Global Video Game Sales Market Analysis and Netflix Content Analysis in power BI

The domain of the Project:

SQL AND POWER BI

Under the guidance of

Ms. Siddhika Shah Software Engineer At HCL Technologies

Prof. Radha kumari Executive Director & Founder SURE Trust

By

SYAM KUMAR APPIKATLA(B.Tech)

Period of the project

May 2025 to August 2025

SURE TRUST

PUTTAPARTHI, ANDHRA PRADESH

Declaration

The project titled "Global Video game sales market and Netflix content analysis using power BI" has been mentored by Ms. Siddhika Shah, organised by SURE Trust, from May 2025 to August 2025, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge the members of the team mentioned below, have worked on it successfully and enhanced their practical knowledge in the domain.

By: **Signature**

Syam kumar Appikatla   A.syam kumar

(Pursuing B.Tech)

Mentors: **Signature**

Ms Siddhika Shah Software Engineer

At HCL Technologies

Seal & Signature

Prof. Radha kumari

Executive Director & Founder
SURE Trust

☐ **Table of Contents**

Global Video Game Sales Market Analysis

Table of Contents

1. Introduction

The video game industry is a dynamic and rapidly evolving sector of the entertainment market, generating billions in revenue annually. This project report provides a comprehensive analysis of global video game sales data, with the primary goal of uncovering key market trends, identifying top-performing segments, and understanding the competitive landscape. The insights derived from this analysis are crucial for stakeholders, including game developers, publishers, and investors, to make informed strategic decisions. The report utilizes a data-driven approach, transforming raw sales figures into actionable intelligence presented through a detailed and interactive dashboard. The data encompasses a wide range of metrics, from total games

released and global sales figures to granular breakdowns by genre and publisher, providing a holistic view of the market.

2. Project Objectives

The core objectives of this project were to:

- **Analyze overall market size and scale:** To quantify the total number of games released and total global sales to understand the overall magnitude of the market.
- **Identify top-performing genres:** To determine which game genres are driving the most significant revenue on a global scale.
- **Evaluate publisher performance:** To identify and rank the leading publishers by their global sales, thereby highlighting the key players in the industry.
- **Examine regional sales patterns:** To analyze how sales are distributed across major geographical regions, including North America (NA), Europe (EU), Japan (JP), and other markets, to understand regional preferences and market maturity.
- **Synthesize findings into a comprehensive dashboard:** To present the analyzed data in a clear, intuitive, and interactive visual format that simplifies complex information and facilitates easy interpretation.

3. Methodology

The project's methodology was structured in three key phases: data acquisition, data processing, and data visualization.

Data Acquisition and Sourcing: The data for this analysis was sourced from a comprehensive dataset of video game sales. While the original source is not explicitly stated in the provided dashboard, it is assumed to be a large-scale collection of sales records spanning multiple years and platforms. The dataset includes fields such as game title, genre, publisher, year of release, and sales figures broken down by region (NA, EU, JP, and Others).

Data Processing and Analysis: The raw data was processed to calculate aggregated metrics and create relational views necessary for the dashboard. This involved:

- **Aggregation**: Summing total global sales, counting the total number of games, and counting the number of unique publishers.
- **Ranking** sales: Sorting data to identify the top-selling genres and the top-performing publishers by global.
- **Normalization**: Converting raw sales numbers into a consistent unit (millions) for easy comparison across different metrics.

Data Visualization: The processed data was then used to create a comprehensive, single-page dashboard. The dashboard was designed to be a "single source of truth" for the project's key findings. It includes various visualization components, each tailored to present a specific insight:

- **Card-based Metrics:** Large, clear cards were used to display key performance indicators (KPIs) like Total Games Released, Total Global Sales, and Total Publishers.

- **Bar Charts:** Bar charts were employed to show the distribution of global sales by genre and to rank the top five publishers. This visual format is effective for comparing discrete categories.
- **Stacked Bar Chart:** A stacked bar chart was used to illustrate the regional sales distribution across different genres, providing insight into which regions contribute most to a genre's success.
- **Filtering and Interactivity:** The dashboard was equipped with filters for game name and year of release, allowing users to interact with the data and explore specific subsets.

4.Findings and Analysis

The analysis of the video game sales data dashboard revealed several significant insights into the market's structure and performance.

4.1. Overall Market Summary

The dashboard provides a high-level overview of the market's scale. With a **Total of 1000 games released** in the analyzed period, the market demonstrates a robust and active development environment. This volume of releases contributes to a **Total Global Sales figure of $25.20K (million)**, which is equivalent to $25.2 billion. The market is supported by a relatively concentrated group of key players, with a **Total of 6 publishers** identified in the dataset. This suggests a market structure where a few major companies hold a significant share of the publishing landscape.

4.2. Genre Performance Analysis

A detailed look at global sales by genre reveals a highly competitive top tier. The **Shooter, Racing, and Puzzle genres are effectively tied for the**

top spot, each generating approximately $3.4K (million) in sales. This parity indicates that these genres have a broad and consistent appeal worldwide. Following closely are **Sports and Role-Playing games**, both with sales around $3.2K (million), demonstrating their sustained popularity. The **Adventure genre** rounds out the top categories with sales of $3.1K (million). This shows a clear hierarchy of genres, with a small number of categories dominating the market.

4.3. Top Publisher Dominance

The analysis of the top five publishers by global sales highlights the industry's concentrated power structure. **Sony leads the pack with sales of $4.5K (million)**, solidifying its position as a market leader. **Microsoft and Nintendo follow closely**, with sales of $4.4K (million) each, underscoring the strong competition among the three console giants. **Ubisoft and Activision** complete the top five with sales of $4.3K (million) and $4.0K (million), respectively. The small differences in sales among these top players demonstrate the intense competition for market share. These findings confirm that a select group of established companies holds a powerful influence over the global video game market.

