**CIS 467 final group project (due by Thursday, October 26 at 11:59 PM).**

**This is a group project (total 300 points). The groups have been created on Blackboard. Please make only one submission per group and put all your team members’ full names into this Word document. Please also submit a Tableau Workbook file .twb into the Final Team Project folder on Blackboard together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations.**

The script file final\_project\_database\_cis467.sql creates a database which contains the 11 tables (the database schema is below), with transactional data related to some company operations.

**Please check early that you can create the final project database on your machine.**

**Please review Panopto video for week 7 for a description of this dataset.**

**Very Important! All three parts of the final project should be on One topic/subject of the data warehouse. For example, if you decide to track customers as your topic/subject, part 1 (Data Warehouse), part 2 (Queries) and part 3 (visualizations) should only be related to customers and should NOT include any other topics.**

**If you use Chat GPT, please use the “Share” button (looks like ‘upward arrow’) in the right corner of ChatGPT chat, and ‘copy link’ and share the link to that chat in this Word document and briefly explain how you used it for your Final Project (for each question if you used it). No points off will be taken for using ChatGPT (it is allowed to use it for Final Project) but you are required to share the link to a chat if you used it.**

**More information on how to share a chat here:** <https://help.openai.com/en/articles/7925741-chatgpt-shared-links-faq>

Timeline

Description automatically generated

Please put all of your work into **this** **single Word doc and also submit a Tableau Workbook file .twb and the Excel file of your Data Warehouse that you used for Tableau visualizations**. Please see instructions for Tableau below in the question 3.

1. **(121 points)** Design and create a data warehouse for the provided database. The decisions about which fields to include and how to aggregate the data are left to you. You do not need to include every single data point from the 11 tables given. Use your judgement as to what will be interesting/useful for the organization. But please make sure that you pull (combine) data from **at least six tables** and compute relevant aggregate statistics. Please compute relevant aggregate statistics for each table that you join. In your queries later in part 2, you may join your Data Warehouse with other tables to answer useful questions. Please see many examples from class lectures and you may adapt those codes for your purpose (for this dataset).

**Submit a screenshot of the first 25 rows of your data warehouse (paste into this Word document) and the SQL code that you used to create it. Please copy and paste your SQL code into this Word document. If your PC does not show 25 rows of data, please submit what you have (i.e., rows you can see on a screenshot) with a comment that you cannot show 25 rows of data. Please add a description of what your Data Warehouse will be tracking for a company.**

USE cis467\_final\_project;

CREATE OR REPLACE VIEW sales\_dw AS

SELECT

ROW\_NUMBER() OVER () AS RowNumber,

ep.EmployeeID,

TotalOrders.OrderID, TotalOrders.CustomerID, TotalOrders.UnitPrice AS SalePrice,

TotalOrders.Quantity,TotalOrders.Discount,TotalOrders.Region,TotalOrders.Country,TotalOrders.OrderDate,

TotalOrders.ShippedDate,PC.ProductID,PC.ProductName,PC.CategoryID,PC.CategoryName, PC.UnitsInStock

