Capstone Project

Exploratory Data Analysis

Play Store App Review Analysis

Team:
Nandeesh Umesha
Vilas Sonawane
Vikas Singh
Asad Mohammed

Introduction

formerly, google playstore is a digital distribution service operated and developed by google.

- Developers code the android applications and publish them in the google play platform for customers(android mobile users) to download.
- Some applications charge a premium to allow downloading and some are free to download without any payment.
- Developers earn the revenue through the apps they developed in the following ways:
  - Revenue source for paid apps:
    - One time payment to purchase the app
    - Monthly or yearly subscription to use the app
    - Or to use some special features of app the user have to pay

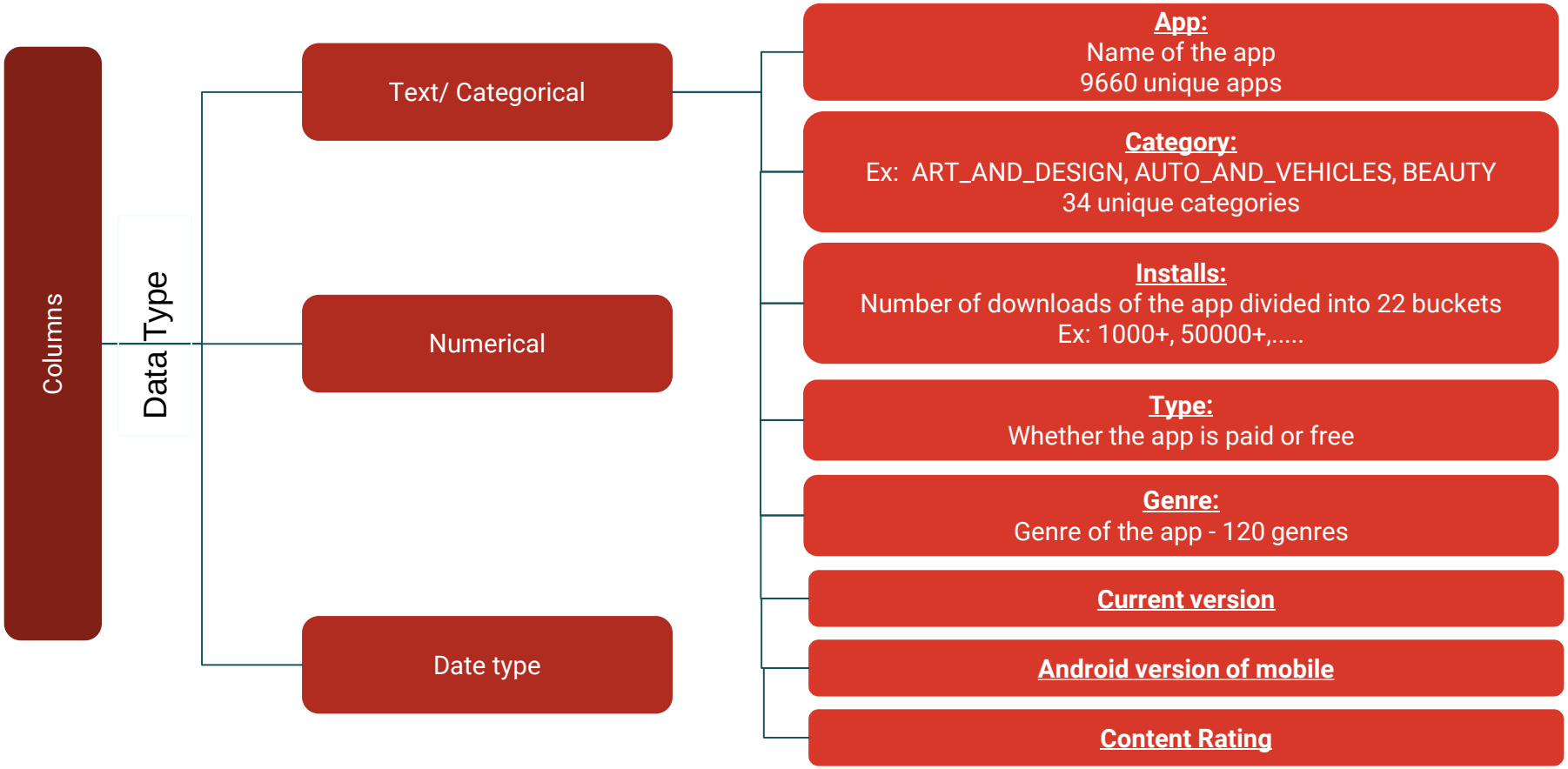
Introduction

- Revenue source for free apps:
  - Advertisements ,affiliate marketing and sponsorships
  - Earnings depend upon various factors like geography, app genre and time of the year.

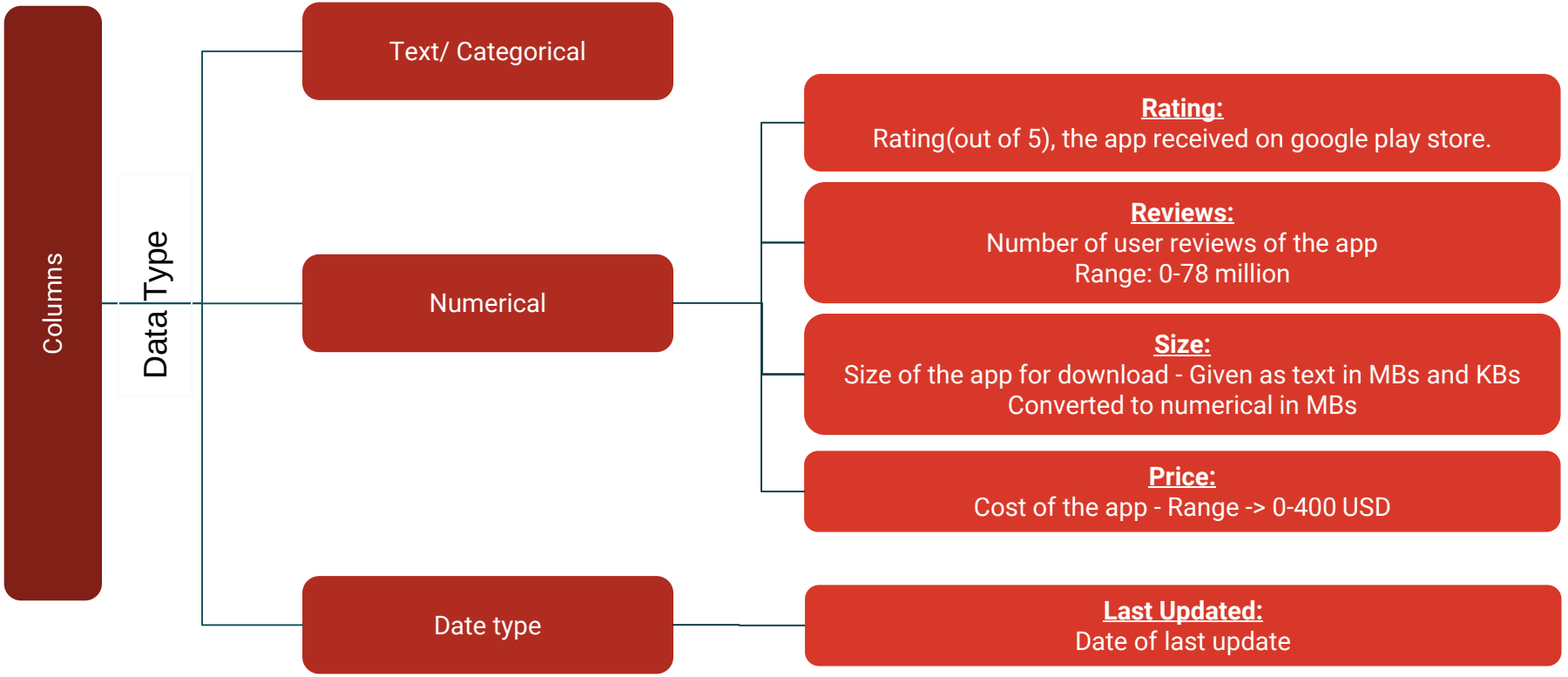
Problem Statement:

