



PLAY MARKET 2025



Date – 10/07/2025

Prepared By: Anurag Yadav

2. Abstract

The *Play_Market_2025* project is a comprehensive data analysis and visualization initiative focused on exploring the mobile application and game ecosystem from a digital distribution platform. Leveraging a structured dataset containing detailed information such as app names, categories, download counts, user ratings, review scores, and content ratings, this project aims to uncover trends, patterns, and insights into user engagement and market dynamics.



Table of Contents

1. Title Page & Cover.....	1
2. Abstract.....	2
3. Table of Contents	3
4. Introduction	6
○ Background	6
○ Project Overview	6
○ Project Objective	6
5. Data Description	7
○ Dataset Source & Acquisition	7
○ Dataset Characteristic	7
6. Methodology	8
○ Data Cleaning & Preprocessing (Python).....	8
○ Database Design & Querying (MySQL/SQL).....	8
○ Data Exploration & Summarization.....	8
○ Data Visualization & Dashboarding	9
○ Outcome	9
7. Play_Market_2025 Analysing App & Game Trend for better market understanding	10
○ Datasets explored in this project	10
■ App Information Dataset.....	10
■ Review Summary Dataset	10
○ Datasets explored in this project	11
■ Product & Category Strategy	11
■ User Management & Feedback Management	11
■ Performance Benchmarking	11
■ Market Segmentation	11
■ Informed Decision-Making	11
8. Python for Data Cleaning – Play_Market_2025	12
○ Data Cleaning Workflow in Python.....	12
■ Importing & Inspecting the data	12
■ Handling Missing & Null Values	13
■ Standardizing Categorical Columns	13
■ Data Type Conversion	14
■ Duplicate Removal	14
■ Exporting the Cleaned Data	14
○ Tools & Libraries used	15
○ Outcome	15

9. Database setup using python	16
○ Establishing Connection to MySQL	16
○ Creating new database	17
○ Creating tables for the datasets	18
○ Uploading Clean Data into tables	19
10. Data extraction for Analysis using Python	20
○ Establishing Database Connection	20
○ Running SQL Queries & Exporting results	21
○ Exporting CSV or Excel	22
○ Multiple Queries without loops (Optimized Approach)	22
○ Outcome	22
11. Insightful Analytics through Excel Dashboards	23
11.1 Excel Dashboards – App Performance & Download	23
○ Overview	23
○ Dashboard Components	23
■ Apps (Downloads Category Wise)	23
■ Category-Wise: Total Review Vs Average Review Score	23
■ Average Review Score (Across Dates)	23
■ Play Game Review Summary	24
○ Insights Derived	24
○ Business Impact	24
11.2 Excel Dashboards – Game Insights	25
○ Overview	25
○ Dashboard Components	25
■ Games (Downloads Vs Rating)	25
■ Number of Games Vs Average Score (by Type)	25
■ Average Review Score (Monthly)	25
○ Insights Derived	26
■ Game Type Performance	26
■ User Sentiment Over Time	26
■ Popular Individual Games	26
○ Business Impact	26
■ Product Strategy	26
■ Content Expansion	26
■ Sentiment Monitoring	27

11.3	Excel Dashboards – User Review Analysis	27
○	Overview	27
○	Key Visualization	27
■	Apps vs Games- Helpful Count & Total Review	27
■	Total Review Volume (Apps)	27
■	Total Review Volume (Games)	27
○	Insights Gained	28
11.4	Dynamic Data Visualization	29
○	Dynamic Components Implemented	29
■	App Performance by Category	29
■	Games – Downloads vs Average Score	29
■	App Review Scores by Section	29
■	Monthly Review Volume Analysis	29
○	Business Impact	30
○	Insights Gained	31
12.	Unleashing Data Through Visual Insights	32
12.1	Landing Page	32
○	Overview	32
○	Insights Gained	33
12.2	Overall Market	34
○	Overview	34
○	Insights Gained	34
12.3	Top Categories by Popularity	35
○	Overview	35
○	Insights Gained	35
12.4	App/Category Review Analysis	37
○	Overview	37
○	Insights Gained	37
12.5	Top Apps & Games	38
○	Overview	38
○	Key Visualization	38
○	Insights Gained	39
13.	Discussion	40
14.	Future work/Potential	40
15.	Conclusion	41
16.	Analyst Profile	41

4. Introduction

- **Background**

The mobile application industry has seen exponential growth over the past decade, with millions of apps and games being hosted on digital distribution platforms. Understanding how these apps perform across different categories, their user engagement, and overall market presence is critical for developers, analysts, and business stakeholders. To address this, *Play_Market_2025* presents a data-driven approach to visualize and analyse the app ecosystem using Power BI.

- **Project Overview**

Play_Market_2025 is a data analytics and visualization project aimed at exploring and evaluating mobile applications and games available on a digital marketplace. The project integrates multiple tools and technologies — including **Python for data cleaning**, **MySQL for structured storage and querying**, **Excel for preprocessing and export**, and **Power BI** for dashboard creation and visualization.

The dataset includes key information such as app names, categories, user ratings, review scores, helpfulness, downloads, and content ratings. Insights are presented through thematic dashboard pages, each addressing a specific angle of market understanding — such as overall trends, category performance, review behaviour, and top-performing apps and games.

- **Project Objectives**

The main goals of the *Play_Market_2025* project are:

- ✓ To clean, structure, and analyse mobile app data using Python and SQL.
- ✓ To explore category-wise trends in app ratings, reviews, and downloads.
- ✓ To identify top-performing apps and games based on user engagement metrics.
- ✓ To export query results into Excel and use them to build interactive, theme-aligned dashboards.
- ✓ To demonstrate end-to-end data analysis and visualization skills across multiple tools and technologies.

5. Data Description

- **Dataset Source & Acquisition**

The dataset used for the *Play_Market_2025* project was sourced from **Kaggle.com**, a popular platform for sharing and exploring datasets. It was originally uploaded by **Dmytro Buhai**, who compiled comprehensive information about apps and games available on a digital distribution platform, including their reviews, ratings, and metadata.

The dataset was downloaded in CSV format and then processed using tools such as **Python (Pandas)** and **SQL** for cleaning, transformation, and integration before analysis.

- **Dataset Characteristics**

The dataset consists of multiple columns capturing critical information about mobile applications and games, categorized by type, content rating, and user feedback. Below are the key characteristics:

- ✓ **App Metadata:** Includes app name, category, section, and type (App/Game)
- ✓ **Performance Metrics:** Number of downloads, total review scores, and user ratings
- ✓ **User Feedback:** Number of reviews, helpful scores, and average rating
- ✓ **Categorical Info:** Content rating (e.g., Everyone, Teen), and category groupings (e.g., Tools, Education)

The dataset supports both **quantitative** and **categorical** analysis, making it ideal for exploring:

- ✓ Top-performing apps
- ✓ Review and rating trends
- ✓ Engagement metrics across different app types and categories

The structured format and diverse feature set enable a wide range of insights using data analysis and visualization tools such as Power BI and Excel.

6. Methodology

The **Play_Market_2025** project followed a structured, end-to-end data analysis pipeline involving data extraction, cleaning, transformation, analysis, and visualization. A combination of tools and technologies — including **Python**, **SQL**, **Excel**, and **Power BI** — was used at different stages to ensure accuracy, scalability, and interactivity of the final output.

- **Data Cleaning & Preprocessing (Python)**

- Imported the dataset using **Pandas**.
- Identified and handled missing values, inconsistencies, and data type mismatches.
- Normalized text-based columns like **category**, **content_rating**, and **type**.
- Converted numerical fields such as **downloads**, **score**, and **rating_count** into appropriate formats for aggregation and comparison.
- Exported the cleaned data to CSV for database loading.

- **Database Design & Querying (MySQL/SQL)**

- Created a structured database and tables using MySQL.
- Imported the cleaned dataset into the database.
- Performed data analysis using SQL queries to:
 - ✓ Aggregate downloads, scores, and reviews
 - ✓ Rank apps by performance
 - ✓ Join and filter data for deeper insights (e.g., category-level performance, rating analysis)
- Exported query results into Excel for further visualization and dashboard integration.

- **Data Exploration & Summarization (Excel)**

- Loaded SQL query outputs into **Excel**.
- Conducted initial exploratory analysis using pivot tables and charts.
- Used Excel to validate data trends, totals, and category-wise distributions before visualization in Power BI.

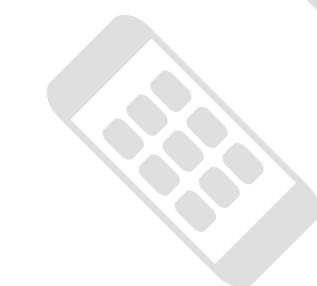
- **Data Visualization & Dashboarding (Power BI)**

- Imported final datasets into **Power BI** for interactive visualization.
- Created multiple dashboards/pages:
 - ✓ Landing Page
 - ✓ Overall Market Overview
 - ✓ Top Apps & Games
 - ✓ Category & Review Analysis
- Applied consistent theming using a custom colour palette and layout components.
- Designed slicers, filters, buttons, and drill-throughs for better navigation and user experience.
- Used **DAX** to calculate KPIs such as rankings, average scores, helpful review percentages, and monthly review trends.

