Analysing Google Apps Store

Table of Contents

Problem Statement

With the advancement of technology, people are shifted more towards mobile. Further, Google Play Store Apps are being used by android users and they are being used for entertainment, education, business, payment, etc. There are various apps which we can download for our entertainment like game apps, song apps, video players and a lot more. People are being shifted towards apps for educational purposes too as they can use apps while travelling easily. Many businesses are making money through apps. We download different payment apps for ease of payment. Some apps are free and some are paid. For some apps, we see advertisements which is source of income for businesses.

Our whole life is largely dependent on these apps and hence, their analysis is extremely important. We need to analyse which factors affect installation numbers.

Objective and Benefits

Objective:

- The objective of the project is to perform data visualization techniques to understand insights of the data. This project aims to apply various Business Intelligence tools such as Tableau to get a visual understanding of the data.

Benefits:

- The data analysis will reveal some common and unique patterns in the dataset related to googleplaystore and googleplaystore_user_reviews datasets.
- Data visualizations will enhance the understanding of the effect of these features on Installs of an app and give a better chance to make business

Steps Followed

1. Data Extraction: This step involves extracting the data from different sources relevant to the problem statement or obtaining data from the client.
2. Data Pre-Processing: Once the raw data is obtained, we need to ensure that the data is free from errors. We perform Exploratory Data Analysis followed by Data Cleaning which involves imputing missing values, removing duplicates, finding anomalies or outliers, and treating them.
3. Data Modification: Data is modified wherever necessary.
4. Data Analysis and visualization: Data is analysed and visualizations are made to make conclusions out of them. Data analysis is done using Python(Jupyter Notebook) and SQL (SQLite) both.
5. Data Loading in Tableau: Cleaned data is loaded into tableau and visualizations are made. Various inferences are made and dashboard is created.
6. Deployment: The prepared visualizations are deployed on the Tableau Online Software where they will be available to public.

1. Data Extraction

- Two datasets are provided – googleplaystore.csv and googleplaystore_user_reviews.csv
- Both datasets are available in repository.

2. Data Pre-Processing

- **Android Ver** and **Current Ver** columns had some null values but they held very less percentage of the total. So, rows corresponding to those null values were simply dropped.
- For **Type** column, wherever we found null values, price corresponding to it was zero. So, we filled Free for those null values of **Type** column.
- For **Rating** column, we found some null values. Those were large in percentage. So, we detected outliers and finally imputed missing values by median of the column.

3. Data Modification

- We dropped some duplicates from **App** column in **googleplaystore** dataset.
- We also modified format of values inside **Reviews**, **Size**, **Installs** and **Price** columns. Also, we modified datatypes of these columns along with **Last Updated** .column
- We also created some SQL tables in SQLite database. We created three tables – one for googleplaystore dataset, second for googleplaystore_user_reviews dataset and last one after merging the last two tables.

4. Data Analysis and visualization

- Data was analysed in Jupyter Notebook and SQLite using Python and SQL.
- Using analysis results, we created some visualizations inside Jupyter Notebook and made some inferences from them.

