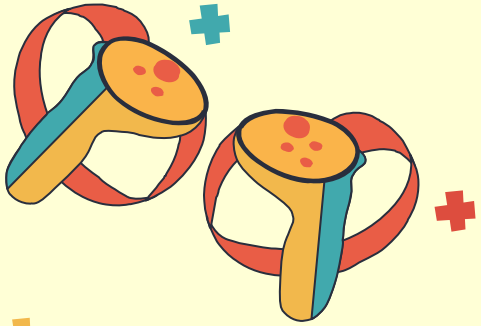
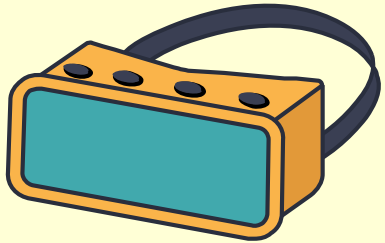
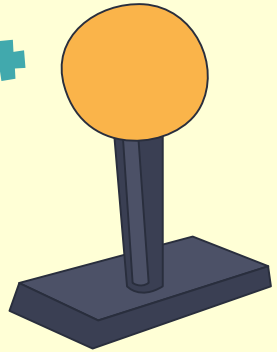
# VIDEO GAMES DATA ANALYSIS REPORT 2024

SQL PROJECT BY ANKIT JADLI







# Purpose of Data Analysis in the Gaming Industry

The gaming industry has been growing rapidly worldwide, and data analysis plays a crucial role in helping companies and developers make better decisions.

Data analysis helps:

- Understand what players want and how they behave, allowing developers to create games that keep players engaged.
- Improve marketing efforts by identifying when and how to reach potential players with the right content.
- Spot trends in game types, platforms, and player preferences, helping developers focus on what's working.
- Enhance player retention and monetization by using insights to improve game features, updates, and overall experiences.

Global Data Highlights (2024 Projections):

- Revenue: The video game industry is expected to generate US$282.30 billion in 2024.
- Growth: The market is projected to grow at an annual rate of 8.76%, reaching US$363.20 billion by 2027.
- Top Market: China is set to lead, with US$94.49 billion in revenue in 2024.
- Average Revenue Per User (ARPU): Globally, each user is projected to contribute US$215.20 in 2024.

Indian Market Data (2024 Projections):

- Revenue: In India, the video game market is expected to reach ₹263.40 billion in 2024.
- Growth: It's predicted to grow at a rate of 12.18% per year, reaching ₹371.80 billion by 2027.
- Key Segment: Online games dominate the market, with a revenue of ₹121.80 billion in 2024.
- ARPU: In India, the average revenue per user is expected to be ₹1.80k in 2024.

Source: Statista

# Why I Chose "Video Game Analysis" As My Project

I've been passionate about video games since I was a kid. They've always been a huge part of my life, from playing ***Counter Strike 1.6*** with friends on school nights to exploring countless ethereal worlds on my own in ***Skyrim***. When I started working on this project, I knew I wanted to focus on something that genuinely excited me — and video games were the obvious choice.

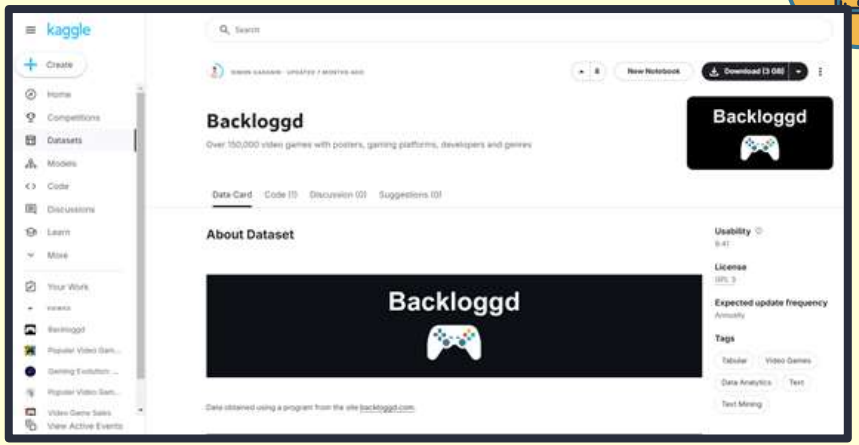
A Growing Industry: The gaming industry is evolving fast, with massive growth year after year. As the industry becomes more competitive, data analysis plays a key role in shaping the future of gaming. Understanding this market from the inside out was an exciting challenge for me.