- **Outcome**

The methodology ensured a seamless flow from raw data to insightful dashboards. Each technology played a specific role:

- **Python** for data shaping,
- **SQL** for querying and filtering,
- **Excel** for tabular analysis, and
- **Power BI** for delivering a polished, interactive, and shareable dashboard experience.



7. Play_Market_2025 – Analysing App & Game Trends for Better Market understanding

The Play_Market_2025 project leverages the power of data analytics and visualization to uncover meaningful patterns in the mobile application and gaming ecosystem. With millions of apps competing for visibility and user engagement, understanding how users interact with these apps has become crucial for developers, marketers, and platform owners.

• Datasets Explored in This Project

The analysis in this project is based on a rich and structured dataset sourced from **Kaggle**, originally compiled by **Dmytro Buhai**. This dataset includes two core components:

- **App Information Dataset**

Contains static metadata about each app or game, including:

- ✓ **App name, type** (App/Game), and **category**
- ✓ **Download count**
- ✓ **Average user score**
- ✓ **Content rating** (e.g., Everyone, Teen)
- ✓ **Section** (subcategory like Tools, Action, Education)

- **Review Summary Dataset**

Provides engagement insights, such as:

- ✓ **Number of reviews**
- ✓ **Total review score**
- ✓ **Helpful score** (indicator of how valuable users found the reviews)

Together, these datasets enable both **categorical segmentation** and **quantitative performance analysis**, which are essential for strategic business decision-making.

- **Business Impact & Insight Potential**

The visualizations and metrics derived from this dataset deliver insights with strong real-world applications:

- **Product & Category Strategy**

Helps app developers and publishers identify **high-performing categories**, align new product launches with user demand, and adjust marketing focus based on popularity and engagement.

- **User Experience & Feedback Management**

By analysing **review trends and helpfulness scores**, companies can detect user pain points, assess sentiment quality, and respond to feedback more effectively.

- **Performance Benchmarking**

Ranking apps by downloads, ratings, or helpfulness allows product teams to **benchmark their performance** against top competitors in the market.

- **Market Segmentation**

Stakeholders can assess content rating distribution and section-specific performance to **target the right demographics** with personalized marketing or app features.

- **Informed Decision-Making**

Interactive dashboards empower business users to explore data with filters and visuals, enabling **real-time insights** without the need for technical intervention.

By combining structured data with intuitive visual storytelling, Play_Market_2025 demonstrates the business value of integrated data analytics across the mobile app lifecycle — from development and review monitoring to strategic planning and execution.

8. Python for Data Cleaning – Play_Market_2025

Before performing any analysis or visualization, it is essential to ensure that the dataset is clean, consistent, and analysis-ready. In the *Play_Market_2025* project, Python (using the pandas library) was used to perform all key data cleaning operations on the raw dataset downloaded from Kaggle.

This step was crucial for identifying issues such as missing values, inconsistent formatting, incorrect data types, and redundant or irrelevant entries. Below is a summary of the approach taken during the cleaning process.

- **Data Cleaning Workflow in Python**

- **Importing and Inspecting the Data**

- ✓ Used **pandas.read_csv()** to load the dataset.
 - ✓ Displayed the first few rows using **df.head()** and examined the structure using **df.info()** and **df.describe()**.



```
import pandas as pd

app_info = pd.read_csv("C:/Users/anura/OneDrive/Desktop/Play Market/Datasets/apps_info.csv")

app_info.head()

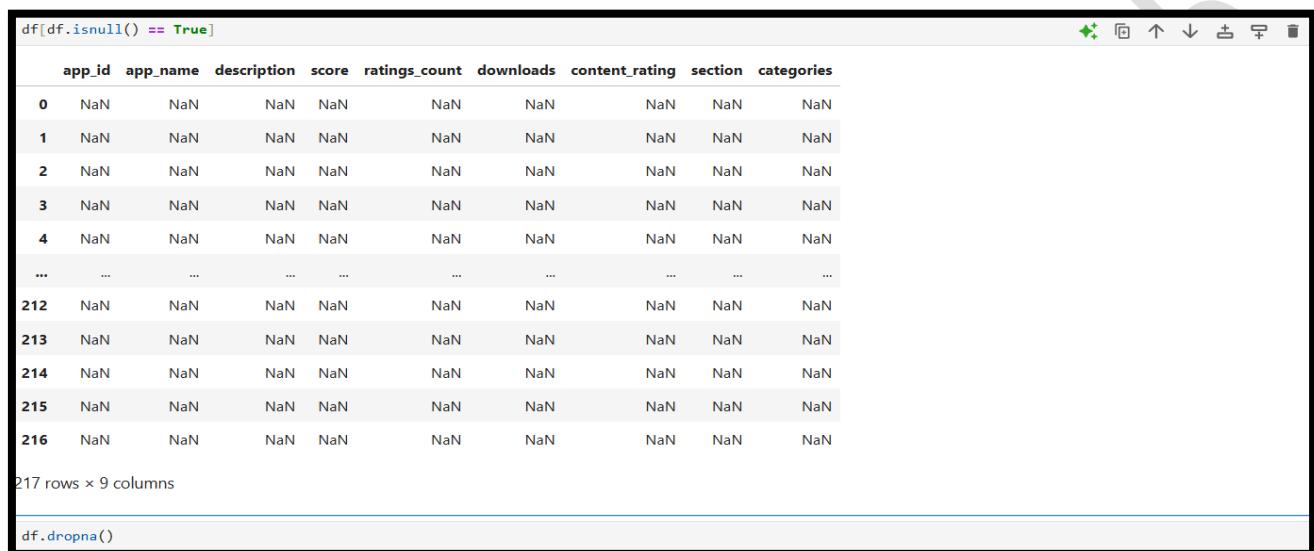
  Type app_id          app_name           description  score  ratings_count  downloads  content_rating    section  categories
0  App     1  Western Union Send Money Now  Enjoy a $0 transfer fee* on your next online i...    4.6        317000  10000000  Everyone  Budgeting tools  Finance
1  App     2  Priceline: Hotel, Flight & Car  Save on your next trip with exclusive deals on...    4.7        502000  10000000  Everyone  Business tools  Travel & Local
2  App     3            Slack  Slack brings team communication and collaborat...
3  App     4        MyWalmart  Introducing MyWalmart, the one app designed fo...
4  App     5  Fidelity Investments  Invest at a firm invested in you. Fidelity's s...

df = pd.DataFrame(app_info)
df.head()
```

Figure 1

○ Handling Missing and Null Values

- ✓ Identified missing values using `df.isnull().sum()`.
- ✓ Dropped rows with excessive missing data or filled them using appropriate methods (e.g., `fillna()` for numerical columns).
- ✓ For example, missing **scores** or **ratings** were either dropped (if minimal) or filled using median or mode.



```
df[df.isnull() == True]

  app_id app_name  description  score  ratings_count  downloads  content_rating  section  categories
0      NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
1      NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
2      NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
3      NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
4      NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
...    ...
212     NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
213     NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
214     NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
215     NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN
216     NaN      NaN          NaN  NaN           NaN      NaN          NaN      NaN          NaN

217 rows × 9 columns

df.dropna()
```

Figure 2

○ Standardizing Categorical Columns

- ✓ Converted all string columns to lowercase and removed extra spaces using `.str.lower().str.strip()`.
- ✓ Standardized inconsistent entries in fields like:
 - type (e.g., "app", "App ", "APP" → "app")
 - content_rating (e.g., "everyone", "Everyone", "EVERYONE" → "Everyone")

○ Data Type Conversion

- ✓ Converted numerical columns like downloads, review_score, helpful_score from string/object to integer or float using `pd.to_numeric()`.
- ✓ Used `astype()` to enforce correct types (e.g., categorical types for type, section).

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 217 entries, 0 to 216
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   app_id            217 non-null    int64  
 1   app_name          217 non-null    string  
 2   description       217 non-null    string  
 3   score             217 non-null    float64 
 4   ratings_count     217 non-null    int64  
 5   downloads          217 non-null    int64  
 6   content_rating    217 non-null    string  
 7   section            217 non-null    string  
 8   categories         217 non-null    string  
dtypes: float64(1), int64(3), string(5)
memory usage: 15.4 KB

df['app_name'] = df['app_name'].astype('string')

df['description'] = df['description'].astype('string')

df['content_rating'] = df['content_rating'].astype('string')

df['section'] = df['section'].astype('string')

df['categories'] = df['categories'].astype('string')
```

Figure 3

○ Duplicate Removal

- ✓ Checked for and removed duplicate app entries using `df.duplicated()` and `df.drop_duplicates()`.

○ Exporting the Cleaned Data

- ✓ After completing the cleaning, the dataset was exported as a cleaned CSV file using:

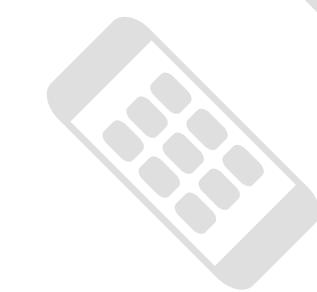
```
df.to_csv("cleaned_play_market_data.csv", index=False)
```

- **Tools & Libraries Used**

- ✓ **Python 3.11+**
- ✓ **pandas** for data manipulation
- ✓ **numpy** for numerical operations
- ✓ **Jupyter Notebook / VS Code** for execution and tracking

- **Outcome**

The cleaned dataset was free from inconsistencies and ready for further analysis in SQL, Excel, and Power BI. This ensured accuracy in metrics like downloads, ratings, and reviews, and allowed for seamless querying and visualization downstream.



9. Database Setup Using Python

After cleaning the dataset using Python, the next step in the *Play_Market_2025* project was to set up a relational database for structured querying and further analysis. This was done using Python in combination with the MySQL Connector package.

- **Establishing Connection to MySQL**

Using **mysql.connector**, a secure connection to the MySQL server was established:

```
[ ]: import mysql.connector

[ ]: # MySQL Credentials
    host = 'localhost'
    user = 'root'
    password = '*****'
    port = 3306
    database_name = 'Play_Market_2025'

[ ]: conn = mysql.connector.connect(
        host = host,
        user = user,
        password = password,
        allow_local_infile = True
    )
    cursor = conn.cursor()
```

Figure 4

- **Creating a New Database**

A new database named play_market_2025 was created to organize the app and review datasets:

```
cursor.execute("CREATE DATABASE play_market_2025")
```

Once created, the connection was switched to this database for all further operations:

```
cursor.execute("USE play_market_2025")
```

```
[ ]: #Drop Already Existing Database
cursor.execute("DROP DATABASE IF EXISTS play_market_2025")
print("Database play_market_2025 has been Dropped!")

