Visualization Project

2023-09-29

Google Play Store

Deepanshu Mittal

Roll No: MDS202320

DG1107: Visualization

Email: deepanshu.mds2023@cmi.ac.in

Chennai Mathematical Institute

Table of Contents

0.1   INTRODUCTION

Google Play Store and formerly Android Market, is a digital distribution service operated and developed by Google. It serves as the official app store for certified devices running on the Android operating system and its derivatives, as well as Chrome-OS, allowing users to browse and download applications developed with the Android software development kit (SDK) and published through Google.

Currently,in 2023, Google play store caters 3.553 million applications and has 2.5 billion active monthly users, much higher then its competitors Apple's app store with 1.642 million and Amazon's app store with 0.822 million applications.

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. In this project we will to do a basic EDA on the google app store data set.

0.2   DATASET DESCRIPTION

Click to find the source for the data set. This dataset has 13 features related to apps like rating, review, size, installs etc. and 10841 rows of data. After eliminating anomalies and data cleaning, our final working dataset has 8886 rows of data.

0.2.1   Detailed description for the columns of the dataset:

- App : Contains name of the application.
- Category : Contains the application category like (Art and Design, Beauty, Games, Family, etc)
- Rating : Contains ratings given to the application.
- Reviews : Contains no of reviews given to the application.
- Size : Contains the size of the application in MBs.
- Installs : Contains the no of installs of the application.
- Type: Gives information about type of application (Free/ Paid).
- Price : If the application is paid, tells the price of the application.
- Content Rating : Tells about target audience of the application.
- Genres : Very similar to category, but gives more detail about the genre of the application.
- Last Updated : Gives the date of the last update of the application.
- Current Version : Tells about the current version of the application.
- Android Version : Tells about the Android version of the application.

0.2.1.1   Full Forms for the abbreviations used in App's category:

- A & D : Art and Design
- A & V : Auto and Vehicles
- B & V : Books and Reference
- BUS : Business
- COMM : Communication
- DATE : Dating
- EDU : Education
- ENTERTAIN : Entertainment
- FIN : Finance
- F & D : Food And Drinks
- H & F : Health and Fitness
- H & H : House and Home
- L & D : Library and Demo
- SHOP : Shopping

- PHOTO : Photography
- T & L : Travel and Locals
- PERSONAL : Personalization
- PROD : Productivity
- N & M : News and Magazines
- M & N : Maps and Navigation

0.2.2 Objectives

Using the above dataset, we are trying to determine the following:

- No of applications per category.
- Average rating for each category.
- How may apps are installed in a certain range.
- Count of the apps based on Content rating.
- Correlation among the features (Rating, Reviews, Size, Installs, Type)
- Relation between Installs and Reviews.
- Relation between Size and Installs.