FROM

(SELECT

e.EmployeeID,

emp.TerritoryID,

t.RegionID,t.TerritoryDescription,

r.RegionDescription

FROM Employees e

JOIN Employeeterritories emp

ON e.EmployeeID = emp.EmployeeID

JOIN Territories t

ON emp.TerritoryID = t.TerritoryID

JOIN Region r

ON t.RegionID = r.RegionID

) AS ep

JOIN EmployeeTerritories et

ON et.EmployeeID = ep.EmployeeID

JOIN

(SELECT

o.OrderID, o.CustomerID, o.EmployeeID, o.OrderDate,o.ShippedDate,

od.UnitPrice,od.Quantity,od.Discount,od.ProductID,

cm.Region,cm.Country

FROM Orders o

JOIN Order\_Details od

ON o.OrderID = od.OrderID

JOIN Customers cm

ON o.CustomerID = cm.CustomerID

) AS TotalOrders ON ep.EmployeeID = TotalOrders.EmployeeID

JOIN

(SELECT

p.ProductID,p.ProductName,p.UnitsInStock,

c.CategoryID,c.CategoryName

FROM Products p

JOIN Categories c

ON p.CategoryID = c.CategoryID

) AS PC ON TotalOrders.ProductID = PC.ProductID

GROUP BY

ep.EmployeeID,

TotalOrders.OrderID, TotalOrders.CustomerID, TotalOrders.UnitPrice,TotalOrders.Quantity,

TotalOrders.Discount,TotalOrders.Region,TotalOrders.Country,TotalOrders.OrderDate,

PC.ProductID,PC.ProductName,PC.CategoryID,PC.CategoryName,PC.UnitsInStock;



2. **(104 points)** Create **eight** SQL queries **on your data warehouse** (not on the original dataset) that answer interesting questions. At least **6** queries should be more complex queries. For example, more complex queries could include Joins, a Group By, UNION elements or a subquery or use some aggregate functions and summary calculations (see examples in the class lectures’ slides).

**Submit a copy of each query SQL code (paste into this Word document), and the screenshot of each query results (or the first 25 rows if it is longer or how many rows you can get on your PC) and full description of the question your SQL code was addressing and what you found in the results. The question that each query answers should be useful for a company to make decisions and act upon.**

**Question1: Find the yearly change of sales for Ikura and Boston Crab Meat during the years mentioned, and analyze the information behind the data.**

**Code:**

SELECT

ProductID, ProductName, CategoryName,

YEAR(STR\_TO\_DATE(OrderDate,"%m/%d/%Y")) AS FixedYear,

ROUND(SUM(SalePrice\*Quantity\*(1-Discount)),2) AS Sales

FROM sales\_dw

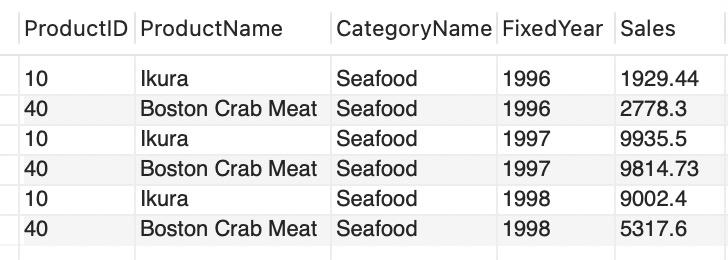
WHERE STR\_TO\_DATE(OrderDate,"%m/%d/%Y") Between '1996-01-01' And '1998-12-31'

AND ProductName = "Ikura" OR ProductName = "Boston Crab Meat"

GROUP BY FixedYear, ProductID, ProductName, CategoryName

ORDER BY FixedYear, ProductID;

**Answer:**



**Introduction：**

In this analysis, we aim to assess the year-over-year (YoY) sales change of two seafood brands Ikura and Boston Crab Meat from 1996 to 1998. This analysis is vital for understanding the performance of our products and identifying trends in sales of the brands.

We obtained the necessary data from our DateWarehouse sales\_dw, which includes records of sales revenues and timestamps. This data is crucial for comparing sales between different time periods.

**Analysis by year:**

* 1996: The sale of Ikura was 1929.44 and the sale of Boston Crab Meat was 2778.3. Boston Crab Meat is a little bit higher, but did not differ a lot.
* 1997: The sale of Ikura was 9935.5 and the sale of Boston Crab Meat was 9814.73. Boston Crab Meat is a little bit higher, but did not differ a lot. This year, both of them achieved a steep leap in sales; it may be due to a great chance of the whole seafood market.
* 1998: The sale of Ikura was 9002.4 and the sale of Boston Crab Meat was 5317.6. Ikura did a much better job this time. The bonus of the market may have reduced so the sales also decreased a little bit but Ikura in total kept the advantage. Boston Crab Meat did a better job than 1996 but may face some problems of its own, the decrease in its sales was also rapid.

