GOOGLE PLAY STORE APP ANALYSIS

(BDA Project)

TEAM-

Riyanchhi Agrawal (1BM17CS076)
Shehyaaz Khan Nayazi (1BM17CS094)

Outline

- Introduction
- Technologies
- Dataset
  - Attributes
- Application
  - Features
- Data Visualization and Analysis

Introduction

Analysis of data examines large and different types of data and study their growth and development.

It involves three major stages-

1. Pre-processing the data
2. Performing Analytics over data
3. Visualizing the data.

Here we analyze the data on apps present in google playstore and based on certain logic we process the data from above mentioned stages.
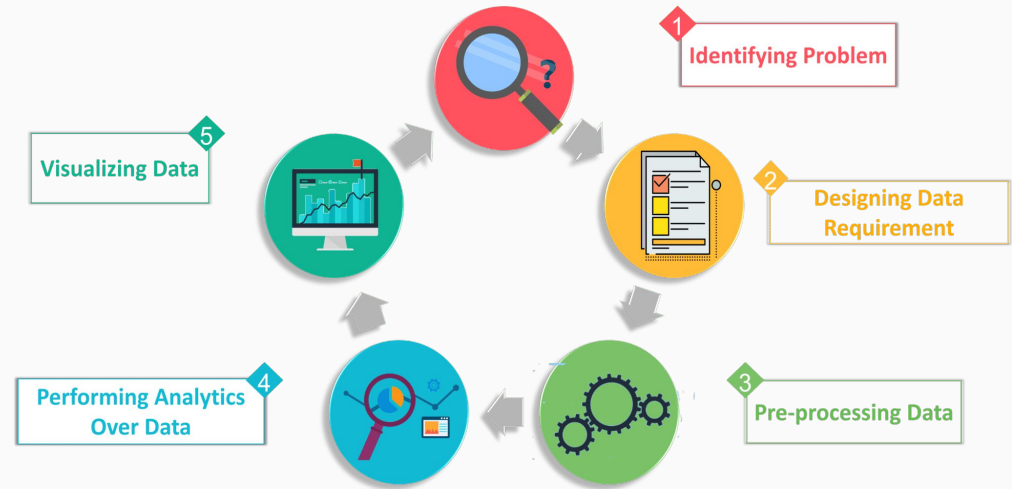
Technologies

Libraries used-

- Pymongo (MongoDB driver for Python)
- PyQt5
- Pandas
- Numpy
- Matplotlib
- MongoDB (Backend NoSQL Database)

Dataset

Description & Reference

An structured data set which is used to perform various analysis and operation on 4470 records with 13 attributes.

Source -

Data Stored

Attributes

The dataset contains following columns -

1. **App** - Application name on Google Playstore
2. **Category** - General category to which the app belongs.
3. **Rating** - App rating ranges between 0 to 5.
4. **Reviews** - No. of user reviews.
5. **Size** - Size of the app in MB..
6. **Installs** - No.of Installs.
7. **Type** - App is capitalized or not.
8. **Price** - Based on Type charging amount of the particular app. (in $)
9. **Content Rating** - Based on user age period access given by developer.
10. **Genres** - A more detailed category to which the application belongs
11. **Last Updated** - Date on which the app was last updated
12. **Current Ver** - Current version of the app
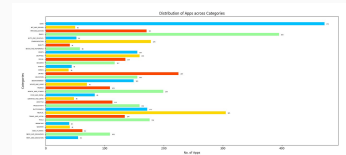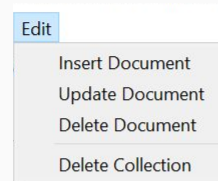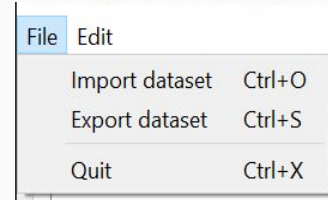13. **Android Ver** - Android version supported by the app