The above objectives can be achieved using histograms, bar plots, piechart and box plots for univariate analysis, scatter plots and line plots for bivariate analysis, and gradient scatter plots for multivariate analysis.
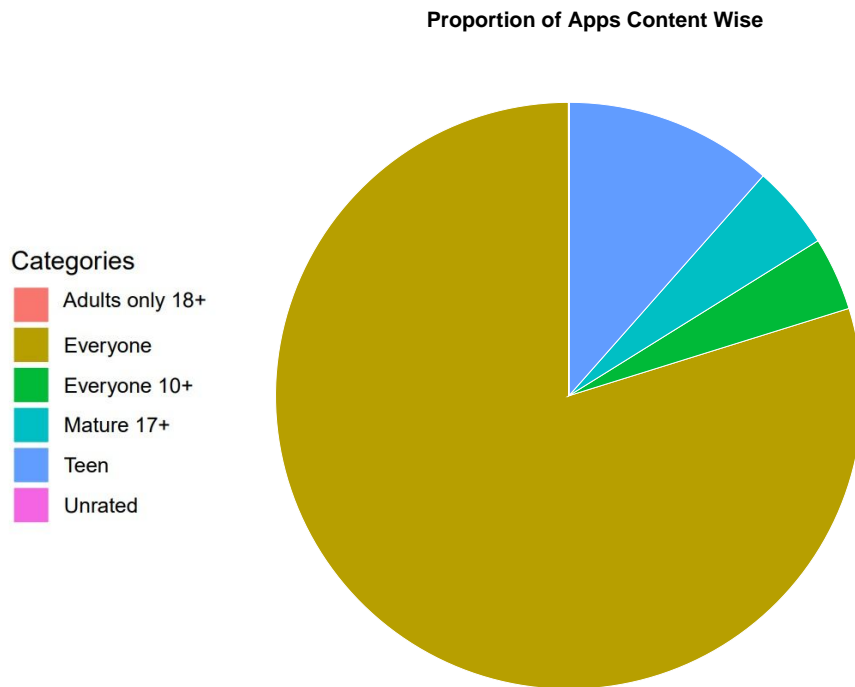
0.3 EXPLORATORY DATA ANALYSIS
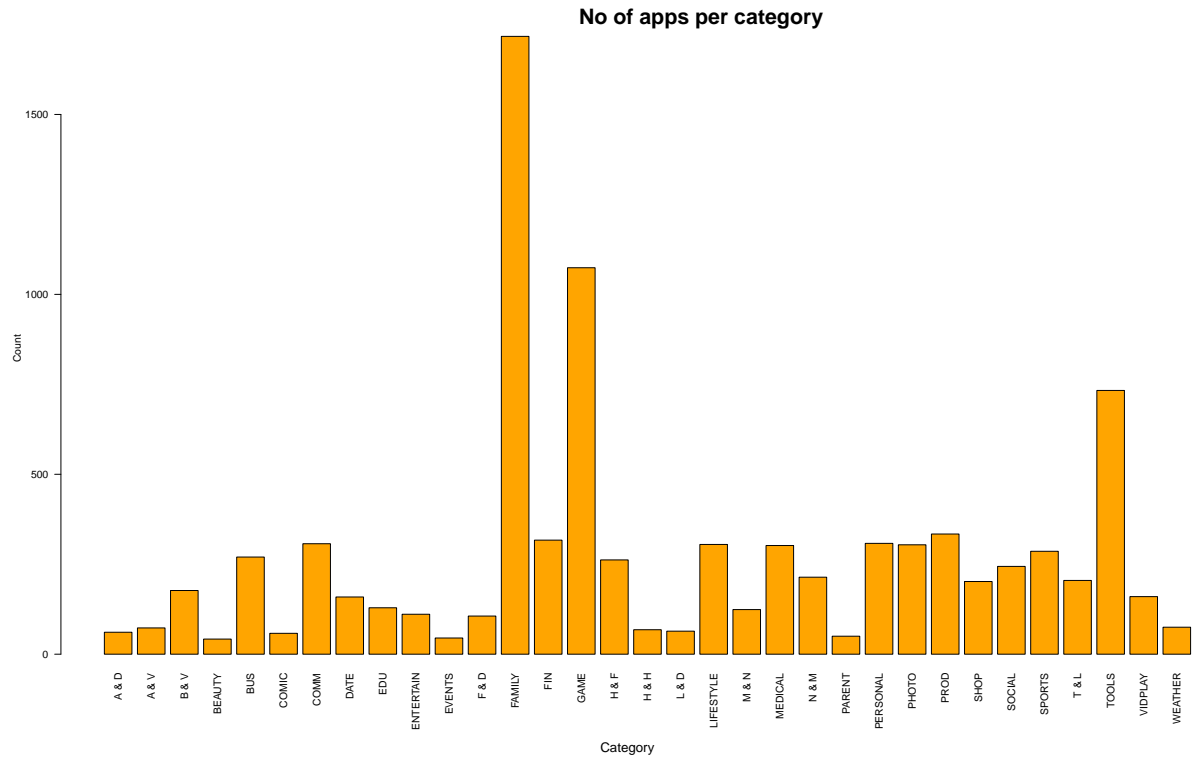
0.3.0.0.1 ***FIG 1: Barplot Representation:*** In the above bar graph we got the counts of apps per category. As per the graph, Apps belonging to Family, Games and Tools category are large in number.

