

REVIEW Assignment 2 - Analysing Company Performance with SQL

Due Oct 7 by 23:59 **Points** 100

Analysing Company Performance with SQL



The Brief:

The Northwind database contains the sales data for a fictitious company called Northwind Traders, which imports and exports specialty foods from around the world.

You are tasked to perform an analysis of the performance of this company using SQL.

Description:

In this individual assignment, you will have to load the provided dataset into a Postgres database and create SQL queries to answer some business questions.

You can download the dataset here: [link](https://drive.google.com/file/d/1xfUF79LxsYte9Wevce-Fhpw50uiS_0Jx/view?usp=sharing) [\(https://drive.google.com/file/d/1xfUF79LxsYte9Wevce-Fhpw50uiS_0Jx/view?usp=sharing\)](https://drive.google.com/file/d/1xfUF79LxsYte9Wevce-Fhpw50uiS_0Jx/view?usp=sharing)

The dataset is composed of 13 tables:

Table	Rows
Categories	8
Customers	91
Employees	9

Employee_Territories 49

Order_Details 2155

Orders 830

Region 4

Products 77

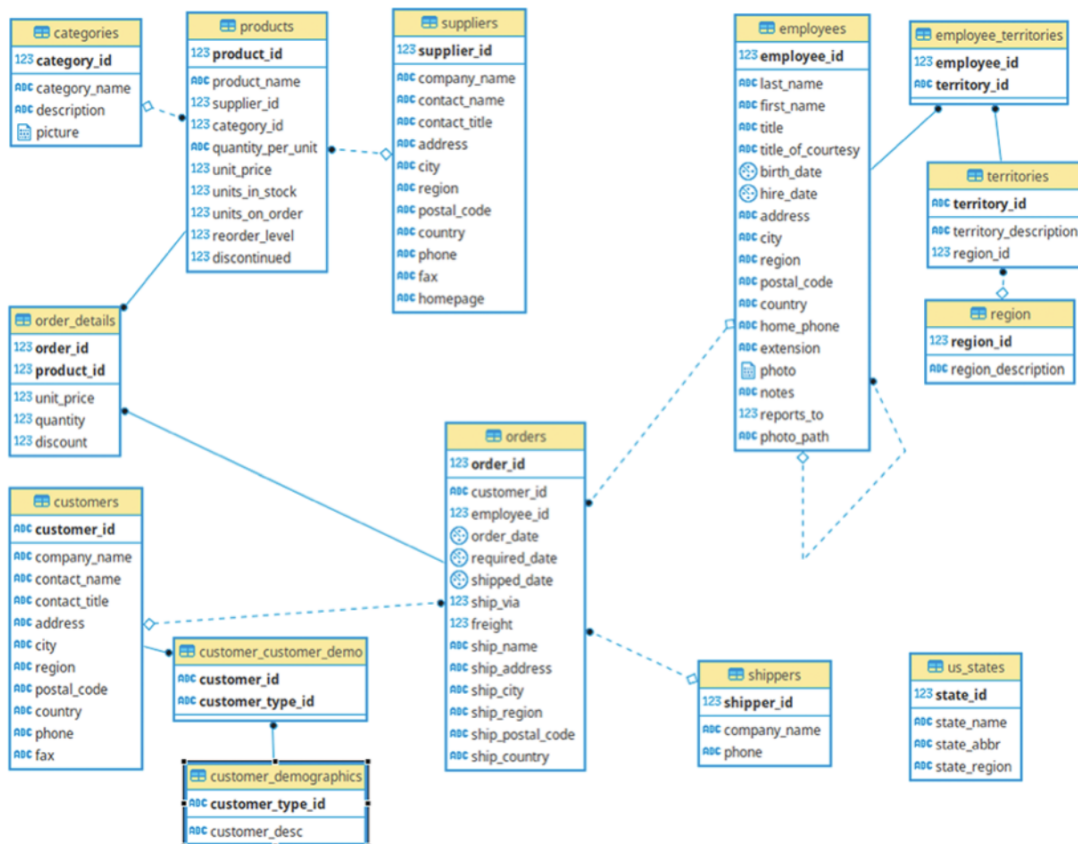
Shippers 6

Suppliers 29

Territories 53

Us_states 51

Here is the entity relationship diagram (ERD) for this dataset



Business Questions:

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:

- A. their name
- B. 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.

