**Steam Data Analysis**

**Introduction**

Welcome to my first SQL project, where I dived into the intriguing world of gaming data using MySQL. I employed MySQL to analyze, query, and derive meaningful insights from this gaming dataset. The goal was not only to gain a deeper understanding of the data but also to showcase the power and versatility of SQL in handling real-world scenarios.

**Data Model**

Table: games

Columns:

app\_id (Primary Key, INT): Native product ID on Steam.

title (VARCHAR(255)): Product title.

date\_release (DATE): Product release date.

win (TEXT): Supports Windows?

mac (TEXT): Supports MacOS?

linux (TEXT): Supports Linux?

rating (TEXT): Product rating category.

positive\_ratio (DOUBLE): Ratio of positive feedbacks.

user\_reviews (INT): Number of user reviews.

price\_final (DOUBLE): Price in US dollars $ calculated after the discount.

price\_original (DOUBLE): Original price in US dollars $.

discount (INT): Percentage of discount applied to the original price.

steam\_deck (TEXT): Supports Steam Deck?

Table: users

Columns:

user\_id (Primary Key, INT): User's auto-generated ID.

products (INT): Number of games/add-ons purchased by the user.

reviews (INT): Number of reviews published.

Table: recommendations

Columns:

app\_id (Foreign Key, INT): Native product ID on Steam (References games.app\_id).

funny (INT): How many users found a recommendation funny.

helpful (INT): How many users found a recommendation helpful.

hours (DOUBLE): How many hours played by the user.

is\_recommended (TEXT): Is the user recommending the product?

date (DATE): Date of publishing.

review\_id (INT): Auto-generated ID that was generated for a review.

user\_id (Foreign Key, INT): User's anonymized ID (References users.user\_id).

ERD can be found in ERD.drawio file.

**Data Cleaning**

Date Columns Formatting:

- In preparation for MySQL requirements (YYYY-MM-DD), date columns in Excel were formatted accordingly.

Title Column Cleaning:

- Utilized the CONCAT function in Excel to concatenate and sanitize "title" column fields.

- Quotations were added to encapsulate the values and mitigate potential issues with special characters in SQL scripts.

- The modified values were then displaced back into the "title" column.

Empty Fields Review:

- All Excel columns were reviewed using filters to ensure there are no empty fields. This step contributes to maintaining data integrity.

**MySQL Database Preparation**

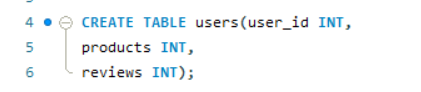
Database and Table Creation:

-- Create a new database

CREATE DATABASE First\_Project;

USE First\_Project;

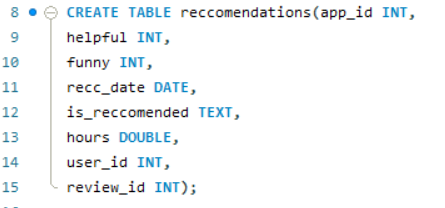
-- Create the 'users' table

CREATE TABLE users (user\_id INT,

products INT,

reviews INT);

-- Create the 'recommendations' table

CREATE TABLE recommendations (app\_id INT,

helpful INT,

funny INT,

rec\_date DATE,

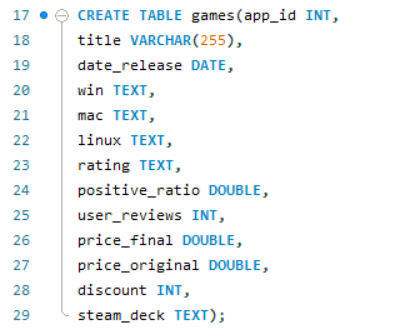
is\_recommended TEXT,

hours DOUBLE,

user\_id INT,

review\_id INT);

-- Create the 'games' table

CREATE TABLE games (app\_id INT,

title VARCHAR(255),

date\_release DATE,

win TEXT,

mac TEXT,

linux TEXT,

rating TEXT,

positive\_ratio DOUBLE,

user\_reviews INT,

price\_final DOUBLE,

price\_original DOUBLE,

discount INT,

steam\_deck TEXT);

-- Load data into the 'recommendations' table

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/recommendations.csv'

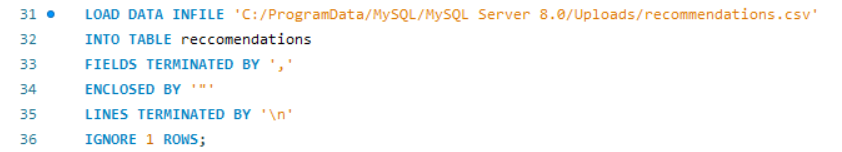
INTO TABLE recommendations

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;



-- Load data into the 'games' table

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/games.csv'

