



---

SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE  
WARSAW SCHOOL OF ECONOMICS

Student's Name and Surname: **Ngoc Bao Dao**

Student No: **64588**

Specialization: **Big Data – Advanced Analytics**

## **Shiny Project**

# **Analysis and Visualization of Google Play Store User Reviews**

Subject: *Querying, Data presentation and Visualization*

Lecturer: **Dr Jarosław Olejniczak**

## I. About the Dataset

### *Context*

While many public datasets provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, I found out that on Kaggle page, there was a dataset which was nicely indexed appendix-like structure to allow for simple and easy web scraping. On the other hand, Google Play Store uses sophisticated modern-day techniques (like dynamic page load) using JQuery making scraping more challenging.

*Dataset source: <https://www.kaggle.com/tanetbosshow-to-get-high-rating-on-play-store/data>*

### *Content*

The set is a collection of 10,841 applications and consists of 13 different variables, which are as following:

- 1. App** - a qualitative variable, contains the names of application.
- 2. Category** - a qualitative variable, contains the category of each application (33 categories).
- 3. Rating** - a quantitative variable, concerns the individual reviews of each application (rating in scale 1-5).
- 4. Reviews** - a quantitative variable, contains the number of reviews of applications.
- 5. Size** - a quantitative variable, describing the size of applications.
- 6. Installs** - a quantitative variable, showing the number of installation.
- 7. Type** – a qualitative variable shows whether the app is free or paid.
- 8. Price** - a quantitative variable, showing the price of application (in USD).
- 9. Content rating** - a qualitative variable, revealing types of people who have rated for applications.
- 10. Genres** - a qualitative variable, indicating the application's genres.
- 11. Last update** – a date variable shows the date when applications were updated.
- 12. Current Ver** – a qualitative variable, reflecting current version of applications.
- 13. Android Ver** - a qualitative variable, revealing the requirement of Android version for downloading application.

## *Inspiration*

I decided to analyze this dataset because I usually download applications from Google Play Store for my Samsung phone and it is interesting for me to discover which applications have the top rating, and which factors influence this.

Additionally, 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.

## II. Application description

The application and all charts used in this project were created using the R programming language with Shiny package. The application I created consists of 1 side bar and 4 main tabs which are “About”, “Google play”, “Kaggle dataset” and “Data visualization”.


The side bar is on the left side and contains logo, key word and basic information about Google Play.

### *Tab 1: ABOUT*

In this tab, there are 2 main parts including information about App developer and Web application, which were highlighted in red.

I also included my photo under App developer section and 2 hyperlinks which directly go to my LinkedIn page and to the Kaggle page where I downloaded the dataset for this analysis.

### Google Play Store User Review



Google Play

Play Your Heart Out

Google Play is a digital distribution service operated and developed by Google LLC.

It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android software development kit.


ABOUT

GOOGLE PLAY

KAGGLE DATASET

DATA VISUALIZATION

### App developer



**Ngoc B. Dao** is a Vietnamese student who are pursuing her Master's program in Big Data at **SGH Warsaw School of Economics**. She had a background in Finance and Accounting with 2 years experience working in a bank.

For more information, visit her [LinkedIn](#).

### Web App

This app is created as part of final project for subject Querying, data presentation, data visualisation lead by **Dr. Jaroslaw OLEJNICZAK**.


This is the first application created by Ngoc using programing language R.

The dataset was taken from the **Kaggle website** which can be downloaded [HERE](#).

## Tab 2: GOOGLE PLAY

In the second tab, I decided to enclose a short video about Google Play Store with the length of 25 seconds to make the application more interesting.

### Google Play Store User Review



Google Play

Play Your Heart Out

Google Play is a digital distribution service operated and developed by Google LLC. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android software development kit.


Google Play also serves as a digital media store, offering music, books, movies, and television programs.

ABOUT

GOOGLE PLAY

KAGGLE DATASET

DATA VISUALIZATION



Cloudy With A Chance Of Meatballs 2 available now on Google Play  
© Sony Pictures Rated PG