Output example:

product_name	product_unit_price
Product X	45.33
Product Y	32.99
Product Z	29.34

Question 2

The Logistics Team wants to do a retrospection of their performances for the year 1998, 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:

- A. their average days between the order date and the shipping date (formatted to have only 2 decimals)
- B. 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).

Output example:

shipping_country	average_days_between_order_shipping	total_volume_orders
Country Z	12	200
Country X	10.12	100
Country Y	9.43	400

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:

- A. their full name (first name and last name combined in a single field)
- B. their job title
- C. their age at the time they were hired
- D. their tenure in years until current date
- E. their manager full name (first name and last name combined in a single field)
- F. their manager job title

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

Output example:

employee_full_name	employee_title	employee_age	employee_tenure	manager_full_name	manager_title
Leslie Knope	Sales Representative	24	1	Ron Swanson	Vice President, Sales
John Doe	Sales Manager	34	7		

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:

- A. their year/month as single field in a date format (e.g. "1996-01-01" for January 1996)
- B. their total number of orders
- C. 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).

Output example:

year_month	total_number_orders	total_freight
1996-04-01	300	18903
1997-01-01	422	9056
1996-03-01	233	2266

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:

- A. their product name
- B. their current unit price (formatted to have only 2 decimals)
- C. their initial unit price (formatted to have only 2 decimals)
- D. their percentage increase (formatted to have only 4 decimals) as:

$(Current\ Unit\ Price \div Initial\ Unit\ Price) - 1$

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).

product_name	current_price	previous_unit_price	percentage_increase
Product X	1.1	1.1	10
Product Y	4.5	2.25	100
Product Z	7.5	2.5	200

Output example:

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:

- A. their category name
- B. their price range as:
 1. "Below \$10"
 2. "\$10 - \$20"
 3. "\$20 - \$50"
 4. "Over \$50"

C. their total amount (formatted to have only 2 decimals) taking into account the offered discount (i.e. subtracting the discounted amount)

D. 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).

Output example:

category_name	price_range	total_amount	total_number_orders
Category X	Below \$10	2997	575
Category X	\$20 - \$50	47474	7865
Category X	Over \$50	46422	4228
Category Y	1. Below \$10	26961	36
Category Y	\$10 - \$20	65110	22

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:

A. their supplier region" as:

1. "America"
2. "Europe"
3. "Asia"
4. "Oceania"

B. their category name

C. their total units in stock

D. their total units on order

E. their total reorder level

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

Output example:

category_name	supplier_region	units_in_stock	units_on_order	reorder_level
Category X	America	393	180	10
Category X	Asia	64	140	15
Category X	Europe	15	0	0
Category Y	America	59	90	40
Category Y	Oceania	26	0	155
Category Y	Europe	396	70	110

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:

- A. their category name
- B. their product name
- C. their unit price
- D. their category average unit price (formatted to have only 2 decimals)
- E. their category median unit price (formatted to have only 2 decimals)
- F. their position against the category average unit price as:
 - 1. "Below Average"
 - 2. "Average"
 - 3. "Over Average"
- G. 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).