- Explore and analyze the Google play store data to discover key factors responsible for app engagement and success and obtain key insights.
- The dataset included two csv files:
  - **Play Store Data.csv:**
    - **Rows:** 10841 instances. Each row has info corresponding to an app
    - **Columns:** 13 columns
  - **User Reviews.csv:**
    - **Rows:** 64295 instances. Each row has info corresponding to a user review.
    - **Columns:** 5 columns

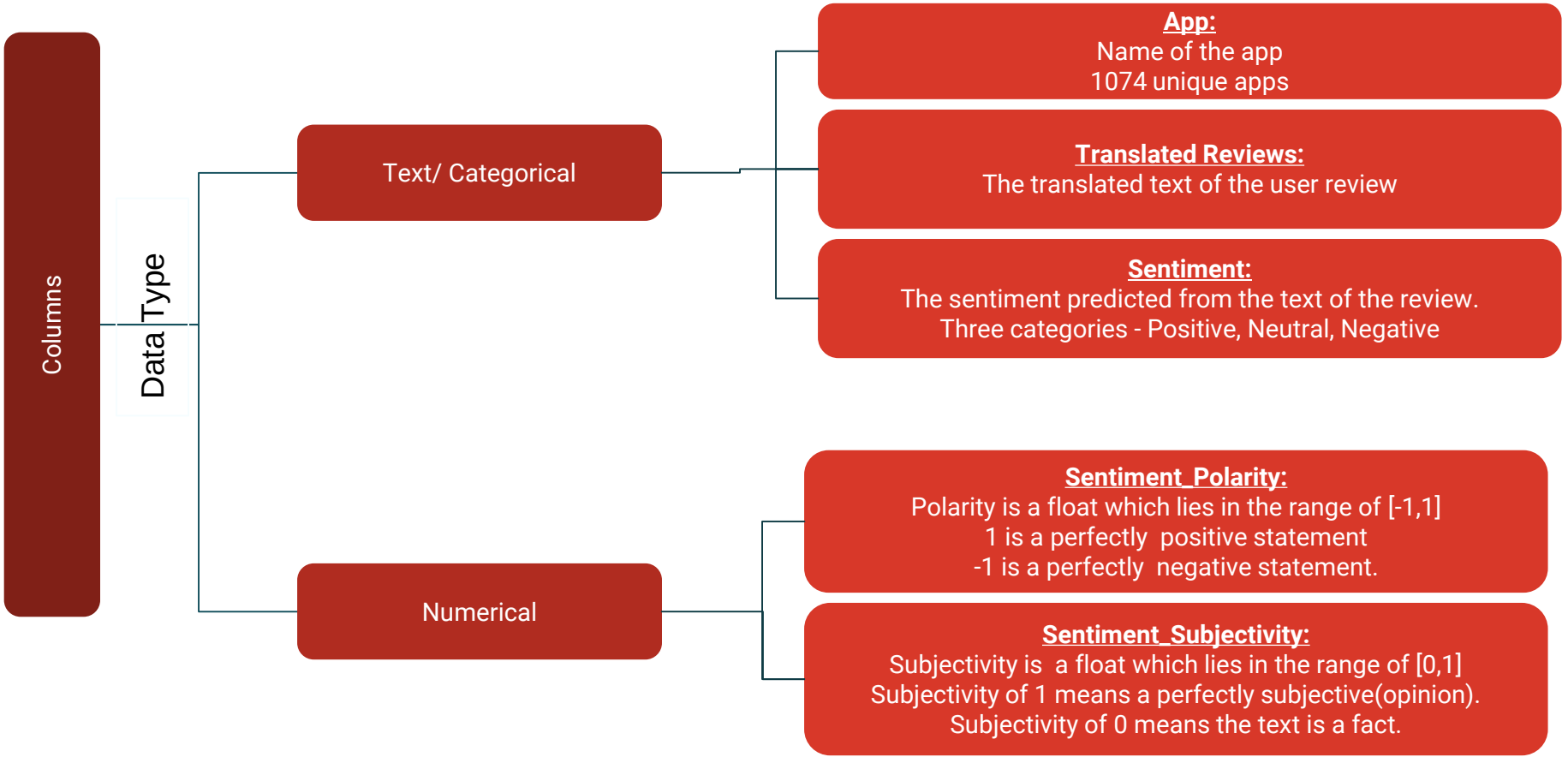
Data Dictionary- Play Store Data.csv



Data Description - Play Store Data.csv



Data Description - User Reviews.csv

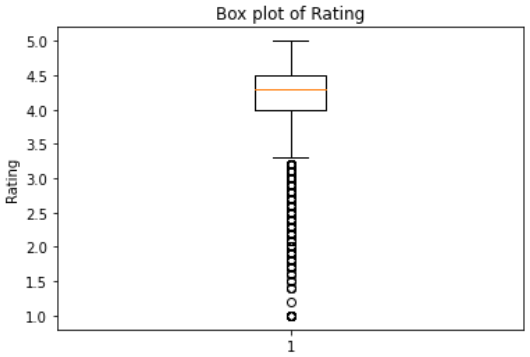


Data Cleaning - Play Store Data.csv

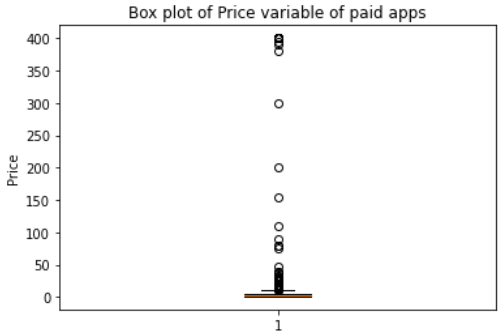
- **Rating:**
  - Only one instance had a rating above 5. So that row was dropped.
    - ~1470 missing values in Rating. Dropped since EDA only.
- **Reviews:**
  - String format converted to integer
- **Size:**
  - Apps with sizes listed as ‘Varies with device’ were dropped
  - String format converted to float in MBs.
- **Installs:**
  - String format converted to integer(new column added)
- **Price:**
  - String format converted to float(‘$0.99’ to 0.99)

