DATA MANAGEMENT PROJECT REPORT

App Store Dashboard

Submitted by:

Katreddy Venkateswara Reddy

11802996

KMO73

Under the guidance of:

Miss. Ashu (Assistant Professor)

Schools of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Katreddy Venkateswara Reddy bearing Registration number 11802996 has completed Data Management (INT 217) project titled, "App Store Dashboard" under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Date:

DECLERATION

I, Katreddy. Venkateswara Reddy, student of Computer Science and Engineering 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:5-12-2020 Signature:

Registration No:11802996

ACKNOWLWDGEMENT

I take this opportunity to present our votes of thanks to all those guideposts who really acted as lightening pillars to enlighten my way throughout this Project that has led to successful and satisfactory completion of this Project. I am grateful to Lovely Professional University for providing us with an opportunity to undertake this Project and providing us with all the facilities. I am highly thankful to All for their active support, valuable time and advice, whole-hearted guidance, sincere cooperation and painstaking involvement during the project and in completing the assignment of preparing the said project within the time stipulated. Lastly, I am thankful to all those, particularly the various friends, who have been instrumental in creating proper, healthy and conductive environment and including new and fresh innovative ideas for me during the project, without their help, it would have been extremely difficult for me to complete the project in a time bound framework.

INDEX

1. INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information conclusions and supporting decision making. Data analysis has multiple facts and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science and social science domains.

App Store analysis is one of the most useful analysis for Apps who want to install the best App applications of every part of the world which lies in their budget. This analysis can also be useful to find the most preferable app applications in various types.

The collected data has been stored in the comma separated value in App Store.csv. Every app contains the following variables.

- App : Name of the apps
- Category : App in which Category
- Rating : Average rating of the apps out of 5
- Reviews : Reviews of the apps
- Size : Size of the apps
- Type of Downloads : Types of downloads in the app store
- Downloads : Number of downloads of the app in AppStore
- Content Rating : Ratings of the customer as text

2.SCOPE OF THE ANALYSIS

- The users who wants to download any application first choose the application based on their size of the app. If any user wants to know more about the size, number of downloads of that particular app. User can check on the slicer to get know about it.

    1. Analysing the best app based on aggregate rating.
    2. Analysing the apps from the customer reviews.
    3. Analysing the app with a greater number of downloads.
    4. Check whether apps have free or paid.
    5. Analyse the number of apps for paid in the app store.

Analyse the app category which get a greater number of reviews.

3.EXISTING SYSTEM

The existing system is currently used to make the dashboard on "Microsoft Excel for Office 365 MSO 64-bit" of version "16.0.11328.20438" and for data cleaning "Tableau Prep Builder" of version "19.31.19.0923.1415" is used.

- DRAWBACKS

    1. Manual Effort: It is more time consuming and takes lots of efforts of the user to manipulate and analyse tons of data in an excel workbook with lots of sheets. With huge amount of data excel slows down the process and sometimes it does not respond.
    2. Human Error: As the amount of data grows within an Excel spreadsheet, there is increased room for human error and formula errors because of the manual process of copying and pasting data into the workbook.
    3. Security: User gets the full access to the dashboard and data within it. This can become a major issue when sensitive information is being shared both internally and externally because there is no

way to secure this data.

4. SOURCE OF DATASET

The source of the App Store dataset is Kaggle. Link is provided below:

https://www.kaggle.com/lava/google-play-store_apps?select = googleplaystore.csv

5. ETL PROCESS

ETL is defined as a process that extracts the data from different source systems, then transforms the data and finally loads the data into the Data Warehouse system. ETL is known as Extract, Transform and Load.

1. Extraction:

   For data extraction we need to know which variables are not required to fulfil the analysis. In this case googleplaystore.csv has such columns like App ID, category ID, last updated time does not need for the analysis. So, we discard or extract it.

2. Transformation:

   In this process we need to check for the data mistakes like spelling mistakes such as “Infinate Paintar” and “Infinite Painter” or “ibs Painte X” and “ibis Paint X”, removing null values, grouping same category data such as “ART_AND_DESIGN” and “Art & Design” or “FOOD_AND_DRINK” and “Food & Drink”. We need to transform the data in such a way that it is suitable for the required analysis. All these processing is done in “Tableau Prep Builder”.

