**INT217 PROJECT REPORT**

(Project Semester August-December 2020)

***STATISTICAL ANALYSIS OF GOOGLE PLAY STORE***

Submitted by :-

***Peddapuram shiva Kumar***

Registration No: ***11811598***

Programme and Section: CSE(Hons) **KM074**

Course Code **INT217**

Under the Guidance of :-

**Komal Arora: 17783**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**



**CERTIFICATE**

##### This is to certify that *Peddapuram Shiva Kumar* Bearing Registration no. 11811598 has completed *INT217* project titled, *“STATISTICAL ANALYSIS OF GOOGLE PLAYSTORE ANALYSIS*” 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 and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, **Peddapuram *Shiva Kumar*,** student of Computer Science 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: Signature

Registration No.11811598 *Peddapuram Shiva 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 **Komal Arora: 17783** 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

1. [Introduction](#Introduction)
2. [**Objectives/Scope of the Analysis**](#Objectives)
3. [**Source of dataset**](#Source)
4. [**ETL process**](#ETL)
5. [**Analysis on dataset**](#Data_Analysis) **(for each analysis)**
6. **Introduction**
7. **General Description**
8. **Specific Requirements, functions and formulas**
9. **Analysis results**
10. **Visualization**
11. [**List of Analysis with results**](#Results)
12. [**References**](#References)
13. **Bibliography**

##### 

##### *Introduction :-*

In this project, I had made an analysis on play store data.

***(1.) App name :***

The name of the app (title of the application)

**(2.)Category :-**

Category to which the app belongs so proper explanation has given below about the categories

(Art & design, Auto and Vehicles, Beauty, Books and Reference, Business, comics, Communication, Education, Entertainment, Dating, Events, Family, Finance, Food and Drink, Game, Health and Fitness, House and Home decors, Lifestyle, Maps and Navigation , Medical , News and Magazine , Parenting Personalization, Photography ,Productivity, Shopping, Social Sports ,Tools, Travels and Local , Video players ,Weather)

**(3.) Rating :-**

Rating value of the app (Ratings are numbered values given by the users after their experience or usage

Ratings are numbered values from 1-5 in stars format)

**(4.)Reviews :-**

Review count

The number of reviews provided by the users after their mind-blowing performance and experience

***(5.) Installs :-***

Number of verified installs (in this option it contains the number of verified customers installed this app and used it for a while)

***(6.) Size :-***

Size of the app (coming to the point of size it is the memory space that is been consumed by the mobile phone while installing a particular app and the way It consumes the storage or memory in the disk or mobile storage)

***(7.) Price :-***

Price of the app in $ dollars(coming to the point of pricing some apps can be downloaded freely and some are paid . the pricing list is totally dependent on the developers end)

***(8.) Content Rating :-***

Intended audience or age group targeted

***(9.) Last Updated :-***

Last Updated date (after bug fixing or fixing any issue and resolving it they will update those app)

***(10.) Minimum Version :-***

Minimum Android version required to run the app (particular android version required to run a particular aaplication . as some of the applications does not support the older versions due to issues like android level support )

***(11.) Latest Version :-***

Current version of the app

***Objectives/Scope of the Analysis*** :-

1. Analyzing the good app based on the following 6 objectives
2. Calculating the top 10 number ratings for a particular app from play store dataset
3. Calculating the Top ranking of applications based on content wise
4. Calculating top 10 Application based on Installs.
5. Number of apps that are most recently updated
6. top maximum sized app category wise
7. Calculating the average no of ratings for a particular app from play store dataset
8. Calculating the average one-off categories rank for a particular application

The scope of main point is to find the most suitable application for a particular category for people based on the given rating review

**Source of dataset**

I had Collected all resources for my project from Kaggle and data.gov.in

1. ***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.

1. ***Open Government Data (OGD)*** :- Open Government Data (OGD***)*** Platform India - data.gov.in - is a platform for supporting the Open Data initiative of the Government of India. The portal is intended to be used by the Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give

***Kaggle Link:-*** https://www.kaggle.com/mrwolfgang/google-playstore-eda



***ETL process:***

**ETL**stands for “extract, transform, and load.”

ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL full-form is Extract, Transform and Load.