4.4. Regional Sales Distribution

The regional sales distribution chart provides a nuanced perspective on genre popularity across different geographical markets. The chart, which breaks down sales into North America (NA), Europe (EU), Japan

(JP), and Other regions, shows that sales contributions vary significantly by genre. For instance, the **Shooter and Action genres appear to have a strong presence in the North American market**, while other genres may show more balanced distribution across regions. This breakdown is critical for publishers, as it can inform marketing and localization strategies. Understanding these regional differences is key to tailoring content and campaigns to specific audiences to maximize sales potential. The data indicates that a one-size-fits-all approach to sales and marketing is not effective in a global market, necessitating a tailored, region-specific strategy.

5. Conclusion and Recommendations

The analysis of the Global Video Game Sales Dashboard has successfully met all project objectives, providing valuable insights into the current state of the video game market. The report has quantified the market's scale, identified the most lucrative genres and dominant publishers, and highlighted critical regional sales patterns.

The key takeaways are:

- The market is large and active, with a high volume of game releases and significant revenue.
- The Shooter, Racing, and Puzzle genres are the most dominant, suggesting broad consumer appeal.
- The market is highly concentrated, with a few major publishers controlling a significant share of global sales.

Regional preferences play a crucial role, necessitating localized strategies for successful market penetration.

Based on these findings, we recommend that companies in the video game industry conduct further analysis to:

- **Deep Dive into Genre Sub-segments:** Explore specific sub-genres within the top categories to identify emerging trends.
- **Publisher-Specific Analysis:** Conduct a more detailed analysis of the top publishers to understand their individual strategies and market dynamics.
- **Historical Trend Analysis:** Extend the analysis to include historical data to identify long-term trends and predict future market shifts

Project Report: Netflix Content Analysis using Power BI

Submitted by:

Syam Kumar Appikatla

Project for:

Sure Trust Internship

Project Report: Netflix Content Analysis using Power BI

A Power BI Capstone Project Report

Abstract

This report details a data analysis project focused on the Netflix content library. The primary objective was to use Microsoft Power BI to develop an interactive dashboard that provides insights into the composition ofmovies and TV shows available on the platform. By processing and visualizing a comprehensive dataset of Netflix titles, the project uncovers key trends and patterns related to content types, genres, viewer ratings, and movie durations. The analysis reveals that the Netflix library is predominantly composed of movies, with a strategic focus on content for mature audiences (TV-MA and TV-14 ratings). The most prominent genres include Dramas, International Movies, and Comedies. Furthermore, the data shows a clear preference for feature- length films (90+ minutes). The resulting dashboard serves as a powerful tool for understanding Netflix's content strategy and provides a foundation for data-driven decision-making in content acquisition and production.

Introduction

1.1 Background

In the highly competitive streaming market, content is the primary driver of subscriber acquisition and retention. Netflix, as a global leader, possesses a vast and diverse library of movies and TV shows. Understanding the composition of this library is crucial for making strategic decisions, identifying content gaps, and catering to audience preferences. Data analysis provides the tools to transform raw content data into actionable insights, helping stakeholders understand what is available on the platform and how it is categorized.

By systematically examining the composition of Netflix's library, organizations can uncover valuable insights into genre distribution, regional representation, and the balance between original and licensed content. Such analysis highlights strengths in the catalog while also identifying potential gaps that may leave certain audience segments underserved. Moreover, comparing these findings with emerging market trends and competitor strategies enables Netflix to refine its content portfolio and ensure continued relevance in a crowded industry. This data-driven approach not only supports smarter investment in new titles but also enhances personalization, helping to deliver the right content to the right audience at the right time. In turn, this strengthens customer satisfaction, improves retention, and sustains Netflix's competitive edge in the global streaminglandscape.

1.2 Problem Statement

The sheer volume of content on Netflix makes it challenging to get a clear, high-level overview of its library without proper analytical tools. There is a need to systematically analyze and visualize key metrics—such as the distribution of movies versus TV shows, the most common genres, and the prevalence of different content ratings—to understand the platform's strategic focus.

1.3 Objectives

The main objectives of this project were:

- To design and build a dynamic, interactive dashboard using Microsoft Power BI.
- To analyze and visualize the distribution of movies and TV shows in the Netflix library.
- To identify the most prominent genres and content ratings on the platform.
- To examine the distribution of movie runtimes.

- To extract and present key insights that summarize Netflix's content strategy.

1.4 Scope

The scope of this project is centered on the analysis of a static Netflix dataset. It involves the entire process of data cleaning, transformation,and visualization within Power BI to create a multi-page dashboard. The analysis focuses on four primary areas: content type, content rating, genre, and movie duration.

and visualization within Power BI to create a multi-page dashboard. The analysis focuses on four primary areas: content type, content rating, genre, and movie duration.

2. Methodology and Tools

2.1 Primary Tool: Microsoft Power BI

Microsoft Power BI was the exclusive tool used for this project. It is a powerful business intelligence platform that enables end-to-end data analysis, from connecting to data sources to creating compelling, interactive visualizations. Its capabilities were leveraged for data transformation (using Power Query), data modeling, and the creation of the final dashboards.

2.2 Data Source

The project utilized a dataset containing details of Netflix titles. The key data fields included:

- show_id: Unique identifier for each title.
- type: Category (Movie or TV Show).
- title: Name of the movie or show.
- country: Country of production.
- release_year: The year the content was released.

- rating: The content rating (e.g., TV-MA, PG-13).
- duration: The runtime of the movie or number of seasons for a TV show.
  isted_in: The assigned genres.

2.3 Project Workflow

1. **Data Loading and Cleaning:** The dataset was imported into Power BI. The Power Query Editor was used to handle inconsistencies and missing values (e.g., nulls in the country field).

2. **Data Transformation:** New columns and measures were created to facilitate analysis. This included creating "Duration Buckets" (e.g., '0-30 min', '30-60 min', '60-90 min', '90+ min') to categorize movies based on their runtime.

3. **Data Visualization:** A variety of visuals were created to represent the findings, including:

   ○ **Key Performance Indicator (KPI) Cards:** To display top-line numbers like Total Titles, Total Movies, and Total TV Shows.

   ○ **Donut and Bar Charts:** To show the distribution and comparison of categorical data like content ratings and genres.

