Data Visualization
Final Project Report
Google Play Store Analysis
Vishruth Guda

Contents

Introduction

Google Play store was launched in ’08, with the name Android Market. It is a platform that allows millions of users to download third-party apps created by hundreds of thousands of developers. These users register personal information with Google to download the applications. However, there is only a little data regarding the information about these million apps as application source code is available only the third-party developers. After the launch, the number of apps for the first few years was in mere thousands. Later, there was exponential growth in both the download and development of the new apps. It was in the year ’13 when the play store reached a million available apps, which grew to an all-time high of 3.6 million in March of ’18. After that, the number of available apps remained stagnant for the past few years as Google removes low-quality apps regularly. According to recent data, it is estimated that 2.64 million apps were on the play store at the end of June ’22. In the year ’21, users downloaded 111.3 billion mobile apps. Since there is high demand for apps in every category and niche, the number of apps released and downloaded is continually rising. It is estimated that 3,700 apps are being added to the Play store daily. However, the commonly used and popular ones are very few compared to the total number. Only 142,000 apps have a rating of 4.5 and above, and 1.44 million apps have some rating, while the rest of 2.1 million have less than three user ratings.

With many firms established to make android applications and gain their revenue through downloads and in-app-revenue, this project intends to explore different categories of applications and compare their metadata which includes revenue, cost, number of downloads, ratings and have performed sentiment analysis through existing user reviews. On little research I found that the revenue (in app or advertisement) is directly proportional to number of downloads and the rating of the application. Ambition to solve a business problem and provide valuable insights on which genre, cost and other factors will yield most results in terms of downloads and ratings thus leading to better revenue and reach with Tableau as a visualization tool.

Research questions addressed

1. Which categories have highest number of average downloads?
2. Which category groups have highest play store ratings?
3. Number of applications with latest updates by year and their average rating to see if updates effect the rating.
4. How many applications have most downloads in different countries, to establish if focused applications based on population yield better results?

5. How sentiment analysis based on user reviews is across different categories of applications?
6. Which price group is the most popular in terms of paid applications?
7. Based on the user reviews of some applications, finding the count of apps with different sentiments after sentiment analysis in their respective genres.
8. How many applications are generating revenue and if generating we compare what for of revenue is generated.

Methodology

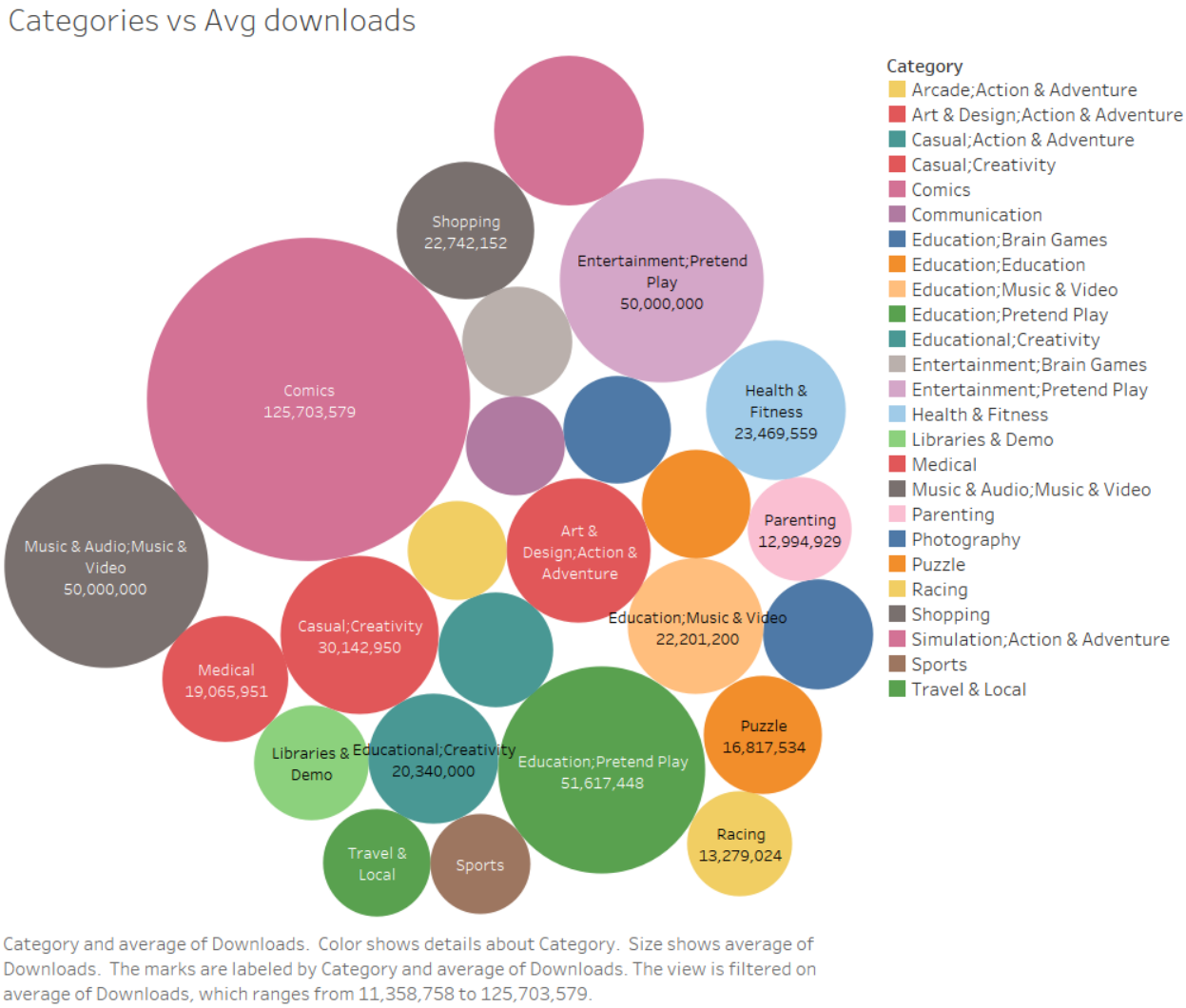
Visualization is representation of data in pictorial or graphical format. It enables us to absorb and analyse information quickly to make faster decisions. Data visualizations help in understanding hidden patterns through visual representation of data and help understanding and make sense of complex data which cannot be understood directly. Good data visualizations are accurate, clear empowering and easy to understand.

