Assignment 1  
Analyzing Company Performance with SQL

short line

Pham Ngoc Quang Vinh

Student ID: 25100660  
August 22, 2023

94692 - Data Science Practice

Master of Data Science and Innovation

University of Technology of Sydney

Table of Contents

[**1. Executive Summary 2**](#_xvtyf97iwmd0)

[**2. Introduction 3**](#_if9iwijz2pg3)

[**3. Data Understanding 4**](#_yg03y3w5hig9)

[**4. Business Question 1 5**](#_tj77rrmixzhi)

[a. Description 5](#_qgmf1pwe8g7)

[b. Results 5](#_hqaliy1gjse9)

[c. Key insights and Findings 5](#_6j15sr92em8)

[**5. Business Question 2 6**](#_gu2c17h82qkp)

[a. Description 6](#_1gaz8kpeqrur)

[b. Results 6](#_x8v0s95f65od)

[c. Key insights and Findings 6](#_o8clu52rymsf)

[**6. Business Question 3 7**](#_3ms610bpkg46)

[a. Description 7](#_ny15vtb5bt7n)

[b. Results 7](#_fud8x1tdupks)

[c. Key insights and Findings 7](#_oxtgc9sbqm5j)

[**7. Business Question 4 8**](#_jdj4xuezrby3)

[a. Description 8](#_kzwwwr3bu5ie)

[b. Results 8](#_r4klgonxy6y)

[c. Key insights and Findings 8](#_jz1tpf9ipdfr)

[**8. Business Question 5 9**](#_tovr0q566rj)

[a. Description 9](#_6v3xib6i85f1)

[b. Results 9](#_s01265i9g5gs)

[c. Key insights and Findings 9](#_25ov5a7hym46)

[**9. Business Question 6 10**](#_e72dw0340hf8)

[a. Description 10](#_whicttvw8dc3)

[b. Results 10](#_b0dpd7nejesa)

[c. Key insights and Findings 10](#_tlowrdqf3uc9)

[**10. Business Question 7 11**](#_k04dc6gvlt52)

[a. Description 11](#_ueiuhlgps20j)

[b. Results 11](#_gw9bu9wapyo)

[c. Key insights and Findings 11](#_wtet14raba2v)

[**11. Business Question 8 12**](#_yfqftwku46ae)

[a. Description 12](#_bsapchik5fg9)

[b. Results 12](#_f37b6ersnfhb)

[c. Key insights and Findings 12](#_tw4yoooy6kr8)

[**12. Business Question 9 13**](#_o9wq12glklsa)

[a. Description 13](#_98dzsgqv007w)

[b. Results 13](#_jbnh0jn595u)

[c. Key insights and Findings 13](#_j38ijxi528kh)

[**13. Business Question 10 14**](#_qeorzpkx1ldx)

[a. Description 14](#_uq41su3mggzf)

[b. Results 14](#_2uq9fclr3diy)

[c. Key insights and Findings 14](#_i1wxr5prlwy5)

[**14. Conclusion 15**](#_9hr11g79asdx)

[**15. References 16**](#_ma3y8ytjvp1y)

# 

# Introduction

The initiative centers around the Northwind database, which contains sales data for the fictional corporation Northwind Traders, which engages in the global import and export of specialty foods. This project's primary objective was to apply SQL query and data science mindset to analyze and gain insights from the Northwind database's data.

The project's primary objectives included:

* Gain insight into product performance, consumer behavior, and market trends.
* Customer Segmentation
* Optimize the supply chain
* Identifying areas for enhancement in various aspects of a business's operations

Problem Statement: The context of this initiative was Northwind Traders' need to enhance its business operations through data-driven decisions. The problem statement entailed utilizing the Northwind database's sales data to resolve crucial business challenges, such as comprehending customer requirements, optimizing inventory, and improving supplier relationships.

The objective of the project was to bridge the distance between raw data by using PostgreSQL and crawl insights so that Northwind Traders could remain competitive and profitable in the specialty foods market.

The objectives can satisfy the stakeholders’ requirements which included the insights to make informed decisions, applying the suitable marketing strategies.

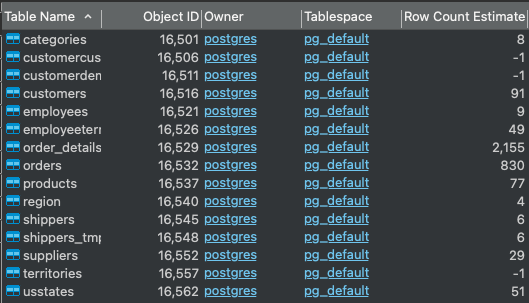
# Data Understanding

* + - 1. **Dataset**

The data is stored in a relational database management system (RDBMS) in a Postgres Database. Data is extracted using SQL queries from various database tables which are stored in the ‘public’ schema. The Northwind database is a fictitious dataset developed for educational purposes and not founded on actual data. As a result, it may not adequately represent the complexity of actual business operations. To visualize the relationship between the tables, we can use an Entity Relationship Diagram (ERD).

(ERD Table)

As described in the ERD above, there are 13 tables in the diagram which all are connected by at least 1 attribute. The data contains about 2155 transactions including all the important information about the transaction like clients, employees, expenditures, items, and orders.



With that being shown, we need to answer 10 different questions by query method and apply our business analysis techniques to gain meaning full insights to guide the department teams understand the status of the business performance. There are 4 department teams are asking these questions:

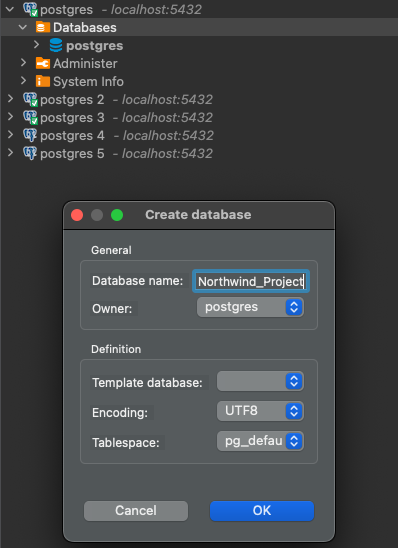
* The Product Team
* The Logistic Team
* The HR Team
* The Pricing Team
  + - 1. **Instruction to create and connect to database**

First and foremost, we need to create our own database and then load the data from the given materials. To create a new database, we need to right click the ‘Databases’ list in the Database Navigator and click on ‘Create New Databases’.

**A screenshot of a computer

Description automatically generated**

After that, a small window will appear and show us the general and definition information about the new database which we can modify. I named the new database ‘Northwind Project’ with the owner is ‘postgres’ and encoding = ‘UTF8’.