Curiosity Meets Passion: This project gave me the chance to merge my love for gaming with my interest in data analysis. I wanted to explore how games are made, how they succeed, and where the industry is headed — not just as a fan, but from a business and analytical perspective.

# Data Collection and Preparation

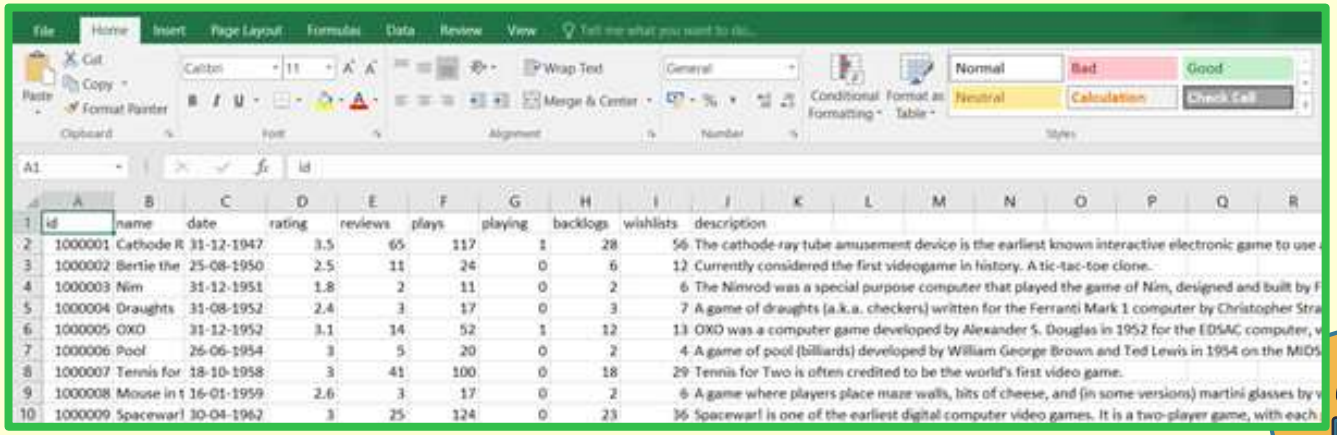
Data Source:

- I found a comprehensive dataset on Kaggle named Backloggd. The dataset provides insights into various video games, including their release dates, genres, developers, and player engagement metrics.




Verification & Data Download:

- I carefully checked the source of the data collection to confirm its credibility. After confirming the dataset's trustworthiness, I downloaded the CSV files to begin my analysis.



| id | name | date | rating | reviews | plays | playing | backlogs | wishlists | description |
|---|---|---|---|---|---|---|---|---|---|
| 1000001 | Cathode R | 31-12-1947 | 3.5 | 65 | 117 | 1 | 28 | 56 | The cathode-ray tube amusement device is the earliest known interactive electronic game to use |
| 1000002 | Bertie the | 25-08-1950 | 2.5 | 11 | 24 | 0 | 6 | 12 | Currently considered the first videogame in history. A tic-tac-toe clone. |
| 1000003 | Nim | 31-12-1951 | 1.8 | 2 | 11 | 0 | 2 | 6 | The Nimrod was a special purpose computer that played the game of Nim, designed and built by |
| 1000004 | Draughts | 31-08-1952 | 2.4 | 3 | 17 | 0 | 3 | 7 | A game of draughts (a.k.a. checkers) written for the Ferranti Mark 1 computer by Christopher Stra |
| 1000005 | OXO | 31-12-1952 | 3.1 | 14 | 52 | 1 | 12 | 13 | OXO was a computer game developed by Alexander S. Douglas in 1952 for the EDSAC computer, |
| 1000006 | Pool | 26-06-1954 | 3 | 5 | 20 | 0 | 2 | 4 | A game of pool (billiards) developed by William George Brown and Ted Lewis in 1954 on the MIDS |
| 1000007 | Tennis for | 18-10-1958 | 3 | 41 | 100 | 0 | 18 | 29 | Tennis for Two is often credited to be the world's first video game. |
| 1000008 | Mouse in t | 16-01-1959 | 2.6 | 3 | 17 | 0 | 2 | 6 | A game where players place maze walls, bits of cheese, and (in some versions) martini glasses by |
| 1000009 | Spacewar! | 30-04-1962 | 3 | 25 | 124 | 0 | 23 | 36 | Spacewar! is one of the earliest digital computer video games. It is a two-player game, with each |

Exploratory Data Analysis (EDA):

- I performed EDA to uncover trends and patterns within the dataset, examining factors such as player engagement, game genres, and release dates.

