INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2018)

Analysis on Google Playstore Apps

Submitted by

Name: Poka Venkata Anil Kumar

Registration No: 11701346

Roll no: A03

Section: KMO64

Course Code: INT 217 (Introduction to Data Management)

Under the Guidance of

Maneet Kaur Ma'am

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Poka Venkata Anil Kumar** bearing Registration no.**1171346** has completed **INT 217** project titled**, 'Analysis on Google Playstore Apps'** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science & Technology
Lovely Professional University
Phagwara, Punjab.

Date:

DECLARATION

I, **Poka Venkata Anil Kumar** , student of **B.tech** under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date : Signature :

Registration No. 11701346 Name : P.V Anil Kumar

ACKNOWLEDGEMENT

I hereby express my heartiest thanks to all the sources who have contributed to the making of this project. I oblige thanks to all those who have supported, provided their valuable guidance and helped for the accomplishment of this project. I also extend my hearty thanks to my family, friends, college teachers, and all the well-wishers.

I also would like to thanks my project guide Ms.Maneet kaur for her guidance and timely suggestion and the information provided by her on this particular topic.

It is a matter of utmost pleasure to express my indebt and deep sense of gratitude to various people who extended their maximum help to supply the necessary information for the present thesis, which became available on account of the most selfless cooperation.

Above all its sincere thanks to the **LOVELY PROFESSIONAL UNIVERSITY** for which this project is given consideration and was done with utmost seriousness.

TABLE OF CONTENT

Introduction :-

This report is about the data analysis on **Google playstore apps** and various visualizations related to this and has been elaborated. In this project , I had done many analysis on playstore apps ,Which were released in years 2010-2018.

For this project I had used Google playstore dataset, which consists of many values in it. They are App name ,App category, Rating , Size , Reviews , Installs , Android version ,Last updated, Type, Content rating and Price.

App Name:

This field contains name of the app. Ex : Facebook , Instagram etc..

App Category:

This field contains information about the app to which category it belongs to.

Ex : Family Category , games, Communication etc..

Rating :

Every product in playstore will have some rating to it, Which helps new user to know about it.

We can rate an app from 1 -5 based on its performance . Ex : 4.4, 4.5 etc..

Size:

This field contains information about size of the apps.

Reviews:

This field contains count of reviews for each Individual App.

Installs:

This field contains Count of Installs in handsets.

Android version

This field describes the android version that the app supports.

Type:

This field says about its type, like Free app or Paid app.

Content rating:

This field says about app restriction. Like for Everyone, Teen, Mature 17+.

Scope of the Analysis:-

Here in this project ,We have analyzed and discussed about the following objectives.

1. To Find Top 10 best Apps in playstore and also to find top apps according to our selection.
2. To find category wise installs.
3. To find out type of apps based on content category.
4. To find out apps that were modified in year wise
5. To find out apps that were modified month wise in a year.
6. To find out apps based on their Android version.
7. To have count on price of paid apps.
8. To find apps with more reviews.
9. To find out app with our selection along with its size.
10. To have count on installs according to category based on their rating.

Source of Dataset:-

I had Collected all resources for my project from Kaggle

KAGGLE:-

Kaggle is an online community of data scientists and machine learners, owned by Google. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and short-form AI education. On 8 March 2017, Google announced that it was acquiring Kaggle.

Kagglelink : https://www.kaggle.com/lava18/google-play-store-apps

ETL Process

Before Doing the Analysis following steps are to be followed :

1. Extract Data

Before proceeding with the analysis, the data needed by the Analyzer must be extracted from the various sources required. During extraction, the desired data is identified and extracted from many different sources, including database systems and applications. Very often, it is not possible to identify the specific subset of interest, therefore more data than necessary has to be extracted, so the identification of the relevant data will be done at a later point in time. Depending on the source system's capabilities (for example, operating system resources), some transformations may take place during this extraction process. The size of the extracted data varies depending on the source and the business situation. After data is extracted, it has to be physically transported to the target or to an intermediate system for further processing.In this report all the datasets are extracted from official website of Kaggle.com . The complete source of dataset and their links is given in Source of dataset section.

2. Transformation

Data extracted from source server is raw and not usable in its original form. Therefore it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful. In this step, you apply a set of functions on extracted data.

Validations done during this stage :

• Filtering – Select only certain columns to load
• Use lookups to merge data
• Converting Table into Pivot Table
• Summarization into pivot table
• Data Cleaning (Removing Redundancy)
• Required Fields should not be blank
• Sorting
• Conversion of Units of Measurements like Date Time Conversion, currency conversions, numerical conversions, etc.