1. ***Data Extraction:-***

In this step, I had removed all the duplicated values, false values, etc…

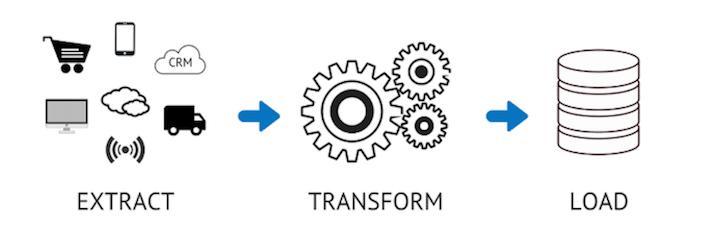
1. ***Data Transformation:-***

In this step, I had replaced null values with NAN and transformed percentage values to numerical values, calculated some row values which were not given using Tableau.

1. ***Data Load:-***

In this step, I had transformed data according to my project and I had calculated the required data and created pivot tables after data load.

The process of ETL plays a key role in data integration strategies. ETL allows businesses to gather data from multiple sources and consolidate it into a single, centralized location. ETL also makes it possible for different types of data to work together.



A typical ETL process collects and refines different types of data, then delivers the data to a data warehouse such as Redshift, Azure, or Big Query.

ETL also makes it possible to migrate data between a variety of sources, destinations, and analysis tools. As a result, the ETL process plays a critical role in producing business intelligence and executing broader data management strategies.

Three steps make up the ETL process and enable data to be integrated from source to destination. These are data extraction, data transformation, and data loading.

Graphical user interface, website

Description automatically generated

A picture containing graphical user interface

Description automatically generatedGraphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

**Analysis on dataset:**

1.This analysis helps in finding out the best application based on the reviews. Reviews describe the application features and the user views to that application. This pivot chart represents the top reviews of the applications based on the category.

1. **Introduction**

This objective is about the basic information of the application like below.

1. Application name,
2. Category
3. Chart
4. Review
5. Ratings
6. Feedback

Graphical user interface, application

Description automatically generated

**General Description**

1. Nowadays due to hike in the technology many of them are creating specialized application mainly developers are building apps as per the public need and many of the people are getting confused about the better and suitable app . so to overcome this issue I have visualized the following data set

**Graphical user interface, application, PowerPoint

Description automatically generated**

***Visualzation-1:***

**This analysis helps in finding out the best application based on the reviews. Reviews describe the application features and the user views to that application. This pivot chart represents the top reviews of the applications based on the category.**

1. **Introduction:**

This objective is about finding best application based on the best application based on the reviews.

1. **General Description**

As we use many applications in our day to day life many customers review the application after experience .this helps us a lot in analysing the best application.

1. **Specific Requirements, functions and formulas:**

Coming to the point of specific requirements we need the requirement of reviews of lot of people which helps us a lot in analysing the data and analysing the best application.

1. **Analysis resultsGraphical user interface, application

   Description automatically generated**

**Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generatedGraphical user interface, application, PowerPoint

Description automatically generated**

**Visualization**

Figure 9: Pivot Chart CO

**Graphical user interface, application, Excel, PowerPoint

Description automatically generated**

**Graphical user interface, application

Description automatically generated**

***Visualization-2:-***

This analysis helps in finding the number of applications available based on the category of the type (free or paid) for the users based on their age.

