Data Analysis to Answer Questions

Session (1)

AL/ Alaa Abdelmoaty Elbiomy

E-Commerce Data Analysis Project

❖ Dataset Overview

The dataset represents simulated e-commerce transactions from an online retail store. It includes the following fields:

1. CustomerID.
2. OrderID.
3. Product.
4. Quantity.
5. UnitPrice.
6. PurchaseDate.
7. Country.

❖ Tasks: [Using Python and SQL]

1. Data Cleaning and Preparation

- Check for missing or inconsistent data (e.g., null values, duplicates) and handle them.
- Convert PurchaseDate to a datetime object and ensure proper date formatting.

2. Filtering Data

2.1 Customer and Order Filters:

1. Filter all transactions made by customers from a specific country (e.g., 'USA').
2. Extract orders where the total spend (Quantity * UnitPrice) exceeds $500.
3. Identify customers who purchased more than 3 different products.

2.2 Time-Based Filters:

1. Filter transactions that occurred in July 2022.

2. Extract orders placed during weekends.

3. Identify transactions during specific sales events, like Black Friday or Cyber Monday.

3. Sorting Data

1. Sort transactions by:

- Total spend in descending order.

- Purchase date in ascending order.

- Product name alphabetically.

2. Rank customers based on their total spending.

4. Aggregated Analysis: Extra grad

4.1 Customer Insights:

1. Calculate the total amount spent by each customer and find the top 10 spenders.

2. Group data by Country and analyze the total revenue generated per country.

3. Identify which country has the highest average transaction value.

4.2 Product Analysis:

1. Find the most purchased product and its total quantity sold.

2. Identify the product that generated the highest revenue.

3. Determine the top 3 least popular products based on sales quantity.

4.3 Time-Based Insights:

1. Analyze the total revenue generated per day in July 2022.

2. Identify peak shopping hours based on the number of transactions.

3. Find the day with the highest total revenue.

5. Advanced Insights: Extra grad

1. Create a column for total spend per transaction (Quantity * UnitPrice).

2. Use quantiles to identify the top 10% of transactions based on total spend.

3. Analyze purchasing trends by country:

- Identify the most popular product in each country.

- Determine the average order value per country.

6. Visualization(using matplotlib) :Extra grad

1. Plot the total revenue per country using a bar chart.
2. Visualize daily revenue trends with a line chart.
3. Use a pie chart to show the distribution of products sold.