

GOOGLE PLAY STORE APP ANALYSIS

(BDA Project)



TEAM-

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

Outline

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

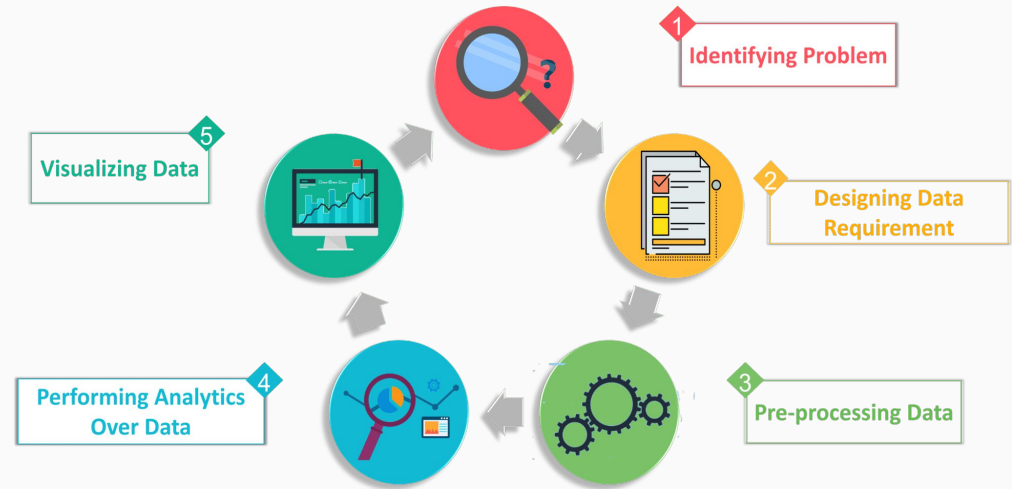
Introduction

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

googleplaystore

File Edit View Insert Format Data Tools Add-ons Help

Last edit was seconds ago

100%

\$

%

123

Arial

10

B

I

T

S

A

<

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

Source - <https://www.kaggle.com/lava18/google-play-store-apps/download>

Data Stored

The screenshot shows the MongoDB Compass web interface. The left sidebar displays the database structure: 8 DBS, 20 COLLECTIONS, and a list of collections including 'google' and 'mydb.google_playstore_apps'. The main panel shows the 'mydb.google_playstore_apps' collection with 4.5k documents, 1.4MB total size, and 319B average size. The 'Documents' tab is active, showing a list of documents. The first document is for 'Photo Editor & Candy Camera & Grid & ScrapBook' and the second is for 'U Launcher Lite - FREE Live Cool Themes, Hide Apps'.

MongoDB Compass - localhost:27017/mydb.google_playstore_apps

Connect View Collection Help

Local

8 DBS 20 COLLECTIONS

☆ FAVORITE

HOST
localhost:27017

CLUSTER
Standalone

EDITION
MongoDB 4.0.20-36-gf8006ee Community

Q google

mydb

google_playstore_apps

mydb.google_playstore_apps Documents

Documents 4.5k TOTAL SIZE 1.4MB AVG. SIZE 319B INDEXES 1 TOTAL SIZE 48.0KB AVG. SIZE 48.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER OPTIONS FIND RESET

ADD DATA VIEW

Displaying documents 1 - 20 of 4470 REFRESH

```
{
  "_id": ObjectId("5fb4ddd0be9292e6b1a73471"),
  "App": "Photo Editor & Candy Camera & Grid & ScrapBook",
  "Category": "ART_AND_DESIGN",
  "Rating": 4.1,
  "Reviews": 159,
  "Size": "19M",
  "Installs": "10,000+",
  "Type": "Free",
  "Price": 0,
  "Content Rating": "Everyone",
  "Genres": "Art & Design",
  "Last Updated": "January 7, 2018",
  "Current Ver": "1.0.0",
  "Android Ver": "4.0.3 and up"
}
```

```
{
  "_id": ObjectId("5fb4ddd0be9292e6b1a73472"),
  "App": "U Launcher Lite - FREE Live Cool Themes, Hide Apps",
  "Category": "ART_AND_DESIGN",
  "Rating": 4.7,
  "Reviews": 87510,
  "Size": "8.7M",
  "Installs": "5,000,000+",
  "Type": "Free",
  "Price": 0,
  "Content Rating": "Everyone"
}
```

