# Combined SQL & Power BI Project – Dataset Assessment Form

1. Your Name:

Jibin Rehman

2. Dataset Location (Drive link or database details):

[Github Repo](https://github.com/d-binu163/game_sales_analysis_project/)

3. Project Goal:

Analyze video game sales data to identify trends in global and regional sales, best-performing platforms, publishers, and genres over time. Build an interactive Power BI dashboard for stakeholders to explore sales performance by various dimensions.

4. Data Source and Format

**- Source of the dataset:** [**Gigasheet Video Game Sales 1978-2024 Data**](https://app.gigasheet.com/spreadsheet/video-game-sales-1978---2024/b3ac0e84_314a_4900_a79d_d2fd8dc49381)

**- Storage Format:** CSV → SQL database (MySQL / PostgreSQL) → imported to Power BI.

5. Data Size and Structure

- Number of tables/files involved: 4 main CSV tables (games, platforms, publishers, sales).

- Primary table(s) for analysis: sales table (contains game\_id, region sales, global sales).

- Relationships between tables (keys, joins):

games.game\_id → sales.game\_id

platforms.platform\_id → games.platform\_id

publishers.publisher\_id → games.publisher\_id

- Number of records (rows) in the primary table: 60k rows (Pre-Cleaned), 20k (Cleaned)

- Number of columns in the primary table: 10 columns

6. Data Content and Quality

- List key columns and their data types:

game\_id (integer, key)

game\_name (text)

platform\_id (integer)

publisher\_id (integer)

year (integer)

global\_sales, na\_sales, pal\_sales, jp\_sales, other\_sales (decimal)

- Are data types and values consistent across columns?

Yes; numeric sales columns are floats/decimals, IDs are integers, names are strings.

- Are there missing values? How will they be handled (SQL or Power BI)?

Yes. Regional Sales had many null values. They were kept as it is.

- Are there any calculated fields to derive (in SQL or Power BI)?

Yes; e.g., Total Sales (in millions), Rank measures for Top N analysis, aggregated sales by game across platforms.

- Does the dataset include categorical variables (e.g., region, product category)?

Platform, Publisher, Genre, Region.

- Does the dataset include date/time columns? What format are they in?

Year (integer)

- Are there geographical columns (e.g., country, city)?

Region (derived from sales columns).

- Any numerical columns with large ranges or outliers? How will you handle them?

Popular games like Wii Sports have very high sales — no removal, but visual scaling is used.

- Are there hierarchical fields (e.g., category → subcategory)?

Genre → Game, Publisher → Game, Platform → Game

- Fields suitable for filtering/grouping (e.g., customer segment, product ID)?

Year, Genre, Platform, Publisher, Region.

7. Analytical Opportunities

- Time-based trends to explore (e.g., monthly sales, year-over-year growth)?

Global & regional sales trends by year, platform lifecycle trends.

- Comparisons or aggregations (e.g., sales by region or product)?

Sales by platform, publisher, genre, region.

- Correlations or patterns (e.g., behavior by customer segments)?

Best-selling games by region, top publishers by year, genre popularity shifts.

8. Performance & Optimization

- Dataset size (MB or rows):

1.5 MB

- Will indexing or query optimization be needed in SQL?

Primary keys & joins handled in SQL; dataset small enough that indexing is minimal. But indexing is still done for faster processing.

- Any planned use of SQL subqueries, joins, CTEs, or window functions?

Joins, aggregate functions, grouping, filtering, normalization, and possibly CTEs for intermediate summaries.

- Will data be pre-aggregated in SQL or loaded raw into Power BI?

Pre-aggregated some data in SQL for faster Power BI visuals; some measures calculated in DAX.

9. Power BI Specifics

- Which visuals do you plan to use (e.g., bar, pie, line, slicers)?

Card KPIs, Line Chart, Donut Chart, Stacked Column Chart, 100% Stacked Area, Treemap, Matrix Table, Clustered Bar, Tables, Slicers.

- Will you use custom measures (DAX) ? Give examples.

Total Sales (In Millions)

Rank Measures

Aggregated Sales across platforms

- Are there any filters, slicers, or drill-through features you will add?

Slicers for Year, Genre, Platform, Publisher, Region

- How will you format and organize report pages?

Page 1: Global Overview

Page 2: Platform Insights (Top 10 only)

Page 3: Publisher Insights

Page 4: Genre & Game Insights

Home Page with navigation buttons to all pages.