Data Cleaning:

- Before diving into the analysis, I checked for duplicates, handled empty values, and corrected misspellings in key fields using Excel. I ensured consistency in formatting for dates, ratings, and numeric values, validating data types to maintain data integrity. This thorough cleaning prepared the dataset for accurate analysis and insights.

# Database Setup

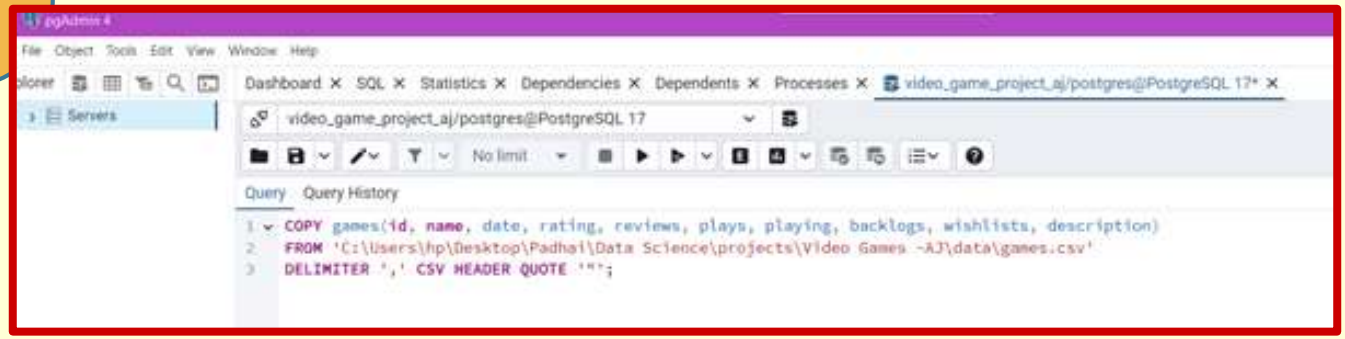
To get started with my project, I downloaded and installed PostgreSQL and PGAdmin, which is a solid tool for managing databases. Once I had it set up, I created multiple tables to organize all the data I had gathered from all the CSV files named Developers, Games, Genres and Platforms respectively.

I made sure to include all the important columns in the tables, such as game names, release dates, ratings, developers, platforms, genres etc. with the “id” acting as the Foreign Key for all tables.


```sql
CREATE TABLE platforms (
  id INT REFERENCES games(id),
  platform VARCHAR(255)
);
```

```sql
CREATE TABLE developers (
  id INT REFERENCES games(id),
  developer VARCHAR(255)
);
```

```sql
CREATE TABLE genres (
  id INT REFERENCES games(id),
  genre VARCHAR(255)
);
```




This structure helped ensure that every bit of information was captured properly and accurately.

After I finished creating the tables, I uploaded the cleaned data from my CSV files into the database using the COPY command.

Setting everything up this way made it much easier for me to access and analyze the data as I moved forward with my project.

# Problem Statements

The following problem statements were created with the assistance of ChatGPT, specifically tailored for an imaginary client seeking an in-depth analysis of the video game industry.

These statements focus on relevant questions that have emerged over the last five years, addressing key trends, player engagement, and market dynamics.

These problem statements were solved using SQL queries, and the outputs provided valuable insights into various aspects of the gaming industry. The analysis helped me understand the gaming landscape on a more integral level, revealing patterns in player behavior, the impact of game releases, and the performance of different platforms and genres.

This deeper understanding has equipped me with the knowledge needed to navigate the complexities of the industry.

# 1. Which year had the highest volume of video game releases?

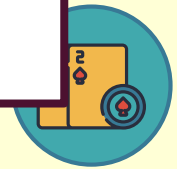
Query | Query History

```
1  SELECT EXTRACT(YEAR FROM date) AS release_year, COUNT(*) AS total_releases
2  FROM games
3  WHERE date IS NOT NULL
4  GROUP BY release_year
5  ORDER BY total_releases DESC
6  LIMIT 1;
```

Data Output | Messages | Notifications

SQL

| | release_year numeric | total_releases bigint |
|---|---|---|
| 1 | 2021 | 10086 |



# 2. What are the most popular release days and months for video games?

Query | Query History

```
1  SELECT EXTRACT(MONTH FROM date) AS release_month,
2  COUNT(*) AS total_releases
3  FROM games
4  WHERE date IS NOT NULL
5  GROUP BY release_month
6  ORDER BY total_releases DESC
7  LIMIT 5;
8
```

