THE CHARACTERISTICS OF ACHIEVING SUCCESS

Google Play is an app store for users who are using Android operating system devices to download applications. There is a huge number of applications published and downloaded by public in Google Play store, so we are going to discuss the characteristics of different application.

The number of users installed the application is the definition of success. The data set shows the characteristics and the relationship with achieving success.

1. <u>Summary about application data</u>

A data set, which is "googleplaystore.csv", includes all detail information of the applications on Google Play. The data set includes the following variables:

1. App - Application name
2. Category - Category the app belongs to ss
3. Rating - Overall user rating of the app (as when scraped)
4. Reviews - Number of user reviews for the app (as when scraped)
5. Size - Size of the app (as when scraped)
6. Installs - Number of user downloads/installs for the app (as when scraped)
7. Type - Paid or Free
8. Price - Price of the app (as when scraped)
9. Content Rating - Age group the app is targeted at - Children / Mature 21+ / Adult
10. Genres - An app can belong to multiple genres (apart from its main category). For example, a musical family game will belong to Music, Game, Family genres.
11. Last Updated - Date of last updated for the app in Play Store (as when scraped)
12. Current Ver - Current version of the app available in Play Store (as when scraped)
13. Android Ver - Min required Android version (as when scraped)

There are 10841 rows and 16 variables of the application's information form dataset.

Since we will mainly use the amount of installation, we use the factor function to sperate the same range of amount into different group.

(Original dataset summary)

```
> summary(store)
     App              Category            Rating         Reviews            Size
 Length:10841      Length:10841       Min.   : 1.000   Min.   :       0   Length:10841
 Class :character  Class :character   1st Qu.: 4.000   1st Qu.:      38   Class :character
 Mode  :character  Mode  :character   Median : 4.300   Median :    2094   Mode  :character
                                      Mean   : 4.193   Mean   :  444153
                                      3rd Qu.: 4.500   3rd Qu.:   54776
                                      Max.   :19.000   Max.   :78158306
                                      NA's   :1474     NA's   :1

    Installs            Type              Price          Content Rating        Genres
 Length:10841      Length:10841       Length:10841      Length:10841       Length:10841
 Class :character  Class :character   Class :character  Class :character   Class :character
 Mode  :character  Mode  :character   Mode  :character  Mode  :character   Mode  :character




 Last Updated      Current Ver        Android Ver
 Length:10841      Length:10841       Length:10841
 Class :character  Class :character   Class :character
 Mode  :character  Mode  :character   Mode  :character
```
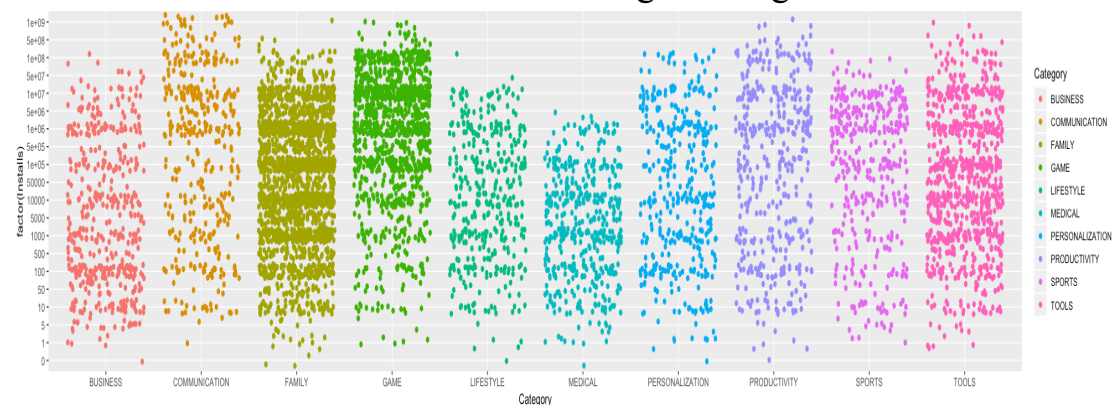
2. <u>Data Analysis</u>

The following analysis will be processed through Category, Type, Rating, Reviews and the number of users installed the application.