3. Load:

   After finishing the transformation of the data, we need to get the output of the cleaned dataset and new data is saved in a new excel file named Appstore_Dataset.csv. We need to save the data in xlsx format so that we can use all the features of the excel. New file name is saved with named Appstore_Dashboard.xlsx. After that we need to open the dataset in Excel for further analysis.

6. ANALYSIS ON DATASET

1. Cost of App which it is Free or Paid

- Introduction

In order to get the best apps with free or without free. This Analysis helps in finding the best Application on category having a cost of App which it is Free or Paid.

- General Description

We have the columns of Type of Downloads and Downloads for this analysis. We need to make a table for Type of Downloads. So, we create a table with column names Category, Type of Download and Downloads.

- Specific Requirements

We need some functions to get this job done. In column Type of Download we apply LARGE function and for the Downloads we apply INDEX and MATCH functions. And also we apply Value field setting for showing values in Percentage (%). For Category we apply FILTERS.

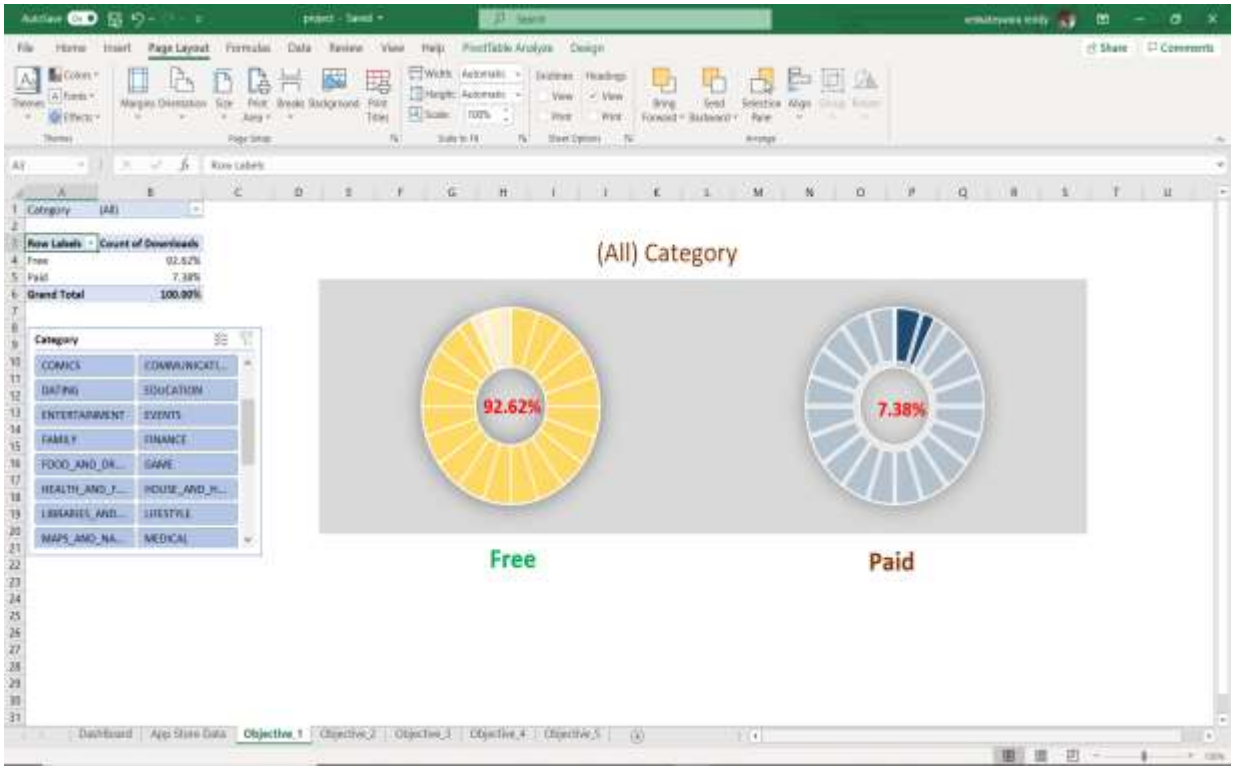
- Analysis Result

Here we analysed cost of app with free (92.62%) and for paid (7.38%). The category with 100% free are “1.9”, “COMICS”, “BEAUTY” and “HOUSE_AND_HOME” Category.

- Visualization

From the pie chart, we conclude that cost of apps with free has value 62.62% and cost of app with paid has 7.38%. We can also select Category manually using Slicers. In this slicer we select randomly app Category and it reflected on the pie chart as well.