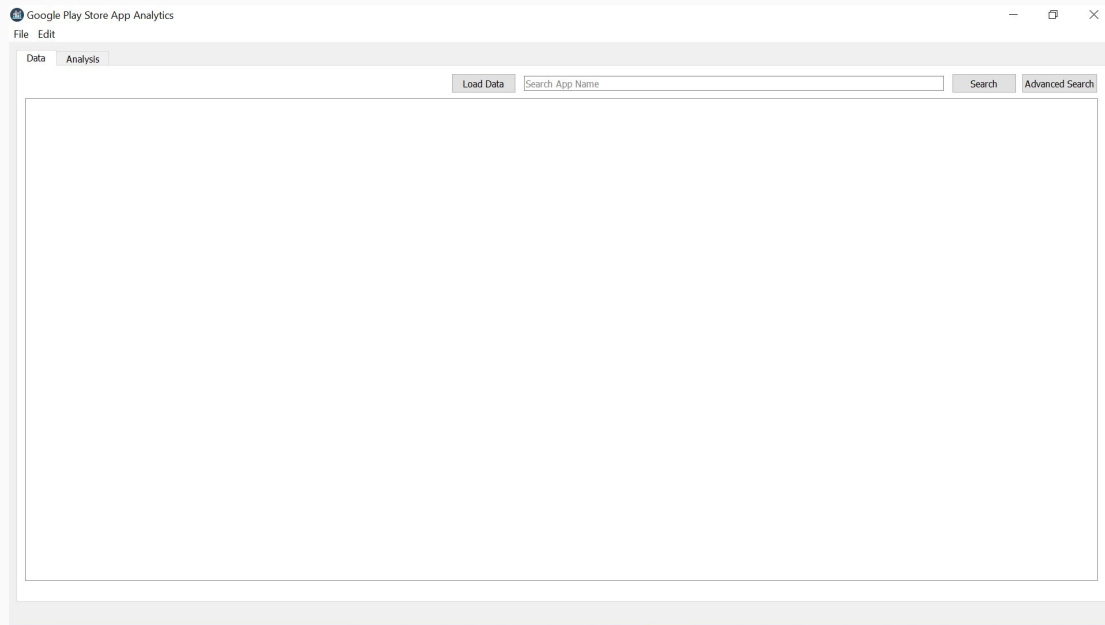
> _MongoSH Beta

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

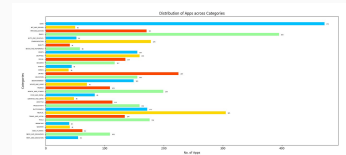
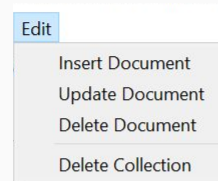
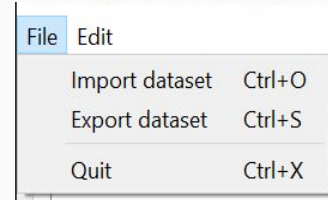
Application



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)



Loading Data

Google Play Store App Analytics

File Edit

Data		Analysis											
Load Data													
Search App Name													
Search													
Advanced Search													
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver		
Photo Editor & ...	ART_AND_DESL...	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0	
U Launcher Lite ...	ART_AND_DESL...	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0	
Sketch - Draw &...	ART_AND_DESL...	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2	
Paper flowers ...	ART_AND_DESL...	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1	2.3	
Smoke Effect ...	ART_AND_DESL...	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0	
Infinite Painter	ART_AND_DESL...	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1	4.2	
Garden Colorin...	ART_AND_DESL...	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	September 20, ...	2.9.2	3.0	
Kids Paint Free ...	ART_AND_DESL...	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design	July 3, 2018	2.8	4.0	
Name Art Photo...	ART_AND_DESL...	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	July 31, 2018	1.0.15	4.0	
Tattoo Name O...	ART_AND_DESL...	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	April 2, 2018	3.8	4.1	
Text on Photo - ...	ART_AND_DESL...	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	October 27, 2017	1.0.4	4.1	
Mandala Colori...	ART_AND_DESL...	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	June 26, 2018	1.0.4	4.4	
3D Color Pixel b...	ART_AND_DESL...	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	August 3, 2018	1.2.3	2.3	
Learn To Draw ...	ART_AND_DESL...	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	June 6, 2018	nan	4.2	
Photo Designer ...	ART_AND_DESL...	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	July 31, 2018	3.1	4.1	
350 Diy Room ...	ART_AND_DESL...	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	November 7, ...	1	2.3	
FlipaClip - ...	ART_AND_DESL...	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	August 3, 2018	2.2.5	4.0	