Category & Type

Install distribution among 10 categories

The above scatter plot shows the distribution of the installation among ten categories. We can observe that the largest number of applications is from category of family, and the greatest number of installed applications is from communication's category. It implies that there is no strong relationship between category and installation as it only represents the number of applications published among different categories, but it does not mean that family's category is one of the categories having the most installation.

Number of applications among categories in top installation group

Type among categories in top installation group
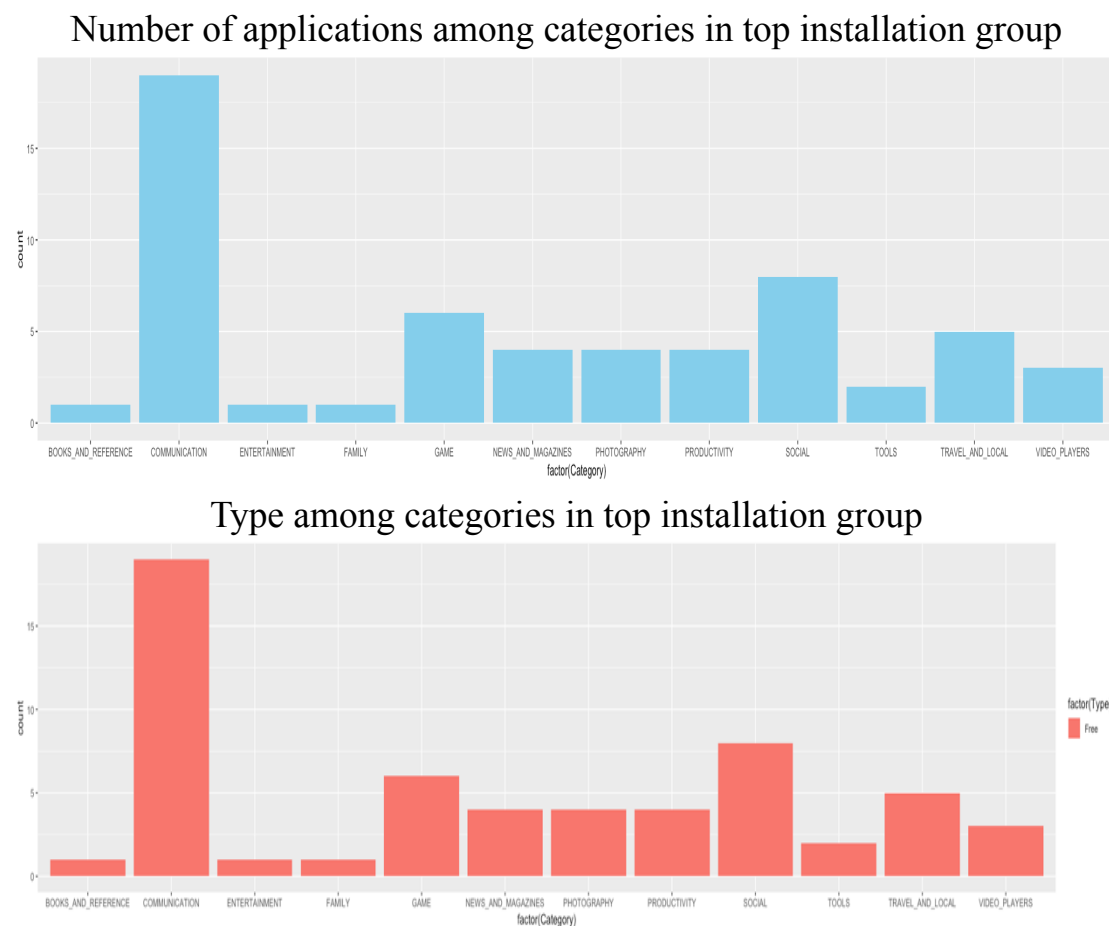
The upper bar chart shows all the applications that have 1,000,000,000+ amount of installation and their related category. 1,000,000,000+ is the largest amount of the installation. The chart also shows that the communication, social and game are the top three categories in this largest installation

group. However, these categories cannot reflect the relationship with achieving success.

