Problem Statement for SQL Project: Customer Behavior and Churn Analysis

Project Overview

The project aims to analyze customer behavior, order patterns, churn rates, and various other metrics using the Northwind database. It explores trends in order frequency, customer segmentation, churn prediction, and customer lifetime value (CLTV). The goal is to derive actionable insights that could help businesses understand customer behavior, identify at-risk customers, and improve retention strategies.

Key Objectives:

1. **Order Frequency Analysis**: Investigate the frequency of customer orders over time and compute the average time between orders for each customer.
2. **Customer Segmentation**: Segment customers based on their purchasing frequency (Frequent, Regular, and Infrequent Buyers) and analyze the average time gap between their orders.
3. **Churn Analysis**: Determine the average order value for customers who have churned vs. those who have remained active.
4. **Order Value Distribution**: Categorize customers based on their total order value (Low, Medium, High) and identify the distribution of these categories across the customer base.
5. **Churn Impact on Product Categories**: Identify product categories that are frequently purchased before and after a churn event.
6. **Churn by Region**: Investigate churn rates across different customer locations (cities, states, and countries) to identify regions with high churn.
7. **Purchase Behavior by Region**: Examine the correlation between customer location and purchase behavior to assess any geographical influences on buying patterns.
8. **Customer Risk Scoring**: Assign a risk score to each customer based on their order frequency, spending behavior, and product category preferences.
9. **Order Frequency in the Last 6 Months**: Measure the number of orders placed by customers in the last 6 months to identify any recent changes in order behavior.
10. **Decreasing Order Frequency**: Identify customers with a decrease in order frequency over the last 6 months compared to the average for all customers.
11. **Customer Lifetime Value (CLTV)**: Calculate the Customer Lifetime Value (CLTV) for each customer based on their spending, order frequency, and other relevant factors to prioritize retention efforts.
12. **Customer Acquisition Trends**: Trends in customer acquisition over time, such as how many new customers are added each month.

Approach and Methodology:

1. **Data Preparation and Cleaning**:
    - Use SQL queries to extract data from relevant tables such as orders, order_details, customers, and products.
    - Handle missing or inconsistent data by using appropriate joins, aggregation, and filtering techniques.
2. **Trend Analysis**:
    - Utilize window functions like LAG() to calculate the time gap between customer orders and derive trends over time.
    - Analyze customer order patterns to understand the frequency and distribution of orders.
3. **Segmentation and Classification**:
    - Categorize customers into segments (Frequent, Regular, Infrequent) based on their order frequency.

    - Analyze the order behavior within these segments to understand their purchasing habits.
4. **Churn Analysis**:
    - Identify customers who have churned (by checking order status).
    - Calculate metrics like average order value for churned vs. non-churned customers to understand how churn impacts revenue.
5. **Geospatial Analysis**:
    - Analyze churn rates by region (country, state, city) and compare them to determine which areas are experiencing the most churn.
    - Investigate purchasing behavior across regions to identify regional patterns.
6. **Risk Scoring**:
    - Develop a risk model that assigns customers to "Low", "Medium", or "High" risk categories based on their purchase frequency and order value.
    - This can be used for targeted marketing and retention strategies.
7. **Lifetime Value Calculation**:
    - Calculate CLTV by considering total order value, discounts, taxes, and shipping fees for each customer.
8. **Trend Analysis**:
    - Trends in customer acquisition over time, such as how many new customers are added each month.

Tools and Technologies:

- SQL: For querying and analyzing the Northwind database.
- Aggregation functions: SUM(), COUNT(), AVG(), GROUP BY, etc.
- Window functions: LAG(), LEAD(), etc.
- CASE statements for segmentation and categorization.

Expected Outcomes:

- Understanding customer purchasing behaviors and trends over time.
- Identifying customers who are at risk of churning and implementing retention strategies.
- Providing insights into the average order value, customer segments, and product categories that influence churn.
- Identifying high-value customers and regions with higher purchasing behavior to help businesses target marketing efforts more effectively.

Deliverables:

1. SQL scripts with detailed queries to perform the above analyses.
2. A summary of key business insights derived from the analysis, including recommendations for improving customer retention and targeting.