Data Output | Messages | Notifications

SQL

| | release_month numeric | total_releases bigint |
|---|---|---|
| 1 | 12 | 25903 |
| 2 | 10 | 12178 |
| 3 | 11 | 11676 |
| 4 | 9 | 10911 |
| 5 | 3 | 10893 |

Query | Query History

```
1  SELECT EXTRACT(DAY FROM date) AS release_day,
2  COUNT(*) AS total_releases
3  FROM games
4  WHERE date IS NOT NULL
5  GROUP BY release_day
6  ORDER BY total_releases DESC
7  LIMIT 5;
8
```

Data Output | Messages | Notifications

SQL

| | release_day numeric | total_releases bigint |
|---|---|---|
| 1 | 31 | 18180 |
| 2 | 1 | 10784 |
| 3 | 30 | 4754 |
| 4 | 28 | 4454 |
| 5 | 15 | 4338 |

# 3. What are the top three most common genres of video games, and how do these genres impact player engagement (total players)?

Query | Query History

```
1  SELECT genre, COUNT(g.id) AS total_games, SUM(plays) AS total_players
2  FROM genres AS ge
3  JOIN games AS g ON ge.id = g.id
4  GROUP BY genre
5  ORDER BY total_games DESC
6  LIMIT 3;
7
```

Data Output | Messages | Notifications

SQL

| | genre character varying (99) | total_games bigint | total_players bigint |
|---|---|---|---|
| 1 | Indie | 50501 | 3727277 |
| 2 | Adventure | 49653 | 12081012 |
| 3 | Simulator | 22828 | 2120077 |

# 4. Which developers have been the most prolific in releasing games?

Query | Query History

```
1  SELECT developer, COUNT(g.id) AS total_games
2  FROM developers d
3  JOIN games g ON d.id = g.id
4  GROUP BY developer
5  ORDER BY total_games DESC
6  LIMIT 5;
7
```

Data Output | Messages | Notifications

SQL

| | developer text | total_games bigint |
|---|---|---|
| 1 | Nintendo | 1926 |
| 2 | Konami | 1575 |
| 3 | Sega | 1476 |
| 4 | Electronic Arts | 1157 |
| 5 | Capcom | 1003 |

# 5. Do specific developers prefer making particular genres (e.g., RPGs, Action)?

Query | Query History

```
1  SELECT developer, genre, COUNT(g.id) AS total_games
2  FROM developers d
3  JOIN games g ON d.id = g.id
4  JOIN genres ge ON g.id = ge.id
5  GROUP BY developer, genre
6  ORDER BY total_games DESC
7  LIMIT 5;
```

Data Output | Messages | Notifications

| | developer text | genre character varying (99) | total_games bigint |
|---|---|---|---|
| 1 | Square Enix | RPG | 505 |
| 2 | Nintendo | Adventure | 449 |
| 3 | Square Enix | Adventure | 382 |
| 4 | Konami | Sport | 380 |
| 5 | Nintendo | Platform | 371 |

# 6. Which games had the highest player engagement (total players and active players) ever, excluding null values?

Query | Query History

```
1  SELECT name, plays AS total_players, playing AS active_players
2  FROM games
3  WHERE plays IS NOT NULL AND playing IS NOT NULL
4  ORDER BY total_players DESC, active_players DESC
5  LIMIT 5;
6
```

Data Output | Messages | Notifications

SQL

| | name<br>character varying (255) | total_players<br>integer | active_players<br>integer |
|---|---|---|---|
| 1 | Minecraft | 61444 | 3589 |
| 2 | Grand Theft Auto V | 57464 | 1703 |
| 3 | The Legend of Zelda: Breath of the Wild | 55899 | 4089 |
| 4 | Portal 2 | 53463 | 989 |
| 5 | Portal | 50646 | 527 |

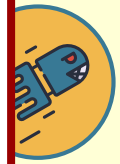
# 7. Which games have the most “favorites” or wish lists?

Query | Query History

```
1  SELECT name, wishlists
2  FROM games
3  WHERE wishlists IS NOT NULL
4  ORDER BY wishlists DESC
5  LIMIT 5;
6
```

Data Output | Messages | Notifications

| | name<br>character varying (255) | wishlists<br>integer |
|---|---|---|
| 1 | Elden Ring | 8311 |
| 2 | The Legend of Zelda: Tears of the Kingdom | 7604 |
| 3 | Baldur's Gate 3 | 7378 |
| 4 | Hollow Knight: Silksong | 7212 |
| 5 | God of War Ragnarök | 6911 |