**Overall Insights:**

* + From 1996 to 1998, the entire tendency of sales for the brands was increasing.
  + Ikura has a better sustainability for further development due to these years’ data. Our company can consider buying more products from it.

**Question2: Where is the most potential region in the USA for Chai?**

**Code:**

SELECT

ProductID, ProductName, Region, Country,

AVG(SalePrice) AS avg\_price,

AVG(Discount) AS avg\_discount,

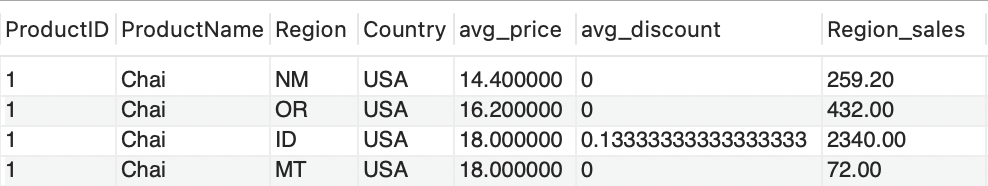
SUM(Quantity) AS Region\_sales

FROM sales\_dw

WHERE Country = "USA" AND ProductName = "Chai"

GROUP BY ProductID, ProductName, Region, Country;

**Answer:**

****

In this analysis, we explored the most promising regions for Chai within the United States. We assessed these regions based on the average sales price, average discount, and the total quantity of products purchased from DataWarehouse sales\_dw. This analysis will help identify regions with potential growth opportunities.

**Analysis and Insights:**

* + The ID and MT region have the highest average prices, both are 18.
  + The ID region has nearly a 13.3% discount, while other regions have no discount.
  + The ID region has the highest sales, 2340.00, while the MT region only had 72 with the same price.

**Conclusion:**

* + The ID region seems the most potential one for Chai, we should mainly selling Chai here.
  + Compared with the ID region, other regions did not have an obvious competitive advantage, so we only need to maintain the current condition and properly decrease the price in the MT region. If it is necessary, we should quit from MT.

**Question3: What is the best seller of our company?**

**Code:**

SELECT

ProductID,

ProductName,

ROUND(SUM(SalePrice\*Quantity\*(1-Discount)),2) AS sales\_revenue

FROM sales\_dw

GROUP BY ProductID

ORDER BY sales\_revenue DESC

LIMIT 10;

**Answer:**



**Introduction:**

In this analysis, we aim to find the top 10 best sellers of the company.

**Datasource：**

Retrieve data from our DataWarehouse that includes information on product names,

**Findings:**

* + From the result we can see that our best seller is the Ce de Blaye with a sales volume of 141396.73, which is almost twice the volume of the second one.

**Question4: What are the top five customers that purchase the most?**

**Code:**

SELECT

CustomerID,

ROUND(SUM(SalePrice\*Quantity\*(1-Discount)),2) AS sales\_revenue

FROM

sales\_dw

GROUP BY

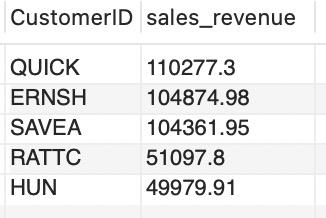
CustomerID

ORDER BY

sales\_revenue DESC

LIMIT 5;

**Answer:**



**Introduction:**

In this analysis, we will find the customers that contribute the most sales volume of our company. So we

**Datasource：**

Retrieve data from DataWarehouse, including information on sales volume and customer id.

**Findings:**

* + from the result we can see that QUICK, ERNSH, SAVEA are our top 3 customers that generate around 100000 dollars sales volume. They are the most valuable customers in our company.