Average Rating of Each Category
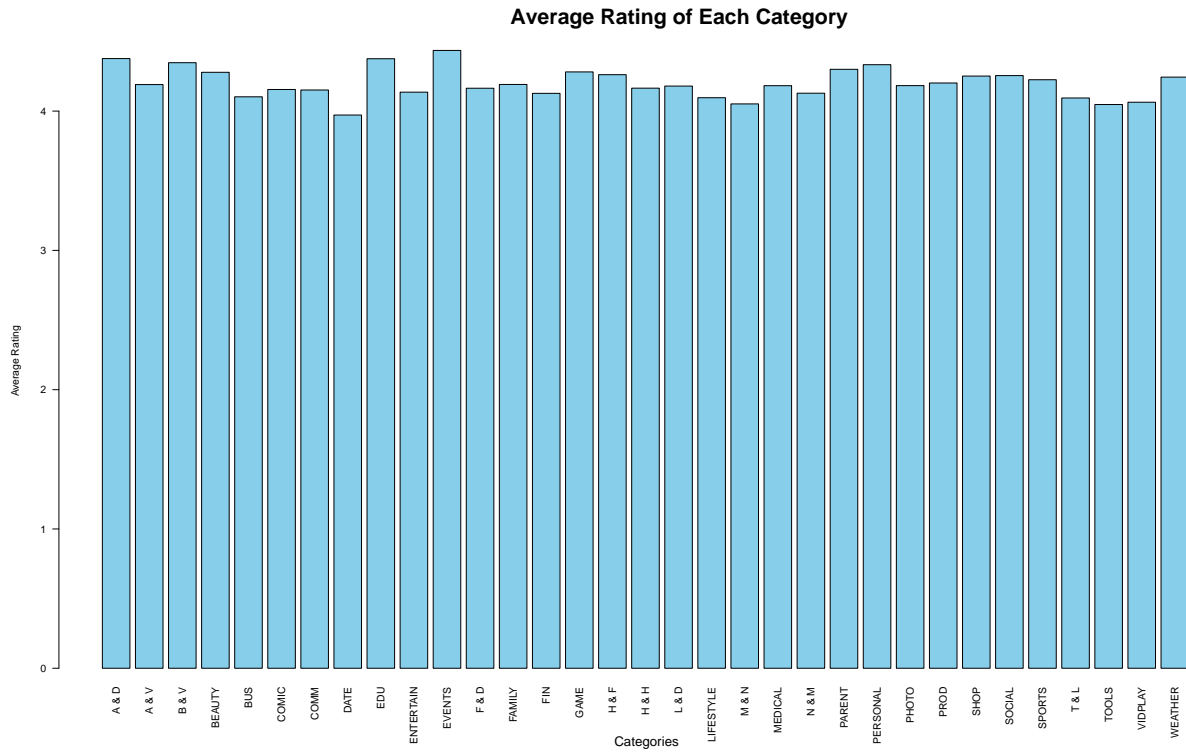
0.3.0.0.2 *FIG 2: Barplot Representation:* In the above bar graph we can see average rating for each category. As per the graph, more or less average rating for each category is similar.