#Create a new database
cursor.execute("CREATE DATABASE IF NOT EXISTS play_market_2025")
print("Database play_market_2025 is successfully created")

#Use newly created database
cursor.execute("USE play_market_2025")
print("Database Switched!")
```

Figure 5

- **Creating Tables for the Datasets**

Separate tables were created to store the app information and review metrics. Each table was designed with appropriate data types based on the cleaned dataset:

```
# Table creation SQL
table_name = 'apps_info'
table_schema = f"""
CREATE TABLE IF NOT EXISTS {table_name}(
app_id INT PRIMARY KEY,
app_name VARCHAR(255),
description VARCHAR(255),
score DECIMAL(3,2),
ratings_count INT,
downloads INT,
content_rating VARCHAR(50),
section VARCHAR(100),
categories VARCHAR(255));
"""

table_name_2 = 'apps_reviews'
table_schema_2 = f"""
CREATE TABLE IF NOT EXISTS {table_name_2}(
app_id INT,
review_text VARCHAR(100),
review_score DECIMAL(3,2),
review_date DATE,
helpful_count INT);
"""

table_name_3 = 'games_info'
table_schema_3 = f"""
CREATE TABLE IF NOT EXISTS {table_name_3}(
game_id INT PRIMARY KEY,
game_name VARCHAR(255),
description VARCHAR(255),
score DECIMAL(3,2),
ratings_count INT,
downloads INT,
content_rating VARCHAR(50),
section VARCHAR(100),
categories VARCHAR(255));
"""

table_name_4 = 'games_reviews'
table_schema_4 = f"""
CREATE TABLE IF NOT EXISTS {table_name_4}(
game_id INT,
review_text VARCHAR(100),
review_score DECIMAL(3,2),
review_date DATE,
helpful_count INT);
"""
```

Figure 6

- **Uploading Cleaned Data into Tables**

The cleaned CSV files were read back into Python using pandas, and each row was inserted into the MySQL tables:

```
query = """
LOAD DATA INFILE 'C:/ProgramData/MySQL/
MySQL Server 8.0/Uploads/apps_info.csv'
INTO TABLE apps_info
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY ''
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
"""

query_2 = """
LOAD DATA INFILE 'C:/ProgramData/MySQL/
MySQL Server 8.0/Uploads/apps_reviews.csv'
INTO TABLE apps_reviews
FIELDS TERMINATED BY ','
ENCLOSED BY ''
LINES TERMINATED BY '\n';
"""

query_3 = """
LOAD DATA INFILE 'C:/ProgramData/MySQL/
MySQL Server 8.0/Uploads/games_info.csv'
INTO TABLE games_info
FIELDS TERMINATED BY ','
ENCLOSED BY ''
LINES TERMINATED BY '\n';
"""

query_4 = """
LOAD DATA INFILE 'C:/ProgramData/MySQL/
MySQL Server 8.0/Uploads/games_reviews.csv'
INTO TABLE games_reviews
FIELDS TERMINATED BY ','
ENCLOSED BY ''
LINES TERMINATED BY '\n';
"""

cursor.execute(query)
print(f"Data successfully uploaded to {table_name}")

cursor.execute(query_2)
print(f"Data successfully uploaded to {table_name_2}")

cursor.execute(query_3)
print(f"Data successfully uploaded to {table_name_3}")

cursor.execute(query_4)
print(f"Data successfully uploaded to {table_name_4})
```

Figure 7

The data was then committed to make the changes permanent:

conn.commit()

- **Outcome**

At this stage:

- ✓ A structured database (play_market_2025) was successfully created.
- ✓ Cleaned data was inserted into normalized tables.
- ✓ The setup allowed for seamless SQL querying and later export to Excel for Power BI visualization.

This phase played a key role in preparing the backend for scalable, filterable, and relational data exploration.

10. Data Extraction for Analysis Using Python

After setting up the database and uploading the cleaned data, the next crucial step in the *Play_Market_2025* project was to extract data from MySQL for analysis and visualization. This was accomplished through a Python script designed to run custom SQL queries and export the results to Excel or CSV files for reporting and use in Power BI.

- **Establishing Database Connection**

The script begins by establishing a connection to the MySQL database using **mysql.connector**:

```
[ ]: import mysql.connector
      import pandas as pd

[ ]: # MySQL Credentials
      host = 'localhost'
      user = 'root'
      password = '*****'
      port = 3306
      database_name = 'Play_Market_2025'

[ ]: conn = mysql.connector.connect(
      host = host,
      user = user,
      password = password,
      allow_local_infile = True
    )
      cursor = conn.cursor()
```

Figure 8

- **Running SQL Queries and Exporting Results**

Each analytical query was written directly in Python and executed using pandas' `read_sql()` method to fetch the results into DataFrames. Examples of such analysis include:

- ✓ Top 10 apps by downloads
- ✓ Average score by category
- ✓ Monthly trend of reviews
- ✓ Most helpful reviews

```
table_1 = 'apps_info'
table_2 = 'apps_reviews'
table_3 = 'games_info'
table_4 = 'games_reviews'

# App performance & Downloads
# Top 10 apps by total downloads
query_1A = """
    select app_id, app_name, categories,
    downloads from {table_1}
    group by 1,2
    order by 4 desc
    limit 10;
"""

# Average Score & number of reviews per category
query_2A = """
    select t1.categories, count(t2.review_text) as "Total Reviews",
    avg(t2.review_score) as "Average Review Score" from {table_1} t1
    join {table_2} t2 on t1.app_id = t2.app_id
    group by 1
    order by 2 desc;
"""

# Monthly trend of app reviews
query_3A = """
    select date_format(review_date, '%y-%m') as "Month_Name",
    avg(review_score) as "Average_Review_Score" from {table_2}
    where monthname(review_date) is not null
    group by 1
    order by 1;
"""

# Top 5 most helpful app reviews
query_4A = """
    select t1.app_id, t1.app_name,
    count(t2.review_score) as "Number of Review Scores",
    sum(t2.helpful_count) as "Total helpful app reviews"
    from {table_1} t1
    join {table_2} t2 on t1.app_id = t2.app_id
    group by 1,2
    order by 4 desc
    limit 5;
"""
```

Figure 9

- **Exporting to CSV or Excel**

Once the results were loaded into a DataFrame, they were exported for reporting and visualization:

```
#App Performance & Downloads (Query_1) [Top 10 apps by total downloads (Bar/Column chart : app_name Vs downloads)]
df1A = pd.read_sql(query_1A, conn)
df1A.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/App Performance & Downloads/query_1.csv", index = False)
print("File Exported for query_1A")

#App Performance & Downloads (Query_2) [Average Score & number of reviews per category (Combo chart: bar for review count + Line for avg score)]
df2A = pd.read_sql(query_2A, conn)
df2A.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/App Performance & Downloads/query_2.csv", index = False)
print("File Exported for query_2A")

#App Performance & Downloads (Query_3) [Monthly trend of app reviews (Line chart - To show distribution over time)]
df3A = pd.read_sql(query_3A, conn)
df3A.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/App Performance & Downloads/query_3.csv", index = False)
print("File Exported for query_3A")

#App Performance & Downloads (Query_4) [Top 5 most helpful app reviews (summary table with conditional formatting on Total helpful app reviews)]
df4A = pd.read_sql(query_4A, conn)
df4A.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/App Performance & Downloads/query_4.csv", index = False)
print("File Exported for query_4A")

#Game Insights (Query_1) [Top 10 games with highest rating and more than 10,000 downloads]
df1G = pd.read_sql(query_1G, conn)
df1G.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/Game Insights/query_1.csv", index = False)
print("File Exported for query_1G")

#Game Insight (Query_2) [Game section with average score and number of games (Clustered bar chart: category vs Average Score and Total Games/Treemap)
df2G = pd.read_sql(query_2G, conn)
df2G.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/Game Insights/query_2.csv", index = False)
print("File Exported for query_2G")

