Project Brief: Instacart Market Basket Analysis

Overview:

The Instacart Market Basket Analysis project aims to explore and analyze customer shopping behavior using transactional data from the Instacart platform. This dataset provides a rich source of information that can be used to derive valuable insights for optimizing operations and enhancing customer experiences.

Dataset Description:

The project utilizes the following datasets:

- **aisles.csv** - Contains information about different product categories (aisles).
- **departments.csv** - Provides details about various departments within the store.
- **order_products__prior.csv** - Includes information about products included in prior customer orders.
- **order_products__train.csv** - Contains details about products in the training set of customer orders.
- **orders.csv** - Provides information about individual orders and customers.
- **products.csv** - Contains details about products, including aisle and department IDs.

Data Dictionary:

Refer to the provided data dictionary for a detailed description of each dataset and its columns.

Project Tasks:

Task 1:

- Data Import and Cleaning
- Import the datasets into SQL Server.
- Identify and handle missing values.
- Ensure data consistency and integrity.
- Generate the ERD for the various tables showing the relationship between each table.
- Apply necessary data transformations.

Task 2:

- Exploratory Data Analysis (EDA) and Descriptive Statistics
- Conduct exploratory data analysis to understand the characteristics of the data.
- Generate descriptive statistics to gain initial insights.

Task 3: Data Analysis and Insights

1. Market Basket Analysis:

Analysis: Identify frequently co-occurring products in orders to improve store layout and marketing strategies.

Questions:

- What are the top 10 product pairs that are most frequently purchased together?
- What are the top 5 products that are most commonly added to the cart first?
- How many unique products are typically included in a single order?

2. Customer Segmentation:

Analysis: Group customers based on their purchasing behavior for targeted marketing efforts.

Questions:

- Can we categorize customers based on the total amount they've spent on orders?
- What are the different customer segments based on purchase frequency?
- How many orders have been placed by each customer?

3. Customer Lifetime Value (CLV) Prediction:

Analysis: Estimate the potential value a customer will bring to the business over their entire relationship.

Questions:

- What is the average CLV for different customer segments (e.g., high spenders, frequent buyers)?
- How has the CLV changed over the past year?

4. Seasonal Trends Analysis:

Analysis: Identify seasonal patterns in customer behavior and product sales.

Questions:

- What is the distribution of orders placed on different days of the week?
- Are there specific months with higher order volumes?

5. Customer Churn Prediction:

Analysis: Predict which customers are most likely to stop using the service in the near future.

Questions:

- Can we identify customers who haven't placed an order in the last 30 days?
- What percentage of customers have churned in the past quarter?

6. Product Association Rules:

Analysis: Identify rules or patterns in customer behavior indicating which products are frequently bought together.

Questions:

- What are the top 5 product combinations that are most frequently purchased together?
- Can we find products that are often bought together on weekends vs. weekdays?

Task 4: Visualization and Reporting

- Create visualizations to present key findings.
- Generate reports summarizing insights.
- Deliverables:
- SQL scripts for data import, cleaning, and analysis.
- Report summarizing key findings and insights.
- Optional: Visualizations to support the analysis.

Note:

- Ensure that all SQL queries are well-documented for easy replication.
- Also add comments to your codes.

Goodluck!!!!!