Case Study: Raw & Relational: Exploring Game Sales with SQL

TL;DR

This project showcases how structured SQL queries and database design alone can surface meaningful, decision-ready business insights, without relying on external BI tools. It simulates an end-to-end analytics workflow using only SQL, from schema creation to insight generation, demonstrating that a well-structured relational model and effective querying can serve as a powerful foundation for analysis.

Key Wins

- Xbox 360 and Nintendo dominated global sales
- The strategy genre underperformed significantly
- 2008-2010 was the peak sales era
- Only 189 games reached “blockbuster” status

Project Objective

To replicate a complete business intelligence workflow using only SQL, starting from raw data and normalization to querying for strategic insights. The aim is to demonstrate how SQL can be used not just for data extraction but also for analytical storytelling and visual simulation.

Background

In traditional analytics workflows, BI dashboards often carry the burden of delivering insights. This project challenges that model: What if dashboards were removed entirely?
Using only MySQL and a normalized schema, this analysis explores global video game sales, simulates common BI visualization in code, and presents key business observations based solely on SQL outputs.

Tools Used

- Microsoft Excel- Initial data cleaning
- MySQL- Relational schema design and SQL querying
- dbdiagram.io- Entity Relationship Diagram (ERD) planning

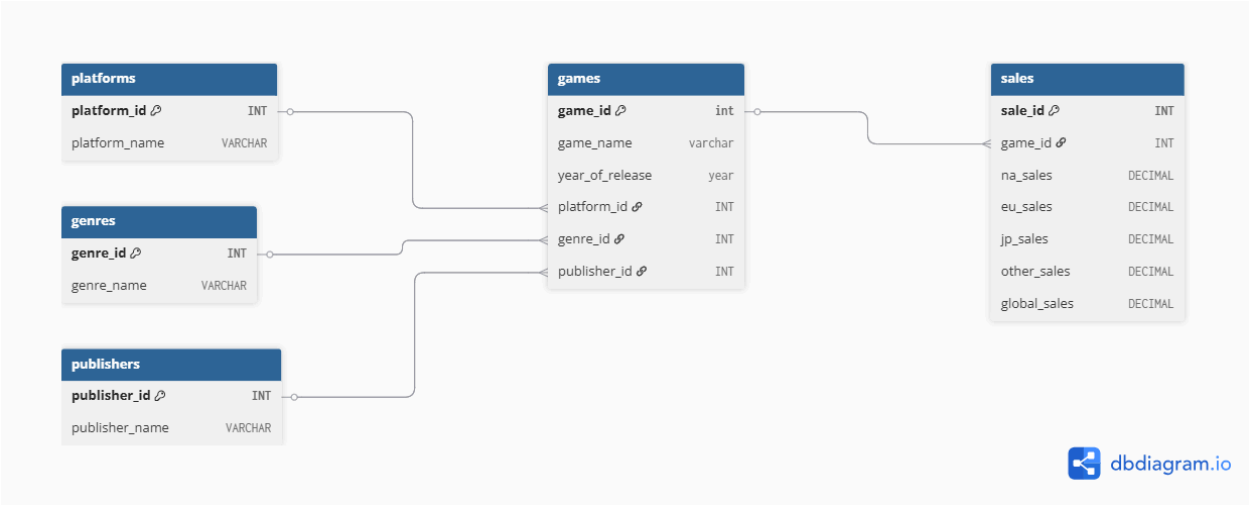
Schema Design

The original flat dataset was decomposed into five interrelated tables to ensure referential integrity and improve query efficiency.

- **Games:** Game titles and metadata
- **Genres:** Game genre classification
- **Publishers:** Publisher information
- **Platforms:** Console/platform details

- **Sales:** Regional and global sales figures

All foreign keys were manually defined to maintain relational accuracy.



"Entity Relationship Diagram (ERD) designed manually in dbdiagram.io based on the normalized schema."

Key Business Questions

This analysis explores questions relevant to strategic decision-makers in the gaming industry, including:

- Which platform achieved the highest global sales?
- Which publisher led the market in total units sold?
- What are the least profitable genres?
- Are blockbuster titles rare or frequent in the dataset?

Insights & Strategic Takeaways

- **Top Platform: Xbox 360**
  With over 10,800 global units sold, Xbox 360 consistently led in both North America and Europe. This suggests that regional market dominance may outweigh platform specifications, an important factor for platform rollout strategies.

- **Top Publisher: Nintendo**

Nintendo outperformed all competitors in total sales. Given its multi-generational presence and strong franchise ecosystem, this trend highlights the value of IP longevity and brand trust in the gaming sector.

- **Lowest-Performing Genre: Strategy**
  Strategy games recorded the least global engagement, indicating a potentially oversaturated or niche market segment.

- **Sales Distribution: Skewed**
  The majority of titles sold under 500K units, with only a handful achieving “blockbuster” status. The power-law distribution suggests that most games should be developed with lean investment unless early indicators point toward breakout success.