1. ***Introduction:-*** This objective is about the number of applications available based on the category of the type

Whether paid or for free based on their age group

1. **General Description :-**

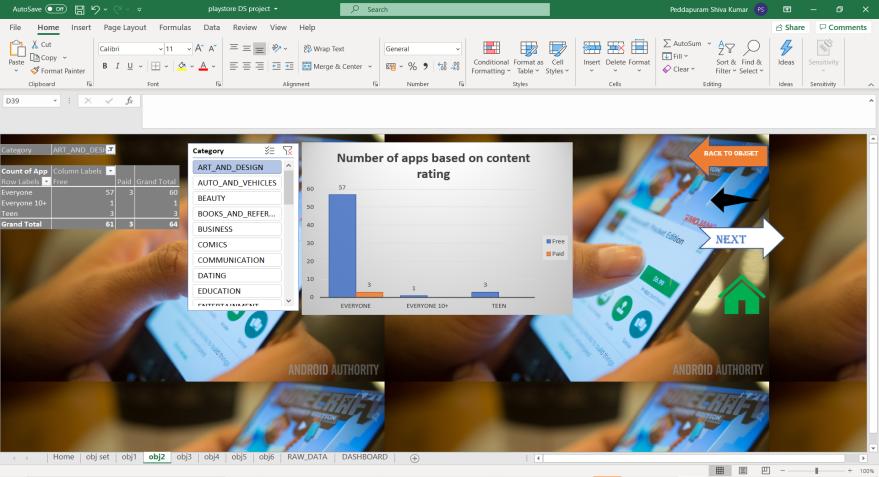
Many of the users shows interest for free services but when we go through the following visualization, we get an overall view about the type of users

1. ***Specific Requirements, functions and formulas :-***

Coming to the point of specific requirements functions and formula the main specific requirement I have used here the type of category for a particular application whether it is paid or for freely available in play store. and second most important thing I have compared is the age group the customers or audience specifically belongs to

1. ***Analysis results***

This pivot chart represents number of applications based on the content rating.