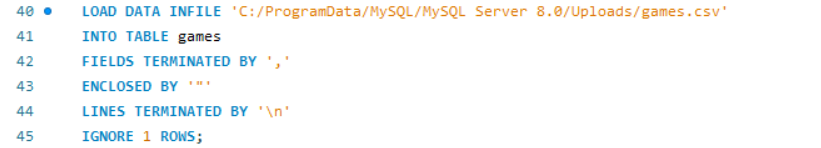
INTO TABLE games

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;



-- Load data into the 'users' table

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users.csv'

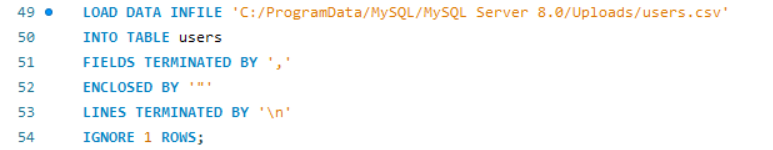
INTO TABLE users

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;



-- Verify data in the 'recommendations' table

SELECT \* FROM recommendations;

-- Verify data in the 'games' table

SELECT \* FROM games;

-- Verify data in the 'users' table

SELECT \* FROM users;

**KPI’s and Query Examples**

1. Release Year Analysis:

Examining historical data, I assessed the distribution of game releases between 2000 and 2022, revealing trends and patterns in the gaming industry over the years.

SELECT YEAR(date\_release) AS release\_year, COUNT(\*) AS num\_releases

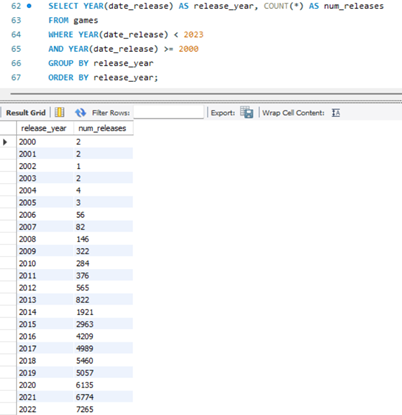
FROM games

WHERE YEAR(date\_release) < 2023

AND YEAR(date\_release) >= 2000

GROUP BY release\_year

ORDER BY release\_year;



2. Platform Adoption Rate:

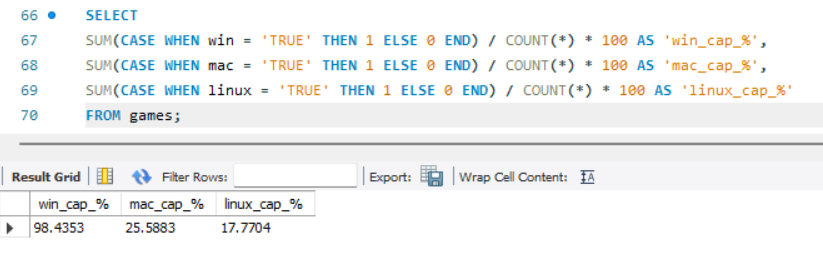
Through the analysis of our dataset, I determined the percentage distribution of games supporting Windows, MacOS, and Linux. This analysis provides insights into prevailing platform preferences among game developers.

SELECT

SUM(CASE WHEN win = 'TRUE' THEN 1 ELSE 0 END) / COUNT(\*) \* 100 AS 'win\_cap\_%',

SUM(CASE WHEN mac = 'TRUE' THEN 1 ELSE 0 END) / COUNT(\*) \* 100 AS 'mac\_cap\_%',

SUM(CASE WHEN linux = 'TRUE' THEN 1 ELSE 0 END) / COUNT(\*) \* 100 AS 'linux\_cap\_%'

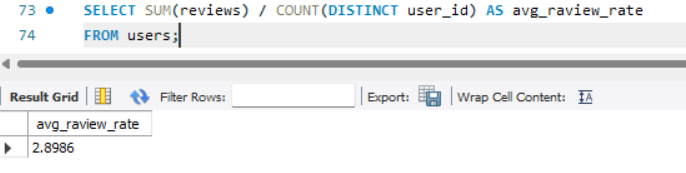
FROM games;

3. User Engagement:

By delving into user profiles data, I calculated the average review rate per user, offering valuable insights into the level of user engagement and feedback within our gaming community.

SELECT SUM(reviews) / COUNT(DISTINCT user\_id) AS avg\_review\_rate

FROM users;



4. Recommendation Effectiveness:

Through analysis of the dataset, I assessed the total positive recommendations, total recommendations, and the resulting recommendation rate. This analysis provides a comprehensive view of the effectiveness of user recommendations within our gaming ecosystem.