   ○ **Treemap:** To visualize the hierarchy and weight of different genres.

4. **Dashboard Design:** The visuals were organized into three distinct, user-friendly dashboard pages: an overall content overview, a detailed analysis page, and a final summary of key insights.

3. Data Analysis and Findings

The analysis of the Netflix dataset yielded several key findings, presented across the interactive dashboards.

2.4 Overall Content Library Composition

The dashboard reveals that the Netflix library is extensive, with a clear focus on one type of content.

- **Total Titles:** 8,792 ● **Total Movies:** 6,126 (~70%) ● **Total TV Shows:** 2,664 (~30%)

This shows a significant strategic inclination towards movies, which outnumber TV shows by more than two to one.

2.5 Content Ratings Distribution

The analysis of content ratings highlights a clear target audience.

- **TV-MA (Mature Audience):** The most common rating, with **3,210 titles (36.5%)**.
- **TV-14 (Parents Strongly Cautioned):** The second most common, with **2,160 titles (24.5%)**.

  TV-PG (Parental Guidance Suggested): Accounts for **860 titles (9.8%)**. ● **R (Restricted):** Accounts for **800 titles (9.0%)**.

Collectively, content aimed at mature audiences (TV-MA, TV-14, R) makes up over 70% of the library, indicating a focus on adult and older teenage viewers.

2.6 Top Genres

The analysis of genres, based on the "listed_in" category, shows a focus on specific types of content.

Top Genre Combinations:

1. Dramas, International Movies: **362 titles**
2. Documentaries: **359 titles**
3. Stand-Up Comedy: **334 titles**
4. Comedies, Dramas, International Movies: **274 titles**
5. Dramas, Independent Movies, International Movies: **252 titles**

Dramas, Comedies, and Documentaries are the dominant high-level categories. The frequent appearance of "International Movies" in top combinations also underscores Netflix's global content strategy.

2.7 Movie Duration Analysis

The analysis of movie runtimes reveals a preference for traditional, feature-length films.

- **90+ min:** This is the largest category, with **4,137 movies**.
- **60-90 min:** The second-largest category.
- **30-60 min & 0-30 min:** These shorter formats are significantly less common.

This suggests that Netflix's movie acquisition and production strategy prioritizes the standard cinematic experience over short films or mid-length features.

3. Conclusion and Key Insights

This project successfully demonstrated the use of Power BI to analyze a large content library and extract meaningful, strategic insights. The dashboard provides a clear, multi-faceted view of the Netflix catalog.

Key Insights:

- **A Movie-Dominant Platform:** Netflix's content strategy is heavily weighted towards movies, which comprise roughly 70% of its entire library.
- **Targeting Mature Viewers:** With TV-MA and TV-14 being the top two ratings, Netflix's primary target audience is adults and older teens.
  Drama is the Core Genre: Dramas, often blended with international and comedy elements, form the backbone of the Netflix catalog, complemented by a strong offering of documentaries and stand-up comedy.
- **Focus on Feature Films:** The platform overwhelmingly favors standard feature-length films (90+ minutes), investing less in shorter-form movie content.

4. Future Scope

This project can be expanded in several ways to provide even deeper insights:

Time-Series Analysis: Analyze trends over time by incorporating release dates to see how content strategy has evolved year over year.

- **Performance Metrics:** Integrate viewership data or audience ratings (e.g., from IMDb or Rotten Tomatoes) to measure the popularity and success of different types of content.

- **Competitive Analysis:** Expand the dataset to include content from other streaming platforms (e.g., Disney+, HBO Max) to perform a comparative analysis of content libraries.

- **Predictive Modeling:** Utilize the existing data to build a model that can predict which attributes (genre, cast, director, etc.) are most likely to correlate with a title's success.
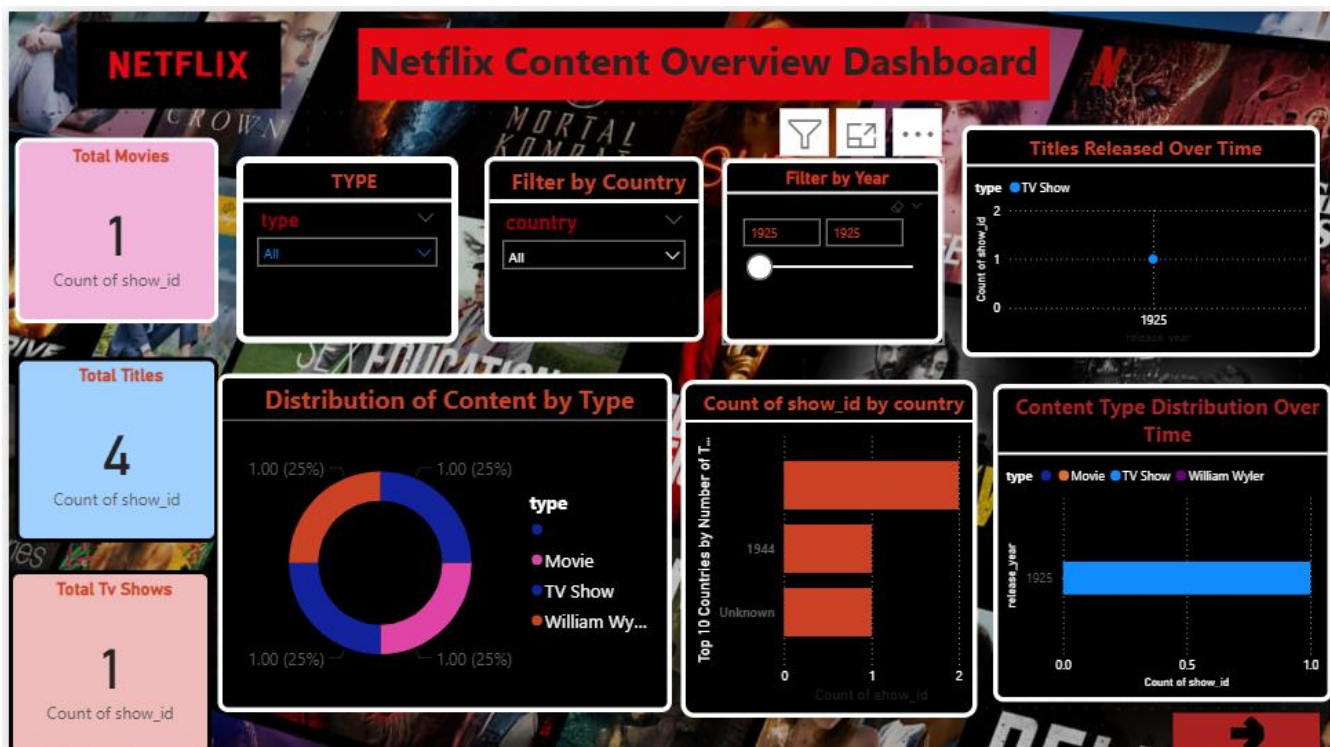
Netflix Content Analysis

