2022-10-7

Huanwei Xu 14090118

**Company Performance with SQL**

Data science Practice |Autumn 2022 | AT2

Table of Contents

[Project Overview 3](#_Toc116071222)

[Data introduction 3](#_Toc116071223)

[Business Questions: 6](#_Toc116071224)

[Question 1 6](#_Toc116071225)

[Answer 1 6](#_Toc116071226)

[The instruction for query 6](#_Toc116071227)

[The result of query 6](#_Toc116071228)

[Question 2 8](#_Toc116071229)

[Answer 2 8](#_Toc116071230)

[The instruction for query 8](#_Toc116071231)

[The result of query 9](#_Toc116071232)

[Question 3 12](#_Toc116071233)

[Answer 3 12](#_Toc116071234)

[The instruction for query 12](#_Toc116071235)

[The result of query 12](#_Toc116071236)

[Question 4 13](#_Toc116071237)

[Answer 4 14](#_Toc116071238)

[The instruction for query 14](#_Toc116071239)

[The result of query 14](#_Toc116071240)

[Question 5 17](#_Toc116071241)

[Answer 5 17](#_Toc116071242)

[The instruction for query 17](#_Toc116071243)

[The result of query 18](#_Toc116071244)

[Question 6 19](#_Toc116071245)

[Answer 6 19](#_Toc116071246)

[The instruction for query 19](#_Toc116071247)

[The result of query 20](#_Toc116071248)

[Question 7 23](#_Toc116071249)

[Answer 7 23](#_Toc116071250)

[The instruction for query 23](#_Toc116071251)

[The result of query 24](#_Toc116071252)

[Question 8 27](#_Toc116071253)

[Answer 8 28](#_Toc116071254)

[The instruction for query 28](#_Toc116071255)

[The result of query 28](#_Toc116071256)

[Question 9 37](#_Toc116071257)

[Answer 9 37](#_Toc116071258)

[The instruction for query 37](#_Toc116071259)

[The result of query 38](#_Toc116071260)

[Question 10 44](#_Toc116071261)

[Answer 10 44](#_Toc116071262)

[The instruction for query 44](#_Toc116071263)

[The result of the query 45](#_Toc116071264)

[Appendix 46](#_Toc116071265)

[The coding of Visualization in R 46](#_Toc116071266)

# Project Overview

This project aims to answer 10 questions and to find out insightful answers to improve the business questions properly.

To make the explore processing interpretable, the visualization tool from R was introduced.

# Data introduction

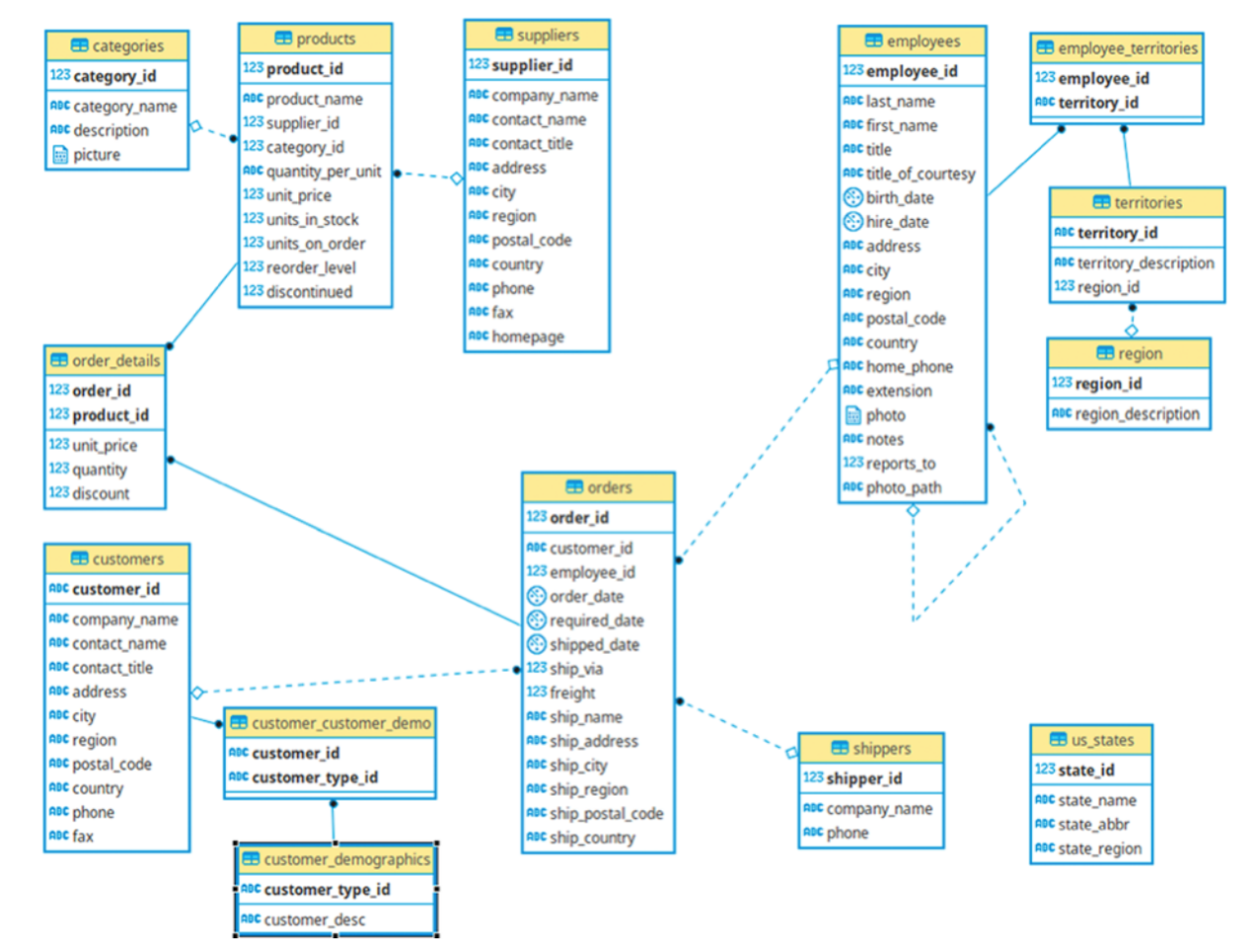


Figure 1. The schema of datasets.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table** | **Rows and Columns** | **Table description** | **Data issues** |
| categories | **8\*4** | The categories of product details | - |
| order\_details | **2155\*5** | Order  information, including product id,  Unit price,  quantity, and  discount | - |
| products | **77\*10** | Product information, including  supplier,  current price,  quantity, stocking, continued status, etc. | - |
| suppliers | **29\*12** | Supplier information, including name, contact details, etc. | There are 20,10 and 24 missing values in columns region, fax and home\_page respectively. |
| orders | **830\*14** | Orders information, including  Customer ID,  The employee of the order belongs,  date, address, etc | There are 21, 507 and 19 missing values in columns shipped\_date, ship\_region, and ship\_postal\_code respectively |
| employees | **9\*18** | Employee information, | There are 4 and 1 missing values in columns region and report\_to respectively |

Figure 2. The description of tales that were used in the analysis

# 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:

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 to 50)
2. they are not discontinued

Finally, order the results by product name in alphabetical order.