The lower bar chart shows the type of applications of the most installation group. It shows that all the applications in this installation group are free of charge. All the applications can be separated into free or paid, and the free applications are more popular because it has a larger amount of installation than paid applications.

Type of application among different installation groups
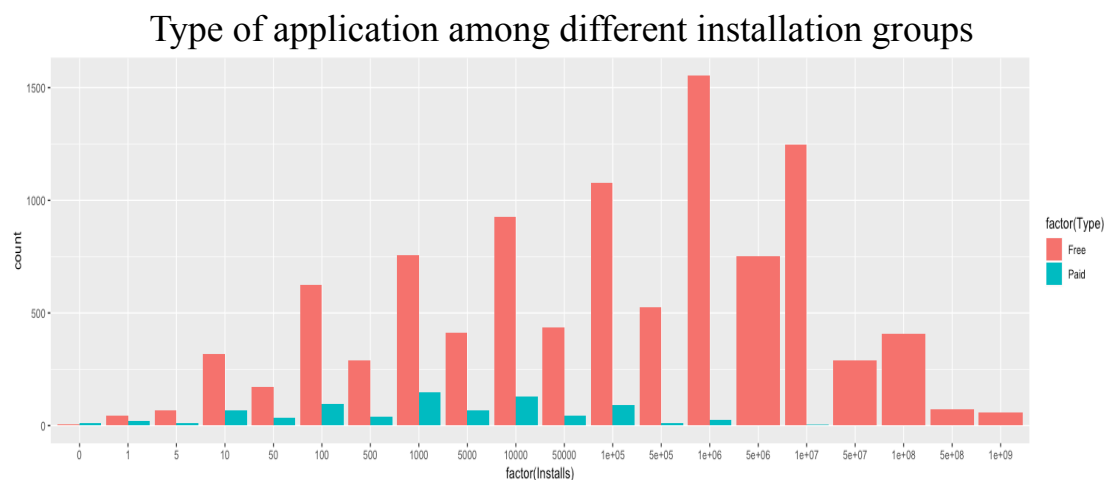
According to the above graph, all the types of installation groups included free applications, and free applications are nearly more than paid applications, except the installation group which is zero amount. Also, some of the groups only have free application, like 5,000,000+(5e+06), 500,000,000(5e+08) and 1,000,000,000(1e+09) etc. Thus, we can observe that the proportion of free application is much larger than paid application and the top four installation groups are only from free application.

To conclude, the free application is one of the characteristics of achieving success.

Type & Rating
Rating among different installation groups and type
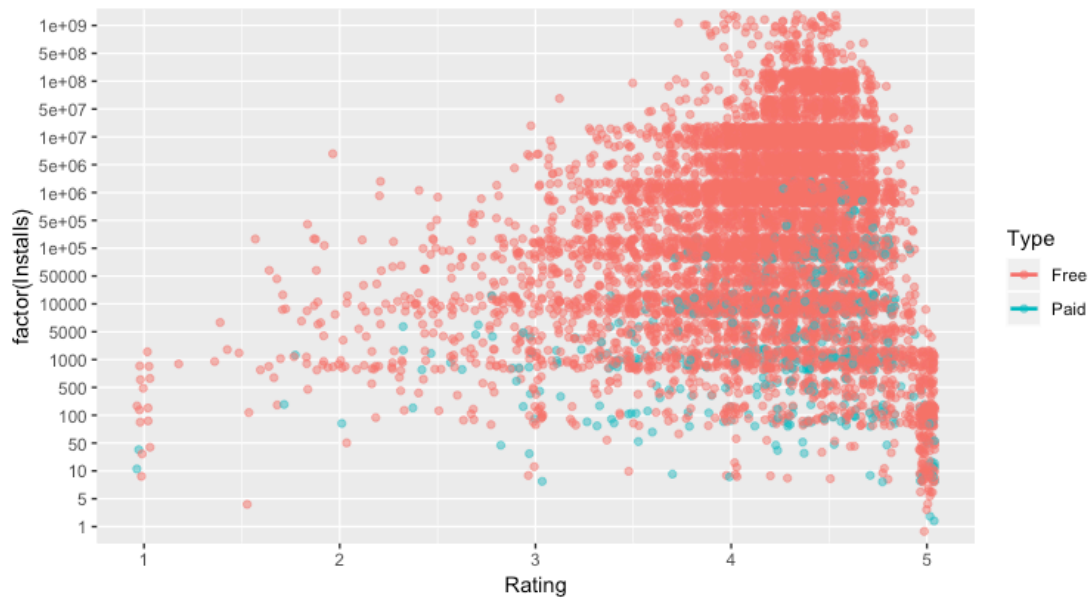
The above graph indicates the users' rating for applications. The amount of rating from free application is much more than paid application. For the level of rating, it mainly centralizes in between rating four and rating five. We can observe that the larger amount of installation, the higher rating is voted for both free and paid applications.