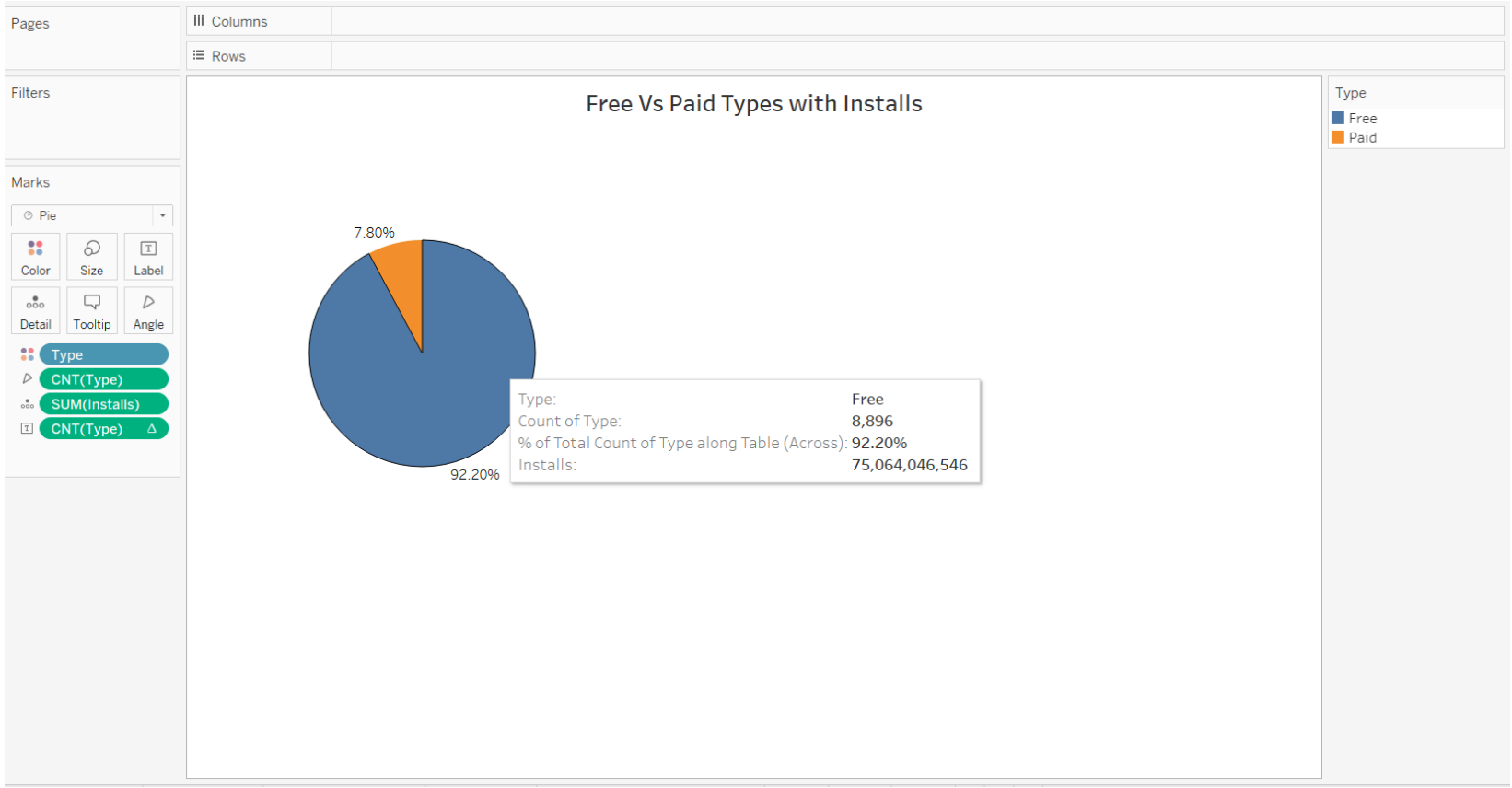
5. Data Loading in Tableau

- Cleaned data is loaded into tableau after lot of pre-processing steps and analysis in Jupyter Notebook.
- This cleaned data is then analysed in Tableau and we made dashboard and various conclusions.

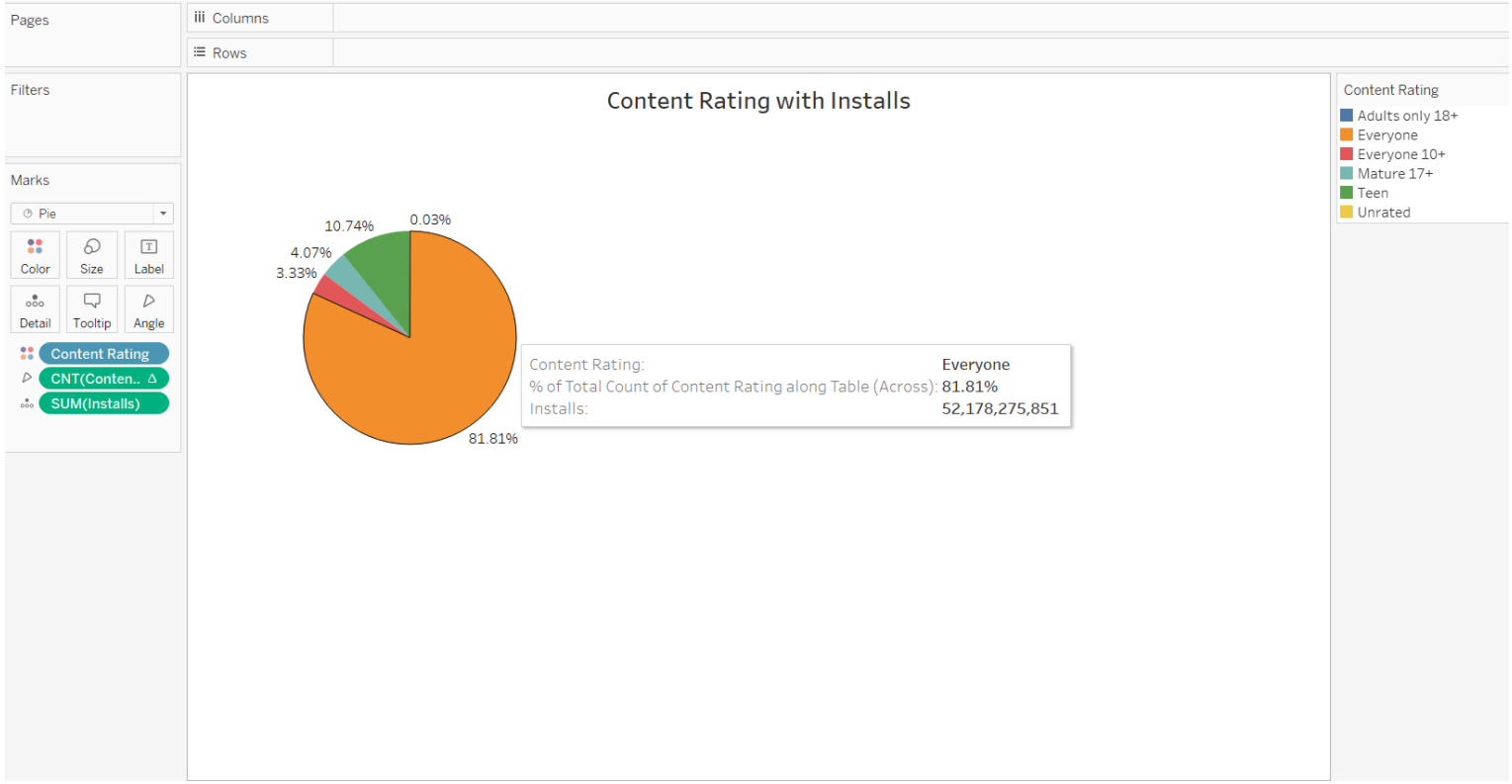
6. Deployment

- All the different worksheets that have been created are compiled together into aTableau workbook. Each worksheet is named based on the type of visualization performed in the chart.
- When we save all the worksheets on Tableau software, it connects to the TableauPublic Software via personal email id and credentials. The dashboard is uploadedonto the Tableau Public Software on personal profile and this is visible to public.
- URL for worksheets and dashboard : https://public.tableau.com/app/profile/sakshi.bhalothia/viz/Dashboard_GooglePlay/Dashboard1