I have kept in mind the above principles and followed 5 step process to produce my visualizations:

1. Clear Ambition: I have decided on the topic of play store analysis and understood need for it and decided to visualize play store data.
2. Understand and clean data: I have used google play store datasets from Kaggle with multiple sheets, I cleaned the data, removed ambiguous values, and transformed data into one data source. Source: Google Play Store Apps| Kaggle, Google Play Store Apps 1 | Kaggle
3. Transform data: I have observed some factors like year and month converted it to date time format, grouped prices of applications, created calculated fields for distinguishing applications with revenue, without revenue and type of revenue. Created hierarchies where needed.
4. Visualize data: Keeping the aspects of effective visualizations in mind I have created easy to read and understand visualizations using related chart types, animations, and filters.
5. Attention to detail: Proper labelling, colour scheming, titles for individual visualizations and corresponding dashboards.
   Thus, keeping in mind lectures on effective visualizations, I have created my visualizations.

Analysis

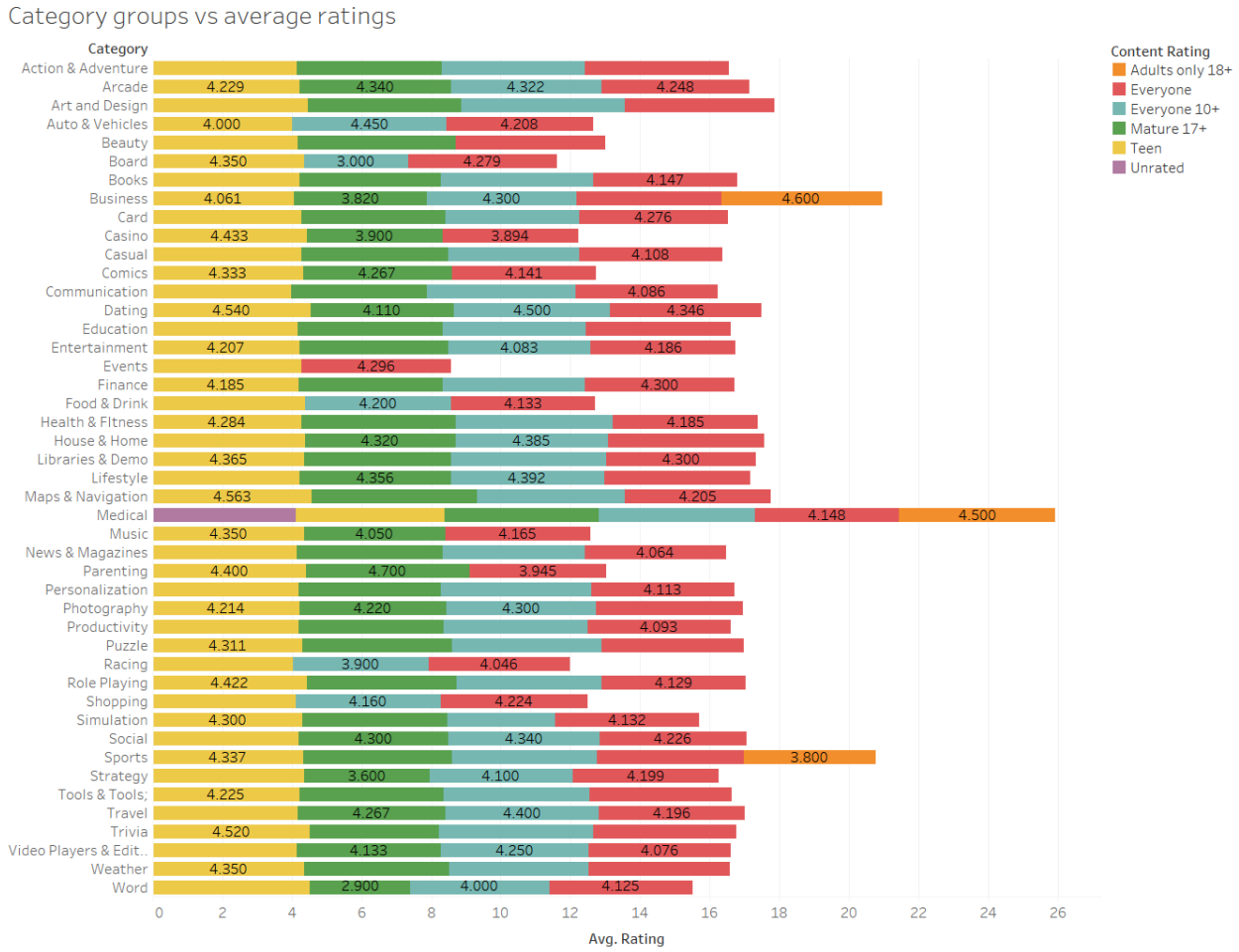
1. **Which categories have highest number of average downloads?**



Category and average of Downloads. Color shows details about Category. Size shows average of Downloads. The marks are labeled by Category and average of Downloads. The view is filtered on average of Downloads, which ranges from 11,358,758 to 125,703,579.

