Video Game Sales & User Engagement Analysis

Project Overview

This project analyses video game data to understand:

- Business performance (Sales trends, platform dominance, regional revenue)
- User engagement (Ratings, wishlist trends, genre popularity)
- Market behaviour over time
- Relationship between ratings and sales

The project integrates **Python, SQL, and Power BI** to create a structured and interactive analytics solution.

Data Sources

The project uses two primary datasets:

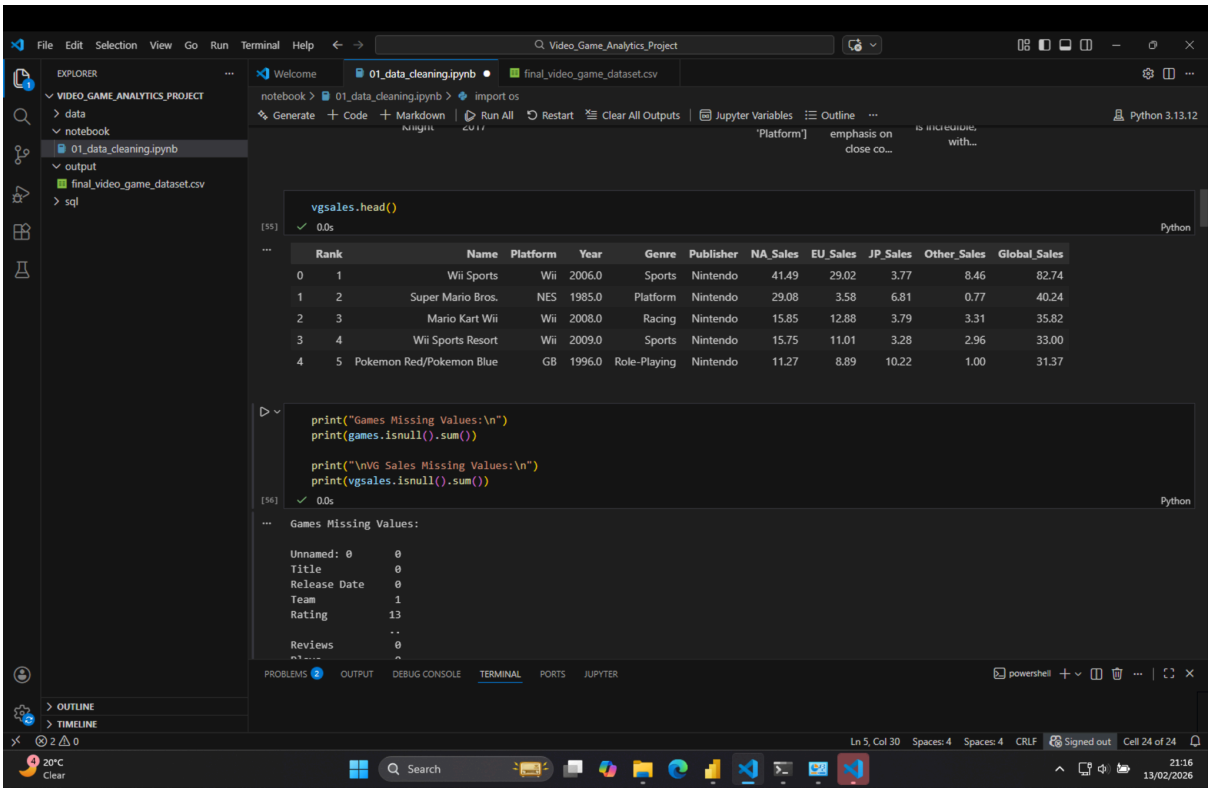
1. games.csv

- Game ID
- Title
- Release Year
- Team / Developer
- Rating
- Plays
- Backlogs
- Wishlist
- Genres

2. vgsales.csv

- Game ID
- Platform
- Publisher
- NA Sales
- EU Sales
- JP Sales
- Other Sales
- Global Sales

Data Cleaning (Python Preprocessing)



Data preprocessing was performed using **Python (Pandas)**.