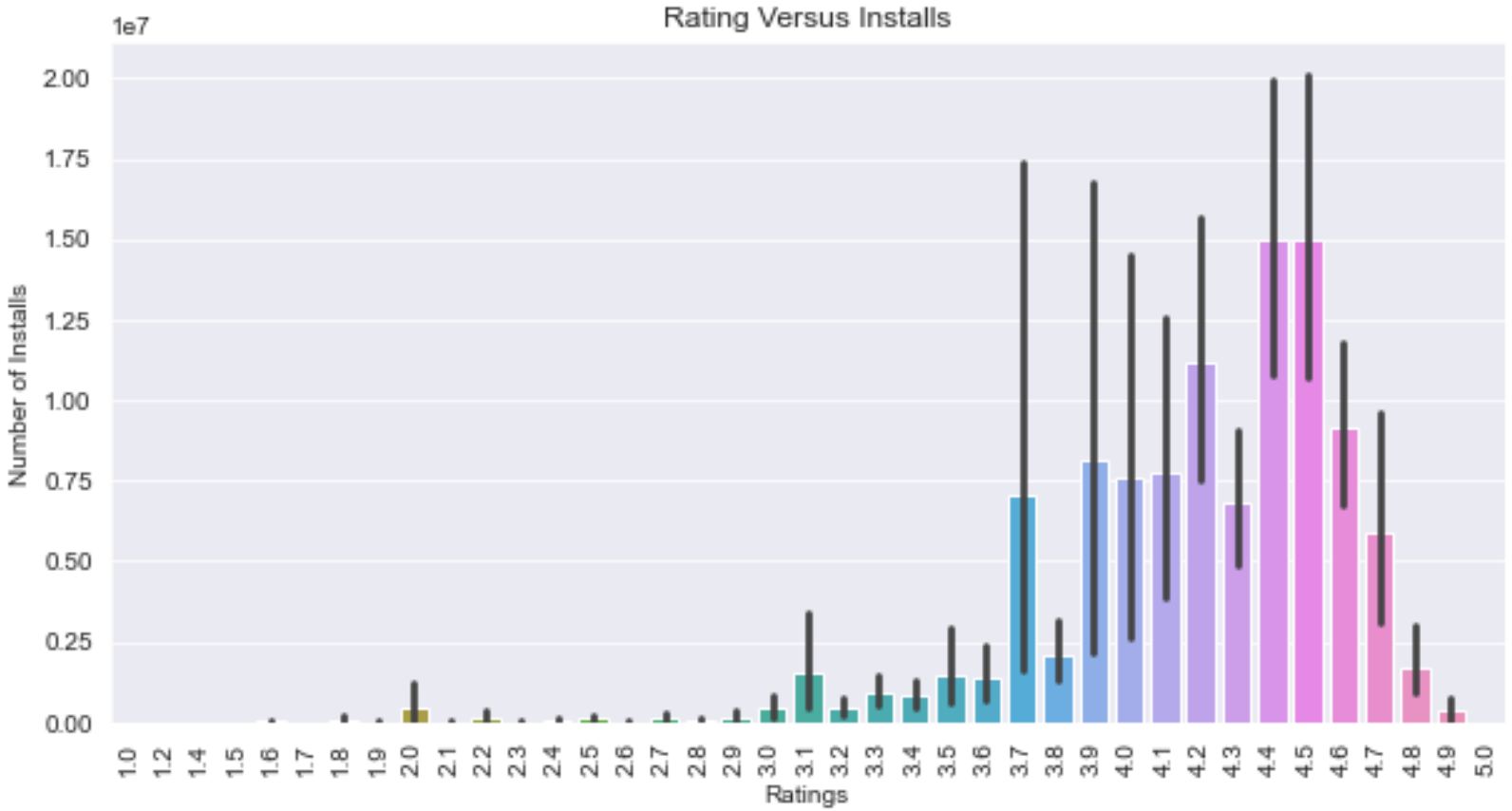
Visualizations



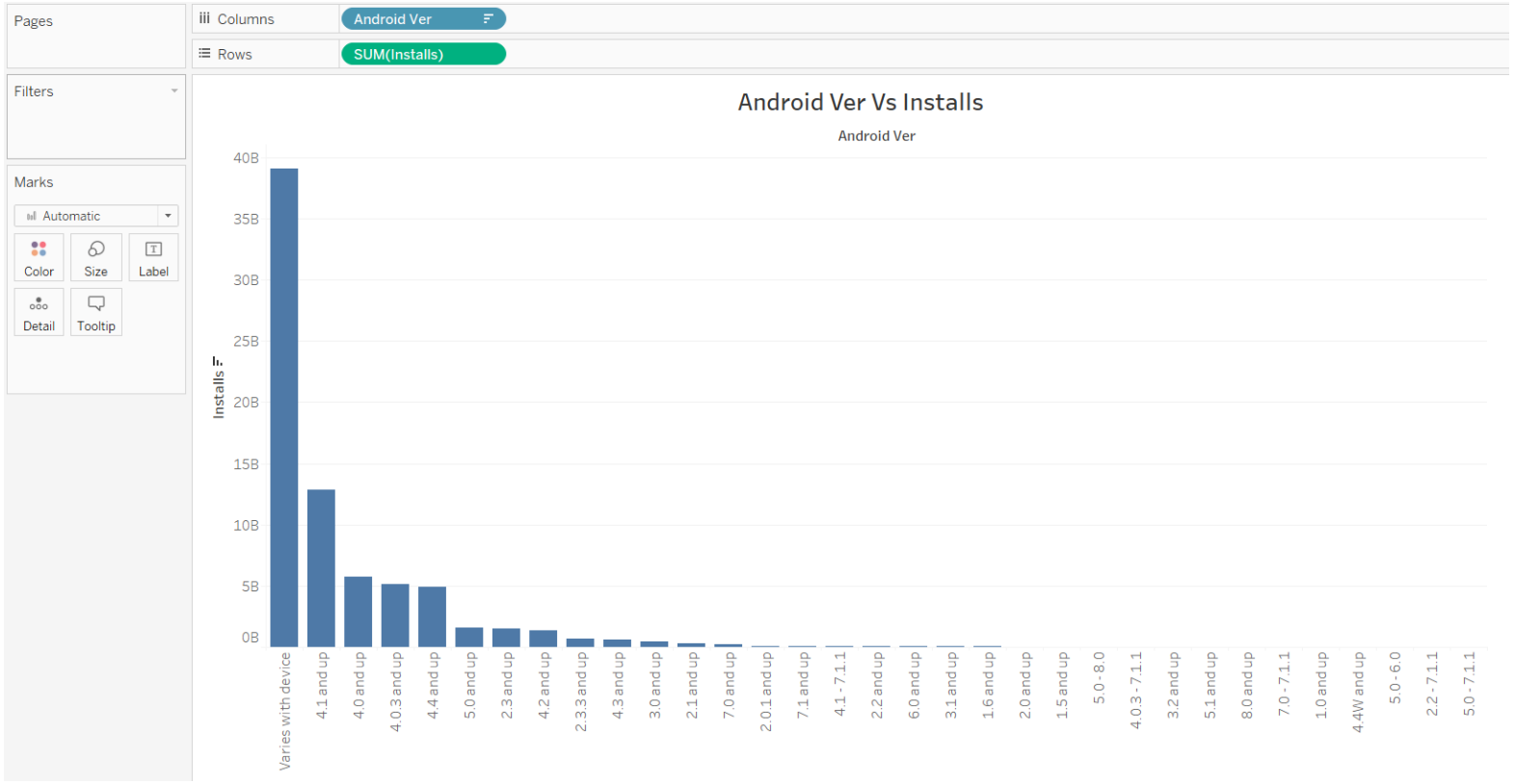
- 92.2% apps are Free and rest are Paid. So, users install Free Apps more than they install Paid Apps.