Data Cleaning

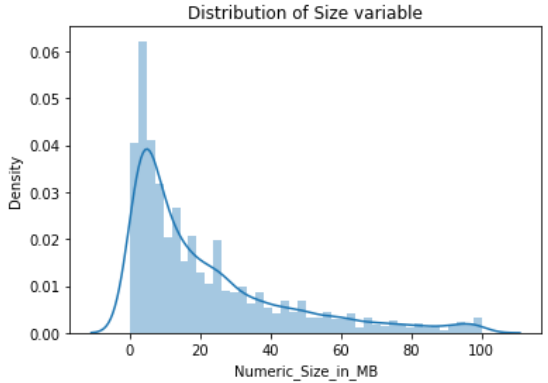
Ratings

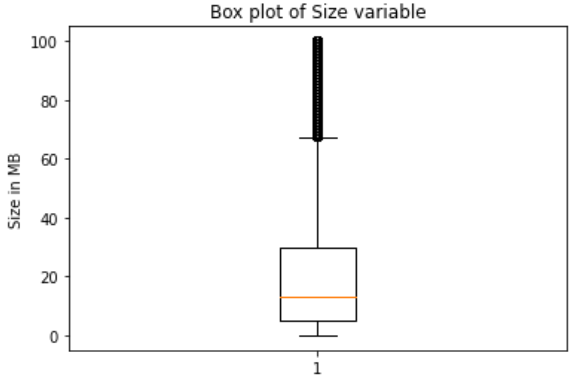


Price



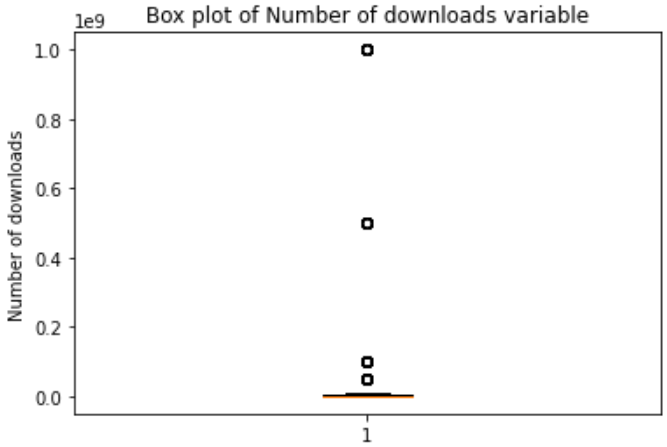
Size

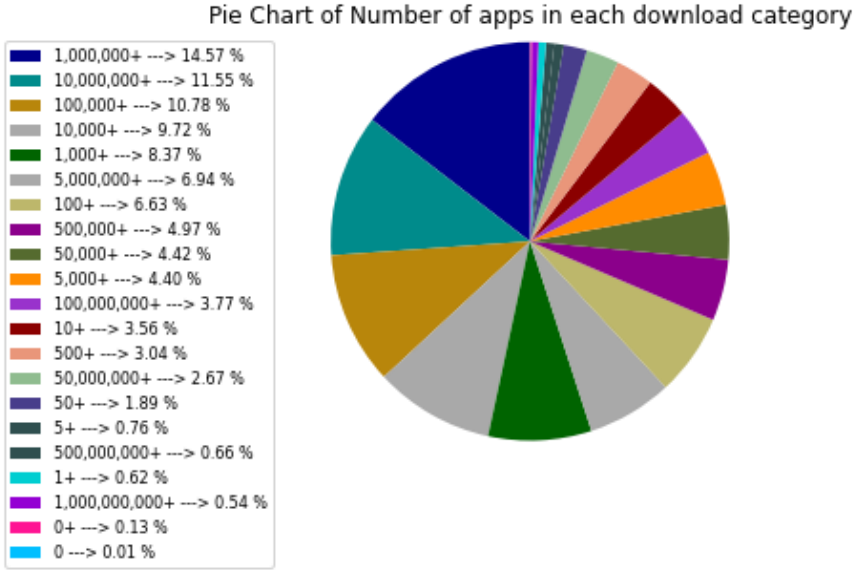




Data Cleaning

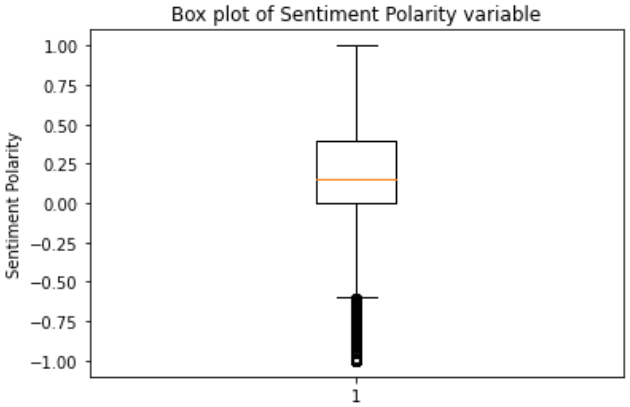
Installs

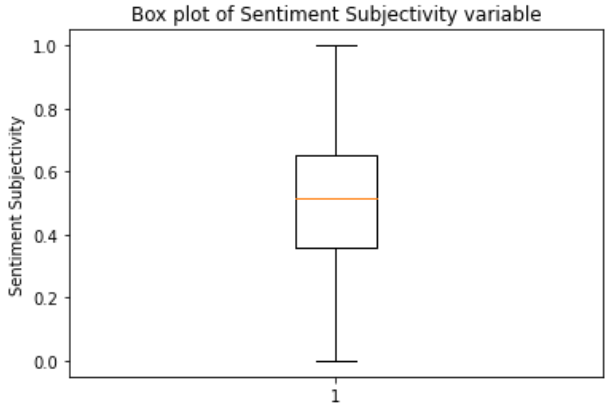




Data Cleaning

- **User Reviews.csv:**
  - ~27k missing ‘Translated user reviews’
  - Therefore, sentiment polarity, sentiment subjectivity and sentiment columns were also null values.
  - All those instances had to be dropped because those instances only had data about name of the app. They have no practical use.



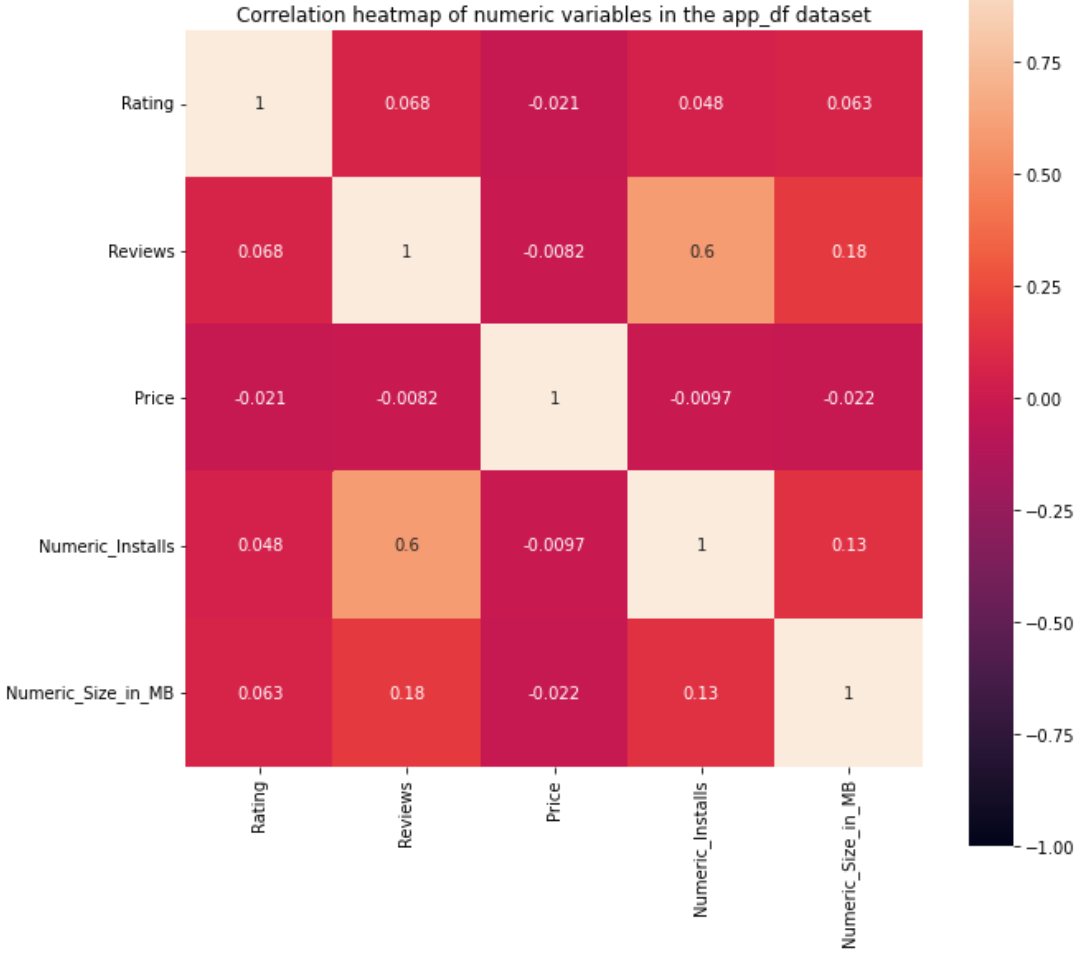


Let the investigation begin!!

I will find the answers at any cost

Mr. Data.csv

- Let’s see if there are any numerical variables which are collinear.
- The only significant correlation is between number of downloads and number of reviews.
- That is expected since the apps having more downloads will be used by more users and will have more reviews eventually.



