Technology in Business

Project Report

Topic: Google Play Analysis (using SQL and PowerBI)

Division B

Submitted by : Yusra Sayed

PRN : 23030141082

Problem Statement

The goal is to analyse the Google Play Store Apps dataset to derive insights into the app market. We aim to understand the factors that contribute to an app's success, including its user ratings, reviews, and category. We want to explore user sentiments towards apps by analysing the user reviews dataset. Additionally, we can see insights into the popularity of app categories based on the total number of installs and the sentiment polarity of user reviews. The ultimate objective is to provide recommendations for app developers to enhance their app's performance and user satisfaction.

Overview of dataset:

• Provide the overview of dataset by how many total unique apps and categories in our dataset.

Explore App Categories and Counts:

• Retrieve the unique app categories and the count of apps in each category.

Top-rated Free Apps:

• Identify the top-rated free apps.

Most Reviewed Apps:

• Find the apps with the highest number of reviews.

Average Rating by Category:

• Calculate the average rating for each app category.

Top Categories by Number of Installs:

• Identify the app categories with the highest total number of installs.

Average Sentiment Polarity by App Category:

• Analyse the average sentiment polarity of user reviews for each app category.

Sentiment reviews by App Category

• Provide the distribution of sentiments across different app categories

About the Dataset

Source of the dataset :

Dataset 1 – googleplaystore.csv

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Upd	Curre	Android Ver
Photo Editor &	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Des	7.Jan.18	1.0.0	4.0.3 and up
Coloring book	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Des	15.Jan.18	2.0.0	4.0.3 and up
U Launcher Lit	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000	Free	0	Everyone	Art & Des	1.Aug.18	1.2.04	4.0.3 and up
Sketch - Draw	ART_AND_DESIGN	4.5	215644	25M	50,000,00	Free	0	Teen	Art & Des	8.Jun.18	Varies	4.2 and up
Pixel Draw - Nu	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Des	20.Jun.18	1.1	4.4 and up
Paper flowers i	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Des	26.Mar.17	1	2.3 and up
Smoke Effect P	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Des	26.Apr.18	1.1	4.0.3 and up
Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000	Free	0	Everyone	Art & Des	14.Jun.18	6.1.61	4.2 and up
Garden Colorin	ART_AND_DESIGN	4.4	13791	33M	1,000,000	Free	0	Everyone	Art & Des	20.Sep.17	2.9.02	3.0 and up
Kids Paint Free	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Des	3.Jul.18	2.8	4.0.3 and up
Text on Photo	ART_AND_DESIGN	4.4	13880	28M	1,000,000	Free	0	Everyone	Art & Des	27.Oct.17	1.0.4	4.1 and up

App : name of the app

Category: category of the app based on its nature.

Rating: rating out of 5

Reviews: number of reviews received by the app

Size: download size of the app

Installs: total number of installations done by users

Type: (free or paid app)

Price: price of the app (if it is paid)

Content rating: app for audience

Genre : genre of the app

Last updated: last released update of the app

Current ver: current app version

Android ver: android version required for the app to be installed

Dataset 2 – googleplaystore_user_reviews.csv

App	Translate	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
10 Best Fc	I like eat del	Positive	1	0.533333333
10 Best Fc	This help eat	Positive	0.25	0.288461538
10 Best Fc	nan	nan	nan	nan
10 Best Fc	Works great	Positive	0.4	0.875
10 Best Fc	Best idea us	Positive	1	0.3
10 Best Fc	Best way	Positive	1	0.3
10 Best Fc	Amazing	Positive	0.6	0.9
10 Best Fc	nan	nan	nan	nan
10 Best Fc	Looking forw	Neutral	0	0
10 Best Fc	It helpful site	Neutral	0	0
10 Best Fc	good you.	Positive	0.7	0.6
10 Best Fc	Useful inforn	Positive	0.2	0.1
10 Best Fc	Thank you! C	Positive	0.75	0.875
10 Best Fc	Greatest eve	Positive	0.9921875	0.866666667
10 Best Fc	Good health.	Positive	0.55	0.511111111
10 Best Fc	nan	nan	nan	nan

App: name of the app

Translated review: User review (Pre-processed and translated to English)

Sentiment: Positive/Negative/Neutral (Pre-processed)

