Assignment 1

Analysing Company Performance with SQL

Balakumaran Sivanesan
Student ID: 24883306

07 - 09 - 2023

94692 - Data Science Practice
Master of Data Science and Innovation
University of Technology of Sydney

Table of Contents

1. Executive Summary **2**
2. Introduction **3**
3. Data Understanding **4**
4. Business Question 1 **5**
a. Description 5
b. Results 5
c. Key insights and Findings 5
5. Business Question 2 **6**
a. Description 6
b. Results 6
c. Key insights and Findings 6
6. Business Question 3 **7**
a. Description 7
b. Results 7
c. Key insights and Findings 7
7. Business Question 4 **8**
a. Description 8
b. Results 8
c. Key insights and Findings 8
8. Business Question 5 **9**
a. Description 9
b. Results 9
c. Key insights and Findings 9
9. Business Question 6 **10**
a. Description 10
b. Results 10
c. Key insights and Findings 10
10. Business Question 7 **11**
a. Description 11
b. Results 11
c. Key insights and Findings 11
11. Business Question 8 **12**
a. Description 12
b. Results 12
c. Key insights and Findings 12
12. Business Question 9 **13**
a. Description 13

b. Results 13
c. Key insights and Findings 13
13. Business Question 10 **14**
a. Description 14
b. Results 14
c. Key insights and Findings 14
14. Conclusion **15**
15. References **16**

1. Executive Summary

The Northwind database is a sample database used for this assessment within the given database, my goal in this assignment is to use PostgreSQL to analyze the performance and answer important business questions for a fictitious company named Northwind Trader and its main objectives and significance of the project are:

- Data Analysis
- Inventory Management
- Customer Relationship Management
- Order Processing Fulfillment
- Employee Performance Analysis

The goal of the Northwind database project is to use the available data within the Northwind database to address numerous operational and strategic difficulties encountered by Northwind Traders, a made-up trading firm. These difficulties include:

- Inventory Management
- Customer Relationship Management
- Sales Analysis
- Employee Performance

2. Introduction

The Northwind database is a sample database used for this assessment within the given database, my goal in this assignment is to use PostgreSQL to analyze the performance and answer important business questions for a fictitious company named Northwind Trader.

3. Dataset presentation

In this assessment, there are numerous tables and rows are given in the database, with the help of Entity Relationship Diagram – ERD I, can able to visualize the given tables in the database.

As shown in the ERD there is a total of 13 tables and all are interconnected. The data comprises more than 1000+ transactions, including the information about the company's employees, products, orders, payments and customers. Throughout the whole given data, different department tends to know about their performance to have a positive impact on business growth and in the given assessment there are 10 business questions asked from different department. In this assessment, I used Postgres and Dbeaver as a tool to answer the given business questions.

4. Business Question 1

a. Description

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 ($20 to $50). In order to help them they asked you to provide them with a list of products with the following information:

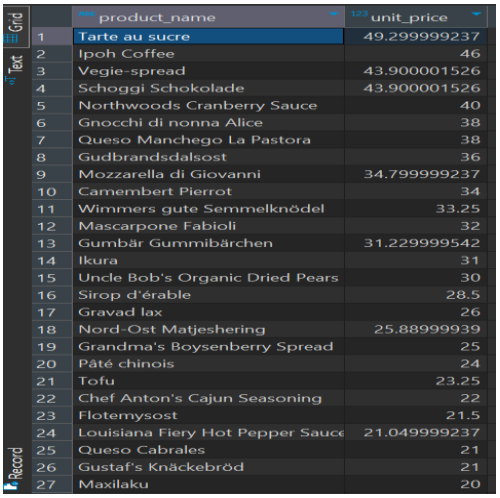
1. their name
2. their unit prices

Filtered on the following conditions:

1. their unit price is between 20 and 50 (greater or equal to 20 but less or equal than 50)
2. they are not discontinued

Finally order the results by unit price in a descending order (highest first).