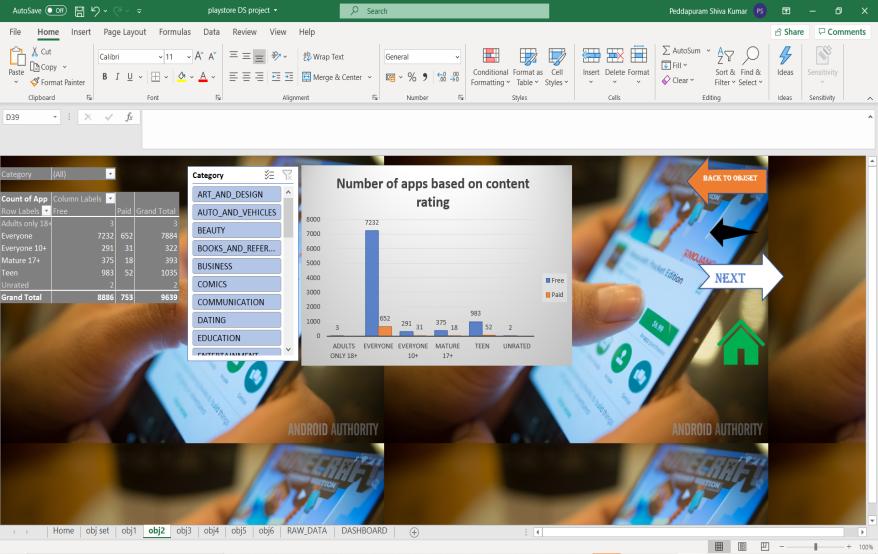


A picture containing text, screenshot, indoor

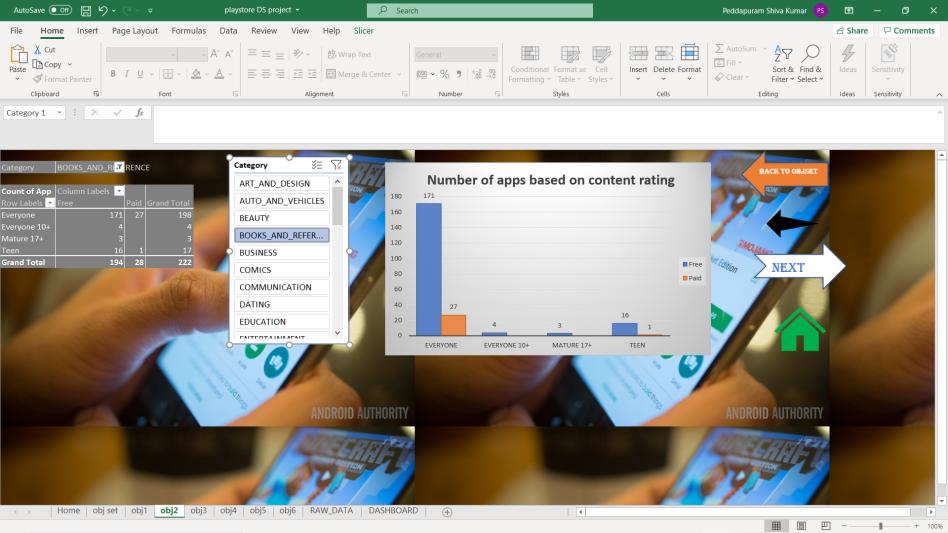
Description automatically generatedA picture containing text, indoor, screenshot

Description automatically generatedA picture containing text, screenshot, indoor

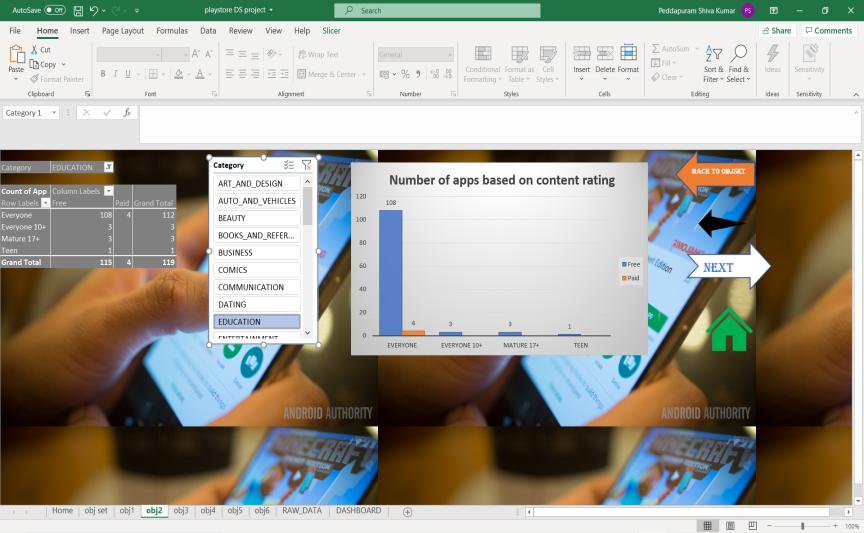
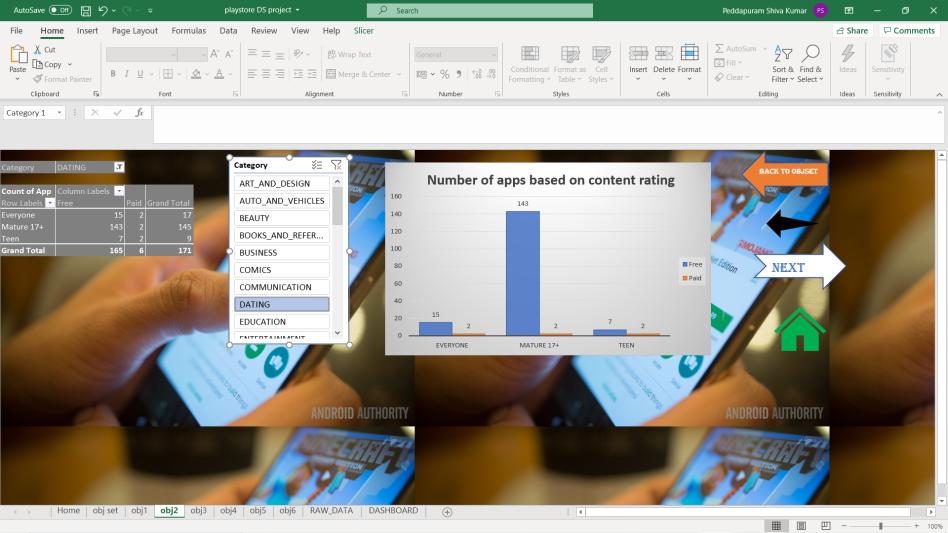
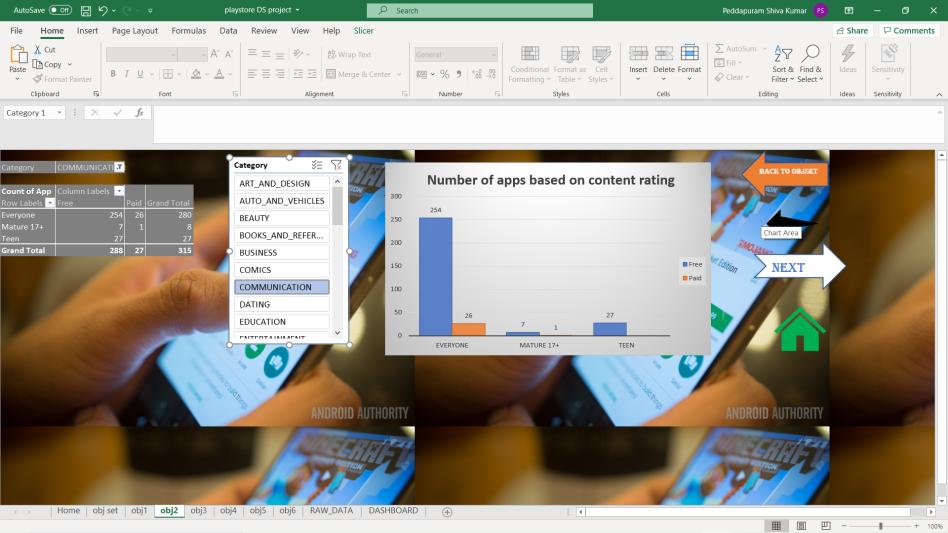
Description automatically generated A picture containing text, indoor, screenshot