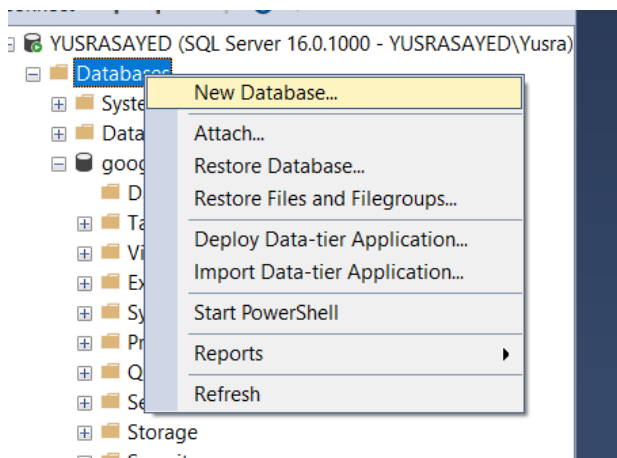**Sentiment Polarity:** Sentiment polarity score

Sentiment Subjectivity: Sentiment subjectivity score
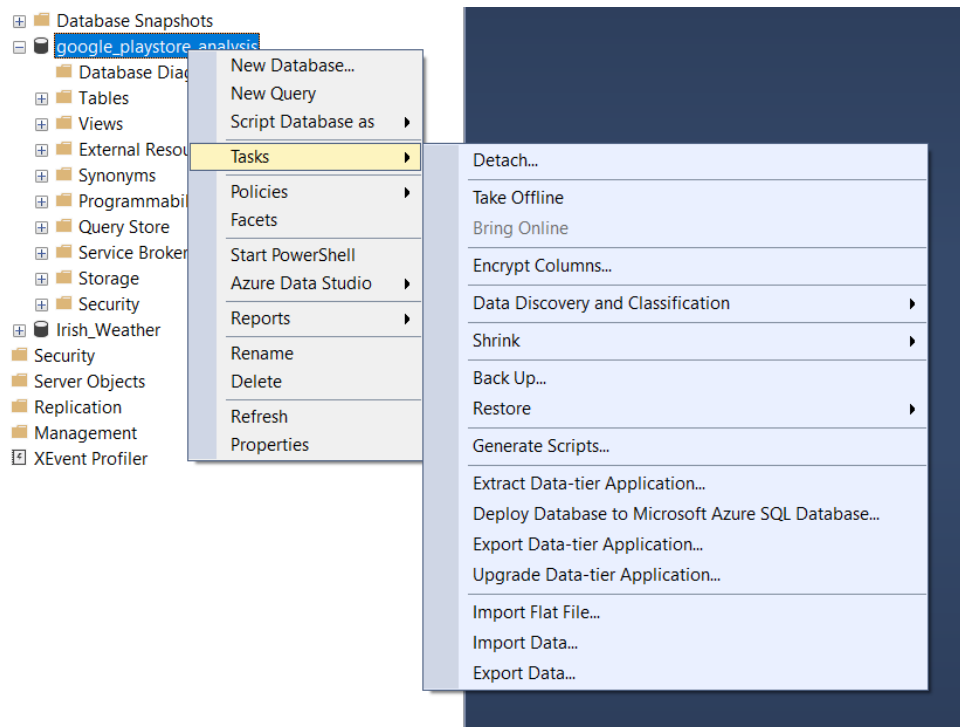
# SQL in SSMS