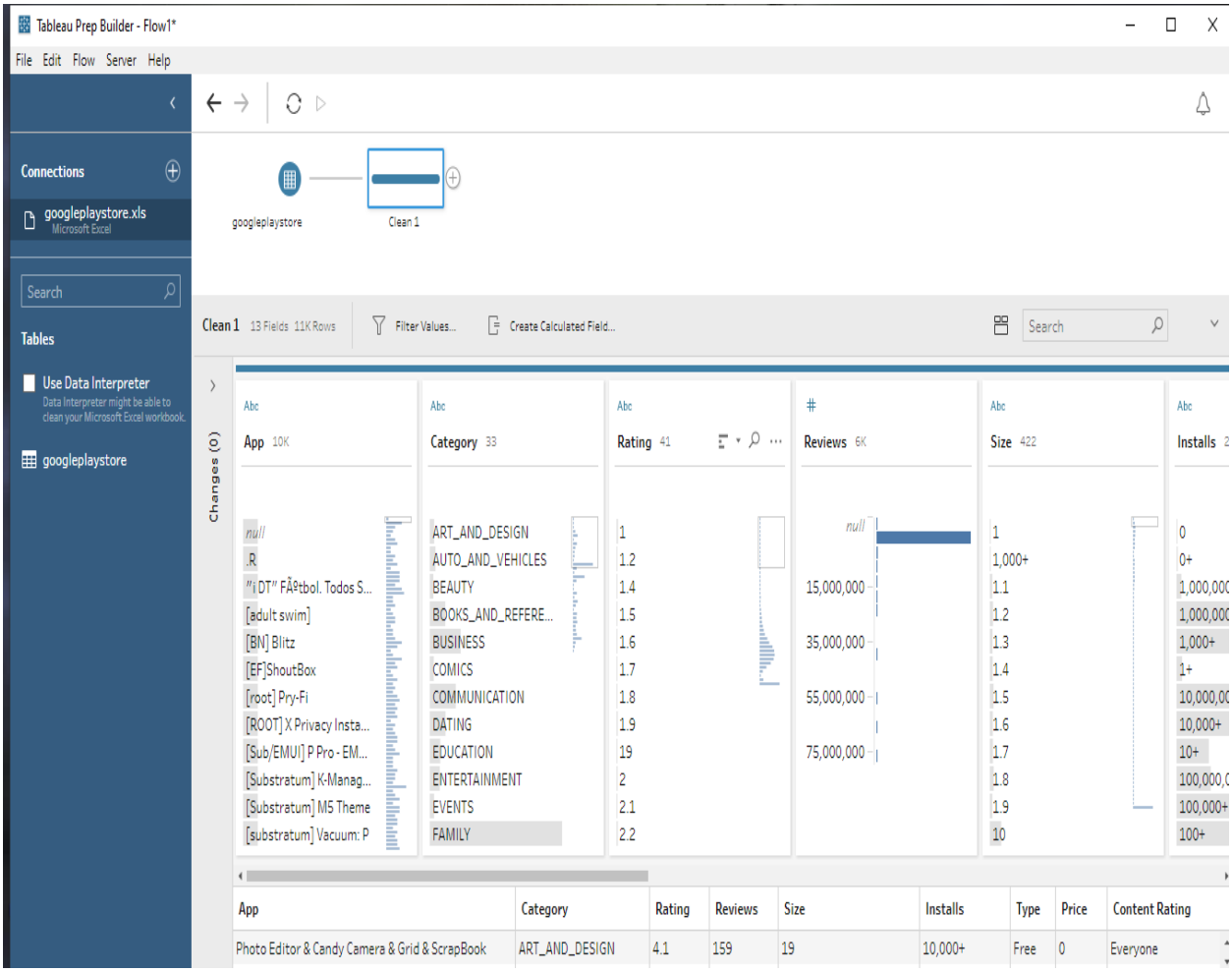
3. Loading

Loading data into the target datawarehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period (nights). Hence, load process should be optimized for performance.In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance .

Load Verification

• Ensure that the key field data is missing
• Test modeling views based on the target tables.
• Check that combined values and calculated measures.

I had done my ETL process with the help of tableau



ANALYSIS ON DATASET

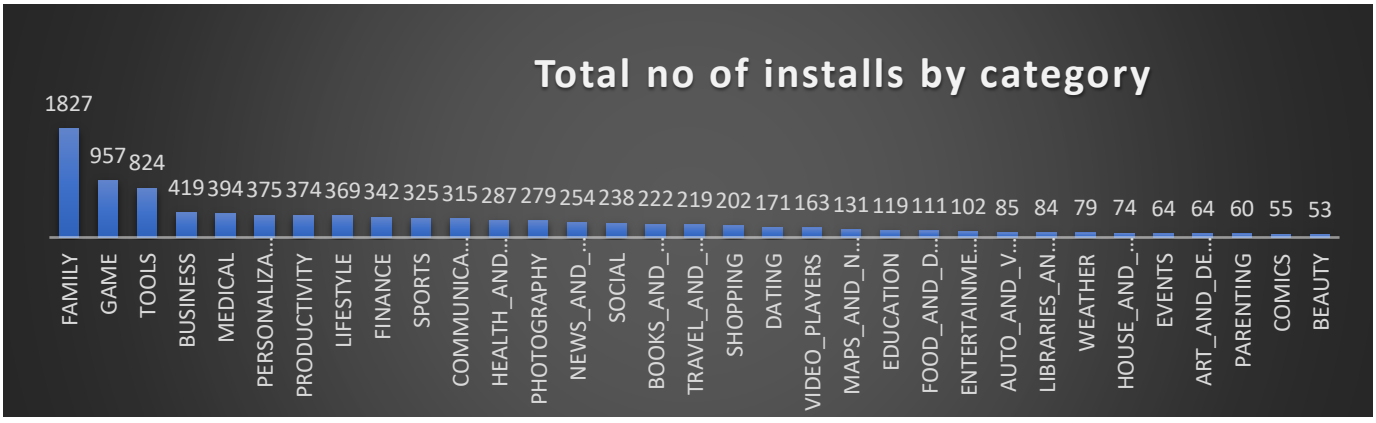
➢ Category wise installs:

General Description : This analysis is all about to find out no of installs per Category.We have many category in apps among them family category apps are having more installs than others and games stood next to Family category. Beauty category is having less installs than all other category.

Analysis Result:

Category	Total Installs
FAMILY	1827
GAME	957
TOOLS	824
BUSINESS	419
MEDICAL	394
PERSONALIZATION	375
PRODUCTIVITY	374
LIFESTYLE	369
FINANCE	342
SPORTS	325

Visualization:



➢ Top 10 Apps Based On Their Category :

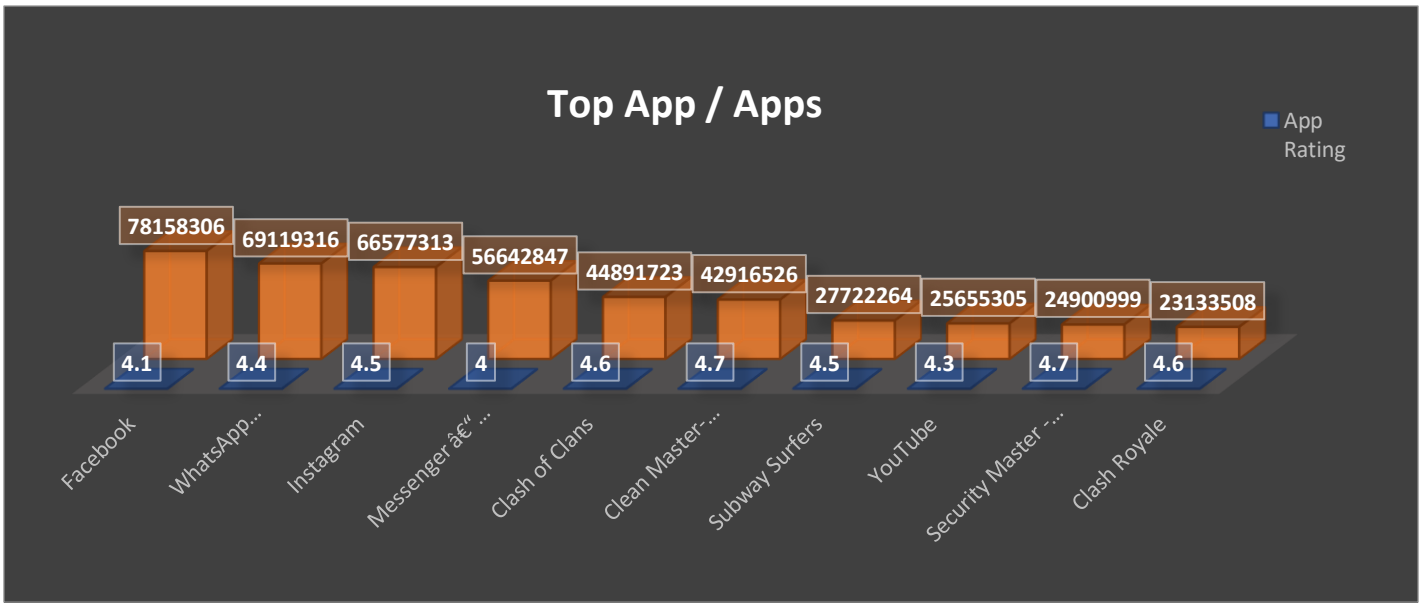
General Description : In this analysis, We are going to find out Top 10 best apps in playstore based on their installs and rating and review.