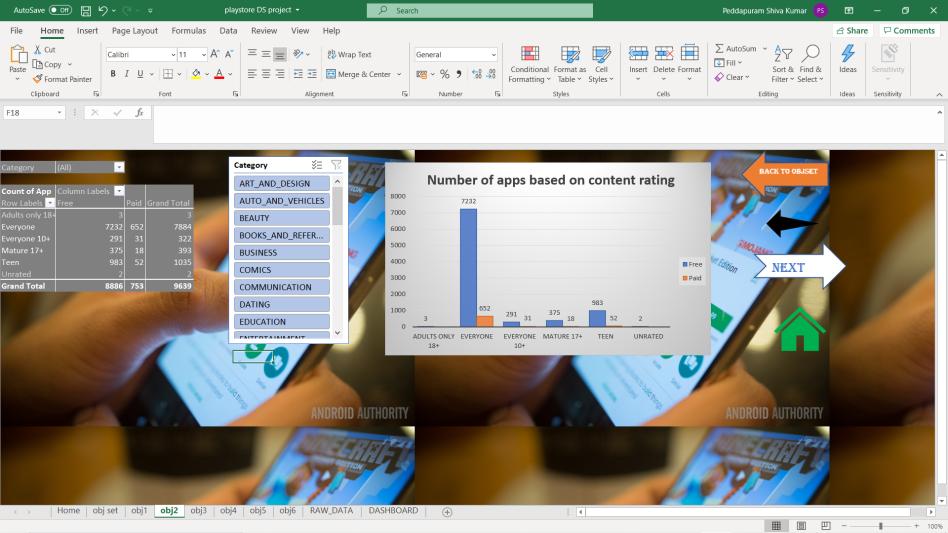
Description automatically generated Graphical user interface, application, PowerPoint

Description automatically generatedA picture containing text, screenshot, indoor

Description automatically generatedGraphical user interface, application, PowerPoint

Description automatically generatedGraphical user interface, application, PowerPoint

Description automatically generated



***Total overall Visualization:-***

Graphical user interface, application, PowerPoint

Description automatically generated

***Visualization-3:-***  **Top RANKING of apps based on INSTALLS**

1. **Introduction**

This objective is about “ **Top RANKING of apps based on INSTALLS ”**

***General Description :-***

In these objective we can get the information about the ranking of top 10 aap based on installs.

