PublishingDate: 16 Dec 2023 Submission Date: 08th January 2024 - 23:59

1. Business Context

In the context of e-commerce sales analysis, we have a dataset containing information on customer orders from an online store. The dataset includes details such as order ID, order date, SKU (stock-keeping unit), color, size, unit price, quantity, and revenue. Our objective is to gain insights and make informed decisions regarding product popularity, customer preferences, and revenue generation based on this dataset.

2. Business Problem Understanding

The e-commerce company is seeking assistance in understanding and improving its sales strategies. They have provided us with a comprehensive dataset containing information on customer orders. Our objective is to analyze this dataset and provide actionable insights to enhance product offerings, optimize pricing, and increase overall revenue.

Your team has been appointed to take a closer look at the records of the sales dataset and analyze the patterns and trends in customer orders.

a) Identify factors contributing to high-revenue orders.

b)Recommend strategies for improving product popularity and customer satisfaction.

2. Data Understanding

For this analysis, the company is expecting your team to explore the usage of MongoDB for the storage and querying of the sales data. The data is available in the provided sample dataset. The data dictionary can also be referred to for understanding the attributes in the dataset.

The e-commerce sales dataset provides valuable information on the frequency and impact of customer orders. It includes data on the unit price, quantity, revenue, and various product attributes.

Sales Dataset Schema:

order_id: (Int32) Unique identifier for each order.

order_date: (Date) Date and time when the order was placed.

sku: (String) Stock-keeping unit, a unique identifier for each product.

color: (String) Color of the product.

size: (String) Size of the product.

unit_price: (Int32) Unit price of the product.

quantity: (Int32) Quantity of the product ordered.

revenue: (Int32) Total revenue for the specific order.

3. Data preparation and Exploratory Data Analysis

You are supposed to utilize appropriate data pre-processing techniques on the given data set. If required, make appropriate assumptions and make it explicitly known while using them in the query. Make appropriate selection of the attributes with sound justification for the same. The data set allows for several new combinations of attributes and attributes exclusions, or the modification of the attribute type (categorical, integer, or real) depending on the purpose of the analysis.

4. Expected Outcomes

You are expected to find out the answers to following questions

1. **What is the total revenue generated by the e-commerce store in June 2022?**
2. **What is the highest revenue of a SKU in a single order?**
3. **Find the average unit price per SKU for the "Dark Blue" color.**
4. **Identify the top 5 best-selling colors based on quantity sold.**
5. **Calculate the total revenue for orders with "Dark Blue" products in XL size.**
6. **What is the average quantity of products per order for "Dark Blue" SKUs?**
7. **Find the total revenue for each day in June 2022 and identify the highest revenue day.**
8. **Calculate the average unit price for each SKU.**
9. **Identify the SKU with the highest total quantity sold.**

10. **Find the revenue contribution of "Dark Blue" products compared to "Navy Blue" products.**

11. **Calculate the total revenue for each size category (e.g., XL, M, 3XL).**

12. **Identify the top 3 SKUs with the highest average revenue per order.**

5. Instructions for solving the assignment

a) **The assignment consists of 12 questions.Question 1 to 11 carries 1 mark each and Question 12 is for 2 marks (Total 13 Marks)**

b) **Each team consists of 3 students.**

The Team members are expected to solve the questions as given below

First student of the batch is expected to solve q.no1 to q.no4,
Second student of the batch is expected to solve q.no5 to q.no8,
Third student of the batch is expected to solve q.no9 to q.no12

The Group leader will be giving a contribution weightage of each group members and marks will be awarded accordingly

c) **The submission should consist of 2 files:**

A PDF file containing answers to the twelve questions based on the analysis that you have carried out earlier along with the supporting MongoDB queries that you have written to extract the answers.

- Name the PDF file in format like "Grp_<your_group_number>.doc" only. Don't add anything into the file names. Add the group member names in the PDF.
- Make sure that you upload the file well ahead of deadline. At last moments, we have seen several groups have faced issues while doing the submissions.
- Note - Since it is a group assignment, only one submission is expected from each group. Unnecessarily don't upload the solution on individual basis. If it's observed, then a penalty (25% reduction) will be applicable on it.

d) **Every group should record a mp4 video which should contain the executions with queries/answers.**

- Name the video file in format like "Grp_<your_group_number>.mp4" only. Don't add anything into the file names.

e) **Plagiarism will be strictly dealt with and if found will result in cancellation of the Assignment and 0 marks being awarded to all the group members**.

f) **Submissions done as mail attachments will be rejected.**

g) **Images/photographs of text/PDF documents will not be evaluated.**

h) **The last date of submission will not be extended in any case.**

7. MongoDB instance: You can use any instance in lab, local, on Cloud. Some example pointers given below.

a) https://www.mongodb.com/try/download/community

b) https://www.mongodb.com/cloud/atlas/register

8. References

- Women Clothing Ecommerce Sales Data(Import Required Fields according to the queries)
- MongoDB documentation
- Groups information