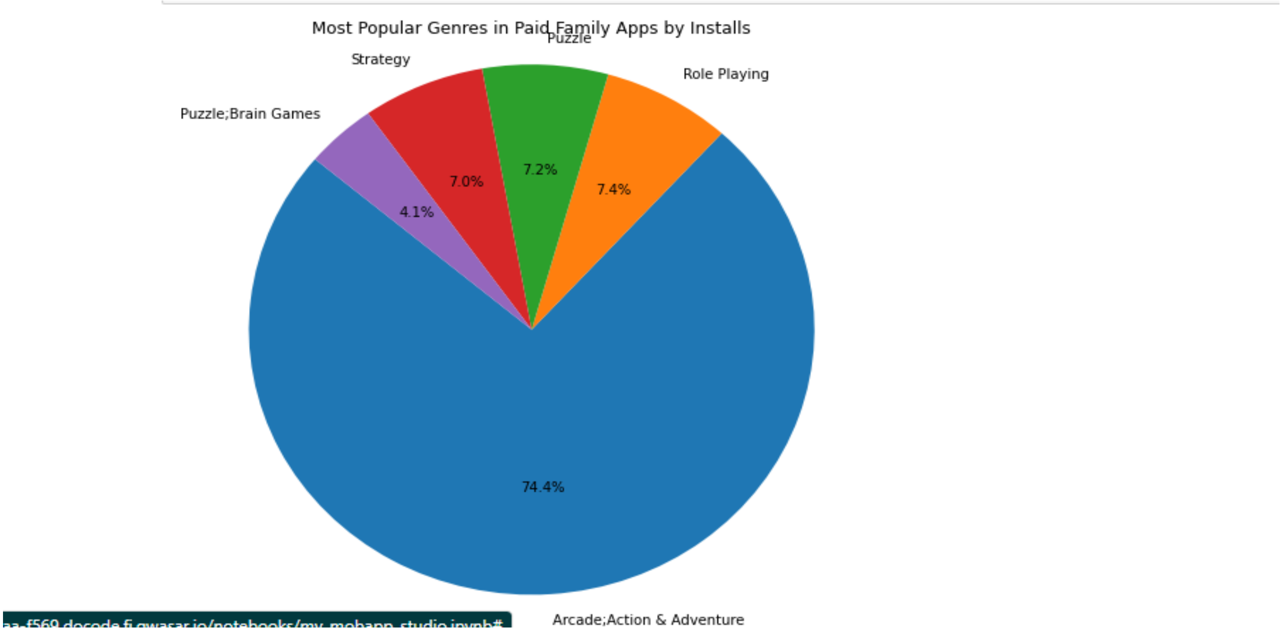
Analysis of Popular Paid Apps in the Family Category

We examined paid applications in the 'Family' category. We selected the top ten paid applications based on installations and showed their popularity with a horizontal bar plot. This graph assists us in determining which paid family applications are the most popular among customers.



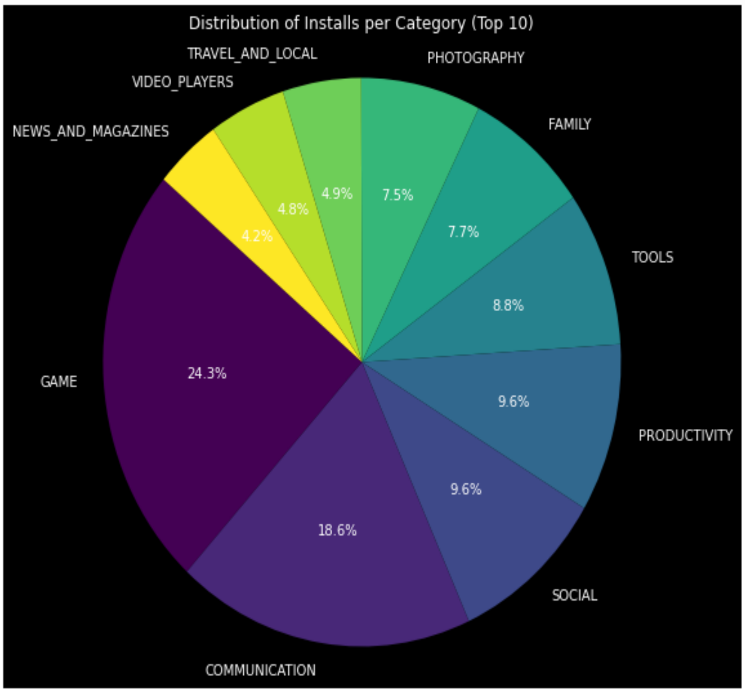
An Investigation into the Most Popular Genres in Paid Family Apps

We looked into the family category further and discovered the most popular genres inside paid family applications. We generated a pie chart that depicts the distribution of installs in the family category across the top genres. The pie chart shows the distribution of installs among the top genres in the 'Family' category.



