
Problem Statement:

You have been given the role of a business analyst for an E-Commerce company and have been asked to prepare a basic report on the data. Here, we are analysing different aspects of E-Commerce Company by doing Exploratory Data Analysis.

Domain: Sales Analytics

Tasks/ Questions to be Answered:

1. To get familiar with the data (9 marks)
    a. Print out the first 10 and the last 10 records of the data. (2 marks)
    b. How many rows and columns are present in the dataset? Use any two different methods to extract this information. (2 + 2 + 1 marks)
    c. How many object data types are there? (1 mark)
    d. Is there any Boolean data type? (1 mark)

2. Eliminating the non-informative columns. (4 marks)
    a. Drop the columns product_specifications and description. (2 marks)
    b. Which method or function is used to permanently delete the columns mentioned in part (b)? Write the code explicitly (2 marks)

3. Here we summarize the data at brand level. (3 Marks)
    a. How many unique Brands are there? (1 mark)
    b. Note that each brand contains multiple products. Show the average rating of the products within each Brand (2 marks)

4. Next, we study the main categories of the products. (9 Marks)
    a. Create an appropriate plot to show the count of items ordered for each product_main_category. (5 Marks)
    b. From the plot identify for which two product_main_category(s) the maximum and the minimum orders were placed (2 Marks)
    c. Write code to print out the top 5 product_main_category(s) in descending order? (2 Marks)

5. Find the net revenue generated by the E-Commerce company over all orders placed. (6 Marks)

6. Calculate the BrandRevenue for each brand and list the top 10 brands having maximum revenue in descending order (5 Marks)
    a. Calculate BrandRevenue for each brand (2 Marks)
    b. List the top 10 brands having maximum revenue in descending order (3 Marks)

7. Compare prices for each product (5 Marks)
    a. Draw boxplots of retail_price & discount_price. (2 Marks)
    b. Are there any outliers? (Yes/No) (1 Mark)
    c. Create a scatterplot retail_price (x-axis) & discounted_price (y-axis) (2 Marks)

8. Create a new dataframe to include the brand specific information as stated (3 Marks)
    a. Total number of orders placed per brand
    b. Total retail_price per brand
    c. Total discount_price per brand
    d. Total BrandRevenue generated per brand.
    e. Pairplot using these four features.

9. Compare performance regionwise (6 Marks)
    a. Draw a lineplot for the monthly Revenue of ECom Company for each region separately. (4 Marks)
    b. Identify the best and the worst performing months for each region. (2 Marks)

Part II – SQL (40 Marks)

Problem Statement:

This is a Store wise Sales and inventory datasets for a retail store in the United States of America. The object of this project is to use the data and analyse the sales at various layers (such as product, store, city, state etc.) We shall also be answering a few key questions that shall help in pricing and product placement decisions.

Domain: Retail Analytics

Data Description: The data contains three tables related to orders placed in a super store.

- **OrderDetails**: The table contains 5000 different Order IDs, the order date, Property ID, Product ID and the quanities
- **Products:** The table contains 94 different products, the category they belong to and the price at which they are sold.
- **PropertyInfo**: The table contains 20 cities where the stores are based, along with the state names.

Tasks/Questions to be Answered:

Please Note: You are required to answer the following questions in SQL and paste screenshots of the query along with the output in case of SQL statements in the submission document:

1. What is the maximum quantity of any order ID in the data? Also, determine the number of orders placed which have this maximum quantity.(2 marks)
2. Find the number of unique products that are sold. (2 marks)
3. List the different types of “Chair” that are sold by using product table (Hint:TR_Products) (2marks)
4. What is the average price of each of these chair listed in the output of previous question? (2 marks)
5. Find the details of the Properties where the state names are more than 10 characters in length? (2 marks)
6. Find the details of the Properties where the second character of the city name is “e”.(2 marks)
7. Find the minimum and maximum prices for products in the “Office Supplies” category (2 marks)
8. What is the purpose of using GROUP BY in SQL? (Hint: This is a theoretical question and needs to be explained with an clear example other than the application given in this project) (2 marks)
9. List the different states in which sales are made and count how many orders are there in each of the states? (Hint: Consider order details as the primary table) (2 marks)
10. Find the average price of items sold in each Product Category and sort it in a decreasing order. (2 marks)
11. Find the Product Category that sells the least number of products? Something for the management to focus on. (2 marks)

12. What is the difference between a WHERE v/s HAVING clause in SQL? (Hint: This is a theoretical question and needs to be explained with an clear example other than the application given in this project) (2 marks)
13. Select the Product categories where the average price is more than 25 (2 marks)
14. Find the top 5 products IDs that sold the maximum quantities? (2 marks)
15. For the above question, print the product names instead of Product IDs. (2 marks)
16. Mention the different types of joins in SQL? Give simple examples of each. Also represent them using Venn diagrams (Hitnt: This is a theoretical question, the explanation needs to be in detail along with an example other than the one given in this project) (2 marks)
17. Determine the 5 products that give the overall minimum sales? (Hint: Sales = Quantity * Price) (2 marks)
18. Repeat the above query for the City of “Orlando”. (2 marks)
19. What is the difference between Drop, Truncate and Delete? Explain with examples. (2 marks)
20. Which are the cities that belong to the same states? (2 marks)