DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2019)

on

ANALYSIS OF
“GOOGLE PLAY STORE
APPS”

Submitted by: Pritam Dhoke

Registration No.: 11709119

Programme and Section: KM066

Course Code: INT217

Under the Guidance of

Shaina Gupta

School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Pritam Dhoke** bearing Registration no. **11709119** has completed **Data Management (INT217)** project titled, **“Analysis of Google play store 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 and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 20-11-2019

DECLARATION

I, **Pritam Dhoke**, 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: 20-11-2019

Registration No.: 11709119

Signature

Name: Pritam Dhoke

ACKNOWLEDGEMENT

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.

Thanks & Regards

TABLE OF CONTENT

1. Introduction

Android is the dominant mobile operating system today with about 85% of all mobile devices running Google’s OS. The Google Play Store is the largest and most popular Android app store.

The purpose of this project is to gather and analyse detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.

Some key observations at first glance include how the performance of the App can be improved from the reviews obtained and different patterns that could be found to get more business values out of the same.

DASHBOARD:

Dashboards are basically created to get a quick glance about how effectively a company is achieving its business objectives. It keeps a record about KPI, metrics and another data keywords in a visual and central place. It enhances the real time performance and tells about the current status and simplifies the data which is complex. Main feature provided by the dashboard is that collects the important data at a single place that increases the decision-making speed and keeps everyone up-to-date.

Dashboard could be made using tables, charts, gauge and numbers. They can be used almost everywhere where data play crucial role, including company, industry etc. We can make dashboard of following types:

Project dashboard, financial dashboard, marketing dashboard and many more.

Process for creating dashboard:

Before creating dashboard, we need to research, question and things to consider.

1) First thing we need to do is to think the need for dashboard, what will be the served purpose of dashboard, from where we will gather the data, what all the capabilities we need and we don’t need. It will help to create the replica of the model in excel on a piece of paper. Draw different boxes for different data types to get a rough layout and sketch graphs you want to include in your dashboard. It will get all your things on the single page and you could get approval from stakeholders before investing money and time on a project.

2) **Questions to ask yourself:**

i) Why are you creating the dashboard? - This question will answer the prime reason to create a dashboard; like is it to show the status of the project, or to show the increasing performance or growth. Letting us know the reason of creating dashboard will guide us the design and data.

ii) Who needs to see the dashboard? - We need to have a clear knowledge about who are the observers, to whom we have showcase our dashboard is it: colleagues, stakeholder, manager or external vendor. We have to think about how much time they will take to digest the information, how they will prefer to digest the information. So, we have to keep in mind about the preferences of the one for whom we have to create the dashboard.

iii) Where will the data come from? - This includes the source from where you will get the data. We could enter the data manually or we can import it into our dashboard. It also includes what re the tools we are using to gather the data.

iv) How up-to-date your dashboard is? - This will answer about the dashboard if it can be updated monthly or weekly or show the real time data.

v) What format does the dashboard need to be in? - Whether your dashboard is static, or you are providing a dynamic link of your dashboard. Do you let the dashboard in read-only mode or provide editing capabilities to some of the users? This includes all these things.

3) **Things to Consider: How to Design the Dashboard**

• Dashboard elements: It refers to all the things included in a dashboard, we can choose from static tables, pivot tables, dynamic charts or from non-charting objects. We can include different small charts or one big chart. This will help you to gather all the similar data together and get a rough layout of the dashboard.

• Dashboard background color: In this you have to choose whether you want to add a color in background of dashboard or not. Whether we want to create charts of same color or of different color.

• Dashboard user- interface: It includes the ease of access. We could add hierarchy to the layout for easy navigation, drop- down menu or add labels to the charts.

2. Scope of Analysis

1. Size of Applications greater than 10mb having rating more than 4.5.

2. No. of Paid Applications in each Category and No. of Teens use paid Application.

3. Most popular category, by number of installs.

4. What is the percentage of paid and free apps in Play Store?

5. Rating Vs Review count. (Slicer, Bar Graph)

6. If an app has high number of installs, does it mean that it gets the more reviews from the users?

When it comes to getting the word out for your Android app's superiority in terms of performance, there are a number of options that are given to you by Google, besides getting featured on Google Play.

Popular

- Top Free: Most popular free apps of all-time
- Top Paid: Most popular paid apps of all-time
- Top Grossing: Apps and games that generate the most revenue, including app purchases and in-app payments
- Trending Apps: Apps showing installation growth in the last 24 hours

New

To determine new apps that google play features and are less than 30 days old, Top New lists use the date of first availability; this includes the first time an app was published, or the first time an app was made available in a location.

The time period when an app is exclusively in Alpha or Beta (without a Production APK) has no effect on its first availability date or its ability to be on a "Top New" list.