TYPE
type
All

COUNTRY
country
All

GENRE
Genres
☐
☐ Action & Adventure
☐ Action & Adventure, Anime Fe...

TOTAL TV SHOWS
2664
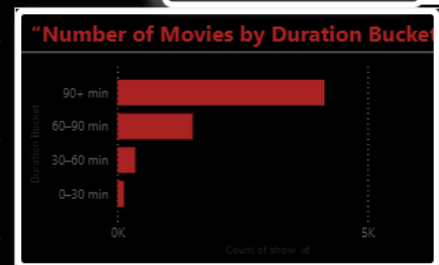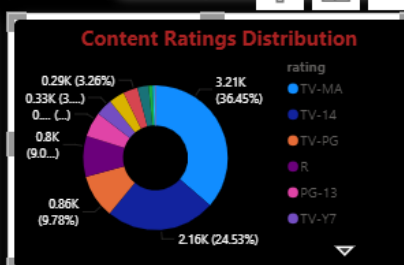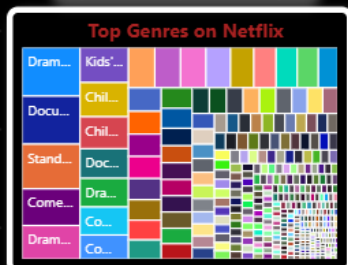Count of show_id

TOTAL TITLES
8792
Count of show_id

TOTAL MOVIES
6126
Count of show_id

Top Genres on Netflix

Content Ratings Distribution
0.29K (3.26%)
0.33K (3...)
0... (...)
0.8K (9.0...)
0.86K (9.78%)
2.16K (24.53%)
3.21K (36.45%)
rating
● TV-MA
● TV-14
● TV-PG
● R
● PG-13
● TV-Y7

"Number of Movies by Duration Bucket
90+ min
60–90 min
30–60 min
0–30 min
0K   5K
Count of show_id

Netflix Content Summary & Key Insights

Total Titles
8792
Count of show_id

Total Movies
6.13K
Count of show_id

Total TV Shows
2664
Count of show_id

Movie Counts by Duration Category

Count of show_id	Duration Bucket
2797	0–30 min
355	30–60 min
1503	60–90 min
4137	90+ min
8792	

Top Genres

Genre	Count of title
Dramas, Int...	362
Document...	359
Stand-Up C...	334
Comedies, ...	274
Dramas, In...	252
Kids' TV	219
Children & ...	215
Children & ...	201

Ratings Share
0.3... (...)
0.86K (9.7...)
2.16K (24.53%)
3.21K (36.45%)
rating
● TV-MA
● TV-14
● TV-PG
● R
● PG-13
● TV-Y7
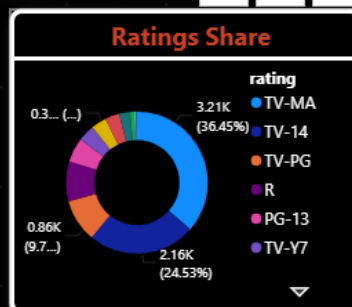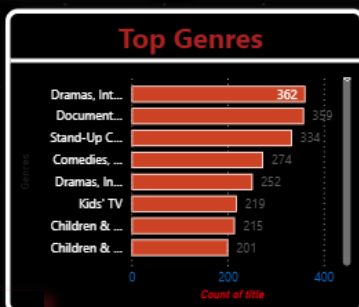
Prepared By: SYAM KUMAR APPIKATLA

Internship Project – Sure Trust

🎬 *Power BI Capstone Project*

Feedback & Attendance Management System using MySQL

Submitted By:

Syam kumar Appikatla

Submitted To:

SureTrust

/

Abstract

This project integrates **Google Form feedback data** with **meeting attendance records** using MySQL. The main objective is to validate whether the feedback submitted belongs to a user who actually attended the meeting.

The process involves:

- Collecting feedback via Google Forms
- Importing responses into MySQL
- Matching responses with user attendance
- Running a stored procedure that automatically validates and logs the results

If the feedback is valid, it is marked as **processed**; otherwise, it is stored with an **error message**. This project demonstrates **database design, SQL programming, data validation, and reporting**, which can be extended to realtime systems in future.

A stored procedure in MySQL ensures that only feedback from users who actually attended the meeting is marked as valid, while others are logged as errors.

The system also maintains a processing log for transparency and accountability.

By combining real-time data collection (Google Forms) and backend validation (MySQL), this project demonstrates a hybrid approach to feedback management that can be applied to educational institutions, corporate organizations, and training programs. This solution improves data integrity, efficiency, and transparency.

.

Introduction:

Feedback plays a vital role in improving the quality of meetings, training sessions, and organizational processes. Traditionally, feedback collection and validation are carried out manually, which is time-consuming, error-prone, and difficult to scale.

In this project, we address these challenges by designing a *database-driven system* that automates the process. The workflow is as follows:

1. Users submit feedback via Google Forms.

2. The responses are stored in Google Sheets.

3. Data is imported into a MySQL database.

4. A stored procedure validates responses against attendance records.

5. Logs and reports are generated for analysis and decision-making.

This approach ensures that:

Feedback is only considered if the user attended the session.