b. Results



	product_name	unit_price
1	Tarte au sucre	49.299999237
2	Ipoh Coffee	46
3	Vegie-spread	43.900001526
4	Schoggi Schokolade	43.900001526
5	Northwoods Cranberry Sauce	40
6	Gnocchi di nonna Alice	38
7	Queso Manchego La Pastora	38
8	Gudbrandsdalsost	36
9	Mozzarella di Giovanni	34.799999237
10	Camembert Pierrot	34
11	Wimmers gute Semmelknödel	33.25
12	Mascarpone Fabioli	32
13	Gumbär Gummibärchen	31.229999542
14	Ikura	31
15	Uncle Bob's Organic Dried Pears	30
16	Sirop d'érable	28.5
17	Gravad lax	26
18	Nord-Ost Matjeshering	25.88999939
19	Grandma's Boysenberry Spread	25
20	Pâté chinois	24
21	Tofu	23.25
22	Chef Anton's Cajun Seasoning	22
23	Flotemysost	21.5
24	Louisiana Fiery Hot Pepper Sauce	21.049999237
25	Queso Cabrales	21
26	Gustaf's Knäckebröd	21
27	Maxilaku	20

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the product name and unit price and ordered it in a descending order.

5. Business Question 2

a. Description

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:

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 1998
2. their average days between the order date and the shipping date is greater or equal 5 days
3. their total number of orders is greater than 10 orders

Finally order the results by country name in an ascending order (following alphabetical order).

b. Results

country	unique_orders
Austria	11
Brazil	28
France	23
Germany	34
Spain	12
Sweden	14
UK	16
USA	39
Venezuela	18

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the country and unique orders and ordered it in a ascending order.

6. Business Question 3

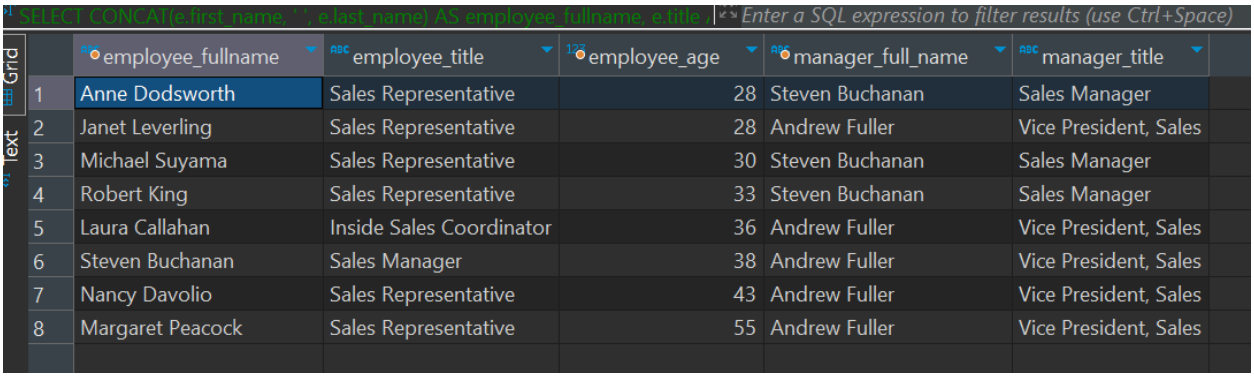
a. Description

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 manager full name (first name and last name combined in a single field)
5. their manager job title

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

b. Results



	employee_fullname	employee_title	employee_age	manager_full_name	manager_title
1	Anne Dodsworth	Sales Representative	28	Steven Buchanan	Sales Manager
2	Janet Leverling	Sales Representative	28	Andrew Fuller	Vice President, Sales
3	Michael Suyama	Sales Representative	30	Steven Buchanan	Sales Manager
4	Robert King	Sales Representative	33	Steven Buchanan	Sales Manager
5	Laura Callahan	Inside Sales Coordinator	36	Andrew Fuller	Vice President, Sales
6	Steven Buchanan	Sales Manager	38	Andrew Fuller	Vice President, Sales
7	Nancy Davolio	Sales Representative	43	Andrew Fuller	Vice President, Sales
8	Margaret Peacock	Sales Representative	55	Andrew Fuller	Vice President, Sales

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the employee's full name, employee title, employee age, manager full name and manager title and ordered it in an ascending order.

