**REVIEW Assignment 2**

**Analyzing Company Performance with SQL**

***Student name: SHIVANI NANDKISHOR NIPANE***

***Student id: 24622969***

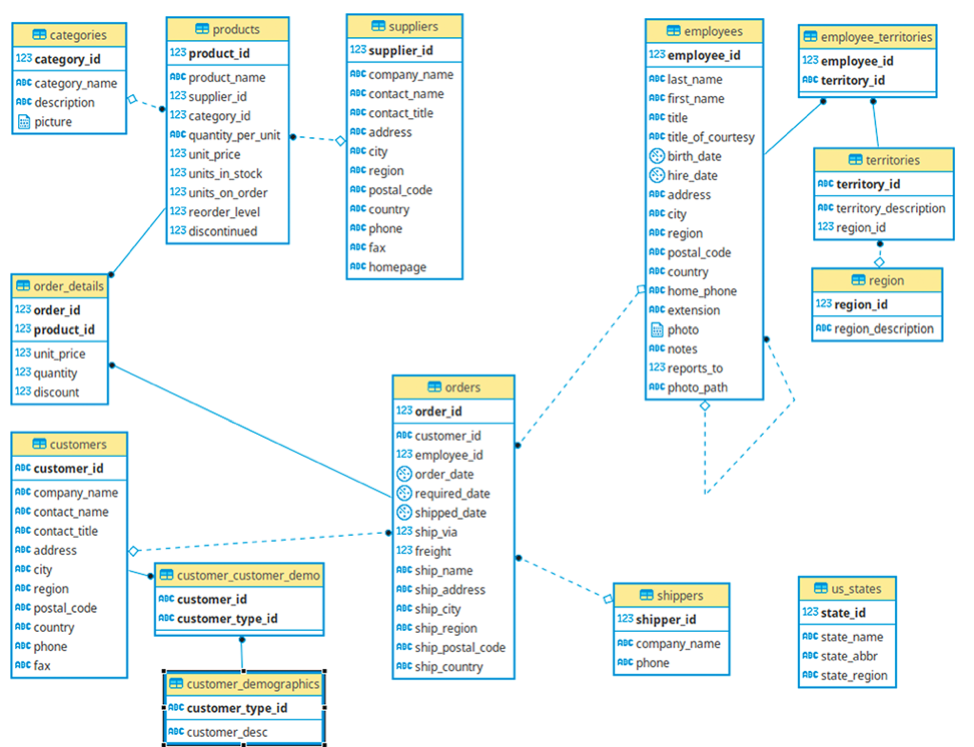
**INTRODUCTION**

The Northwind database is a sample database that was originally created by Microsoft and used as the basis for their tutorials in a variety of database products for decades. The Northwind database contains the sales data for a fictitious company called “Northwind Traders,” which imports and exports specialty foods from around the world. The Northwind database is an excellent tutorial schema for a small-business ERP, with customers, orders, inventory, purchasing, suppliers, shipping, employees, and single-entry accounting. The Northwind database has since been ported to a variety of non-Microsoft databases, including PostgreSQL.

The Northwind dataset includes sample data for the following.

* **Suppliers**: Suppliers and vendors of Northwind
* **Customers**: Customers who buy products from Northwind
* **Employees**: Employee details of Northwind traders
* **Products**: Product information
* **Shippers**: The details of the shippers who ship the products from the traders to the end-customers
* **Orders and Order\_Details**: Sales Order transactions taking place between the customers & the company

The Northwind sample database includes 14 tables and the table relationships are showcased in the following entity relationship diagram.



Instructions for connecting to the database

First, Install the PostgreSQL and pgadmin should be started. Open pgadmin server and on the left hand side there is the postgres version where there are tables, schemas present in the server.For creating a database, right-click on the databases tab and there will be an option for creating a new database. So a database can be created like this. We need to connect to the database using \c which means connection to the database.

A file is downloaded in the computer under the name “**northwind.postgres.sql**”

LOAD the sample database using \i ‘ Path/ Location’

So, I loaded the file using \i command and followed the syntax like this as

\i ‘F:/Downloads/northwind.postgres.sql’

The file is now successfully loaded on my laptop.

Now, the tasks are resolved based on the queries provided you can a see a proper analysis of the questions which needs to be resolved.

Question 1

For their annual review of the company pricing strategy, the Product Team wants to look at the products that are currently being offered for a specific price range ($10 to $50). In order to help them they asked you to provide them with a list of products with the following information:

1. their name
2. their unit price

Filtered on the following conditions:

1. their unit price is between 10 and 50 (greater or equal to 10 but less or equal than 50)
2. they are not discontinued

Finally order the results by product name in alphabetical order.