From here we can conclude that “1.9”, “COMICS”, “BEAUTY” and “HOUSE_AND_HOME” Category has the highest aggregate percentage over the value 100% from the selected Category.



2. Finding the best Application depends on content rating through Count of Rating:

- Introduction

To get to know about which content has the highest percentage of count of rating on which category we need to analysis this one. In this dataset we are provided with content rating and category. This will tell about the age wise analysis. How percentage of adults are using this app. How much percentage of teenagers are using this application.

- General Description

We have the columns of Content Rating, Category and Count of Rating for this analysis. Let's create Bar chat with the Content Rating and Count Rating.

- Specific Requirements

We need LARGE function to find the largest value from the total number of Count of Rating. INDEX function is also needed to match the content rating corresponding to the value of the Count of Rating columns. And also we apply Value field setting for showing values in Percentage (%). For Category we apply FILTERS.

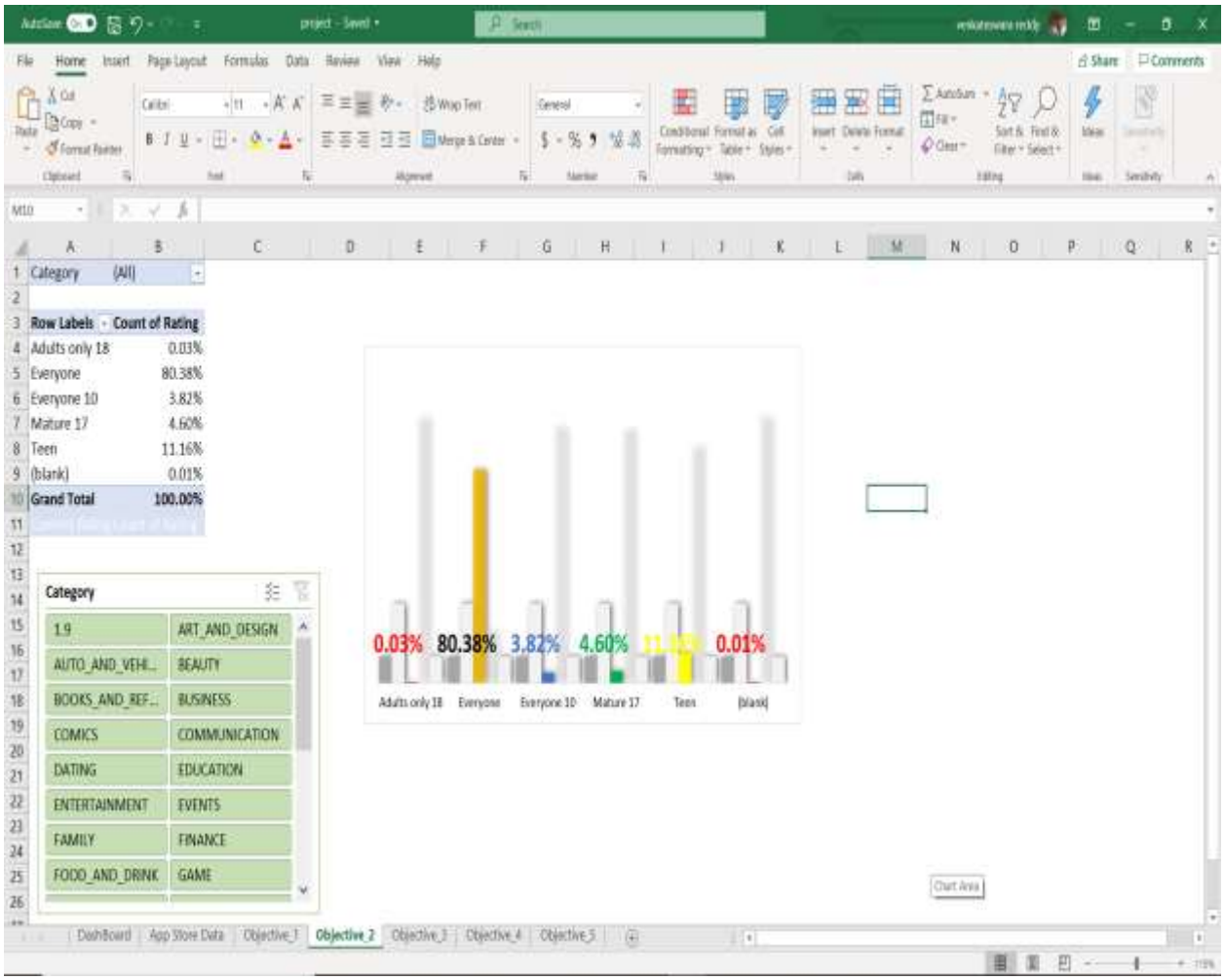
- Analysis Result

Here has the highest percentage (%) of count of rating with content rating.