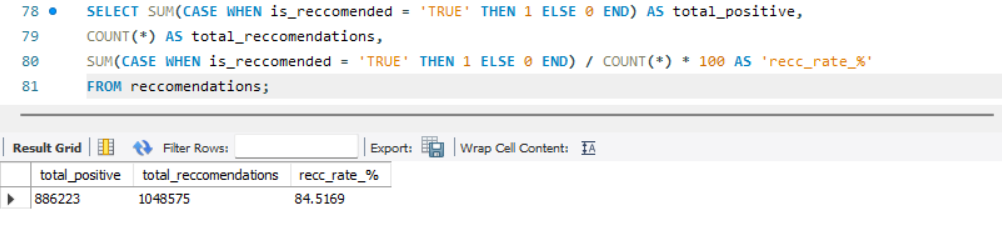
SELECT

SUM(CASE WHEN is\_recommended = 'TRUE' THEN 1 ELSE 0 END) AS total\_positive,

COUNT(\*) AS total\_recommendations,

SUM(CASE WHEN is\_recommended = 'TRUE' THEN 1 ELSE 0 END) / COUNT(\*) \* 100 AS 'recc\_rate\_%'

FROM recommendations;



5. Top Recommended Games by Playtime:

Utilizing recommendation and games data, I identified the top 10 recommended games based on playtime. This information unveils user preferences and highlights the most engaging titles in our gaming platform.

SELECT DISTINCT g.title, r.hours

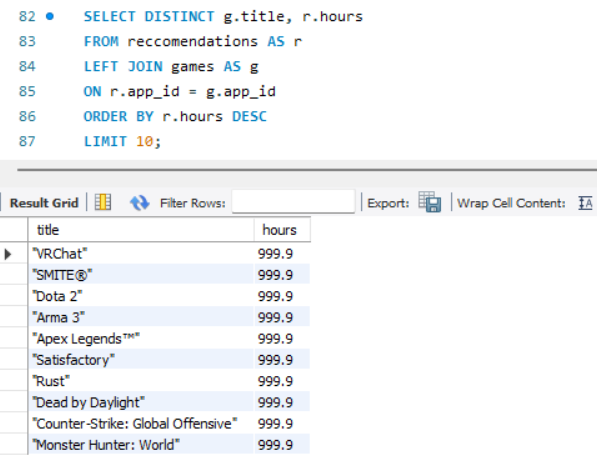
FROM recommendations AS r

LEFT JOIN games AS g

ON r.app\_id = g.app\_id

ORDER BY r.hours DESC

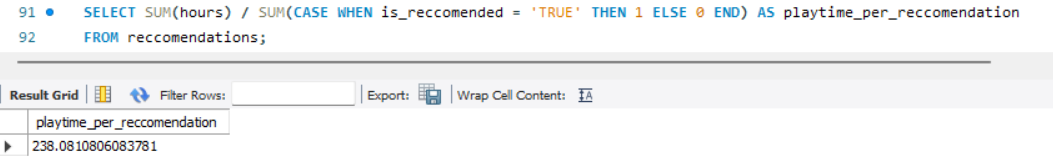
LIMIT 10;



6. Playtime per Recommendation:

By examining playtime data in conjunction with recommendations, I determined the average playtime per recommended game. This metric offers insights into user involvement with recommended titles.

SELECT SUM(hours) / SUM(CASE WHEN is\_recommended = 'TRUE' THEN 1 ELSE 0 END) AS playtime\_per\_recommendation

FROM recommendations;

7. Highly Rated Games Analysis:

Leveraging data on highly-rated games released in 2023 with positive reviews and above-average user engagement, I identified and analyzed successful game launches within the current year.