Installs for apps
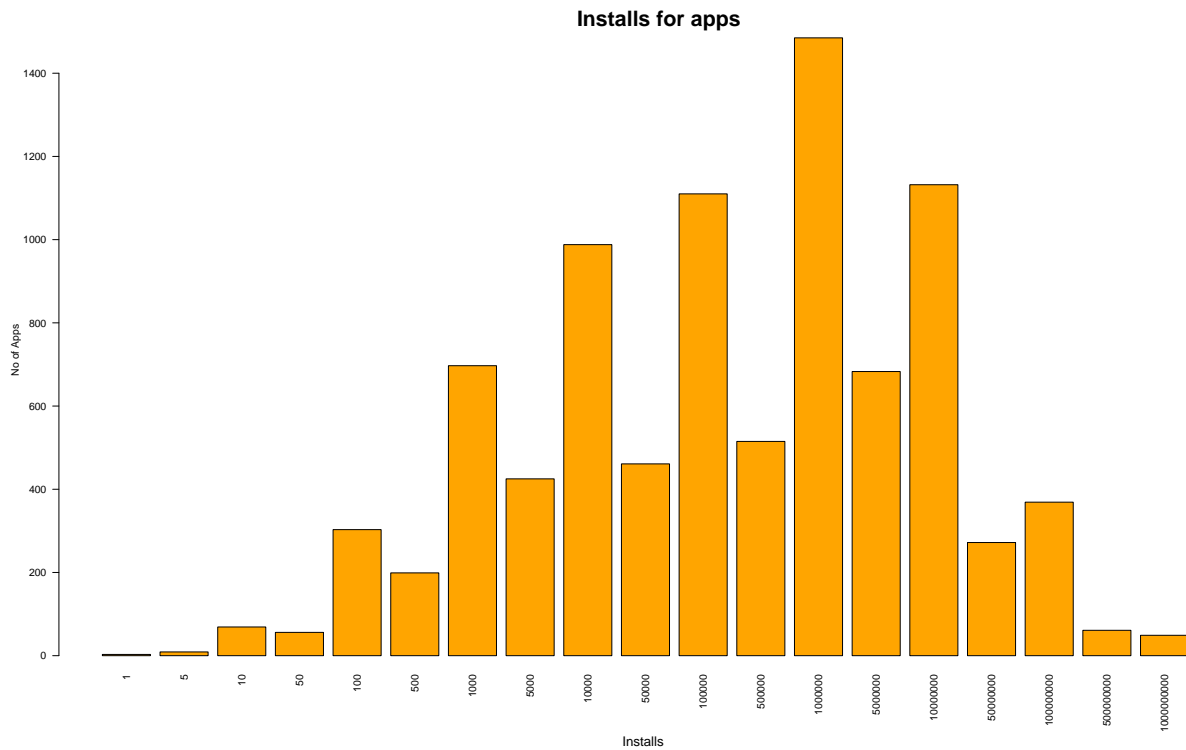
0.3.0.0.3 *FIG 3: Barplot Representation:* In the above bar graph we can observe no of apps with no of installs they received.

Proportion of Apps Content Wise

Categories

- Adults only 18+
- Everyone
- Everyone 10+
- Mature 17+
- Teen
- Unrated

0.3.0.0.4 _**FIG 4: Pie Chart Representation:**_ In the above Piechart, we can observe that most of the applications on play store are meant to use by everyone, very few applications have adult tag associated to them.

Heatmap among the variable specified
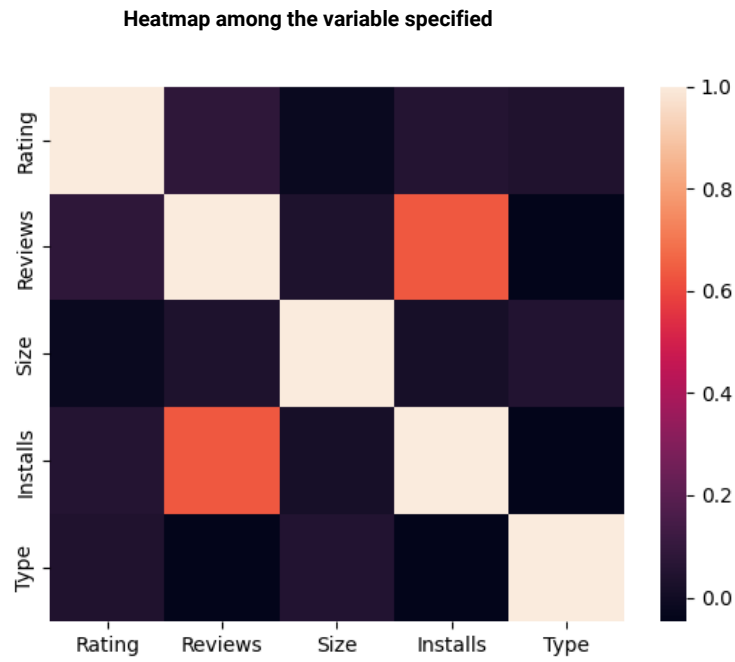
0.3.0.0.5 *FIG 5: Heatmap Representation:* In the above graph, we can observe that there is a strong positive correlation between Installs and Reviews.