This simple visualization shows the average downloads for each category of applications. This visualization contains the categories with the greatest number of average downloads. We can see that Comics category holds the greatest average downloads followed by Education: Pretend Play. We can infer that apps in these categories are the most popular.

2. Which category groups have highest play store ratings?

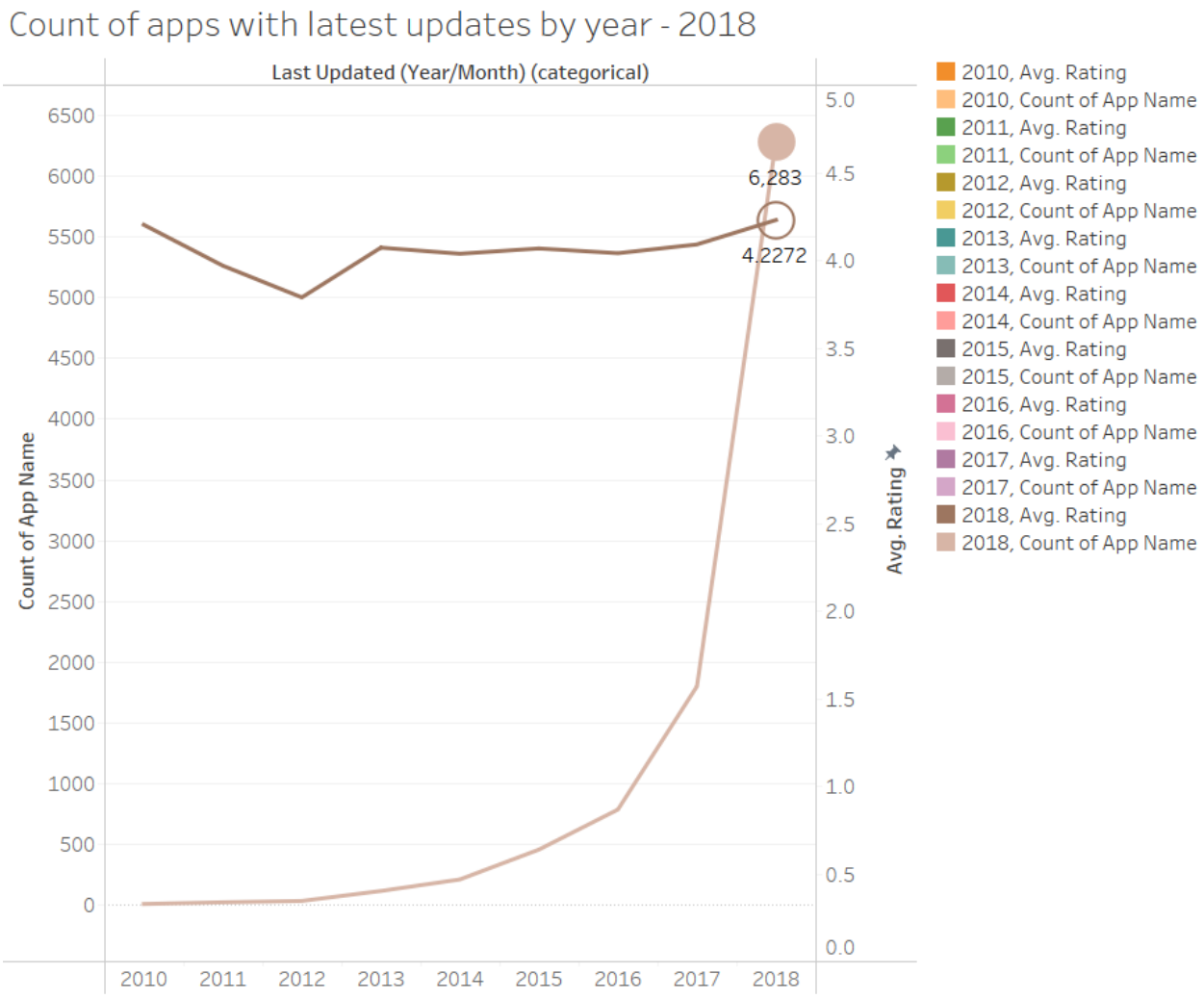


Average of Rating for each Category. Color shows details about Content Rating. The view is filtered on Category and Content Rating. The Category filter excludes February 11, 2018. The Content Rating filter excludes Null.

This visualization shows the Genre vs content rating of the of the applications in that category. This is a drill down chart which furthers divides genres into different categories.