▶ 00:07

00:25

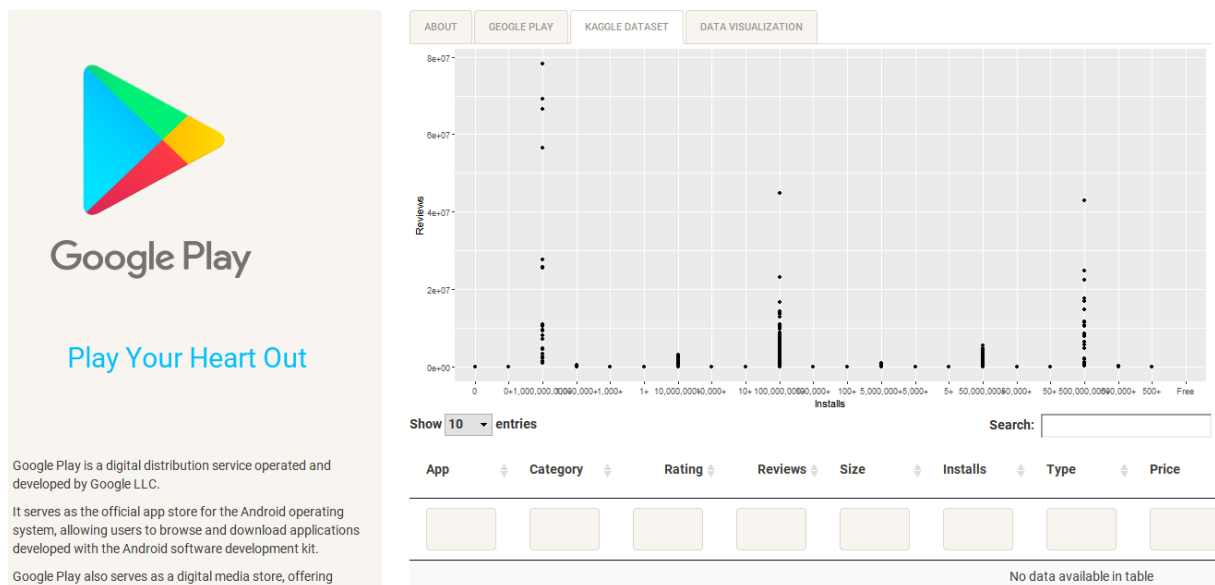
🔊

🖥️

🔄

## Tab 3: KAGGLE DATASET

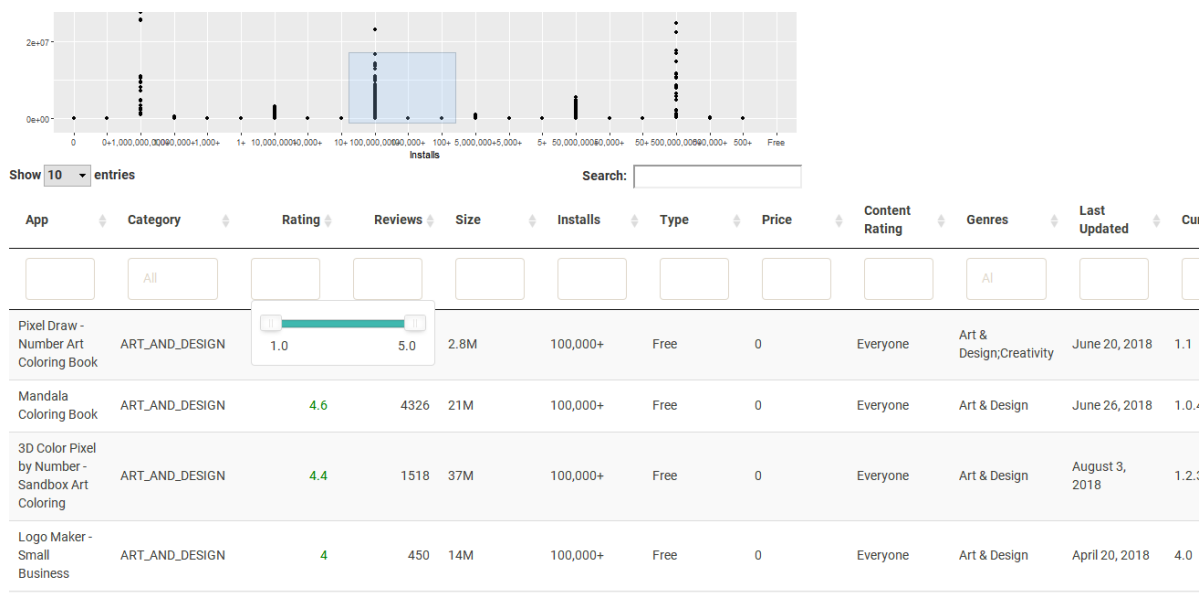
The 3<sup>rd</sup> tab in my application consists of a plot and a dataset below with the help of user\_brush function.



