

Emerald Cohort SQL Data Analysis Capstone Project

Project Title: Multi-Table Data Exploration and Insights

Overview

Dear SQL Data Analyst,

Congratulations on embarking on this exciting journey to showcase your SQL expertise! This capstone project is designed to mirror real-world challenges and provide you with the opportunity to work with multi-table datasets in a professional context. By diving into a sector of your choice, you will demonstrate your ability to analyze data, uncover actionable insights, and communicate your findings effectively.

Your task is to explore a dataset consisting of at least three related tables, with the main table containing a minimum of five columns. This dataset should focus on a specific sector such as healthcare, security, e-commerce, etc. allowing you to address relevant questions and trends within that field.

Through this project, you will hone your SQL skills, develop business or research questions, and present your insights using data visualizations. It is an opportunity to practice end-to-end data analysis, from cleaning and transformation to analysis and professional presentation.

We look forward to seeing your creativity, analytical rigor, and storytelling skills come to life in this project.

Best of luck,

The Projects Team,

The Data Immersed Community

Your Mission

Using SQL, your mission is to:

1. Explore and understand the dataset to uncover meaningful insights.
2. Develop and answer key business or research questions relevant to your chosen sector.
3. Create data-driven visualizations using Excel to support your findings.
4. Present your work professionally to an audience for review and feedback.

This project is individual, and all data analysis tasks, including cleaning, querying, and transformation, must be performed in SQL by you alone.

Visualizations should complement your insights.

Project Objectives

1. Dataset Selection & Familiarization

- Identify and obtain a high quality dataset with **at least three (3) related tables**.
- Ensure the **main table has at least five columns**, covering key attributes of the dataset.
- Explore relationships between the tables and document how they are linked (e.g., foreign keys, IDs).

2. SQL Utilization

- Perform all data cleaning, transformation, and analysis in SQL.
- Utilize essential SQL techniques such as joins, aggregate functions, conditional logic, and window functions.

3. Data Cleaning & Preparation

- Ensure the dataset is clean and ready for analysis:
 - Handle missing or inconsistent data.
 - Remove duplicates.
 - Format and structure data for meaningful insights.
- Define relationships between the tables and use them effectively in your analysis.

4. Research Question Development

- **Formulate at least three key questions** to guide your analysis.
- Ensure your questions address trends, performance, or actionable insights.

5. Detailed Analysis & Insights

- Perform in-depth analysis using SQL queries to:
 - Analyze trends, patterns, and outliers.
 - Uncover relationships between different tables.

- Provide actionable recommendations based on findings.
- Use advanced SQL techniques, including CASE statements, CTEs, and subqueries.

6. **Final Deliverables** Prepare the following deliverables to showcase your work:

I. SQL Scripts

- A comprehensive SQL script documenting all data cleaning, transformation, and analysis processes.
- Include comments for clarity and documentation.

II. Presentation

- Create a concise PowerPoint presentation summarizing your findings.
- Include key visuals and actionable recommendations.
- Ensure your presentation is polished and professional.

III. Detailed Report Prepare a written report that documents: - Dataset overview and structure. - Research questions and methodology. - Key findings, insights, and recommendations. - Challenges faced and solutions applied. - Include SQL queries as an appendix.

IV. Visualizations

- **Use Microsoft Excel** to create a meaningful interactive dashboard.
- Ensure visuals align with your narrative and clarify your insights.

Submission and Timeline

This capstone project is expected to last **2 weeks from the date it is given**. Also take note of the following:

1. Submission:

- Submit your SQL scripts, report, visualizations, and presentation materials.

2. Presentation:

- Prepare to present your findings to an audience for review and feedback.

3. Revisions:

- Make necessary corrections based on feedback and submit the final polished version.

Evaluation Criteria

Your project will be assessed based on:

1. Data Understanding & Preparation:

- Depth of data exploration.
- Effectiveness of cleaning and preparation steps.

2. Insightfulness of Analysis:

- Relevance and clarity of insights.
- Alignment with research questions.

3. SQL Mastery:

- Effective use of SQL for cleaning, querying, and transforming data.
- Complexity and efficiency of SQL queries.

4. Visualization & Presentation:

- Clarity and quality of visualizations.
- Professionalism and coherence of presentation and report.

5. Documentation:

- Thoroughness and structure of your SQL scripts, report, and presentation.

Conclusion

This capstone project provides an opportunity to showcase your SQL expertise and analytical thinking. By working with a real-world dataset, you will demonstrate your ability to derive meaningful insights and communicate them effectively to stakeholders. Success in this project depends on attention to detail, problem-solving, and clear communication.

Good luck and enjoy your data analysis journey!