- Top New Free: Most popular free apps less than 30 days old
- Top New Paid: Most popular paid apps less than 30 days old

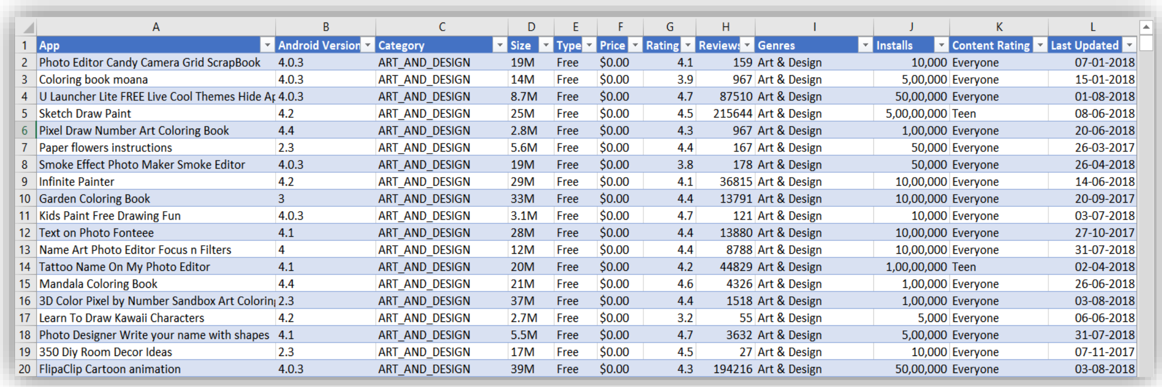
3. Source of Dataset

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. The dataset is chosen from Kaggle. It is the web scraped data of 10k Play Store apps for analysing the Android market. It consists of in total of *10841* rows and 13 columns.

At the time of the data collection, the Google Play Store broke apps down into 41 general categories. Education apps were the most common individual category, comprising 8% of the total number of apps available for download.

Dataset includes the following Columns:

- **App:** Application name
- **Category:** Category the app belongs to
- **Rating:** Overall user rating of the app (as when scraped)
- **Reviews:** Number of user reviews for the app (as when scraped)
- **Size:** Size of the app (as when scraped)
- **Installs:** Number of user downloads/installs for the app (as when scraped)
- **Type:** Paid or Free
- **Price:** Price of the app (as when scraped)
- **Content Rating:** Age group the app is targeted at - Children / Mature 21+ / Adult
- **Genres:** An app can belong to multiple genres (apart from its main category). For e.g. a musical family game will belong to Music, Game, Family genres.
- **Last Updated:** Date when the app was last updated on Play Store (as when scraped)
- **Current Ver:** Current version of the app available on Play Store (as when scraped)
- **Android Ver:** Min required Android version (as when scraped)



	A	B	C	D	E	F	G	H	I	J	K	L
1	App	Android Version	Category	Size	Type	Price	Rating	Reviews	Genres	Installs	Content Rating	Last Updated
2	Photo Editor Candy Camera Grid ScrapBook	4.0.3	ART_AND_DESIGN	19M	Free	$0.00	4.1	159	Art & Design	10,000	Everyone	07-01-2018
3	Coloring book moana	4.0.3	ART_AND_DESIGN	14M	Free	$0.00	3.9	967	Art & Design	5,00,000	Everyone	15-01-2018
4	U Launcher Lite FREE Live Cool Themes Hide Ap	4.0.3	ART_AND_DESIGN	8.7M	Free	$0.00	4.7	87510	Art & Design	50,00,000	Everyone	01-08-2018
5	Sketch Draw Paint	4.2	ART_AND_DESIGN	25M	Free	$0.00	4.5	215644	Art & Design	5,00,00,000	Teen	08-06-2018
6	Pixel Draw Number Art Coloring Book	4.4	ART_AND_DESIGN	2.8M	Free	$0.00	4.3	967	Art & Design	1,00,000	Everyone	20-06-2018
7	Paper flowers instructions	2.3	ART_AND_DESIGN	5.6M	Free	$0.00	4.4	167	Art & Design	50,000	Everyone	26-03-2017
8	Smoke Effect Photo Maker Smoke Editor	4.0.3	ART_AND_DESIGN	19M	Free	$0.00	3.8	178	Art & Design	50,000	Everyone	26-04-2018
9	Infinite Painter	4.2	ART_AND_DESIGN	29M	Free	$0.00	4.1	36815	Art & Design	10,00,000	Everyone	14-06-2018
10	Garden Coloring Book	3	ART_AND_DESIGN	33M	Free	$0.00	4.4	13791	Art & Design	10,00,000	Everyone	20-09-2017
11	Kids Paint Free Drawing Fun	4.0.3	ART_AND_DESIGN	3.1M	Free	$0.00	4.7	121	Art & Design	10,000	Everyone	03-07-2018
12	Text on Photo Fonteee	4.1	ART_AND_DESIGN	28M	Free	$0.00	4.4	13880	Art & Design	10,00,000	Everyone	27-10-2017
13	Name Art Photo Editor Focus n Filters	4	ART_AND_DESIGN	12M	Free	$0.00	4.4	8788	Art & Design	10,00,000	Everyone	31-07-2018
14	Tattoo Name On My Photo Editor	4.1	ART_AND_DESIGN	20M	Free	$0.00	4.2	44829	Art & Design	1,00,00,000	Teen	02-04-2018
15	Mandala Coloring Book	4.4	ART_AND_DESIGN	21M	Free	$0.00	4.6	4326	Art & Design	1,00,000	Everyone	26-06-2018
16	3D Color Pixel by Number Sandbox Art Colorin	2.3	ART_AND_DESIGN	37M	Free	$0.00	4.4	1518	Art & Design	1,00,000	Everyone	03-08-2018
17	Learn To Draw Kawaii Characters	4.2	ART_AND_DESIGN	2.7M	Free	$0.00	3.2	55	Art & Design	5,000	Everyone	06-06-2018
18	Photo Designer Write your name with shapes	4.1	ART_AND_DESIGN	5.5M	Free	$0.00	4.7	3632	Art & Design	5,00,000	Everyone	31-07-2018
19	350 Diy Room Decor Ideas	2.3	ART_AND_DESIGN	17M	Free	$0.00	4.5	27	Art & Design	10,000	Everyone	07-11-2017
20	FlipaClip Cartoon animation	4.0.3	ART_AND_DESIGN	39M	Free	$0.00	4.3	194216	Art & Design	50,00,000	Everyone	03-08-2018