Basically there are some kind of misinformation like users downloads apps and then just uninstall them

1. **Specific Requirements, functions and formulas**

Basically we deal about the top 10 installs for some websites based on their category and finalize the better ones as per the requirements

1. **Analysis results**

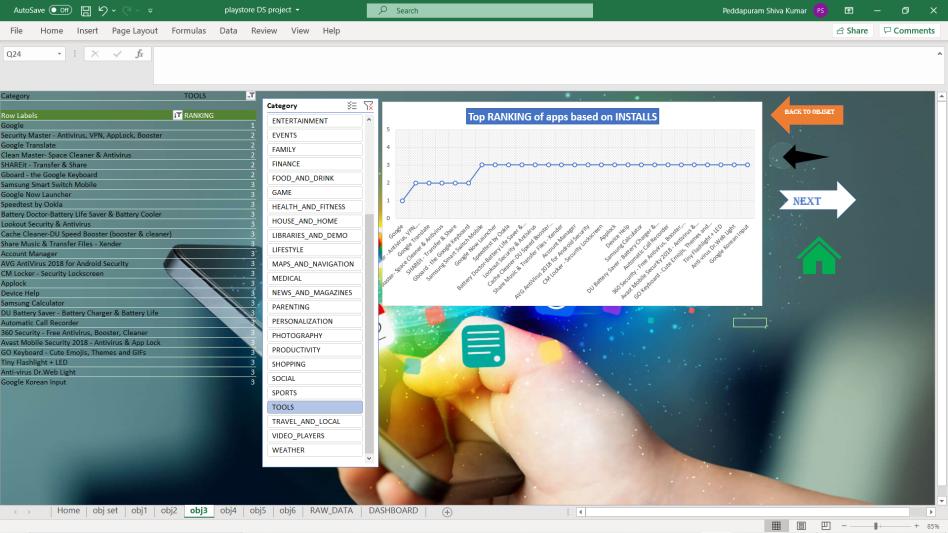
**Graphical user interface, application

Description automatically generated**

**Graphical user interface, application, table, Excel, PowerPoint

Description automatically generatedGraphical user interface, application, table, Excel

Description automatically generated**

**Graphical user interface, text, application, website

Description automatically generated**

**Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated**

***Visualization-4:-* *number of apps based on rating***

1. ***Introduction:-***

This objective is about the **number of apps based on rating**.