Steps Performed:

1. Removed Duplicates

```
df = df.drop_duplicates()
```

2. Handled Missing Values

- Ratings: filled or validated
- Plays, Wishlist, Backlogs: converted to numeric
- Null sales replaced with 0 where appropriate

```
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')
df.fillna(0, inplace=True)
```

3. Standardised Formats

- Release year converted to an integer
- Genre names cleaned (trimmed spaces)
- Platform & Publisher names normalised

```
df['release_year'] = df['release_year'].astype(int)
df['genres'] = df['genres'].str.strip()
```

4. Merged Datasets

Games and Sales data merged using game_id.

```
merged_df = pd.merge(games_df, sales_df, on='game_id')
```

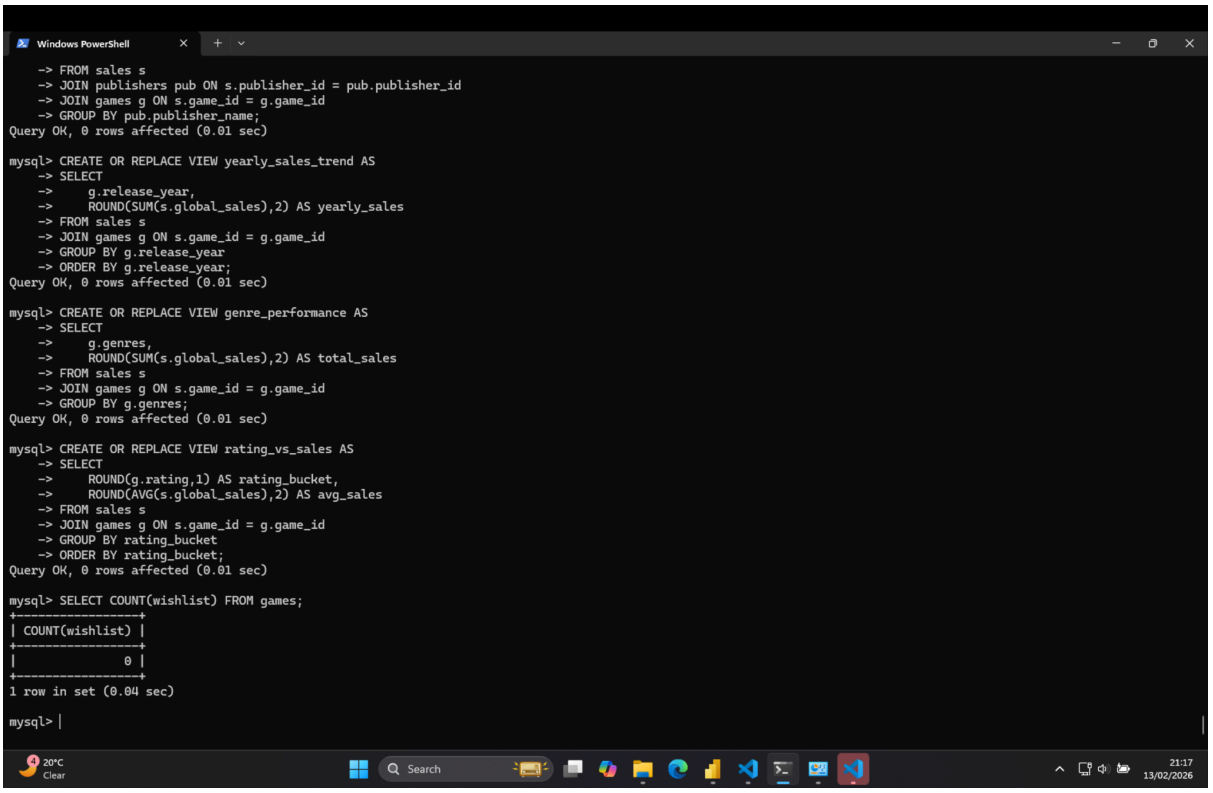
5. Export Clean Data

Cleaned dataset exported for SQL import.

```
merged_df.to_csv("final_video_game_dataset_clean.csv", index=False)
```

SQL Database Setup

Database: **MySQL**



Tables Created:

1. games

- game_id (Primary Key)
- title
- release_year
- rating
- plays
- backlogs

- wishlist
- genres

2. sales

- sale_id (Primary Key)
- game_id (Foreign Key)
- platform_id
- publisher_id
- na_sales
- eu_sales
- jp_sales
- other_sales
- global_sales

3. platforms

- platform_id (Primary Key)
- platform_name

4. publishers

- publisher_id (Primary Key)
- publisher_name

Foreign Key Enforcement

```
ALTER TABLE sales
ADD CONSTRAINT fk_game
FOREIGN KEY (game_id)
```

```
REFERENCES games(game_id);
```

Ensures referential integrity between metadata and sales tables.

