Database Management Systems
Project 2

Vikas Thammanna Gowda

Due: 03/07/2025

1 Project Overview

1.1 Context

You are acting as a data analyst for a fictional DVD rental business represented by the Sakila sample database (`https://dev.mysql.com/doc/sakila/en/`), and table details (`table details`). Your mission is to extract actionable insights from the database by writing a series of SQL queries (ranging from simple to complex), integrating those queries into a Python workflow, and visualizing your findings with matplotlib. Your final deliverable will demonstrate not only your technical prowess but also your ability to craft a compelling data-driven story.

1.2 Tools & Technologies

- **Database:** MySQL (Sakila sample database)

- **Programming Language:** Python

- **Libraries:** A MySQL connector (such as `mysql.connector`, `pymysql`, or `SQLAlchemy`), pandas for data manipulation, and matplotlib for visualizations.

1.3 Objectives

- Master the creation and optimization of SQL queries—including advanced joins, aggregations —to extract meaningful insights from a relational database.

- Establish robust connections between Python and the MySQL Sakila database to retrieve and process data effectively.

- Use matplotlib to design clear and impactful visualizations that enhance the understanding of complex data analyses.

- Develop a coherent story that translates raw data into actionable business insights and strategic recommendations.

- Combine SQL, Python, and visualization skills to tackle real-world data challenges and derive advanced analytical insights

1.4 Learning Outcomes

By completing this project, students will be able to:

- Demonstrate the ability to write and optimize sophisticated SQL queries that extract accurate and relevant data.

- Show competence in connecting Python with a MySQL database and efficiently retrieving data using appropriate libraries.

- Create clear, well-labeled visualizations that communicate the results of your data analysis effectively.

- Analyze and interpret data to identify trends, patterns, and anomalies that can inform business decisions.

- Collaborate effectively in a group and document individual contributions.

- Produce a well-organized analytical report that integrates methodology, visualizations, and narrative discussions of the findings.

Portfolio Update

Below are some examples of how this project can be added to your portfolio.

Data Analysis Project Developed advanced SQL queries and integrated Python with MySQL using `mysql.connector` to extract and analyze data from the Sakila sample database, resulting in actionable business insights and a comprehensive analytical report.

<div align="center">OR</div>

End-to-End Data Pipeline: Created an end-to-end data analysis workflow by connecting Python to MySQL using `mysql.connector`, processing query results with pandas, and presenting findings through compelling visualizations, which informed strategic decision-making for a fictional DVD rental business.

2 Project Structure & Tasks

Your project is divided into several analytical sections. Each section must include the following components:

- **Task Description:** A clear statement of what is to be analyzed.

- **Queries:** Develop at least three SQL queries (per section) that extract the necessary data.

- **Visualizations:** Produce at least one visualization per query (or query set) using matplotlib.

- **Narrative Report:** Provide a written analysis discussing the results and what they imply for the business. Interpret the visualizations and any patterns or anomalies that emerge.

2.1 Section 1: Customer and Rental Insights

Task: Analyze customer behavior by extracting and visualizing data related to:

- Total rentals per customer

- Average rental duration (calculated from rental and return dates)

- Distribution of rental counts (e.g., segmented by month or by customer group)

Queries Requirement: Develop a minimum of three queries to extract:

- Overall customer rental counts

- Computed average rental durations (only including returned rentals)

- Rental activity during a specified period (e.g., a selected month)

Report & Visualization: For each query:

- Create a corresponding visualization (e.g., bar chart, scatter plot, histogram) that clearly presents the data.

- Discuss what the results reveal about customer engagement, rental behavior, and any emerging trends or anomalies.

2.2 Section 2: Film and Category Analysis

Task: Evaluate film performance by analyzing:

- Total rentals per film category

- Average rental rate for each category

- The top 5 most rented films within one or more specific categories

Queries Requirement: Develop a minimum of three queries to extract:

- Aggregated rental counts for each film category

- Average rental rate figures grouped by category

- A ranking of films by rental count within a chosen category

Report & Visualization: For each query:

- Include a visualization (such as a bar chart or line graph).

- Provide a detailed narrative discussing how different film categories perform, any relationships between rental rates and rentals, and what the top rented films indicate about customer preferences.

2.3 Section 3: Actor and Film Performance

Task: Assess actor performance by linking the rental performance of films to the actors featured in them. Your analysis should address:

- Overall rental counts associated with each actor.

- Comparative performance of actors within a specific film category.

- Identification of the top 10 actors based on aggregated rental counts.