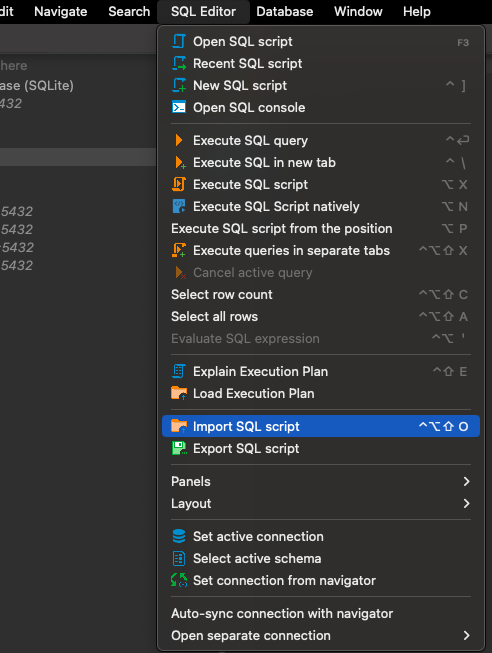
****

When we finished creating the new database, we will need to right click on the Northwind Project database and open a new SQL script.

A screenshot of a computer

Description automatically generated

In the navigation bar, we choose ‘SQL Editor’ and then click on ‘Import SQL script’.



Now, we locate the given script file and open it to import the script in the database script.

A screenshot of a computer

Description automatically generated

Then select the ‘Execute’ button to start loading the data into the schema.

When the data loading process is finished, we can finally start to use the information to answer all the business questions.

# Business Question 1

## 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 price

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).

To facilitate their pricing strategy evaluation, the Product Team requires a complete list containing the names and unit prices of all products falling within the specified price range ($20 to $50). This information will assist them in making data-driven pricing and positioning decisions for the company's products.

### Results

|  |  |
| --- | --- |
| **product\_name** | **unit\_price** |
| Tarte au sucre | 49,3 |
| Ipoh Coffee | 46 |
| Vegie-spread | 43,9 |
| Schoggi Schokolade | 43,9 |
| Northwoods Cranberry Sauce | 40 |
| Gnocchi di nonna Alice | 38 |
| Queso Manchego La Pastora | 38 |
| Gudbrandsdalsost | 36 |
| Mozzarella di Giovanni | 34,8 |
| Camembert Pierrot | 34 |
| Wimmers gute Semmelkn�del | 33,25 |
| Mascarpone Fabioli | 32 |
| Gumb�r Gummib�rchen | 31,23 |
| Ikura | 31 |
| Uncle Bob's Organic Dried Pears | 30 |
| Sirop d'�rable | 28,5 |
| Gravad lax | 26 |
| Nord-Ost Matjeshering | 25,89 |
| Grandma's Boysenberry Spread | 25 |
| P�t� chinois | 24 |
| Tofu | 23,25 |
| Chef Anton's Cajun Seasoning | 22 |
| Flotemysost | 21,5 |
| Louisiana Fiery Hot Pepper Sauce | 21,05 |
| Queso Cabrales | 21 |
| Gustaf's Kn�ckebr�d | 21 |
| Maxilaku | 20 |

There are 27 products that have the unit price falling in the price range of $20 and $50 which is described in a descending order in the table above.

## Key insights and Findings

As we can see the products in the price range is diversely distributed regardless of the types of the product. The price range is $20 to $50 because if you want to sell something below $20, it might be difficult to generate a profit after adjusting for the Cost of goods sold (CoGS) and advertising cost. On the other hand, if the price exceeds $50, it becomes extremely challenging to convert customers due to their value-based consideration. This would help the Product Team know which products to focus on and make informed decisions.

# Business Question 2

## 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).

In short, the Logistics Team seeks to identify countries whose 1998 orders had lengthier processing and shipping times (greater than or equal to 5 days) and a substantial number of orders (more than 10 orders). This information will enable them to evaluate their logistics performance for the specified year and make adjustments as necessary.

### Results

|  |  |  |
| --- | --- | --- |
| **ship\_country** | **average\_days\_between\_order\_shipping** | **total\_number\_orders** |
| Austria | 5,89 | 11 |
| Brazil | 8,12 | 28 |
| France | 9,43 | 23 |
| Germany | 5,38 | 34 |
| Spain | 7,83 | 12 |
| Sweden | 13,29 | 14 |
| UK | 6,25 | 16 |
| USA | 7,89 | 39 |
| Venezuela | 8,73 | 18 |

## There are totally 9 countries that have the bad performances in the year of 1998

## Key insights and Findings

As in the graph above, in 1998, Austria, Germany and UK have the fastest average delivery days which respectively are 5.89, 5.38, and 6.25 days. At the same time, the order shipping days delay among the rest variate around the range of 8 to 9 day except for the value of Sweden, which is extremely high and considered the slowest shipping speeds of all the countries. This is the alert for the need of improvement and enhancement in the delivery quality in these countries which accounted to the hand of the Logistic Team.

## The chart above shows us the correlation between the average days between order and shipping vs the total number of orders. This data indicates a negative correlation between the time required to process and dispatch orders and the quantity of orders. In other words, there is a tendency for the total number of orders to decrease when orders take longer to dispatch. This result is consistent with your previous assertion regarding the negative relationship between shipping days and order quantity in terms of logistics performance.

# Business Question 3

## 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).

This data will provide the HR Team with a comprehensive overview of employee details, their age at the time of hire, and their current manager, facilitating HR management and analysis.

### Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **employee\_full\_name** | **employee\_title** | **employee\_age** | **manager\_full\_name** | **manager\_title** |
| Anne Dodsworth | Sales Representative | 28 | Steven Buchanan | Sales Manager |
| Janet Leverling | Sales Representative | 28 | Andrew Fuller | Vice President, Sales |
| Michael Suyama | Sales Representative | 30 | Steven Buchanan | Sales Manager |
| Robert King | Sales Representative | 33 | Steven Buchanan | Sales Manager |
| Laura Callahan | Inside Sales Coordinator | 36 | Andrew Fuller | Vice President, Sales |
| Steven Buchanan | Sales Manager | 38 | Andrew Fuller | Vice President, Sales |
| Nancy Davolio | Sales Representative | 43 | Andrew Fuller | Vice President, Sales |
| Margaret Peacock | Sales Representative | 55 | Andrew Fuller | Vice President, Sales |

This table illustrate the 8 out of 9 employees in the company and the staff is missing is Andrew Fuller which is the Vice President.