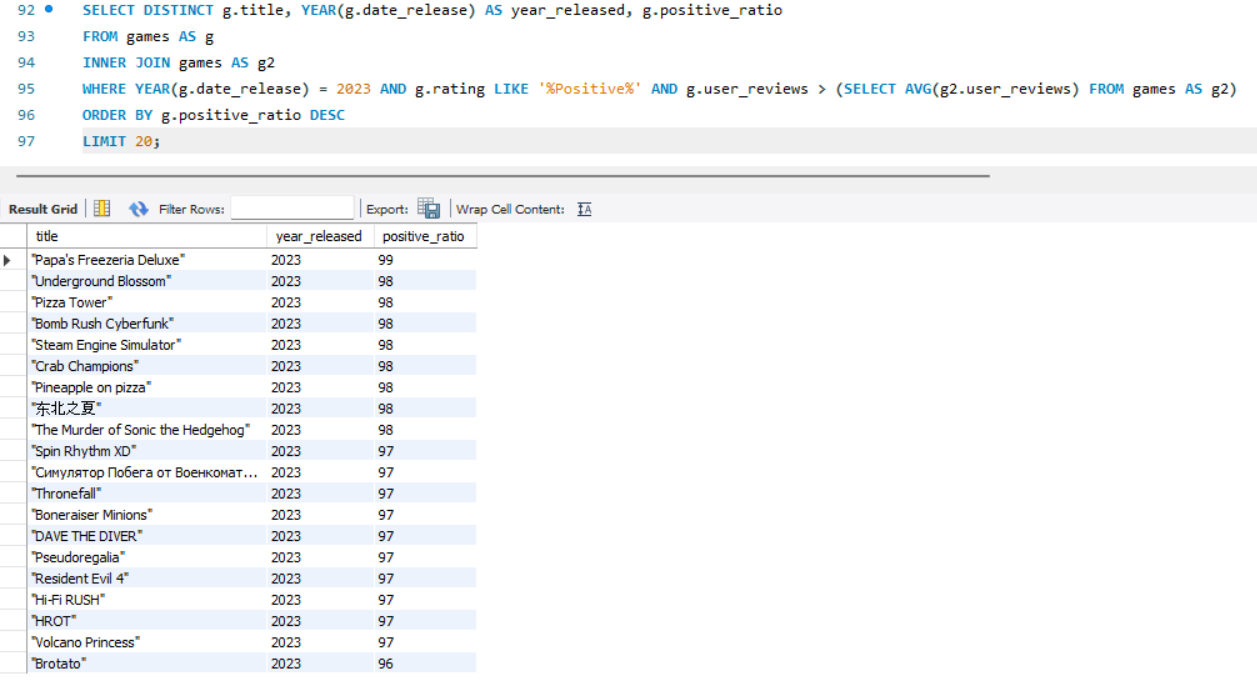
SELECT DISTINCT g.title, YEAR(g.date\_release) AS year\_released, g.positive\_ratio

FROM games AS g

INNER JOIN games AS g2

WHERE YEAR(g.date\_release) = 2023 AND g.rating LIKE '%Positive%' AND g.user\_reviews > (SELECT AVG(g2.user\_reviews) FROM games AS g2)

ORDER BY g.positive\_ratio DESC

LIMIT 20;

**Troubleshooting**

While working on the project, an inconsistency was identified in the naming convention of tables. The original table name "reccomendations" contained a spelling error. To maintain a consistent and accurate schema, the following corrective actions were taken:

ALTER TABLE reccomendations RENAME recommendations;

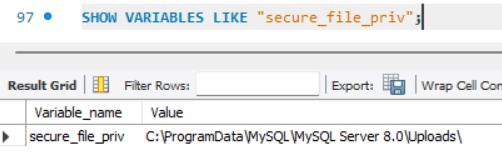
ALTER TABLE recommendations RENAME COLUMN is\_reccomended TO is\_recommended;

ALTER TABLE recommendations RENAME COLUMN recc\_date TO rec\_date;

This adjustment ensures a standardized naming convention across the database, enhancing clarity and consistency in the project documentation. Please ensure that these corrections are applied for a seamless and error-free database schema.

During the execution of the LOAD DATA INFILE command, a security error was encountered, and the following message was received:

“The MySQL server is running with the --secure-file-priv option so it cannot execute this statement”

To address this security restriction, it was necessary to move the Excel sheets to a designated safe folder. The safe folder location was determined using the following SQL command:

SHOW VARIABLES LIKE "secure\_file\_priv";

By relocating the files to the secure folder, the security constraint was addressed, ensuring the seamless execution of the LOAD DATA INFILE command without compromising the server's security settings.

**Future Work**

Power BI Integration:

I am eager to leverage the insights gathered from MySQL, seamlessly integrating them into Power BI. This enhancement promises to deliver more sophisticated visualizations, offering stakeholders a user-friendly interface for exploring the data.

Skill Enhancement:

A dedicated focus on advancing my SQL proficiency is in the pipeline. Anticipate the deployment of more intricate queries and optimized database operations, contributing to a refined and efficient data management process.

Python Exploration:

Looking ahead, there's a deliberate plan to embark on a journey into Python. Beyond coding, this endeavor is poised to broaden the spectrum of possibilities for data manipulation and analysis.

These forthcoming initiatives align with a commitment to elevate the project's capabilities, ensuring a harmonious blend of professionalism and a dynamic approach to data management and analysis.

**References**

<https://www.kaggle.com/datasets/antonkozyriev/game-recommendations-on-steam/data?select=games.csv>

<https://dev.mysql.com/doc/>

<https://chat.openai.com>

<https://app.diagrams.net>

**License**

This project is released under an open and permissive license. You are free to use, modify, and distribute the code as you see fit. There are no copyright restrictions imposed on this project. The aim is to encourage collaboration and innovation, allowing individuals to adapt and benefit from the code in a manner that aligns with their specific needs and goals.

**Contact Information**

Should you have any inquiries, suggestions, or interest in collaboration, feel free to connect with me on [LinkedIn](http://www.linkedin.com/in/nkurlianskas). I welcome the opportunity to engage in meaningful discussions and potential collaborations.