## Answer 1

### The instruction for query

This question only used table **products** in a single query**.**

### The result of query

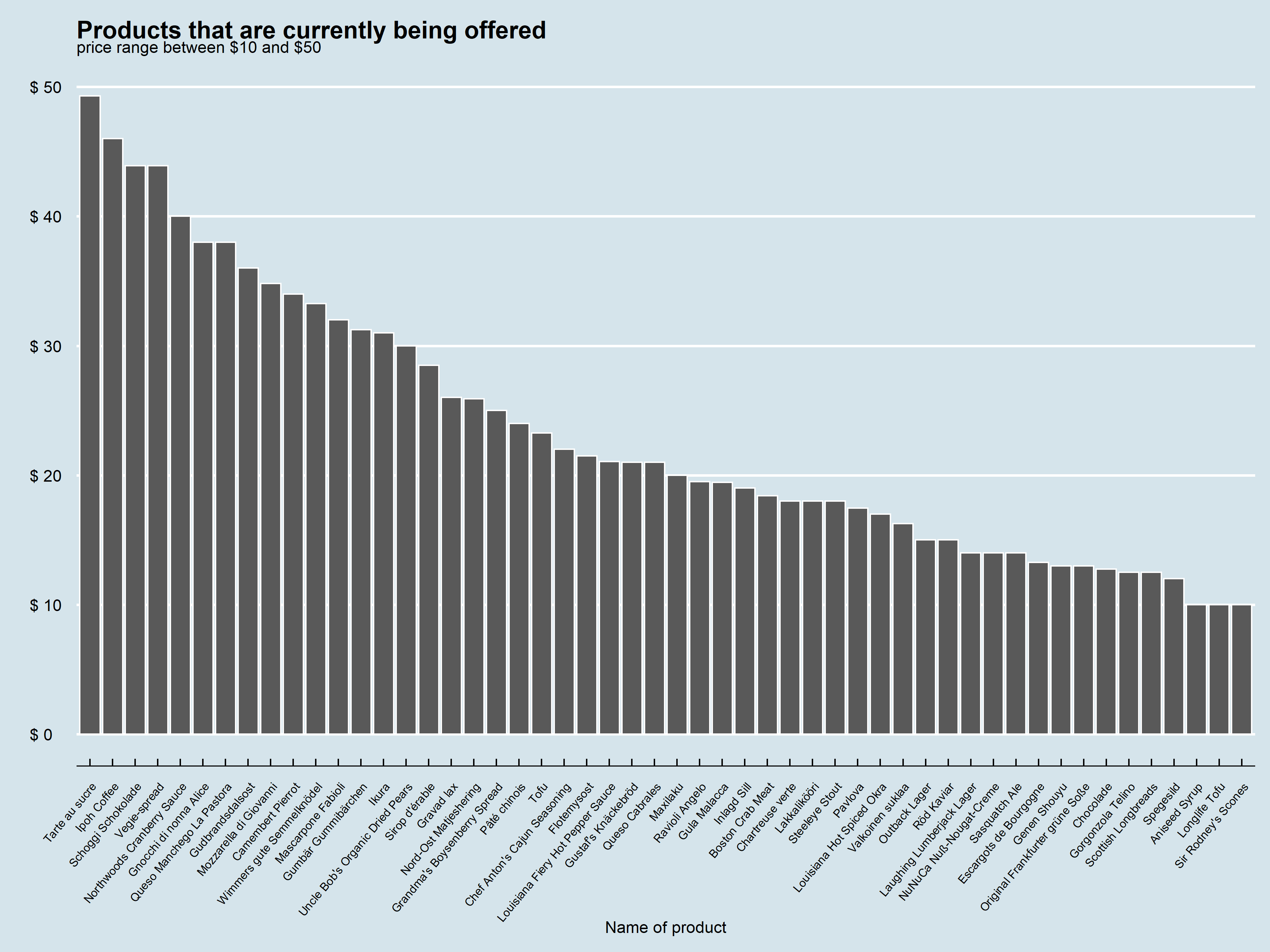


Figure 3.

The highest price of the product is Tarte au sucre which is $49.3, and the lowest price of products are Aniseed Syru, Longlife Tof and Sir Rodney's Scones which both are $10.

## 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 with 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 descending order (higher value first).

## Answer 2

### The instruction for query

This question only used table **orders** in a single query**.**

### The result of query

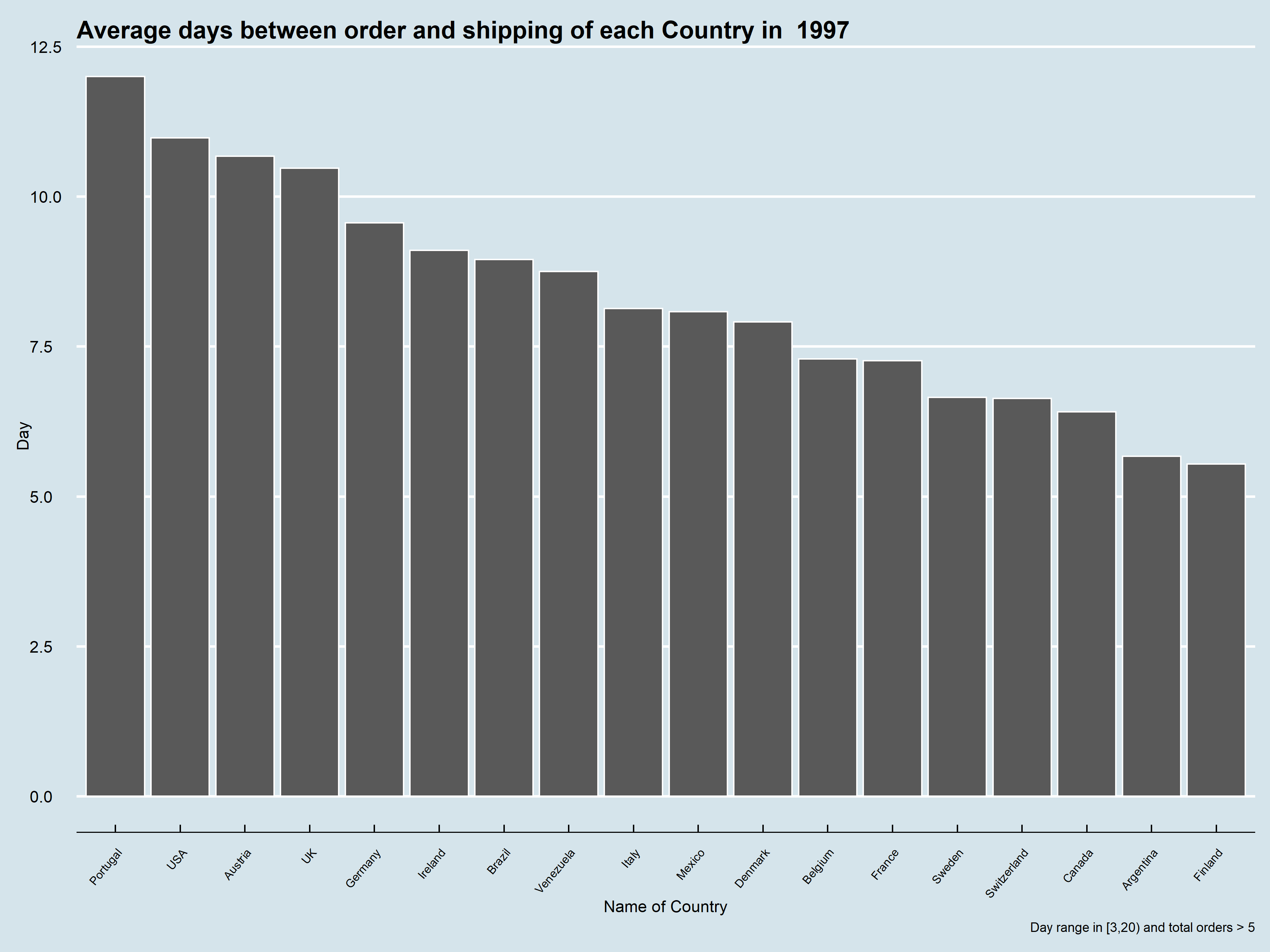


Figure 4.

As shown in figure 4, the worst performance in 1997 is Portugal, where the average time consumption between order and shipping is 12 days. The best one is Finland, where is 5.54 days.

Chart, histogram

Description automatically generated

Figure 5.

As shown in figure 5, the most total orders Country in 1997 is Germany, which is 64. The lowest total orders Country is Argentina, which is 6.