**Question 5: What are the top 3 employees who generate the highest sales?**

**Code:**

SELECT

e.EmployeeID,

s.LastName,

s.FirstName,

ROUND(SUM(SalePrice\*Quantity\*(1-Discount)),2) AS sales\_revenue

FROM

Employees s

join sales\_dw e

on s.EmployeeID = e.EmployeeID

GROUP BY

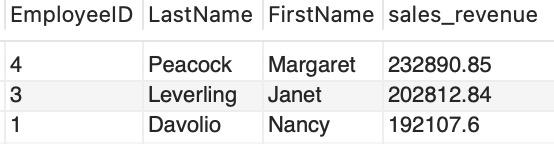
EmployeeID

ORDER BY

sales\_revenue DESC

LIMIT 3;

**Answer:**



**Introduction:**

In this analysis, we will determine the top 3 employees who have generated the highest sales revenue. This analysis helps recognize the most effective salespeople within the organization.

**Datasource：**

Retrieve data from DataWarehouse, including employee information and sales revenue.

**Analysis and Insights:**

* + From the result we can see the top 3 most effective employees we have are Margaret, Janet and Nancy.

**Question6: How long does it take for the company to ship the product?**

**Code:**

SELECT

Country,

AVG (datediff(STR\_TO\_DATE(ShippedDate,"%m/%d/%Y"),STR\_TO\_DATE(OrderDate,"%m/%d/%Y"))) AS avg\_shiptime

FROM

sales\_dw

GROUP BY

Country

ORDER BY

avg\_shiptime ASC;

**Answer:**



**Introduction:**

In this analysis, we will calculate the average shipping time for each company region to identify the most efficient region in terms of delivery times. This analysis will help us understand which regions perform well in terms of order fulfillment.

**Analysis and Insights:**

* + Present the region with the lowest average shipping time, indicating the highest efficiency in order fulfillment. So, we can see that Finland is the most efficient country in terms of shipping products, and the Ireland takes longest time to ship products.

**Question7: Who was the best seller in 1997 due to the sale quantity?**

ChatGPT: Help us achieve how to select the best seller by using 'WITH AS'

LINK:<https://chat.openai.com/share/1b50c9fe-c6d4-4716-b603-552e6420a92b>

**Code:**

WITH QuarterlySales AS (

SELECT

ProductID,ProductName,CategoryName,

YEAR(STR\_TO\_DATE(OrderDate, '%m/%d/%Y')) AS FixedYear,

QUARTER(STR\_TO\_DATE(OrderDate, '%m/%d/%Y')) AS FixedQuarter,

SUM(Quantity) AS total\_sales,

ROW\_NUMBER() OVER (PARTITION BY QUARTER(STR\_TO\_DATE(OrderDate, '%m/%d/%Y')) ORDER BY SUM(Quantity) DESC) AS RowNum

FROM sales\_dw

WHERE YEAR(STR\_TO\_DATE(OrderDate, '%m/%d/%Y')) = 1997

GROUP BY FixedYear,FixedQuarter,ProductID, ProductName, CategoryName

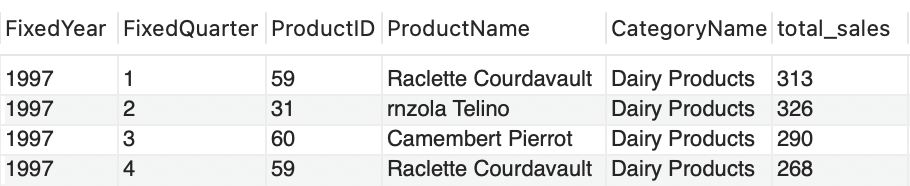
)

SELECT FixedYear,FixedQuarter,ProductID,ProductName,CategoryName,total\_sales

FROM QuarterlySales

WHERE RowNum = 1;

**Answer:**



**Introduction:**

In this analysis, we will determine the product with the highest sales revenue in each of the four quarters of 1997. We retrieve data from DateWarehouse sales\_dw, including information on product sales, products names, timestamps, and products categories.