Figure 1: Dataset

The most critical thing from which patterns could be obtained is data. It may be a single review or a bundle of them. Whatever data comes in, could be used to draw value out of it. Data comes with unexpected values too, which should be handled before it affects the performance of trained models that predict the outcome.

Here is the first step to clean the data that will make the results “more” accurate.

By finding all unique values of each row the inappropriate values can be identified. Different methods can then be used for removing them or to change those values accordingly to use them to make predictions better.

4. ETL Process

Data cleaning is done using **Tableau Prep Builder** Software:

The process of extracting data from multiple source systems, transforming it to suit business needs, and loading it into a destination database is commonly called ETL, which stands for extraction, transformation, and loading. While ETL is usually explained as three distinct steps, this simplifies it too much as it is truly a broad process that requires a variety of actions.

The following tasks are the main actions that happen in the ETL process:

ETL is a 3-step process:

Step – 1: Extraction

Step – 2: Transformation

Step – 3: Loading

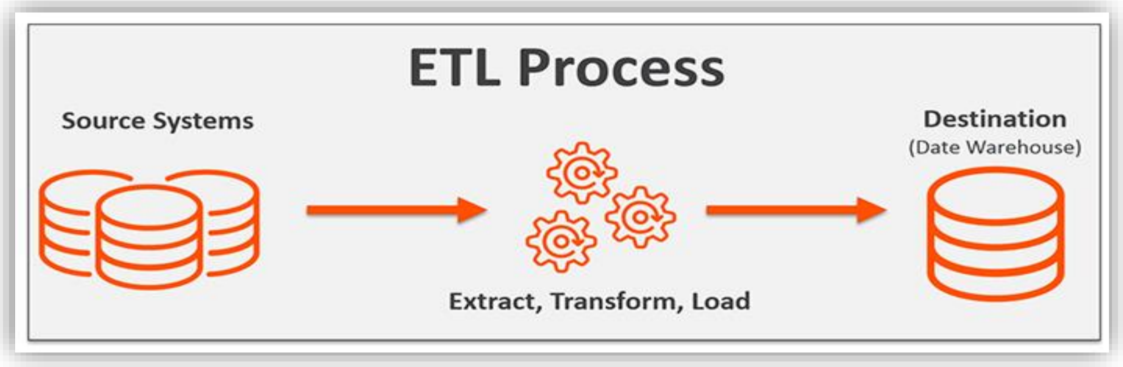


Figure 2: ETL Process

4.1. Extraction of Data

The first step in ETL is extraction. During extraction, data is specifically identified and then taken from many different locations, referred to as the Source. The Source can be a variety of things, such as files, spreadsheets, database tables, a pipe, etc. It is not typically possible to pinpoint the exact subset of interest, so more data than necessary is extracted to ensure it covers everything needed.

4.2. Transformation of Data

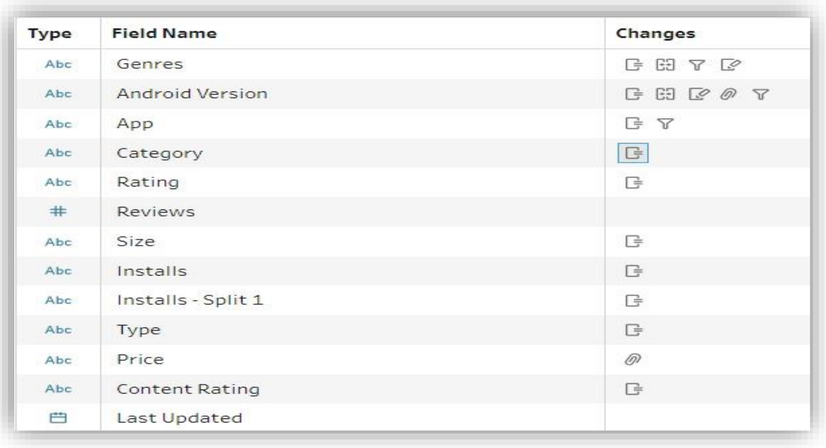
The next step in the ETL process is transformation. After data is extracted, it must be physically transported to the target destination and converted into the appropriate format. This data transformation may include operations such as cleaning, joining, and validating data or generating calculated data based on existing values.