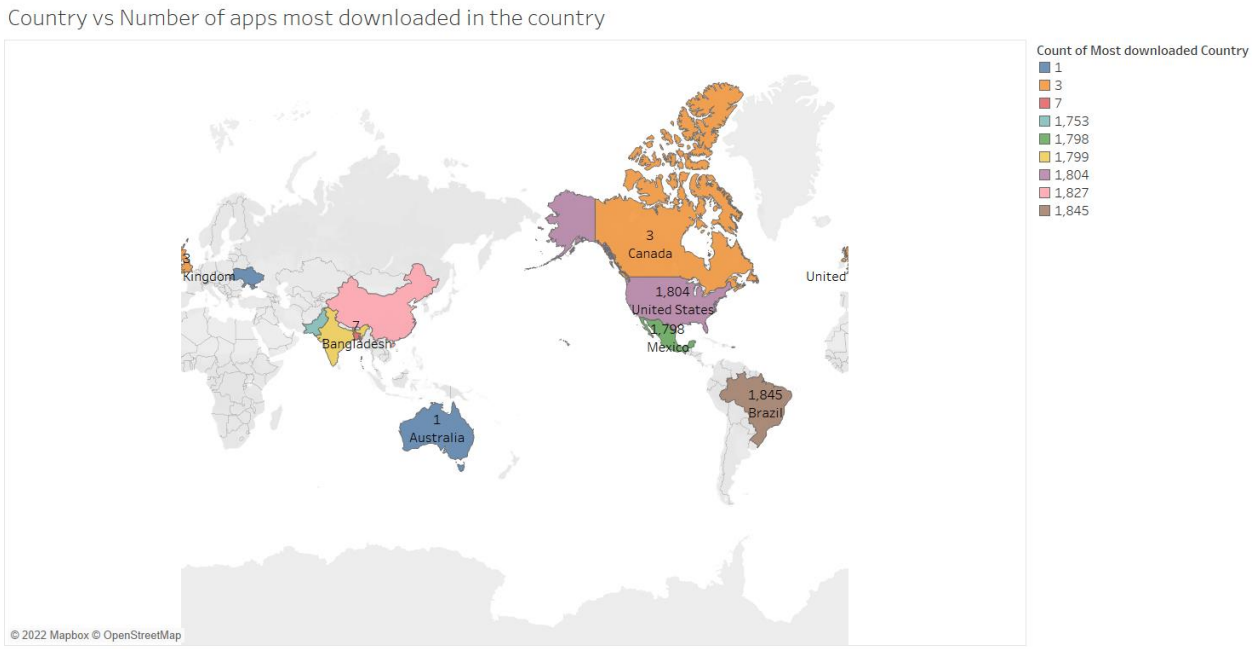
We can observe that Parenting apps with an average rating of 4.7 take the largest piece of the cake. In subcategories, Lifestyle: education takes the lead with 5.0 average rating.

3. **Number of applications with latest updates by year and their average rating to see if latest updates effect the rating.**



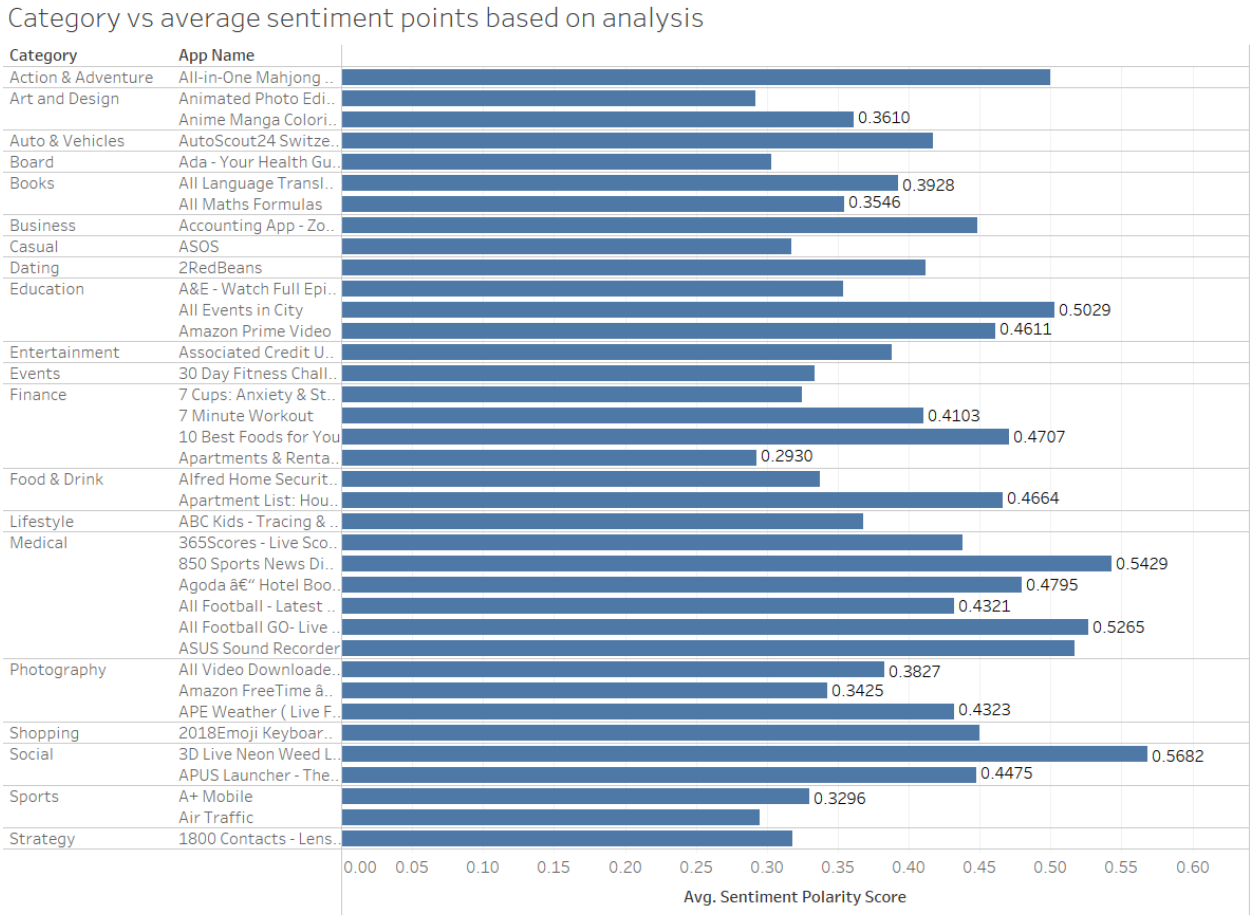
This visualization contains animation which explains the relationship between year/month of latest updates vs count and average rating of applications. We can observe that most of the applications received latest updates as per the dataset, the updates did not have a great effect on the rating. We can say that periodical updates are required but are not a necessity for high app ratings.