1. **General Description**

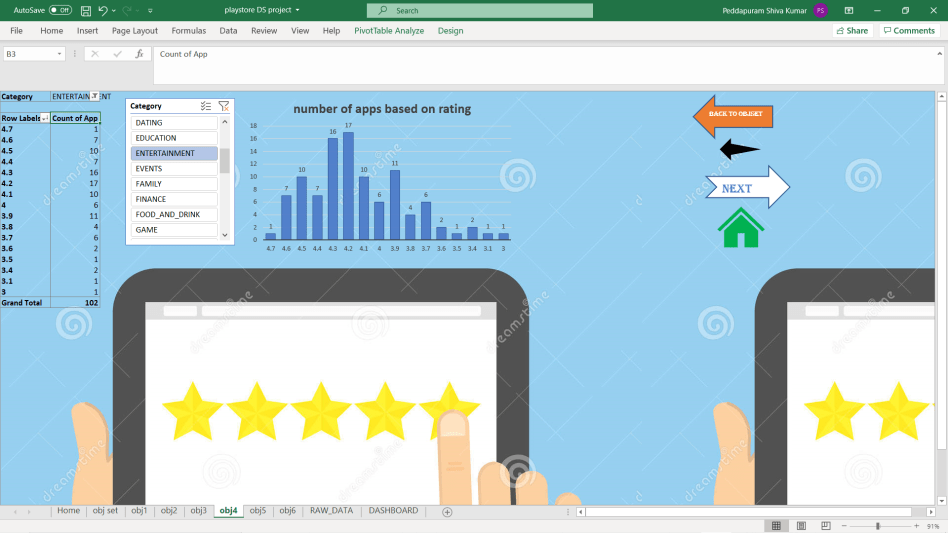
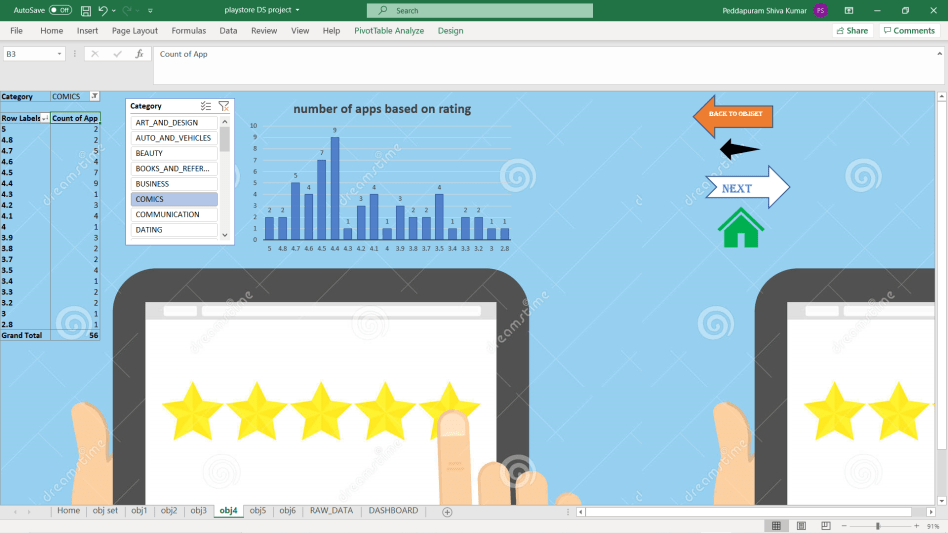
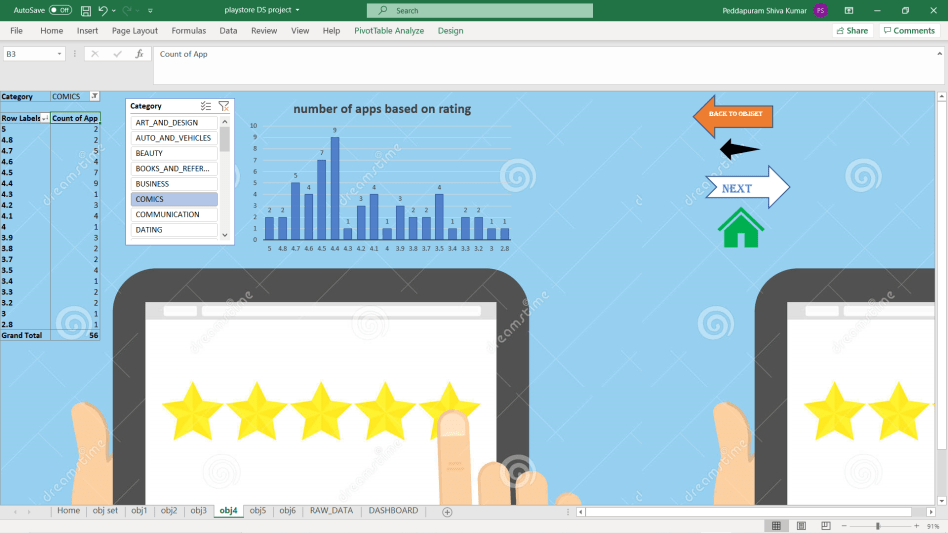
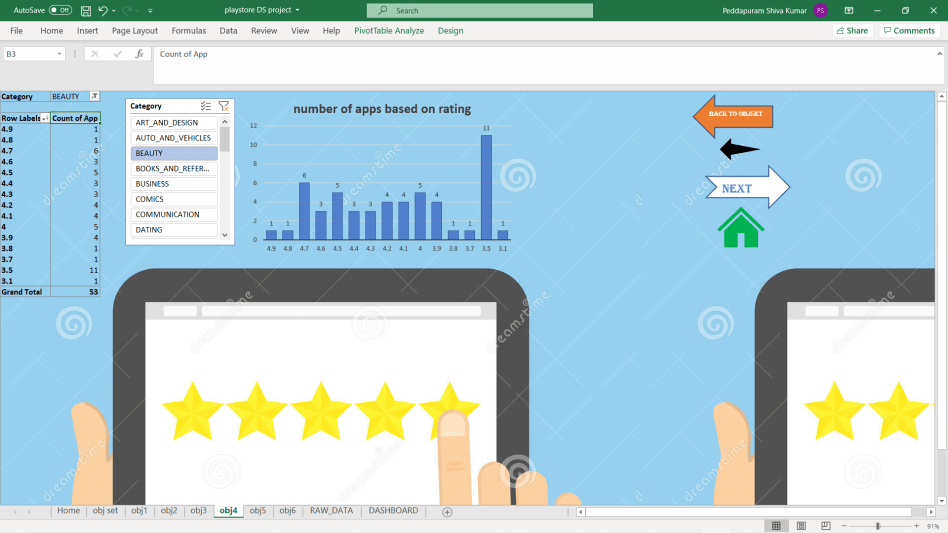
Rating is the main requirement for the application development and promotion and it helps lot of the customers.