Application
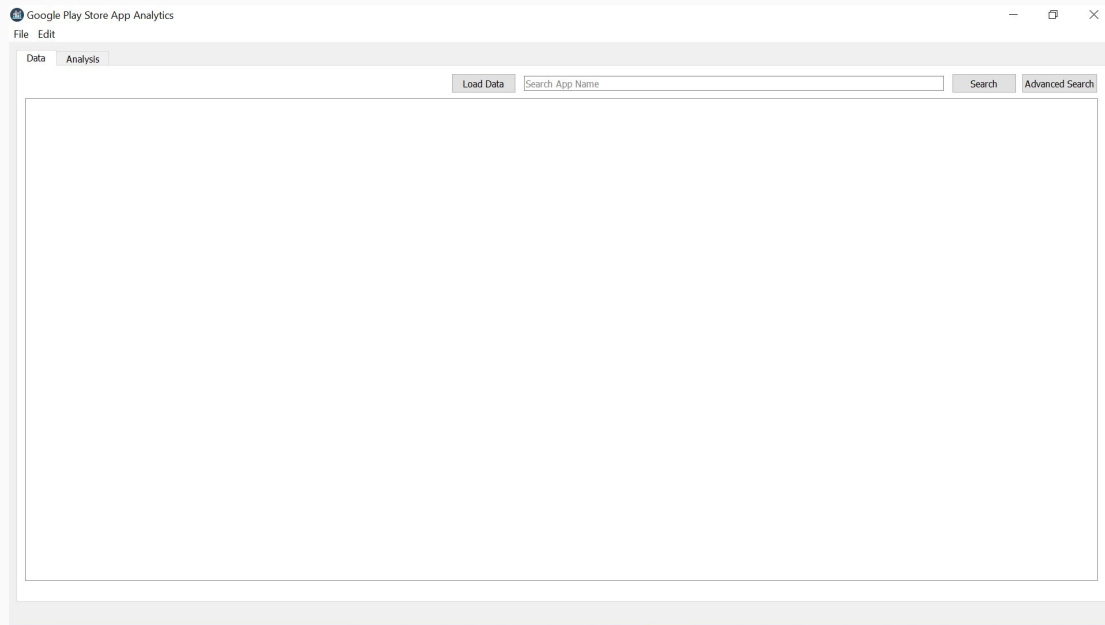
With the help of QT Designer front-end of Application is built to give better user experience on data manipulation & analysis.

Features

- Import / Export Dataset
- Insert Document
- Delete Document(s)
- Update Document(s)
- Delete Collection
- Perform Analysis (using graphs)
- Recommendations (using Demographic Filtering)

File	Edit	
	Import dataset	Ctrl+O
	Export dataset	Ctrl+S
	Quit	Ctrl+X

Edit
Insert Document
Update Document
Delete Document
Delete Collection

Loading Data

Data is loaded from mongoDB in the table format.

Insert Document

On Selecting "Insert Document", a form will open to take values as an input for attributes and create a record on submitting.

Update/Delete Document

Data Visualization Analysis
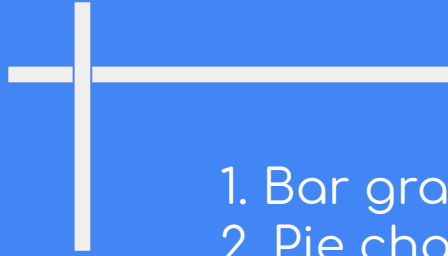
1. Bar graph of distribution of apps across Category.
2. Pie chart of free and paid apps.
3. Pie chart of Content rating in each category.
4. Bar graph of android version.
5. Histogram of Rating.
6. Boxplot of Installs.
7. Ratings vs Size.