#Game Insight (Query_3) [Monthly trend of game reviews (Line chart - To show distribution over time)]
df3G = pd.read_sql(query_3G, conn)
df3G.to_csv("C:/Users/anura/OneDrive/Desktop/Play Market/New folder/Excel Dashboard Export/Game Insights/query_3.csv", index = False)
print("File Exported for query_3G")
```

Figure 10

- **Multiple Queries Without Loops (Optimized Approach)**

To streamline exports without using loops, each query was written and executed separately, with clear naming conventions for the output files. This improves readability and maintains transparency in what each output file contains.

- **Outcome**

This step enabled:

- ✓ Seamless extraction of SQL-based insights
- ✓ Clean, structured outputs ready for Power BI dashboards and Excel reports
- ✓ The flexibility to tweak queries and regenerate analysis quickly

11. Insightful Analytics through Excel Dashboards

Excel remains one of the most versatile tools for rapid and interactive data visualization. In this project, Excel dashboards serve as a powerful medium to transform cleaned and structured data into accessible, decision-ready insights. With the use of pivot tables, slicer, charts, and conditional formatting, these dashboards provide a clear narrative on app performance, user engagement, and category-wise trends. Their simplicity, flexibility, and widespread usability make excel dashboards a crucial step in showcasing key findings before progressing to advanced BI tools like Power BI.

11.1 Excel Dashboard – *App Performance & Download*

○ Overview

After successfully extracting cleaned data from the SQL database using Python, the next phase involved developing analytical dashboards using Microsoft Excel. This dashboard titled "App Performance & Download" focuses on visualizing user engagement, app popularity, and rating trends across various categories.

○ Dashboard Components

The dashboard consists of multiple interactive and insight-rich charts:

- **Apps (Downloads Category-wise)**
 - ✓ Bar chart showing total downloads across major app categories.
 - ✓ Highlights which categories dominate user installs.
 - ✓ Clear indication that **Communication** and **Productivity** apps lead in volume.
- **Category-Wise: Total Review vs Average Review Score**
 - ✓ Combo chart (bar + line) showing:
 - Total review counts (bar).
 - Average review scores (line).
 - ✓ Useful in assessing whether volume correlates with user satisfaction.
- **Average Review Score (Across Dates)**
 - ✓ Line chart visualizing **time-series trend** of user ratings.
 - ✓ Helps analyze how sentiment or performance fluctuates over time.

▪ Play Games Review Summary

- ✓ Key performance metrics for *Play Games*:

- Total Helpful App Reviews.
- Number of Review Scores.

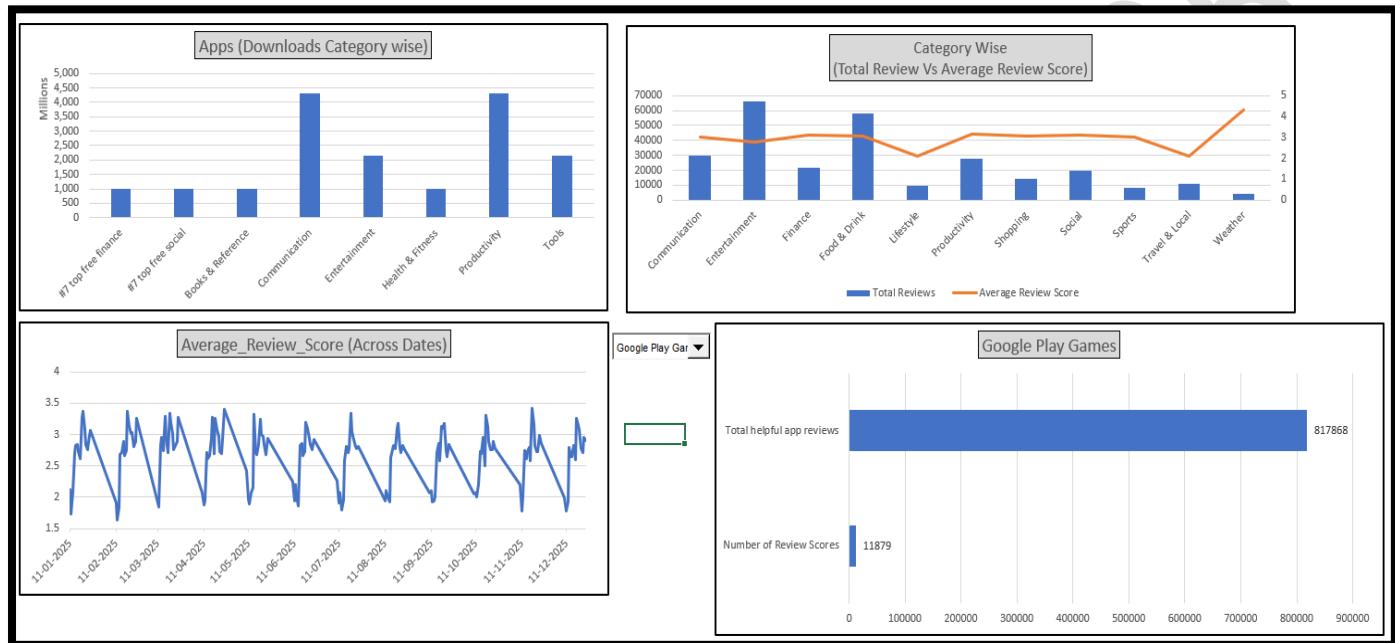


Figure 11

○ Insights Derived

- Communication and Productivity categories are leading in total downloads, confirming their high relevance to users.
- Categories like Weather or Finance show relatively lower engagement but maintain higher average ratings, signalling niche satisfaction.
- The time-based review score chart shows fluctuations that may relate to user experience changes or feature rollouts.
- One app stands out with over 800K helpful reviews, showing strong user involvement and usefulness.

○ Business Impact

This Excel-based dashboard enables data-driven decision-making by:

- Identifying the most and least engaging app categories.

- Highlighting user satisfaction through sentiment analysis.
- Guiding stakeholders on **where to invest or improve** based on user feedback and category performance.
- Providing historical review patterns for performance tracking.

11.2 Excel Dashboard – *Game Insight*

○ Overview

The "Game Insight" dashboard is crafted to explore deeper patterns within the mobile gaming segment of the dataset. After extracting and cleaning the relevant data, this Excel dashboard analyses download trends, rating behaviour, and the distribution of game types based on user feedback and engagement.

○ Dashboard Components

▪ Games (Download vs Rating)

- ✓ A combination bar and line chart showcasing:
 - Number of downloads per game (bars).
 - Corresponding highest ratings (line).
- ✓ Highlights the popularity of games like **Words of wonder: Crossword**, **Vita Mahjong**, and **Simon's Cat Match**.

▪ Number of Games vs Average Score (by Type)

- ✓ Horizontal bar chart comparing:
 - Number of games per category.
 - Average score for each category.
- ✓ Clearly indicates that categories like **Explore games on PC**, **Casual games**, and **Stylized games** have the most content and consistent user satisfaction.

▪ Average Review Score (Monthly)

- ✓ A time series line chart displaying how average review scores for games trend over time.

- ✓ Enables tracking of how user sentiment changes across months—potentially due to new game launches or update.

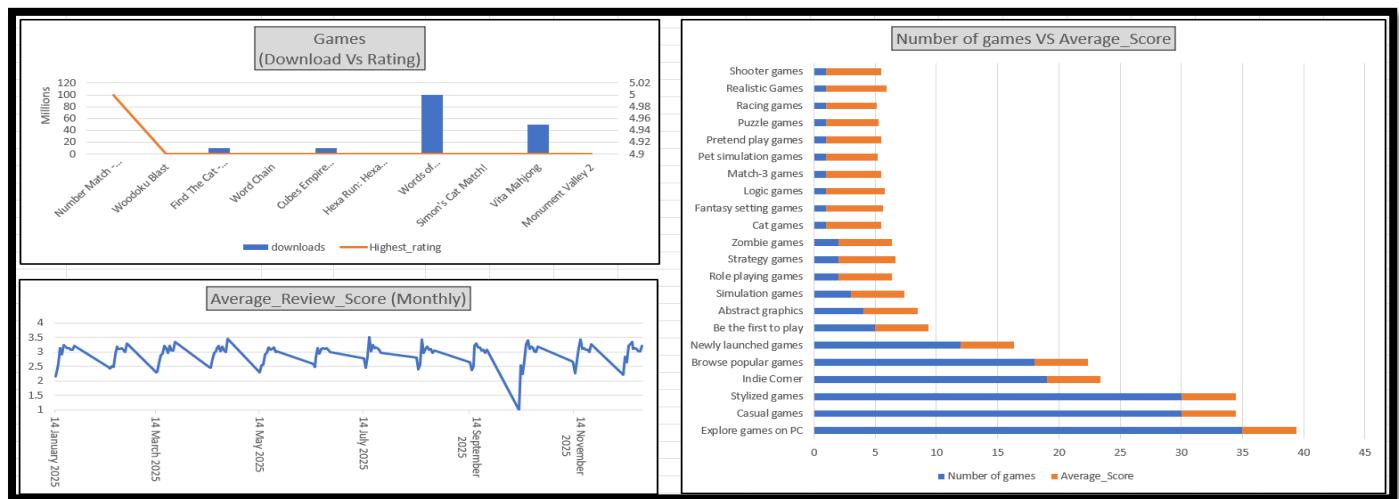


Figure 12

○ Key Insights