Output example:

category_name	product_name	unit_price	average_unit_price	median_unit_price	average_unit_price_position	median_unit_price_position
Category X	Product X	18	22.99	21.05	Below Average	Below Median
Category X	Product Y	263.5	22.99	21.05	Over Average	Over Median
Category Y	Product Z	46	28.73	32	Over Average	Over Median

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:

- A. their full name (first name and last name combined in a single field)
- B. their job title
- C. their total sales amount excluding discount (formatted to have only 2 decimals)
- D. their total number of unique orders
- E. their total number of orders
- F. their average product amount excluding discount (formatted to have only 2 decimals)
- G. their average order amount excluding discount (formatted to have only 2 decimals)
- H. their total discount amount (formatted to have only 2 decimals)
 - I. their total sales amount including discount (formatted to have only 2 decimals)
 - J. Their total discount percentage (formatted to have only 2 decimals)

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

Output example:

employee_full_name	employee_title	total_sale_amount_excluding_discount	number_unique_orders	number_orders	average_product_amount	average_order_amount	total_discount_amount	total_sale_amount_including_discount	total_discount_percentage
John Doe	Sales Representative	250187.45	43	107	802.82	1962.44	11211.51	166537.76	6.73
Leslie Knope	Inside Sales Coordinator	213051.3	67	168	775.36	1929.4	6438.75	126862.28	5.08
Ben Dover	Sales Representative	202143.71	42	176	465.46	1929.4	16727.76	124568.23	13.43

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:

- their categories name
- their full name (first name and last name combined in a single field)
- their total sales amount including discount (formatted to have only 2 decimals)
- 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)
- 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).

Output example:

category_name	employee_full_name	total_sale_amount	percent_of_employee_sales	percent_of_category_sales
Category X	John Doe	19642.55	22.45	10.44
Category X	Leslie Knope	17897.85	25.41	7.33
Category Y	John Doe	11000.53	14.11	6.68
Category Y	Leslie Knope	9450.2	15.99	4.11

Submission:

You will submit the following **2 files** to Canvas:

- your final report (in Word document or PDF),
- a zip file containing your code:
 - SQL scripts (1 text file with .sql extension per business question) saved in a folder called 'queries'
 - their resulting tables as CSV files saved in a folder called 'csv'

Note: Each SQL script SHOULD NOT contain any comments, only the SQL statements.

Use this template for the structure of your zip file: [template](https://drive.google.com/file/d/1gX6kgri8V_wfVZN8D2dEjFmx2glaQggY/view?usp=sharing)

(https://drive.google.com/file/d/1gX6kgri8V_wfVZN8D2dEjFmx2glaQggY/view?usp=sharing)

The report will need to contain at least the following topics:

- Presentation of data
- Instructions to create and connect to database
- Answers to business questions with explanations, tables and/or graphs (as if you would have communicated your results back to the stakeholders)

All assignments need to be submitted before the due date on Canvas. Penalties will be applied for late submission.

Assessment Criteria:

- Efficiency and conciseness of SQL queries
- Readability and consistency of coding style
- Insightfulness and clarity of written explanations for business questions
- Clarity and quality of visualizations and written report

Data Dictionary:

Categories Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Category_ID	int			Categories of Northwind products.
		IX	Category_Name	varchar (15)			Name of food category.
			Description	text	Yes		Full description of the category. Provide samples.

Customers Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
-------------	-------------	------------	-----------	--------	----------	---------	-------------

PK	IX	Customer_ID	varchar (5)		Unique five-character code based on customer name.
	IX	Company_Name	varchar (40)		
		Contact_Name	varchar (30)	Yes	
		Contact_Title	varchar (30)	Yes	
		Address	varchar (60)	Yes	Street or post-office box.
	IX	City	varchar (15)	Yes	
	IX	Region	varchar (15)	Yes	State or province.
	IX	Postal_Code	varchar (10)	Yes	
		Country	varchar (15)	Yes	
		Phone	varchar (24)	Yes	Phone number includes country code or area code.
		Fax	varchar (24)	Yes	Phone number includes country code or area code.