Reviews

Number of reviews among different installation groups
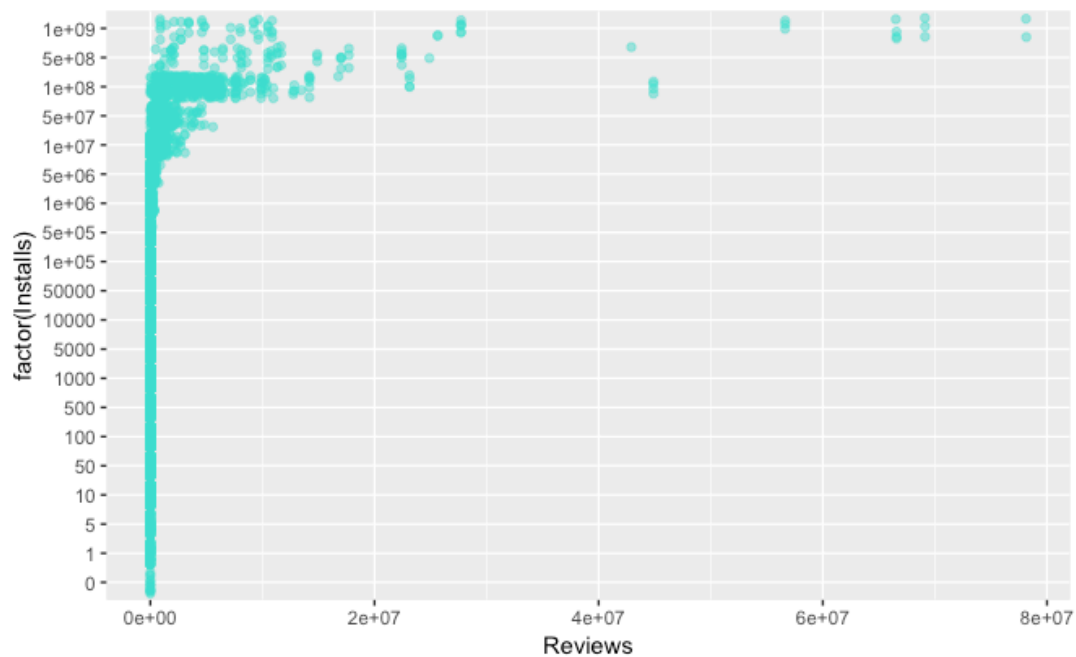
The graph above shows that the number of reviews among different installation groups. It is mentioned in the above that in overall there are more fewer reviews. It is also observed that a few but larger number of reviews centralized on higher amount of installation. We conducted that those apps which have the larger amount of installation would receive more reviews from users.

3. Conclusion

The type (paid / free) of applications has a larger effect on the number of users downloaded for the apps than category. Free applications basically indicated that it is easier to have larger number of installations from users. Larger number of reviews as well.

The category which has a mass number of apps in between does not equal larger number of user download. Part of the reason may be due to more current popularity or trend but perhaps it may be more related to user's common preference indication in lifestyle.