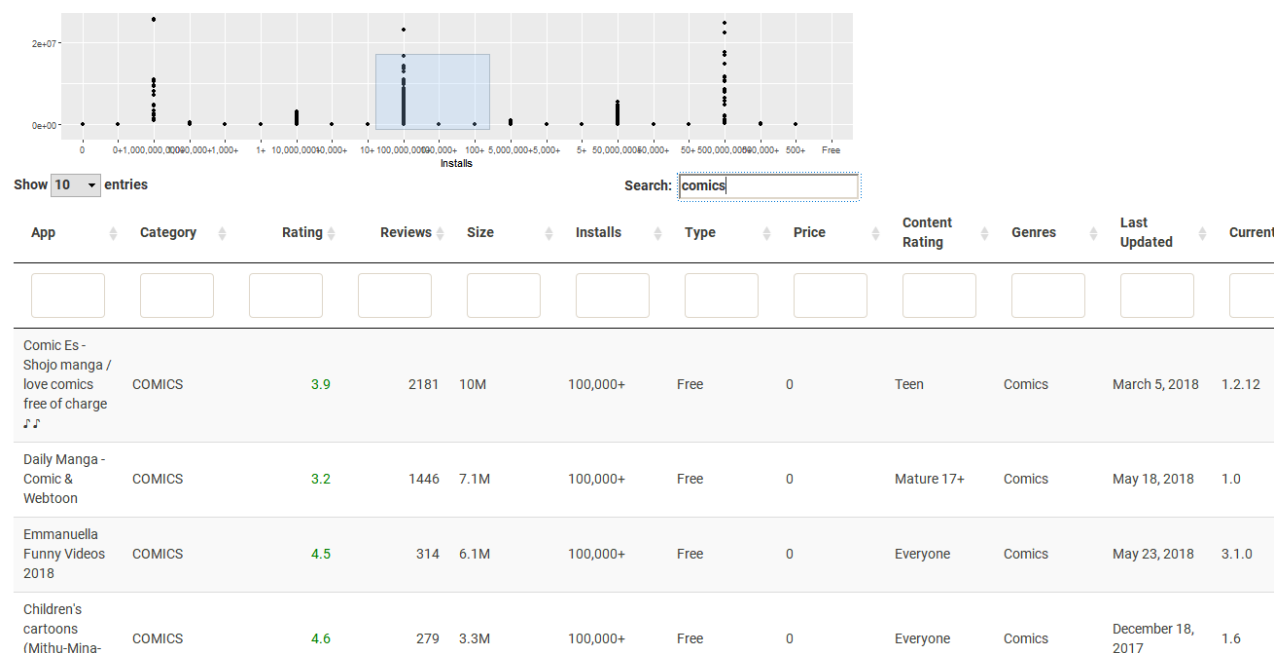
The plot helps the user to see the correlation between the number of installations and the number of reviews. It is not surprising that the higher number of installations results in the higher number of reviews.

When the user click on a particular area on the plot, the table below will show all the variables inside the square. On the top left of table, the user can choose how many entries per page they would like to see. Beside each variables in the first row, there is also filter function.

Especially the rating column was highlighted in green to make the information more visible for the user.



On the top right of the table, we have a search box, where you can type a key word and the table will only shows variables that match your search key words.



On the bottom left of the page, we have a black button named “Download Table”.

Cartoonish cartoons (Mithu-Mina- Raju)	COMICS	4.6	279	3.3M	100,000+	Free	0	Everyone	Comics	December 18, 2017	1.6
Faustop Sounds	COMICS	4.7	1677	40M	100,000+	Free	0	Everyone	Comics	May 26, 2018	1.7.7
Laftel - Watching and Announcing Snooping, Streaming	COMICS	4.6	4895	35M	100,000+	Free	0	Everyone	Comics	August 2, 2018	2.4.25
Q Avatar (Avatar Maker)	COMICS	4.4	2012	3.6M	100,000+	Free	0	Everyone	Comics	March 4, 2017	1.3.5
AF Comics Reader - Free	COMICS	2.8	5	4.1M	100+	Free	0	Everyone	Comics	November 10, 2017	1.0
Comics Reader	COMICS	3.9	4496	609k	100,000+	Free	0	Everyone	Comics	October 29, 2017	2.3.216
Daily Manga - Comic & Webtoon	COMICS	3.2	1462	7.1M	100,000+	Free	0	Mature 17+	Comics	May 18, 2018	1.0

Showing 1 to 10 of 12 entries (filtered from 2,289 total entries)

Previous 1 2 Next

DOWNLOAD TABLE

When clicking on this button, the dataset will be automatically downloaded and then the user will have options to save or open the file.

Google Play also serves as a digital media store, offering music, books, movies, and television programs.

Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design/Creativity
Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design
3D Color Pixel by Number - Sandbox Art Coloring	ART_AND_DESIGN	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design
Logo Maker - Small Business	ART_AND_DESIGN	4	450	14M	100,000+	Free	0	Everyone	Art & Design
Boys Photo Editor - Six Pack & Men's Suit	ART_AND_DESIGN	4.1	654	12M	100,000+	Free	0	Everyone	Art & Design
Mcqueen Coloring pages	ART_AND_DESIGN		61	7.0M	100,000+	Free	0	Everyone	Art & Design/Action & Adventure
Animated Photo Editor	ART_AND_DESIGN	4.1	203	6.1M	100,000+	Free	0	Everyone	Art & Design
Easy Realistic Drawing Tutorial	ART_AND_DESIGN	4.1	223	4.2M	100,000+	Free	0	Everyone	Art & Design
Pink Silver Bow Keyboard Theme	ART_AND_DESIGN	4.2	1120	9.2M	100,000+	Free	0	Everyone	Art & Design
Anime Manga Coloring Book	ART_AND_DESIGN	4.5	5035	11M	100,000+	Free	0	Everyone	Art & Design