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.

File Edit

Data Analysis

Load Data BMSlib Search Advanced Search

App	Category	Rating	Reviews	Size	Installs	Type	Price
BMSlib	EDUCATION	4.9	34500	100M	50,000+	Free	0.0

Insert Document

Application information

App Name

BMSlib

Category

Education

Genre

Books

Rating

4.9

Reviews

34500

Size in MB

100

Installs

50000+

Type

Free

Price in \$

0 if free, else price...

Content Rating

Everyone

Last Updated

5/13/2000

Current Version

1.0

Android version

May 2000

Save

Cancel

Jun

Mon

Tue

Wed

Thu

Fri

Sat

30

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

1

2

3

4

5

6

7

8

9

10

Update/Delete Document

Update Document ? X

Search

App Name X

Category

Update Options

Options :

Update information

App Name	<input type="text"/>		
Category	<input type="text" value="eg: Education, P..."/>	Genre	<input type="text" value="eg : Arcade, Creativity"/>
Rating	<input type="text" value="0.0"/>	Reviews	<input type="text" value="5000"/> X
Size in MB	<input type="text" value="in MB"/>	Installs	<input type="text" value="eg: 100+, 500+"/>
Type	<input type="button" value="Free"/>	Price in \$	<input type="text" value="0 is free, else price in \$"/>
Content Rating	<input type="button" value="Everyone"/>	Last Updated	<input type="text" value="1/1/2000"/>
Current Version	<input type="text"/>	Android version	<input type="text" value="eg: 1.0.0 and up"/>

Delete Document ? X

Search

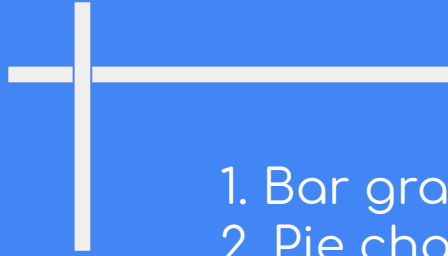
App Name

Category

Delete Options

Options :