Employees Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Employee_ID	int			Number automatically assigned to new employee.
		IX	Last_Name	varchar (20)			

	First_Name	varchar	(10)		
	Title	varchar	(30)	Yes	Employee's title.
	Title_Of_Courtesy	varchar	(25)	Yes	
	BirthDate	datetime		Yes	
	HireDate	datetime		Yes	
	Address	varchar	(60)	Yes	Street or post-office box.
	City	varchar	(15)	Yes	
	Region	varchar	(15)	Yes	Street or post-office box.
IX	Postal_Code	varchar	(10)	Yes	
	Country	varchar	(15)	Yes	
	Home_Phone	varchar	(24)	Yes	Phone number includes country code or area code.
	Extension	varchar	(4)	Yes	Internal telephone extension number.
	Notes	text		Yes	General information about employee's background.
FK	ReportsTo	int		Yes	Employee's supervisor.

Employee_Territories Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK	FK	IX	Employee_ID	int			

PK FK IX Territory_ID varchar (20)

Order_Details Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK	FK	IX	Order_ID	int			Same as Order_ID in Orders table.
PK	FK	IX	Product_ID	int			Same as Product_ID in Products table.
			Unit_Price	decimal		(0)	
			Quantity	smallint		(1)	
			Discount	decimal		(0)	

Orders Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Order_ID	int			
	FK	IX	Customer_ID	varchar	(5)	Yes	
	FK	IX	Employee_ID	int	Yes		Same entry as in Employees table.
		IX	Order_Date	datetime	Yes		
			Required_Date	datetime	Yes		
		IX	Shipped_Date	datetime	Yes		
	FK	IX	Ship_Via	int	Yes		Same as Shipper ID in Shippers table.
			Freight	decimal	Yes	(0)	

	Ship_Name	varchar (40)	Yes	Name of person or company to receive the shipment.
	Ship_Address	varchar (60)	Yes	Street address only -- no post-office box allowed.
	Ship_City	varchar (15)	Yes	
	Ship_Region	varchar (15)	Yes	State or province.
IX	Ship_Postal_Code	varchar (10)	Yes	
	Ship_Country	varchar (15)	Yes	

Products Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Product_ID	int			Number automatically assigned to new product.
		IX	Product_Name	nvarchar (40)			
	FK	IX	Supplier_ID	int	Yes		Same entry as in Suppliers table.
	FK	IX	Category_ID	int	Yes		Same entry as in Categories table.
			Quantity_Per_Unit	nvarchar (20)	Yes		(e.g., 24-count case, 1-liter bottle).
			UnitPrice	money	Yes	(0)	
			Units_In_Stock	smallint	Yes	(0)	

Units_On_Order	smallint	Yes	(0)	
Reorder_Level	smallint	Yes	(0)	Minimum units to maintain in stock.
Discontinued	bit		(0)	Yes means item is no longer available.

Region Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Region_ID	int			
			Region_Description	varchar	(50)		

Shippers Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Shipper_ID	int			Number automatically assigned to new shipper.
			Company_Name	varchar	(40)		Name of shipping company.
			Phone	varchar	(24)	Yes	Phone number includes country code or area code.

Suppliers Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX	Supplier_ID	int			Number automatically assigned to new supplier.

IX	Company_Name varchar (40)			
	Contact_Name	varchar (30)	Yes	
	Contact_Title	varchar (30)	Yes	
	Address	varchar (60)	Yes	Street or post-office box.
	City	varchar (15)	Yes	
IX	Region	varchar (15)	Yes	State or province.
	Postal_Code	varchar (10)	Yes	
	Country	varchar (15)	Yes	
	Phone	varchar (24)	Yes	Phone number includes country code or area code.
	Fax	varchar (24)	Yes	Phone number includes country code or area code.
	Home_Page	ntext	Yes	Supplier's home page on World Wide Web.

Territories Table

Primary Key	Foreign Key	Index Name	Data type	Length	Nullable	Default	Description
PK		IX Territory_ID	varchar	(20)			
			Territory_Description	varchar	(50)		
	FK	Region_ID	int				