Whether the transformation takes place in the data warehouse or beforehand, there are both common and advanced transformation types that prepare data for analysis. Some of these include:

- Basic transformations:
  - Cleaning
  - Format revision
  - Restructuring
  - Deduplication
- Advanced transformations:
  - Filtering
  - Joining
  - Splitting
  - Derivation
  - Summarization
  - Integration

4.3. Loading Data:

The final step in the ETL process involves loading the transformed data into the destination target. This target may be a database or a data warehouse. There are two primary methods for loading data into a warehouse: full load and incremental load. The full load method involves an entire data dump that occurs the first time the source is loaded into the warehouse. The incremental load, on the other hand, takes place at regular intervals. These intervals can be streaming increments (better for smaller data volumes) or batch increments (better for larger data volumes).



Type	Field Name	Changes
Abc	Genres	⇨ ⊟ ▽ ✎
Abc	Android Version	⇨ ⊟ ✎ ⌀ ▽
Abc	App	⇨ ▽
Abc	Category	⇨
Abc	Rating	⇨
#	Reviews	
Abc	Size	⇨
Abc	Installs	⇨
Abc	Installs - Split 1	⇨
Abc	Type	⇨
Abc	Price	⌀
Abc	Content Rating	⇨
📅	Last Updated	

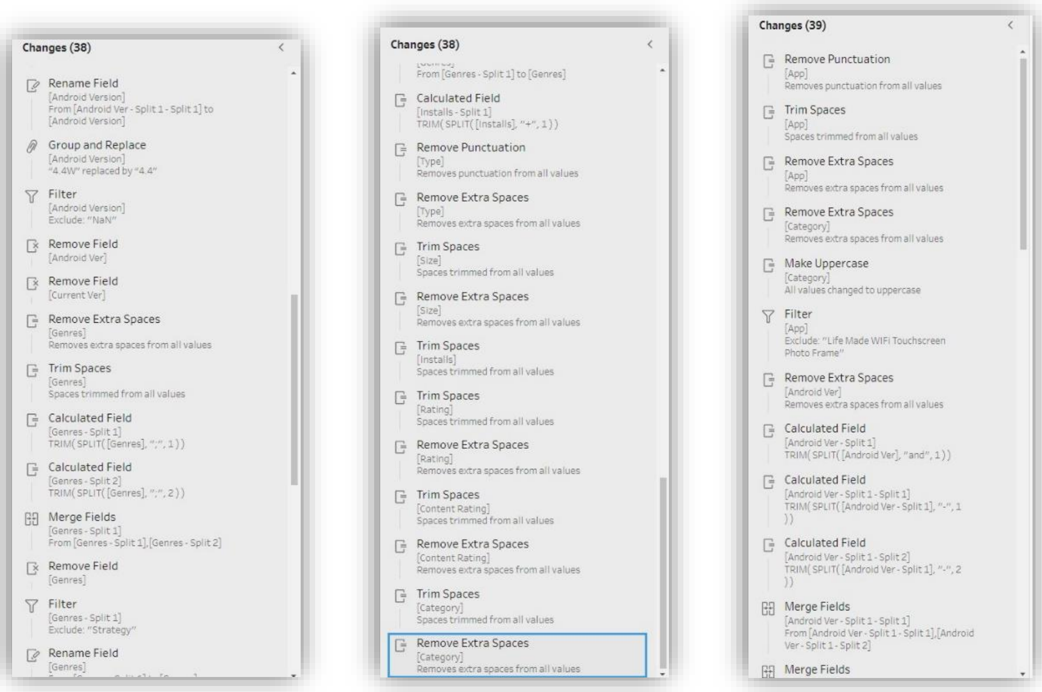
Figure 3: Overall Cleaning.

Converting our data into appropriate forms

1) Data Cleaning by removing NULL values, removing entities which is not of the same type in same column:
   a) Cleaning the App Column By removing punctuation.
   b) Trim Space, Remove Extra Spaces.
   c) Removed Null Values.
2) Changed the type of Rating to String type.
3) All values of Category changed to UPPERCASE.
4) **Android Version:** Split by '**and**' -> removed Extra spaces -> Split by '**-**' -> Group and Replaced "4.4W" by "4.4" -> Remove Field Android Split-1 and Created New Android Version -> Removed old Android Ver.
5) **Genres:** Grouped and Replace -> Trim Spaces -> Split by '**;**' and Created New Calculated Field -> Merged old Genres into New Genres
6) Split install by removing '+' and create New Calculation Field.
7) **Size:** For example, the size of the app is in “string” format. We need to convert it into a numeric value. If the size is “10M”, then ‘M’ was removed to get the numeric value of ‘10’. If the size is “512k”, which depicts app size in kilobytes, the first ‘k’ should be removed, and the size should be converted to an equivalent of ‘megabytes’.
8) **Installs:** The value of installs is in “string” format. It contains numeric values with commas. It should be removed. And also, the ‘+’ sign should be removed from the end of each string.

9) **Category and Content Rating:** The Category and Content Rating consists of categorical values that should be converted to numeric values if we need to perform regression. So, these were converted to numeric values.

10) **Price:** The price is in “string” format. We should remove the dollar sign (**$**) from the string to convert it into numeric form.