1. Connect to the SQL Server
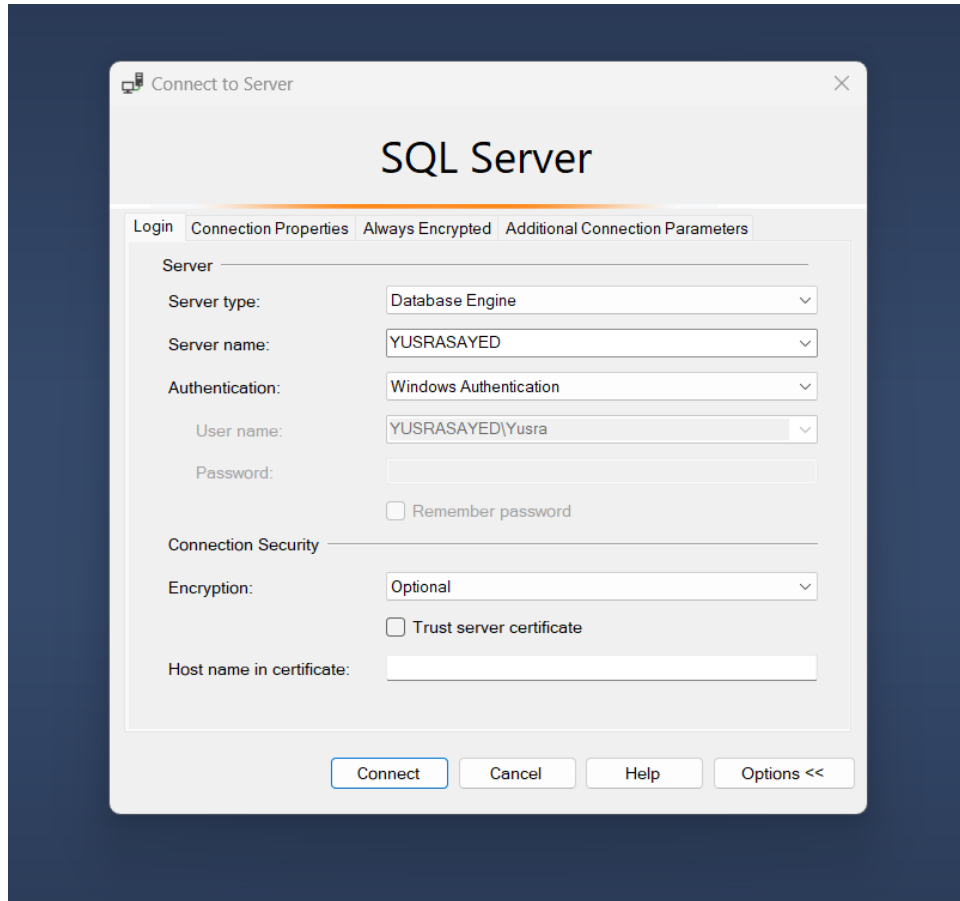

2. Create a new database

3. Name the database and import the csv files into the database by right clicking on the database > Tasks > Import Flat File
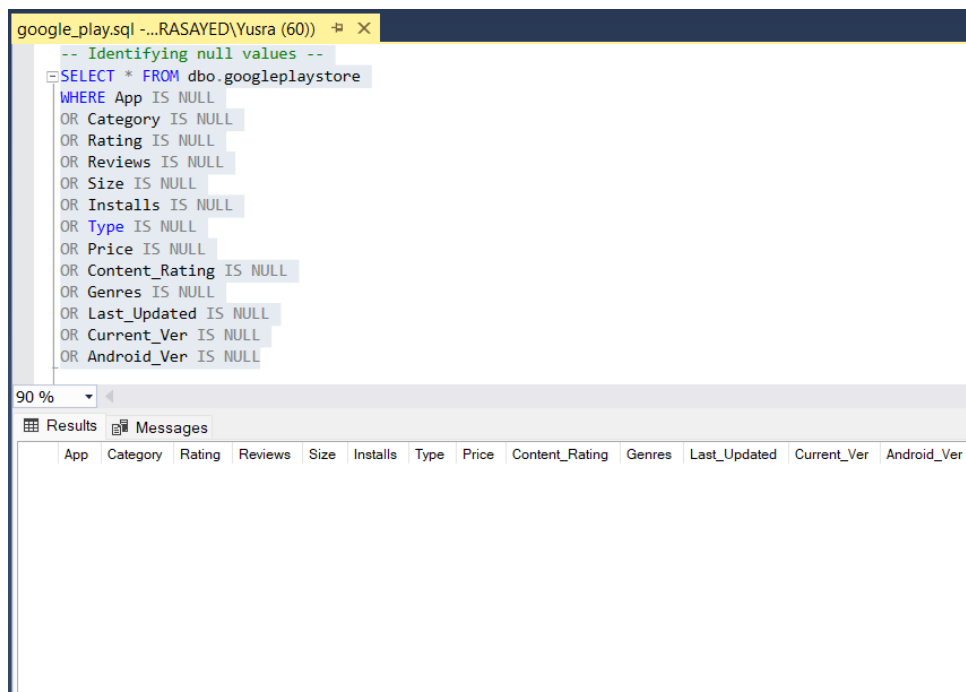