ASCII Chart Representation

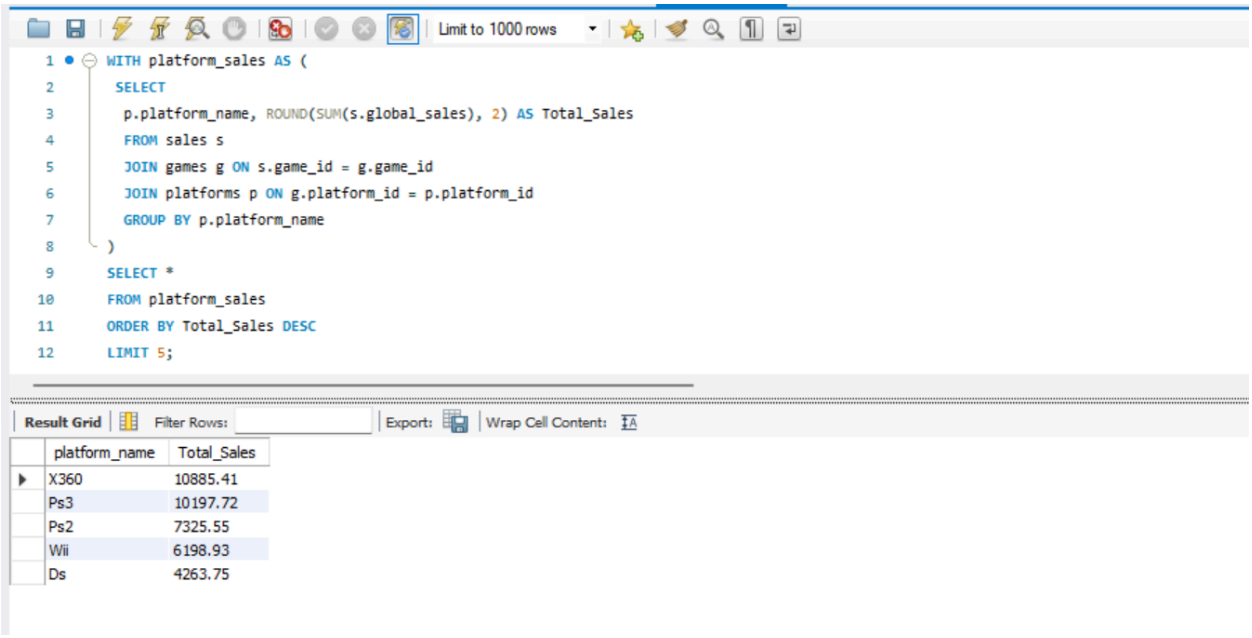
An ASCII-based bar chart was constructed using raw SQL to simulate a visual comparison between platforms.

- Legend: Each "▓" ≈ ~1,000 units sold
- Chart scaled manually using query output

(Chart included in GitHub markdown, visualized purely in code)

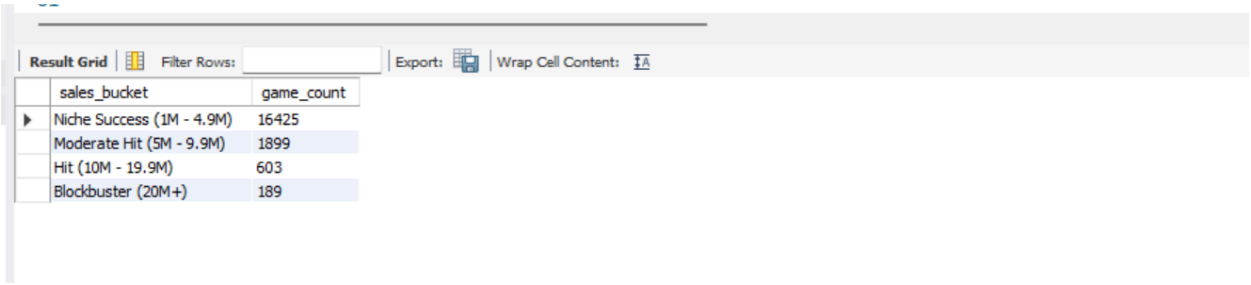
Sample Query Previews:

- **Top 5 platforms by Global Sales**
  This query replicates a dashboard-style summary tool using Common Table Expressions (CTEs) to filter, rank, and present platform performance metrics.



- **Sales Buckets by Game Volume**
  Sales volume buckets were created using SQL **CASE WHEN** logic to simulate a segmented chart view, mimicking user-friendly dashboard visuals through code alone.

```
64
65  WITH sales_buckets AS (
66   SELECT g.game_name,
67    CASE
68     WHEN s.global_sales >= 20 THEN 'Blockbuster (20M+)'
69     WHEN s.global_sales >= 10 THEN 'Hit (10M - 19.9M)'
70     WHEN s.global_sales >= 5 THEN 'Moderate Hit (5M - 9.9M)'
71     WHEN s.global_sales >= 1 THEN 'Niche Success (1M - 4.9M)'
72     ELSE 'Low Impact (< 1M)'
73  END AS sales_bucket
74  FROM sales s
75  JOIN games g ON s.game_id = g.game_id
76  )
77  SELECT sales_bucket, COUNT(*) AS game_count
78  FROM sales_buckets
79  GROUP BY sales_bucket
80  ORDER BY game_count DESC;
81
```



sales_bucket	game_count
Niche Success (1M - 4.9M)	16425
Moderate Hit (5M - 9.9M)	1899
Hit (10M - 19.9M)	603
Blockbuster (20M+)	189

Why It Matters

This project validates SQL as more than a backend tool; it is a storytelling mechanism. By combining strong schema design, analytical logic, and clear presentation, SQL can serve as a standalone solution for deriving and communicating insights. This approach is especially valuable for analysts working in data-constrained or early-stage environments where BI tools may not yet be implemented.

What I'd Explore Next

If scaled further, this project could evolve in the following ways:

- **Price vs Sales Analysis**- Explore how pricing strategies impact regional revenue
- **Ratings vs Performance**- Correlate game reviews with sales outcomes
- **Seasonality Trends**- Examine quarterly release patterns and their effects
- **Advanced SQL Techniques**- Implement window functions for better ranking and cohort analysis

Explore the Project

- GitHub Repo
- Portfolio