Changes (38)

Rename Field
[Android Version]
From [Android Ver - Split 1 - Split 1] to [Android Version]

Group and Replace
[Android Version]
"4.4W" replaced by "4.4"

Filter
[Android Version]
Exclude: "NaN"

Remove Field
[Android Ver]

Remove Field
[Current Ver]

Remove Extra Spaces
[Genres]
Removes extra spaces from all values

Trim Spaces
[Genres]
Spaces trimmed from all values

Calculated Field
[Genres - Split 1]
TRIM(SPLIT([Genres], ";", 1))

Calculated Field
[Genres - Split 2]
TRIM(SPLIT([Genres], ";", 2))

Merge Fields
[Genres - Split 1]
From [Genres - Split 1],[Genres - Split 2]

Remove Field
[Genres]

Filter
[Genres - Split 1]
Exclude: "Strategy"

Rename Field
[Genres]

Changes (38)

From [Genres - Split 1] to [Genres]

Calculated Field
[Installs - Split 1]
TRIM(SPLIT([Installs], "+", 1))

Remove Punctuation
[Type]
Removes punctuation from all values

Remove Extra Spaces
[Type]
Removes extra spaces from all values

Trim Spaces
[Size]
Spaces trimmed from all values

Remove Extra Spaces
[Size]
Removes extra spaces from all values

Trim Spaces
[Installs]
Spaces trimmed from all values

Trim Spaces
[Rating]
Spaces trimmed from all values

Remove Extra Spaces
[Rating]
Removes extra spaces from all values

Trim Spaces
[Content Rating]
Spaces trimmed from all values

Remove Extra Spaces
[Content Rating]
Removes extra spaces from all values

Trim Spaces
[Category]
Spaces trimmed from all values

Remove Extra Spaces
[Category]
Removes extra spaces from all values

Changes (39)

Remove Punctuation
[App]
Removes punctuation from all values

Trim Spaces
[App]
Spaces trimmed from all values

Remove Extra Spaces
[App]
Removes extra spaces from all values

Remove Extra Spaces
[Category]
Removes extra spaces from all values

Make Uppercase
[Category]
All values changed to uppercase

Filter
[App]
Exclude: "Life Made WIFi Touchscreen Photo Frame"

Remove Extra Spaces
[Android Ver]
Removes extra spaces from all values

Calculated Field
[Android Ver - Split 1]
TRIM(SPLIT([Android Ver], "and", 1))

Calculated Field
[Android Ver - Split 1 - Split 1]
TRIM(SPLIT([Android Ver - Split 1], "-", 1))

Calculated Field
[Android Ver - Split 1 - Split 2]
TRIM(SPLIT([Android Ver - Split 1], "-", 2))

Merge Fields
[Android Ver - Split 1 - Split 1]
From [Android Ver - Split 1 - Split 1],[Android Ver - Split 1 - Split 2]

Merge Fields

Figure 4: Changes Made in Tableau

5. Analysis of Dataset

5.1. No. of Apps Greater than Rating>=4.5 According to Size.

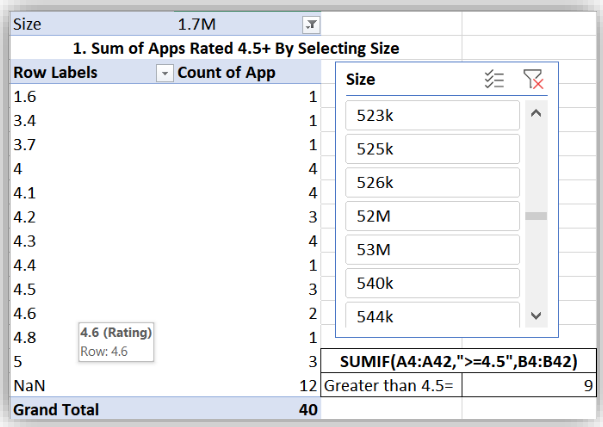
The most obvious reason, why your app’s rating in mobile stores matter is that a great rating and a bunch of positive reviews impresses potential users. And to be honest, who doesn’t want to be honored for great work? In fact, app store reviews and ratings can give your app a competitive edge.

Description:

From the dataset of google play store, we can make the list of apps greater than 4.5 Rating or count of apps accordingly.

Formulas:

=SUMIF(Calculation!A4:A42,">=4.5",Calculation!B4:B42)



Size 1.7M

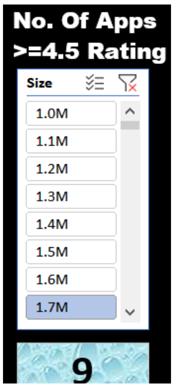
1. Sum of Apps Rated 4.5+ By Selecting Size

Row Labels	Count of App
1.6	1
3.4	1
3.7	1
4	4
4.1	4
4.2	3
4.3	4
4.4	1
4.5	3
4.6	2
4.8	1
5	3
NaN	12
Grand Total	40

SUMIF(A4:A42,">=4.5",B4:B42)	
Greater than 4.5=	9

Figure 5: Calculation-1

Visualization:



No. Of Apps >=4.5 Rating

9

Figure 6: Visualization-1

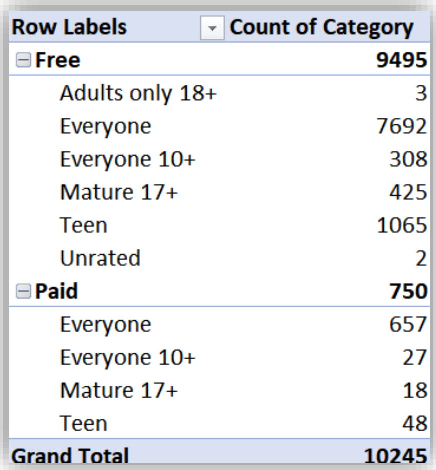
- ✓ In this screenshot the number 9 into box below the slicer shows that count of apps having size greater or equal to 1.7M **and** Rating more than 4.5.
- ✓ Dashboard user can also select multiple entities to get the count accordingly.