Invalid entries are flagged and logged for review.
Administrators can view real-time participation reports.

The project showcases how **SQL databases** can be leveraged for data validation, automation, and reporting.
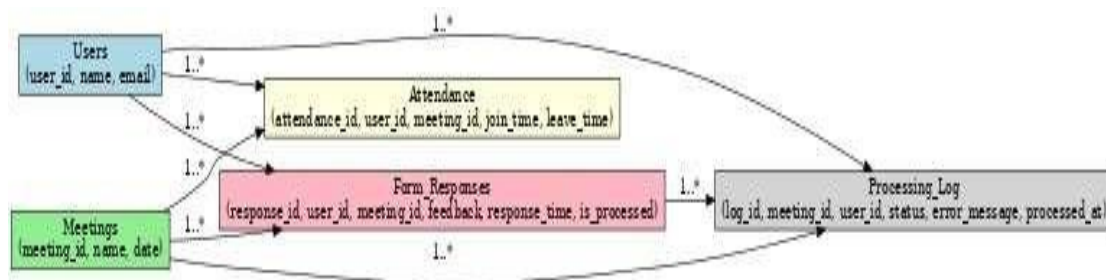
Content:

System Requirements

Hardware Requirements

- Processor: Intel i3 or higher
- RAM: Minimum 4 GB
- Storage: 100 MB free space

Software Requirements

- MySQL Workbench 8.0 or above
- MySQL Server 8.0 or above
- Python 3.10+ (optional for automation)
- Google Forms & Google Sheets

```sql
CREATE TABLE users (
  user_id INT AUTO_INCREMENT PRIMARY KEY,
  name VARCHAR(100) NOT NULL,
  email VARCHAR(150) UNIQUE
);

-- Meetings Table
CREATE TABLE meetings (
  meeting_id INT AUTO_INCREMENT PRIMARY KEY,
  meeting_name VARCHAR(100) NOT NULL,
  meeting_date DATE NOT NULL
);

-- Attendance Table
CREATE TABLE attendance (
  attendance_id INT AUTO_INCREMENT PRIMARY KEY,
  user_id INT,
  meeting_id INT,
  join_time DATETIME,
  leave_time DATETIME,
  FOREIGN KEY (user_id) REFERENCES users(user_id),
  FOREIGN KEY (meeting_id) REFERENCES meetings(meeting_id)
);

-- Form Responses Table
CREATE TABLE form_responses (
  response_id INT AUTO_INCREMENT PRIMARY KEY,
  user_id INT,
  meeting_id INT,
  feedback TEXT,
  response_time DATETIME,
  is_processed TINYINT(1) DEFAULT 0,
  FOREIGN KEY (user_id) REFERENCES users(user_id),
  FOREIGN KEY (meeting_id) REFERENCES meetings(meeting_id)
);

-- Processing Log Table
CREATE TABLE processing_log (
  log_id INT AUTO_INCREMENT PRIMARY KEY,
  meeting_id INT,
  user_id INT,
  status ENUM('processed','error'),
  error_message VARCHAR(255),
  processed_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);
```

log_id	meeting_id	user_id	status	error_message	processed_at
1	1	1	processed	NULL	2025-09-01 14:35:43
2	1	2	error	No attendance record	2025-09-01 14:35:43
3	2	2	processed	NULL	2025-09-01 14:35:43
4	1	2	error	No attendance record	2025-09-01 14:35:43
5	2	3	error	No attendance record	2025-09-01 14:35:43
6	2	3	error	No attendance record	2025-09-01 14:35:43
7	1	2	error	No attendance record	2025-09-01 14:35:43
8	2	3	error	No attendance record	2025-09-01 14:35:43
NULL	NULL	NULL	NULL	NULL	NULL

Database Schema

The system consists of 6 main tables:

- **users** – stores participant information
- **meetings** – stores meeting details
- **attendance** – records attendance of users
- **form_responses_raw** – stores raw Google Form data
- **form_responses** – stores processed form responses
- **processing_log** – logs validation results

Data Insertion

```sql
sql

-- Add Users
INSERT INTO users (name, email) VALUES
('Rahul Sharma', 'rahul@example.com'),
('Priya Nair', 'priya@example.com'),
('Amit Kumar', 'amit@example.com');

-- Add Meetings
INSERT INTO meetings (meeting_name, meeting_date) VALUES
('Team Sync', '2025-08-20'),
('Project Review', '2025-08-22');

-- Add Attendance
INSERT INTO attendance (user_id, meeting_id, join_time, leave_time) VALUES
(1, 1, NOW(), NOW()),    -- Rahul attended meeting 1
(3, 1, NOW(), NOW()),    -- Amit attended meeting 1
(2, 2, NOW(), NOW());    -- Priya attended meeting 2
```

log_id	meeting_id	user_id	status	error_message	processed_at
1	1	1	processed	NULL	2025-09-01 14:43:45
2	1	2	error	No attendance record	2025-09-01 14:43:45
3	1	2	error	No attendance record	2025-09-01 14:43:45
4	2	2	processed	NULL	2025-09-01 14:43:45
5	2	2	processed	NULL	2025-09-01 14:43:45
6	2	3	error	No attendance record	2025-09-01 14:43:45
7	2	3	error	No attendance record	2025-09-01 2025-09-01 14:43:45
8	1	2	error	No attendance record	2025-09-01 14:43:45
9	2	3	error	No attendance record	2025-09-01 14:43:45
NULL	NULL	NULL	NULL	NULL	NULL

Importing Google Form Data

Import Google Form Data

The **feedback** is collected through **Google Forms**, which automatically stores responses into a **Google Sheet**.
To use this data inside MySQL, the sheet is exported as a **CSV file** and then imported into MySQL using the **Table Data Import Wizard** in MySQL Workbench.

Step 1: Export from Google Sheets

- Open the Google Form responses sheet.
- Click on **File → Download → CSV**.

Step 2: Import into MySQL

- In MySQL Workbench → Right-click database `suretrust_project`.
- Select **Table Data Import Wizard**. · Choose your CSV file.
- Import it into a new table called **`form_responses_raw`**.