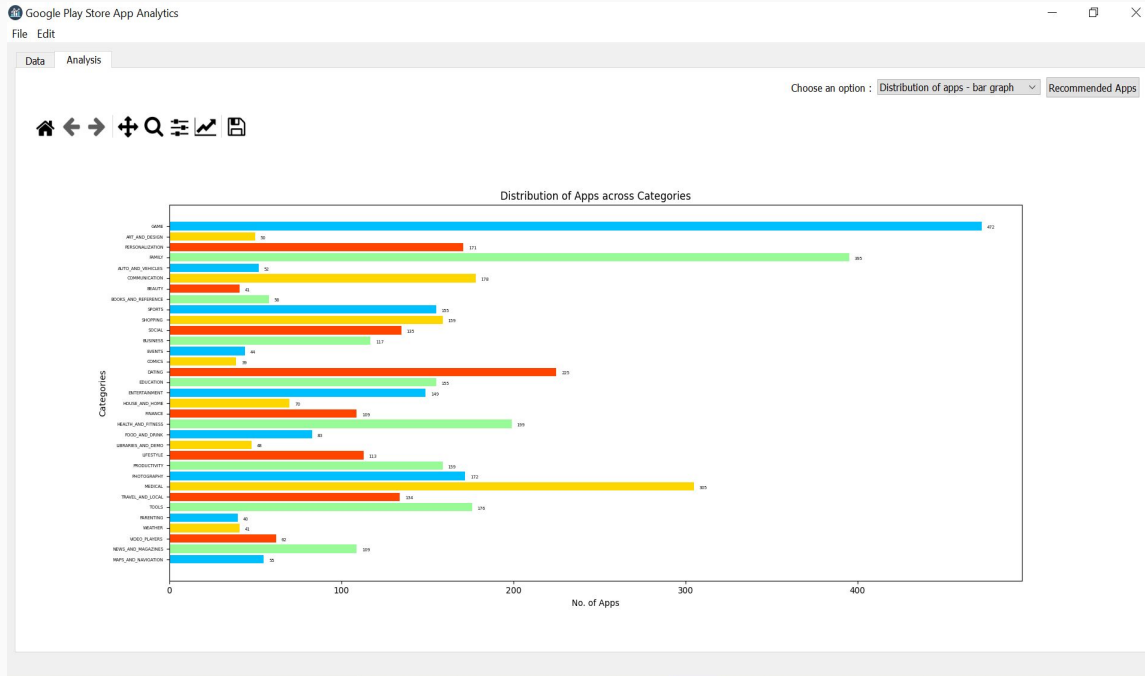
Data Visualization Analysis



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

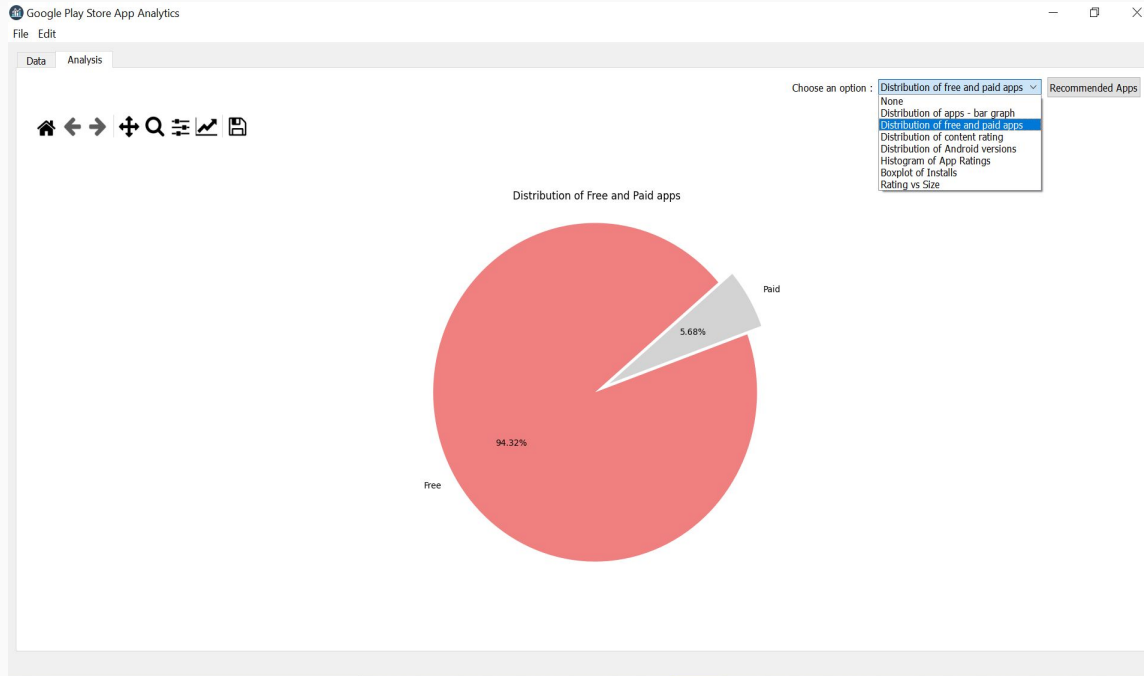
-Bar Graph



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

Distribution of Apps based on Type

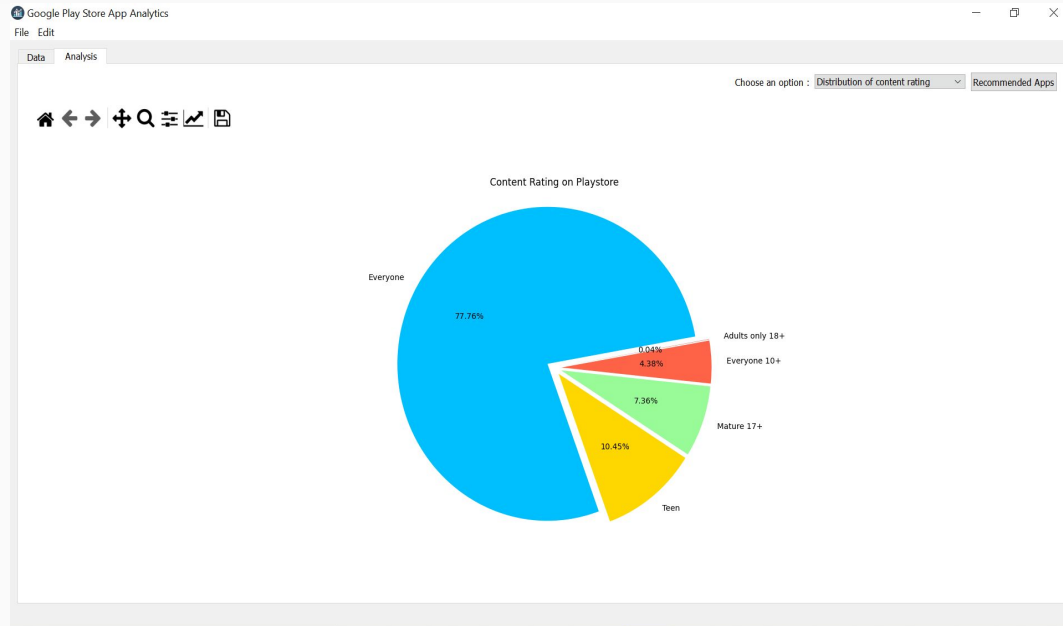
-Pie Chart



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%).