5.2. No. of teens uses Paid or Free app:

Free Apps generally receive more downloads than paid Apps. Leading on from this, free Apps make it easier to get tons more users. With free Apps, users tend to have lower expectations. Let's say they download your App, use it for a little while and but it isn't helping them in their everyday life.

Description:

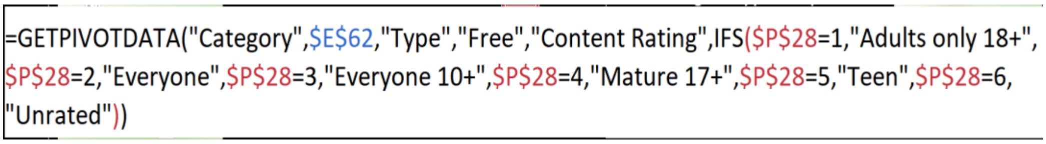
There are many users using free apps for their daily needs such as Fitness, music, Arts, etc. and some are organisations apps only used by specific users who works within organisation.



Row Labels	Count of Category
⊟Free	**9495**
Adults only 18+	3
Everyone	7692
Everyone 10+	308
Mature 17+	425
Teen	1065
Unrated	2
⊟Paid	**750**
Everyone	657
Everyone 10+	27
Mature 17+	18
Teen	48
Grand Total	**10245**

Figure 7: Calculation-2

Formula:



```
=GETPIVOTDATA("Category",$E$62,"Type","Free","Content Rating",IFS($P$28=1,"Adults only 18+",
$P$28=2,"Everyone",$P$28=3,"Everyone 10+",$P$28=4,"Mature 17+",$P$28=5,"Teen",$P$28=6,
"Unrated"))
```

Visualization:

This shows 1065 Apps are used by Category "Teen" User.

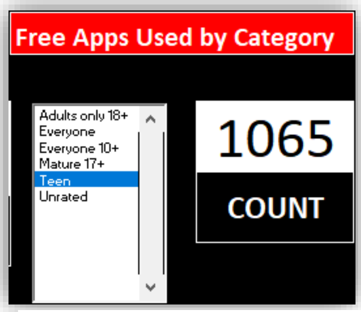


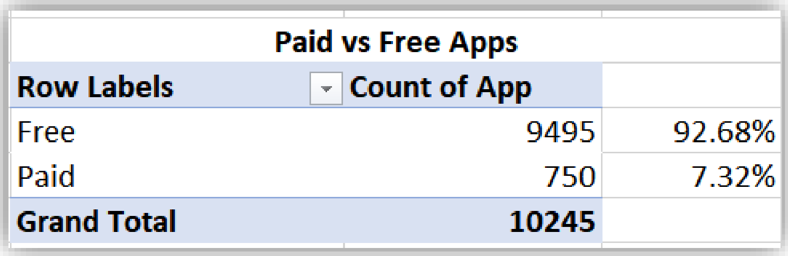
Figure 8: Free Apps Used by category

5.3. Percentage of Free and Paid Apps:

This statistic shows the distribution of free and paid apps in Google Play as of October 2018. As of the measured period, 92.7 percent of apps in the Google Play were freely available.

Description:

This Pivot chart is Easily understandable to draw some conclusions.

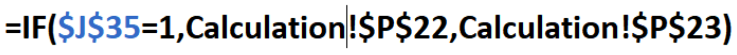


Paid vs Free Apps

Row Labels	Count of App	
Free	9495	92.68%
Paid	750	7.32%
Grand Total	**10245**	

Figure 9: Calculation-3

Formula:

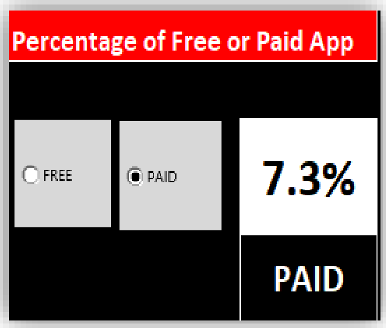


```
=IF($J$35=1,Calculation!$P$22,Calculation!$P$23)
```

Visualization:

By calculating the percentage of the count of the apps we can get the following result.

All the Field in this Visualization is dynamic in nature.