Step 3: Verify Data Import

```
-- Check first 10 records from raw responses
SELECT * FROM form_responses_raw LIMIT 10;
```

1	Timestamp ⌄	Name ⌄	Email ⌄	Meeting ID ⌄	Feedback ⌄
3	28/08/2025 21:51:45	Badarla Bala siva teja	badarlabalu5@gmail.cor	101	Very Good
4	29/08/2025 11:19:52	Samuel	samue22@gmail.com	101	Good
5	30/08/2025 21:43:30	Devesh Kushwaha	deveshkushwaha1256@	101	Excellent
6	01/09/2025 10:19:55	Raju	raju59342@gmail.com	101	Very Good
7	01/09/2025 10:20:44	Santhosh	santhosh4234@gmail.cc	101	Excellent
8	01/09/2025 10:21:31	Likith	likitha643@gmail.com	101	Bad
9	01/09/2025 10:22:08	Ramesh	ramesh1234@gmail.con	101	Very Good
10	01/09/2025 10:23:06	Ramana	ramana04@gmail.com	101	Excellent
11	01/09/2025 10:24:24	Aman	iaman9846@gmail.com	101	Very Good
12	01/09/2025 10:25:50	pawan kalyan	kalyansep2@gmail.com	101	Good
13	01/09/2025 10:28:25	Thaman	thaman6978@gmail.cor	101	Very Good
14	01/09/2025 10:37:08	tharun	tharun4321@gmail.com	101	Very Good

Moving Data to Main Table

Step 4: Insert into Processed Table

We move data from `form_responses_raw` to `form_responses`.
The **user_id** is matched using the **email ID**.

INSERT INTO form_responses (user_id, meeting_id, feedback, response_time)

SELECT u.user_id, 1, r.Feedback, r.Timestamp

FROM form_responses_raw r

JOIN users u ON u.email = r.Email;

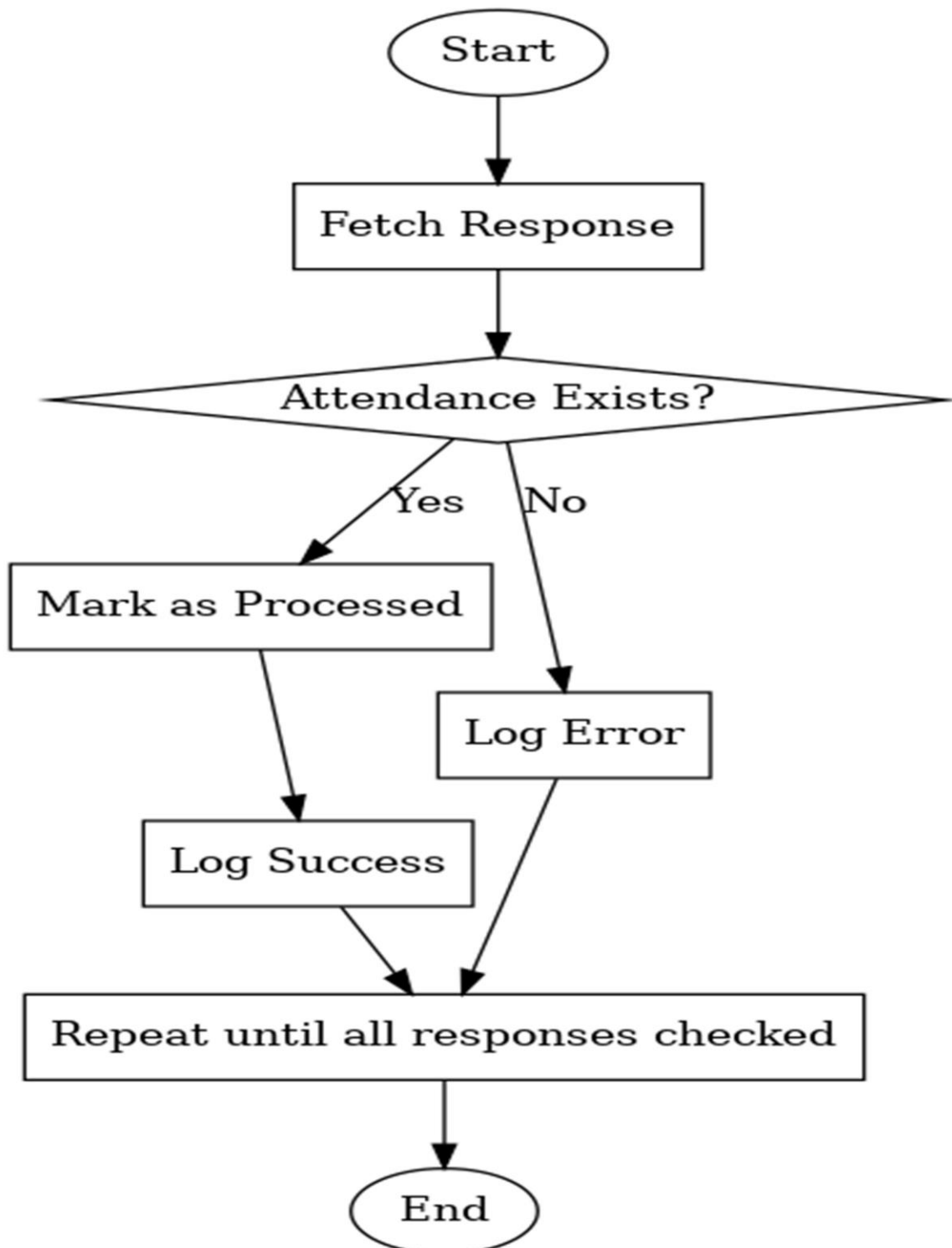| log_id	meeting_id	user_id	status	error_message	processed_at
1	1	1	processed	NULL	2025-09-01 15:50:31
2	1	2	error	No attendance record	2025-09-01 15:50:31
3	2	2	processed	NULL	2025-09-01 15:50:31
4	1	2	error	No attendance record	2025-09-01 15:50:31
5	2	3	error	No attendance record	2025-09-01 15:50:31
6	2	3	error	No attendance record	2025-09-01 15:50:31
7	1	2	error	No attendance record	2025-09-01 15:50:31
8	2	3	error	No attendance record	2025-09-01 15:50:31

Stored Procedure for Processing Responses

Stored Procedure Logic:

- **Fetch unprocessed responses** from `form_responses`.
- **Check attendance** of the user for the corresponding meeting.
- If valid → mark response as **processed** and log it as **success**.
- If invalid → log it as **error** with message *"No attendance record"*.