4. Specify the destination of the csv files, click on next and then change the datatypes of each column as necessary and allow null values and click on next. You should see a "Success" message and you are good to go.
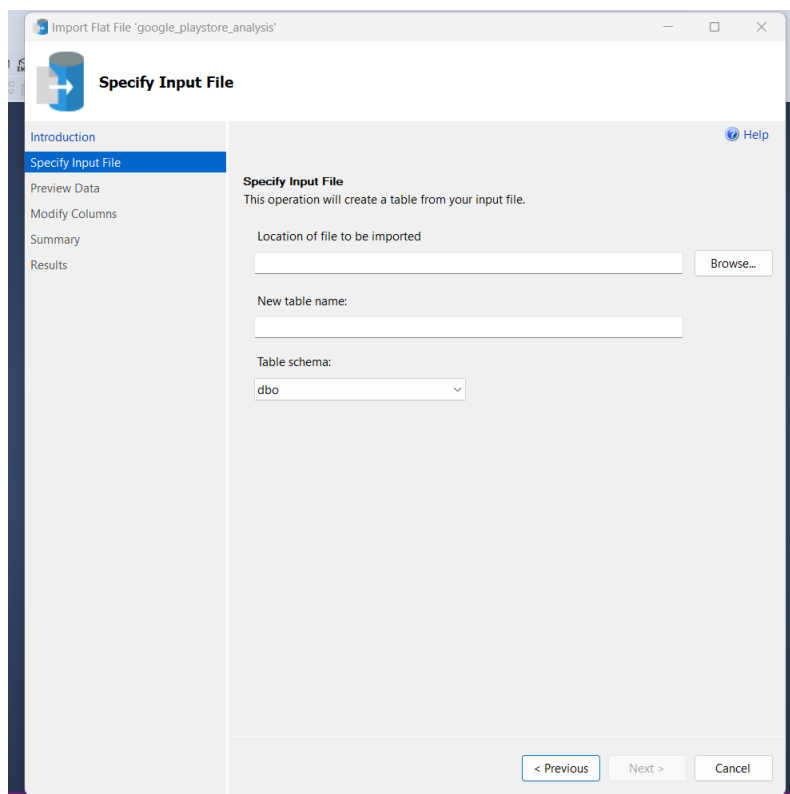
5. Click on new query and start writing SQL queries
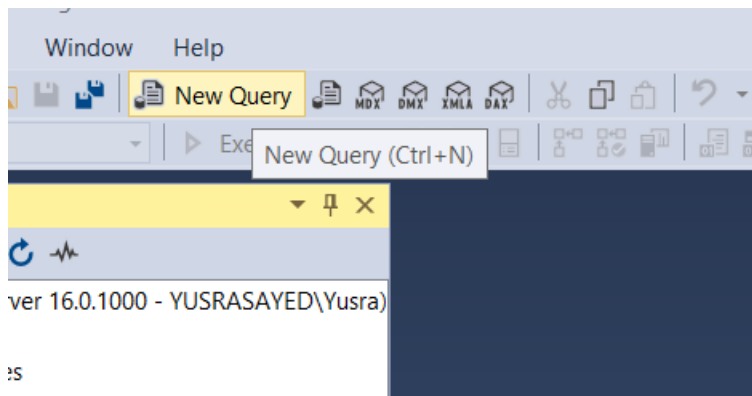
5.1 Identifying Null Values

5.2 Removing the Null Values

```sql
DELETE FROM dbo.googleplaystore
WHERE App IS NULL
OR Category IS NULL
OR Rating IS NULL
OR Reviews IS NULL
OR Size IS NULL
OR Installs IS NULL
OR Type IS NULL
OR Price IS NULL
OR Content_Rating IS NULL
OR Genres IS NULL
OR Last_Updated IS NULL
OR Current_Ver IS NULL
OR Android_Ver IS NULL
```

) %

Messages

```
(0 rows affected)

Completion time: 2024-09-01T22:37:30.1880277+05:30
```