Answer 1.

The company wants to know the products which are not discontinued with their respective names and their unit price between $10 and $50. So in this case, we needed products table which contains product name and unit price and by using between operator we filtered the given range between 10 and 50 and for discontinued product we filtered by not operator then arranged the product name in ascending order.

#### Question 2

The Logistics Team wants to do a retrospection of their performances for the year 1997, in order to identify for which countries they didn’t perform well. They asked you to provide them a list of countries with the following information:

1. their average days between the order date and the shipping date (formatted to have only 2 decimals)
2. their total number of unique orders (based on the order id)

Filtered on the following conditions:

1. the year of order date is 1997
2. their average days between the order date and the shipping date is greater or equal to 3 days but less than 20 days
3. their total number of orders is greater than 5 orders

Finally order the results by the average days between the order date and the shipping date in a descending order (higher value first).

Answer 2.

The company wants to know about average days between order date and shipping date and total number of unique orders. So, in this I calculated the average days using average function considering the date by subtracting the shipped date and order date and counted the total volume function using count function and later on I used to\_char fnction to get the year of order date 1997 and group by ship country arranged the average days between order shipping in descending order. Only one table was considered in this case which was orders which contained all the necessary information to resolve the task.

Question 3

The HR Team wants to know for each employee what was their age on the date they joined the company and who they currently report to. Provide them with a list of every employees with the following information:

1. their full name (first name and last name combined in a single field)
2. their job title
3. their age at the time they were hired
4. their tenure in years until current date
5. their manager full name (first name and last name combined in a single field)
6. their manager job title

Finally order the results by employee age and employee full name in an ascending order (lowest first).

Answer 3.

The HR wants to get the employees details which works in their company. The information of the all the employees were present in employee table. So the employee table contained the employee data and the name was given as first name and last name which needs to be renamed as their full name so we concating the first and last name as their full name with their job title. We need to calculate the age when the employee was hired so we extracted year from age and calculated the age using age function between hire date and birth date. The tenure was also required to fill in this section, so for calculating the tenure we extracted the year from current date and hire date in this way we got the employees tenure. We also need to specify the managers under whom they are working and which employee reports to which manager for this report\_to column was present in the employee table where it contains employee\_id. So by using the inner join and we got the manager’s name with there job title and the table was order by employee age and employee full name in ascending order.

Question 4

The Logistics Team wants to do a retrospection of their global performances over 1996-1997, in order to identify for which month they perform well. They asked you to provide them a list with:

1. their year/month as single field in a date format (e.g. “1996-01-01” for January 1996)
2. their total number of orders
3. their total freight (formatted to have no decimals)

Filtered on the following conditions:

1. the order date is between 1996 and 1997 (greater or equal to 1996 but less or equal than 1997)
2. their total number of orders is greater than 20 orders
3. their total freight is greater than 2500

Finally order the results by total freight (descending order).

Answer 4.

To indentify in which month they performed well, all the specified variables were present in the orders table. Here we used DATE\_TRUNC function based for the month and used count function to count the order ids for the total number of orders and used SUM function to calculated the total number of freights . We first filtered the year which was between 1996-1997 using to\_char function and group the month of order date using date trunc function and filtering the condition using having

Count of order id greater than 20 and sum of freights must be greater than 2500 which is order by total freight in descending order.

#### Question 5

The Pricing Team wants to know which products had an unit price increase and the percentage increase was not between 10% and 30%. In order to help them they asked you to provide them a list of products with:

1. their product name
2. their current unit price (formatted to have only 2 decimals)
3. their initial unit price (formatted to have only 2 decimals)
4. their percentage increase (formatted to have only 4 decimals)  as:

*(Current Unit Price ÷ Initial Unit Price)  -  1*

Filtered on the following conditions:

1. their percentage increase is not between 10% and 30%  (lower than 10 or greater than 30)

Finally order the results by percentage increase (ascending order).

Answer 5.

In this case, two tables are required to address this task, products and order\_details table. The product table consists of product name and current unit price of the product whereas order\_details table contains initial unit price and we need to calculate the percentage increase by using the formula [(current unit price/initial unit price )-1]\*100 and then using not between operator to filter the conditions i.e. lower than 10 or greater than30 and arranging the percentage increase in ascending order.

Question 6

The Pricing Team wants to know how each category performs according to their price range. In order to help them they asked you to provide them a list of categories with:

1. their category name
2. their price range as:
   1. “1. Below $10”
   2. “2. $10 - $20”
   3. “3. $20 - $50”
   4. “4. Over $50”
3. their total amount (formatted to have only 2 decimals) taking into account the offered discount (i.e. subtracting the discounted amount)
4. their volume of orders (number of orders in which the category was present)

Finally order the results by category name then price range (both ascending order).