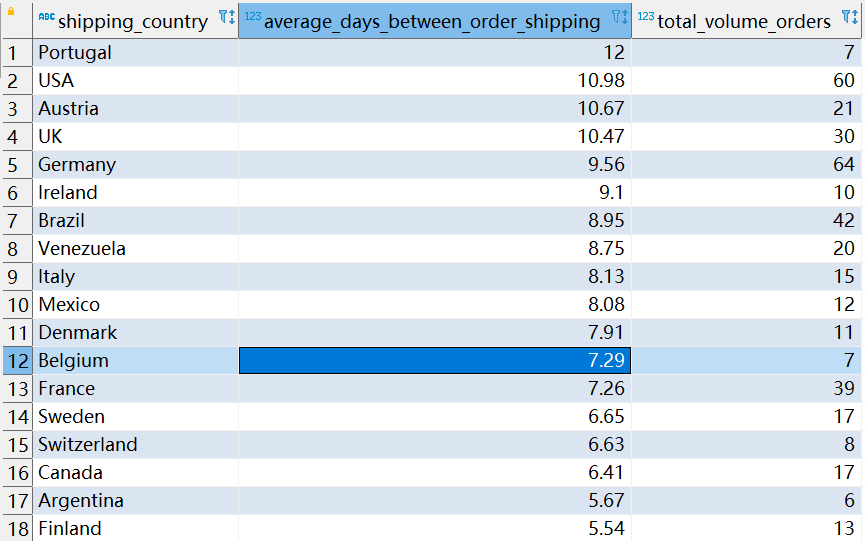


Figure 6. The result of the query

## Question 3

The HR Team wants to know each employee what was age on the date they joined the company and who they currently report to. Provide them with a list of every employee 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 the current date
5. their manager’s full name (first name and last name combined in a single field)
6. their manager’s job title

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

## Answer 3

### The instruction for query

Question 3 only used the table **employees** in a single query**.** The extra step is to combine the rows of employees and the row of the boss - Fuller Andrew as a single table since the boss Fuller Andrew has a missing value in the column *refer\_to.*

### The result of query

Table

Description automatically generated

Figure 7. The result of the query. Tenure calculated by 2022 – the age when hired

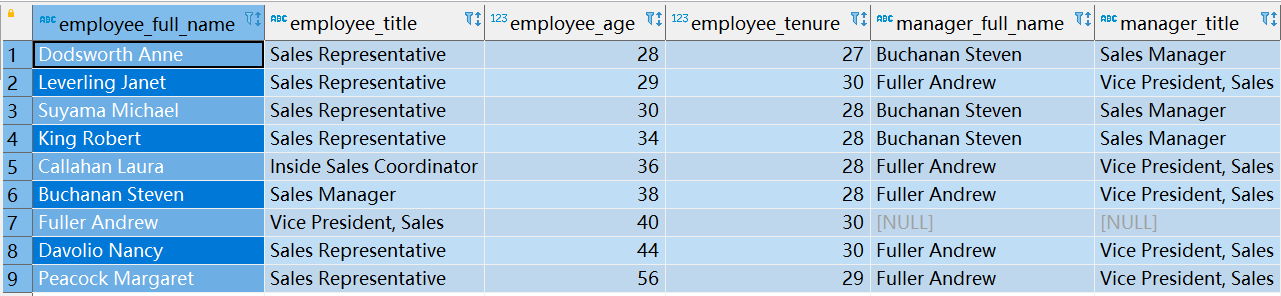


Figure 8. The result of the query. Tenure calculated by (2022-10-01 – hired date) / 365.

Figures 7 and 8 show a subtle difference in “tenure” since given 2 different algorithms.

## 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 a 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 to 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

### The instruction for query

Question 3 only used the table **orders** in a single query.

### The result of query

Table

Description automatically generated

Figure 9. The result of the query.

Chart, line chart

Description automatically generated

Figure 10. Given the filter conditions, the total amount of orders vary in 1996-1997.

Figure 10 shows that, given the total number of orders great than 20 and total freight greater than 2500, the total number of orders is increased steadily before August 1997 and increased dramatically after August 1997.

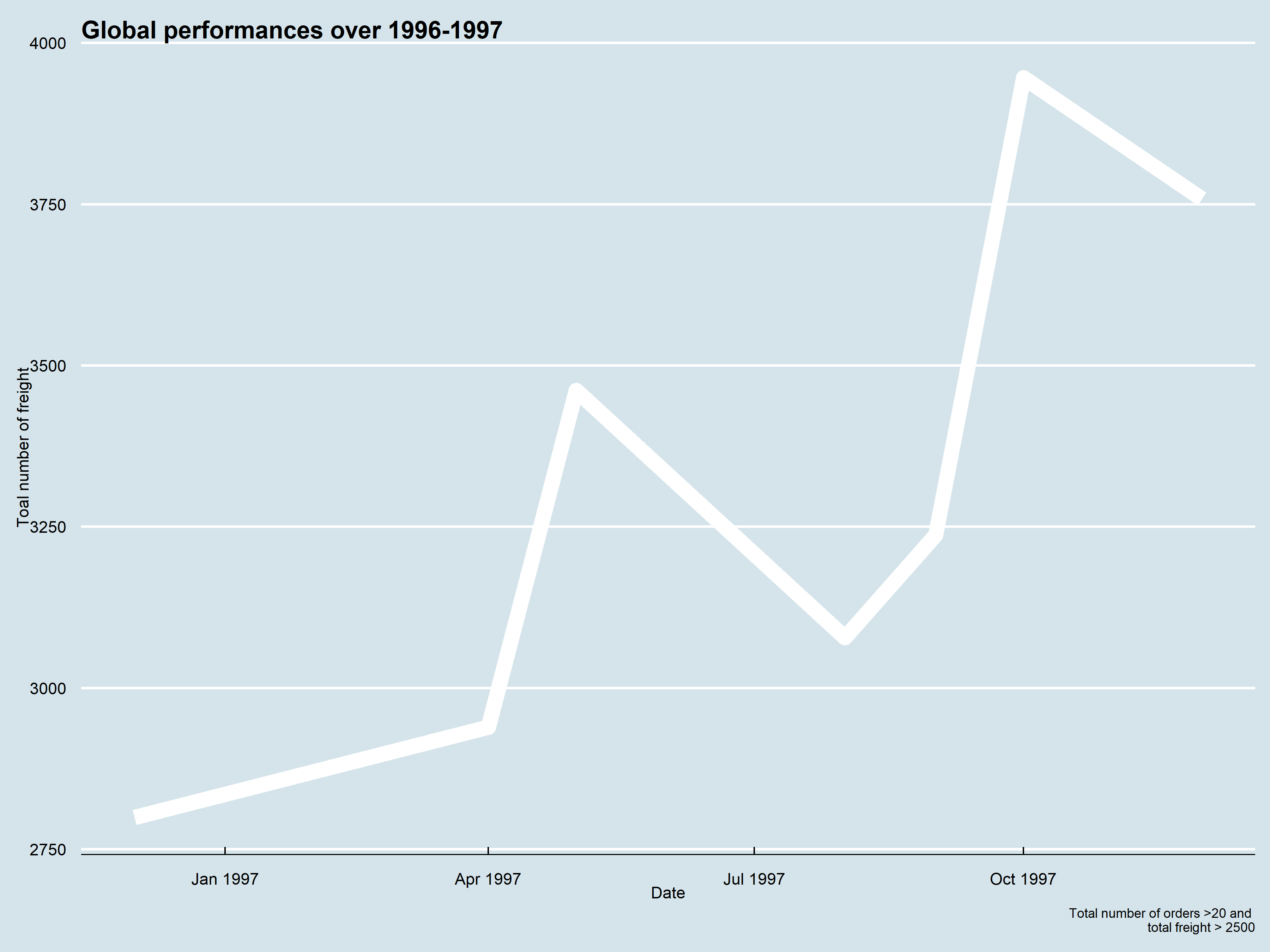


Figure 11. Given the filter conditions, the total amount of freight vary in 1996-1997.

As figure 11 shown, the lowest amount of freight happened in Dec 2016, which is 2,799, and the highest happen in Oct 1997, which is 3,946.