5.3 Overall view of the dataset

```sql
SELECT
COUNT(DISTINCT APP) AS total_apps,
COUNT(DISTINCT Category) AS total_categories
FROM dbo.googleplaystore
```

90 %

Results   Messages

	total_apps	total_categories
1	9635	33

5.4 Exploring app categories

```sql
SELECT TOP 5
Category,
COUNT(App) AS Total_app
FROM dbo.googleplaystore
GROUP BY Category
ORDER BY Total_app DESC
```

90 %

Results   Messages

	Category	Total_app
1	FAMILY	1972
2	GAME	1144
3	TOOLS	843
4	MEDICAL	463
5	BUSINESS	460

5.5 Top rated free apps

```sql
SELECT TOP 10
App,
Category,
Rating,
Reviews
FROM dbo.googleplaystore
WHERE Type ='Free' AND Rating <> 'NaN'
ORDER BY Rating DESC
```

90 %

Results | Messages

	App	Category	Rating	Reviews
1	Hojiboy Tojiboyev Life Hacks	COMICS	5	15
2	American Girls Mobile Numbers	DATING	5	5
3	Awake Dating	DATING	5	2
4	Spine- The dating app	DATING	5	5
5	Girls Live Talk - Free Text and Video Chat	DATING	5	6
6	Online Girls Chat Group	DATING	5	5
7	Speeding Joyride & Car Meet App	DATING	5	3
8	SUMMER SONIC app	EVENTS	5	4
9	Prosperity	EVENTS	5	16
10	Mindvalley U Tallinn 2018	EVENTS	5	1

5.6 Most reviewed apps

```sql
SELECT TOP 10
Category,
APP,
Reviews
FROM dbo.googleplaystore
ORDER BY Reviews DESC
```

90 %

Results | Messages

	Category	APP	Reviews
1	SPORTS	GollerCepte Live Score	9992
2	TOOLS	Ad Block REMOVER - NEED ROOT	999
3	SHOPPING	SnipSnap Coupon App	9975
4	SHOPPING	SnipSnap Coupon App	9975
5	SPORTS	US Open Tennis Championships 2018	9971
6	SPORTS	US Open Tennis Championships 2018	9971
7	TRAVEL_AND_LOCAL	DreamTrips	9971
8	FAMILY	Adult Color by Number Book - Paint Mandala Pages	997
9	VIDEO_PLAYERS	BSPlayer ARMv7 VFP CPU support	9966
10	TOOLS	Easy Resume Builder, Resume help, Curriculum vitae	996

5.7 Average Rating by Category

```sql
SELECT TOP 5
Category,
AVG(TRY_CAST(Rating AS FLOAT)) AS avg_rating
FROM dbo.googleplaystore
GROUP BY Category
ORDER BY avg_rating DESC
```

90 %

Results | Messages

	Category	avg_rating
1	EVENTS	4.43555555555556
2	EDUCATION	4.38903225806452
3	ART_AND_DESIGN	4.35806451612903
4	BOOKS_AND_REFERENCE	4.34606741573034
5	PERSONALIZATION	4.3359872611465

5.8 Average sentiment polarity by app category

```sql
SELECT TOP 10
Category,
AVG(TRY_CAST(Sentiment_Polarity AS FLOAT)) AS avg_sentiment_polarity
FROM dbo.googleplaystore
JOIN dbo.googleplaystore_user_reviews
ON dbo.googleplaystore.App = dbo.googleplaystore_user_reviews.App
GROUP BY Category
ORDER BY avg_sentiment_polarity DESC
```

90 %

Results | Messages

	Category	avg_sentiment_polarity
1	COMICS	0.457055793122866
2	EVENTS	0.377932702552956
3	AUTO_AND_VEHICLES	0.348682739078995
4	PARENTING	0.318880372723734
5	WEATHER	0.282033420287065
6	HEALTH_AND_FITNESS	0.280001939017626
7	MAPS_AND_NAVIGATION	0.267218420452935
8	FOOD_AND_DRINK	0.253250664820985
9	PERSONALIZATION	0.251881938847314
10	EDUCATION	0.251392024242436