Distribution of Apps across Category

-Bar Graph
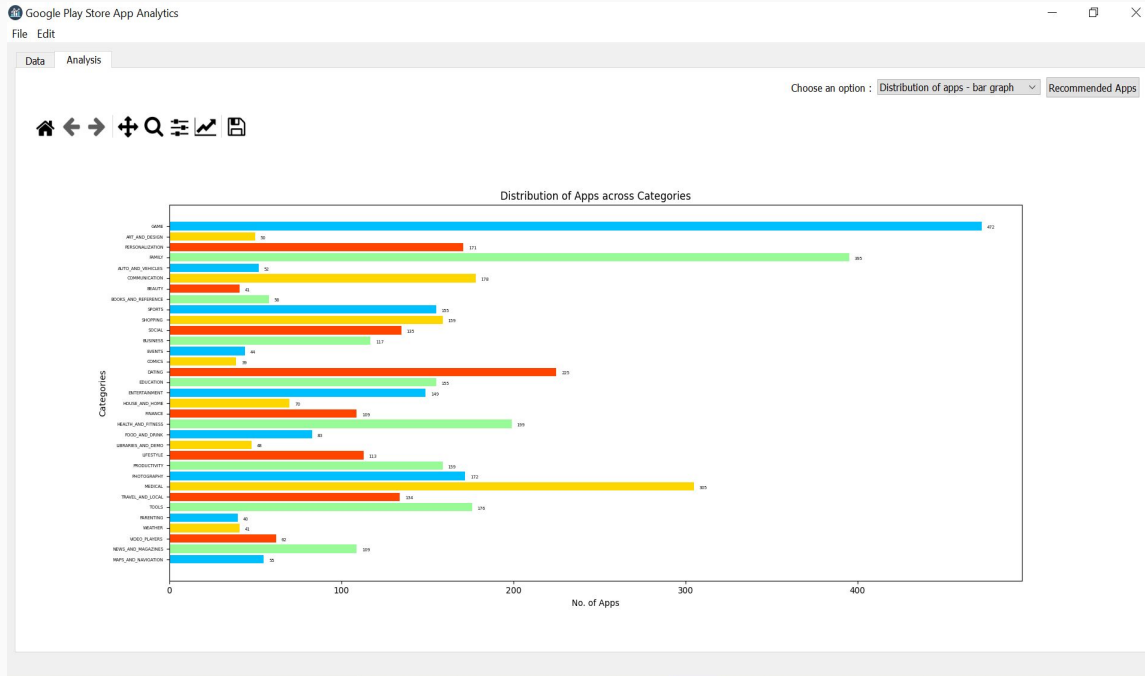
The graph is created with help of pandas and matplotlib library from mongoDB Query and from the graph its shows that highest number of Apps falls in top 3 category - Game, Family, Medical.

Distribution of Apps based on Type

-Pie Chart
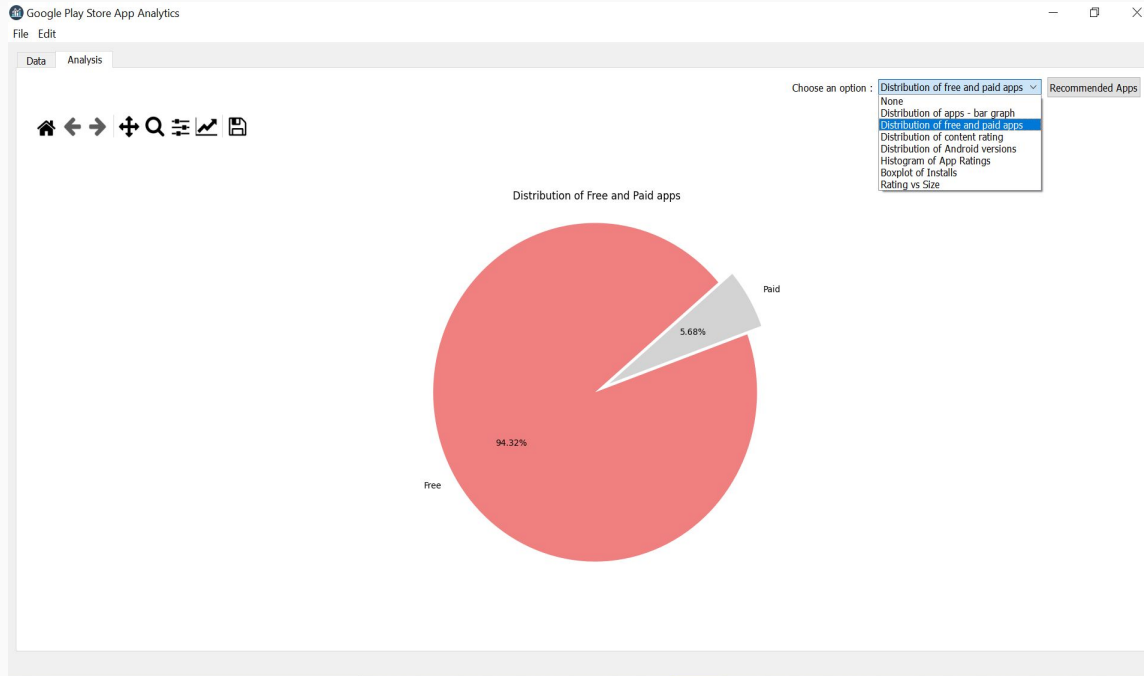
Pie-chart is created with help of pandas and matplotlib library from mongoDB Query and from the graph its shows that largest number of Apps are Free (>94%) and exploding portion consists of paid Apps(<6%).