7. Business Question 4

a. Description

The Logistics Team wants to do a retrospection of their global performances over 1997-1998, in order to identify for which month, they perform well. They asked you to provide them a list with:

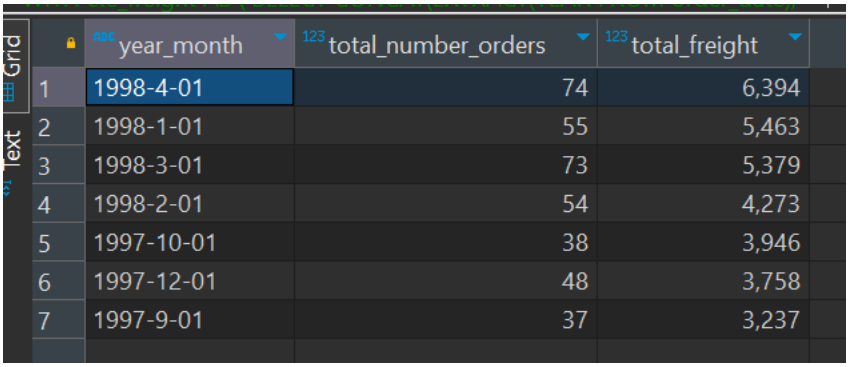
1. their year/month as single field in a date format (e.g., “1990-01-01” for January 1990)
2. their total number of orders
3. their total freight (formatted to have no decimals)

Filtered on the following conditions:

1. the order date is between 1997 and 1998
2. their total number of orders is greater than 35 orders

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

b. Results



Grid | Text

	year_month	total_number_orders	total_freight
1	1998-4-01	74	6,394
2	1998-1-01	55	5,463
3	1998-3-01	73	5,379
4	1998-2-01	54	4,273
5	1997-10-01	38	3,946
6	1997-12-01	48	3,758
7	1997-9-01	37	3,237

c. Key insights and Findings

From the given results I can state that I’ve met the required conditions of the table the business team needed, with the help of the given data I founded the year month, total number orders, total freight and ordered it in an descending order.

8. Business Question 5

a. Description

The Pricing Team wants to know which products had an unit price increase and the percentage increase was not between 20% 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 with the result formatted to an integer (e.g 50 for 50%) using the following calculation:

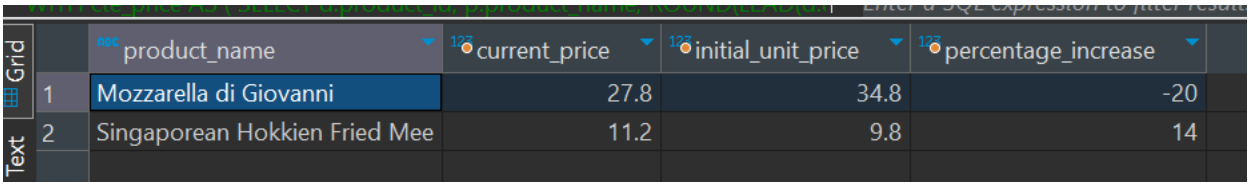
(Current Unit Price - Initial Unit Price) ÷ Initial Unit Price * 100

Filtered on the following conditions:

1. their percentage increase is not between 20% and 30% (lower than 20 or greater than 30)

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

b. Results



Grid | Text

	product_name	current_price	initial_unit_price	percentage_increase
1	Mozzarella di Giovanni	27.8	34.8	-20
2	Singaporean Hokkien Fried Mee	11.2	9.8	14

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the product name, current price, initial unit price, percentage increase and ordered it in an descending order.

9. Business Question 6

a. Description

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. Below $20"

"2. $20 - $50"

"3. Over $50"

3. their total amount (formatted to be integer) 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).

b. Results

	category_name	price_range	total_amount	total_number_orders
1	Beverages	1. Below $20	111,464	317
2	Beverages	2. $20 - $50	25,079	28
3	Beverages	3. Over $50	149,984	24
4	Condiments	1. Below $20	28,622	85
5	Condiments	2. $20 - $50	85,073	121
6	Confections	1. Below $20	57,369	197
7	Confections	2. $20 - $50	96,094	106
8	Confections	3. Over $50	23,636	16
9	Dairy Products	1. Below $20	17,886	81
10	Dairy Products	2. $20 - $50	157,148	204
11	Dairy Products	3. Over $50	76,296	54
12	Grains/Cereals	1. Below $20	25,364	99
13	Grains/Cereals	2. $20 - $50	75,363	91
14	Meat/Poultry	1. Below $20	5,121	36
15	Meat/Poultry	2. $20 - $50	76,504	96
16	Meat/Poultry	3. Over $50	96,563	36
17	Produce	1. Below $20	2,566	13
18	Produce	2. $20 - $50	57,960	81
19	Produce	3. Over $50	44,743	39
20	Seafood	1. Below $20	69,673	217
21	Seafood	2. $20 - $50	39,963	70
22	Seafood	3. Over $50	31,988	27

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the category name, price range, total amount, total number orders and ordered it in a ascending order.