Also we can find Top 10 best apps with our own selection. i.e; based on installs, Category , Rating, reviews etc…

Analysis result:

App	App Rating	Total Reviews
Facebook	4.1	78158306
WhatsApp Messenger	4.4	69119316
Instagram	4.5	66577313
Messenger â€“ Text and Video Chat for Free	4	56642847
Clash of Clans	4.6	44891723
Clean Master- Space Cleaner & Antivirus	4.7	42916526
Subway Surfers	4.5	27722264
YouTube	4.3	25655305
Security Master - Antivirus, VPN, AppLock, Booster	4.7	24900999
Clash Royale	4.6	23133508

Visualization :



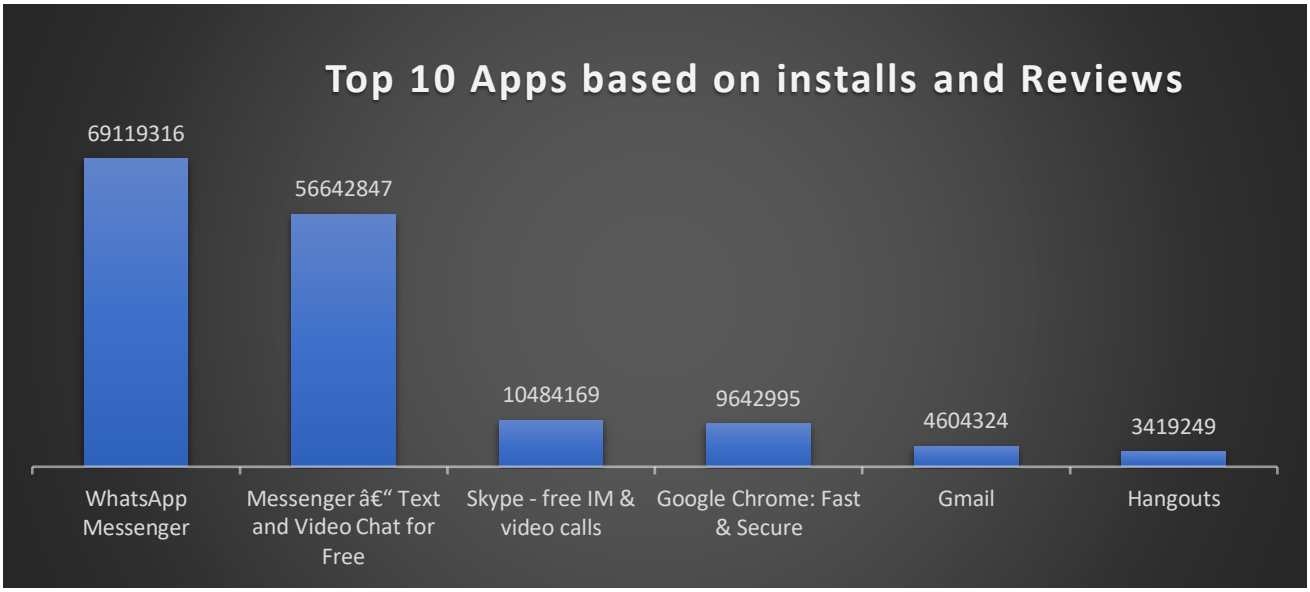
No of Reviews based on Installs :

General description: In this analysis , We are finding the top 10 apps with more reviews in playstore. Also we can find out apps according our selection using slicers.

Analysis Result:

App	Total Reviews
Facebook	78158306
WhatsApp Messenger	69119316
Instagram	66577313
Messenger â€“ Text and Video Chat for Free	56642847
Clash of Clans	44891723
Clean Master- Space Cleaner & Antivirus	42916526
Subway Surfers	27722264
YouTube	25655305
Security Master - Antivirus, VPN, AppLock, Booster	24900999
Clash Royale	23133508

Visualization :



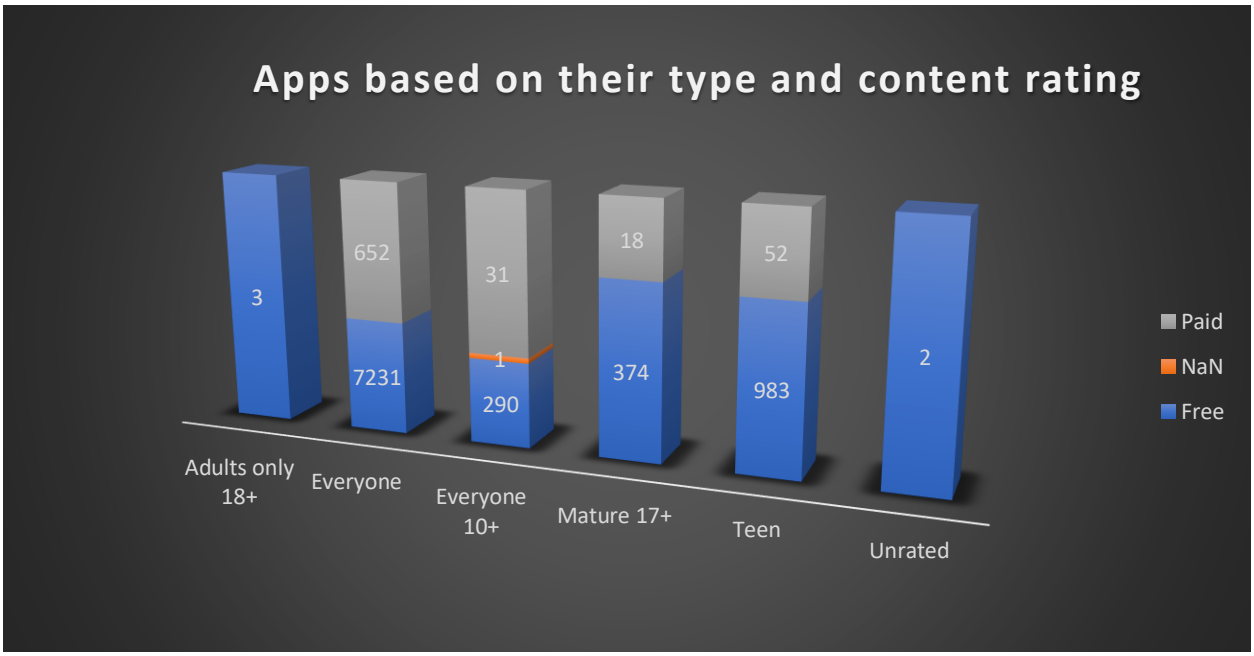
➢ No of installs based on Content Rating And Type :

General Description : In this analysis , weare finding total number of installs based on their content rating and type (whether that app is Free app or paid app) .