Queries Requirement: Develop a minimum of three queries to extract:

- Total rental counts per actor across the entire dataset.

- A focused analysis for a specific category (e.g., Comedy or Action).

- A ranking of the top 10 actors by total rental counts.

Report & Visualization: For each query:

- Create a corresponding visualization (such as a horizontal bar chart or grouped bar chart).

- Discuss which actors drive the most rentals, how actor performance varies across film categories, and the potential business implications for marketing or film promotion.

2.4 Section 4: Revenue and Store Analysis

Task: Perform an advanced analysis of revenue and payment trends by:

- Calculating total revenue per store.

- Analyzing the average payment amount per rental.

- Tracking monthly revenue trends.

Queries Requirement: Develop a minimum of three queries to extract:

- Aggregated revenue figures for each store.

- Average payment details per rental transaction.

- Monthly revenue trends based on payment dates.

Report & Visualization: For each query:

- Produce a visualization (such as a bar chart for store revenue, a box plot for payment distribution, or a line graph for monthly trends).

- Discuss the revenue performance of each store, interpret trends in customer spending, and suggest potential business strategies based on the observed trends.

3 Python Integration Guidelines

In your Python script, include:

- **Database Connection:** Connect to the Sakila database using MySQL.

- **Query Execution and Data Loading:** Execute SQL queries and load results into pandas DataFrames.

- **Visualization:** Use matplotlib to create labeled, insightful visualizations.

- **Comments and Documentation:** Ensure code is well-commented and structured.

4   Write-Up Structure

Your final submission must include:

Title Page

- Course Number and Title

- Project Title

- Names of Group Members

- Date of Submission

Abstract

A concise summary of the write-up, including the purpose, key findings, and future perspectives (100–150 words).

1. Introduction

- Overview of the Sakila database and business context.

- Objectives of your analysis.

2. Methodology

- Description of your Python environment and database connection.

- Overview of the approach for each analysis section.

3. Analysis Sections

For each section:

- **Task Description:** Restate the specific analysis task.

- **Visualizations:** Include each chart with captions and descriptions.

- **Discussion:** Provide a narrative interpreting the results, discussing patterns, anomalies, and business implications.

4. Individual Contributions

- Documentation of each member's role in the project.

- Reflective paragraph from each member on their experience and learning.

5. Conclusion and Recommendations

- Summarize key findings.

- Propose actionable recommendations based on the analysis.

Appendix (Optional)

- Include additional code snippets or extended analyses.

5   Grading Criteria

Criteria	Weightage
SQL Query Development - All specified tasks are addressed with a minimum of three queries per section. - Use of advanced SQL features (joins, aggregations) where appropriate to extract deeper insights. - Queries return accurate and meaningful results.	30%
Clarity and Organization - Well-structured write-up with logical flow. - Proper formatting and cohesive narrative.	15%
Python Integration - Database Connection and Data Extraction. - Code Quality and Organization.	10%
Narrative Analysis and Reporting with Visualizations - The report offers a clear, compelling narrative that interprets the results. - Appropriate chart types are chosen for different types of data analysis. - Visualizations are clearly labeled, easy to interpret, and professional in appearance.	30%
Individual Contributions - Clear documentation of group members' roles. - Reflective paragraphs from each member.	15%
Total	**100%**

Submission Details

1. **Report:** The final write-up must be submitted in `.pdf` format.

   - **File Format:** The final write-up must be submitted in `.pdf` format.

   - **Font Style:** Times New Roman.

   - **Font Sizes:**

     – Section Titles: 14pt, **Bold**. Subsection Titles: 12pt, **Bold**. Body Text: 12pt.

   - **Text Alignment and Page Numbers:** The entire document must be justified. Include page numbers in the footer of each page.

   - **Line Spacing and Margins:** Set line spacing to 1.15 throughout the document. Use standard margins (1-inch on all sides).

   - **Visualizations:** All the plots must be complete, be of the same size, and be centered with a figure number and a figure name.

   - **Word count:** Minimum 2000 words excluding code blocks.

2. **Code:** Compress all the Python scripts into a single ZIP folder.

Academic Integrity Policy

   - All submitted work must be the result of the group's original research, analysis, and writing.

   - The use of AI tools (e.g., ChatGPT, Bard, etc.) to generate or assist in writing the report is strictly prohibited. Any detected use of AI tools will result in an automatic **F grade** for the project.

   - Proper citations and references must be provided for any external sources consulted. Plagiarism will not be tolerated and will be subject to academic penalties.