**Project Instructions: SQL Analysis on a Schema**

**Objective**

The objective of this project is to perform SQL queries on a schema, demonstrating your understanding of SQL concepts and data analysis techniques. You will write and execute queries to extract meaningful insights from the data and present your findings.

**Steps and Guidelines**

1. **Set Up Your Environment**
   * Ensure you have access to a SQL database (MySQL).
   * Download and set up a schema.
2. **Project Setup**
   * Create a new directory for your project.
   * Use a SQL script for your analysis.
   * Create a README file that briefly describes the dataset and the steps you plan to take in your analysis.
3. **Data Import and Schema Overview**
   * Import the data into your SQL database.
   * Provide an overview of the schema with ER diagram.
   * Use DESCRIBE or SHOW COLUMNS commands to understand the structure of each table.
4. **Data Analysis Using SQL**
   * **Descriptive Statistics**:
     + Write queries to calculate basic statistics such as count, sum, average, minimum, and maximum for numerical columns.
     + Example: Calculate the total sales, average order value, and total number of transactions.
   * **Data Cleaning**:
     + Write queries to identify and handle missing or inconsistent data.
     + Example: Find rows with NULL values in important columns and decide how to handle them (e.g., filling, removing).
   * **Aggregation and Grouping**:
     + Use GROUP BY to aggregate data by different dimensions.
     + Example: Calculate total sales by product category, region, or time period.
   * **Joins and Relationships**:[ ]
     + Write queries to join tables to extract meaningful insights.
     + Example: Join the sales with the product table to get product names and categories.
   * **Subqueries and CTEs**:
     + Use subqueries and Common Table Expressions (CTEs) to write complex queries.
     + Example: Use a CTE to calculate the monthly sales and then find the month with the highest sales.
   * **Advanced SQL Functions**:
     + Use window functions to perform calculations across a set of table rows related to the current row.
     + Example: Calculate running totals or moving averages.
5. **Insights and Conclusions**
   * Summarize your key findings from the SQL analysis.
   * Discuss any patterns, anomalies, or interesting relationships you discovered.
   * Highlight any potential areas for further analysis or questions that emerged from your analysis.
6. **Documentation and Presentation**
   * Ensure your SQL script is well-documented. Include comments to explain each query, the rationale behind your choices, and your findings.
   * Prepare a brief presentation (5-10 slides) summarizing your project. Include key findings, interesting queries, and potential next steps.
   * Submit your SQL script, the README file, and the presentation slides.

**Submission Deadline**

* Please submit your project by [Insert Deadline Here].

If you have any questions or need further assistance, feel free to reach out during office hours or via email.

### Evaluation Rubric for SQL Analysis Project

**Total Marks: 20**

| **Criteria** | **Description** | **Marks** | **Scoring Details** |
| --- | --- | --- | --- |
| **1. Environment and Project Setup (2 Marks)** | | | |
| Environment Setup | Ensure access to a SQL database (MySQL) and setup of the schema. | 1 | 1: Fully set up, 0.5: Partially set up, 0: Not set up |
| Project Directory and Script | Creation of a new directory and use of a SQL script for analysis. | 1 | 1: Fully created, 0.5: Partially created, 0: Not created |
| **2. Data Import and Schema Overview (2 Marks)** | | | |
| Schema Overview | Overview of the schema with ER diagram and table descriptions. | 1 | 1: Comprehensive, 0.5: Partial, 0: Missing or unclear |
| Table Structure | Use of DESCRIBE or SHOW COLUMNS to understand table structure. | 1 | 1: Thoroughly understood, 0.5: Partially understood, 0: Not understood |
| **3. Data Analysis Using SQL (8 Marks)** | | | |
| Descriptive Statistics | Queries to calculate basic statistics (count, sum, average, min, max). | 2 | 2: Comprehensive, 1: Partial, 0: Missing or incorrect |
| Data Cleaning | Queries to identify and handle missing or inconsistent data. | 2 | 2: Effectively handled, 1: Partially handled, 0: Not handled |
| Aggregation and Grouping | Use of GROUP BY to aggregate data by different dimensions. | 2 | 2: Effectively used, 1: Partially used, 0: Not used |
| Joins and Relationships | Queries to join tables to extract meaningful insights. | 2 | 2: Comprehensive, 1: Partial, 0: Missing or incorrect |
| **4. Subqueries and CTEs (2 Marks)** | | | |
| Subqueries and CTEs | Use of subqueries and Common Table Expressions (CTEs) for complex queries. | 2 | 2: Effectively used, 1: Partially used, 0: Not used |
| **5. Advanced SQL Functions (2 Marks)** | | | |
| Advanced SQL Functions | Use of window functions to perform calculations across a set of table rows. | 2 | 2: Effectively used, 1: Partially used, 0: Not used |
| **6. Insights and Conclusions (2 Marks)** | | | |
| Insights | Summary of key findings, patterns, anomalies, or interesting relationships. | 1 | 1: Highly significant and relevant, 0.5: Somewhat significant, 0: Not significant |
| Conclusions | Clear and well-supported conclusions based on analysis. | 1 | 1: Clear and well-supported, 0.5: Somewhat clear, 0: Unclear or unsupported |
| **7. Documentation and Presentation (2 Marks)** | | | |
| Documentation | Well-documented SQL script with comments explaining each query. | 1 | 1: Clear and thorough, 0.5: Partially clear, 0: Unclear or missing |
| Presentation | Brief presentation (5-10 slides) summarizing the project with key findings. | 1 | 1: Clear and well-organized, 0.5: Partially clear, 0: Unclear or disorganized |

### Summary of Marks:

* **Environment and Project Setup**: 2 marks
* **Data Import and Schema Overview**: 2 marks
* **Data Analysis Using SQL**: 8 marks
* **Subqueries and CTEs**: 2 marks
* **Advanced SQL Functions**: 2 marks
* **Insights and Conclusions**: 2 marks
* **Documentation and Presentation**: 2 marks