Analysis Result:

Count of Installs	Type		
Content Rating	**Free**	**NaN**	**Paid**
Adults only 18+	3		
Everyone	7231		652
Everyone 10+	290	1	31
Mature 17+	374		18
Teen	983		52
Unrated	2		

Visualization:



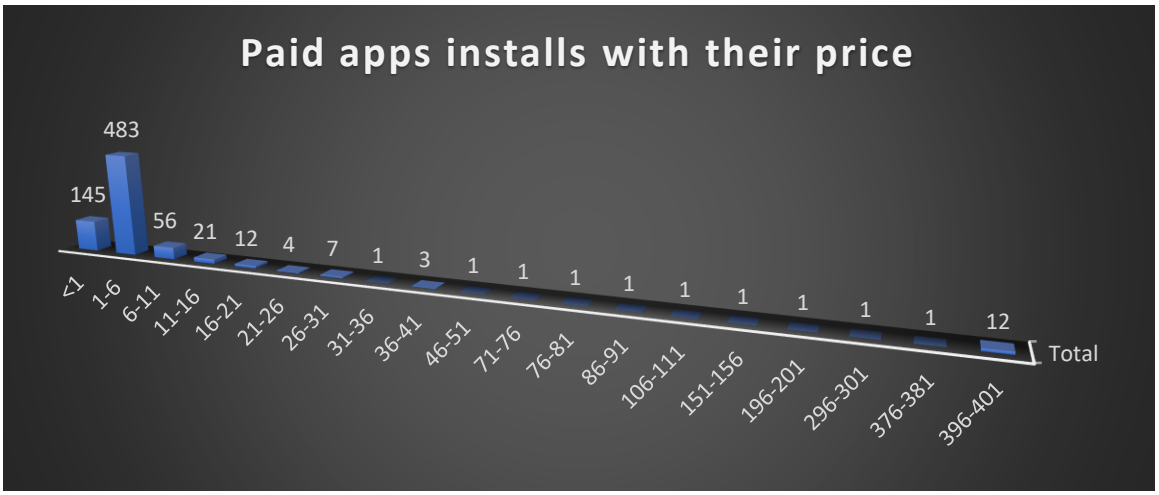
➢ Price of Paid Apps :

General Description: In this analysis, we are going to find No of installs of the paid apps with in a price range in Dollors.

Analysis Result:

Price in dollors	Count of Installs
<1	145
1-6	483
6-11	56
11-16	21
16-21	12
21-26	4
26-31	7
31-36	1
36-41	3
46-51	1
71-76	1
76-81	1

Visualization :



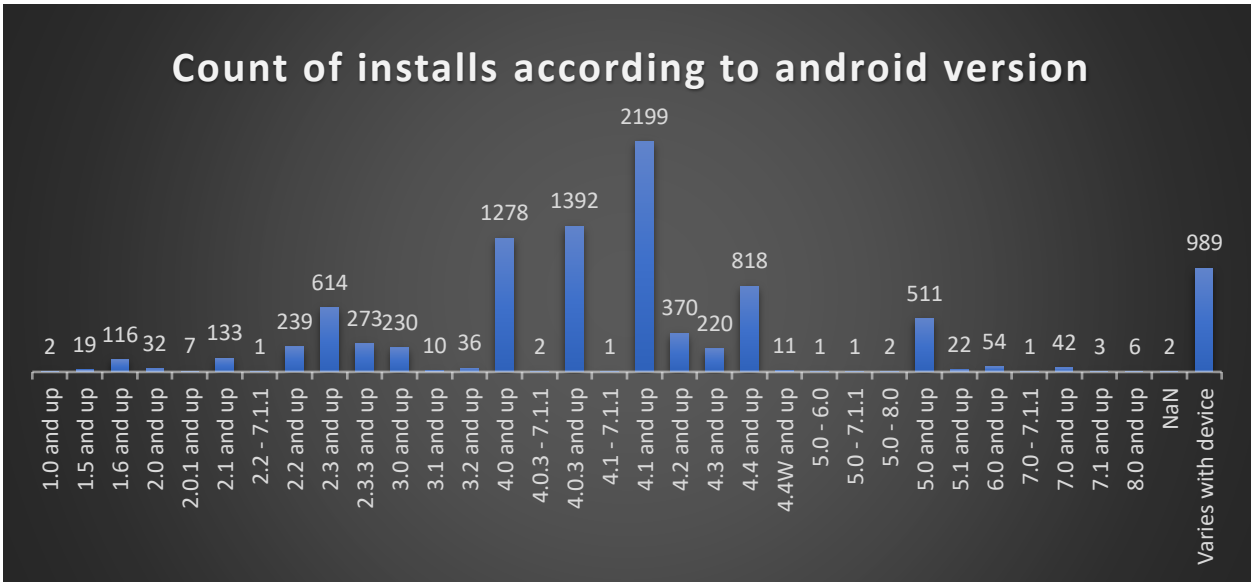
➢ No of Android Wise Installs :

General Description: In this analysis, We are going to find out total No of Apps that run on a specific Android version . Every app is designed according to Android Version . According to my analysis android version 4.1 and up will support more apps.

Analysis Result:

Android Ver	No of Installs
4.1 and up	2199
4.0.3 and up	1392
4.0 and up	1278
Varies with device	989
4.4 and up	818
2.3 and up	614
5.0 and up	511
4.2 and up	370
2.3.3 and up	273

Visualization :