1. Are free apps of poor quality?

- **Indicators of app quality:**
  - Rating
  - Sentiment polarity of user reviews

a. Comparison based on rating:

(All apps)	
Type	**Average rating**
Free	4.19
Paid	4.27

(Apps with more than 100 reviews)	
Type	**Average rating**
Free	4.21
Paid	4.35

1. Are free apps of poor quality?

b. Comparison based on user review:

- **Metrics for user reviews:**
  - Mean sentiment polarity
  - No. of positive reviews

Type	Mean sentiment polarity	No. of positive reviews (as a % of total reviews)
Free	0.20	64.66
Paid	0.18	67.92

- **Note:** The number of paid apps(9) is very low in comparison to the number of free apps that have written reviews.

1. Are free apps of poor quality?

Conclusion:

- In terms of rating, paid apps perform slightly better whereas in terms of sentiment of user review, free apps outperform paid apps slightly.
- In both the cases, the difference is subtle only.
- So, with the given dataset we cannot conclusively prove that paid apps outperform free apps in terms of user satisfaction/ quality.

2. Are free apps more popular?

- **Indicators of app's popularity:**
  - No. of downloads

a. Mean of downloads:

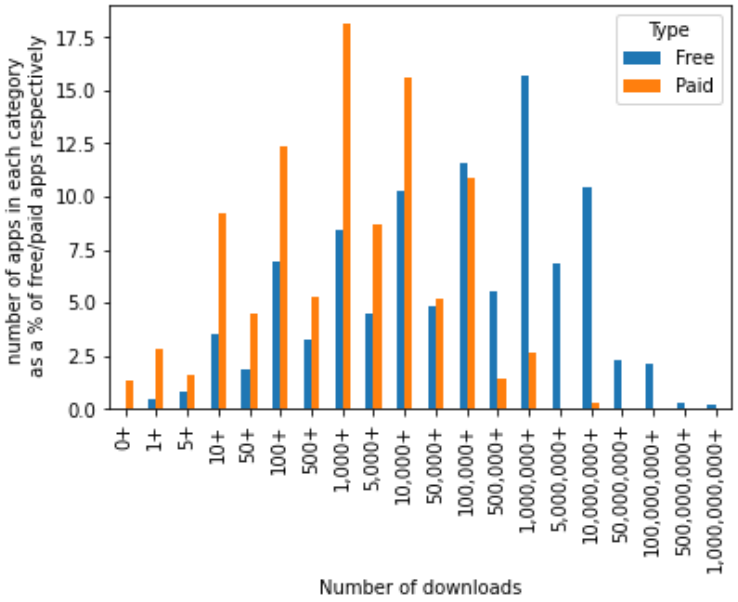
Type	Average downloads
Free	8.46 million
Paid	76,300

a. Median of downloads:

Type	Median downloads
Free	100000
Paid	1000

- A few apps like facebook and twitter may be abnormally impacting the mean downloads. Let us look at the median downloads.

2. Are free apps more popular?

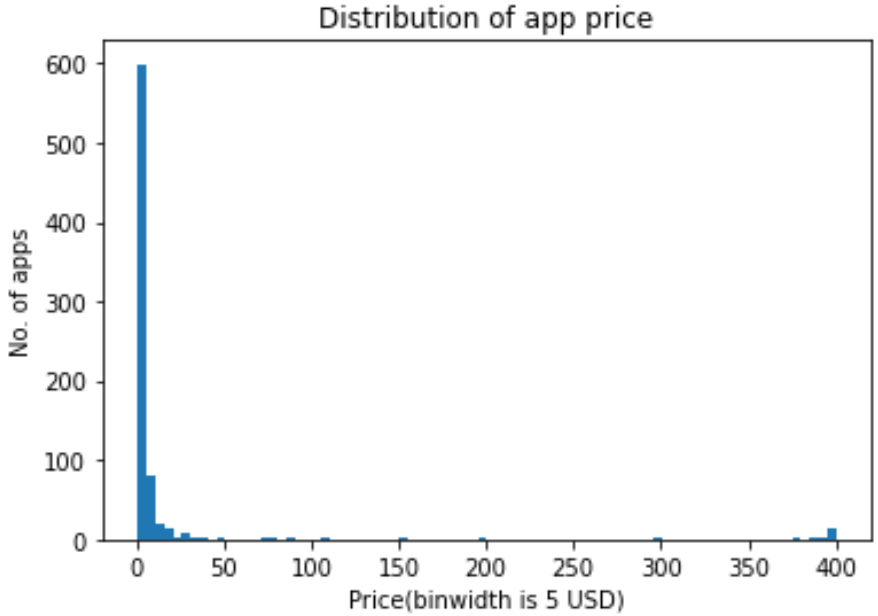


Conclusion:

- Free apps do have significantly more downloads than paid apps
- On average, free apps have 100 times more downloads than paid apps
- There are larger number of free apps in the bigger download buckets.
- There are only two paid apps with a million+ downloads. - Minecraft and Hitman sniper - both gaming apps
- With the given data, we can infer that free apps are indeed more popular

3. What type of apps are users ready to pay for?

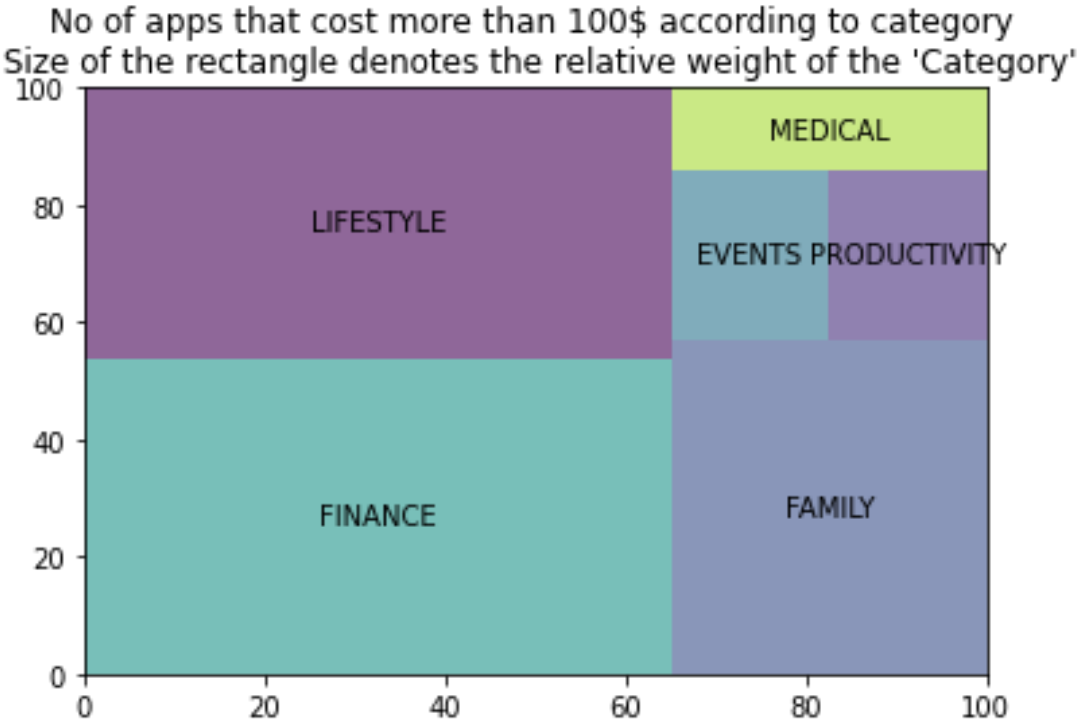
Distribution of app price(only paid apps)



Total number of paid apps	751
Apps costlier than $10	73 (~10%)
Apps costlier than $5	157 (~20%)