▪ Game Type Performance:

Categories such as **Explore games on PC** and **Casual games** not only have the most games but also rank high in average scores, indicating their broad appeal and quality.

▪ User Sentiment Over Time:

Monthly review scores show cyclical behaviour, suggesting periodic engagement peaks (e.g., during holidays or major releases).

▪ Popular Individual Games:

A few games attract massive downloads while maintaining near-perfect ratings—signalling both reach and quality.

○ Business Impact

▪ Product Strategy:

Developers and marketing teams can identify top-performing genres and align product strategy to focus on high-rating categories.

▪ Content Expansion:

Gaps in certain game types with fewer titles but strong average scores indicate potential areas for expansion.

- **Sentiment Monitoring:**

Continuous monitoring of monthly review trends helps in identifying the impact of updates, patches, or market competition.

11.3 Excel Dashboard – *User Review Analysis*

- **Overview:**

The User Review Analysis dashboard focuses on understanding user sentiment and engagement by analysing review volume and helpfulness across various apps and games. This dashboard delivers critical insight into how users interact with mobile products based on their feedback volume and perceived helpfulness.

- **Key Visualizations:**

- **Apps vs Games – Helpful Count & Total Reviews**

This horizontal bar chart compares total reviews and average helpful counts for both apps and games. It highlights which products not only gather more reviews but also provide helpful user experiences, enabling businesses to assess user satisfaction qualitatively.

- **Total Review Volume (Apps)**

A vertical bar chart showcasing the total number of reviews for top apps. This helps identify which applications are receiving the most feedback, indicating higher user engagement and broader reach.

- **Total Review Volume (Games)**

A donut chart representing top games based on review volume. This visualization allows a quick glance at which games are actively receiving attention and evaluation from players, useful for competitive benchmarking.

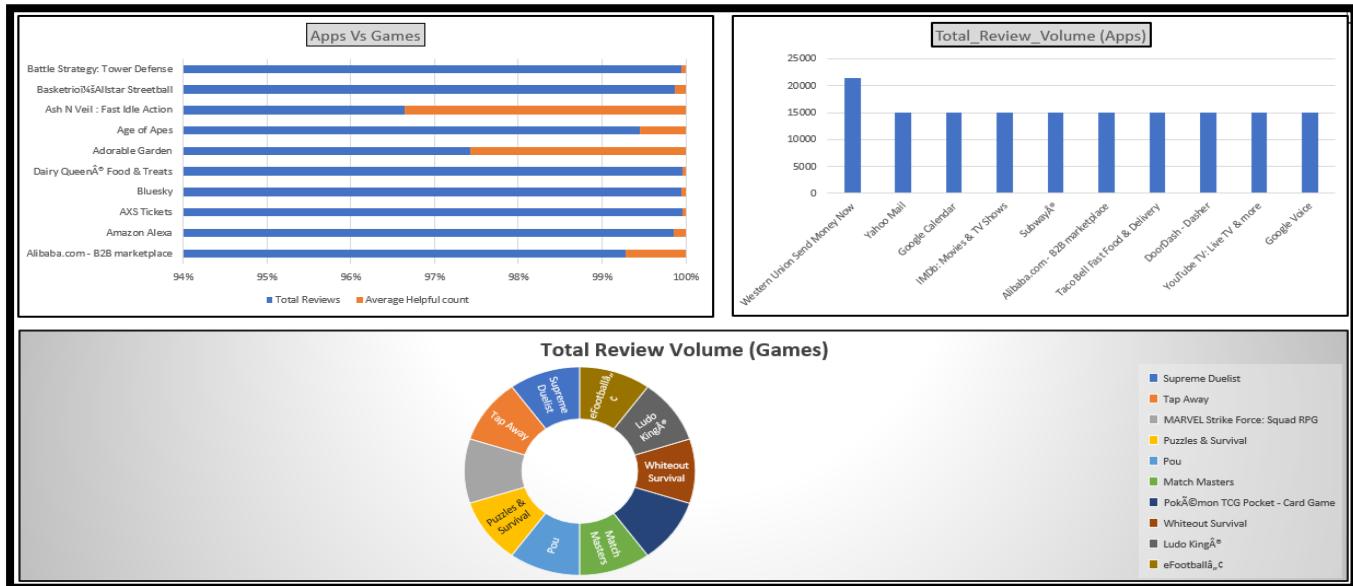


Figure 13

- **Insights Gained:**

- Identified the most reviewed apps and games, useful for prioritizing competitive analysis.
- Assessed which platforms users find most helpful based on helpfulness ratings.
- Recognized review engagement patterns that can be aligned with future marketing or development focus.

11.4 Dynamic Data Visualization

To enhance user interactivity and analytical depth, the project integrates dynamic Excel dashboards driven by dropdown filters. These interactive charts allow users to customize views based on selected categories, sections, or review months, enabling on-the-fly comparative analysis.

- **Dynamic Components Implemented:**

- **App Performance by Category:**

- ✓ A dropdown enables users to select specific app categories (e.g., Entertainment, Finance, Productivity).
 - ✓ Bar charts dynamically reflect download volumes and performance metrics based on the selected category.

- **Games - Downloads vs Average Score:**

- ✓ Section-wise dropdown filters visualize the relationship between game downloads and their average user rating.
 - ✓ Users can compare engagement trends across genres like Casual Games, Indie Corner, or Explore on PC.

- **App Review Scores by Section:**

- ✓ Users select a section (e.g., Music & Audio), which displays the average review score for apps in that section.
 - ✓ Helps identify high-performing apps within specific domains.

- **Monthly Review Volume Analysis:**

- ✓ Dropdown to filter reviews by specific months.
 - ✓ A horizontal bar chart shows which apps received the most feedback in a selected month, helping to uncover peak engagement windows.

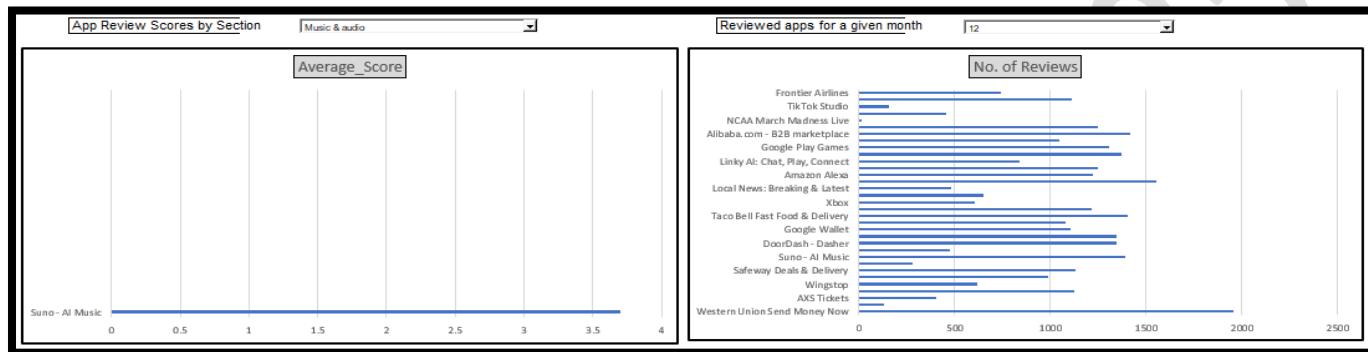
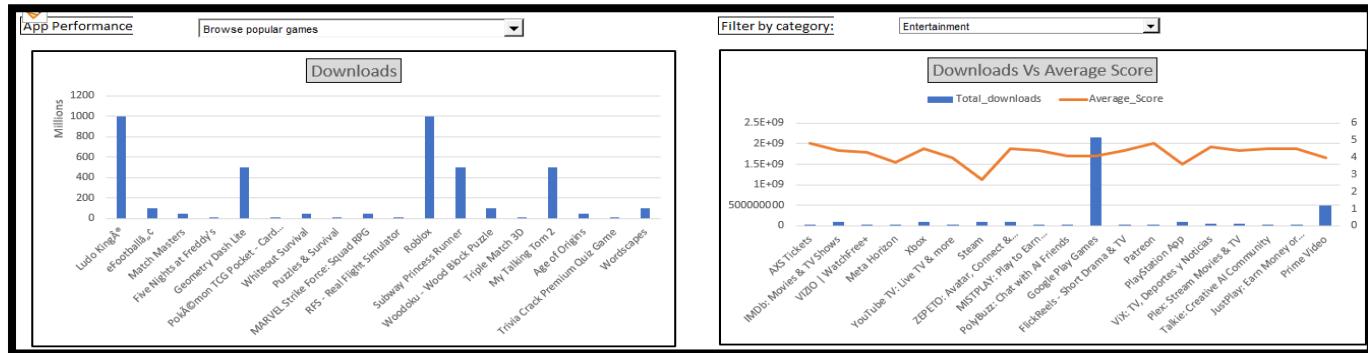


Figure 14 & 15

- **Business Impact:**

These dynamic dashboards empower decision-makers to:

- ✓ Drill down into specific market segments (e.g., high-performing categories or newly launched games).
- ✓ Compare user sentiment across time or between sections.
- ✓ Identify apps or games with peak engagement during certain months or within selected verticals.
- ✓ Make data-driven product decisions based on user preferences and feedback trends.

By offering this level of interactivity, the dashboard not only improves usability but also maximizes the analytical depth and clarity for various business and analytical users.

- **Insights Gained**

The implementation of dynamic visualizations allowed for granular insights into app and game trends, some of which include:

- ✓ **Entertainment and Communication apps** consistently showed higher download volumes, reflecting strong user demand and broader market appeal.
- ✓ **Stylized Games and Casual Games** emerged as the most numerous and highly rated game types, suggesting their popularity across diverse user demographics.
