SQL Problem Statement

Mentorness Internship Program

Project Name: Uncover Insights from Real Sales Data

In this internship, you will work with a database containing total 5 tables related to customer's information, product details, PIN code details, delivery person details and order details. The schema of the database is provided in Dataset Excel File

Database Description:

- **Customers:** Contains customer information, including cust_id, first_name, last_name, email, phone, primary_pincode, gender, dob, joining_date.
- **Products:** Contains product information, including product_id, product_name, brand, category, procurement_cost_per_unit, mrp.
- **Pincode:** Contains pincode-related information, including the pincode, city and state.
- **Delivery Person:** Contains information about delivery personnel, including delivery_person_id, name, joining_date, pincode
- **Orders:** Contains order details, including order_id, order_type, cust_id, order_date, delivery_date, tot_units, displayed_selling_price_per_unit, total_amount_paid, product_id, delivery_person_id, payment_type, delivery_pincode

What you have to do?

Your tasks will involve querying this database using SQL to answer various questions and analyze the data. There are total 19 questions divided in 3 sections:

- Basic (1 mark each)
- Based on SQL Joins (2 marks each)
- Advance (3 marks each)

Basic Questions:

1. How many customers do not have DOB information available?
2. How many customers are there in each pincode and gender combination?
3. Print product name and mrp for products which have more than 50000 MRP?
4. How many delivery personal are there in each pincode?

5. For each Pin code, print the count of orders, sum of total amount paid, average amount paid, maximum amount paid, minimum amount paid for the transactions which were paid by 'cash'. Take only 'buy' order types
6. For each delivery_person_id, print the count of orders and total amount paid for product_id = 12350 or 12348 and total units > 8. Sort the output by total amount paid in descending order. Take only 'buy' order types
7. Print the Full names (first name plus last name) for customers that have email on "gmail.com"?
8. Which pincode has average amount paid more than 150,000? Take only 'buy' order types
9. Create following columns from order_dim data -
   - order_date
   - Order day
   - Order month
   - Order year
10. How many total orders were there in each month and how many of them were returned? Add a column for return rate too.
    return rate = (100.0 * total return orders) / total buy orders
    Hint: You will need to combine SUM() with CASE WHEN

Question On SQL Joins:

11. How many units have been sold by each brand? Also get total returned units for each brand.
12. How many distinct customers and delivery boys are there in each state?
13. For every customer, print how many total units were ordered, how many units were ordered from their primary_pincode and how many were ordered not from the primary_pincode. Also calulate the percentage of total units which were ordered from primary_pincode(remember to multiply the numerator by 100.0). Sort by the percentage column in descending order.
14. For each product name, print the sum of number of units, total amount paid, total displayed selling price, total mrp of these units, and finally the net discount from selling price.
    (i.e. 100.0 - 100.0 * total amount paid / total displayed selling price) & the net discount from mrp (i.e. 100.0 - 100.0 * total amount paid / total mrp)

Advance Questions:

15. For every order_id (exclude returns), get the product name and calculate the discount percentage from selling price. Sort by highest discount and print only those rows where discount percentage was above 10.10%.

16. Using the per unit procurement cost in product_dim, find which product category has made the most profit in both absolute amount and percentage Absolute Profit = Total Amt Sold - Total Procurement Cost Percentage Profit = 100.0 * Total Amt Sold / Total Procurement Cost - 100.0

17. For every delivery person(use their name), print the total number of order ids (exclude returns) by month in separate columns i.e. there should be one row for each delivery_person_id and 12 columns for every month in the year

18. For each gender - male and female - find the absolute and percentage profit (like in Q15) by product name

19. Generally the more numbers of units you buy, the more discount seller will give you. For 'Dell AX420' is there a relationship between number of units ordered and average discount from selling price? Take only 'buy' order types