## Question 5

The Pricing Team wants to know which products had a 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

### The instruction for query

To get the result of question 5, 2 steps of preparation were set. The first step is to create 2 additional tables from the table **orders** and **order\_details** and used the function *rownumber()* to order that contain the latest price of each product and the earliest price of each production respectively**.** The second step is to combine those 2 tables together as a single one and calculated the percentage of the increase by two prices.

### The result of query

To get the final result, 2 steps of preparation were set. The first step is to create 2 additional tables that contain the latest price of each product and the earliest price of each production respectively, filter condition by date. The second step is to combine those 2 tables together as a single one and calculated the percentage of the increase by two prices.

Graphical user interface, application, table

Description automatically generated

Figure 12. The result of query given percentage < 10% or > 30%.

## Question 6

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

1. their category name
2. their price range is 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

### The instruction for query

Question 6 used the tables **order\_details, products,** and **categories** by linking their Primary/foreign key in a single query.

### The result of query

Table

Description automatically generated

Figure 13. Result of the query.

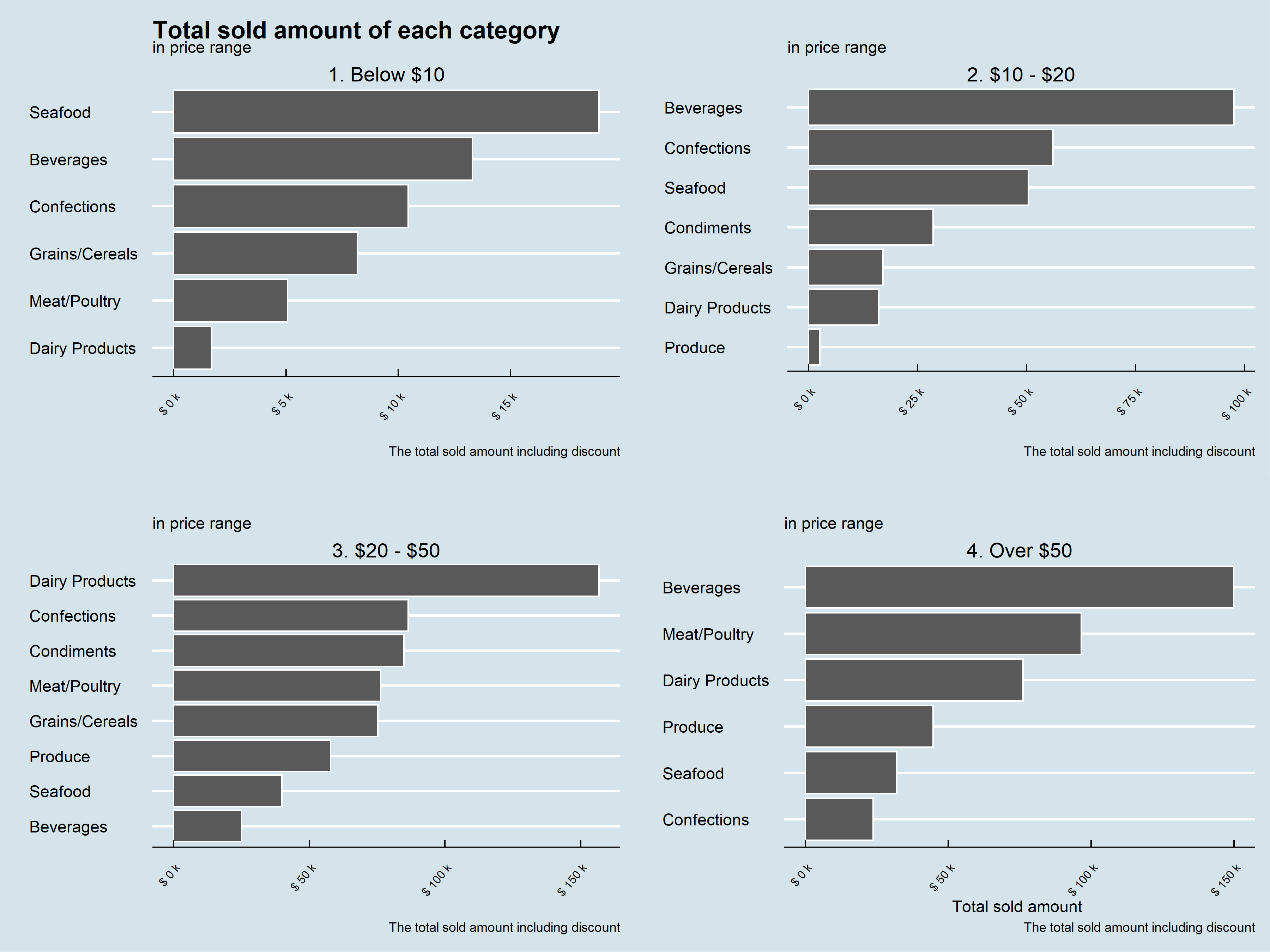


Figure 14. The total sold amount of each category in different price ranges.

Chart, surface chart

Description automatically generated

Figure 15. The total number of orders of each category in different price ranges.

In the interval of the price below $10, the most and least of the total sold amount are Seafood and Diary products respectively, and the most and least of the total number of orders are Seafood and Diary products respectively as well.

In the interval of the price [10,20], the most and least of the total sold amount are Beverages and Produce respectively(except Meat/Poultry, which got $0). The same situation happen in the total number of orders.

In the interval of the price (20,50], the situation of the total sold amount and the total number of orders is a reversion as in the interval of the price below $10. The Diary products got the best of both and the Beverage is the worst one in both.

In the last interval, the price above $50, the most total sold amount is Beverages, and the least one is Confections(except Condiments and Grains/Cereals, which both are $0). The most and least of the total number of orders are Dairy products and Confections.

## 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 with a list of categories:

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

### The instruction for query

Question 7 used the tables **products, suppliers** and **categories** by linking their Primary/foreign key in a single query.

### The result of query

Table

Description automatically generated

Figure 16. The result of query.