Flowchart of Logic

```
                    ┌─────────┐
                    │  Start  │
                    └─────────┘
                         │
                         ▼
              ┌──────────────────────┐
              │   Fetch Response     │
              └──────────────────────┘
                         │
                         ▼
              ╱────────────────────╲
             ╱  Attendance Exists?  ╲
             ╲                      ╱
              ╲────────────────────╱
              Yes              No
               │                │
               ▼                ▼
     ┌──────────────────┐   ┌──────────┐
     │ Mark as Processed│   │Log Error │
     └──────────────────┘   └──────────┘
               │                │
               ▼                │
        ┌──────────────┐        │
        │ Log Success  │        │
        └──────────────┘        │
               │                │
               ▼                ▼
     ┌────────────────────────────────────┐
     │ Repeat until all responses checked │
     └────────────────────────────────────┘
                         │
                         ▼
                    ┌─────────┐
                    │   End   │
                    └─────────┘
```

SCRIPT:

```sql
CREATE PROCEDURE process_responses()
BEGIN
  DECLARE done INT DEFAULT 0;
  DECLARE r_id INT;
  DECLARE u_id INT;
  DECLARE m_id INT;

  DECLARE cur CURSOR FOR
    SELECT response_id, user_id, meeting_id
    FROM form_responses
    WHERE is_processed = 0;

  DECLARE CONTINUE HANDLER FOR NOT FOUND SET done = 1;

  OPEN cur;

  read_loop: LOOP
    FETCH cur INTO r_id, u_id, m_id;
    IF done THEN
      LEAVE read_loop;
    END IF;

    -- Check attendance
    IF (SELECT COUNT(*) FROM attendance
        WHERE user_id = u_id AND meeting_id = m_id) > 0 THEN
      UPDATE form_responses
      SET is_processed = 1
      WHERE response_id = r_id;

      INSERT INTO processing_log (meeting_id, user_id, status, error_message)
      VALUES (m_id, u_id, 'processed', NULL);
    ELSE
      INSERT INTO processing_log (meeting_id, user_id, status, error_message)
      VALUES (m_id, u_id, 'error', 'No attendance record');
    END IF;

  END LOOP;

  CLOSE cur;
END$$
```

Execution:

```sql
CALL process_responses();
```

log_id	meeting_id	user_id	status
1	1	1	processed
2	1	2	error
3	2	2	processed
4	2	3	error
5	1	2	error
6	2	3	error
7	1	2	error
8	2	3	error
NULL	NULL	NULL	NULL

Results and Reports

Reports Generated:

After processing the feedback responses, we analyze the results to identify:

1. Total valid responses per meeting.
2. Total errors (invalid responses) per meeting.
3. User-level participation analysis.