## Key insights and Findings

The table displays the ages of employees when they entered the corporation, which range between 28 and 55 years. The table is sorted by employee age in ascending order, with the youngest employees listed first. This clearly illustrates the various age distribution of employees. Some employees were comparatively youthful at the time of their employment (e.g., 28 years), whereas others joined the company at an older age (e.g., 55 years). Moreover, sales-related positions, including "Sales Representative" and "Inside Sales Coordinator," report to the "Sales Manager," who in turn reports to the "Vice President, Sales." The reporting relationships illustrate this hierarchical structure.

# Business Question 4

## 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” 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).

In this question, from the years 1997 to 1998, the Logistics Team is attempting to identify the months with a high number of orders (greater than 35) and significant total freight which is the total cost of transportation of goods. This data will assist them in evaluating their performance during the specified time period and identifying successful months for further examination or improvement.

### Results

|  |  |  |
| --- | --- | --- |
| **year\_month** | **total\_number\_orders** | **total\_freight** |
| Apr-98 | 74 | 6394 |
| Jan-98 | 55 | 5463 |
| Mar-98 | 73 | 5379 |
| Feb-98 | 54 | 4273 |
| Oct-97 | 38 | 3946 |
| Dec-97 | 48 | 3758 |
| Sep-97 | 37 | 3237 |

The table summarized 7 months of the 2-year period that have the total number of orders greater than 35 orders and significant total freight.

## Key insights and Findings

As the result table from the query of the Logistic Team’s performance, I have plotted the above graph show the Total Freight and total number of orders in the sorted timeline. The linear line shows the overall trend of the graph is an upward trend in performance which indicate great improvement between 1997 and 1998. April 1998 is the month with the greatest total freight and the greatest number of orders (74 orders). This suggests that April 1998 was an exceptionally fruitful month for the logistics team. Moreover, the months that have the most total number of orders tend to fall in late 1997 and early 1998 which show the promising period for the Logistic Team to focus on.

# Business Question 5

## 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).

In summary, the Pricing Team seeks to identify products that have experienced unit price increases outside of the specified range and to provide comprehensive information about these products, including their names, current and initial unit prices, and the calculated percentage increase. In this question, we consider the ‘initial unit price’ is the unit price of the earliest price of a product and the ‘current unit price’ is the latest price, sorted by ‘order\_date’ column.

### Results

|  |  |  |  |
| --- | --- | --- | --- |
| **product\_name** | **current\_price** | **previous\_unit\_price** | **percentage\_increase** |
| Mozzarella di Giovanni | 34,8 | 34,8 | 0 |
| Singaporean Hokkien Fried Mee | 14 | 9,8 | 42,86 |
| Queso Cabrales | 21 | 14 | 50 |

There are only 3 products have that satisfied the requirement query of the Pricing Team which have the unit price increase outside the range from 20% to 30%.

## Key insights and Findings

These products have encountered unit price changes outside the specified 20% to 30% range. This data can be useful for the Pricing Team in evaluating and possibly adjusting pricing strategies for these products based on their percentage price changes. Products with price increases of less than 20% may indicate that the pricing strategy is not keeping up with inflation or cost increases. In contrast, products with percentage increases exceeding 30% may indicate that price increases are negatively impacting demand.

In conclusion, searching for products with percentage increases outside the range of 20% to 30% enables the Pricing Team to identify outliers and potential pricing strategy issues. It enables them to make educated pricing adjustments decisions based on the observed percentage price changes.

# Business Question 6

## 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. “1. Below $20”
   2. “2. $20 - $50”
   3. “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).

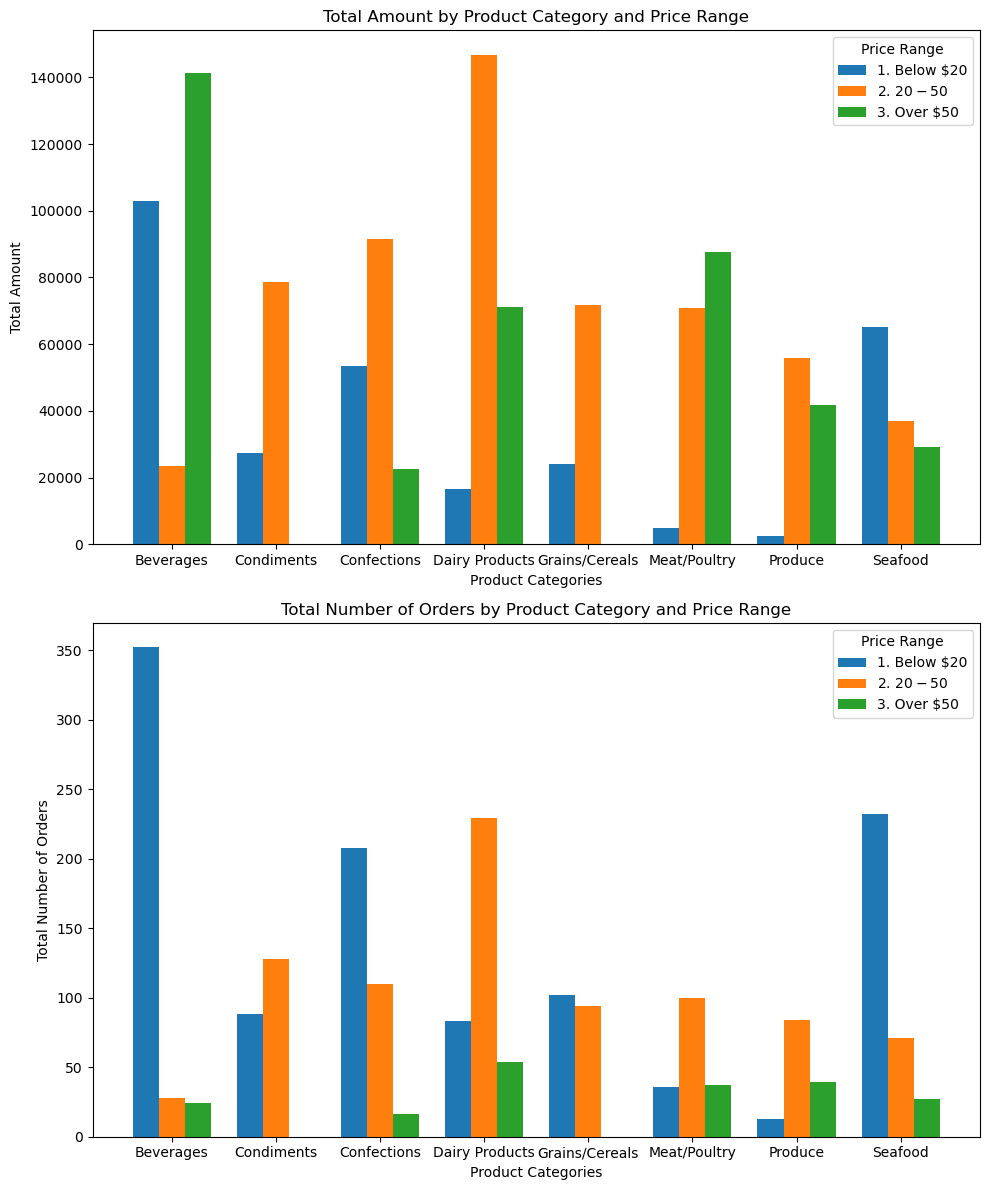
This analysis will shed light on how different product categories perform across various price ranges, allowed the Pricing Team to make informed decisions regarding pricing and discounts within each category.