# 8. What games are most frequently added to player backlogs?

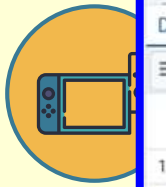
Query | Query History

```
1  SELECT DISTINCT name, backlogs
2  FROM games
3  where backlogs IS NOT NULL
4  ORDER BY backlogs DESC
5  LIMIT 10;
6
```

Data Output | Messages | Notifications

| | name character varying (255) | backlogs integer |
|---|---|---|
| 1 | Hollow Knight | 15647 |
| 2 | The Witcher 3: Wild Hunt | 12406 |
| 3 | Red Dead Redemption 2 | 12202 |
| 4 | Hades | 11926 |
| 5 | Yakuza 0 | 11708 |
| 6 | Bloodborne | 10801 |
| 7 | Cyberpunk 2077 | 10774 |
| 8 | Elden Ring | 10744 |
| 9 | Fallout: New Vegas | 10531 |
| 10 | Outer Wilds | 10304 |

# 9. Do games with high player counts also receive high ratings?



Query | Query History

```
1  SELECT DISTINCT name, rating, plays
2  FROM games
3  WHERE rating IS NOT NULL AND plays IS NOT NULL
4  ORDER BY plays DESC, rating DESC
5  LIMIT 10;
```

Data Output | Messages | Notifications

SQL

| | name<br>character varying (255) | rating<br>numeric (3,2) | plays<br>integer |
|---|---|---|---|
| 1 | Minecraft | 4.30 | 61444 |
| 2 | Grand Theft Auto V | 3.90 | 57464 |
| 3 | The Legend of Zelda: Breath of the Wild | 4.40 | 55899 |
| 4 | Portal 2 | 4.50 | 53463 |
| 5 | Portal | 4.10 | 50646 |
| 6 | Undertale | 4.30 | 49039 |
| 7 | Among Us | 3.10 | 45004 |
| 8 | Super Mario Odyssey | 4.20 | 44730 |
| 9 | Super Smash Bros. Ultimate | 4.10 | 43599 |
| 10 | God of War | 4.30 | 42246 |

# 10. Which platform has the most active players currently, restricted to games released in the last 5 years?

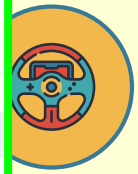
Query | Query History

```
1  SELECT platform, SUM(playing) AS active_players
2  FROM platforms AS p
3  JOIN games AS g ON p.id = g.id
4  WHERE playing IS NOT NULL
5  AND date >= '2019-01-01'
6  GROUP BY platform
7  ORDER BY active_players DESC
8  LIMIT 5;
9
```

Data Output | Messages | Notifications

SQL

| | platform character varying (99) | active_players bigint |
|---|---|---|
| 1 | Windows PC | 223054 |
| 2 | PlayStation 4 | 151400 |
| 3 | PlayStation 5 | 145893 |
| 4 | Nintendo Switch | 141001 |
| 5 | Xbox One | 127713 |



# Final Analysis

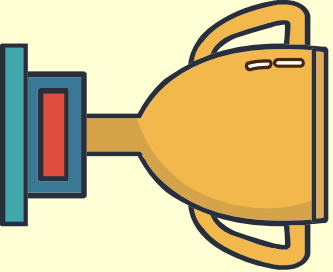

PS1. 2021 saw a surge in game releases, likely driven by increased demand for home entertainment during the pandemic, which led to more game production.

PS2. December is a key release month, aligning with the holiday season when companies target higher sales and player engagement.

PS3. Indie games are abundant due to lower development barriers, but Adventure games tend to attract more players because of their engaging narratives and immersive worlds.

PS4. Nintendo leads in game production, leveraging its strong legacy and diverse game portfolio, while Konami and Sega continue to be industry giants.

PS5. Square Enix dominates RPGs, reflecting their expertise in creating immersive, story-driven worlds, while Nintendo excels in genres that emphasize fun, innovative gameplay.



# Final Analysis

PS6. Games like Minecraft and GTA V have high player engagement due to their open-world design and replayability, which keep players invested over time.

PS7. Anticipated releases like Elden Ring and Zelda generate significant interest before launch, reflecting strong marketing and fan anticipation.

PS8. Games with longer play-times, like Hollow Knight and The Witcher 3, often sit in backlogs due to the time commitment required, though they maintain player interest.

PS9. Highly-rated games such as Portal 2 retain strong player bases, showing that quality and replayability drive continued engagement.

PS10. Windows PC and PlayStation platforms dominate the market due to their extensive game libraries, strong player communities, and versatility in gaming experiences.

# THANK YOU

SQL PROJECT BY ANKIT JADLI

(and hire me pls)