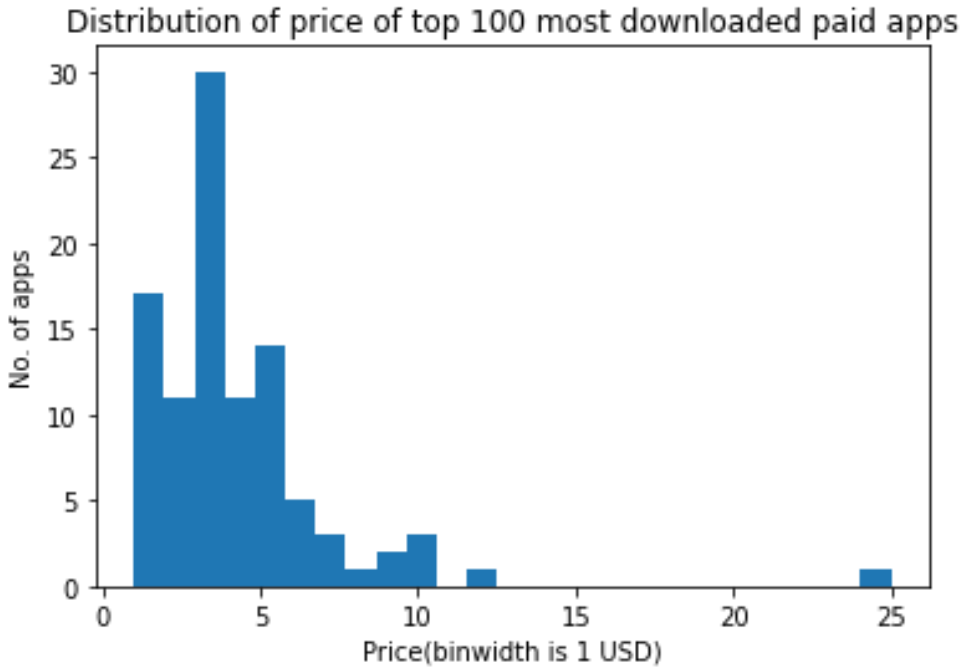
3. What type of apps are users ready to pay for?

categories of high cost apps(>$100)



3. What type of apps are users ready to pay for?

Price distribution of top 100 most downloaded paid apps



Apps costing less than $5	83 (~84%)
Apps costing less than $10	97 (~98%)

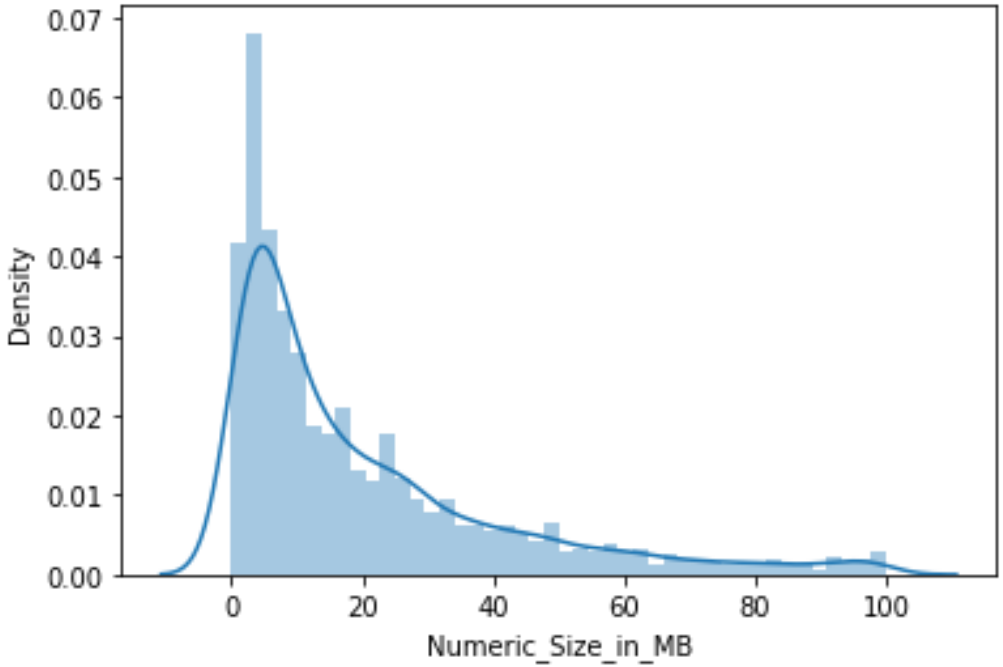
3. What type of apps are users ready to pay for?

Conclusion:

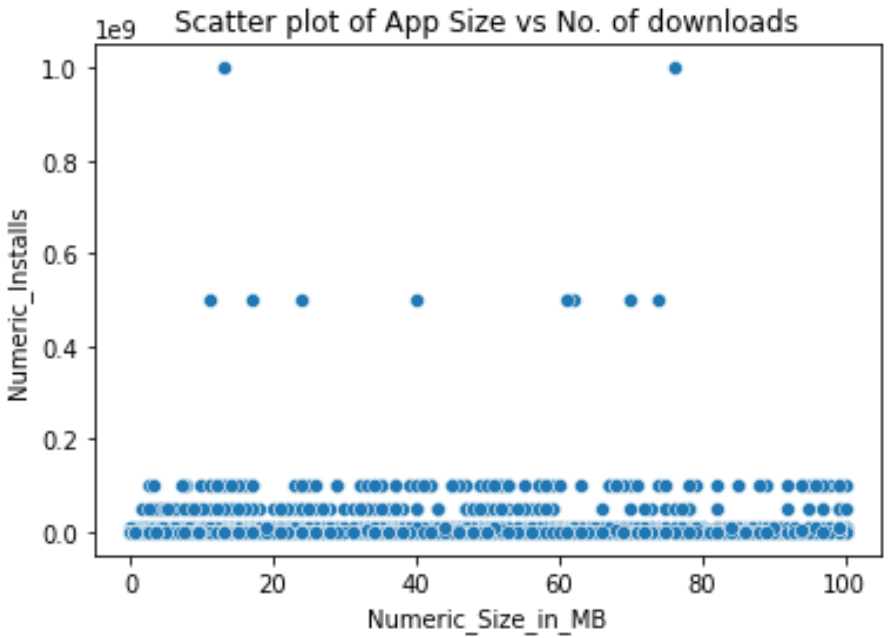
- Android app users does not seem much interested in paying to download an app.
- Among the few paid apps, the most popular ones are those costing less than $5.
- There are a few very costly apps related to finance, lifestyle and family with very few users(10000 on average)
- Note: Here the cost of an app is only the cost to download an app.

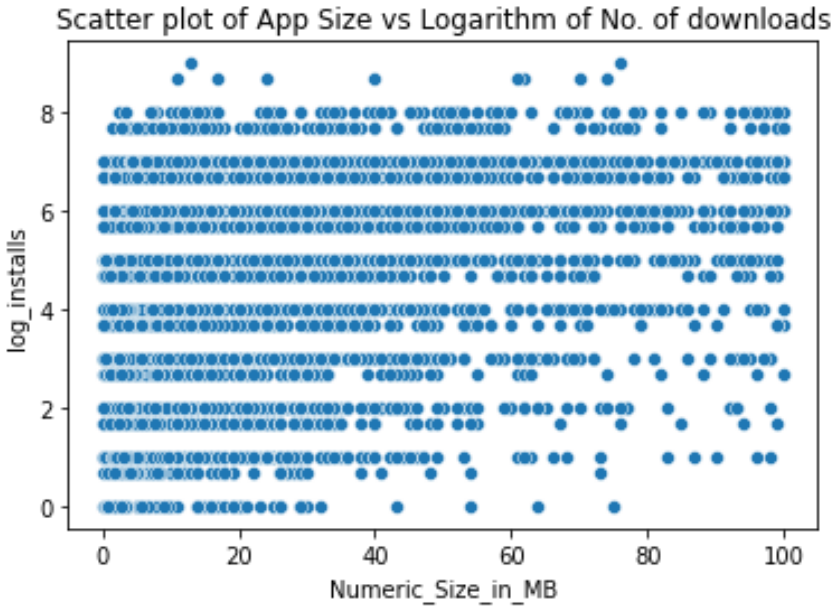
4. Are heavy apps not popular?