### Results

|  |  |  |  |
| --- | --- | --- | --- |
| **category\_name** | **price\_range** | **total\_amount** | **total\_number\_orders** |
| Beverages | 1. Below $20 | 102945 | 352 |
| Beverages | 2. $20 - $50 | 23527 | 28 |
| Beverages | 3. Over $50 | 141397 | 24 |
| Condiments | 1. Below $20 | 27299 | 88 |
| Condiments | 2. $20 - $50 | 78748 | 128 |
| Confections | 1. Below $20 | 53365 | 208 |
| Confections | 2. $20 - $50 | 91429 | 110 |
| Confections | 3. Over $50 | 22563 | 16 |
| Dairy Products | 1. Below $20 | 16569 | 83 |
| Dairy Products | 2. $20 - $50 | 146783 | 229 |
| Dairy Products | 3. Over $50 | 71156 | 54 |
| Grains/Cereals | 1. Below $20 | 24071 | 102 |
| Grains/Cereals | 2. $20 - $50 | 71673 | 94 |
| Meat/Poultry | 1. Below $20 | 4728 | 36 |
| Meat/Poultry | 2. $20 - $50 | 70699 | 100 |
| Meat/Poultry | 3. Over $50 | 87595 | 37 |
| Produce | 1. Below $20 | 2432 | 13 |
| Produce | 2. $20 - $50 | 55732 | 84 |
| Produce | 3. Over $50 | 41820 | 39 |
| Seafood | 1. Below $20 | 65110 | 232 |
| Seafood | 2. $20 - $50 | 36980 | 71 |
| Seafood | 3. Over $50 | 29172 | 27 |

The output data contains 22 rows which includes the price range, total amount, and total number of orders of 8 different categories.

## Key insights and Findings



In the first question, we have known that the price range is $20 to $50 because it may be difficult to generate a profit after adjusting for the Cost of goods sold (CoGS) and advertising cost if you sell something for less than $20. However, if the price transcends $50, it is exceedingly difficult to convert customers due to their value-based consideration. Now, the Pricing Team want to look at the price range distribution from each category rather than from each product.

The highest total amount falls fundamentally on the range ‘$20 - $50’ with the peak is category “Dairy Products” with the amount of $146,783 in sale. However, there’re still some noticeable high amount in sale that appear in the range ‘Below $20’ of the category “Beverages” and “Seafood” which respectively are $102,945 and $65,110 and, in the range “Above $50” are $141,397 of “Beverages” and $87,595 of “Meat/Poultry”.

On the table Total number of orders, the significant numbers of orders were purchased in the range “Below $20” in categories “Beverages”, “Seafood” and “Confections” which all are above 200 orders. Only 1 category has the large number of orders in range of “$20 - $50” is “Dairy Products” and the rest are fluctuated at the range of 100 orders , some even lower. Categories with lesser total amounts or fewer orders may present opportunities for development or targeted marketing. This table delivers an understandable description of how different product categories perform within different price ranges, allowing the team to make pricing and marketing decisions based on data.

# Business Question 7

## 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:
   1. “America”
   2. “Europe”
   3. “Asia-Pacific”
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 category name, then supplier region and reorder level (each in ascending order).

This information will enable the Logistics Team to analyse the unit in stock and unit on orders from regional suppliers for various product categories, facilitating in inventory management and supply chain planning.