Distribution of content Rating of Apps

-Pie Chart
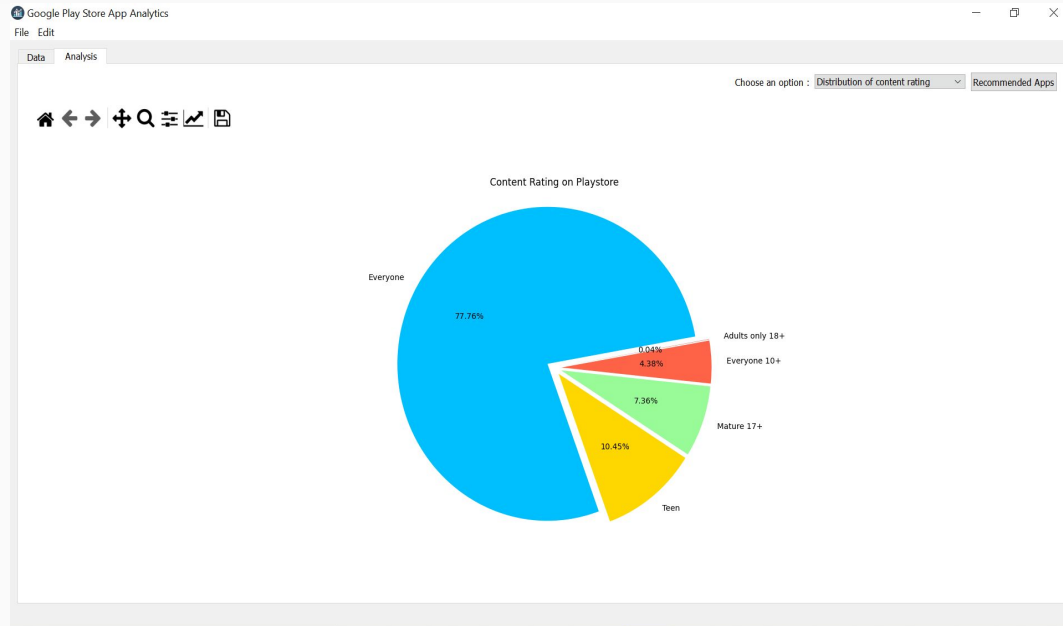
Pie-chart is created with help of pandas and matplotlib library from mongoDB Query and from the graph its shows that largest number of Apps have "Everyone"(including all age group) as content rating (77.76%) and exploding portion have others.

Distribution of Android Version of Apps

-Bar Graph

From the graph its shows that highest number of Apps contains the version 4.1 or above (>900) or have variable version based on device and very few have 2.1 or above (~1).
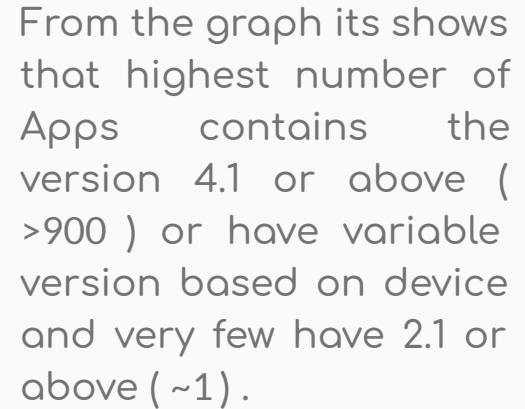
Graph Representation based on App Rating