5.9 Sentiment reviews by app category

```sql
SELECT TOP 10
Category,
Sentiment,
COUNT(*) AS total_sentiment
FROM dbo.googleplaystore
JOIN dbo.googleplaystore_user_reviews
ON dbo.googleplaystore.App = dbo.googleplaystore_user_reviews.App
WHERE Sentiment <> '-nan'
GROUP BY Category, Sentiment
ORDER BY total_sentiment DESC
```

90 %

⊞ Results ⬚ Messages

	Category	Sentiment	total_sentiment
1	GAME	Positive	11058
2	GAME	nan	9315
3	GAME	Negative	7273
4	SPORTS	nan	5336
5	FAMILY	nan	4037
6	FAMILY	Positive	3888
7	HEALTH_AND_FITNESS	Positive	3475
8	PHOTOGRAPHY	nan	3179
9	PRODUCTIVITY	nan	2659
10	DATING	nan	2558

Dashboard

6.1 Click import data from excel
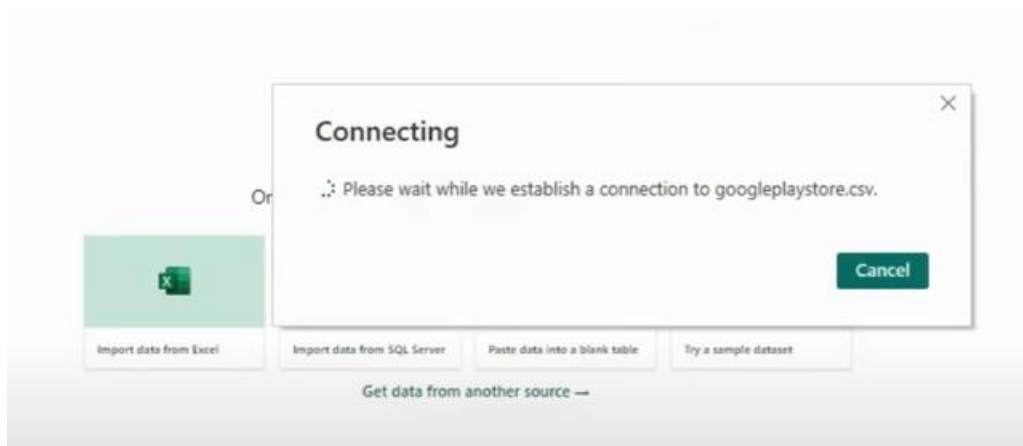
6.2 Click on transform data to clean and process the data to ensure consistency