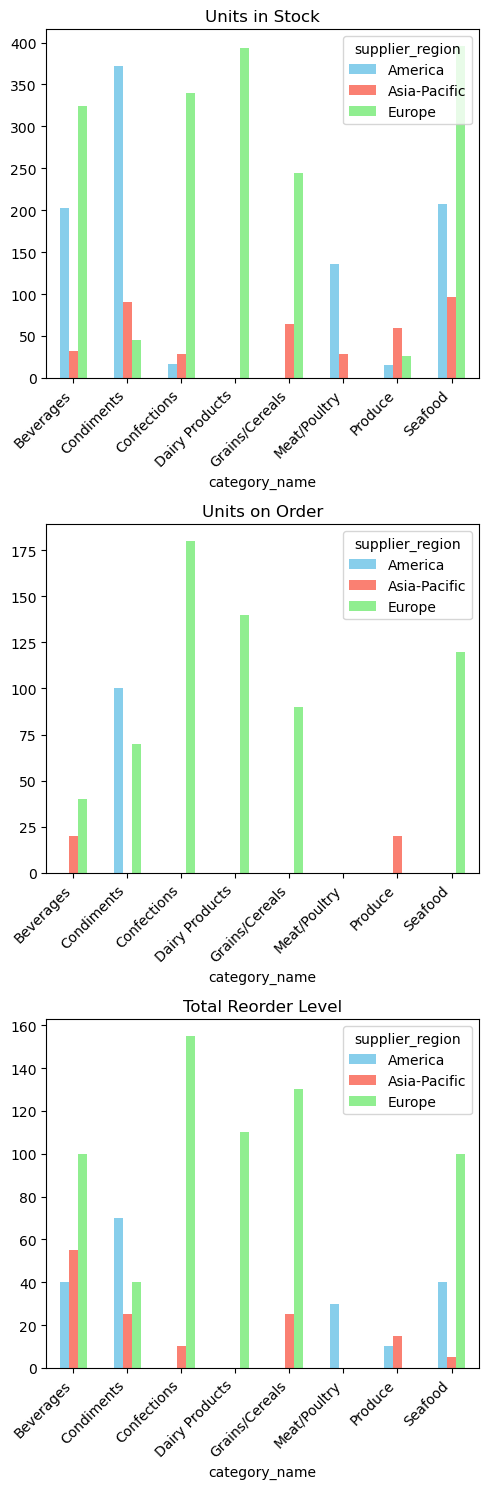
### Results

## 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **category\_name** | **supplier\_region** | **unit\_in\_stock** | **unit\_on\_order** | **reorder\_level** |
| Beverages | America | 203 | 0 | 40 |
| Beverages | Asia-Pacific | 32 | 20 | 55 |
| Beverages | Europe | 324 | 40 | 100 |
| Condiments | America | 372 | 100 | 70 |
| Condiments | Asia-Pacific | 90 | 0 | 25 |
| Condiments | Europe | 45 | 70 | 40 |
| Confections | America | 17 | 0 | 0 |
| Confections | Asia-Pacific | 29 | 0 | 10 |
| Confections | Europe | 340 | 180 | 155 |
| Dairy Products | Europe | 393 | 140 | 110 |
| Grains/Cereals | Asia-Pacific | 64 | 0 | 25 |
| Grains/Cereals | Europe | 244 | 90 | 130 |
| Meat/Poultry | America | 136 | 0 | 30 |
| Meat/Poultry | Asia-Pacific | 29 | 0 | 0 |
| Meat/Poultry | Europe | 0 | 0 | 0 |
| Produce | America | 15 | 0 | 10 |
| Produce | Asia-Pacific | 59 | 20 | 15 |
| Produce | Europe | 26 | 0 | 0 |
| Seafood | America | 208 | 0 | 40 |
| Seafood | Asia-Pacific | 97 | 0 | 5 |
| Seafood | Europe | 396 | 120 | 100 |

The table shows all 21 rows and can be used to monitor and manage the inventory levels of various product categories across multiple supplier regions. It reveals which products are in stock, which are on order, and when resupply may be necessary to maintain optimal inventory levels. This information can be crucial for supply chain management and ensuring customer product availability.

## Key insights and Findings



A graph with different colored lines

Description automatically generated

A graph with different colored lines

Description automatically generated

As is presented in the charts above, the data is divided into 3 segments: Unit in stock, unit on order, and total reorder level. The number of units in stock, units on order or the total reorder level in region “Europe” is tremendously larger than the other regions. However, in 2 categories “Meat/Poultry” and “Produce”, countries in “Europe” hardly purchase so the value of these 2 categories in Europe is next to 0. This could be cause for concern; therefore, the team should investigate the stock situation and determine reorder levels. In Europe, the "Confections" category has the highest reorder rate, indicating a possible need for more frequent replenishment. In Asia-Pacific, the "Beverages" category has a high reorder rate, indicating the need for proactive inventory management in this region. There is a significant number of units are on order for the "Condiments" category in America, while the same category has a high reorder rate in Europe. This suggests that these localities have distinct demand patterns. Categories with low units in stock and high reorder rates like “Beverages” and “Seafood” in Asian-Pacific require special attention to prevent stockouts and effectively meet customer demand.

The logistics team must continue to monitor inventory levels, collaborate with suppliers, and modify reorder points as necessary to ensure that products are available to meet customer demand while minimising excess inventory. In addition, they should investigate and resolve the limited stock situation in Europe's "Meat/Poultry" and “Produce” category.

# Business Question 8

## 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 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. “Equal Average”
   3. “Over Average”
7. their position against the category median unit price as:
   1. “Below Median”
   2. “Equal 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).

This analysis will help the Pricing Team evaluate the competitiveness of each product's pricing within its category.

### Results

## 



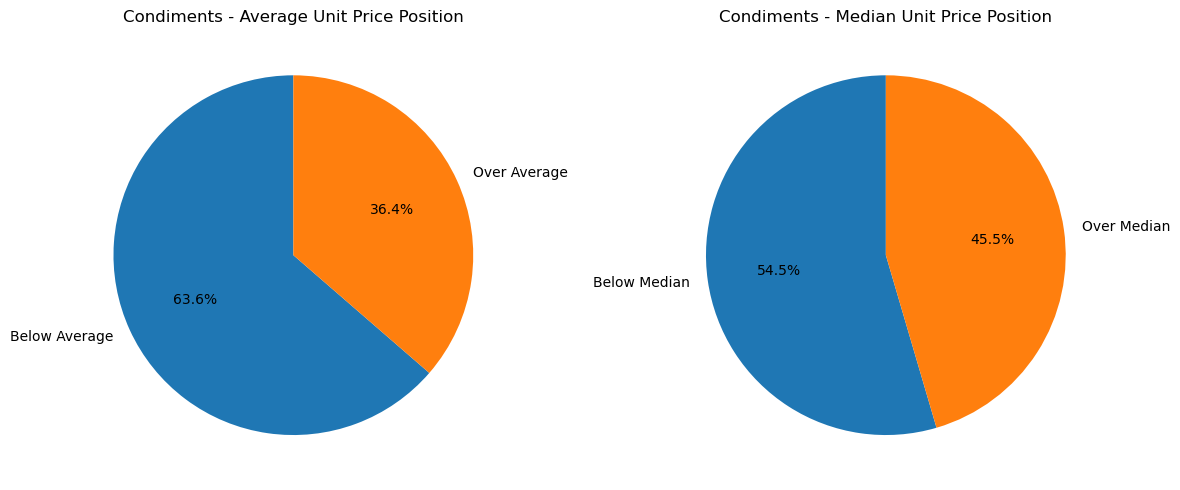
The table includes 67 data rows, and it provides useful insight into the relationship between the unit prices of individual products and the overall pricing characteristics of their respective product categories. It aids in determining whether a product's price is more or less expensive than the average or median of its category, which can be helpful in developing pricing methods and advertising.

## Key insights and Findings

A blue and orange pie chart

Description automatically generated

There are significant price variations within the Beverages category, with some products being priced significantly higher than others. The majority of items in this category are priced below the average for the category.

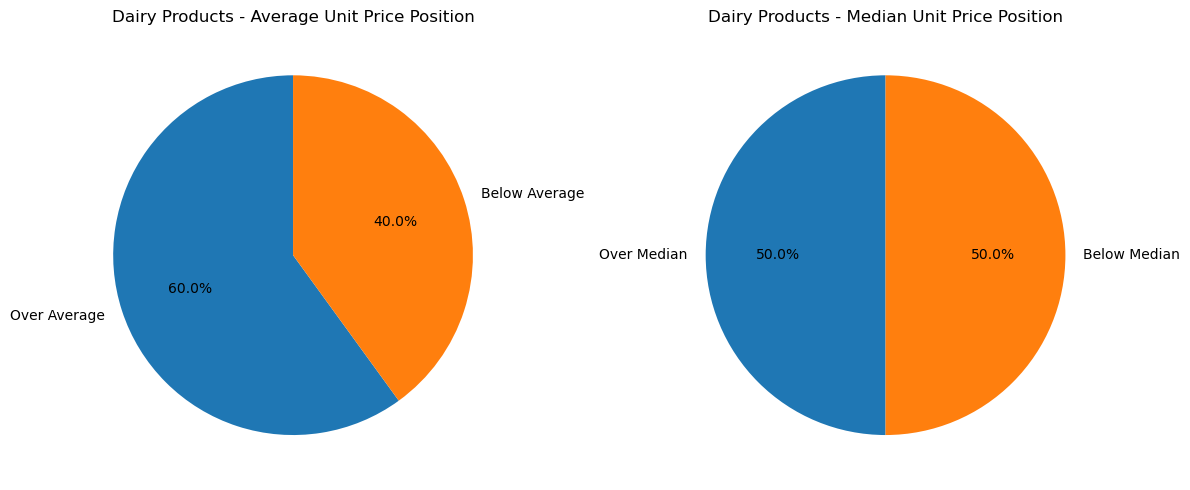


The prices in the Condiments category are relatively consistent, but "Northwoods Cranberry Sauce" is significantly more expensive than the category average.