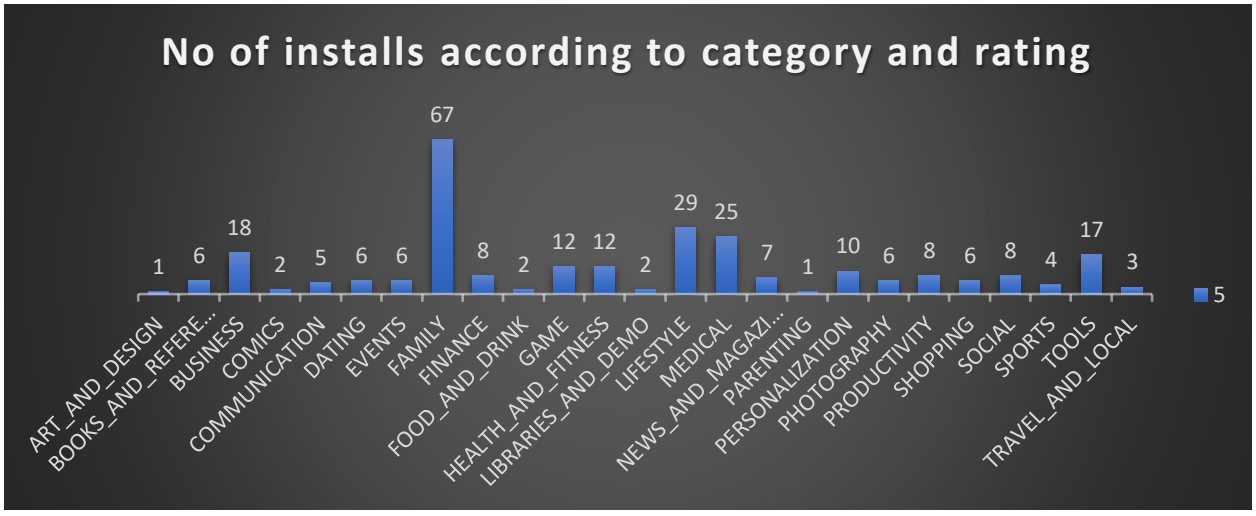
➢ No of apps based on Category and Rating :

General Description: In this analysis , We are having count on apps based on their category by Its Rating. For example : To find out count of installs on all the apps with **5 rating** based on their category.

Analysis Result:

Category	5
FAMILY	67
LIFESTYLE	29
MEDICAL	25
BUSINESS	18
TOOLS	17
HEALTH_AND_FITNESS	12
GAME	12
PERSONALIZATION	10

Visualization :



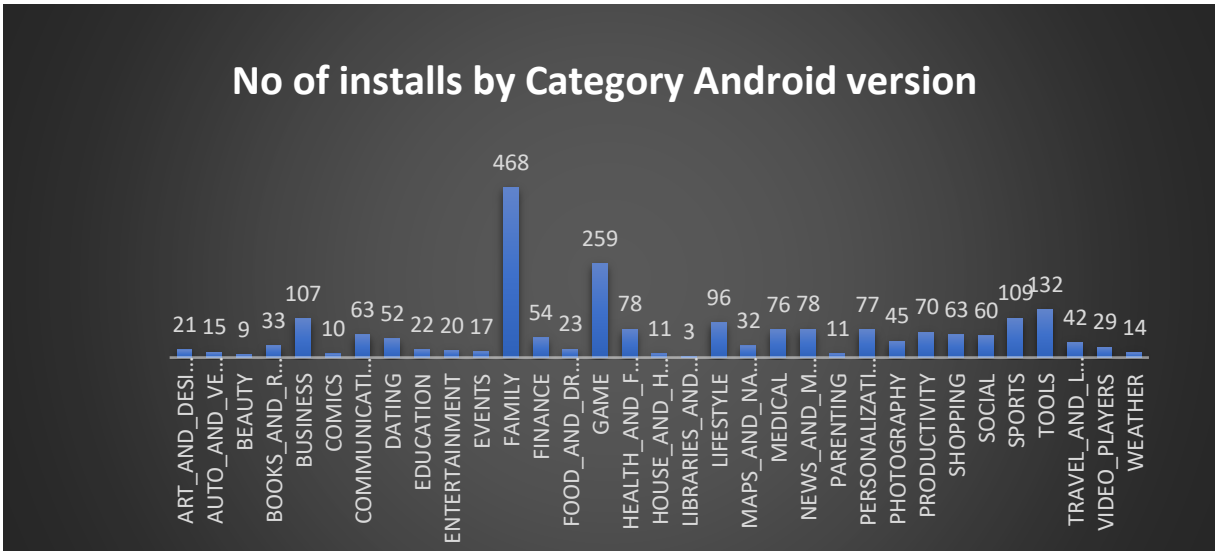
➢ No of apps based on Category and Android Version :

General Description: In this analysis , We are having count on apps based on their category by Its Android version. For example: To find out count of installs on all the apps with Android version **4.1 and up** based on their Category.

Analysis result:

Category	4.1 and up
FAMILY	468
GAME	259
TOOLS	132
SPORTS	109
BUSINESS	107
LIFESTYLE	96
HEALTH_AND_FITNESS	78
NEWS_AND_MAGAZINES	78
PERSONALIZATION	77
MEDICAL	76

Visualization :



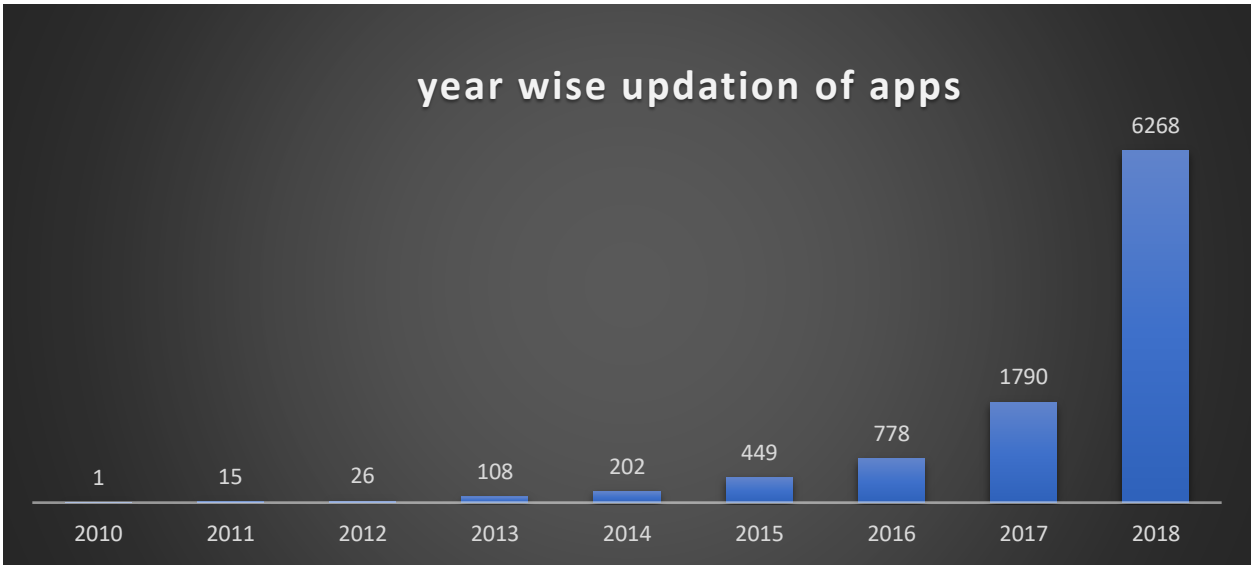
➢ No of apps Updated per Year :

General Description: In this analysis , We are going to find out no of Apps updated in a specific year . It means no of updated in 2010, 2015, 2018 etc…