**Analysis and Insights:**

* + Though the bestsellers were not all from the same brand, they were in the same category.
  + The total sale increased a little from quarter 1 to 2, then decreased from 2 to 4.

**Conclusion:**

* + As an important daily product, the Dairy product was the most popular category in 1997.
  + Raclette Courdavault did a great job in 1997, we can consider buying more products from it.

**Question 8: Find TOP 10 products with the largest price fluctuations,**

**Code:**

SELECT

ProductID, ProductName, CategoryName,

MIN(SalePrice \* (1 - Discount)) AS LowestPrice,

MAX(SalePrice \* (1 - Discount)) AS HighestPrice,

MAX(SalePrice \* (1 - Discount)) - MIN(SalePrice \* (1 - Discount)) AS PriceFlu

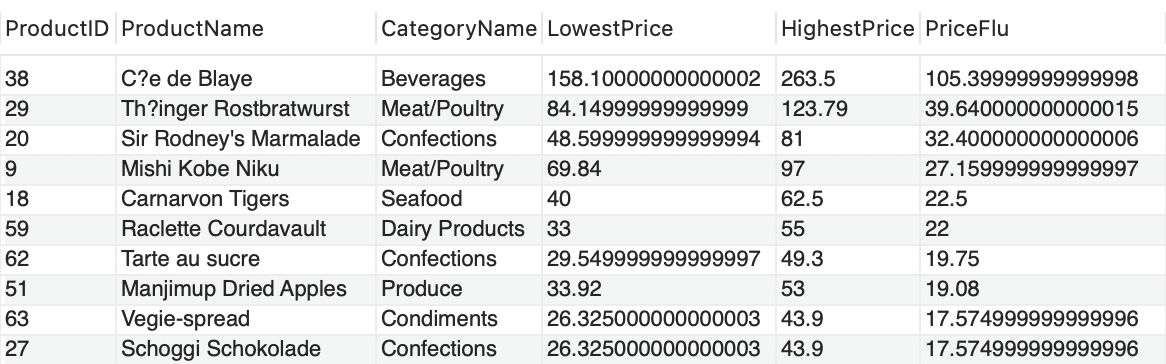
FROM sales\_dw

GROUP BY ProductID, ProductName, CategoryName

ORDER BY PriceFlu DESC

LIMIT 10;

**Answer:**



**Introduction:**

In this analysis, we will determine the top 10 products with the largest price fluctuation from 1996 to 1998. We retrieve data from DateWarehouse sales\_dw, including information on product sales, products names, timestamps, and products’ price fluctuation.

**Analysis and Insights:**

* + Ce de Blaye had the largest price fluctuation, exceeding other brands a lot.
  + Among the 10 products, the confections appeared 3 times and it was the highest frequency.