Percentage of Free or Paid App

○ FREE

● PAID

7.3%

PAID

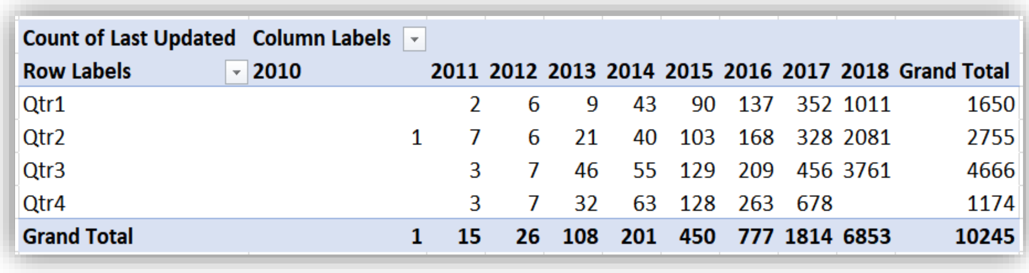
Figure 10: Percentage of Free/Paid Apps

5.4. **Percentage of Apps Last Update:**

Update time can be divided into four quarter and also columns can be the years hence the pivot table can easily be converted into pivot chart.

Description:

Let’s now analyse when applications were last updated by making a time series chart.



Count of Last Updated	Column Labels									
Row Labels	2010	2011	2012	2013	2014	2015	2016	2017	2018	Grand Total
Qtr1		2	6	9	43	90	137	352	1011	1650
Qtr2	1	7	6	21	40	103	168	328	2081	2755
Qtr3		3	7	46	55	129	209	456	3761	4666
Qtr4		3	7	32	63	128	263	678		1174
Grand Total	**1**	**15**	**26**	**108**	**201**	**450**	**777**	**1814**	**6853**	**10245**

Figure 11: Calculation-4

Visualization:

Most applications have been updated within the last 6 months but there are applications that haven’t not seen an update for 5 years!

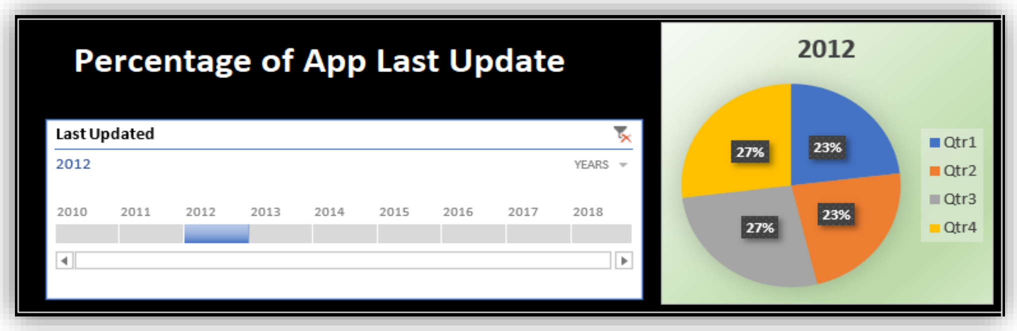


Figure 12: Visualization-2

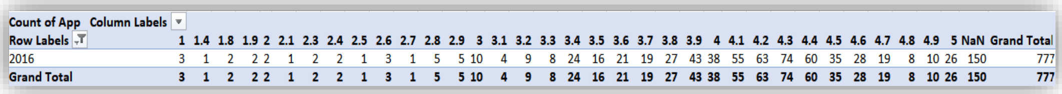
5.5. Distribution of User Ratings by Years:

Ratings matters for all the app provider in order to gain market interest in their apps. The dataset is of 2010 to 2018 that means 8 years.

Description:

Slicers used to change the year for which user want to get the bar chart. For example, if 2016 is selected then the graph shows 74 apps are of 4.5 rating in year 2016.

Formula:



Count of App	Column Labels																
Row Labels	1	1.4	1.8	1.9	2	2.1	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3	3.1	3.2	3.3
2016	3	1	2	2	2	1	2	2	1	3	1	5	5	10	4	9	8
Grand Total	**3**	**1**	**2**	**2**	**2**	**1**	**2**	**2**	**1**	**3**	**1**	**5**	**5**	**10**	**4**	**9**	**8**

Row Labels	3.4	3.5	3.6	3.7	3.8	3.9	4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5	NaN	Grand Total
2016	24	16	21	19	27	43	38	55	63	74	60	35	28	19	8	10	26	150	777
Grand Total	**24**	**16**	**21**	**19**	**27**	**43**	**38**	**55**	**63**	**74**	**60**	**35**	**28**	**19**	**8**	**10**	**26**	**150**	**777**

Figure 13: Calculation-5

Visualization:

It also shows that Null mean 150 apps in 2016 have no ratings.