- By hovering over the pie chart, we can see Installs in number. This number is more for Free Apps.



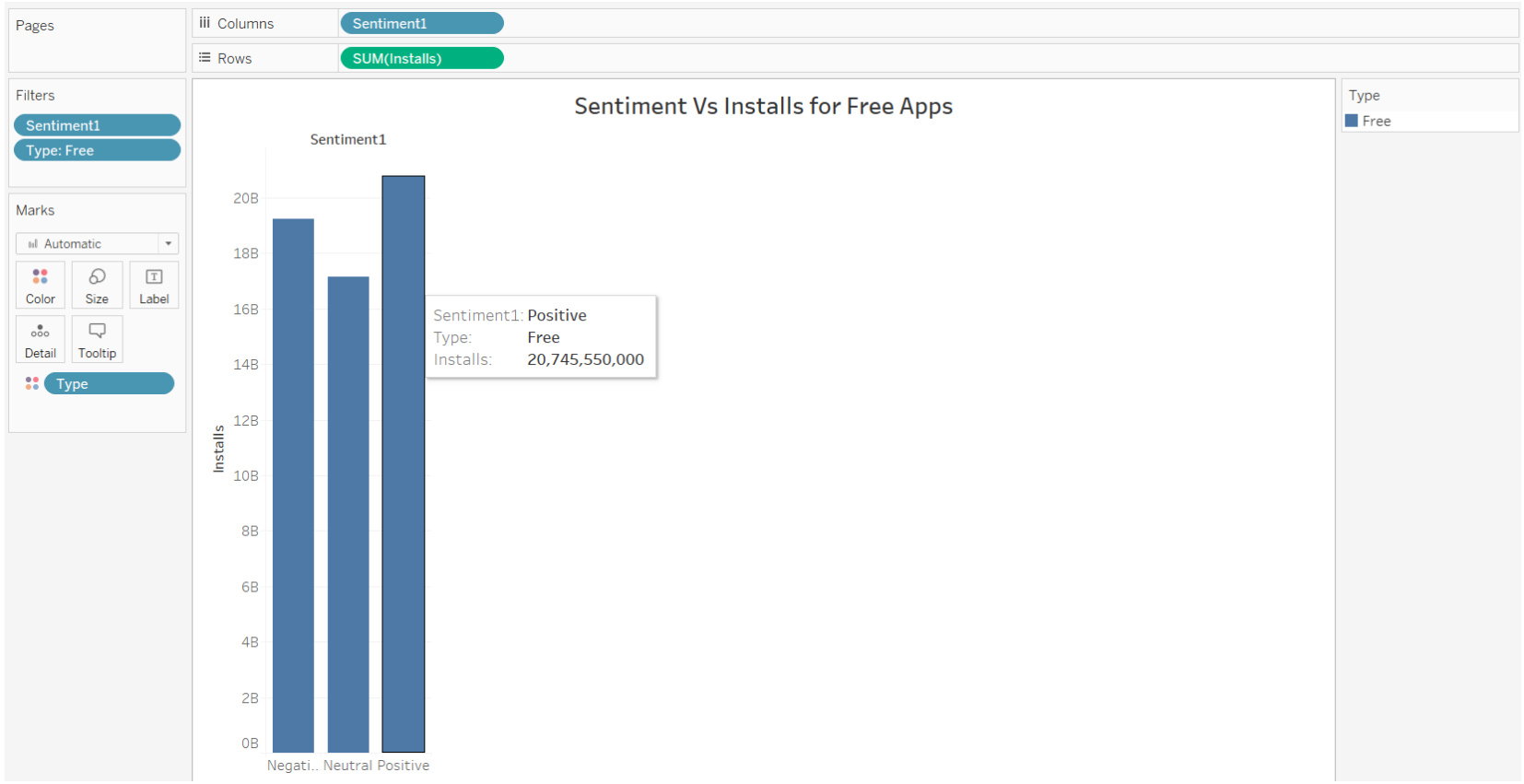
- Apps having Content Rating as Everyone are installed in majority and the percentage is 81.81%.