- ✓ **Newly launched games** often had fewer reviews but competitive average scores, pointing to strong early user satisfaction.
- ✓ The **monthly review volume trend** revealed peaks around certain months, possibly aligned with app updates, events, or seasonal user activity.
- ✓ In some cases, apps had **high download counts but lower average ratings**, indicating usability or satisfaction issues that may need attention.
- ✓ Sections like **Productivity and Food & Drink** displayed fewer downloads but higher average scores, suggesting niche user satisfaction and loyalty.
- ✓ Review analysis by section helped identify **hidden high performers** like "Suno – AI Music," which had high review scores despite a narrow category.

12 Unleashing Data Through Visual Insights

In an era driven by digital consumption, understanding user behaviour, market dynamics, and app performance requires more than raw numbers — it demands clarity through visualization. This project, **Play Market 2025**, transforms complex datasets related to mobile applications and games into clear, interactive dashboards. By leveraging tools like Power BI, Excel, SQL, and Python, we unlock hidden patterns in downloads, user reviews, category trends, and app engagement. These visual insights empower stakeholders to make informed, data-driven decisions in a rapidly evolving digital ecosystem.

12.1 Landing Page

○ Overview

The Landing Page serves as the introductory panel for the "**Play Market 2025**" Power BI dashboard, offering users a snapshot of the entire dataset's scale and composition. This page integrates an **About Dataset** section alongside a suite of compact KPI cards and visuals to quickly orient the viewer with the structure, volume, and key metrics of the app market landscape.

The visual components featured include:

- ✓ **Total Downloads by Type (App vs Game)** – showcasing overall user engagement.
- ✓ **Total Apps and Total Games** – breaking down the product distribution.
- ✓ **Total Score by Type** – visualizing user review engagement.
- ✓ **Overall Downloads and Average App Rating** – representing platform-wide adoption and sentiment.

This high-level view acts as a **dashboard summary**, enabling stakeholders to instantly understand the dataset's magnitude before diving into deeper analytical layers.

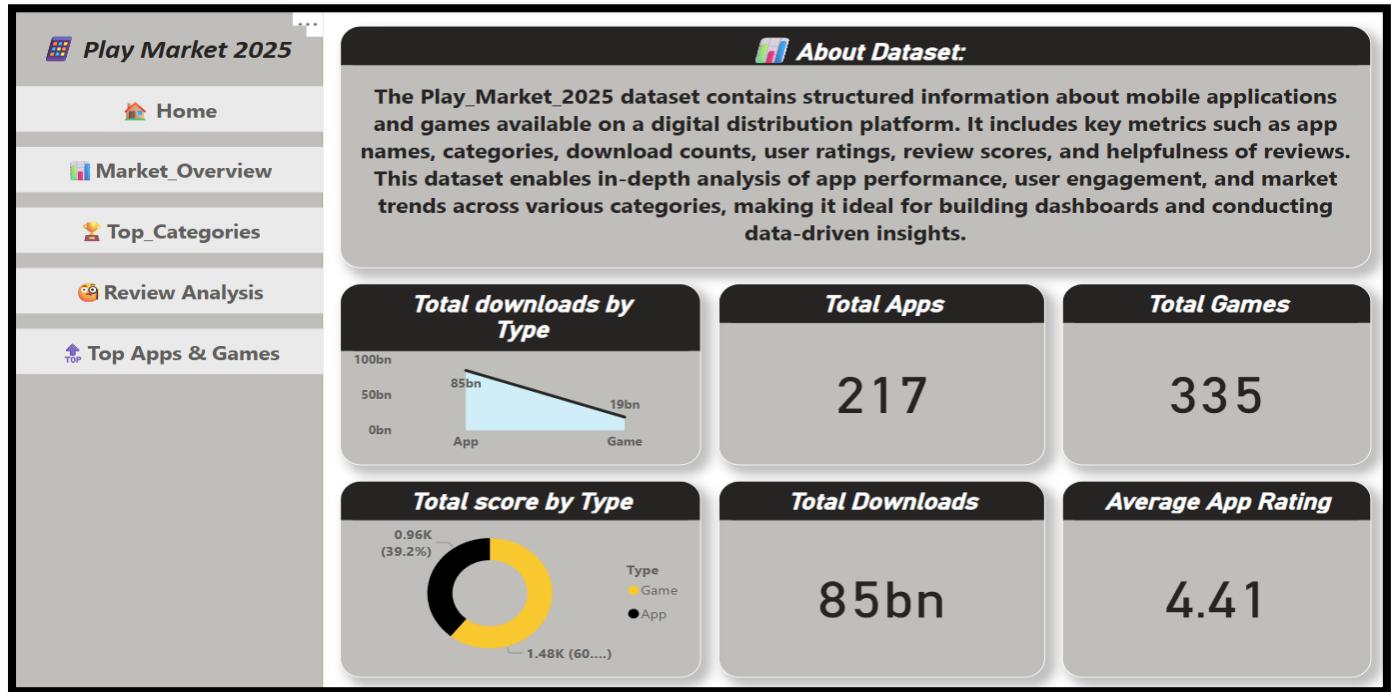


Figure 16

- **Insights Gained**

- ✓ **Market Volume & Distribution:** The platform hosts a significant number of mobile applications, with Apps (217) and Games (335), indicating a healthy and diversified digital ecosystem.
- ✓ **User Engagement through Downloads:** The total download count is exceptionally high, with Apps accumulating 85 billion and Games contributing 19 billion, highlighting apps as the dominant force in terms of user acquisition.
- ✓ **User Feedback & Review Activity:** The Total Score by Type donut chart reflects that both Apps and Games receive substantial review attention, offering ample data for deeper sentiment and engagement analysis.
- ✓ **Performance Snapshot:** The Average App Rating of 4.41 suggests that the quality of applications, on average, meets or exceeds user expectations, a key indicator of app store health.

This Landing Page not only welcomes users into the dashboard experience but also establishes a strong context by quantifying the breadth and engagement level of the dataset. It enables users—whether analysts, business strategists, or developers—to make informed decisions about which areas to explore in greater detail.

12.2 Overall Market

○ Overview

This page provides a comprehensive summary of the digital app market by showcasing download trends and user rating patterns across both applications and games. It offers a high-level view of how different apps perform in terms of total downloads, user preferences by type (App/Game), and rating comparisons. The layout includes interactive slicers for flexible filtering, allowing users to drill down into performance data by app name, type, or section.

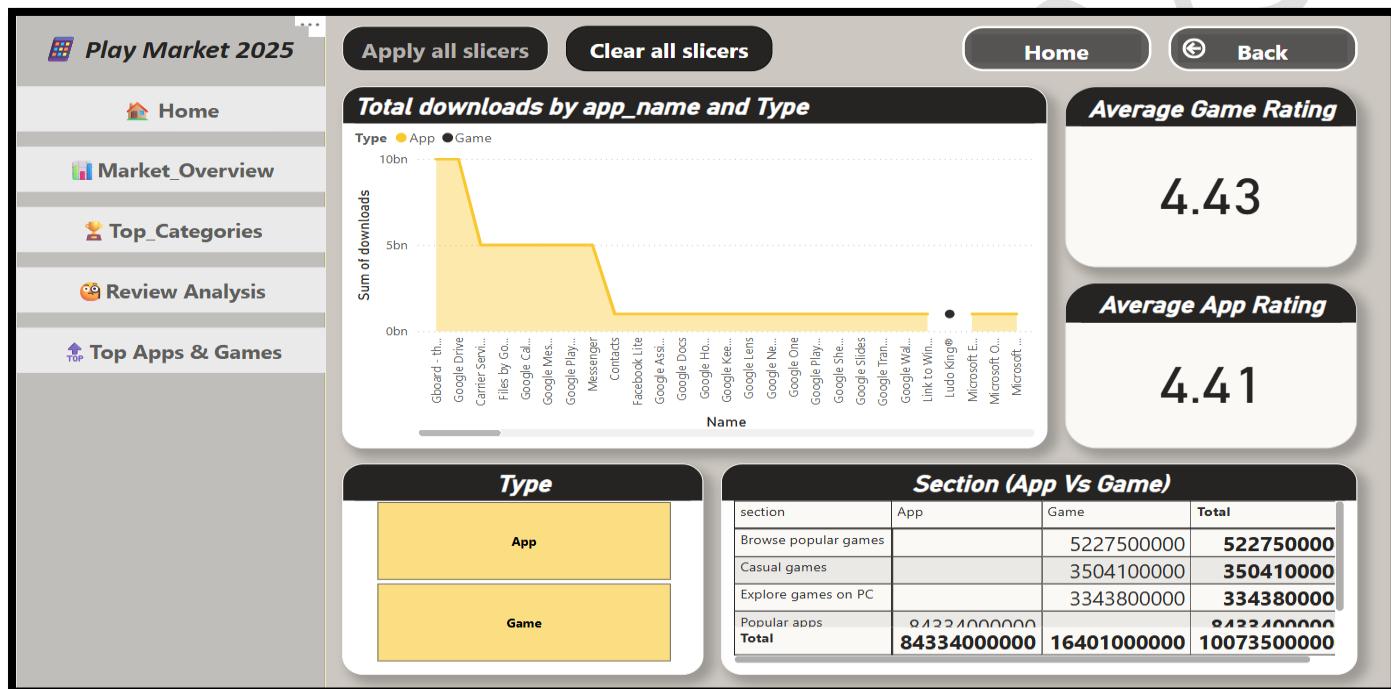


Figure 17

○ Insights Gained

- ✓ **Top Downloaded Apps:** The area chart highlights the most downloaded apps, clearly identifying leading names and their download volumes.
- ✓ **Rating Comparison:** Average ratings for both apps and games are prominently displayed, reflecting user satisfaction and experience across categories.
- ✓ **Type-wise Performance:** The "Type" slicer lets users compare app vs. game performance visually and interactively.
- ✓ **Section Breakdown:** The matrix table segregates downloads by section, giving a side-by-side view of how apps and games are distributed across categories like "Browse popular games", "Casual games", and "Popular apps".

This dashboard tab acts as a market pulse—offering decision-makers a quick but rich insight into user demand and content performance across the ecosystem.

12.3 Top Categories by Popularity