1. **Specific Requirements, functions and formulas**

In this objective I have analysed the number of apps based on rating

1. **Analysis resultsGraphical user interface, application

   Description automatically generated**

**Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated**

1. **Visualization :-**

**Graphical user interface, application

Description automatically generated**

***Visualization-5:-***

1. ***Introduction:-***

This objective is about the **Number of apps that are most recently updated**

1. **General Description**

As customers uses the application many of them gets many issues in order to resolve those issues lot of updates are done. As the application gets updated it gives lot of best experience to the customers. Which further helps the customers to feel happy and satisfied for the usage

1. **Specific Requirements, functions and formulas**

In this objective I have analysed the number of apps that was recently updated.

**Analysis results :-**

**Graphical user interface

Description automatically generated**

**Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface, website

Description automatically generatedGraphical user interface, website

Description automatically generated**

**Visualization-6:-**

1. ***Introduction:-***

This objective is about the “**top maximum sized app category wise “**

1. **General Description**

In these analysis I have analysed about the maximum size of the application as per category wise

As the size of the application increases many of the customers shows less interest to download it.

1. **Specific Requirements, functions and formulas**

In this objective I have analysed about the application which has maximum sized as per category wise.

***Analysis :***

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface

Description automatically generated

**List of Analysis with results**

1. Analyzing the good app based on the following 6 objectives
2. Calculating the top 10 number ratings for a particular app from play store dataset
3. Calculating the Top ranking of applications based on content wise
4. Calculating top 10 Application based on Installs.
5. Number of apps that are most recently updated
6. top maximum sized app category wise
7. Calculating the average no of ratings for a particular app from play store dataset
8. Calculating the average one-off categories rank for a particular application

In the dashboard the combination of the above objectives is also provided to compare and with the slicers and timelines provided we can download the particular app as per our requirements

***Reference/Bibliography :***

1. ***https://www.kaggle.com/kracekumar/indian-elections***
2. https://en.wikipedia.org/wiki/Elections\_in\_India ′
3. https://data.gov.in/search/site?query=elections ′
4. https://eci.gov.in/