A blue and orange pie chart

Description automatically generated

The Confections category contains items with a broad spectrum of prices. "Sir Rodney's Marmalade" is the most expensive item in this category.

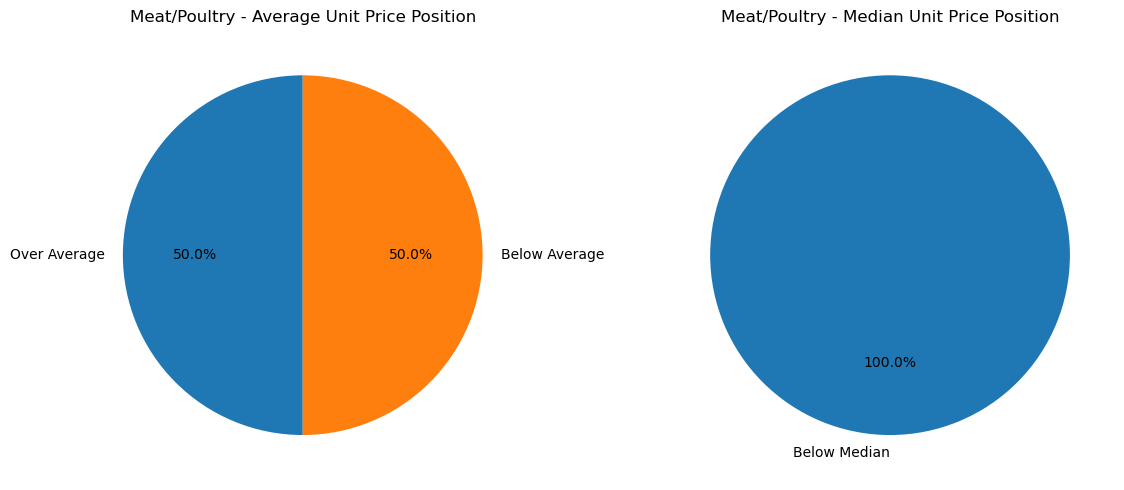


Within the category of Dairy Products, "Raclette Courdavault" has the most expensive price. The majority of products are priced near the category average.

A blue and orange pie chart

Description automatically generated

"Gnocchi di nonna Alice" is a product that stands out with a substantially higher price than others in the Grains/Cereals category. The majority of items in this category are priced below the average for the category.



Meat/Poultry Category: "Paté chinois" is the most expensive item in the Meat/Poultry category, whereas "Tourtière" is the least expensive. However there are only 2 products in this category which we can’t conclude anything.

A diagram of a product and a product

Description automatically generated with medium confidence

In the Produce category, prices are relatively consistent, with "Manjimup Dried Apples" being the most expensive item.

A blue and orange pie chart

Description automatically generated

Within the Seafood category, prices vary considerably. "Carnarvon Tigers" is the most expensive product in this category.

The preceding pie charts illustrate how product unit prices compare to their respective category averages and medians. It helps identify pricing anomalies and products with prices that deviate from category norms. This data is valuable for pricing strategies, promotions, and product positioning within each category.

# Business Question 9

## 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).

Total sales (including discounts) will be used to arrange the results in descending order. These KPIs will assist the Sales Team in evaluating and comparing employee performance based on a variety of sales-related metrics, allowing for greater insight into their efficacy in driving sales.

### Results

|  |  |  |
| --- | --- | --- |
| **employee\_full\_name** | **employee\_title** | **total\_sale\_amount\_excluding\_discount** |
| Margaret Peacock | Sales Representative | 250187,45 |
| Janet Leverling | Sales Representative | 213051,3 |
| Nancy Davolio | Sales Representative | 202143,71 |
| Andrew Fuller | Vice President, Sales | 177749,26 |
| Laura Callahan | Inside Sales Coordinator | 133301,03 |
| Robert King | Sales Representative | 141295,99 |
| Anne Dodsworth | Sales Representative | 82964 |
| Michael Suyama | Sales Representative | 78198,1 |
| Steven Buchanan | Sales Manager | 75567,75 |

|  |  |  |  |
| --- | --- | --- | --- |
| **total\_number\_unique\_orders** | **total\_number\_orders** | **average\_product\_amount** | **average\_amount\_per\_order** |
| 156 | 420 | 595,68 | 1603,77 |
| 127 | 321 | 663,71 | 1677,57 |
| 123 | 345 | 585,92 | 1643,44 |
| 96 | 241 | 737,55 | 1851,55 |
| 104 | 260 | 512,7 | 1281,74 |
| 72 | 176 | 802,82 | 1962,44 |
| 43 | 107 | 775,36 | 1929,4 |
| 67 | 168 | 465,46 | 1167,14 |
| 42 | 117 | 645,88 | 1799,23 |

|  |  |  |
| --- | --- | --- |
| **total\_discount\_amount** | **total\_sale\_amount\_including\_discount** | **total\_discount\_percentage** |
| 17296,6 | 232890,85 | 6,91 |
| 10238,46 | 202812,84 | 4,81 |
| 10036,11 | 192107,6 | 4,96 |
| 11211,51 | 166537,76 | 6,31 |
| 6438,75 | 126862,28 | 4,83 |
| 16727,76 | 124568,23 | 11,84 |
| 5655,93 | 77308,07 | 6,82 |
| 4284,97 | 73913,13 | 5,48 |
| 6775,47 | 68792,28 | 8,97 |

This table provides information regarding the efficacy of sales personnel of total 9 employees, including their total sales, the number of orders they manage, the average product and order quantities, and the discounts they implement. These metrics are crucial for evaluating the efficiency and output of the Sales team and individual employees.

