Customer RFM Analysis with SQL

Goal:
This project aims at designing and implementing an intermediate-level SQL solution to manage and analyze customer data. It focuses on applying advanced SQL queries for data extraction, performing data transformation to enhance usability, and generating meaningful insights to support data-driven decision-making.

Dataset Overview:
The dataset contains detailed customer demographics, purchase history, and responses to marketing initiatives.

Key Tasks:

1. **Database Creation**

   - Build a new database for storing customer records.
   - Assign proper data types for each column and the necessary tables.
   - Import the dataset into the database.

2. **Data Cleaning & Preparation**

   - Detect and address missing values.
   - Adjust incorrect data types (e.g., converting year_birth from text to int).
   - Standardize date formats.

3. **Advanced SQL Queries**

   - Develop complex queries to uncover valuable insights.
   - Apply aggregate functions (SUM, AVG, COUNT, MAX, MIN) for summaries.
   - Use JOINs to merge information across tables.
   - Utilize subqueries and CTEs for advanced data handling.

4. **Customer Segmentation**

   - Segment customers by demographics, spending habits, and campaign engagement.
   - Identify high-value customers based on spending, purchase frequency, and campaign responses.

5. **Marketing Campaign Evaluation**

    - Assess campaign effectiveness through response rates.
    - Identify customers engaging with multiple campaigns and study their behavior.

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

    - Perform RFM analysis to classify customers by value.
    - Create SQL queries to calculate recency, frequency, and monetary value of customers.
    - Group customers into RFM-based segments.

7. **Optimization & Indexing**

    - Enhance query performance.
    - Apply indexing strategies to improve execution speed.

8. **Documentation & Reporting**

    - Document database schema and cleaning process.
    - Provide detailed explanations of queries and their objectives.
    - Deliver a summary report of findings and key insights.

Tools:

- SQL DBMS (MySQL, PostgreSQL, SQL Server, etc.)
- SQL clients (MySQL Workbench, pgAdmin, SSMS, etc.)
- Power BI

Evaluation Criteria:

- Accuracy of data cleaning
- Efficiency and complexity of SQL queries
- Depth of insights generated
- Effectiveness of segmentation and campaign evaluation
- Quality of documentation and reporting
- Query optimization and indexing

Submission:

- SQL scripts (.sql) for all queries
- PDF summary report with insights and recommendations

Timeline:

- Duration: 15 days from project kickoff
- Submission: Upload SQL scripts and report to the assigned platform or email

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.