-Histogram
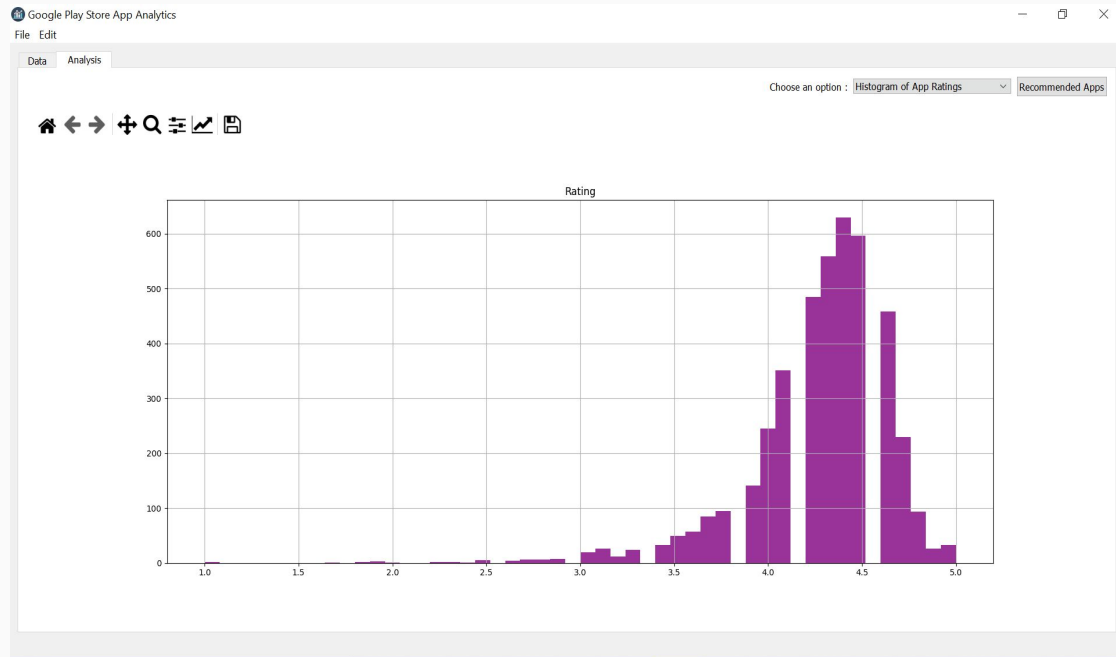
Histogram is a visual representation of the distribution of numerical data in range.

It is created with help of pandas and matplotlib library where data extracted from mongoDB Query .The graph shows that highest number of rated Apps falls in the range of **4.0 - 4.5.**

Distribution of Installs group by App Type

-Box Plot

Box plot is a visual representation of the distribution of numerical data of number of Installs by Type of App in range. It is created by extracting data from mongoDB Query .The graph shows two batches of two types paid and free.

Summary	Free	Paid
Median	6.0	3.7
IQR	(5.7 - 7.0)	(2.7 - 4.0)