## Key insights and Findings

A graph of green bars

Description automatically generated with medium confidence

According to total sales excluding discounts, Margaret Peacock, Janet Leverling, and Nancy Davolio lead. Margaret Peacock leads overall sales. Revenue by Title: Sales Representatives, Vice President of Sales, Inside Sales Coordinator, and Sales Manager are included. It's intriguing to compare sales success across different roles. Employee Average Product Quantity: Some workers have larger average product quantities each order. Robert King has a much higher average product quantity per order than other merchants. The table shows all discounts and their percentages. Robert King may provide greater discounts since he has the most overall discounts. However, percentage discount is also important. Orders: Record the total number of unique and employee orders. This shows their consumer base and order volume. This column shows the total sale, including discounts. Compare that to the entire transaction amount minus discounts to see how discounts effect sales. The discount % is necessary for calculating the percentage of sales that are discounted. Finding a balance between discounts to attract customers and profitability is key.

# Business Question 10

## 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).

These KPIs will provide the Sales Team with insight into how the sales performance of each employee compares to that of their colleagues within each product category and within the context of the complete sales team.

### Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **category\_name** | **employee\_full\_name** | **total\_sale\_amount** | **percent\_of\_employee\_sales** | **percent\_of\_category\_sales** |
| Beverages | Margaret Peacock | 50308,21 | 0,21602 | 0,18781 |
| Beverages | Nancy Davolio | 46599,35 | 0,24257 | 0,17396 |
| Beverages | Janet Leverling | 44757,4 | 0,22068 | 0,16709 |
| Beverages | Andrew Fuller | 40248,25 | 0,24168 | 0,15025 |
| Beverages | Robert King | 27963,83 | 0,22449 | 0,10439 |
| Beverages | Anne Dodsworth | 19642,55 | 0,25408 | 0,07333 |
| Beverages | Laura Callahan | 17897,85 | 0,14108 | 0,06682 |
| Beverages | Steven Buchanan | 11000,53 | 0,15991 | 0,04107 |
| Beverages | Michael Suyama | 9450,2 | 0,12786 | 0,03528 |
| Condiments | Margaret Peacock | 23314,87 | 0,10011 | 0,21985 |
| Condiments | Andrew Fuller | 14850,67 | 0,08917 | 0,14004 |
| Condiments | Laura Callahan | 14637,66 | 0,11538 | 0,13803 |
| Condiments | Nancy Davolio | 13561,56 | 0,07059 | 0,12788 |
| Condiments | Janet Leverling | 13381,64 | 0,06598 | 0,12619 |
| Condiments | Anne Dodsworth | 10125,54 | 0,13098 | 0,09548 |
| Condiments | Robert King | 8851,37 | 0,07106 | 0,08347 |
| Condiments | Michael Suyama | 4648,47 | 0,06289 | 0,04383 |
| Condiments | Steven Buchanan | 2675,3 | 0,03889 | 0,02523 |
| Confections | Janet Leverling | 33622,4 | 0,16578 | 0,2009 |
| Confections | Nancy Davolio | 28568,92 | 0,14871 | 0,17071 |
| Confections | Margaret Peacock | 27768,73 | 0,11923 | 0,16592 |
| Confections | Laura Callahan | 21699,91 | 0,17105 | 0,12966 |
| Confections | Andrew Fuller | 21455,69 | 0,12883 | 0,1282 |
| Confections | Robert King | 14518,99 | 0,11655 | 0,08675 |
| Confections | Anne Dodsworth | 8053,16 | 0,10417 | 0,04812 |
| Confections | Michael Suyama | 6859,63 | 0,09281 | 0,04099 |
| Confections | Steven Buchanan | 4809,8 | 0,06992 | 0,02874 |
| Dairy Products | Nancy Davolio | 36022,98 | 0,18751 | 0,15361 |
| Dairy Products | Margaret Peacock | 33549,8 | 0,14406 | 0,14307 |
| Dairy Products | Janet Leverling | 32320,83 | 0,15936 | 0,13782 |
| Dairy Products | Robert King | 27621,86 | 0,22174 | 0,11779 |
| Dairy Products | Andrew Fuller | 23812,55 | 0,14299 | 0,10154 |
| Dairy Products | Steven Buchanan | 21937,63 | 0,3189 | 0,09355 |
| Dairy Products | Laura Callahan | 21101,47 | 0,16633 | 0,08998 |
| Dairy Products | Anne Dodsworth | 21101,12 | 0,27295 | 0,08998 |
| Dairy Products | Michael Suyama | 17039,04 | 0,23053 | 0,07266 |
| Grains/Cereals | Margaret Peacock | 22579,61 | 0,09695 | 0,23583 |
| Grains/Cereals | Janet Leverling | 21235,01 | 0,1047 | 0,22179 |
| Grains/Cereals | Andrew Fuller | 11172,95 | 0,06709 | 0,1167 |
| Grains/Cereals | Laura Callahan | 11072,05 | 0,08728 | 0,11564 |
| Grains/Cereals | Michael Suyama | 9410,7 | 0,12732 | 0,09829 |
| Grains/Cereals | Nancy Davolio | 8465,9 | 0,04407 | 0,08842 |
| Grains/Cereals | Robert King | 6535,5 | 0,05247 | 0,06826 |
| Grains/Cereals | Steven Buchanan | 4027,56 | 0,05855 | 0,04207 |
| Grains/Cereals | Anne Dodsworth | 1245,3 | 0,01611 | 0,01301 |
| Meat/Poultry | Margaret Peacock | 30867,14 | 0,13254 | 0,18934 |
| Meat/Poultry | Andrew Fuller | 29873,6 | 0,17938 | 0,18325 |
| Meat/Poultry | Robert King | 21176,72 | 0,17 | 0,1299 |
| Meat/Poultry | Janet Leverling | 20502,62 | 0,10109 | 0,12577 |
| Meat/Poultry | Laura Callahan | 16395,28 | 0,12924 | 0,10057 |
| Meat/Poultry | Nancy Davolio | 15038,47 | 0,07828 | 0,09225 |
| Meat/Poultry | Steven Buchanan | 11488,2 | 0,167 | 0,07047 |
| Meat/Poultry | Michael Suyama | 9003,69 | 0,12181 | 0,05523 |
| Meat/Poultry | Anne Dodsworth | 8676,66 | 0,11223 | 0,05322 |
| Produce | Nancy Davolio | 19706,25 | 0,10258 | 0,19709 |
| Produce | Margaret Peacock | 17186,56 | 0,0738 | 0,17189 |
| Produce | Laura Callahan | 12016,52 | 0,09472 | 0,12018 |
| Produce | Janet Leverling | 11960,85 | 0,05897 | 0,11963 |
| Produce | Michael Suyama | 11560,7 | 0,15641 | 0,11562 |
| Produce | Robert King | 10753,38 | 0,08633 | 0,10755 |
| Produce | Andrew Fuller | 9376,48 | 0,0563 | 0,09378 |
| Produce | Steven Buchanan | 7109,02 | 0,10334 | 0,0711 |
| Produce | Anne Dodsworth | 314,81 | 0,00407 | 0,00315 |
| Seafood | Margaret Peacock | 27315,93 | 0,11729 | 0,2081 |
| Seafood | Janet Leverling | 25032,09 | 0,12342 | 0,1907 |
| Seafood | Nancy Davolio | 24144,15 | 0,12568 | 0,18394 |
| Seafood | Andrew Fuller | 15747,57 | 0,09456 | 0,11997 |
| Seafood | Laura Callahan | 12041,54 | 0,09492 | 0,09174 |
| Seafood | Anne Dodsworth | 8148,9 | 0,10541 | 0,06208 |
| Seafood | Robert King | 7146,58 | 0,05737 | 0,05445 |
| Seafood | Michael Suyama | 5940,7 | 0,08037 | 0,04526 |
| Seafood | Steven Buchanan | 5744,25 | 0,0835 | 0,04376 |