Relationship between Installs and Review

Datapoints between Installs and Review
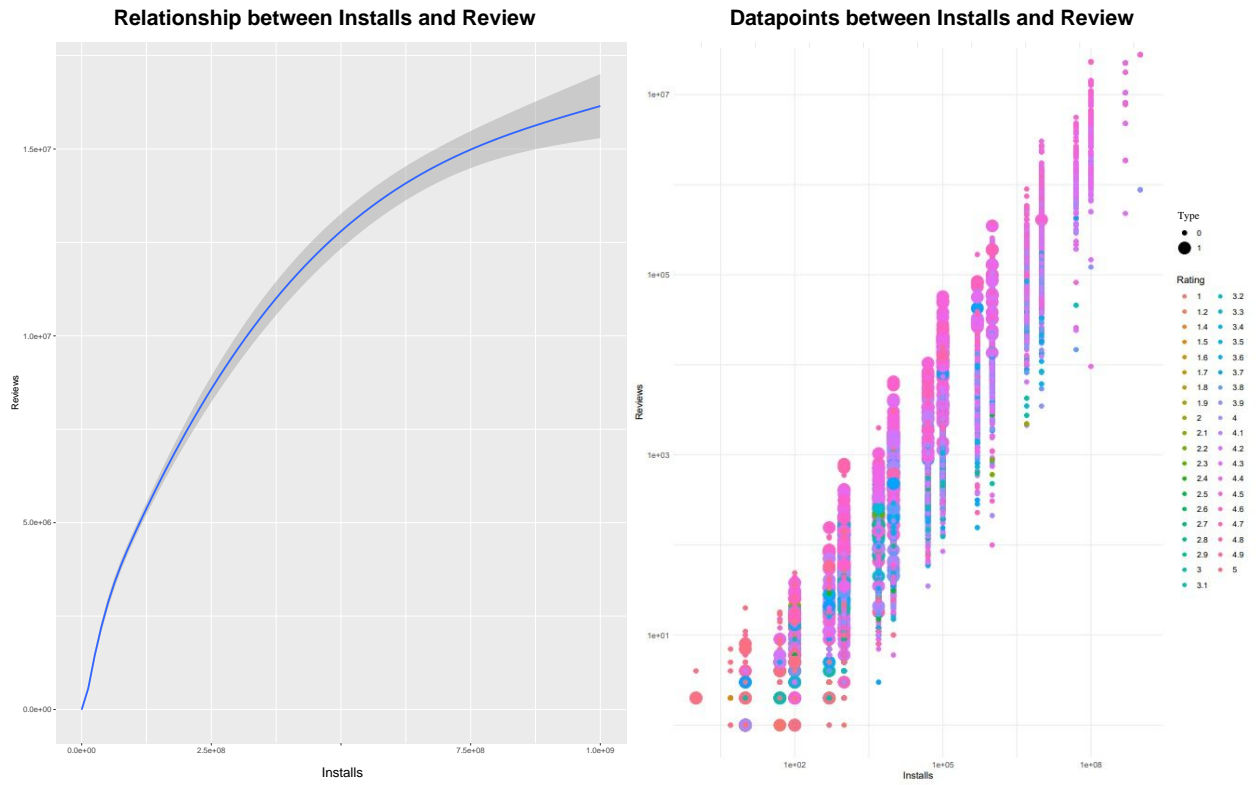
0.3.0.0.6 *FIG 6: Smooth & Point Representation:* In the above graph, we can observe that there is a strong relationship between Installs and Reviews, larger the no of install, larger the no of reviews. But datapoints also suggest it is not true everywhere. Although, initally it is showing strong positive correlation.