Important SQL Queries Used

1. Total Global Sales

```
SELECT SUM(global_sales) AS total_global_sales
FROM sales;
```

2. Sales by Year

```
SELECT release_year, SUM(global_sales) AS yearly_sales
FROM games g
JOIN sales s ON g.game_id = s.game_id
GROUP BY release_year
ORDER BY release_year;
```

3. Top 10 Platforms

```
SELECT p.platform_name, SUM(s.global_sales) AS total_sales
FROM sales s
JOIN platforms p ON s.platform_id = p.platform_id
GROUP BY p.platform_name
ORDER BY total_sales DESC
LIMIT 10;
```

4. Top Publishers

```
SELECT pub.publisher_name, SUM(s.global_sales) AS total_sales
```

```
FROM sales s

JOIN publishers pub ON s.publisher_id = pub.publisher_id

GROUP BY pub.publisher_name

ORDER BY total_sales DESC

LIMIT 10;
```

5. Rating vs Sales Analysis

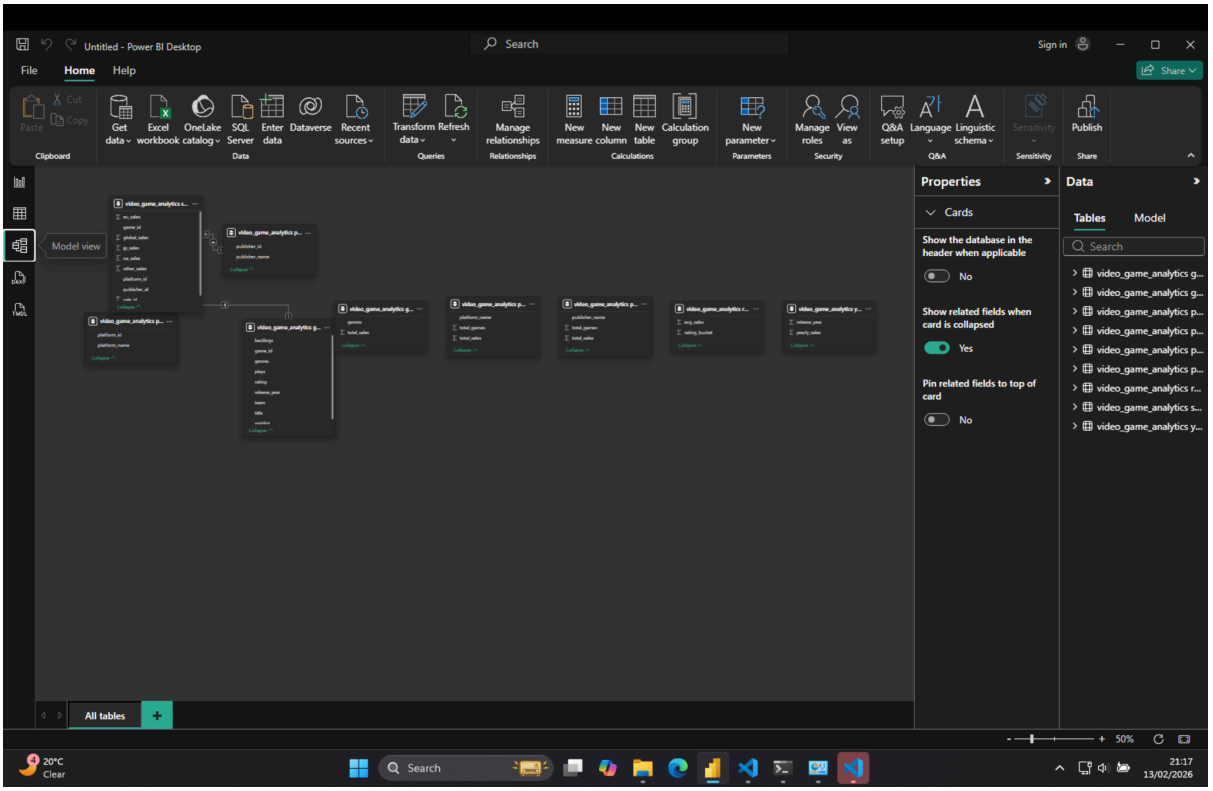
```
SELECT rating, SUM(global_sales) AS total_sales

FROM games g

JOIN sales s ON g.game_id = s.game_id

GROUP BY rating;
```