- **Overview**

This section of the dashboard focuses on analysing app and game categories to uncover which segments dominate in user interest, downloads, and satisfaction. It allows stakeholders to understand which categories are not only the most downloaded but also the highest rated—helping product strategists, marketers, and developers pinpoint growth areas.

Interactive slicers for **Content Rating** and **Section** let users filter data dynamically and gain deeper context into category-level trends across different user age groups and usage types.

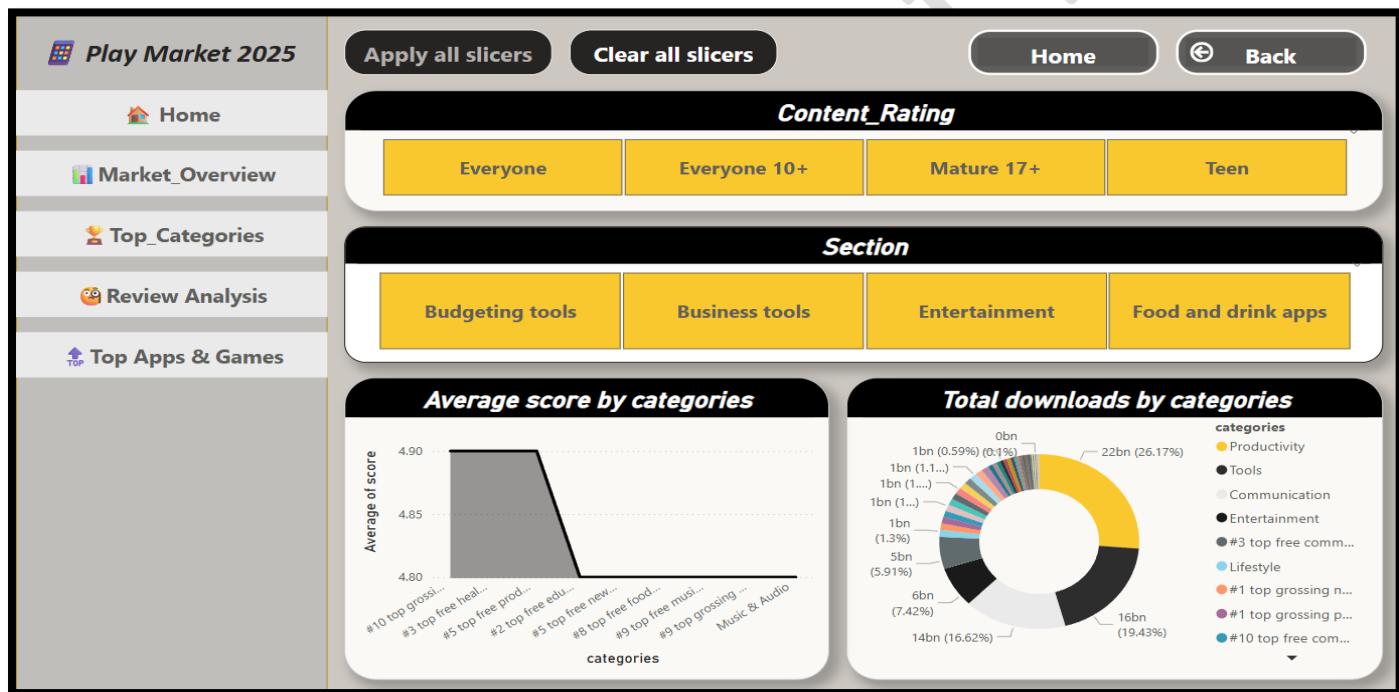


Figure 18

- **Insights Gained**

- ✓ **Category Download Trends:**

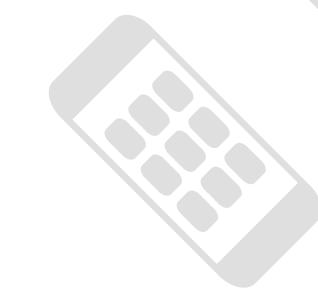
The donut chart shows that **Productivity**, **Tools**, and **Communication** categories lead the market with the highest total downloads—indicating their essential nature in users' daily lives.

✓ **User Satisfaction by Category:**

The average score chart highlights that categories like **#10 Top Grossing**, **#3 Top Free Health**, and **#5 Top Free Productivity** receive consistently higher user ratings, suggesting they offer better user experience and meet expectations more effectively.

✓ **Content Ratings Comparison:**

With filters like **Everyone**, **Mature 17+**, and **Teen**, users can explore how content targeting impacts ratings and downloads, revealing the influence of age-appropriate design and functionality.



12.4 App/Category Review Analysis

○ Overview

This page focuses on understanding user sentiment and engagement by analyzing **review scores** and **rating counts** across apps and games. It provides a quantitative look into how users respond to applications in terms of helpful feedback and satisfaction levels.

The visualizations are structured to give a comparative view by **app type (App vs Game)**, allowing users to instantly grasp the scale and depth of user interactions via reviews.

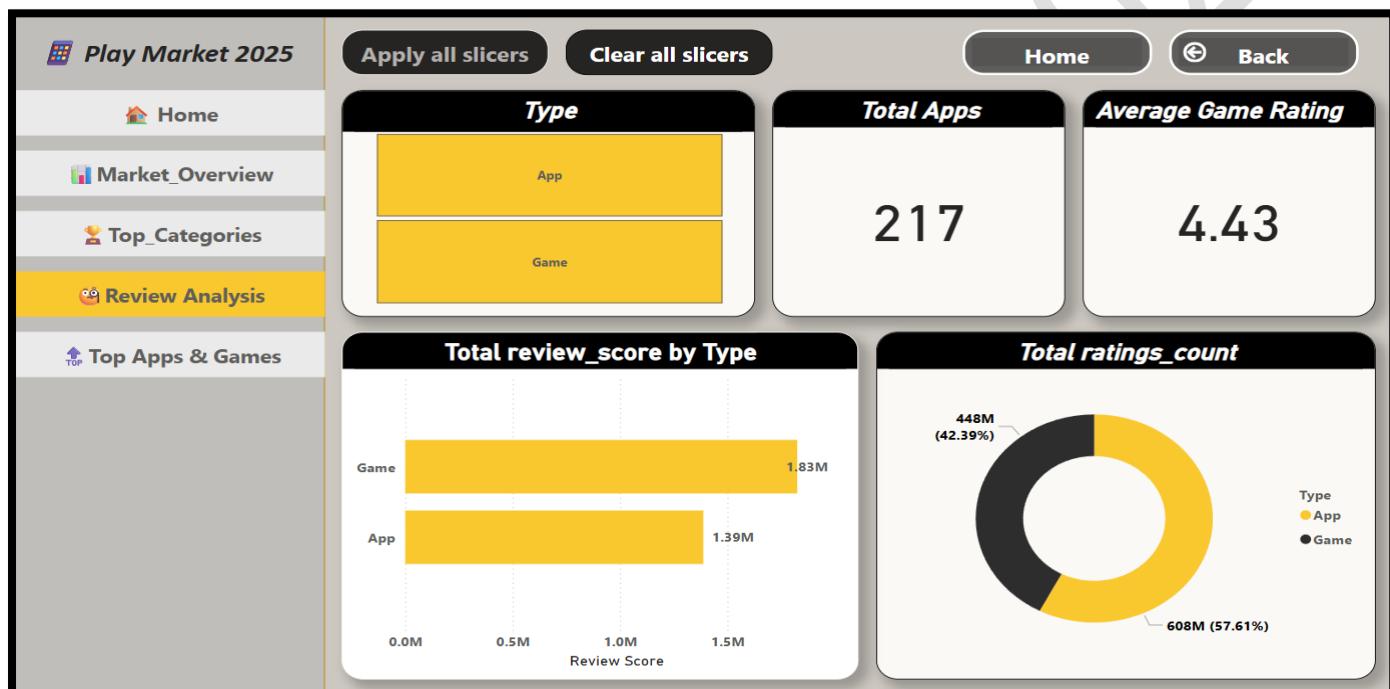


Figure 19

○ Insights Gained:

- ✓ **Volume of Ratings:** Apps account for **57.61%** of total ratings (608 million), while games represent **42.39%** (448 million). This suggests that although games are fewer in number, they still receive significant user attention and engagement.
- ✓ **Total Review Score Comparison:** Despite fewer reviews overall, **games** have a higher cumulative review score (**1.83M**) than apps (**1.39M**). This indicates that users are more expressive or enthusiastic when rating games, possibly due to deeper user experiences or expectations.

- ✓ **App vs Game Review Intensity:** The dashboards reveal that the **engagement per item is higher for games**, meaning users tend to rate and review games more intensely than utility-focused apps.
- ✓ **Rating Quality Distribution:** With average ratings for both games (4.43) and apps (4.41) being very close, it can be inferred that both types deliver satisfactory experiences, though games slightly edge out apps in user satisfaction.

This page equips decision-makers with a better understanding of how users perceive different content types, aiding efforts in improving app features, refining marketing strategies, and enhancing user satisfaction.

12.5 Top Apps & Games

○ Overview

This section highlights the **top-performing mobile apps and games** on the platform using a multi-dimensional approach. It blends metrics such as **download volumes**, **user ratings**, and **review helpfulness** to surface standout performers in the Play Market 2025 dataset.

The layout is designed for **interactive filtering**—users can slice the data by app type, section, or individual app names, enabling targeted analysis across various performance dimensions.

○ Key Components & Visuals:

- ✓ **Top 10 Apps by Downloads:** Displays the most downloaded apps, with names like *Gboard*, *Google Drive*, and *Carrier Services* leading the chart with **10 billion+ downloads**, showcasing their widespread usage and utility.
- ✓ **Downloads vs Rating Comparison:** A scatter plot maps **average rating counts** against **download volumes**, helping identify apps with both popularity and strong user sentiment.
- ✓ **Score by Type (Game vs App):** Reveals that **games outperform apps** in terms of review score metrics—despite having fewer downloads, games gather deeper user engagement.
- ✓ **Ratings Count by Section:** The “**Popular Apps**” section alone accounts for a massive **604 million ratings (57%)**, emphasizing user focus on mainstream apps.
- ✓ **Helpful Score vs Total Review Score by App Name:** Offers granular insights into user feedback quality, not just volume. Apps like *Alibaba.com* and *Google Calendar* show high helpfulness scores, indicating trusted and valued reviews.

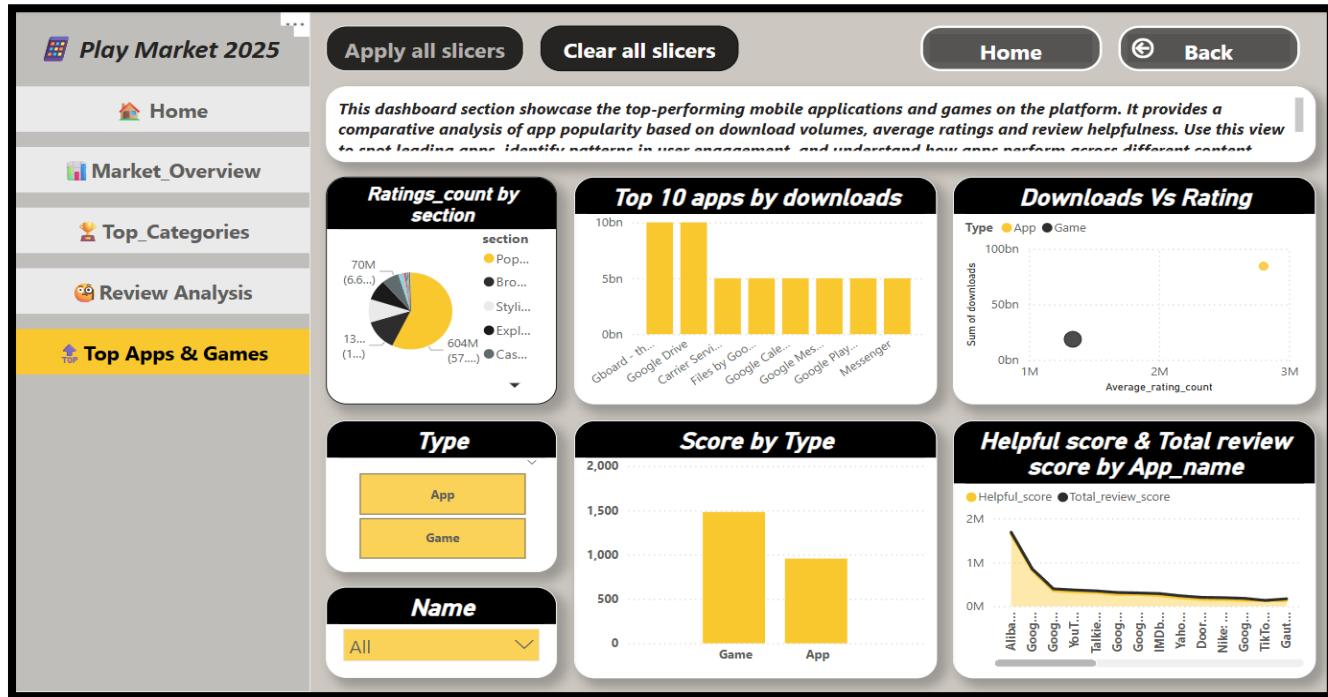


Figure 20

- **Insights Gained:**

- ✓ **Engagement is not just about downloads**—apps like *Google Calendar* and *YouTube* have strong helpful score trends, indicating meaningful user interactions.