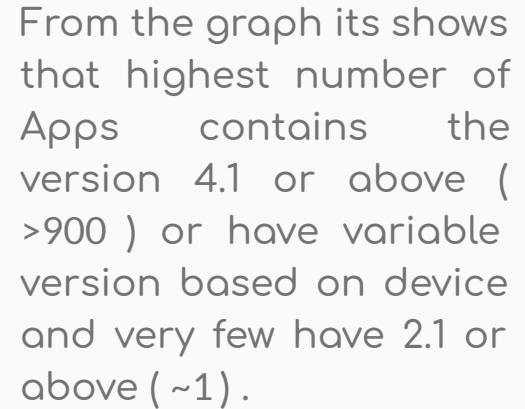
Distribution of content Rating of Apps

-Pie Chart



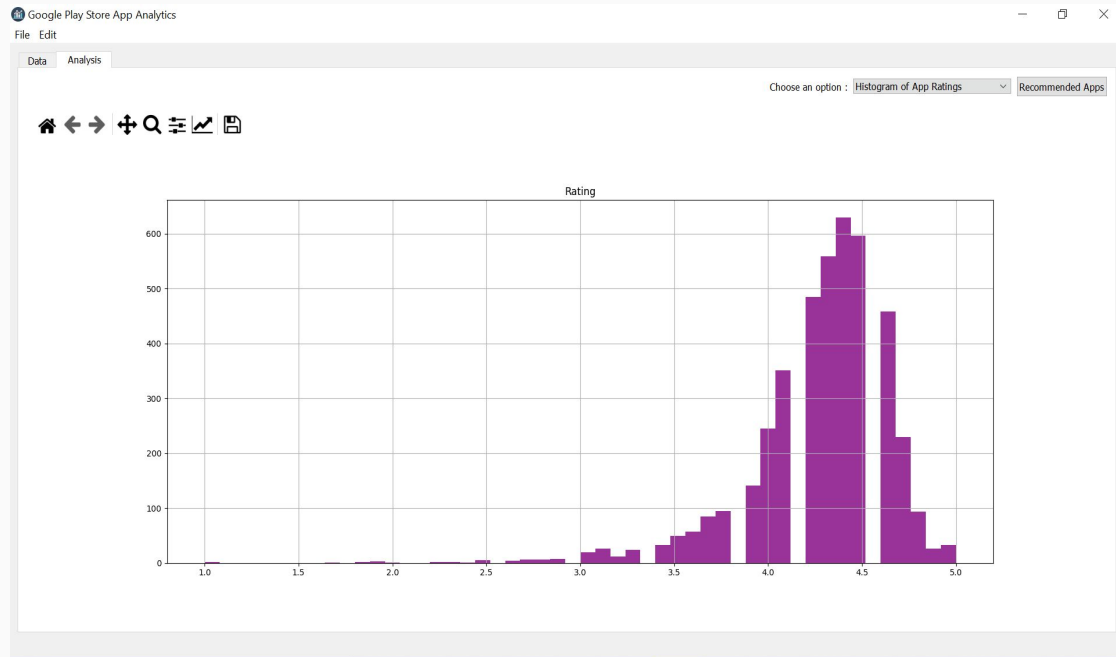
Pie-chart is created with help of pandas and matplotlib library from mongoDB Query and from the graph its shows that largest number of Apps have “Everyone”(including all age group) as content rating (77.76%) and exploding portion have others.