Graph Representation of Rating vs Size

-Scatterplot

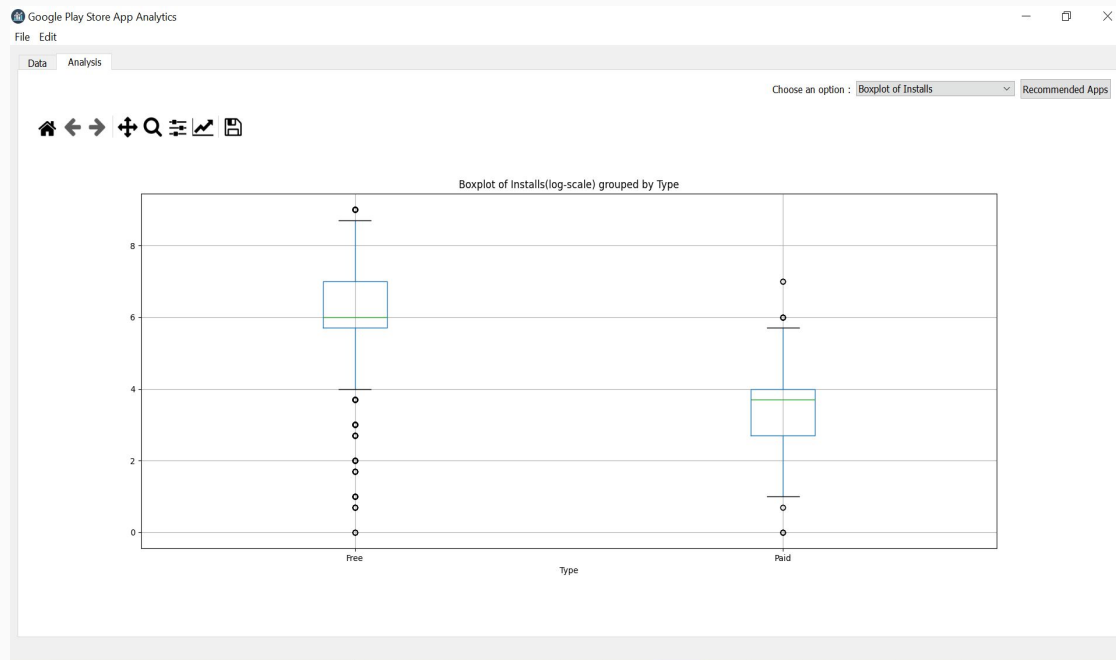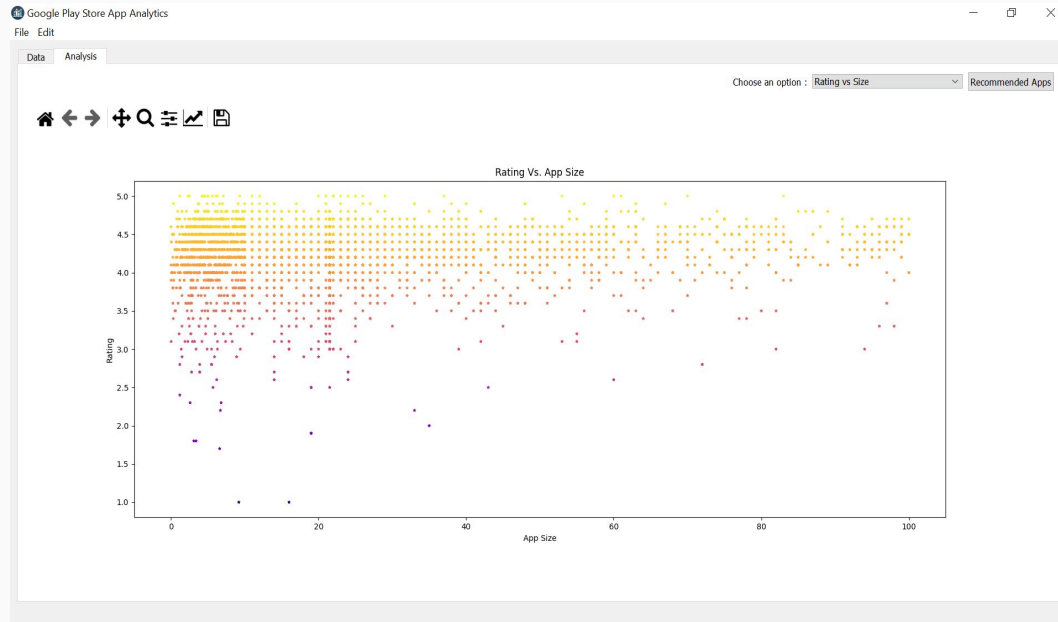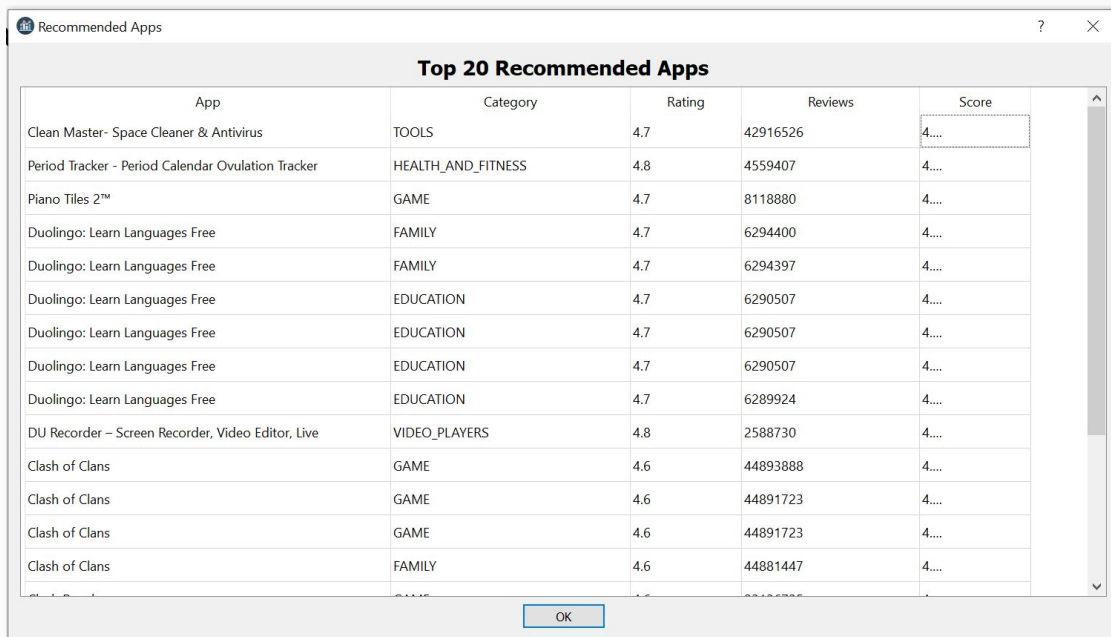Scatterplot is a visual representation of the distribution of numerical data. It is created with help of pandas and matplotlib library where data extracted from mongoDB Query.