- Visualization

From the bar chart we can conclude that Everyone has the highest percentage of count of rating with value 80.38% and (blank), Adult only 18 has the lowest percentage of count of rating with values 0.01% and 0.03% in the following dataset. We can also select category names with the help of slicers and get the

category with maximum percentage of content rating from the selected category.



3. Finding the best Application on category depends on having high Number of Downloads:

- Introduction

To get to know that how many apps category have number of Downloads Facility and how many have not. We need to count the total number of apps that has number of Download facility and same for which has not.

- General Description

We have two columns of Downloads and Count of Apps. We need to create two rows to calculate. Also we have one column category.

- Specific Requirements

We need Count function to calculate the total number of Apps which have and have not downloads. For the category we need FILTERS.

- Analysis Result

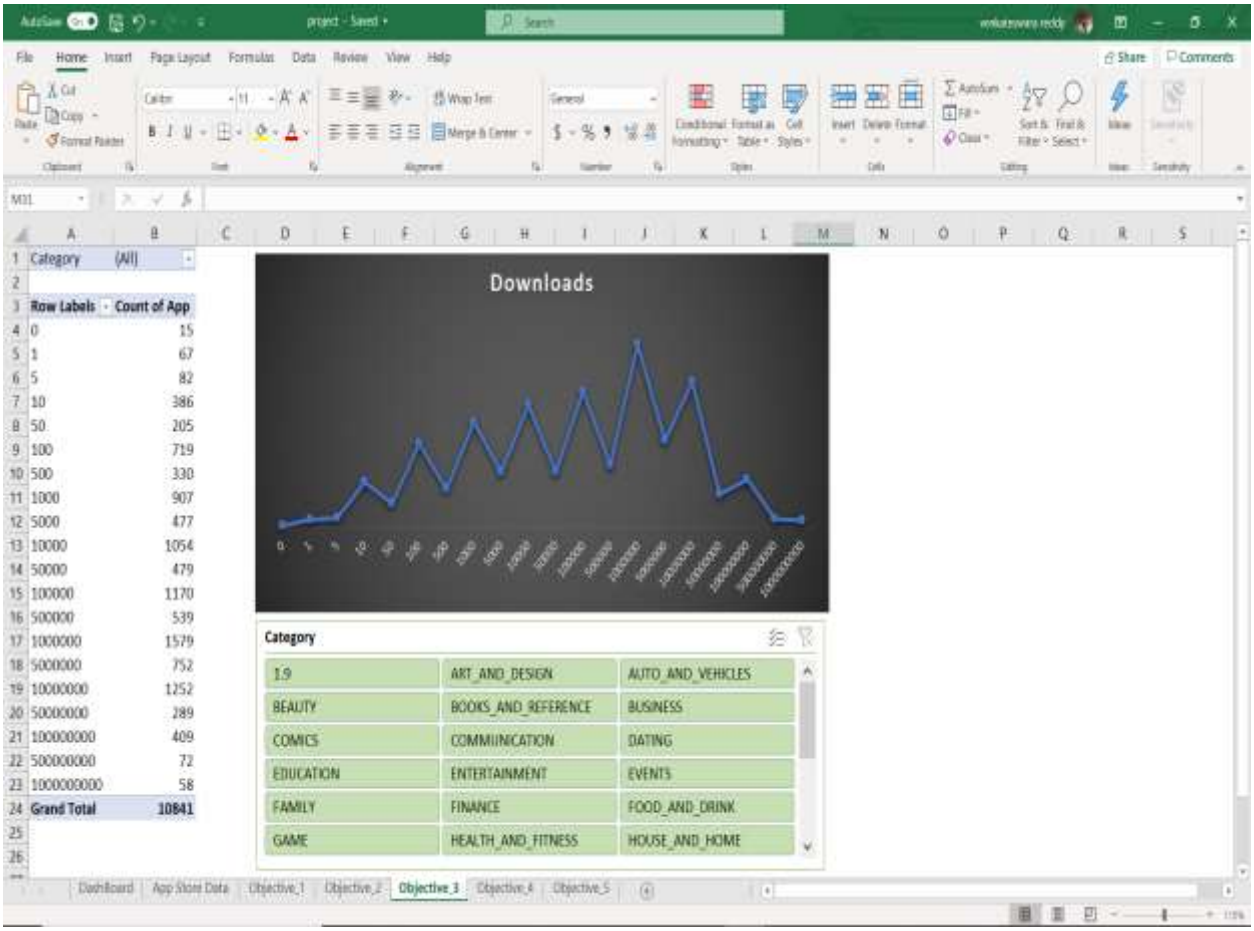
Here we get the two rows with the corresponding value.