Distribution of app size(in MB)

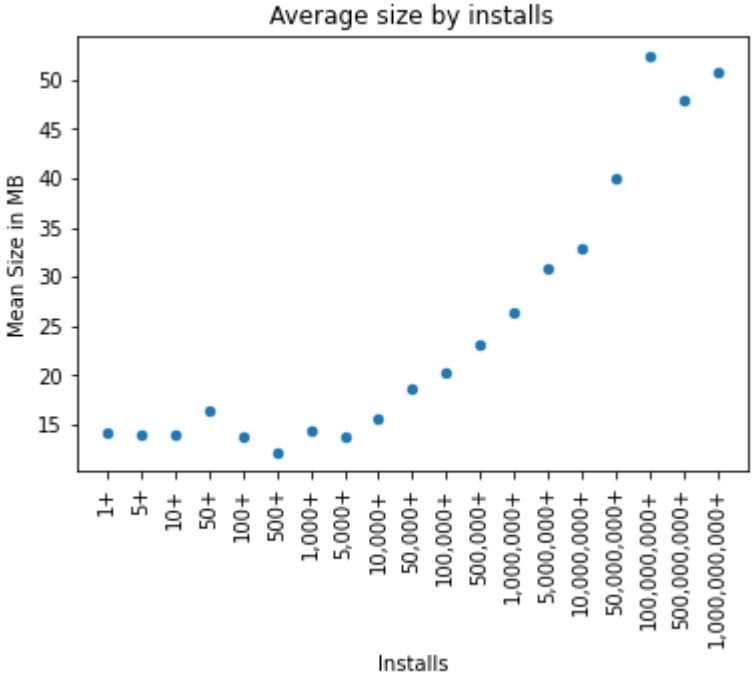


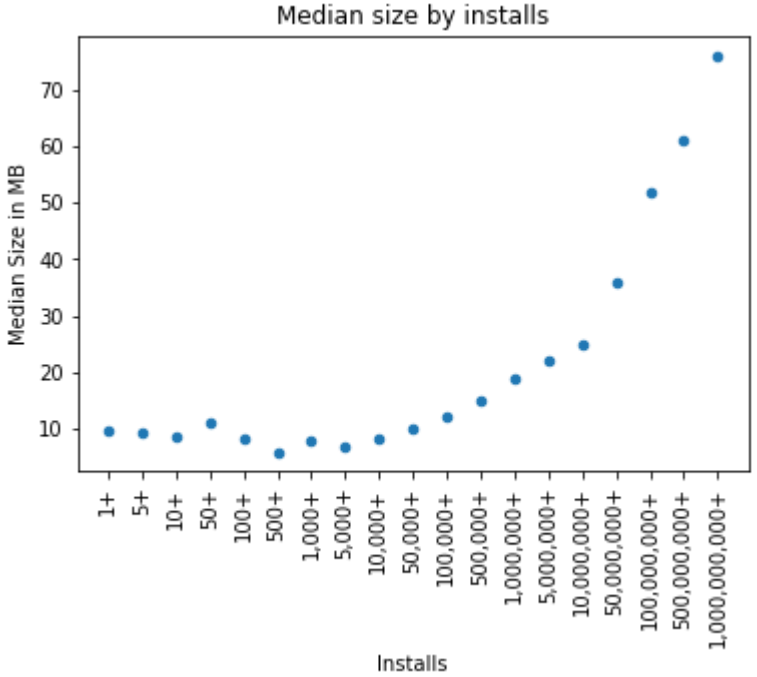
4. Are heavy apps not popular?





4. Are heavy apps not popular?





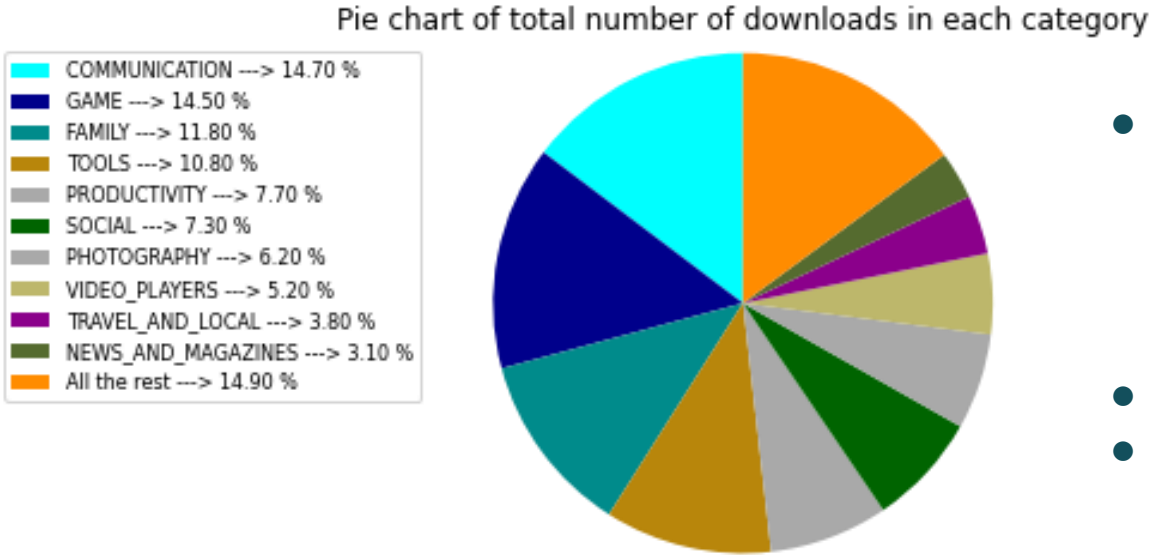
4. Are heavy apps not popular?

Conclusion:

- We can observe a counterintuitive insight that apps with higher downloads have higher average size!
- But we have to note that the max app size is 100 MB.
- So higher app size does not negatively affect an apps popularity if it stays within a reasonable limit.
- This may be attributed to technological developments in low cost storage and processor speeds.

5. Market Research of android apps

Category wise market share

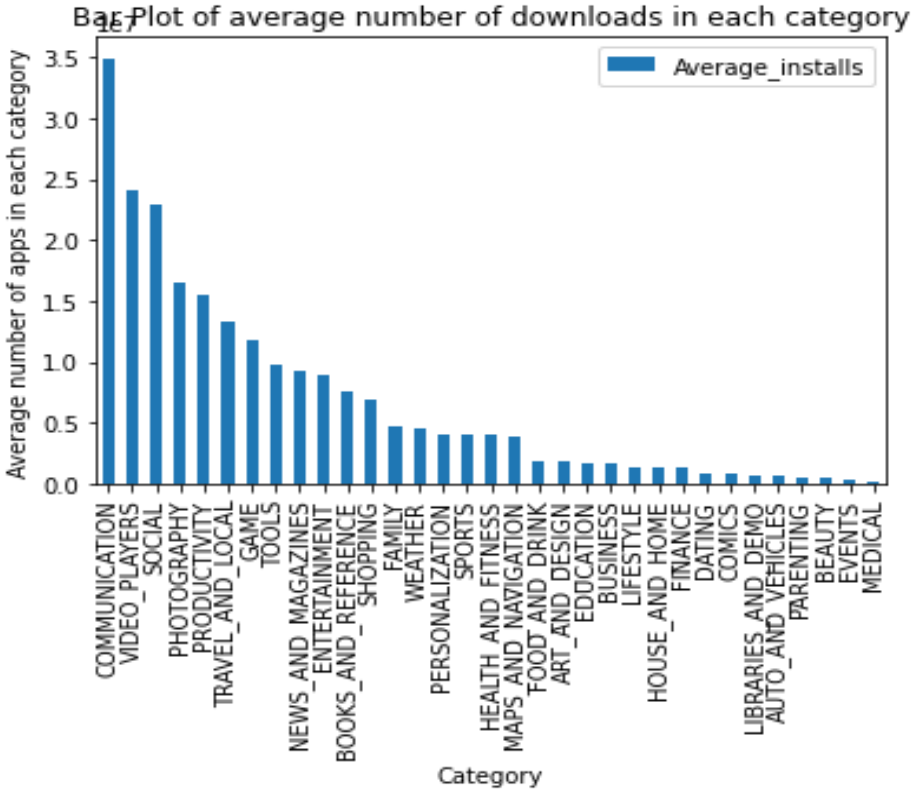


- Communication, game, family, tools, productivity and social category of apps have the most number of users.