4. How many applications have most downloads in different countries, to establish if focused applications based on population yield better results?



This visualization shows count of number of apps with max downloads in a particular country. This visualization uses a map type chart. We can see that most of the maximum downloads are from highly populous countries. So focused marketing with country specific features while developing a app might result in better downloads.

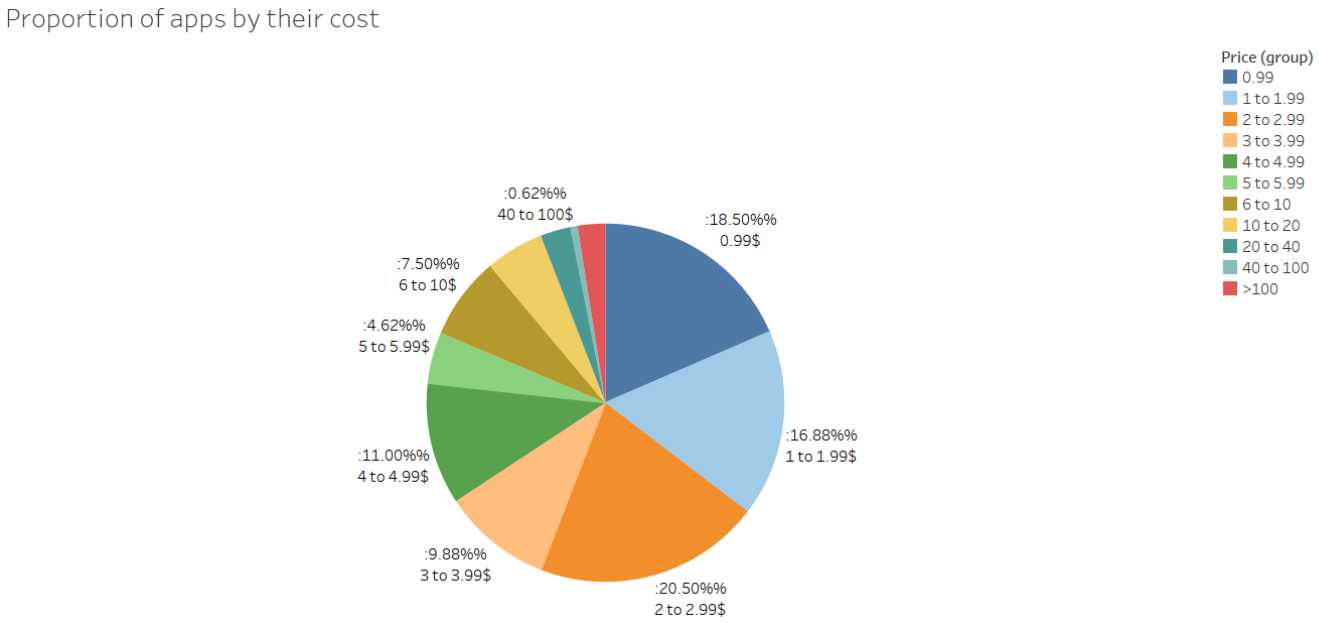
5.How sentiment analysis based on user reviews is across different categories of applications?



Average of Sentiment Polarity Score for each App Name broken down by Category. The view is filtered on average of Sentiment Polarity Score and App Name. The average of Sentiment Polarity Score filter ranges from 0.2890 to 0.5682. The App Name filter excludes Null.

Most of the times, user ratings may be affected by one feature or a news or a market trend. Thus, rating may not always be the correct measure for user liking. So, with this analysis of user reviews and assigning a sentiment score to them, we can see that 3D Live neon app from social category has the highest positive sentiment. The rating is 4.5 but the sentiment is better.