10. Business Question 7

a. Description

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:
   - "America"
   - "Europe"
   - "Asia-Pacific"
2. their category names
3. their total units in stock
4. their total units on order
5. their total reorder levels

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

b. Results

	category_name	suppliers_region	units_in_stock	units_on_order	reorder_level
1	Condiments	America	113	0	25
2	Confections	America	17	0	0
3	Meat/Poultry	America	21	0	10
4	Meat/Poultry	America	115	0	20
5	Beverages	Asia-Pacific	15	10	30
6	Condiments	Asia-Pacific	24	0	5
7	Confections	Asia-Pacific	29	0	10
8	Grains/Cereals	Asia-Pacific	38	0	25
9	Meat/Poultry	Asia-Pacific	0	0	0
10	Meat/Poultry	Asia-Pacific	0	0	0
11	Produce	Asia-Pacific	20	0	10
12	Seafood	Asia-Pacific	42	0	0
13	Beverages	Europe	52	0	10
14	Beverages	Europe	111	0	15
15	Beverages	Europe	20	0	15
16	Condiments	Europe	4	100	20
17	Condiments	Europe	76	0	0
18	Condiments	Europe	0	0	0
19	Condiments	Europe	53	0	0
20	Condiments	Europe	120	0	25
21	Condiments	Europe	6	0	0
22	Dairy Products	Europe	22	30	30
23	Dairy Products	Europe	86	0	0
24	Produce	Europe	15	0	10
25	Seafood	Europe	85	0	10
26	Seafood	Europe	123	0	30

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the Category name, supplier's region, units in stock, units on order, reorder level and ordered it in a ascending order.

11. Business Question 8

a. Description

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:

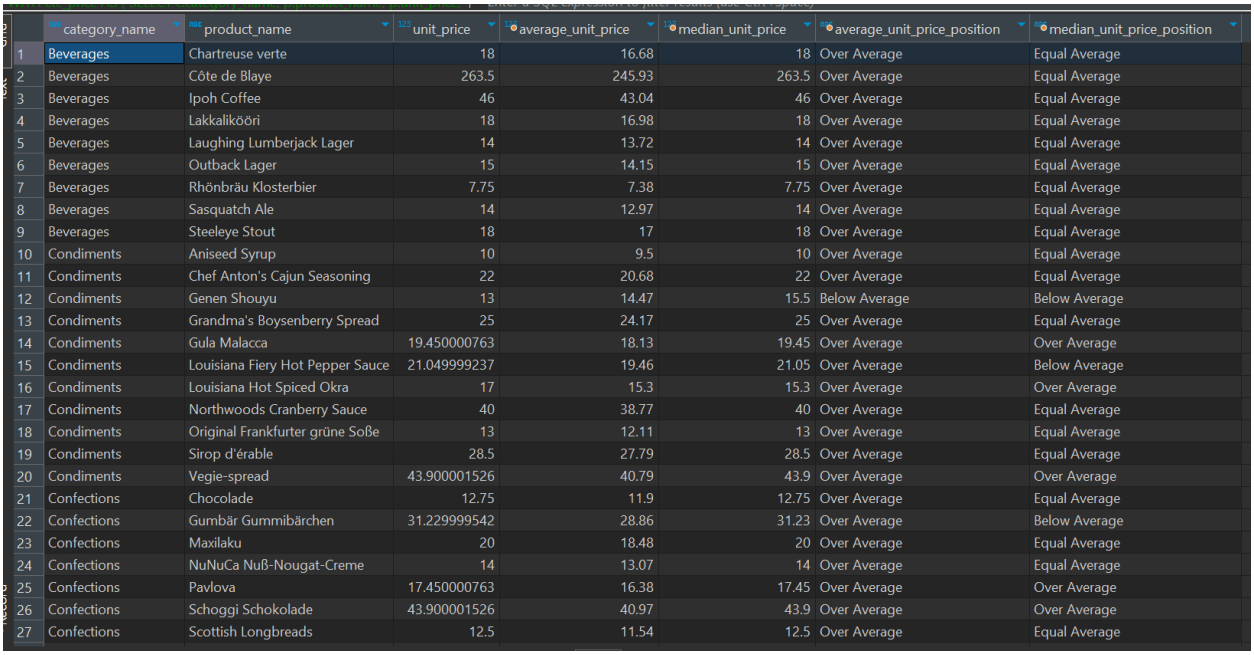
1. their category names
2. their product names
3. their unit prices
4. their category average unit price (formatted to have only 2 decimals)
5. their category median unit price (formatted to have only 2 decimals)
6. their position against the category average unit price as:
   - “Below Average”
   -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