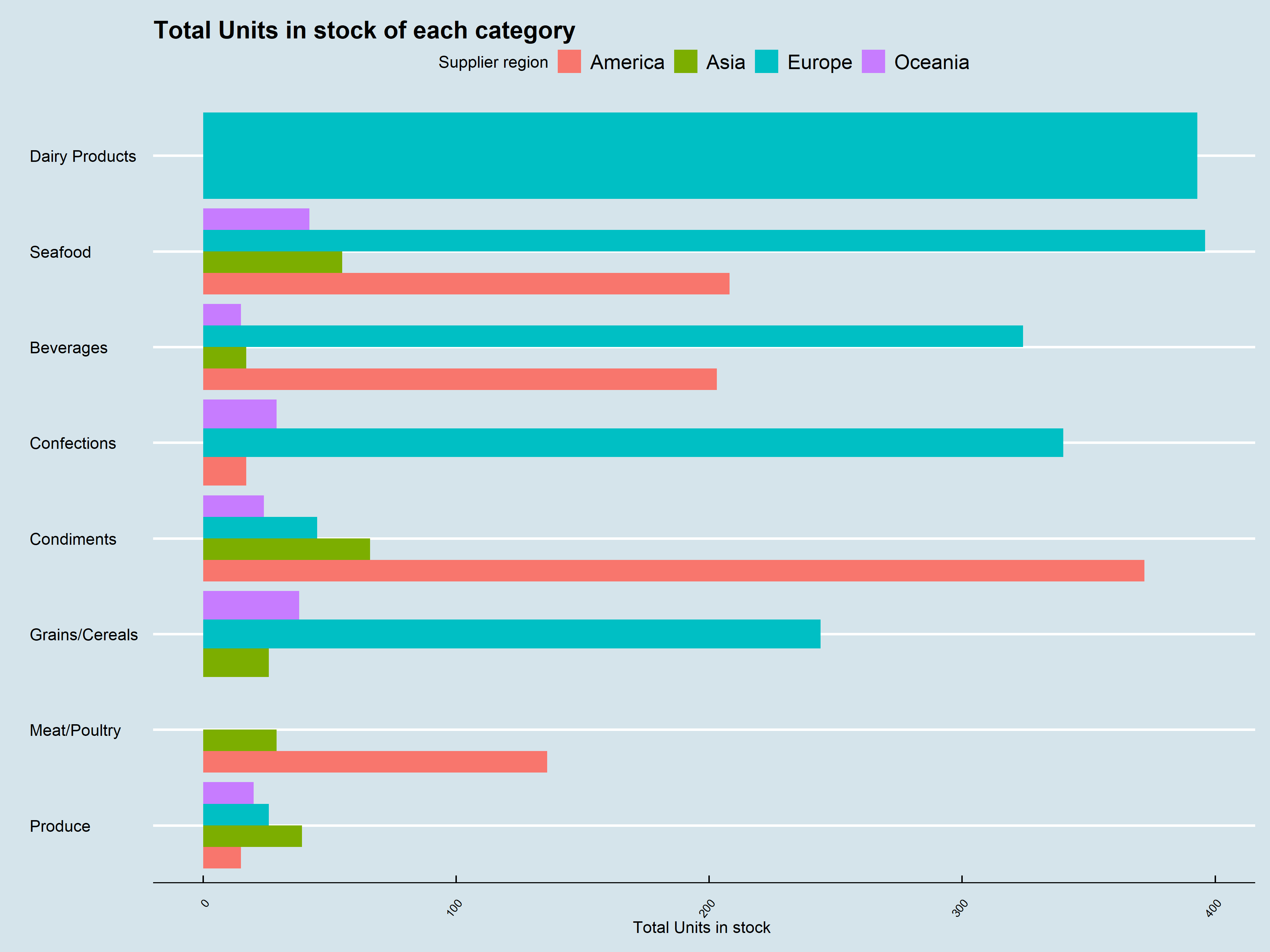


Figure17.

Chart, bar chart

Description automatically generated

Figure 18

Chart, bar chart

Description automatically generated

Figure 19.

As shown in figures 17 to 19, the major supply is from Europe, the second is from America, and the least is from Asia and Oceania.

## Question 8

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

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 and then product name (both ascending).

## Answer 8

### The instruction for query

Question 8 used the tables **products** and **categories** by linking their Primary/foreign key in a single query.

### The result of query

Since is too hard to interpret when only presented as the result of the query as a table, I decided to visualize the difference between the price of the product in separate graphs. Each graph presents the specific category, and the average and the median of the category shows as a *white dash line* and a *blue dash* line respectively.

Chart

Description automatically generated

Figure 20. The products of unit price in Beverage. Côte de Blaye has the higest unit price , which is $263.50, and Laughing Lumberjack Lager has the least unit price, which is $14.

Chart

Description automatically generated

Figure 21. The products of unit price in Condiments. Vegie-spread has the higest unit price, which is $43.9, and Aniseed Syrup has lowest unit pirce, which is $10.

Chart, funnel chart

Description automatically generated

Figure 22. The products of unit price in Confections. Sir Rodney's Marmalade has the highest unit price, which is $81.00,and Teatime Chocolate Biscuits has lowest unit price, which is $9.2.

Chart, funnel chart

Description automatically generated

Figure 23. The products of unit price in Dairy products. Raclette Courdavault has the highest unit price, which is $55, and Geitost has the lowest unit price, which is $2.5.

Chart, funnel chart

Description automatically generated

Figure 24. The products of unit price in Grains/Cereals. Gnocchi di nonna Alice has the highest unit price, which is $38, and Filo Mix has the lowest unit price, which is $7.

Chart

Description automatically generated

Figure 25. The products of unit price in Meat/poultry. The average and median price of Meat/poultry is the same.

Chart

Description automatically generated

Figure 26. The products of unit price in produce. The highest unit price is $53, which is Manjimup Dried Apples , and the lowest unit price is $10, which is Longlife Tofu.

Chart

Description automatically generated

Figure22. The products of unit price in kinds of seafood. Carnarvon Tigers is the highest price of product, the unit price is $62.5, whereas the Konbu is the product has lowest unit price, which is $6.