-Bar Graph



Graph Representation based on App Rating

-Histogram

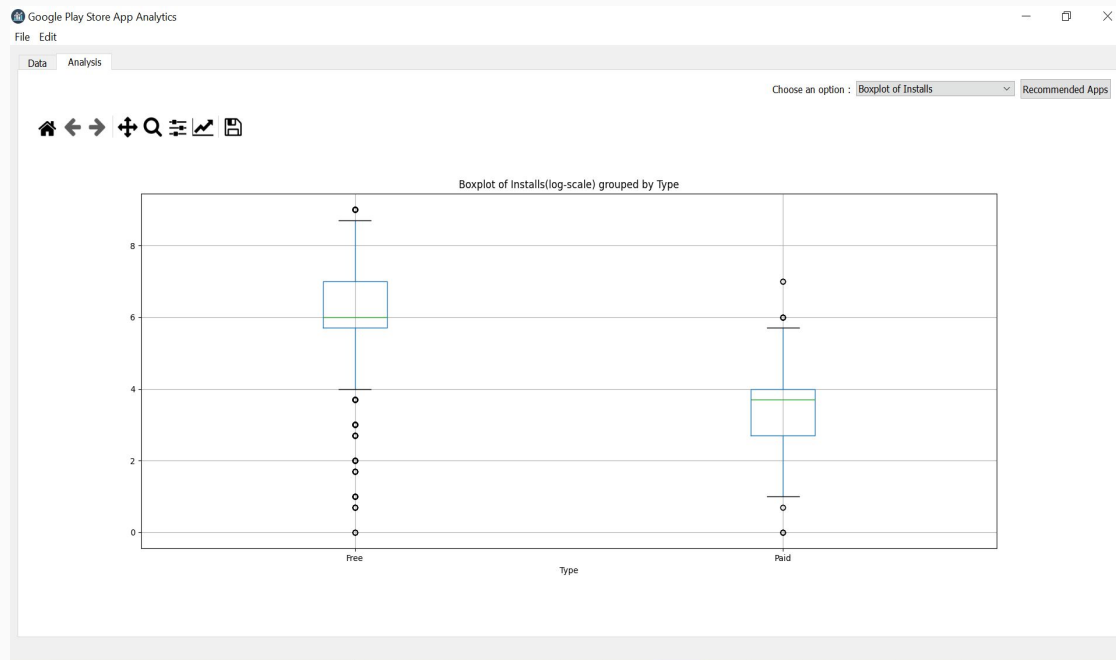


Histogram is a visual representation of the distribution of numerical data in range.