Filtered on the following conditions:

They are not discontinued

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

b. Results



	category_name	product_name	unit_price	average_unit_price	median_unit_price	average_unit_price_position	median_unit_price_position
1	Beverages	Chartreuse verte	18	16.68	18	Over Average	Equal Average
2	Beverages	Côte de Blaye	263.5	245.93	263.5	Over Average	Equal Average
3	Beverages	Ipoh Coffee	46	43.04	46	Over Average	Equal Average
4	Beverages	Lakkalikööri	18	16.98	18	Over Average	Equal Average
5	Beverages	Laughing Lumberjack Lager	14	13.72	14	Over Average	Equal Average
6	Beverages	Outback Lager	15	14.15	15	Over Average	Equal Average
7	Beverages	Rhönbräu Klosterbier	7.75	7.38	7.75	Over Average	Equal Average
8	Beverages	Sasquatch Ale	14	12.97	14	Over Average	Equal Average
9	Beverages	Steeleye Stout	18	17	18	Over Average	Equal Average
10	Condiments	Aniseed Syrup	10	9.5	10	Over Average	Equal Average
11	Condiments	Chef Anton's Cajun Seasoning	22	20.68	22	Over Average	Equal Average
12	Condiments	Genen Shouyu	13	14.47	15.5	Below Average	Below Average
13	Condiments	Grandma's Boysenberry Spread	25	24.17	25	Over Average	Equal Average
14	Condiments	Gula Malacca	19.450000763	18.13	19.45	Over Average	Over Average
15	Condiments	Louisiana Fiery Hot Pepper Sauce	21.049999237	19.46	21.05	Over Average	Below Average
16	Condiments	Louisiana Hot Spiced Okra	17	15.3	15.3	Over Average	Over Average
17	Condiments	Northwoods Cranberry Sauce	40	38.77	40	Over Average	Equal Average
18	Condiments	Original Frankfurter grüne Soße	13	12.11	13	Over Average	Equal Average
19	Condiments	Sirop d'érable	28.5	27.79	28.5	Over Average	Equal Average
20	Condiments	Vegie-spread	43.900001526	40.79	43.9	Over Average	Over Average
21	Confections	Chocolade	12.75	11.9	12.75	Over Average	Equal Average
22	Confections	Gumbär Gummibärchen	31.229999542	28.86	31.23	Over Average	Below Average
23	Confections	Maxilaku	20	18.48	20	Over Average	Equal Average
24	Confections	NuNuCa Nuß-Nougat-Creme	14	13.07	14	Over Average	Equal Average
25	Confections	Pavlova	17.450000763	16.38	17.45	Over Average	Over Average
26	Confections	Schoggi Schokolade	43.900001526	40.97	43.9	Over Average	Over Average
27	Confections	Scottish Longbreads	12.5	11.54	12.5	Over Average	Equal Average

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the category name, product name, unit price, average unit price, median unit price, average unit price position, median unit price position and ordered it in an ascending order.