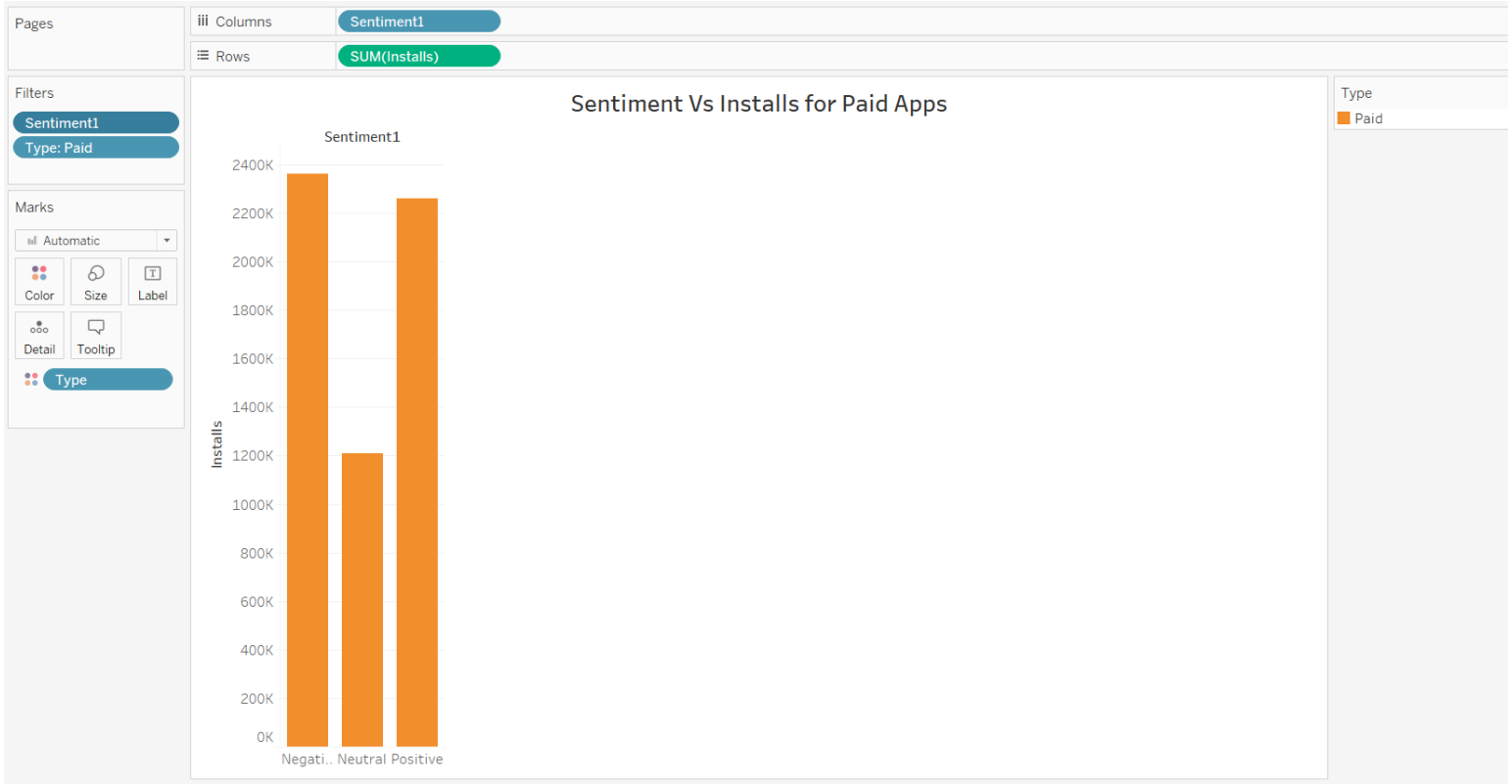
- From the above graph, we can infer that Apps having rating of 4.4 and 4.5 are installed the most.
- Apps having lowest rating are installed the least. Apps having rating less than 3.0 have low installation numbers.