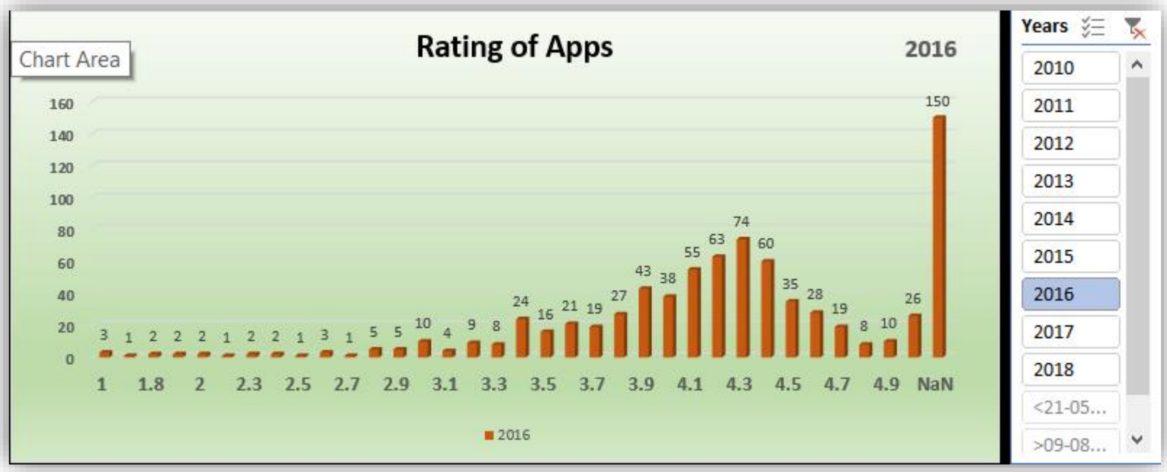


Figure 14: Visualization-3

6. Result of Analysis:

By the end of this project I get to know that free apps have more installs.

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market!

- The **Dataset** has data about more than 10,000 **apps** in the Store.
- Important **Conclusions** from the Analysis are: Maximum Number of **Apps** in the Store are from the "Family' and 'Game' Category.
- Most of the Apps hold a rating of above 4.0 easily.
- A total of 271 Apps have 5.0 Rating.
- The most famous Apps like WhatsApp, Facebook and Instagram are the most reviewed Apps.
- 93% of the Apps are free in the Play Store.
- The costliest App is 'I am Rich- Trump Edition', which is of $400!
- Apps related to Education, Lifestyle and Tools seem to fetch full Ratings.

The Final Dashboard of this Project “**Analysis on Google play Store Apps**”

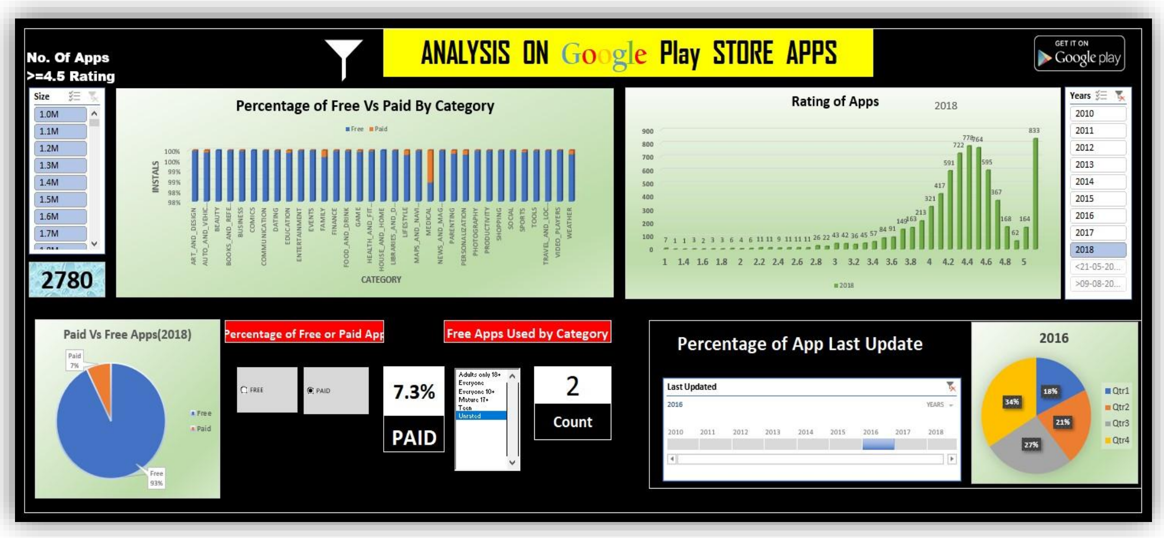


Figure 15: Dashboard

7. Future Scope

The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project. Many other interesting possibilities can be explored using this dataset.

Future work can include

- Optimization of the pie-charts shown above. There are multiple domains in the same slice. The multiple domains could be separated and added to the same field to get a more detailed version of this chart.
- Prediction of the number of reviews and installs by using the regression model.
- Identifying the categories and stats of the most installed apps.
- Exploring the correlation between the size of the app, the version of Android, etc on the number of installs.

The ways in which questions can be asked varies, so does the way of tackling a problem. Only the one that has been minutely observed and tested will provide results worth trusting.

8. Reference

- https://medium.com/the-research-nest/data-science-tutorial-analysis-of-the-google-play-store-dataset-c720330d4903
- https://www.pewresearch.org/internet/2015/11/10/an-analysis-of-apps-in-the-google-play-store/
- https://nycdatascience.com/blog/student-works/web-scraping/analysis-of-apps-in-the-google-play-store/

9. Bibliography

- https://www.kaggle.com/lava18/google-play-store-apps
- https://www.excel-easy.com
- https://www.youtube.com/results?search_query=how+to+make+dashboard+in+excel
- https://www.tableau.com/about/blog/2019/3/heres-4-powerful-tableau-prep-builder-features-you-should-be-using-for-messy-data