- Visualization

After analysis we get that 15 Apps do have Downloads and the 58 Apps have the highest Downloads (1000000000).

By using category we can find on which category have the high number of downloaded Apps from the selected Category.

From this user finding the best application on category depends on having high number of downloads.


project - Saved
File Home Insert Page Layout Formulas Data Review View Help
Share Comments
Downloads
Category (All)
Row Labels Count of App
0 15
1 67
5 82
10 386
50 205
100 719
500 330
1000 907
5000 477
10000 1054
50000 479
100000 1170
500000 539
1000000 1579
5000000 752
10000000 1252
50000000 289
100000000 409
500000000 72
1000000000 58
Grand Total 10841
Category
1.9
ART_AND_DESIGN
AUTO_AND_VEHICLES
BEAUTY
BOOKS_AND_REFERENCE
BUSINESS
COMICS
COMMUNICATION
DATING
EDUCATION
ENTERTAINMENT
EVENTS
FAMILY
FINANCE
FOOD_AND_DRINK
GAME
HEALTH_AND_FITNESS
HOUSE_AND_HOME
Dashboard
App Store Data
Objective_1
Objective_2
Objective_3
Objective_4
Objective_5

4. Finding the best Application on category depends on having Highest Rating to App:

- Introduction

To get to know that how many apps category have highest rating to app and how many have not. We need to count the total number of apps that has number of Download facility and same for which has not.

- General Description

We have two columns of Rating and Count of Apps. We need to create two rows to calculate. Also we have one column category.

- Specific Requirements

We need Count function to calculate the total number of Apps which have and have not rating. For the category we need FILTERS and for columns of Rating we group the values.