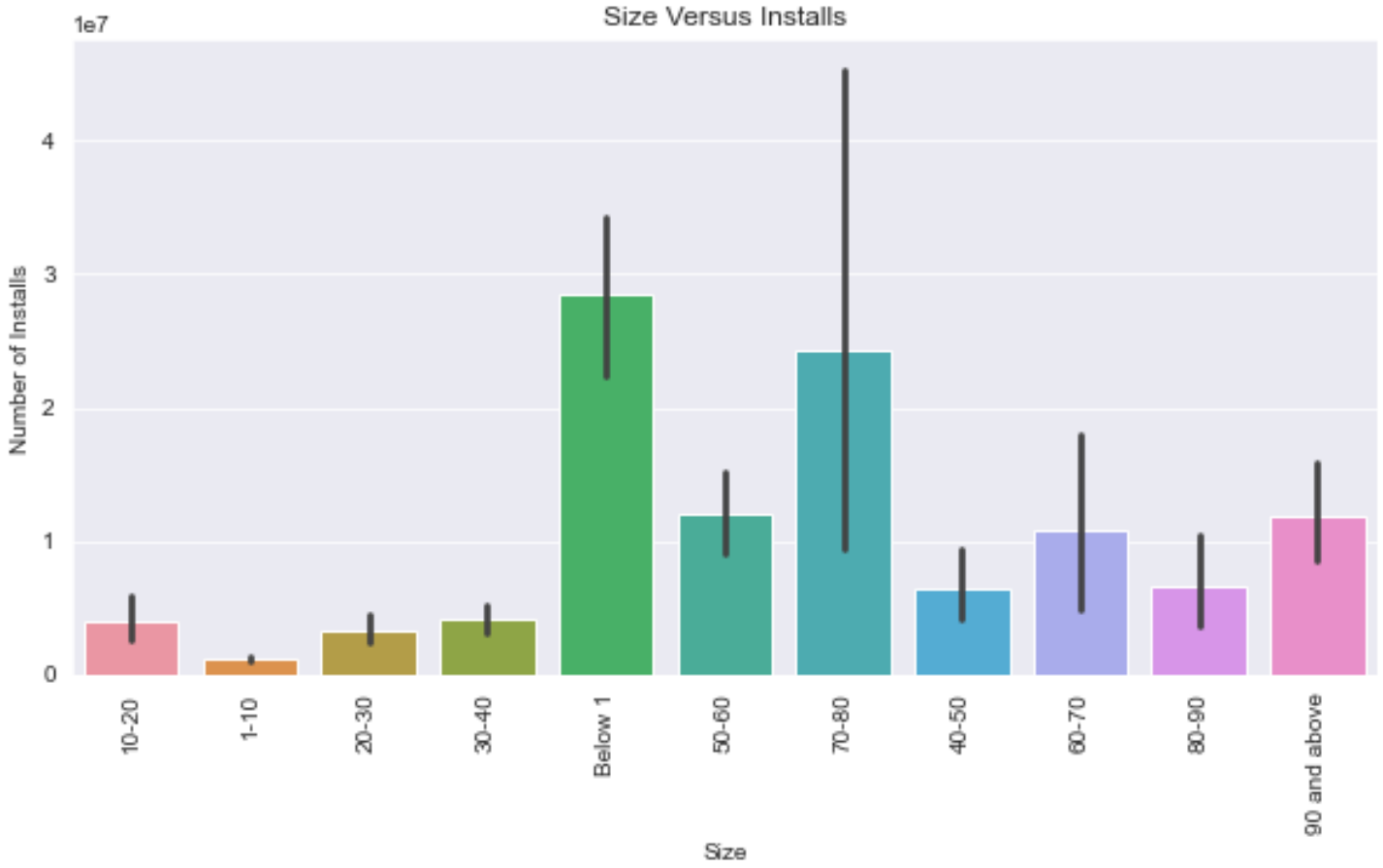
- Apps having lower android version compatibility are installed in lesser numbers than apps having higher android version compatibility.