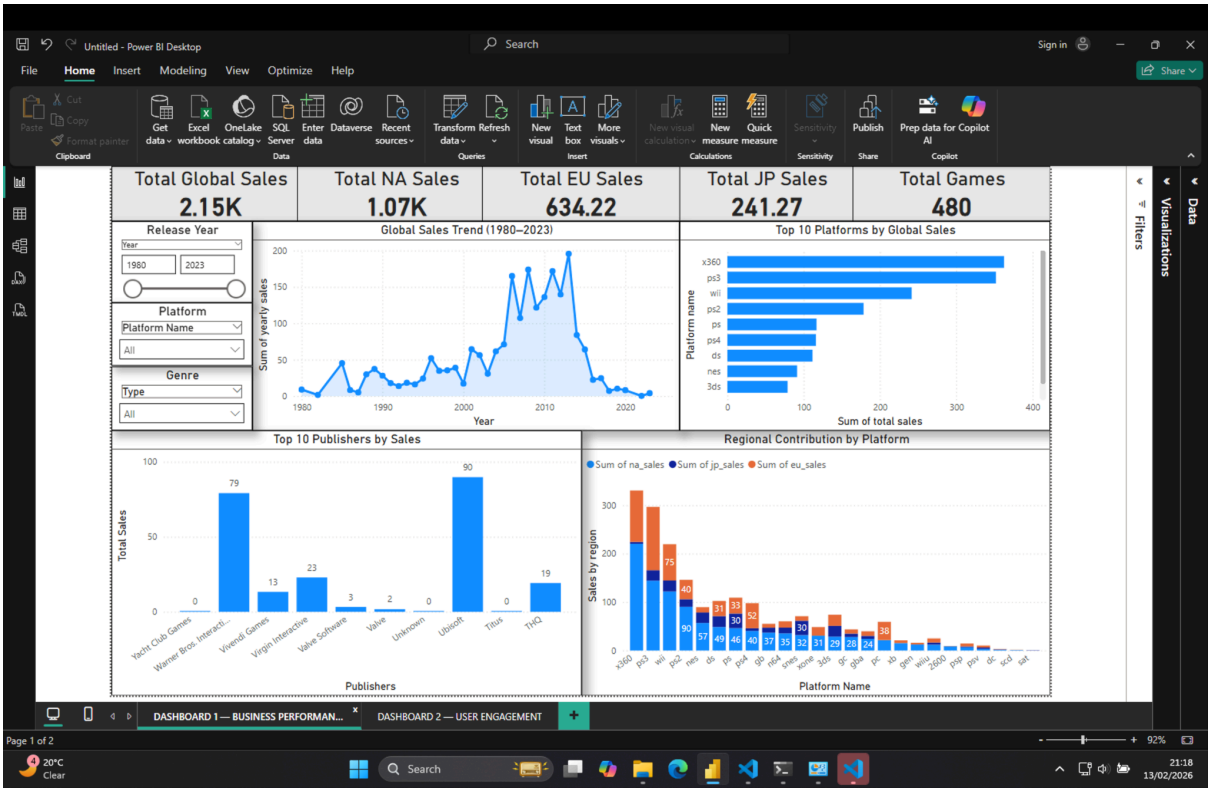
Power BI Integration



Steps:

1. Connected Power BI to MySQL database.
2. Imported structured tables.
3. Created relationships using game_id, platform_id, publisher_id.
4. Built calculated measures:
    - Total Global Sales
    - Average Rating
    - Total Games
5. Applied Top N filters.
6. Added slicers (Year, Platform, Genre).