- Top 5 categories capture 60%
- Top 10 capture 85%

5. Market Research of android apps

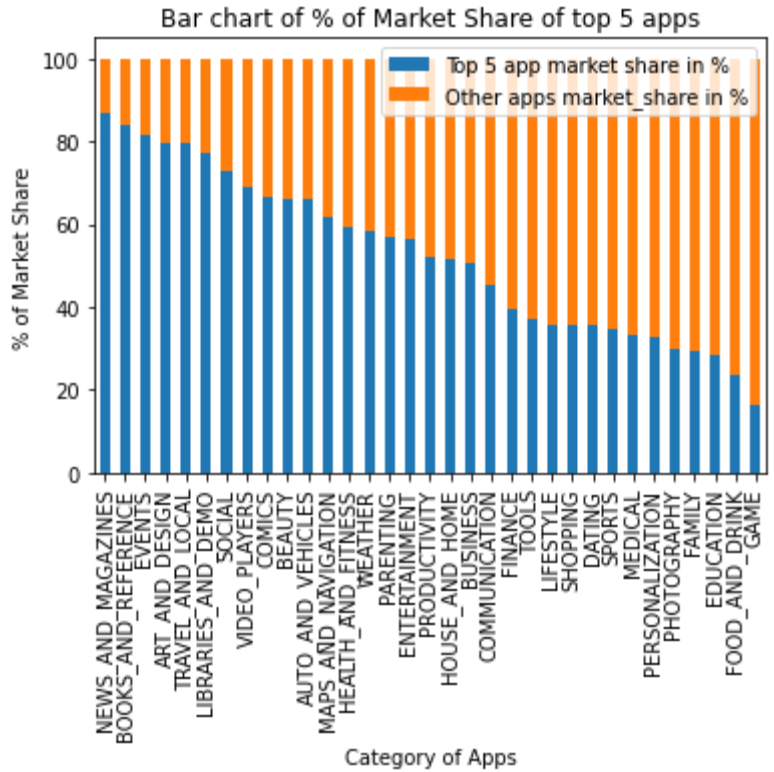
Average number of downloads in each category:



- Communication tops in total number of users as well as in average number of users.
- While the total number of users in game, family, tools categories is high, on an average, they have relatively lesser number of users. This may indicate that the number of apps in these categories is large.
- Photography and travel_and_local categories have more users on average. It means they have fewer apps with good acceptance by customers.

5. Market Research of android apps

Oligopoly?



Oligopoly categories:(2/3rd market by top 5)

News_and_magazines, Books_and_reference, Events, Art_and_design, Travel_and_local, Libraries_and_demo, Social, video_players

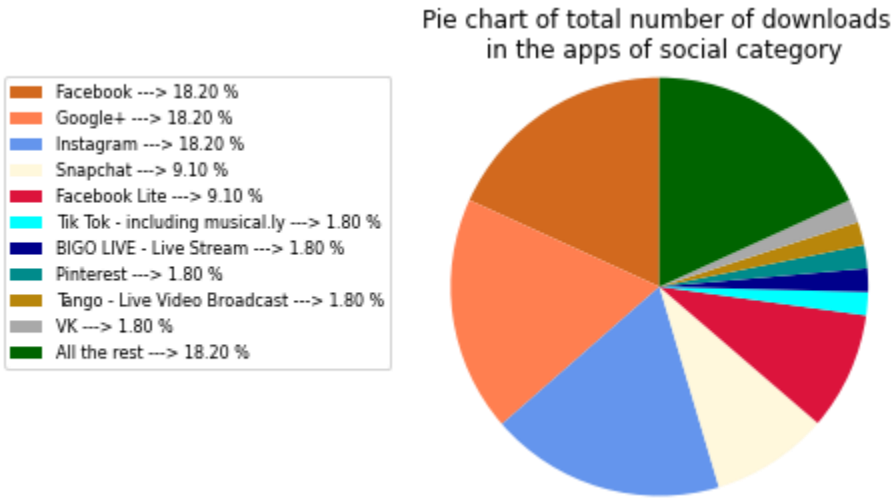
Competitive categories:(less than 1/3rd market by top 5)

Game, Food_and_drink, Education, Family, Photography, Personalisation and Medical

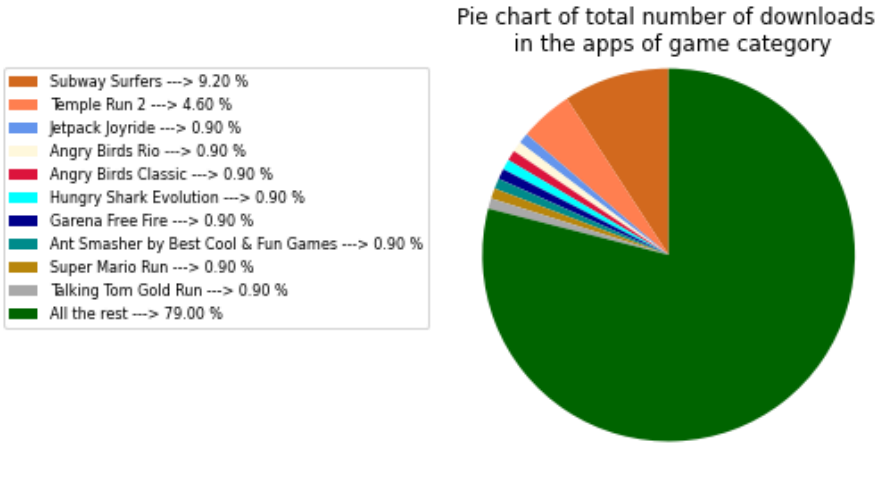
15 out of 33 categories have a google app among the top 5! - Compulsory apps?

5. Market Research of android apps

Social Category:



Game Category:



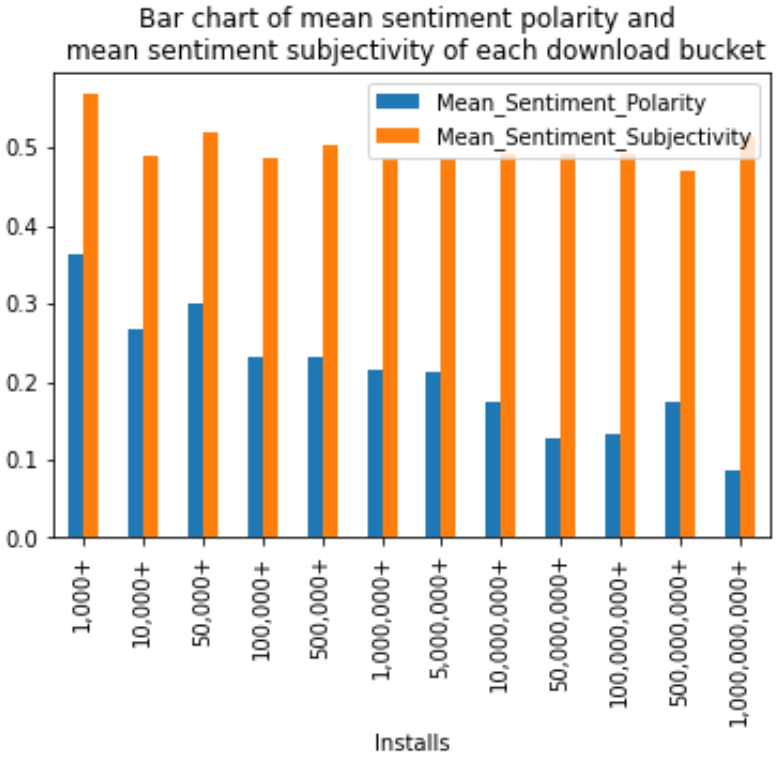
- Facebook's apps -(facebook, facebook lite, instagram) - 45% of the market.
- Google+ and Snapchat capture - 27%. Difficulty for new entrants.
- Apart from subway surfers(9.2%) and Temple Run2(4.6%), all the gaming apps have less than 1% market share. There is diversity in gaming market.