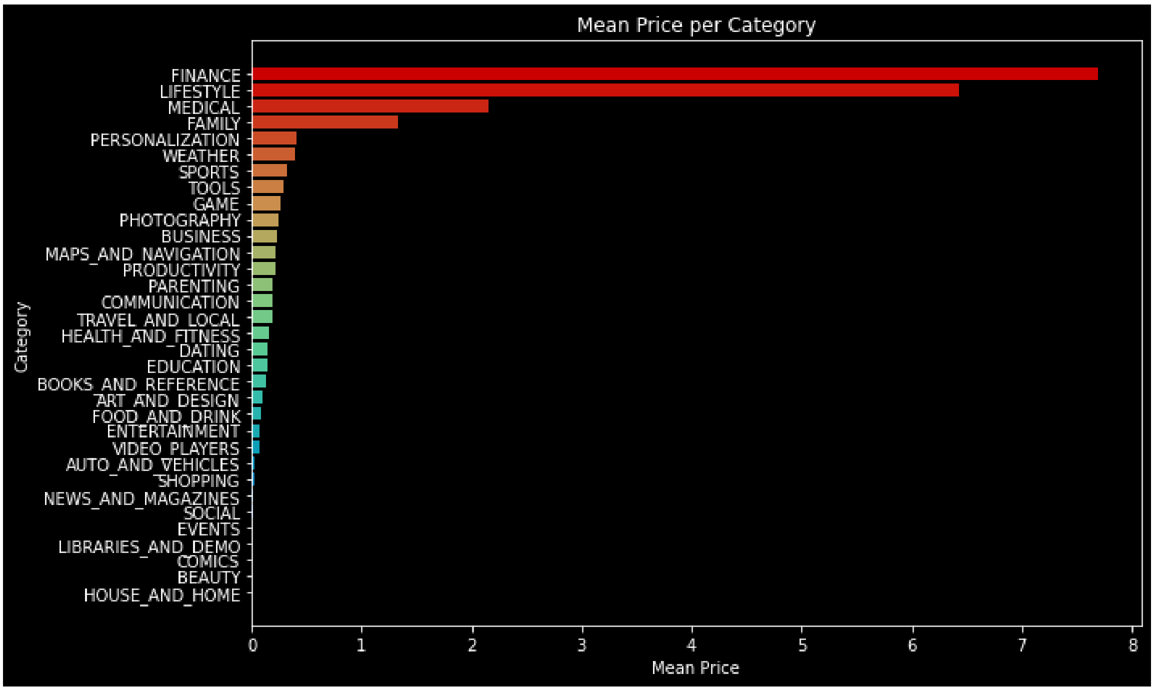
Installs by Category Analysis

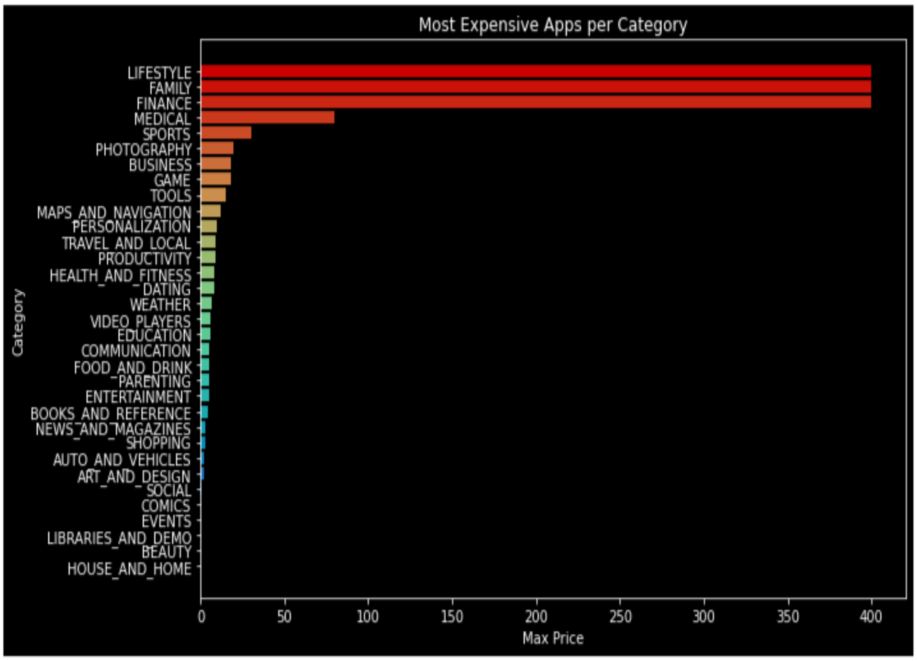
In our analysis of app installations by category, we went a step further by categorizing apps and calculating the total number of installations within each category. Our findings revealed intriguing insights: the game category emerged as the leader with the highest installations, closely followed by the communication and productivity categories. This array of installations across categories highlights user preferences and gives us a glimpse into the app landscape's dynamics.



Analysis of the Mean Price by Category

We estimated the average price of applications in each category and ordered the results in descending order. The investigation revealed that applications in the medical and family categories have higher mean prices. These are snapshots that show the mean price per category and the most expensive app per category analysis:





This visual shows the most expensive apps per category.

Conclusion

The Google Play Store app dataset study revealed information on app categories, user preferences, pricing methods, and other topics. Exploration of histograms, correlation matrices, scatter plots, and visualizations of popular applications and genres helped us better identify patterns and trends within the dataset.

The findings might be useful for app developers, marketers, and decision-makers wanting to enhance their app strategy based on user behavior and preferences. Furthermore, the data found that some app categories, such as communication and social media, were quite popular among users.

Furthermore, the study discovered a positive association between app ratings and the number of installations, implying that higher-rated apps tend to attract more users. These insights might help developers prioritize particular categories. and improving their app quality to increase user engagement and retention.