This table presents insight into how the sales performance of every employee contributes to various product categories. It displays the total sales for each employee within each category, as well as their percentage of sales in that category and the category's percentage of total sales. This data can be utilised to evaluate employee performance and sales performance by category.

## Key insights and Findings

The table provides sales information for various product categories within your organisation. It displays the total quantity of sales for each employee in each category.

A pie chart with numbers and text

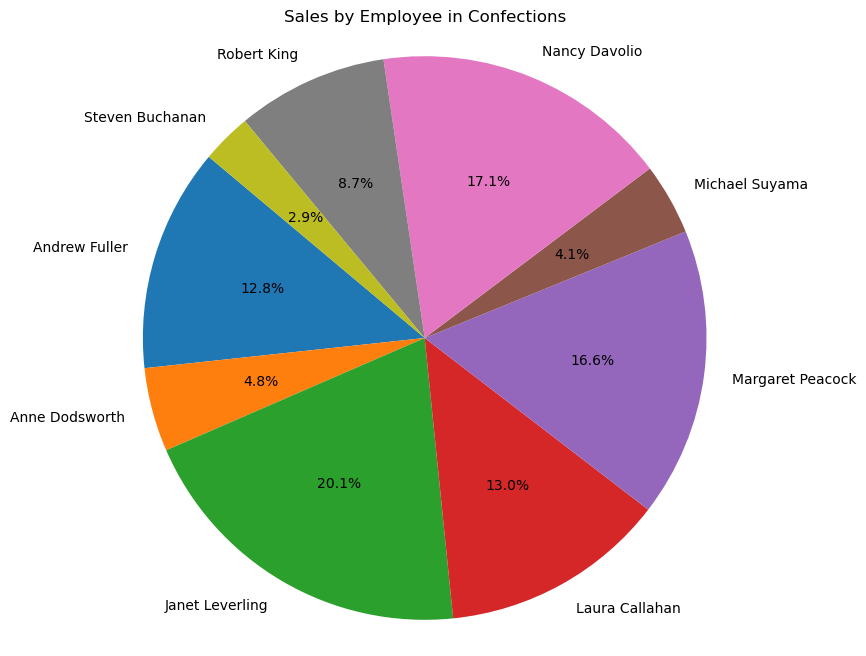
Description automatically generated

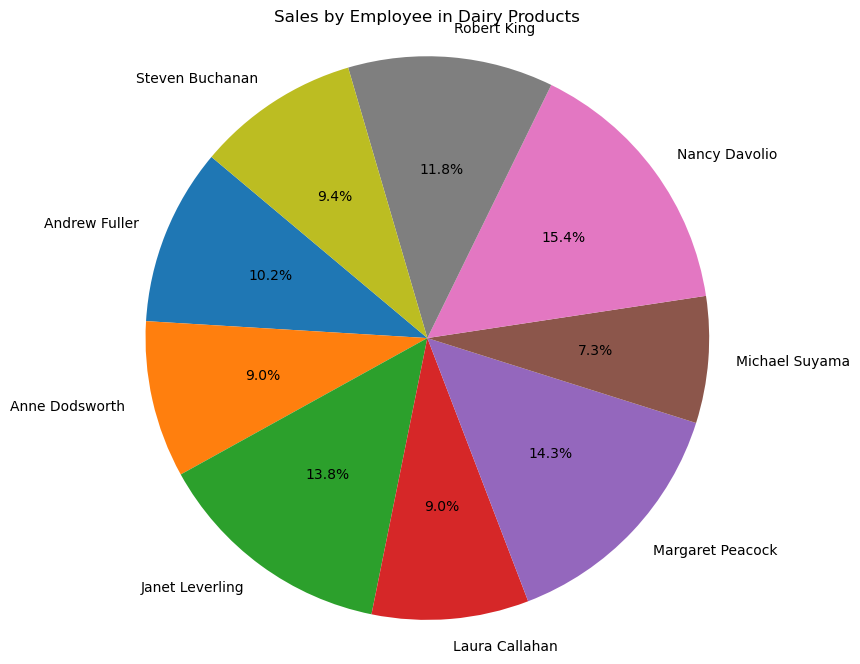
Margaret Peacock, Nancy Davolio, and Janet Leverling are the top sales contributors in the "Beverages" category.

A pie chart with numbers and text

Description automatically generated

Margaret Peacock and Andrew Fuller are significant contributors to the "Condiments" category, while Janet Leverling is the sales leader in the "Confections" category.





Different employees contribute differently to sales across various product categories, indicating a wide variety of product knowledge among employees.

The table also displays the percentage of total sales contributed by each employee within their respective categories, as well as the percentage of total category sales that each employee represents. This information can assist in determining the relative impact of each employee on particular product categories.

A pie chart with numbers and text

Description automatically generated

A pie chart with numbers and text

Description automatically generated

A pie chart with numbers and text

Description automatically generated

A pie chart with numbers and text

Description automatically generated

Category Impact Employees such as Margaret Peacock and Nancy Davolio have a significant impact on multiple categories, demonstrating their adaptability in managing a variety of products. Some employees, such as Steven Buchanan and Anne Dodsworth, have lower sales across multiple categories, which may suggest opportunities for advancement or alternative sales strategies. There is variation in sales contributions within each category, with some employees contributing a greater proportion than others. This information can be used to identify top performers and areas that may require additional support or training.

Overall, the table provides insightful information about the sales performance of employees across various product categories, allowing you to make informed decisions and optimise your sales strategies.

# Conclusion

In conclusion, the Northwind initiative illuminated the company's operations, sales, and consumer behaviour. Use these insights to optimise strategy, make data-driven choices, and boost corporate success. It also showed how data research and visualisation may provide you a commercial advantage. These insights may help the organisation improve its plans for growth and profit.