5. Market Research of android apps

Conclusion:

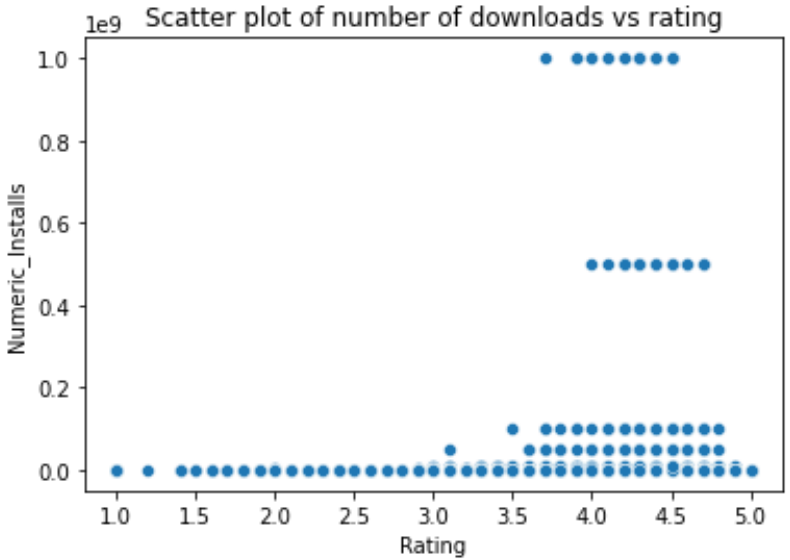
- Just 5 categories of apps have a market share of 60% and top 10 capture almost 85% of the android app market. So the digital space seems to be more relevant and useful for the ordinary customer in few areas of life only.
- Apps belonging to the communication category have the largest number of users in total as well as on average.
- Few big companies dominate heavily in some categories. Ex: Facebook alone has 45% of market share in social apps and google apps dominate 15 out of 33 categories with more than 10 apps.
- Gaming is the most diverse and competitive category.

6. Sentiment polarity, Sentiment subjectivity and Popularity of an app

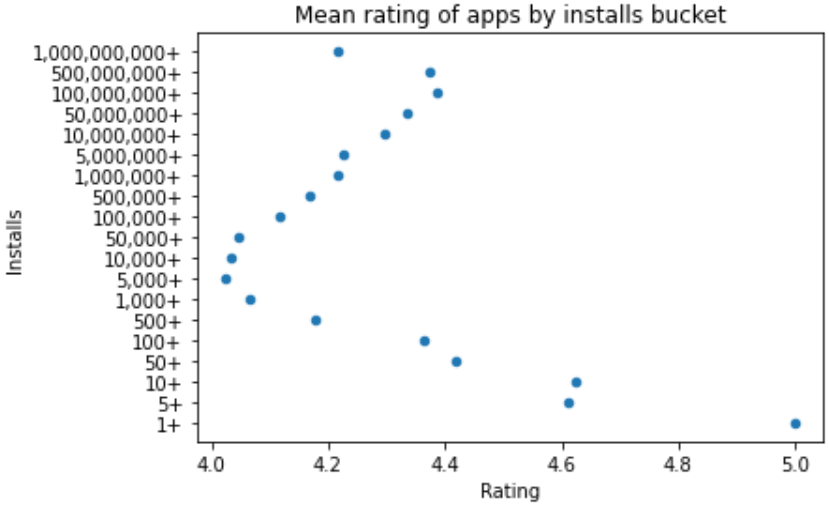


- Mean sentiment subjectivity does not show any interesting pattern.
- But, Mean sentiment polarity shows a negative relation with the number of downloads. The higher the downloads, lower the mean sentiment polarity.
- It means that apps that have more users have less positive written reviews overall.

6. Are apps with good rating more popular?

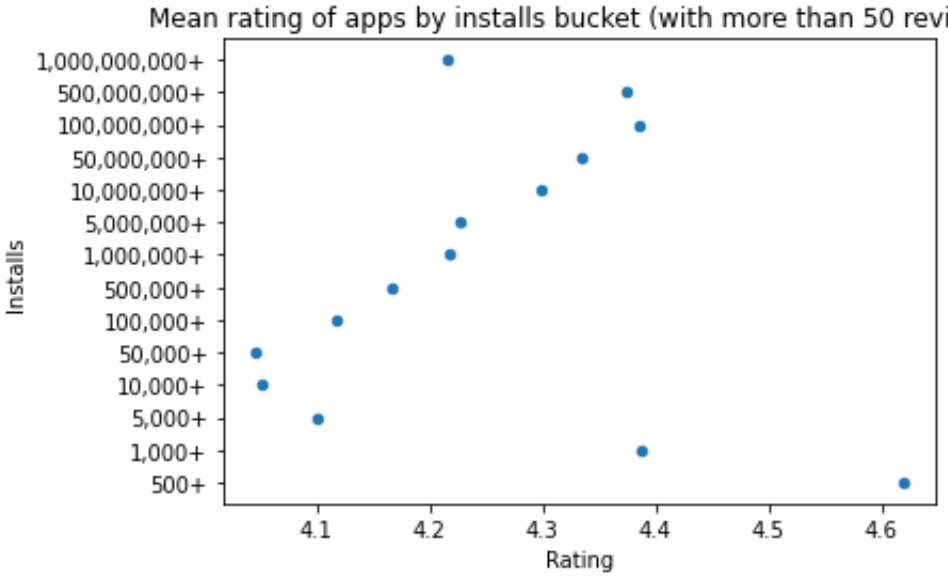


- The scatter plot is not so intuitive since the number of downloads varies by a huge number.



- For apps with more than 5000 users, there is a clear positive correlation
- Apps having less than 5000 downloads may have obtained influenced/ paid reviews.

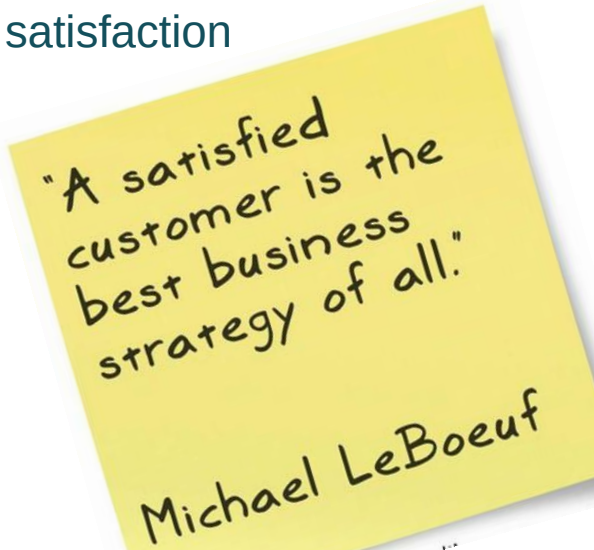
6. Are apps with good rating more popular?



- Apart from the few outliers, there is a clear positive trend between rating of an app and number of installs.

Conclusion:

- Apps with higher rating are indeed more popular.
- The above analysis proves the commonsensical intuition.
- Customer satisfaction matters.



In Summary…

- User satisfaction does not differ between paid apps or free apps. The customers seem equally satisfied with paid apps or free apps.
- The price here corresponds only to the price paid to download an app and there may be paid services offered in free apps also.
- When it comes to popularity of an app, free apps outrank paid apps by a 100 times.
- Android users do not prefer paying to download an application.
- 85% of the most popular paid apps cost less than 5$.
- For a new developer planning to launch an app, it is preferable to make it free for download and earn through ad revenue.
- If ads are not preferred, the cost of the app should be less than $5.
- Few super costly apps are available mostly in the finance, family and lifestyle category but with very few users.

In Summary…

- Larger size of the app does not necessarily lead to low popularity, infact, more popular apps have higher size on average.
- Mobile apps have an opportunity to acquire users in a lot of categories since, currently, only 5 categories - Communication, game, family, tools, productivity and social capture 60% of the android app market.
- In some categories like social, video players, news and travel, the big tech players like google, facebook, tiktok pose a huge entry barrier for new entrants.
- Categories like gaming, education are competitive and are not so dominated by few.
- Customer satisfaction(Rating) does lead to popularity.