12. Business Question 9

a. Description

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).

b. Results

	employee_full_name	employee_title	total_sale_amount_excluding_discount	total_number_orders	total_number_entries	average_amount_per_entry	average_amount_per_order	total_discount_amount	total_sale_amount_including_discount
1	Margaret Peacock	Sales Representative	250,187.45	156	420	595.68	1,603.77	17,296.6	232,890.85
2	Janet Leverling	Sales Representative	213,051.3	127	321	663.71	1,677.57	10,238.46	202,812.84
3	Nancy Davolio	Sales Representative	202,143.71	123	345	585.92	1,643.44	10,036.11	192,107.6
4	Andrew Fuller	Vice President, Sales	177,749.26	96	241	737.55	1,851.55	11,211.51	166,537.76
5	Laura Callahan	Inside Sales Coordinator	133,301.03	104	260	512.7	1,281.74	6,438.75	126,862.28
6	Robert King	Sales Representative	141,295.99	72	176	802.82	1,962.44	16,727.76	124,568.23
7	Anne Dodsworth	Sales Representative	82,964	43	107	775.36	1,929.4	5,655.93	77,308.07
8	Michael Suyama	Sales Representative	78,198.1	67	168	465.46	1,167.14	4,284.97	73,913.13
9	Steven Buchanan	Sales Manager	75,567.75	42	117	645.88	1,799.23	6,775.47	68,792.28

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the employee's full name, employee title, total sale amount excluding discount, total number of orders, total number entries, average amount

per entry, average amount per order, total discount amount, total sale amount including discount, total discount percentage and ordered it in a descending order.

13. Business Question 10

a. Description

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).

b. Results

	category_name	employee_full_name	total_sale_amount_including_discount	percentage_of_employee_sales	percentage_of_category_sales
1	Beverages	Margaret Peacock	52,842.35	21.12	18.44
2	Beverages	Nancy Davolio	48,832	24.16	17.04
3	Beverages	Janet Leverling	46,506.55	21.83	16.23
4	Beverages	Andrew Fuller	42,029.4	23.65	14.67
5	Beverages	Robert King	33,517	23.72	11.7
6	Beverages	Anne Dodsworth	20,513.6	24.73	7.16
7	Beverages	Laura Callahan	18,640.8	13.98	6.51
8	Beverages	Steven Buchanan	13,517.5	17.89	4.72
9	Beverages	Michael Suyama	10,127.75	12.95	3.53
10	Condiments	Margaret Peacock	25,242.45	10.09	22.2
11	Condiments	Andrew Fuller	16,353.1	9.2	14.38
12	Condiments	Laura Callahan	15,447.95	11.59	13.59
13	Condiments	Janet Leverling	14,109.15	6.62	12.41
14	Condiments	Nancy Davolio	14,033.85	6.94	12.34
15	Condiments	Anne Dodsworth	11,106.9	13.39	9.77
16	Condiments	Robert King	9,668.7	6.84	8.5
17	Condiments	Michael Suyama	4,930.2	6.3	4.34
18	Condiments	Steven Buchanan	2,802.45	3.71	2.46
19	Confections	Janet Leverling	34,859.31	16.36	19.68
20	Confections	Nancy Davolio	30,343.54	15.01	17.13
21	Confections	Margaret Peacock	29,543.06	11.81	16.68
22	Confections	Laura Callahan	22,639.5	16.98	12.78
23	Confections	Andrew Fuller	21,696.95	12.21	12.25
24	Confections	Robert King	16,754.45	11.86	9.46
25	Confections	Anne Dodsworth	8,782.94	10.59	4.96
26	Confections	Michael Suyama	6,990.3	8.94	3.95
27	Confections	Steven Buchanan	5,489.05	7.26	3.1

c. Key insights and Findings

From the given results I can state that I've met the required conditions of the table the business team needed, with the help of the given data I founded the Category name, employee full name, total sale amount including discount, percentage of employee sales, percentage of category sales and ordered it in an descending order.

14. Conclusion

- Summarize the key findings, insights, and outcomes of the project.

  The assessment aimed to optimize business operations and decision making using the Northwind database and my goal is to analyze the given data and need to solve every business question for the positive impact on business growth

  - Data analysis
  - Inventory management
  - Customer segmentation
  - Profitability Insights
  - Forecasting
  - Stakeholder Engagement

  And the given assessment successfully achieved its goals and solved every business questions

  In conclusion, the project successfully leveraged the Northwind database to optimize business operations and achieve stakeholder goals. Future work should focus on advancing analytics capabilities, real-time data integration, and ongoing improvements in inventory management and customer experience to maintain and build upon the project's successes.

15. References

- Include a list of references used throughout the project report.

Instructions: Include a list of references used throughout the project report, following the appropriate citation style.