Showing 1 to 10 of 1,569 entries

Previous 1 2 3 4 5 ... 157 Next

What do you want to do with plotextract.csv (237 KB)?  
From: 127.0.0.1

Open Save ^ Cancel X

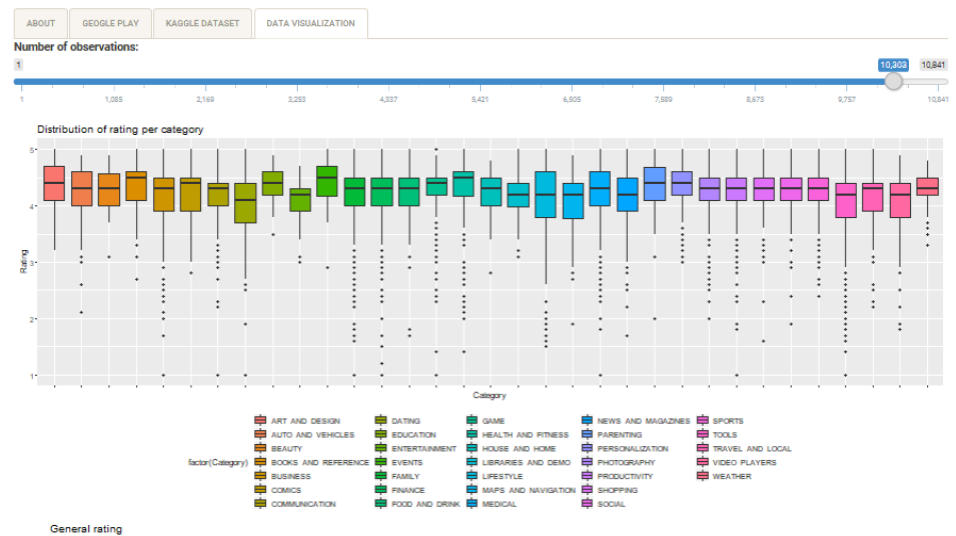
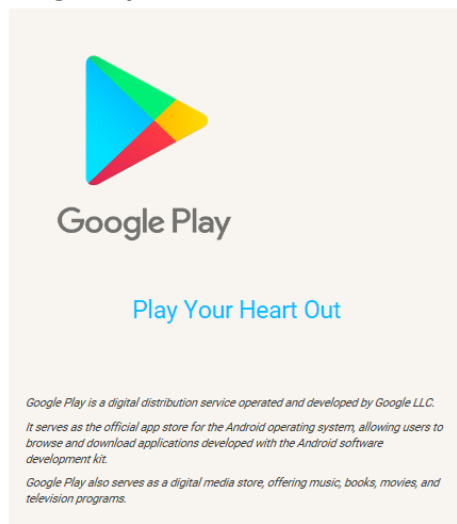
#### Tab 4: DATA VISUALIZATION

In the last tab about data visualization, I inserted here 1 slider and 2 plots visualizing the dataset.

The slider has the value ranging from 1 to 10,841 and has the initial value of 5000. This tool gives the user choice to choose the number of observations you would like to see, and the below 2 plots will change according to your choice.

The first plot is a box plot, which was built with help of ggplot2 package, which shows the distribution of application rating for all 33 categories. It is not hard to see that the highest ratings belong to applications with events and parenting categories and Dating is the category that received the lowest rates.

## Google Play Store User Review



In the second plot, it shows a grid with bar charts for 2 type of applications that are free and paid. But it is clearly that most of applications in Google Play Store got high rating in range from 4.2 to 5.0 disregard of how expensive they are.

developed with the Android software development kit.

Google Play also serves as a digital media store, offering music, books, movies, and television programs.



### **III. Conclusion**

With the fact that Android mobile operating system has been the best-selling OS worldwide on smartphones and on tablets, Google Play Store will still be the largest application store for smartphones. This dataset only features some main characteristics of the total over 3.5 million Android applications (as of 2017). However, we can draw some interesting conclusions about the Google Play Store.

It is obvious that the application store is the key factors that influence the smartphone buyers' decisions. Therefore, many phone brands chose Android as the operating system for their products. The Android applications on Google Play Store has been widely used by the Android users due to the high quality and the variety. Most of applications on Google Play Store were highly evaluated by the users. And even though applications are available through Google Play either free of charge or at a cost, the majority of them are still free. That enables users to wide range of choice and enrich their experience.