SQL Intermediate Project Brief

Project Title: Intermediate Customer Data Management and Analysis using SQL

Objective:

Develop an intermediate-level SQL project to manage and analyze customer data. The project will focus on advanced querying techniques, data manipulation, and generating insights from the dataset provided.

Dataset:

The dataset includes comprehensive customer demographic information, purchasing behavior, and responses to marketing campaigns.

Tasks:

1. Database Setup:

- Create a new database to store the customer data.
- Define appropriate data types for each column and create tables.
- Load the dataset into the database.

2. Data Cleaning and Preparation:

- Identify and handle missing or null values.

- Correct data types where necessary (e.g., convert `Income` from string to numeric).

- Ensure that date fields are in the correct format.

3. Advanced Querying:

- Write complex queries to extract meaningful insights from the data.

- Use aggregate functions (SUM, AVG, COUNT, MAX, MIN) to calculate summary statistics.

- Implement JOIN operations to combine data from multiple tables if necessary.

- Utilize subqueries and common table expressions (CTEs) for more advanced data manipulation.

4. Customer Segmentation:

- Create segments of customers based on demographics, purchasing behavior, and responses to campaigns.

- Write queries to identify top-performing segments based on total spending, frequency of purchases, or campaign responses.

5. Marketing Campaign Analysis:

- Analyze the effectiveness of different marketing campaigns.

- Write queries to calculate the response rates for each campaign.

- Identify customers who have accepted multiple campaigns and analyze their behavior.

6. Recency, Frequency, Monetary (RFM) Analysis:

- Perform an RFM analysis to identify high-value customers.

- Write queries to calculate recency, frequency, and monetary value for each customer.

- Segment customers into different RFM categories.

7. Optimization and Indexing:

- Optimize SQL queries for performance.

- Implement indexing strategies to speed up query execution.

- Analyze query execution plans to identify and resolve performance bottlenecks.

8. Documentation:

- Document the database schema and the data cleaning process.

- Provide detailed explanations of each query and its purpose.

- Include a summary report highlighting the key insights and findings from the analysis.

Tools and Technologies:

- SQL Database Management System: MySQL, PostgreSQL, SQL Server, or any other preferred DBMS.

- SQL Clients: MySQL Workbench, pgAdmin, SQL Server Management Studio (SSMS), or any other SQL client for querying and visualization.

- Jupyter Notebook with SQL extension (optional): For combining SQL queries with explanatory text and visualizations.

Evaluation Criteria:

- Quality and accuracy of data cleaning and preparation.

- Complexity and efficiency of SQL queries.

- Depth of analysis and insights generated from the data.

- Effectiveness of customer segmentation and campaign analysis.

- Quality of documentation and reporting.

- Performance optimization and indexing strategies.

Submission Requirements:

- SQL script files (.sql) containing all the queries used in the project.

- A summary report in PDF format outlining the key insights, findings, and recommendations.

Timeline:

- Deadline: 15 days from the project initiation date.

- Submission Mode: Upload the SQL script files and report to a designated platform or email them to the project coordinator.

Data Dictionary:

- ID: Unique identifier for each customer.

- Year_Birth: Year the customer was born.

- Education: Level of education attained by the customer.

- Marital_Status: Marital status of the customer.

- Income: Annual income of the customer (in dollars).

- Kidhome: Number of children in the customer's household.

- Teenhome: Number of teenagers in the customer's household.

- Dt_Customer: Date when the customer was registered.

- Recency: Number of days since the customer last made a purchase.

- MntWines: Amount spent on wines in the last 2 years.

- MntFruits: Amount spent on fruits in the last 2 years.

- MntMeatProducts: Amount spent on meat products in the last 2 years.

- MntFishProducts: Amount spent on fish products in the last 2 years.

- MntSweetProducts: Amount spent on sweet products in the last 2 years.

- MntGoldProds: Amount spent on gold products in the last 2 years.

- NumDealsPurchases: Number of purchases made with a discount.

- NumWebPurchases: Number of purchases made through the web.

- NumCatalogPurchases: Number of purchases made using a catalog.

- NumStorePurchases: Number of purchases made directly in stores.

- NumWebVisitsMonth: Number of visits to the company's website in the last month.

- AcceptedCmp3: 1 if the customer accepted the offer in the 3rd campaign, 0 otherwise.

- AcceptedCmp4: 1 if the customer accepted the offer in the 4th campaign, 0 otherwise.

- AcceptedCmp5: 1 if the customer accepted the offer in the 5th campaign, 0 otherwise.

- AcceptedCmp1: 1 if the customer accepted the offer in the 1st campaign, 0 otherwise.

- AcceptedCmp2: 1 if the customer accepted the offer in the 2nd campaign, 0 otherwise.

- Response: 1 if the customer accepted the offer in the last campaign, 0 otherwise.

- Complain: 1 if the customer complained in the last 2 years, 0 otherwise.

- Country: Country where the customer resides.