6. Which price group is the most popular in terms of paid applications?

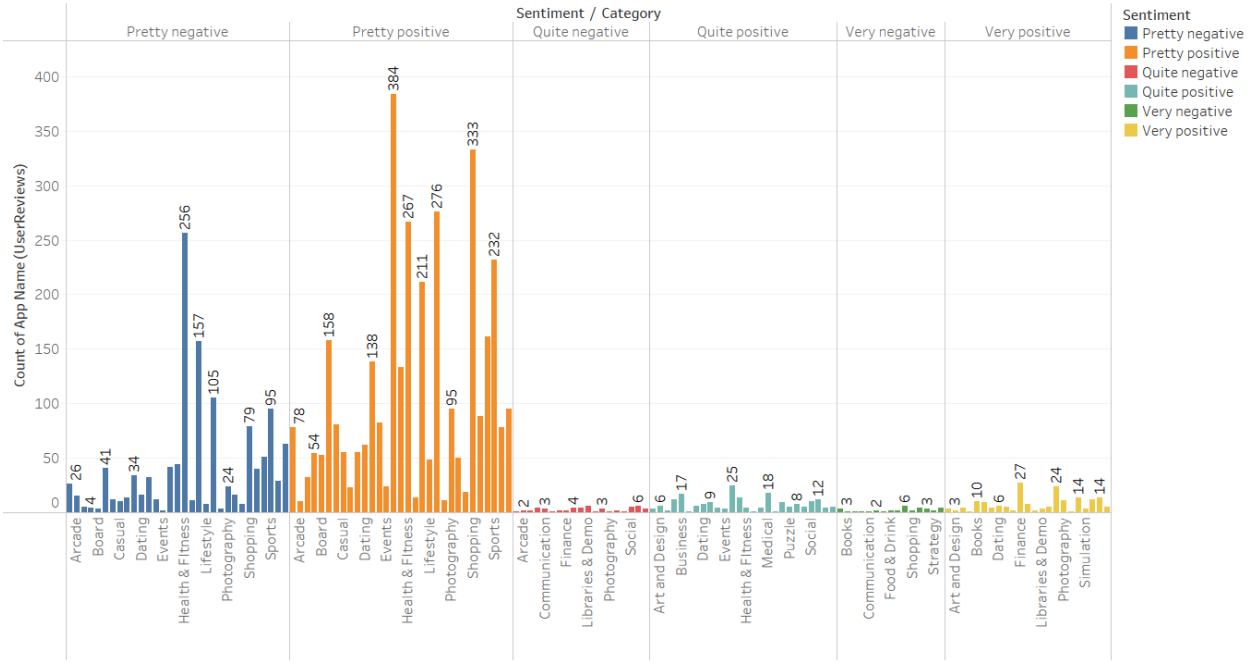


This is a simple visualization to understand in which price range most of the applications fall in.

We can see that 2$ to 2.99$ is the most popular price range of applications. Thus this price range may provide balance between affordability and revenue.

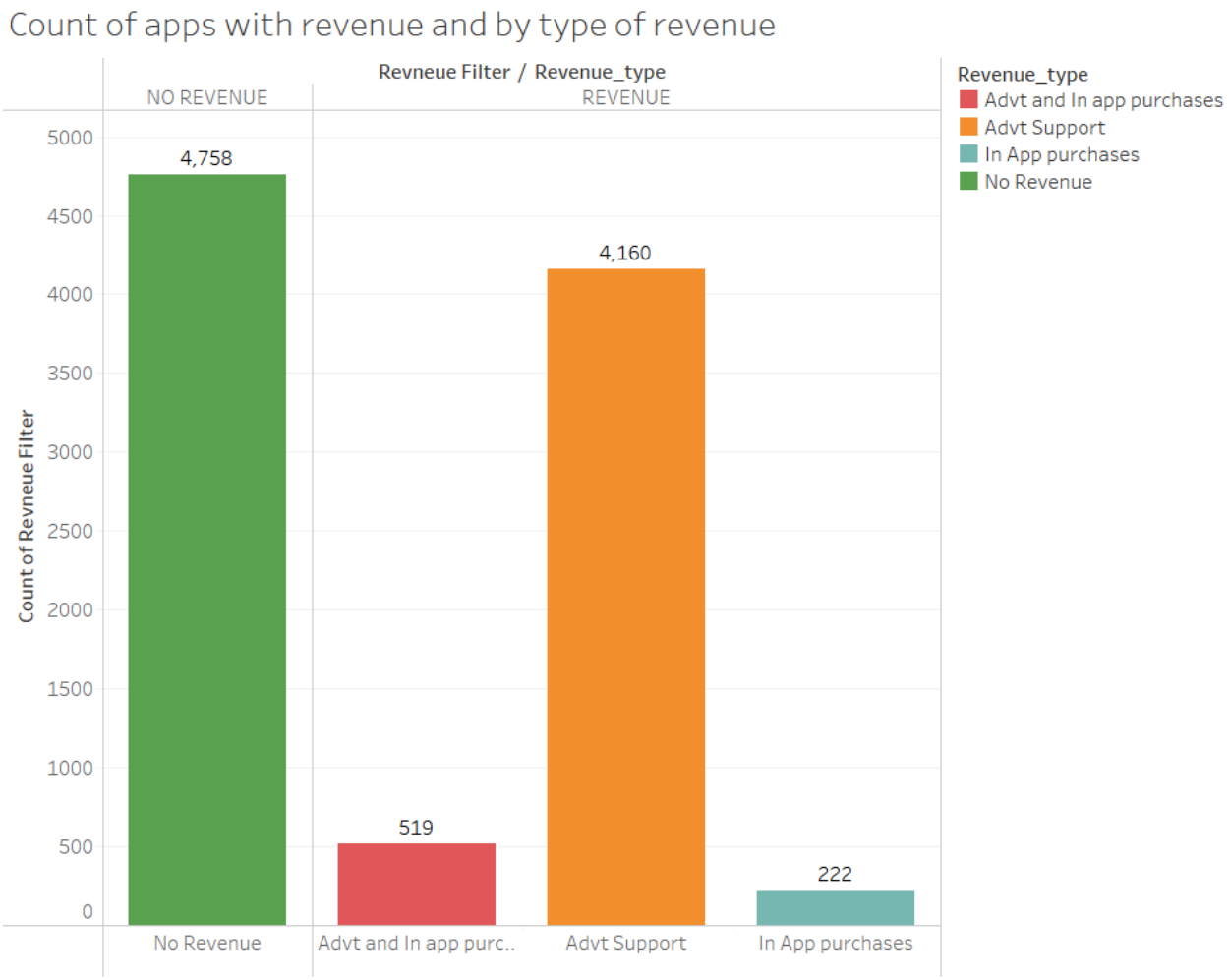
7. **Based on the user reviews of some applications, finding the count of apps with different sentiments after sentiment analysis in their respective genres.**