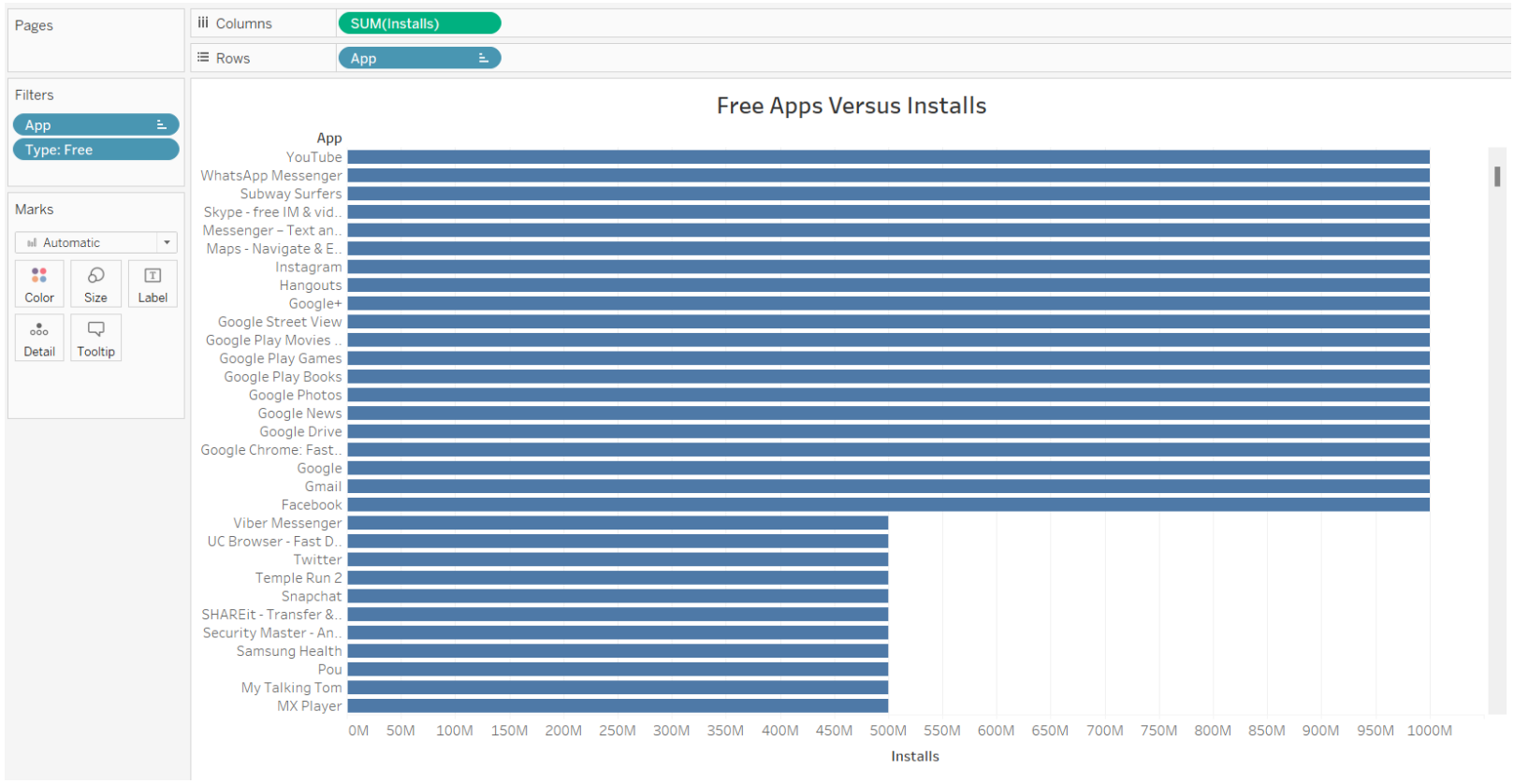
- For majority installs, positive reviews were given. Negative and neutral reviews are also common.



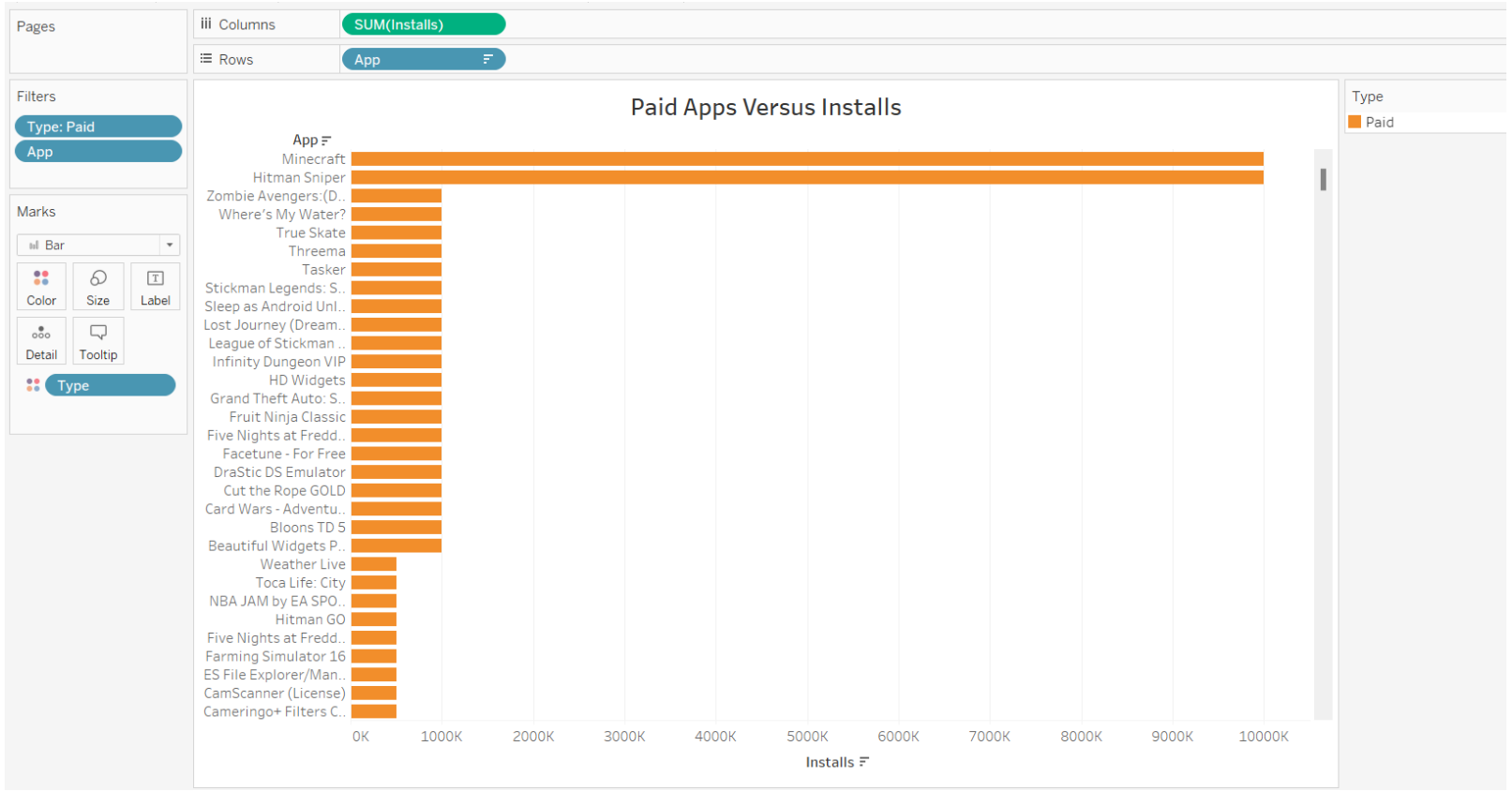
- In case of paid apps, most installations have negative reviews followed by positive and neutral