- Analysis Result

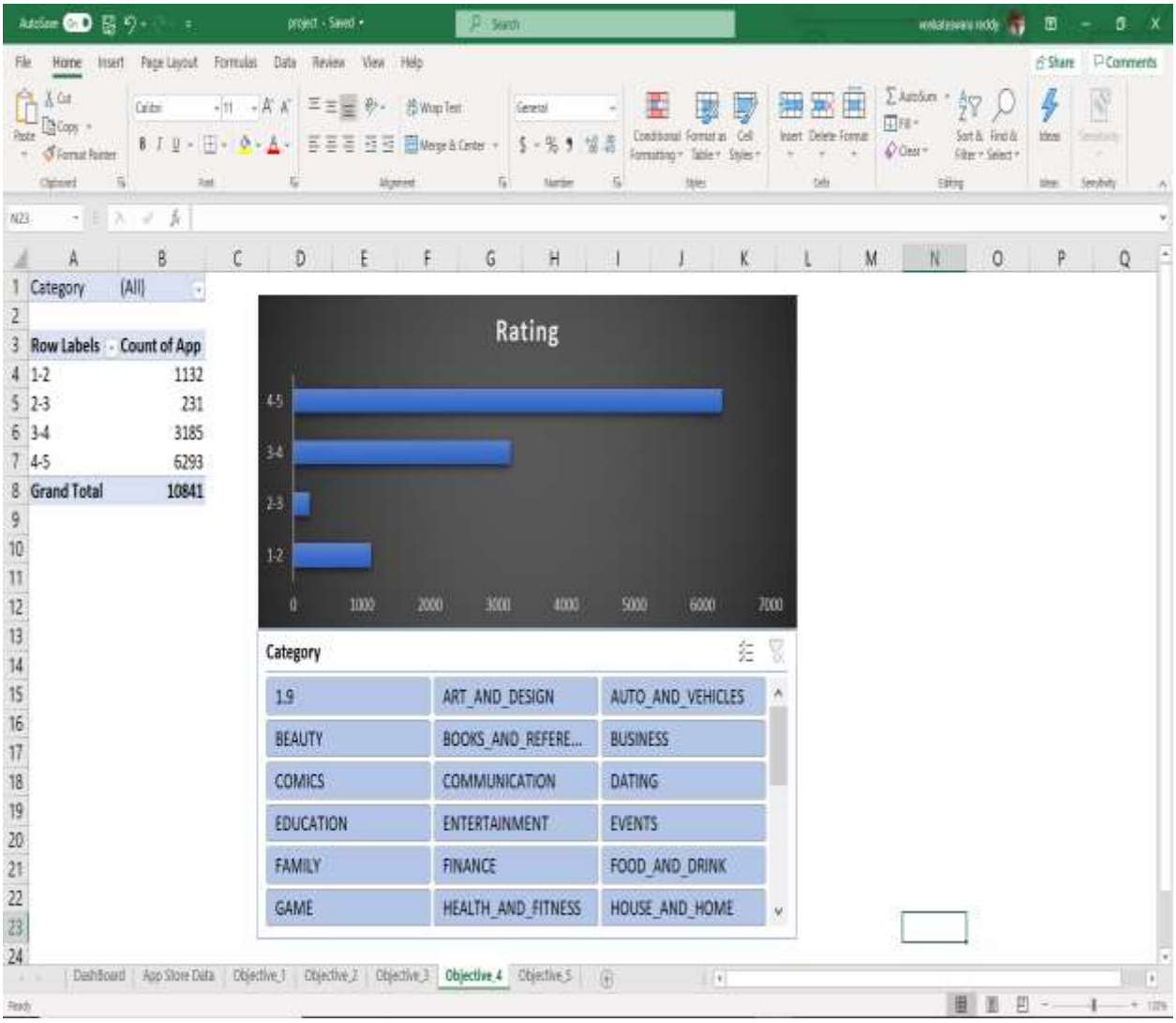
Here we get the two rows with the corresponding value.one is rating of grouped values and other is count of apps (no of apps having the corresponding rating values).

- Visualization

After analysis we get that 6293 Apps have highest Rating and the 1132 Apps have the lowest rating.

By using category we can find on which category have the high rating with count of Apps from the selected Category.

From this user finding the best application on category depends on having highest Rating to app.



Category	(All)
Row Labels	Count of App
1-2	1132
2-3	231
3-4	3185
4-5	6293
Grand Total	10841

The following graph depicts the Rating Distribution for Each application Based on the Rating. It is done by using the slicer. It is inserted from which we can select the app and the respective rating is shown.

5. Finding the best Application on category depends on Size of Application:

- Introduction

To get to know that how many apps category have Size(mb) of app very less and how many have high app Size. We need to count the total number of apps that has app size with high and less size.

- General Description

We have two columns of Size and Count of Apps. We need to create two rows to calculate. Also we have one column category.

- Specific Requirements

We need Count function to calculate the total number of Apps which have and have not size. For the category we need FILTERS and for columns of Rating we group the values.