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

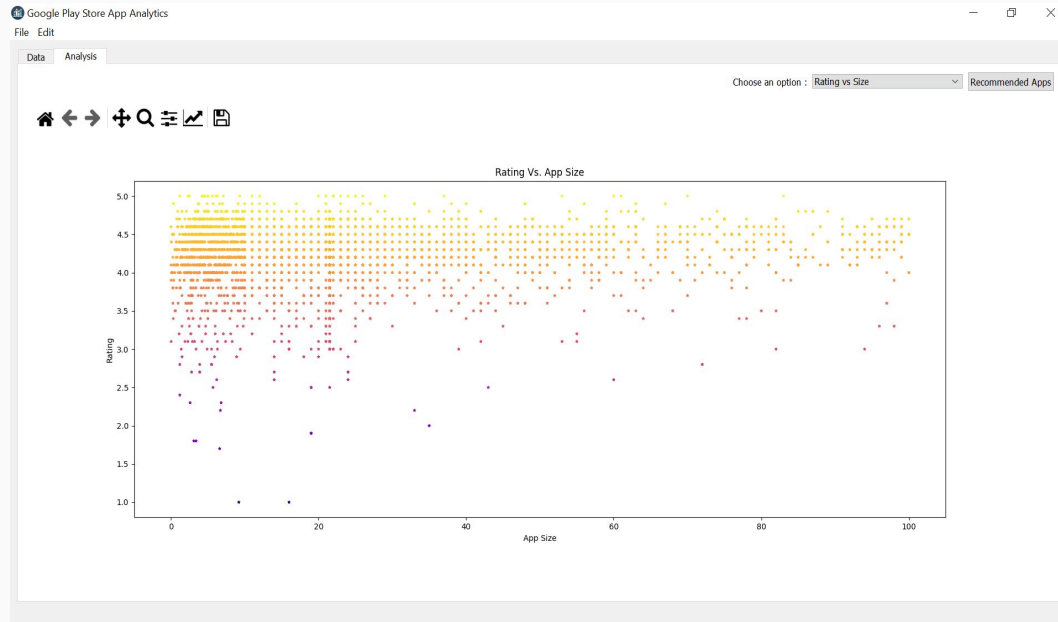


Box plot is a visual representation of the distribution of numerical data of number of Installs by Type of App in range. It is created by extracting data from mongoDB Query .The graph shows two batches of two types paid and free.

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

Graph Representation of Rating vs Size

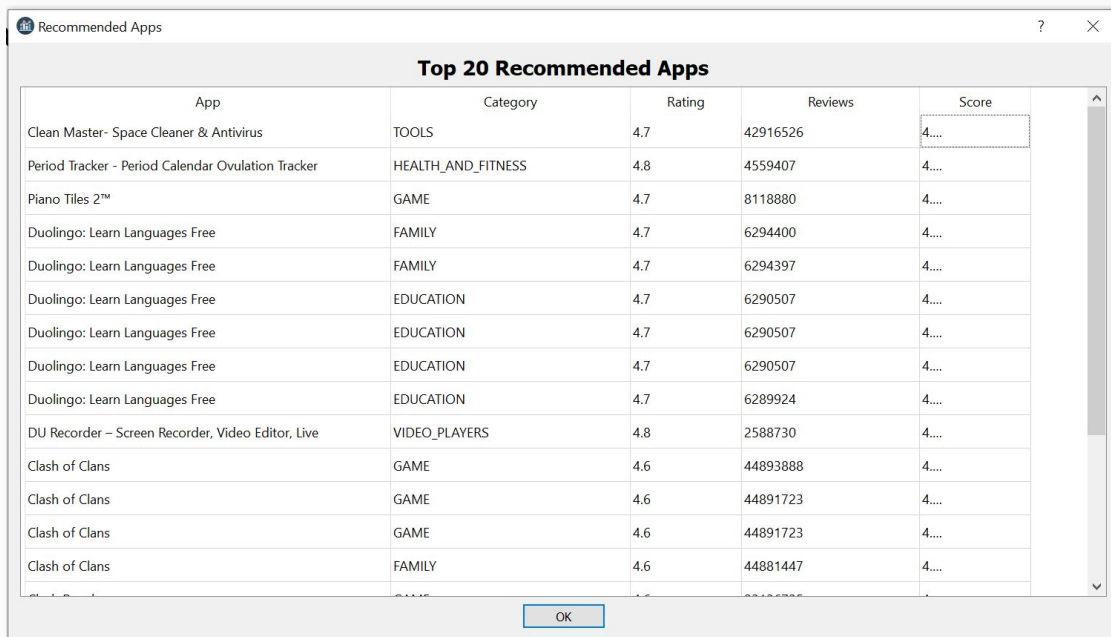
-Scatterplot



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



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....

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>