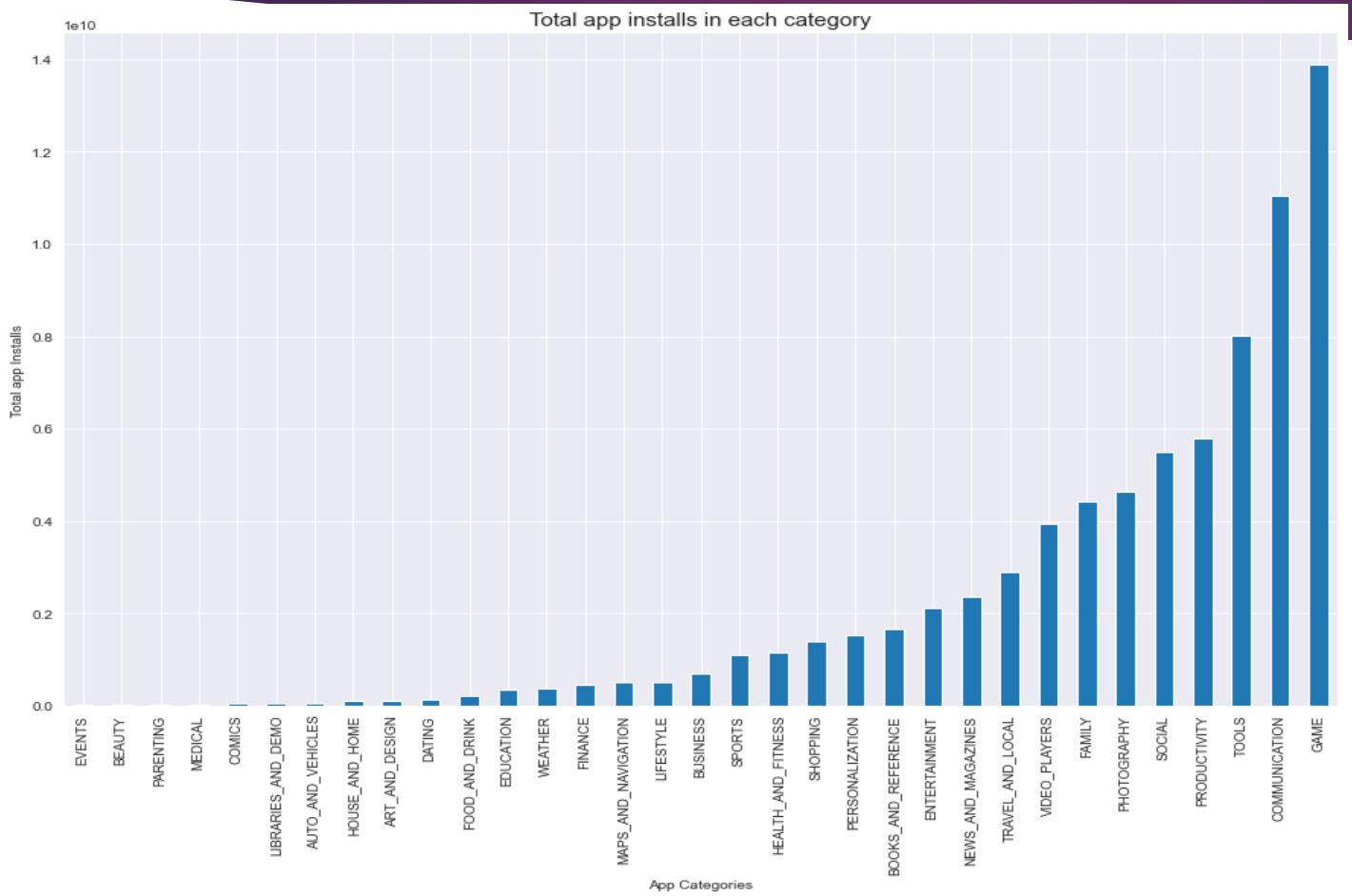
- Maximum installs were for apps whose size was below 1.



- Youtube , Whatsapp, Subway Surfers etc, are some of the apps having most installs.



- Mitman and Hitman Sniper had maximum number of installs while there were many apps with 0 installs


Total app installs in each category
1e10
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Total app installs
EVENTS
BEAUTY
PARENTING
MEDICAL
COMICS
LIBRARIES_AND_DEMO
AUTO_AND_VEHICLES
HOUSE_AND_HOME
ART_AND_DESIGN
DATING
FOOD_AND_DRINK
EDUCATION
WEATHER
FINANCE
MAPS_AND_NAVIGATION
LIFESTYLE
BUSINESS
SPORTS
HEALTH_AND_FITNESS
SHOPPING
PERSONALIZATION
BOOKS_AND_REFERENCE
ENTERTAINMENT
NEWS_AND_MAGAZINES
TRAVEL_AND_LOCAL
VIDEO_PLAYERS
FAMILY
PHOTOGRAPHY
SOCIAL
PRODUCTIVITY
TOOLS
COMMUNICATION
GAME
App Categories

Thankyou