6.3 close and apply the changes when satisfied with the changes
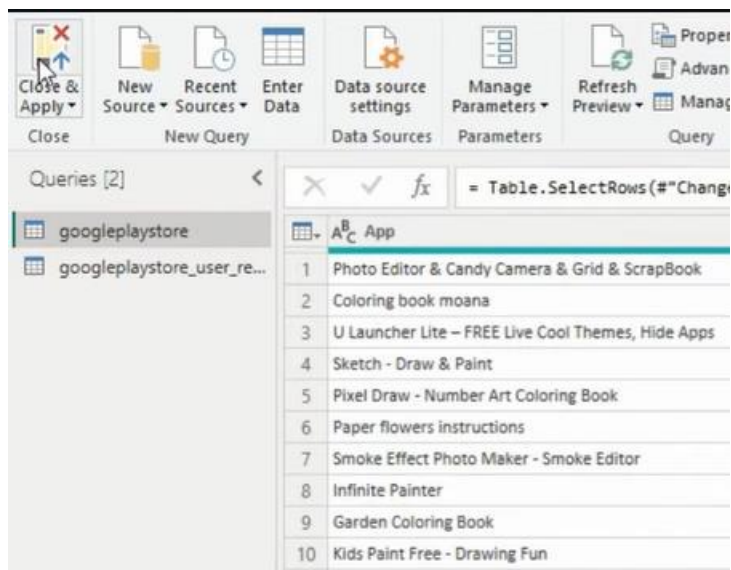
6.4 Load it into PowerBI and begin making the dashboard
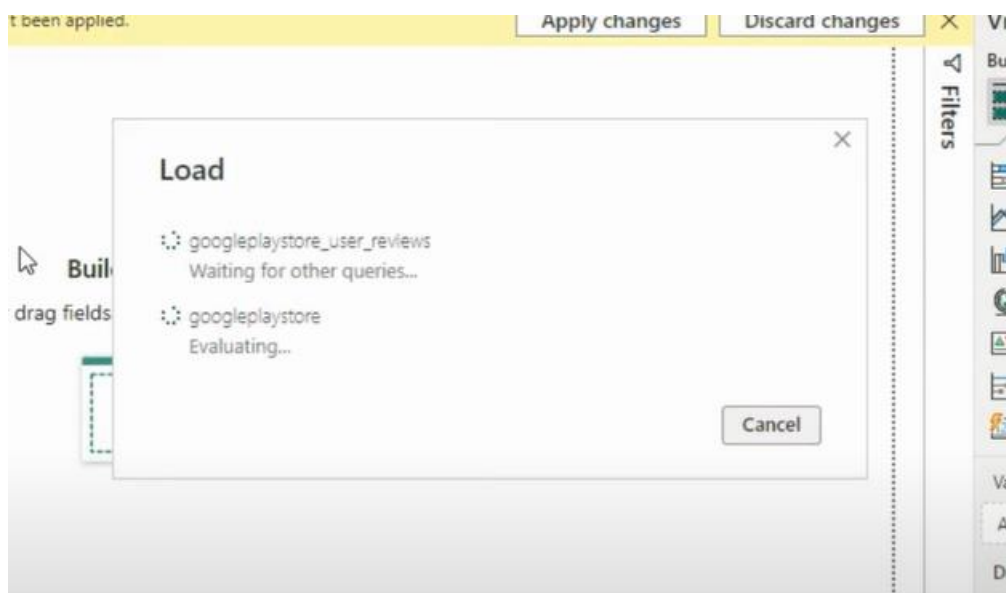
Google Play Store
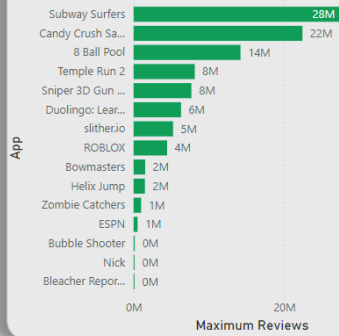
Total Apps
8180

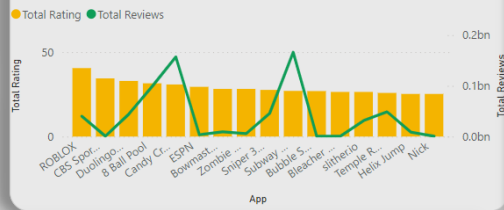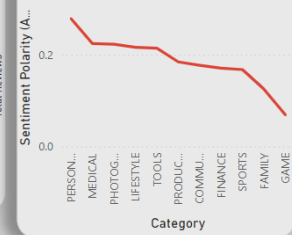**Total Categories**
33

Type
Free Paid

Total Reviews
5bn

Rating and Reviews by App
Total Rating Total Reviews

Sentiment Polarity by Category

Reviews by App

App	Reviews
Subway Surfers	28M
Candy Crush Sa...	22M
8 Ball Pool	14M
Temple Run 2	8M
Sniper 3D Gun ...	8M
Duolingo: Lear...	6M
slither.io	5M
ROBLOX	4M
Bowmasters	2M
Helix Jump	2M
Zombie Catchers	1M
ESPN	1M
Bubble Shooter	0M
Nick	0M
Bleacher Repor...	0M

Maximum Reviews

No. of Apps by Category

Total Installs from each Category

Category by Sentiment
- Positive
- Negative
- Neutral

1.04K (30.69%)
1.22K (35.9...)
1.13K (33.4%)
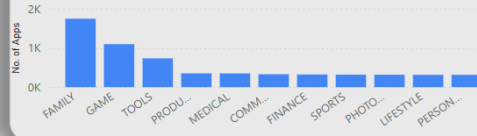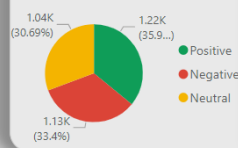
Category	Average of Rating
EVENTS	4.44
EDUCATION	4.39
ART_AND_DESIGN	4.36
BOOKS_AND_REFE...	4.35
PERSONALIZATION	4.34
PARENTING	4.30
GAME	4.29