Chart showing sentiment analysis on apps with category filter



We can see that most of the categories are in positive category by sentiment analysis, hence we can infer that generally most of the categories have positive sentiment. While deigning an application category in Pretty positive, quite positive, and very positive categories may be considered while developing an application.

8. **How many applications are generating revenue and if generating we compare what form of revenue is generated.**

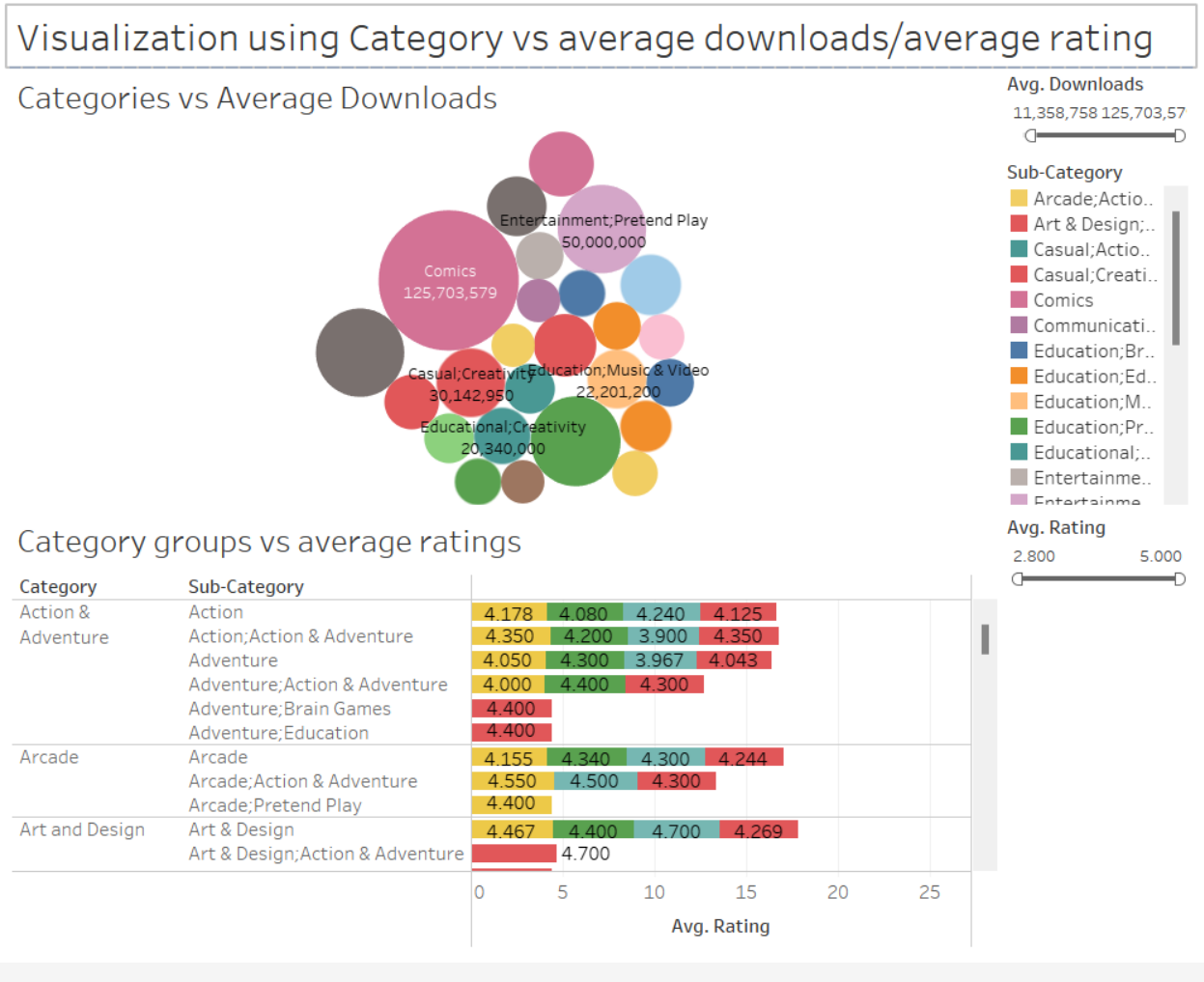


The financial viability of an application depends upon the revenue it generates. Here we have compared the number of apps free from revenue and in paid apps, we have compared the type of revenue generated. It is a drill down chart which shows count of apps with no revenue vs generating revenue and further drilled down to the type of revenue generated.