- Analysis Result

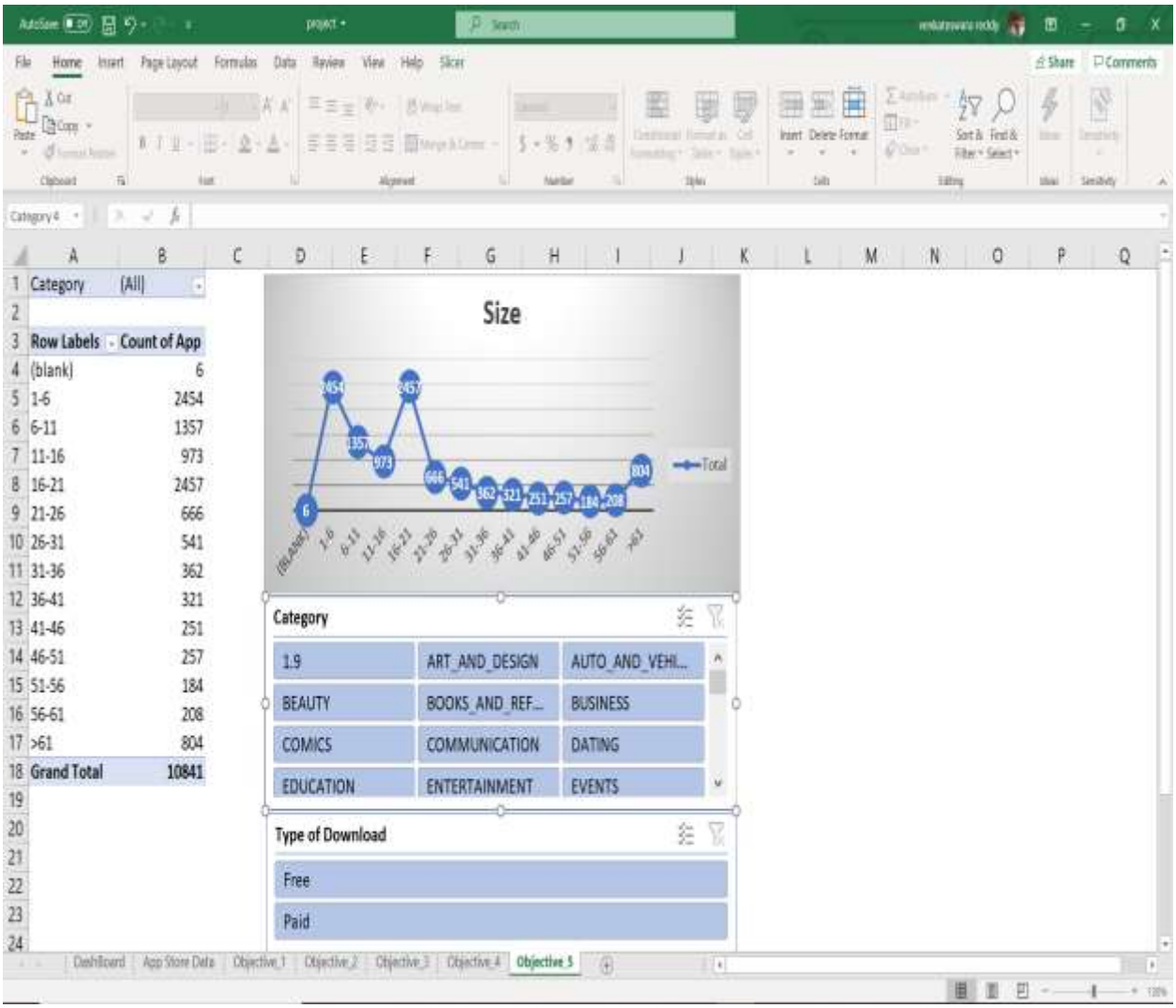
Here we get the two rows with the corresponding value.one is Size of grouped values and other is count of apps (no of apps having the corresponding Size values).

- Visualization

After analysis we get that 804 Apps have highest App Size(>61) and the 2454 Apps have the lowest app Size(1-6).

By using category we can find on which category have the high size apps with count of Apps from the selected Category.

From this user finding the best application on category depends on Size of Application.



Category	(All)
Row Labels	Count of App
(blank)	6
1-6	2454
6-11	1357
11-16	973
16-21	2457
21-26	666
26-31	541
31-36	362
36-41	321
41-46	251
46-51	257
51-56	184
56-61	208
>61	804
Grand Total	10841

The above pivot table will say the size of the application. The line plot will change according to it. We can't select the multiple apps at a time.