## 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).

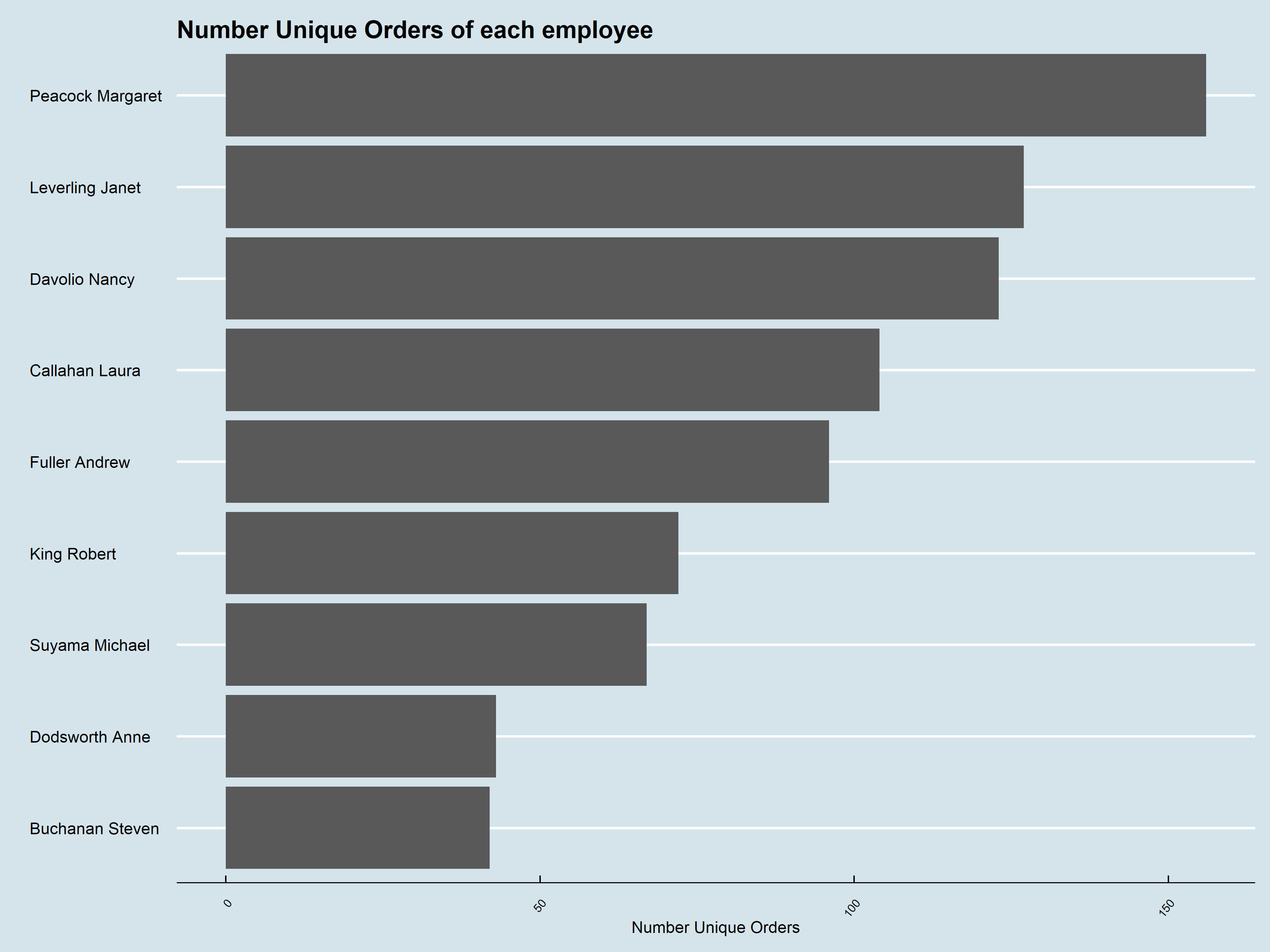
## Answer 9

### The instruction for query

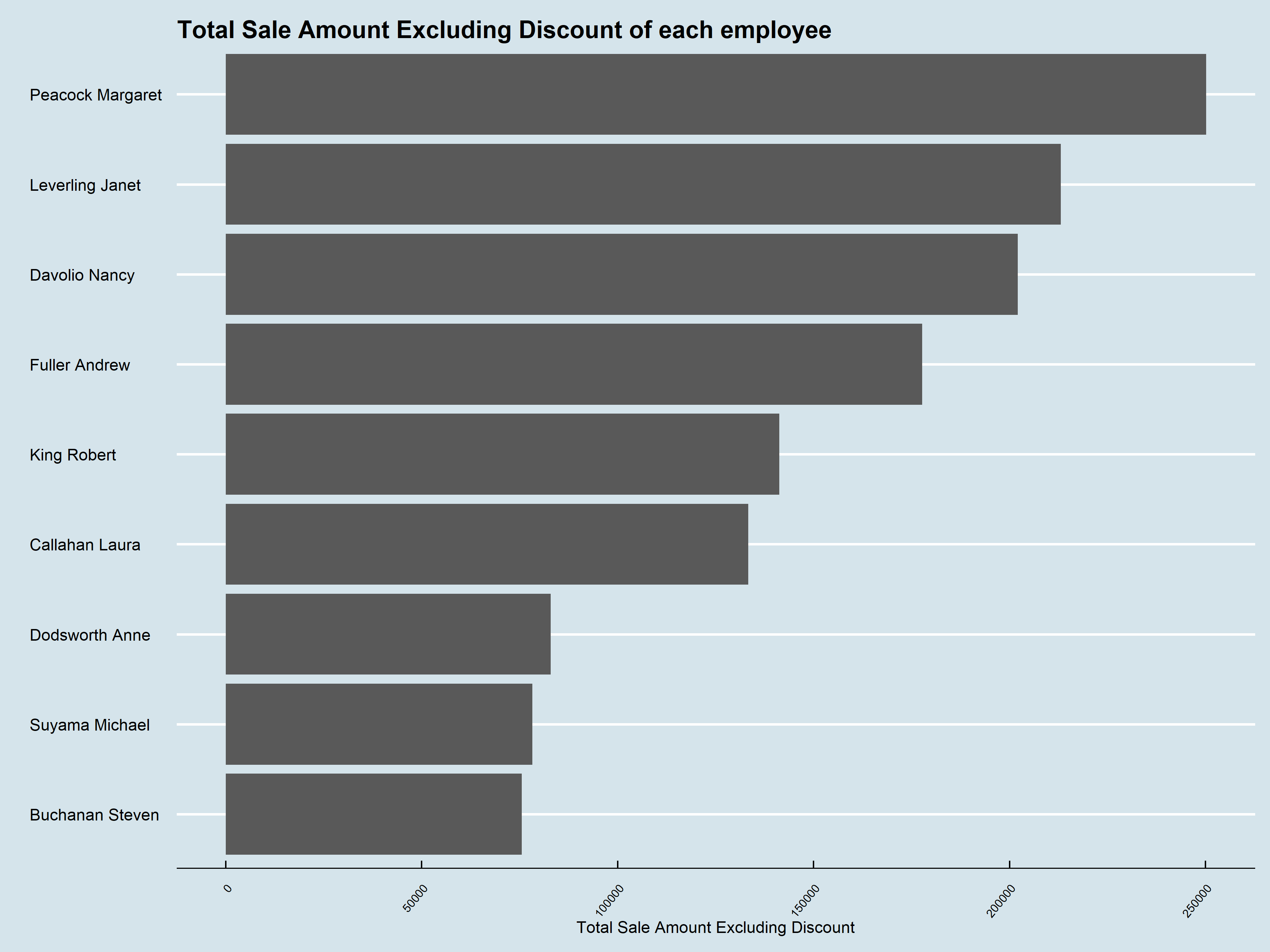
Question 8 used the tables **orders, order\_details** and **employees** by linking their Primary/foreign key in a single query.

### The result of query

Chart, funnel chart

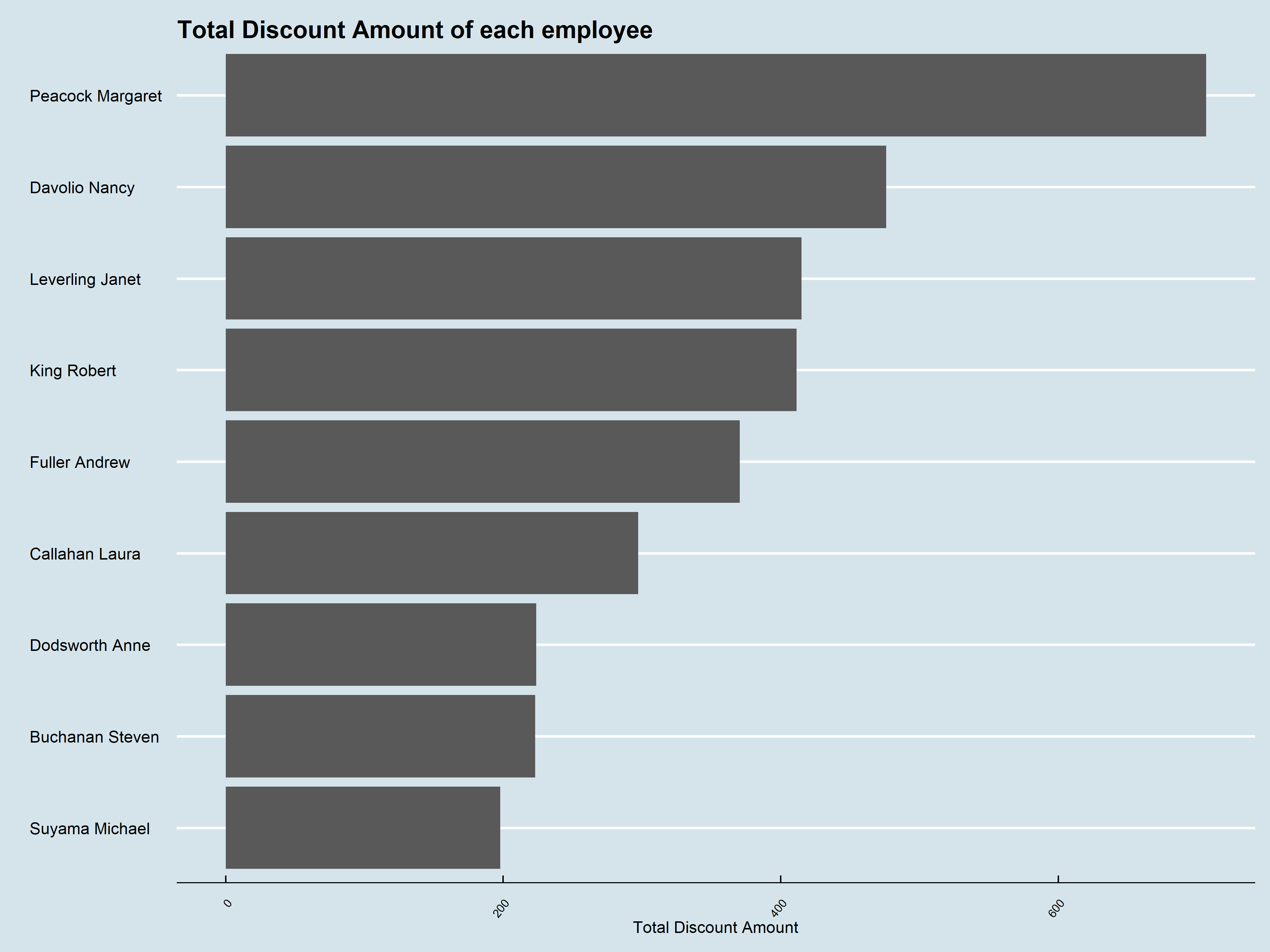
Description automatically generated 

Figures 23 and 24. Peacock Margaret has the most orders whether counting the unique or not. Buchanan Steven has the least number of unique orders.

 Chart, funnel chart

Description automatically generated

Figures 25 and 26. No matter including or excluding the discount is, the order is the same. The most contributor is Peacock Margaret, the least contributor is Buchanan Steven.