Dashboards

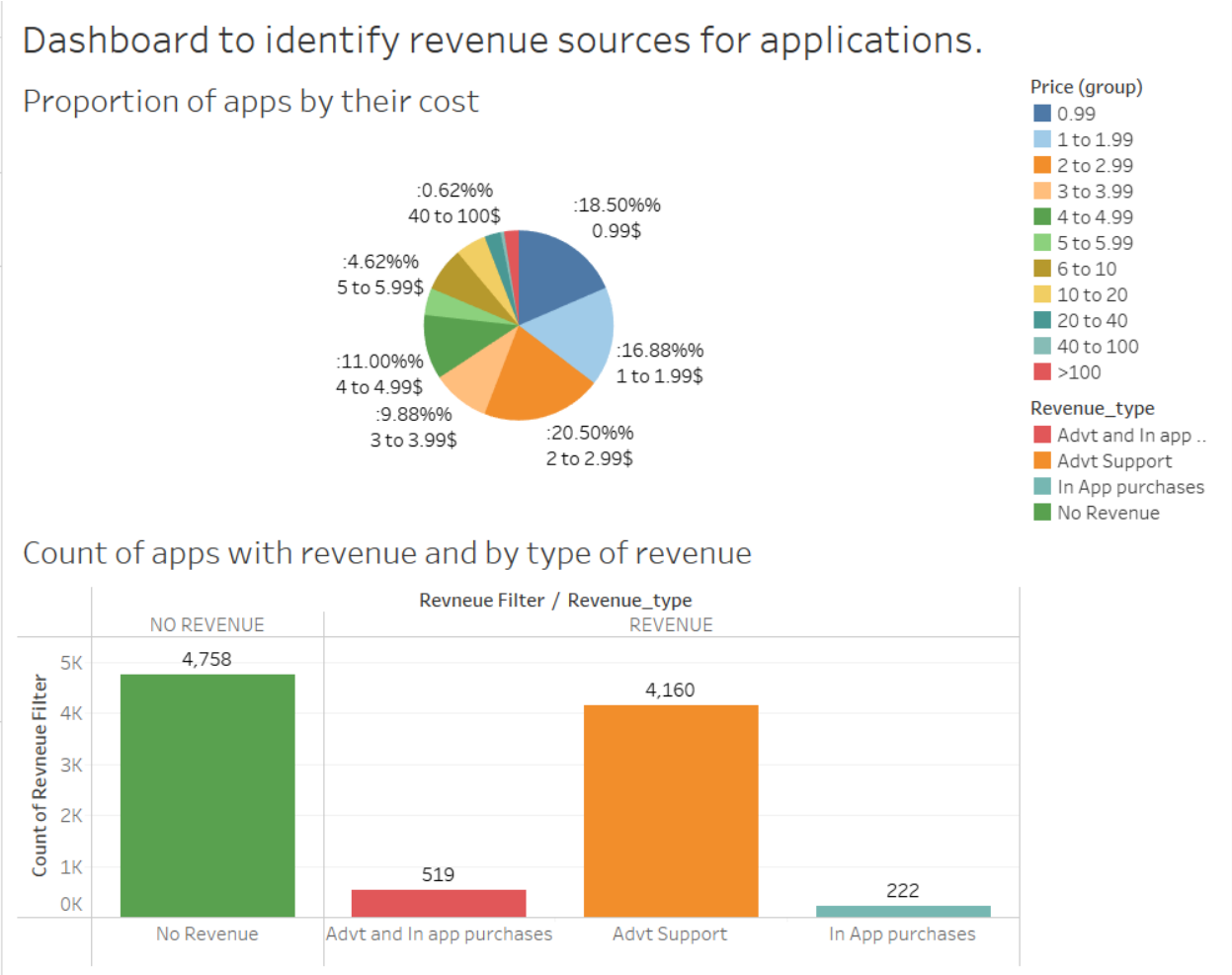
A dashboard is a collection of several views, letting you compare a variety of data simultaneously. This helps in grouping different visualizations which provide answer to a particular business case.

Dashboard 1: It provides high level comparison on the most important attributes of a successful application i.e., downloads, rating vs category.



By this dashboard we can see that apps related to comics with a good rating of 4+ and great average downloads can become a popular app.

Dashboard 2: This dashboard explains how an application can generate revenue and be financially viable.



We can observe that revenue with paid apps in the range of 2-2.99$ and advertising revenue are the most popular forms.

Conclusion:

With the above visualizations we can conclude that apps in Comics and education category, with teen rated content with a price range between 2-2.99$ and revenue through advertisement support would provide high ratings and downloads making the app successful.

Further research questions:

1. Is it possible to create a combination of different genres to create a popular app.

2. Analyse future trends of which category of apps will come into marketplace based on previous trends.

References:

Ceci, L. (January 3, 2022). Global Google Play app downloads 2016-2021.

https://www.statista.com/statistics/734332/google-play-app-installs-per-year/

Ceci, L. (July 13,2022). Google Play: number of available apps 2009-2022.

https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/