Answer 6.

In this case, three tables are required to address this task, categories,products and order\_details table. The categories table contains information for category name while order\_details contains the unit price of the product and we used case for defining the price range of the unit price. Thereafter, we calculated total amount using this formula where we used SUM function and inside it we used the formula [(unit price-unit price\*discount)\*quantity] and used the count function to count the total number of orders and group the table using category name and price range and arranged the table in the ascending order.

#### Question 7

The Logistics Team wants to know what is the current state of our regional suppliers' stocks for each category of product. In order to help them they asked you to provide them a list of categories with:

1. their supplier region” as:
   1. “America”
   2. “Europe”
   3. “Asia”
   4. “Oceania”
2. their category name
3. their total units in stock
4. their total units on order
5. their total reorder level

Finally order the results by supplier region, then category name and reorder level (each in ascending order).

Answer 7.

Here three tables are essential to solve the task, categories,supplier and products. The supplier table contains supplier region information. The unit in stocks, units on order and reorder level are present in products table and category id is present in both categories and product table. To calculate the total units in stocks, total units on stock and reorder level we used sum function and order by supplier region, category name and reorder level in ascending order.

Question 8

The Pricing Team wants to know for each currently offered product how their unit price compares against their categories average and median unit price. In order to help them they asked you to provide them a list of products with:

1. their category name
2. their product name
3. their unit price
4. their category average unit price (formatted to have only 2 decimals)
5. their category median unit price (formatted to have only 2 decimals)
6. their position against the category average unit price as:
   1. “Below Average”
   2. “Average”
   3. “Over Average”
7. their position against the category median unit price as:
   1. “Below Median”
   2. “Median”
   3. “Over Median”

Filtered on the following conditions:

1. They are not discontinued

Finally order the results by category name then product name (both ascending).

Answer 8.

There will be two subqueries. The first will provide the category name and average price, while the second will return the category name and median unit price. Then we must connect both the queries with the category and the product table with the correspondence field. Then, using select, we can get the category name, the product name, the unit price, the average unit price, the median unit price, and the median unit price position. Following that, we can sort data by category and product name.

Question 9

The Sales Team wants to build a list of KPIs to measure employees' performances. In order to help them they asked you to provide them a list of employees with:

1. their full name (first name and last name combined in a single field)
2. their job title
3. their total sales amount excluding discount (formatted to have only 2 decimals)
4. their total number of unique orders
5. their total number of orders
6. their average product amount excluding discount (formatted to have only 2 decimals). This corresponds to the average amount of product sold (without taking into account any discount applied to it).
7. their average order amount excluding discount (formatted to have only 2 decimals). This corresponds to the ratio between the total amount of product sold (without taking into account any discount applied to it) against to the total number of unique orders.
8. their total discount amount (formatted to have only 2 decimals)
9. their total sales amount including discount (formatted to have only 2 decimals)
10. Their total discount percentage (formatted to have only 2 decimals)

Finally order the results by total sales amount including discount (descending).

Answer 9.

Numerous subqueries must be used for this query, and all must ultimately be merged with the employee table. The total sale amount excluding discount is obtained using the first subquery, while the total unique orders is computed using the second subquery. Number of total orders, average product and order amounts, total discount and sale amounts, and total discount % all have their own subqueries. Columns for employee id are common to all subqueries. As a result, it will be linked to the employee table and use the select command to obtain all output. Following that, data will be sorted using the order by command based on total sale amount including discount.

Question 10

The Sales Team wants to build another list of KPIs to measure employees' performances across each category. In order to help them they asked you to provide them a list of categories and employees with:

1. their categories name
2. their full name (first name and last name combined in a single field)
3. their total sales amount including discount (formatted to have only 2 decimals)
4. their percentage of total sales amount including discount against his/her total sales amount across all categories (formatted to have only 5 decimals and maximum value up to 1)
5. their percentage of total sales amount including discount against the total sales amount across all employees (formatted to have only 5 decimals and maximum value up to 1)

Finally order the results by category name (ascending) then total sales amount (descending).

Answer 10.

The entire output is obtained using three subqueries. The first subquery, known as table1, is used to obtain the total selling amount. The total sales for each employee are obtained using the second subquery, which is utilised as table 2. The total sales amount for each Category is obtained using the third subquery, which is known as table3. The employee id column in each sub query links it to the employee table. Data is then sorted using category name in ascending order and total sale amount in decreasing order using the order by command.