SQL Queries

1. Valid Responses per Meeting

```sql
SELECT m.meeting_name, COUNT(fr.response_id) AS valid_responses
FROM form_responses fr
JOIN meetings m ON fr.meeting_id = m.meeting_id
WHERE fr.is_processed = 1
GROUP BY m.meeting_name;
```
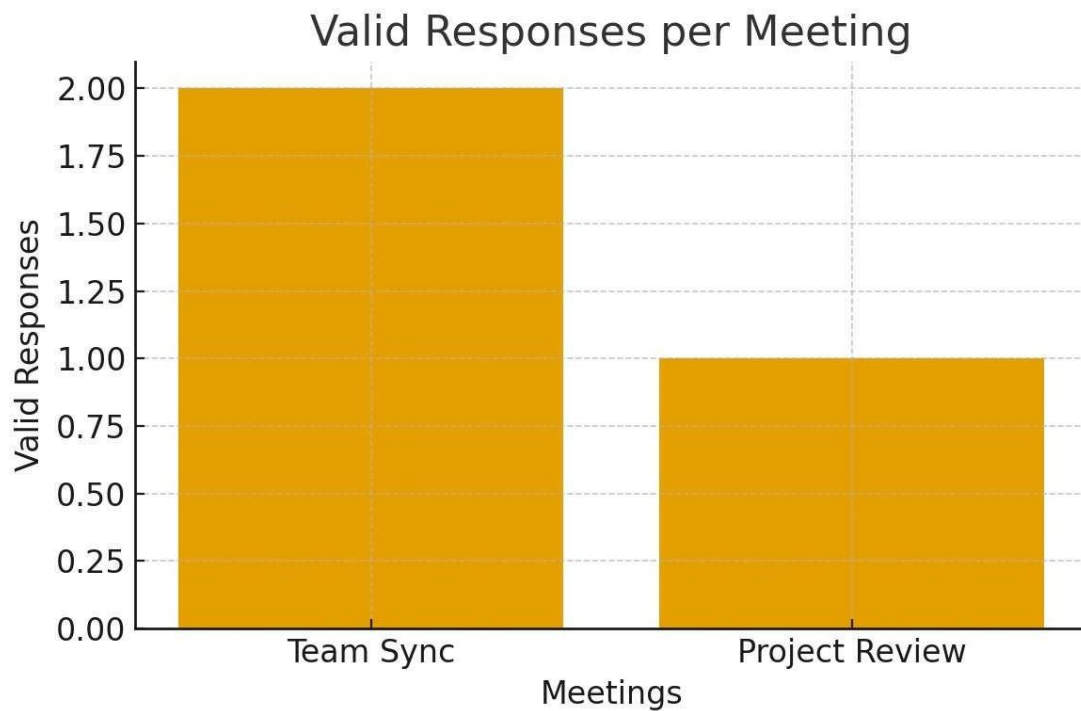
2. Invalid Responses per Meeting

```sql
SELECT m.meeting_name, COUNT(pl.log_id) AS invalid_responses
FROM processing_log pl
JOIN meetings m ON pl.meeting_id = m.meeting_id
WHERE pl.status = 'error'
GROUP BY m.meeting_name;
```
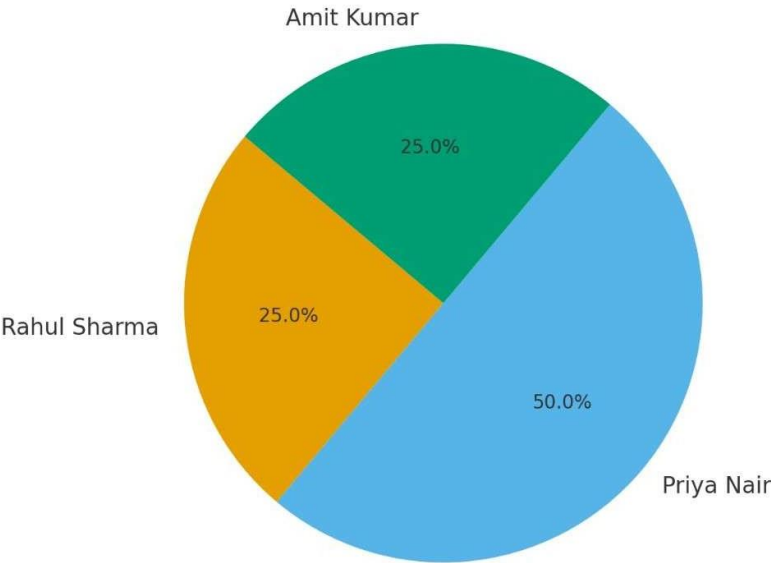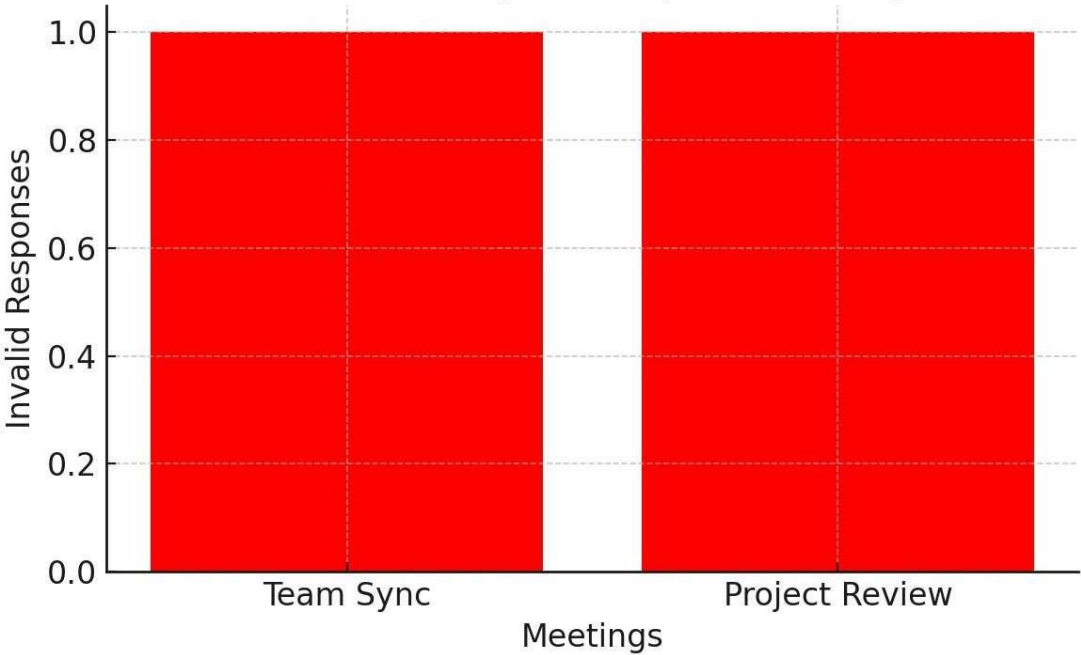
3.User response validation

```sql
SELECT u.name, u.email,
       COUNT(a.attendance_id) AS total_meetings_attended,
       COUNT(fr.response_id) AS feedback_given
FROM users u
LEFT JOIN attendance a ON u.user_id = a.user_id
LEFT JOIN form_responses fr ON u.user_id = fr.user_id
GROUP BY u.user_id;
```

Valid Responses per Meeting

User Participation (Feedback Given)

Amit Kumar — 25.0%
Rahul Sharma — 25.0%
Priya Nair — 50.0%

Invalid Responses per Meeting

Meetings: Team Sync (1.0), Project Review (1.0)

Conclusion and Future Scope

Conclusion

This project successfully demonstrates how feedback validation can be automated using MySQL database design and stored procedures. By integrating Google Form data with attendance records, the system ensures that only genuine participants' feedback is considered valid.

Key achievements:

- Designed a relational database with Users, Meetings, Attendance, and Feedback tables.
- Imported real-time Google Form data into MySQL.
- Implemented a stored procedure to validate responses automatically.
- Generated logs for both valid and invalid feedback entries.
- Produced analytical reports on participation and meeting engagement.

The solution shows that **SQL-based systems** can effectively replace manual validation, reduce errors, and improve transparency in feedback collection.

<u>Future Scope</u>

The project can be further extended with:

1.Automation via Python – Automatically fetch Google Sheets data into MySQL using scripts.

2.Dashboard Integration – Create a real-time dashboard (using Power BI, Tableau, or Flask Web App

3.Email Notifications – Send automatic alerts to participants with invalid feedback.

4.Scalability – Extend system to handle multiple organizations and larger datasets.

Limitations

- · Current implementation requires manual import of Google Form data into MySQL.

- · The validation process works only if emails in Google Form exactly match the database records.

- · Reporting is SQL-based; visualization requires export to external tools.