Dashboard 1 – Business Performance



KPIs:

- Total Global Sales
- NA Sales
- EU Sales
- JP Sales
- Total Games

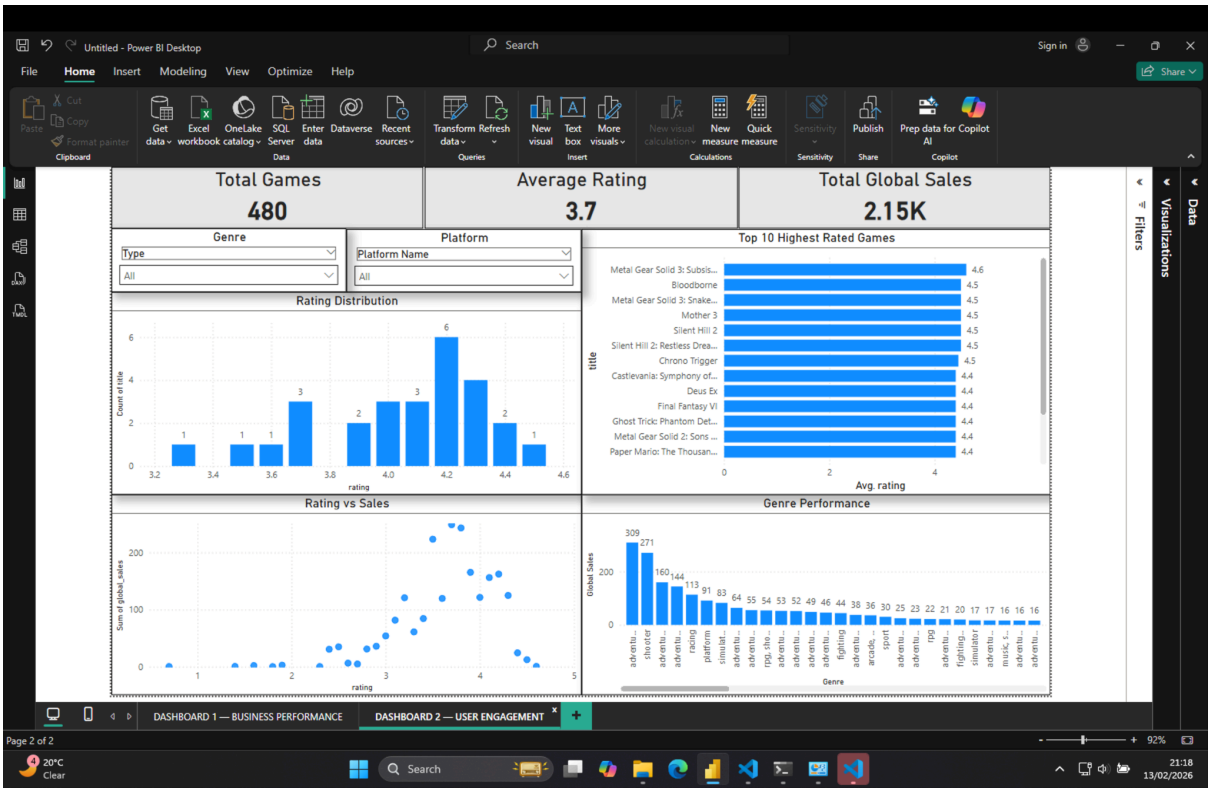
Visuals:

- Global Sales Trend (Line Chart)
- Top 10 Platforms (Bar Chart)
- Top 10 Publishers (Bar Chart)
- Regional Sales by Platform (Stacked Column)

Key Insights:

- Sales peaked between 2008–2012.
- Xbox 360 and PS3 dominate sales.
- North America generates highest revenue.
- Certain publishers consistently outperform others.

Dashboard 2 – User Engagement



KPIs:

- Total Games
- Average Rating
- Total Global Sales

Visuals:

- Rating Distribution
- Top 10 Highest Rated Games
- Rating vs Sales (Scatter Plot)
- Genre Performance

Key Insights:

- Higher-rated games tend to generate higher sales.
- Adventure and Action genres dominate engagement.