Relationship between Size and Installs

Datapoints between Size and Installs
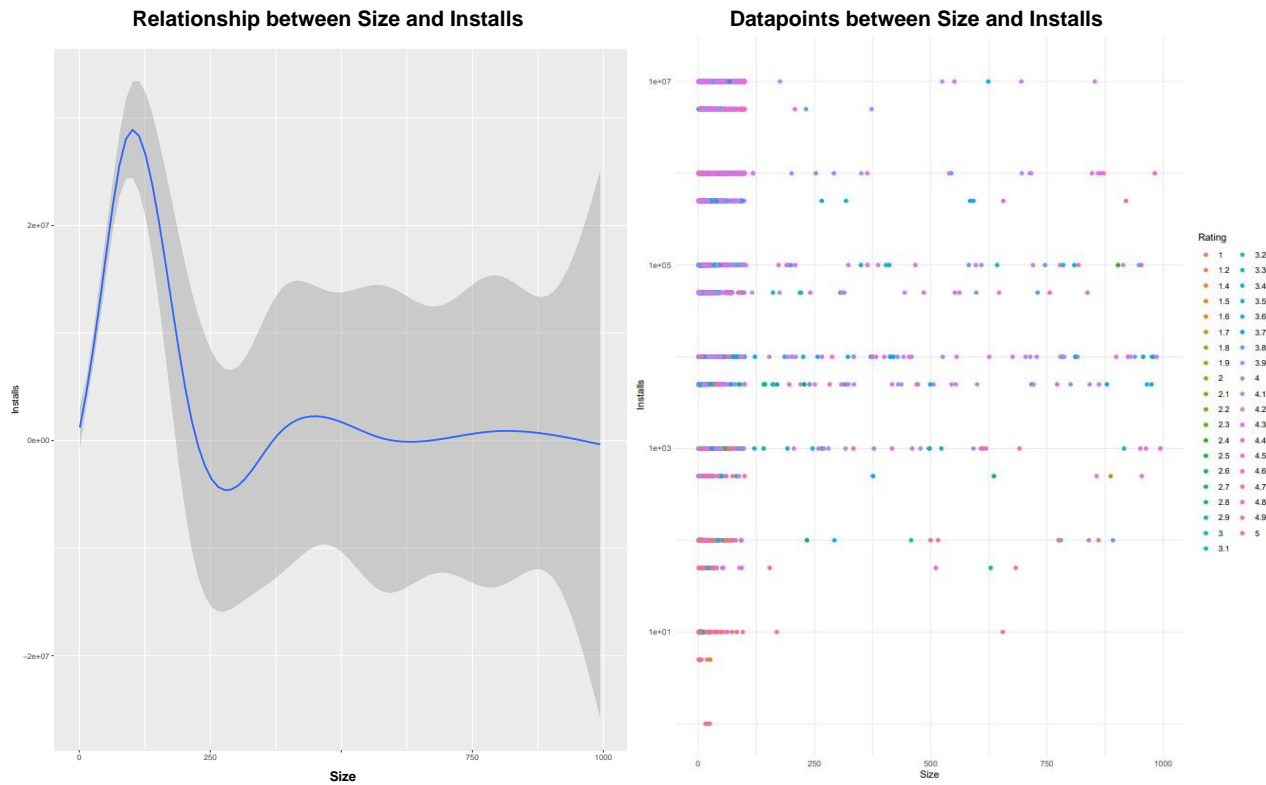
0.3.0.0.7 _FIG7:Smooth&PointRepresentation:_ In the above graph, we can observe a strange yet obvious relationship between Size and Installs of the application. Between 0-100 (MBs) of app, the installs are increasing, then from 100-250 (MBs) the no of installs are decreasing and after that no of installs are constant. Although confidence intervals have increased as the size of application increased.

0.4 RESULTS

- There are millions of application on Google Play Store, and the major dominating category among them are Family, Games and Tools application.
- In our analysis, we took 33 unique categories and found out that more or less average rating for each category is similar.
- We found out that most of the applications on play store are meant to use by everyone, very few applications have adult tag associated to them.
- Using Heatmap, we found out that there is a strong correlation between Installs and Review.
- We found out applications with size between 50-100 MBs usually have maximum downloads and as we keep on increasing the size of application, the installs decreases, also signifying negative correlation.

0.5 CONCLUSION

In this project, we had a dataset from Google play store with 13 features and 10841 rows. Here we tried to establish correlations between different variables in the dataset and how they are dependent on different demographics. During our analysis, we delivered the above mentioned objectives and found out some interesting insights from the data like Google Play Store boasts a vast array of applications, with Family, Games, and Tools being the dominant categories, across 33 unique categories, we observed that the average user ratings for these apps are generally consistent, indicating a relatively uniform user experience, most apps are family-friendly, as only a small proportion carries an adult tag, suggesting a focus on a broad audience, the analysis revealed a strong positive correlation between the number of installs and the number of reviews, highlighting the influence of user engagement on app popularity, interestingly, apps sized between 50-100 MB tend to have the highest number of downloads, while larger apps experience decreasing install rates, illustrating a negative correlation between app size and popularity. Although our analysis is quite basic, it gave insights that have enormous potential to drive app-making businesses to success.