The pattern shows that the lesser the size of the app gives more user satisfaction resulting highest rating (~ 5.0).

Data Recommendation based on Analysis

Recommended Apps

Top 20 Recommended Apps

App	Category	Rating	Reviews	Score
Clean Master- Space Cleaner & Antivirus	TOOLS	4.7	42916526	4....
Period Tracker - Period Calendar Ovulation Tracker	HEALTH_AND_FITNESS	4.8	4559407	4....
Piano Tiles 2™	GAME	4.7	8118880	4....
Duolingo: Learn Languages Free	FAMILY	4.7	6294400	4....
Duolingo: Learn Languages Free	FAMILY	4.7	6294397	4....
Duolingo: Learn Languages Free	EDUCATION	4.7	6290507	4....
Duolingo: Learn Languages Free	EDUCATION	4.7	6290507	4....
Duolingo: Learn Languages Free	EDUCATION	4.7	6290507	4....
Duolingo: Learn Languages Free	EDUCATION	4.7	6289924	4....
DU Recorder – Screen Recorder, Video Editor, Live	VIDEO_PLAYERS	4.8	2588730	4....
Clash of Clans	GAME	4.6	44893888	4....
Clash of Clans	GAME	4.6	44891723	4....
Clash of Clans	GAME	4.6	44891723	4....
Clash of Clans	FAMILY	4.6	44881447	4....

OK

Demographic filtering is used to generate top 20 recommended Apps. Based on Rating and Reviews, a score is generated and using the score we can rank the records.

" Clean Master- Space Cleaner & Antivirus " is the most recommended app with rating 4.7 and highest number of reviews 4,29,16,526.

End of Presentation

THANK-YOU

https://github.com/Shehyaaz/BDA-Project