- ✓ **Games show stronger review intensity** despite lower install bases, reflecting passionate user bases or content depth.
- ✓ **“Popular” and “Casual” sections dominate rating volumes**, reinforcing their critical role in platform strategy and marketing focus.
- ✓ **Top-downloaded apps align with utility and system functions**, whereas high-score apps often involve engagement-heavy experiences (e.g., games, media).

This page empowers product analysts, developers, and marketers to benchmark app performance, uncover user behaviour patterns, and strategize improvements based on what drives ratings, engagement, and trust on the platform.

13 Discussion

- The **Play Market 2025** Power BI dashboard offers an interactive and visual exploration of mobile applications and games, incorporating metrics like downloads, ratings, review scores, and helpfulness count.
- Through the use of dynamic slicers and segmented pages, the dashboard allows users to analyse app performance across different types (App vs Game), content ratings, and categories with ease.
- **Apps** hold a significantly larger share of total downloads, while **games** exhibit higher engagement through better average review scores and helpful ratings—indicating a more emotionally invested user base for gaming content.
- Categories such as **Productivity** and **Entertainment** dominate in terms of download volumes, while sections like **Popular Apps** gather the highest rating counts—highlighting their broad user appeal and high engagement.
- The **Top Apps & Games** view gives a focused comparison of the leading apps by download volume and review quality, helping identify market leaders and areas of opportunity.
- Insights from this dashboard can assist stakeholders in identifying trends, optimizing app development strategies, improving user experience, and effectively positioning products in the market.

14 Future Work / Potential

- **Incorporate Real-Time Data:** Integrating real-time or periodically refreshed data sources can keep the dashboard updated with the latest trends, ensuring timely decision-making.
- **User Segmentation:** Introduce filters or views based on user demographics (age, region, etc.) to understand behaviour patterns across different audience groups.
- **Sentiment Analysis:** Apply natural language processing (NLP) to review texts for sentiment scoring, providing deeper insight into user feedback beyond numeric ratings.
- **Competitor Benchmarking:** Expand the dataset to include competing platforms or apps to perform comparative market analysis.
- **Churn & Retention Metrics:** Introduce user lifecycle metrics such as churn rate, retention rate, and session duration for more advanced behavioural analysis.
- **Predictive Analytics:** Implement machine learning models to predict future download volumes, rating trends, or app success based on historical patterns.
- **Mobile Compatibility:** Optimize the dashboard layout for mobile and tablet views to enhance accessibility for on-the-go stakeholders.
- **Monetization Insights:** Add financial KPIs such as in-app purchases,

15 Conclusion

- The Play Market 2025 project showcases how data visualization can simplify and elevate the understanding of large-scale mobile app data.
- By using Power BI, combined with Python, SQL, and Excel, the project delivers dynamic dashboards that break down performance metrics across apps and games.
- The analysis covers key dimensions such as:
 - ✓ Downloads
 - ✓ Ratings and Review Scores
 - ✓ Category and Section Popularity
 - ✓ User Engagement Patterns
- Each dashboard tab serves a unique analytical purpose—from high-level market overview to in-depth review and category analysis.
- The interactive and filter-enabled visuals empower stakeholders to explore data intuitively and extract relevant insights on demand.
- This work emphasizes the value of business intelligence tools in supporting strategic planning, product optimization, and market understanding.
- The project lays a solid groundwork for future integration of advanced analytics, including predictive modelling and real-time tracking.

16 Analyst Profile

- **Name:** Anurag Yadav
- **Role:** Data Analyst | Business Intelligence Developer
- **Technical Skills:** Power BI, DAX, SQL, MySQL, Python (pandas, numpy, seaborn, streamlit), Excel, Data Cleaning, Data Modelling, Data Visualization
- **Industry Experience:** 7+ years in the IT industry with hands-on experience in data analysis, reporting, and visualization solutions.
- **Project Focus:** Skilled in transforming raw data into actionable insights through interactive dashboards and reports. Passionate about uncovering patterns in user behaviour and improving decision-making through data.
- **Tools Used in This Project:** Power BI Desktop, DAX, Power Query
- **Professional Interests:** Data storytelling, user engagement analytics, dashboard design, predictive analytics, and music data analysis.