Analysis Result:

Years	No of Installs
2010	1
2011	15
2012	26
2013	108
2014	202
2015	449
2016	778
2017	1790
2018	6268

Visualization :



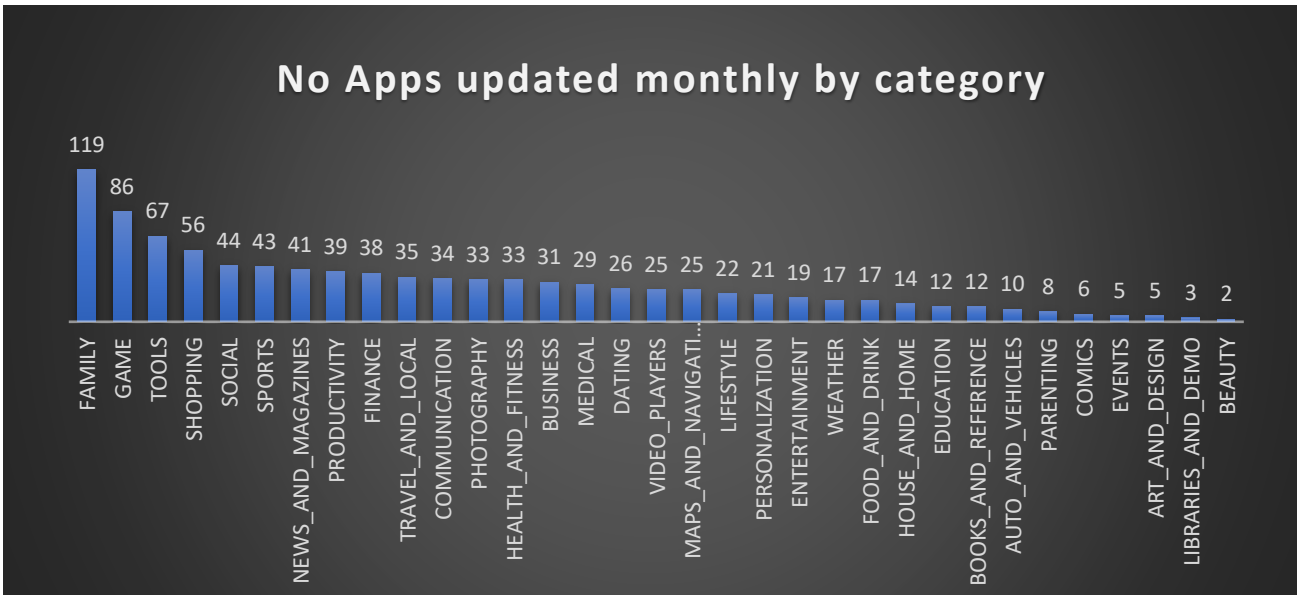
➢ No of apps Updated per Month in a Year:

General Description: In this analysis , We are going to find out no of Apps updated in a specific month in year . It means no of updated in January, May , June etc…

Analysis Result:

Category	Aug
FAMILY	119
GAME	86
TOOLS	67
SHOPPING	56
SOCIAL	44
SPORTS	43
NEWS_AND_MAGAZINES	41
PRODUCTIVITY	39
FINANCE	38

Visualization:



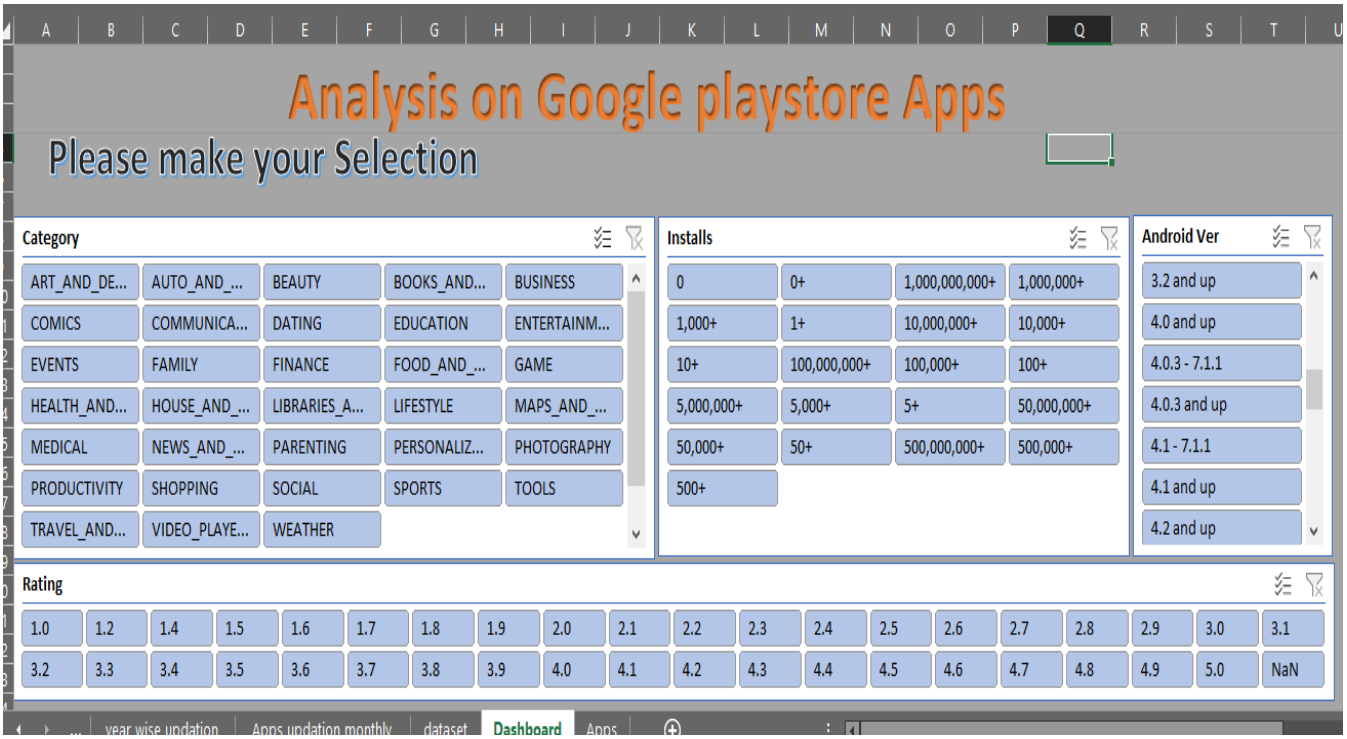
Analysis Result

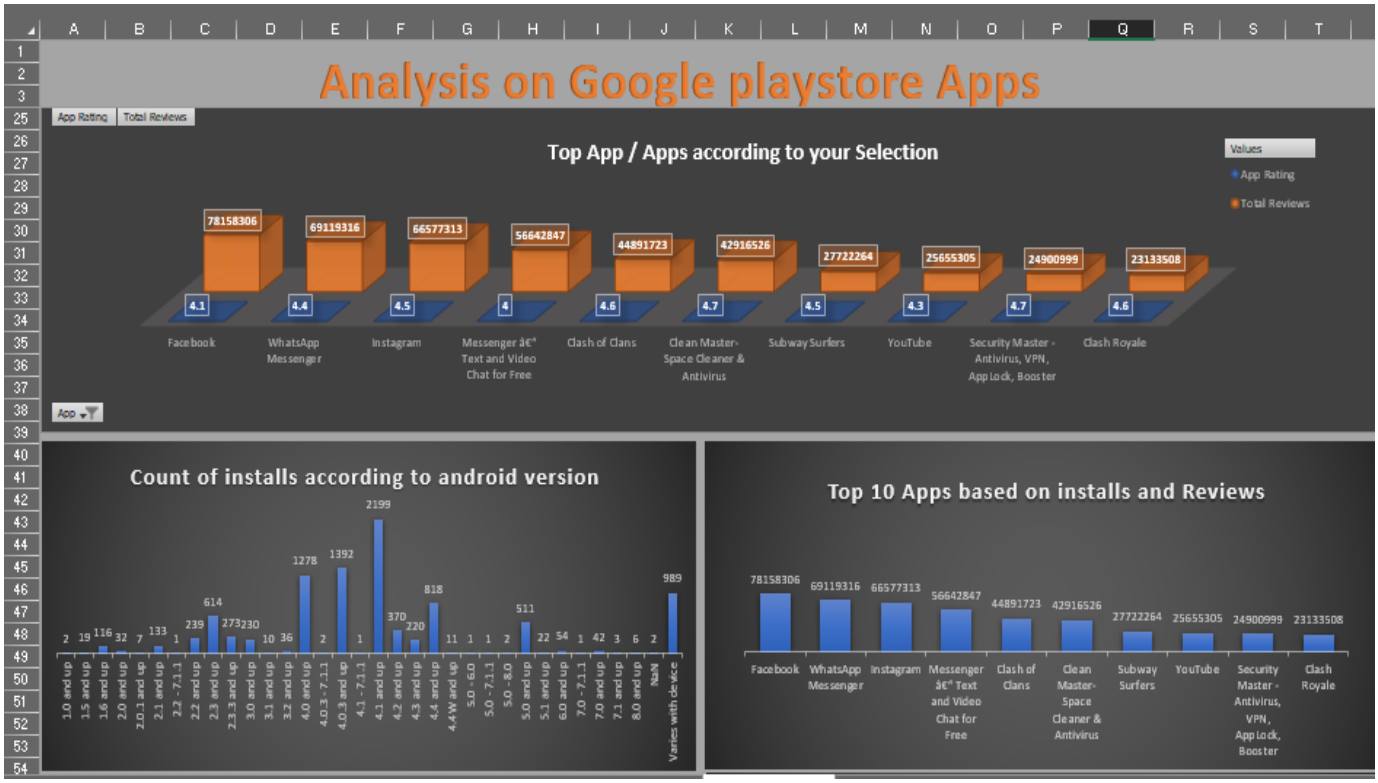
Dashboard 1 :-

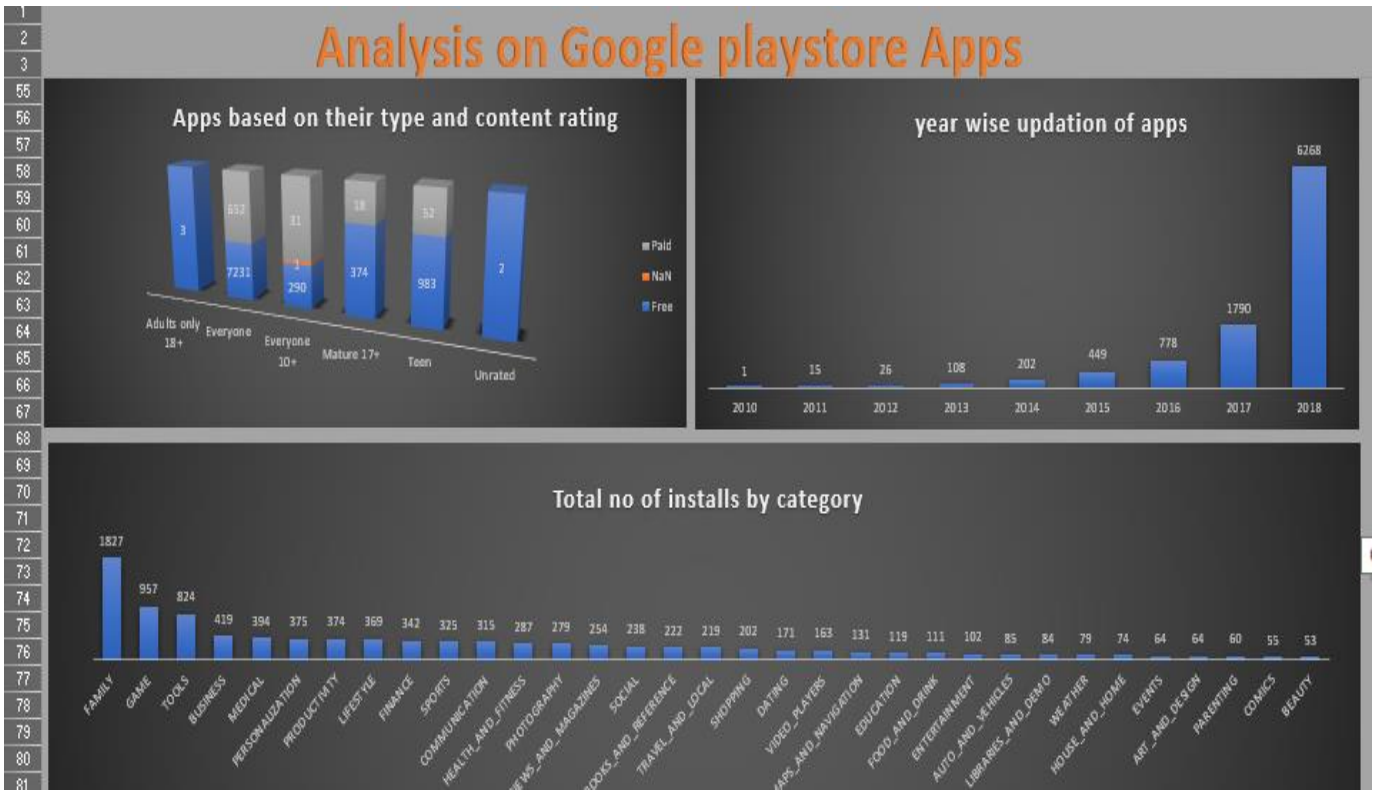
General Description: In Dashboard we are having over all analysis of the project. In my dashboard we can have our selection then it will find out the most popular App in playstore with its rating and reviews according to our selection . Also we can find out no of installs done along with Android version . Moreover we can also find total no of installs along with their category according to selection.We can also find out Apps last updated Year .