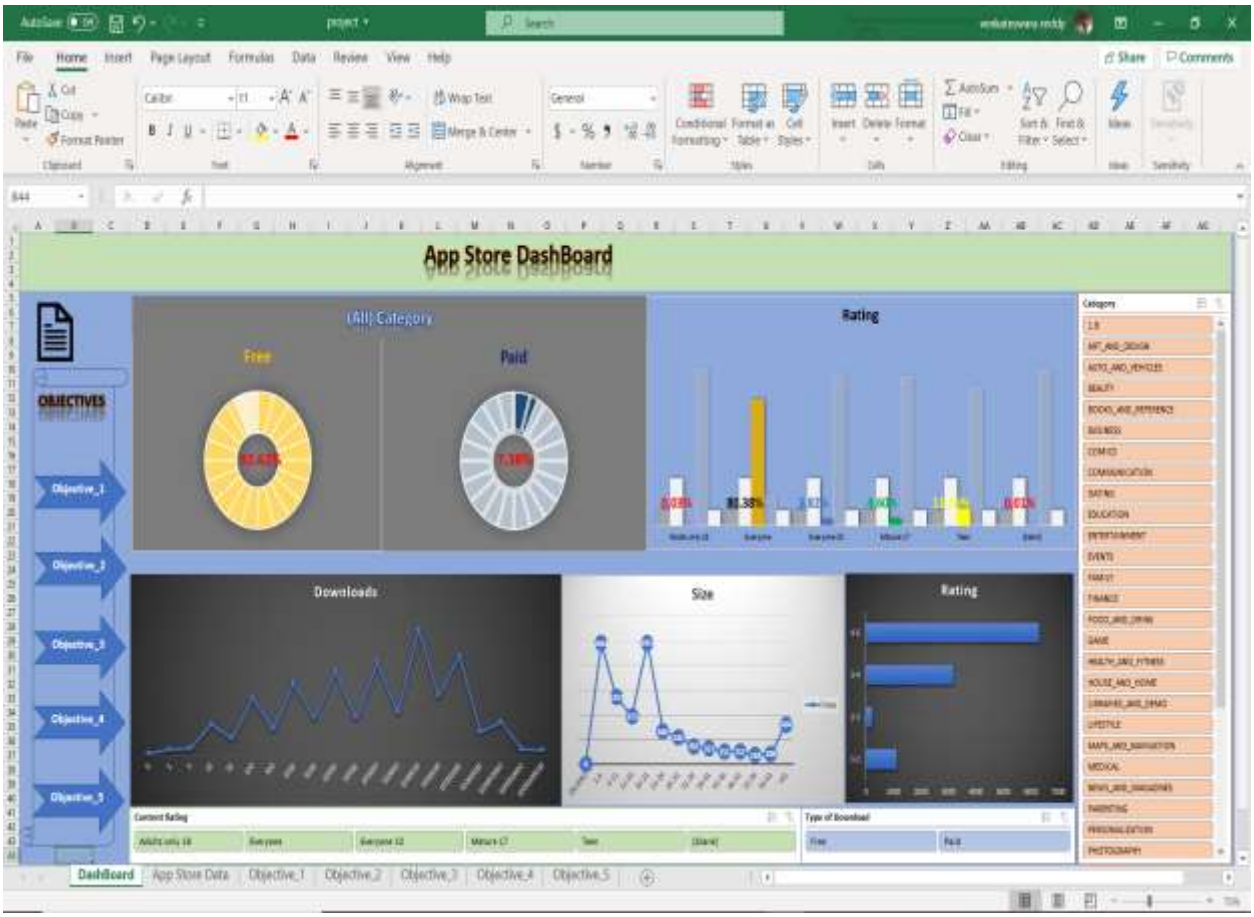
From the above pivot table it is easily showing that size of the applications that used the most.

7.Result of analysis

1.This project will help you in finding the best application based on the size of the app. Users are prefering the least size apps more now a days.so i think minimum size app are most useful.
2.It will least the number of downloads for a particular application. So that maximum downloads can get the updates more that any other less download apps.
3.The users are mostly prefering the free apps than the paid apps. Not every free application is secure. some of them are to be in risky. So user have to download based on the reviews.
4.Mostly womens are using the face beauty application a lot . Some mens are also using compare to men womens are more.
5.Rating is done to all the apps. Highly rated apps are mostly downloaded by the people. Rating is done based on the the user requirements. If the application is reaching the user requirements so we can accept high rating from the user.

8.DASHBOARD

An Excel Dashboard is one pager that helps user and business leaders in tracking key KPIs or metrics and take a decision based on it. It contains charts or tables or slicers that are backed by data.



APP STORE DASHBOARD

9.FUTURE SCOPE

Making dashboard is a way to predict and analyse the future scopes of the data. It makes the manager or the business leaders easy to find problems and get the efficient solution. It helps us to visualize the data in tabular formats or different kinds of charts.

10.REFERENCES

I have taken help from my Data Science teacher Miss. Ashu Madam. I have also taken help from my classmates to complete this project. I am very thankful to them. Apart from these I have researched on the internet for more knowledge about Excel and Dashboard.

11.BIBLIOGRAPHY

- https://en.wikipedia.org/wiki/Dashboard_(business)
- https://www.guru99.com/etl-extract-load-process.html
- https://www.youtube.com/watch?v=K74_FNnlIF8&t=1564s
- https://www.youtube.com/watch?v=sM2mdBu0Owc&t=106s