 Chart, funnel chart

Description automatically generated

Figures 27 and 28. Buchanan Steven has the most discount percentage versus total sold amount which is 0.3%, whereas he/she is a worst seller in the company

Chart, funnel chart

Description automatically generated

Figure 29. King Robert is the first one of the average order amount, and Suyama Michael is the last one of that.

Chart, funnel chart

Description automatically generated

Figure 30. Peacock Margaret has the best performance no matter the different products selling and the total of the selling.

## 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 with a list of categories and employees:

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 instruction for query

Question 8 used the tables **orders**, **order\_details**, **products**, **employees**, **orders** and **categories** by linking their Primary/foreign key in a single query

### The result of the query

Chart, bar chart

Description automatically generated

Figure 31. Each employee has good performance selling both Beverages and Dairy products, on the other hand, Grains/Cereals is the least portion of selling whatever each employee.

Chart, bar chart

Description automatically generated

Figure 32. Again, the graph shows Peacock Margaret has a well balanced of each category selling, and Dodsworth Anne seems only has well performance on Condiments and Beverages, but poor at Produce and Grains/Cereals.

# Appendix

## The coding of Visualization in R

#Q1

```{r}

setwd("D:/datascience/DSP/2022 Spring/AT2/vis")

library(showtext)

library(tidyverse)

library(extrafont)

library(ggthemes)

library(gridExtra)

Sys.setenv(LANG = "en")

Sys.setlocale("LC\_TIME", "en\_US")

q1 = read.csv("Q1.csv")

q1\_plot <- ggplot(q1,aes(reorder(product\_name,-product\_unit\_price),product\_unit\_price)) +

   geom\_bar(stat="identity",colour="white") +

    labs(x = "Name of product", y = "", title = "Products that are currently being offered",subtitle="price range between $10 and $50") +

   theme\_economist()+

   theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +

   scale\_y\_continuous(labels =function(x) {return(paste("$",x))})

  #geom\_text(aes(label = product\_unit\_price), position = position\_dodge(0.9), vjust = -0.8)

ggsave(q1\_plot,filename = "q1\_plot.png",width = 12,height = 9)

```

#Q2

```{r}

q2 = read.csv("q2.csv")

q2\_plot\_1 <- ggplot(q2,aes(reorder(shipping\_country,-average\_days\_between\_order\_shipping),average\_days\_between\_order\_shipping)) +

   geom\_bar(stat="identity",color="white") +

    labs(x = "Name of Country", y = "Day", title = "Average days between order and shipping of each Country in  1997",caption="Day range in [3,20) and total orders > 5") +

   theme\_economist()+

   theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))

  #geom\_text(aes(label = product\_unit\_price), position = position\_dodge(0.9), vjust = -0.8)

ggsave(q2\_plot\_1,filename = "q2\_plot\_1.png",width = 12,height = 9)

q2\_plot\_2 <- ggplot(q2,aes(reorder(shipping\_country,-total\_volume\_orders),total\_volume\_orders)) +

   geom\_bar(stat="identity",colour="white") +

    labs(x = "Name of Country", y = "Toal orders", title = "Total orders of each country in 1997",caption="Day range in [3,20) and totals order > 5") +

   theme\_economist()+

   theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))

  #geom\_text(aes(label = product\_unit\_price), position = position\_dodge(0.9), vjust = -0.8)

ggsave(q2\_plot\_2,filename = "q2\_plot\_2.png",width = 12,height = 9)

```

#Q4

```{r}

q4 = read.csv("q4.csv")

q4$year\_month =as.Date(q4$year\_month)

q4\_plot\_1 = ggplot(q4,aes(x=year\_month,y=total\_number\_orders)) + geom\_line(colour="white",size=5) +labs(x = "Date", y = "Toal number of orders", title = "Global performances over 1996-1997",caption ="Total number of orders >20 and

total freight > 2500") +

   theme\_economist()

ggsave(q4\_plot\_1,filename = "q4\_plot\_1.png",width = 12,height = 9)

q4\_plot\_2 = ggplot(q4,aes(x=year\_month,y=total\_freight)) + geom\_line(colour="white",size=5) +labs(x = "Date", y = "Toal number of freight", title = "Global performances over 1996-1997",caption ="Total number of orders >20 and

total freight > 2500") +

   theme\_economist()

ggsave(q4\_plot\_2,filename = "q4\_plot\_2.png",width = 12,height = 9)

```

#Q6

```{r}

q6 = read.csv("q6.csv")

q611=ggplot(q6%>%filter(price\_range=="1. Below $10")) + geom\_bar(aes(reorder(category\_name,total\_amount),total\_amount),stat="identity",colour="white") + labs(x = "", y = "", title = "Total sold amount of each category ",subtitle = "in price range" ,caption="The total sold amount including discount") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") +

   scale\_y\_continuous(labels =function(x) {return(paste("$",x/1000,"k"))}) + coord\_flip()

q612 <- ggplot(q6%>%filter(price\_range=="2. $10 - $20")) + geom\_bar(aes(reorder(category\_name,total\_amount),total\_amount),stat="identity",colour="white") + labs(x = "", y = "", title = "",subtitle = "in price range",caption="The total sold amount including discount") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") +

   scale\_y\_continuous(labels =function(x) {return(paste("$",x/1000,"k"))})+ coord\_flip()

q613=ggplot(q6%>%filter(price\_range=="3. $20 - $50")) + geom\_bar(aes(reorder(category\_name,total\_amount),total\_amount),stat="identity",colour="white") + labs(x = "", y = "", title = "",subtitle = "in price range",caption="The total sold amount including discount") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") +

   scale\_y\_continuous(labels =function(x) {return(paste("$",x/1000,"k"))})+ coord\_flip()

q614=ggplot(q6%>%filter(price\_range=="4. Over $50")) + geom\_bar(aes(reorder(category\_name,total\_amount),total\_amount),stat="identity",colour="white") + labs(x = "", y = "Total sold amount ", title = "",subtitle = "in price range",caption="The total sold amount including discount") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") +

   scale\_y\_continuous(labels =function(x) {return(paste("$",x/1000,"k"))})+ coord\_flip()

q61 <-grid.arrange(q611, q612,q613,q614, ncol=2)

ggsave(q61,filename = "q61.png",width = 12,height = 9)

q621=ggplot(q6%>%filter(price\_range=="1. Below $10"),aes(reorder(category\_name,total\_number\_orders),total\_number\_orders)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "", title = "Total number of orders of each category",subtitle = "in price range") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") + coord\_flip()

q622=ggplot(q6%>%filter(price\_range=="2. $10 - $20"),aes(reorder(category\_name,total\_number\_orders),total\_number\_orders)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "", title = "",subtitle = "in price range") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") + coord\_flip()

q623=ggplot(q6%>%filter(price\_range=="3. $20 - $50"),aes(reorder(category\_name,total\_number\_orders),total\_number\_orders)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "", title = "",subtitle = "in price range") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") + coord\_flip()

q624=ggplot(q6%>%filter(price\_range=="4. Over $50"),aes(reorder(category\_name,total\_number\_orders),total\_number\_orders)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total number of orders", title = "",subtitle = "in price range") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) +facet\_wrap("price\_range") + coord\_flip()

q62 <-grid.arrange(q621, q622,q623,q624, ncol=2)

ggsave(q62,filename = "q62.png",width = 12,height = 9)