Visualization :

Slicers




Analysis on Google playstore Apps
App Rating
Total Reviews
Top App / Apps according to your Selection
Values
App Rating
Total Reviews
78158306
69119316
66577313
56642847
44891723
42916526
27722264
25655305
24900999
23133508
4.1
4.4
4.5
4
4.6
4.7
4.5
4.3
4.7
4.6
Facebook
WhatsApp Messenger
Instagram
Messenger â€" Text and Video Chat for Free
Clash of Clans
Clean Master- Space Cleaner & Antivirus
Subway Surfers
YouTube
Security Master - Antivirus, VPN, AppLock, Booster
Clash Royale
App
Count of installs according to android version
Top 10 Apps based on installs and Reviews

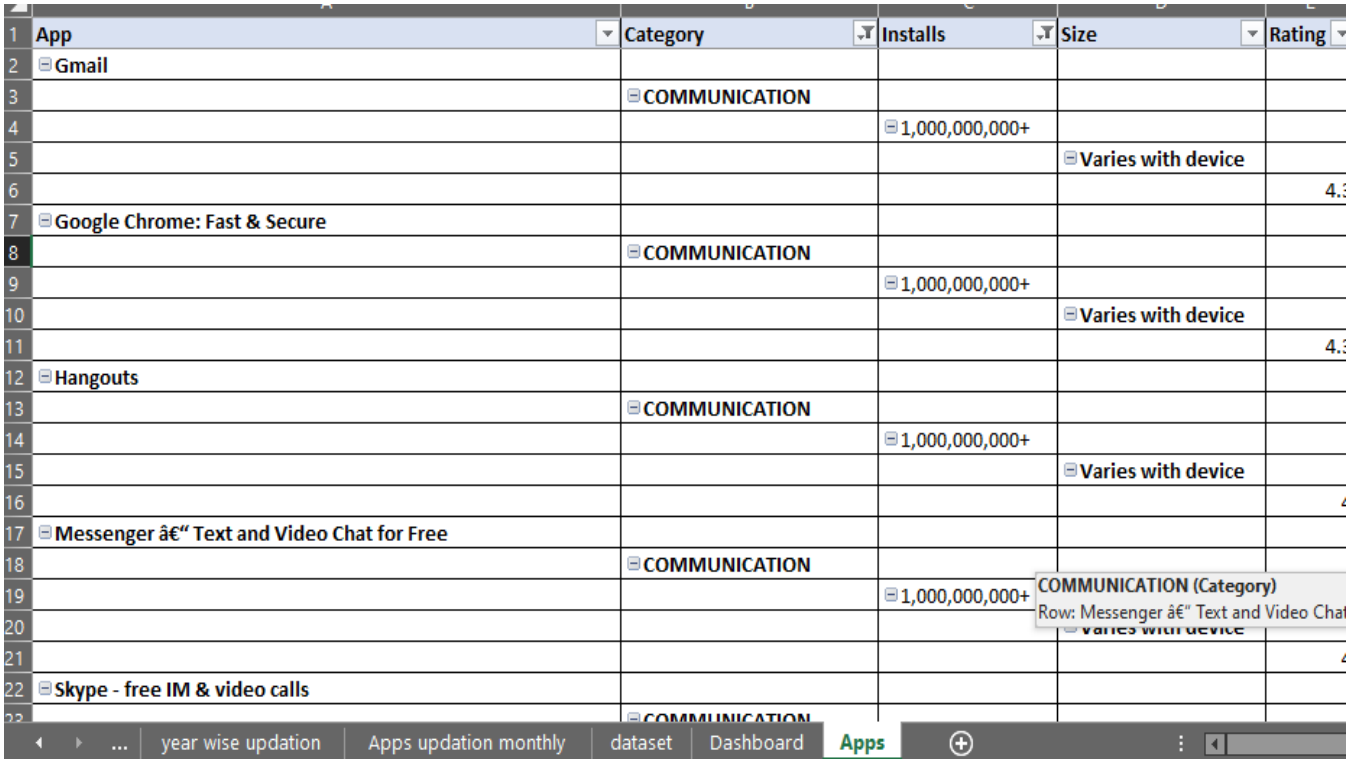

Analysis on Google playstore Apps
Apps based on their type and content rating
Adults only 18+
Everyone
Everyone 10+
Mature 17+
Teen
Unrated
Paid
NaN
Free
year wise updation of apps
Total no of installs by category
FAMILY
GAME
TOOLS
BUSINESS
MEDICAL
PERSONALIZATION
PRODUCTIVITY
LIFESTYLE
FINANCE
SPORTS
COMMUNICATION
HEALTH_AND_FITNESS
PHOTOGRAPHY
NEWS_AND_MAGAZINES
SOCIAL
BOOKS_AND_REFERENCE
TRAVEL_AND_LOCAL
SHOPPING
DATING
VIDEO_PLAYERS
MAPS_AND_NAVIGATION
EDUCATION
FOOD_AND_DRINK
ENTERTAINMENT
AUTO_AND_VEHICLES
LIBRARIES_AND_DEMO
WEATHER
HOUSE_AND_HOME
EVENTS
ART_AND_DESIGN
PARENTING
COMICS
BEAUTY

Dashboard 2 :-

General Description:

In this dashboard we will have Information About the like App name and size, Installs and its Category.

Visualization:



Future Scope

Since the objectives are clear for this report and the complete dataset is available. This report can be helpful for future studies and analysis for students. One can analyze different objective According to his/her choice.

Visualization in this report is in fully detailed manner which can help any individual to understand the data properly. Comparisons among data are easily understandable.

Also the visualization of datasets can be examined to find out an app as per our choices like Size, Rating, Category .

The dataset and the dashboard are dynamically linked, so any changesdone to the dataset will take effect in the graph also.

A power query can be used to add more datasets to this dataset.

References

[1] ETL What it is and why it matters , Wikipedia.

[2] Ralph., Kimball, (2004). *The data warehouse ETL toolkit : practical techniques for extracting, cleaning, conforming, and delivering data*. Caserta, Joe, 1965-. Indianapolis, IN: Wiley. *ISBN 978-0764579233. OCLC 57301227.*

[3] Playstore for apps References.

Bibliography

1. https://www.kaggle.com/lava18/google-play-store-apps
2. https://en.wikipedia.org/wiki/Extract,_transform,_load
3. https://www.guru99.com/etl-extract-load-process.html
4. https://42matters.com/app-market-explorer