- Strong positive correlation between ratings and sales.
- Genre preferences influence commercial success.

Exploratory Data Analysis (EDA) Findings

1. Sales peaked during console boom years.
2. Platform lifecycle impacts sales volume.
3. High ratings moderately correlate with higher revenue.
4. NA region leads market contribution.
5. Certain genres consistently outperform others.

Tools & Technologies Used

- Python (Pandas, NumPy)
- MySQL
- Power BI
- Data Cleaning & Normalization
- Data Visualization
- SQL Aggregation & Joins

Project Outcome

By the end of this project:

- Clean SQL database created
- Structured relational schema implemented

- Two interactive Power BI dashboards developed
- Business insights generated from data
- User engagement patterns identified

Technical Skills Demonstrated

- Data Cleaning
- SQL Normalization
- Foreign Key Design
- Aggregation Queries
- Power BI Dashboard Design
- KPI Creation
- EDA & Insight Generation

📊 Exploratory Data Analysis (EDA)

📁 *games.csv (Game Metadata Only)*

---

1. 🌟 What are the top-rated games by user reviews?

Top-rated games are those with ratings above **4.4**. These include highly acclaimed adventure and RPG titles. The Top 10 Highest Rated Games visual in Dashboard 2 shows the leading titles based on average rating.

---

2. 🧩 What are the most common genres in the dataset?

Adventure, Action, and RPG are the most common genres in the dataset. Adventure appears most frequently, indicating strong representation in the game market.

---

3. 🗓️ What is the game release trend across years?

Game releases increased steadily from the 1990s, peaked between **2008–2012**, and declined after 2015. This trend aligns with the console boom era and is visible in the sales trend analysis.

---

4. 🔎 What is the distribution of user ratings?

Most user ratings fall between **3.5 and 4.3**, showing that the majority of games receive above-average reviews. The rating distribution chart shows a concentration in this range.

💰 *vgsales.csv (Sales Data Only)*

---

5. 🌍 Which region generates the most game sales?

North America generates the highest total game sales, followed by Europe and then Japan. This is clearly visible in the KPI cards and regional sales breakdown chart.

---

6. 🕹️ What are the best-selling platforms?

The best-selling platforms are:

- Xbox 360
- PS3
- Wii
- PS2

These platforms dominate global sales according to the Top 10 Platforms visual.

---

7. 📅 What's the trend of game releases and sales over years?

Global sales peaked between **2008–2012**, followed by a gradual decline. The line chart in Dashboard 1 clearly shows this trend.

---

8. 🏢 Who are the top publishers by sales?

Top publishers include:

- Ubisoft
- Warner Bros
- Nintendo
- Electronic Arts

These publishers consistently generate high global revenue.

---

9. 🧭 How do regional sales compare for specific platforms?

- Xbox platforms perform strongly in North America.
- PlayStation platforms dominate in Europe.
- Nintendo platforms show strong performance in Japan.

The Regional Sales by Platform chart illustrates this comparison.

🔁 *Merged Dataset (Sales + Engagement + Ratings)*

---

10. 🎮 Which game genres generate the most global sales?

Action and Adventure genres generate the highest global sales. These genres dominate revenue across multiple platforms.

---

11. 🎯 How does user rating affect global sales?

The scatter plot shows a **moderate positive correlation** between rating and global sales. Higher-rated games generally tend to generate higher sales.

---

12. 📈 What's the trend of releases and sales over time?

Both releases and sales increased steadily until around 2012, after which the market experienced a gradual decline. This reflects the console lifecycle trend.

---

13. 🎉 What are the top-performing combinations of Genre + Platform?

High-performing combinations include:

- Action + Xbox 360
- Adventure + PS3
- RPG + Nintendo platforms

Certain genres perform better on specific platforms due to audience preference.

🎯 Final Conclusion for Report

- North America is the dominant revenue region.
- Action & Adventure are the strongest revenue genres.
- Ratings positively influence sales performance.
- Console lifecycle directly impacts market revenue.
- Large publishers dominate global revenue.
- The market peaked around 2008–2012.