```

#Q7

```{r}

q7 = read.csv("q7.csv")

q711 = ggplot(q7%>%filter(supplier\_region=="America"),aes(reorder(category\_name,unit\_in\_stock),unit\_in\_stock)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units in stock ",title="Supplier from America", subtitle = "Total Units in stock of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

q712 = ggplot(q7%>%filter(supplier\_region=="America"),aes(reorder(category\_name,unit\_on\_order),unit\_on\_order)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units on order ", subtitle = "Total Units on order of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q713 = ggplot(q7%>%filter(supplier\_region=="America"),aes(reorder(category\_name,reorder\_level),reorder\_level)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Reorder level", subtitle = "Total reorder level of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q71 <-grid.arrange(q711, q712,q713, nrow=3)

q721 = ggplot(q7%>%filter(supplier\_region=="Asia"),aes(reorder(category\_name,unit\_in\_stock),unit\_in\_stock)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units in stock ",title="Supplier from Asia", subtitle = "Total Units in stock of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

q722 = ggplot(q7%>%filter(supplier\_region=="Asia"),aes(reorder(category\_name,unit\_on\_order),unit\_on\_order)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units on order ", subtitle = "Total Units on order of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q723 = ggplot(q7%>%filter(supplier\_region=="Asia"),aes(reorder(category\_name,reorder\_level),reorder\_level)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Reorder level", subtitle = "Total reorder level of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q72 <-grid.arrange(q721, q722,q723, nrow=3)

ggsave(q72,filename = "q72.png",width = 12,height = 9)

q731 = ggplot(q7%>%filter(supplier\_region=="Europe"),aes(reorder(category\_name,unit\_in\_stock),unit\_in\_stock)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units in stock ",title="Supplier from Europe", subtitle = "Total Units in stock of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

q732 = ggplot(q7%>%filter(supplier\_region=="Europe"),aes(reorder(category\_name,unit\_on\_order),unit\_on\_order)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units on order ", subtitle = "Total Units on order of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q733 = ggplot(q7%>%filter(supplier\_region=="Europe"),aes(reorder(category\_name,reorder\_level),reorder\_level)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Reorder level", subtitle = "Total reorder level of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q73 <-grid.arrange(q731, q732,q733, nrow=3)

ggsave(q73,filename = "q73.png",width = 12,height = 9)

q741 = ggplot(q7%>%filter(supplier\_region=="Oceania"),aes(reorder(category\_name,unit\_in\_stock),unit\_in\_stock)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units in stock ",title="Supplier from Oceania", subtitle = "Total Units in stock of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

q742 = ggplot(q7%>%filter(supplier\_region=="Oceania"),aes(reorder(category\_name,unit\_on\_order),unit\_on\_order)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Total Units on order ", subtitle = "Total Units on order of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q743 = ggplot(q7%>%filter(supplier\_region=="Oceania"),aes(reorder(category\_name,reorder\_level),reorder\_level)) + geom\_bar(stat="identity",colour="white") + labs(x = "", y = "Reorder level", subtitle = "Total reorder level of each category") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8)) + coord\_flip()

q74 <-grid.arrange(q741, q742,q743, nrow=3)

ggsave(q74,filename = "q74.png",width = 12,height = 9)

q7\_1 = ggplot(q7,aes(reorder(category\_name,unit\_in\_stock),unit\_in\_stock,fill=supplier\_region)) + geom\_bar(stat="identity",position="dodge") + labs(x = "", y = "Total Units in stock ",title = "Total Units in stock of each category",fill="Supplier region") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

ggsave(q7\_1,filename = "q7\_1.png",width = 12,height = 9)

q7\_2 =ggplot(q7,aes(reorder(category\_name,unit\_on\_order),unit\_on\_order,fill=supplier\_region)) + geom\_bar(stat="identity",position="dodge") + labs(x = "", y = "Total Units on order ",title = "Total Units on order of each category",fill="Supplier region") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

ggsave(q7\_2,filename = "q7\_2.png",width = 12,height = 9)

q7\_3 = ggplot(q7,aes(reorder(category\_name,reorder\_level),reorder\_level,fill=supplier\_region)) + geom\_bar(stat="identity",position="dodge") + labs(x = "", y = "Reorder level ",title = "Total Reorder level of each category",fill="Supplier region") +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

ggsave(q7\_3,filename = "q7\_3.png",width = 12,height = 9)

q7 <-grid.arrange(q7\_1, q7\_2,q7\_3, nrow=3)

ggsave(q7,filename = "q7.png",width = 12,height = 9)

```

#Q8

```{r}

q8 = read.csv("q8.csv")

q8\_plot = function(x){

  mean = mean(q8$average\_unit\_price[q8$category\_name==x])

  median = mean(q8$median\_unit\_price[q8$category\_name==x])

p<-ggplot(q8%>%filter(category\_name==x),aes(reorder(product\_name,unit\_price),unit\_price)) + geom\_bar(stat="identity",position="dodge") +

  labs(x = "", y = "Unit price",title = "Unit price of currently offered product",subtitle = paste("In",x),caption="Blue dash line = median \n white dash line = mean") +

   theme\_economist() +

  theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  +

  coord\_flip() +scale\_y\_continuous(labels=function(x){return(paste("$",x))})+

  geom\_hline(yintercept = mean,

             colour = 'white', size = 1, linetype = 'dashed')  +geom\_hline(yintercept = median,

             colour = 'Sky blue', size = 1, linetype = 'dashed') +geom\_text(aes(x = 1, y = median,

                stat = "unique",

                label = paste(median)),color="black")+geom\_text(aes(x = 1, y = mean,

                stat = "unique",

                label = paste(mean)),color="black")

ggsave(p,filename = paste("q7\_",str\_replace(x,"/"," "),".png"),width = 12,height = 9)

}

q8\_plot("Beverages")

q8\_plot("Condiments")

q8\_plot("Confections")

q8\_plot("Dairy Products")

q8\_plot("Grains/Cereals")

q8\_plot("Meat/Poultry")

q8\_plot("Produce")

q8\_plot("Seafood")

```

#Q9

```{r}

q9=read.csv("q9.csv")

q9$employee\_full\_name <- as.factor(q9$employee\_full\_name)

for (i in 3:10){

  string=colnames(q9)[i]

  string=str\_replace\_all(string,"\_"," ")

  string=str\_to\_title(string)

p=ggplot(q9,aes(reorder(employee\_full\_name,q9[,i]),q9[,i])) + geom\_bar(stat="identity",position="dodge") + labs(x = "", y = string,title = paste(string,"of each employee")) +

   theme\_economist() + theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  + coord\_flip()

ggsave(p,filename = paste("q9\_",string,".png"),width = 12,height = 9)

}

```

#Q10

```{r}

q10 = read.csv("q10.csv")

# q10\_plot = function(x){

#   mean = mean(q8$average\_unit\_price[q8$category\_name==x])

#   median = mean(q8$median\_unit\_price[q8$category\_name==x])

# p<-ggplot(q10,aes(reorder(employee\_full\_name,-total\_sale\_amount),total\_sale\_amount)) + geom\_bar(stat="identity",position="dodge") +

#   labs(x = "", y = "Unit price",title = "Total sold amount each employee",caption="Blue dash line = median \n white dash line = mean") +

#    theme\_economist() +

#   theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8))  +

#   coord\_flip() +scale\_y\_continuous(labels=function(x){return(paste("$",x/1000,"k"))})

# ggsave(p,filename = paste("q7\_",str\_replace(x,"/"," "),".png"),width = 12,height = 9)

#

# }

ggplot(q10) + geom\_bar(aes(employee\_full\_name,percent\_of\_employee\_sales, fill= category\_name),stat="identity",position="stack") +

  labs(x = "", y = "",title = "Sells contrition in different Category of each employee",caption="",fill="category name") +

   theme\_economist() +

  theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8),panel.grid = element\_blank()) +

  coord\_flip() +scale\_y\_continuous(labels=function(x){return(paste(x\*100,"%"))})

ggplot(q10) + geom\_bar(aes(x=category\_name,y=percent\_of\_category\_sales, fill= employee\_full\_name),stat="identity",position="stack") +

  labs(x = "", y = "",title = "Sells contrition of different employee in each category",caption="",fill="Emplyee name") +

   theme\_economist() +

  theme(axis.text.x = element\_text(angle = 50, vjust = 1, hjust = 1, size = 8),panel.grid = element\_blank()) +

  coord\_flip() +scale\_y\_continuous(labels=function(x){return(paste(x\*100,"%"))})

#+ coord\_polar(theta = 'y')

#geom\_text(aes(label = percent\_of\_employee\_sales),

            #vjust = -0.25) #+facet\_wrap(~category\_name)

#+ coord\_polar(theta = 'y')