**Conclusion:**

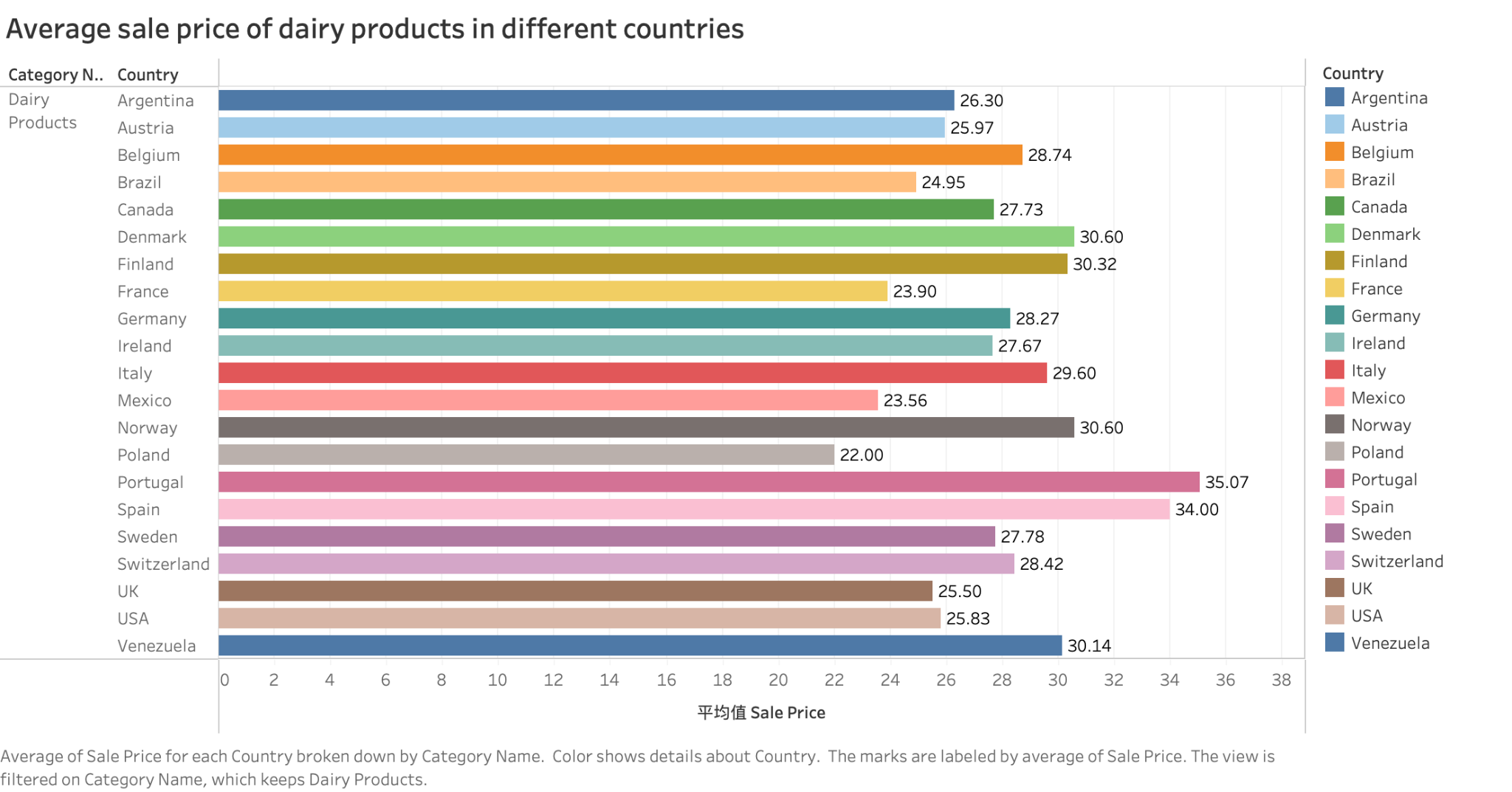
* + During 1996 to 1998, Ce de Blaye had a high price elasticity. According to its high price compared with other brands here, it could be considered as a kind of luxury during 1996 to 1998, we should not have too many stocks of it.
  + The confections had an obvious price elasticity. Though it was not very expensive, it was still unnecessary. If we do not have enough revenue, we can reduce the cost on them.

3. (**75 points**) Create **five** Tableau individual visualizations (graphs) **on your data warehouse** with valuable information to present findings to senior management of the company. Save each visualization as a png file (as we will practice in the lab 5) and paste each individual visualization png file **into this Word** document with the full explanation of what the visualizations show, how they are useful to a company and how company management could make decisions based on what you show. Finally, combine those **five** visualizations into one **Dashboard** (as we will practice in the lab 5), and save this Dashboard as a png file and **paste the Dashboard into this Word** document.

**Please also save the whole Tableau project as a Tableau Workbook file .twb (In Tableau use File - Save as) and submit to the Final Team Project folder on Blackboard together with this Word document and together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations.**

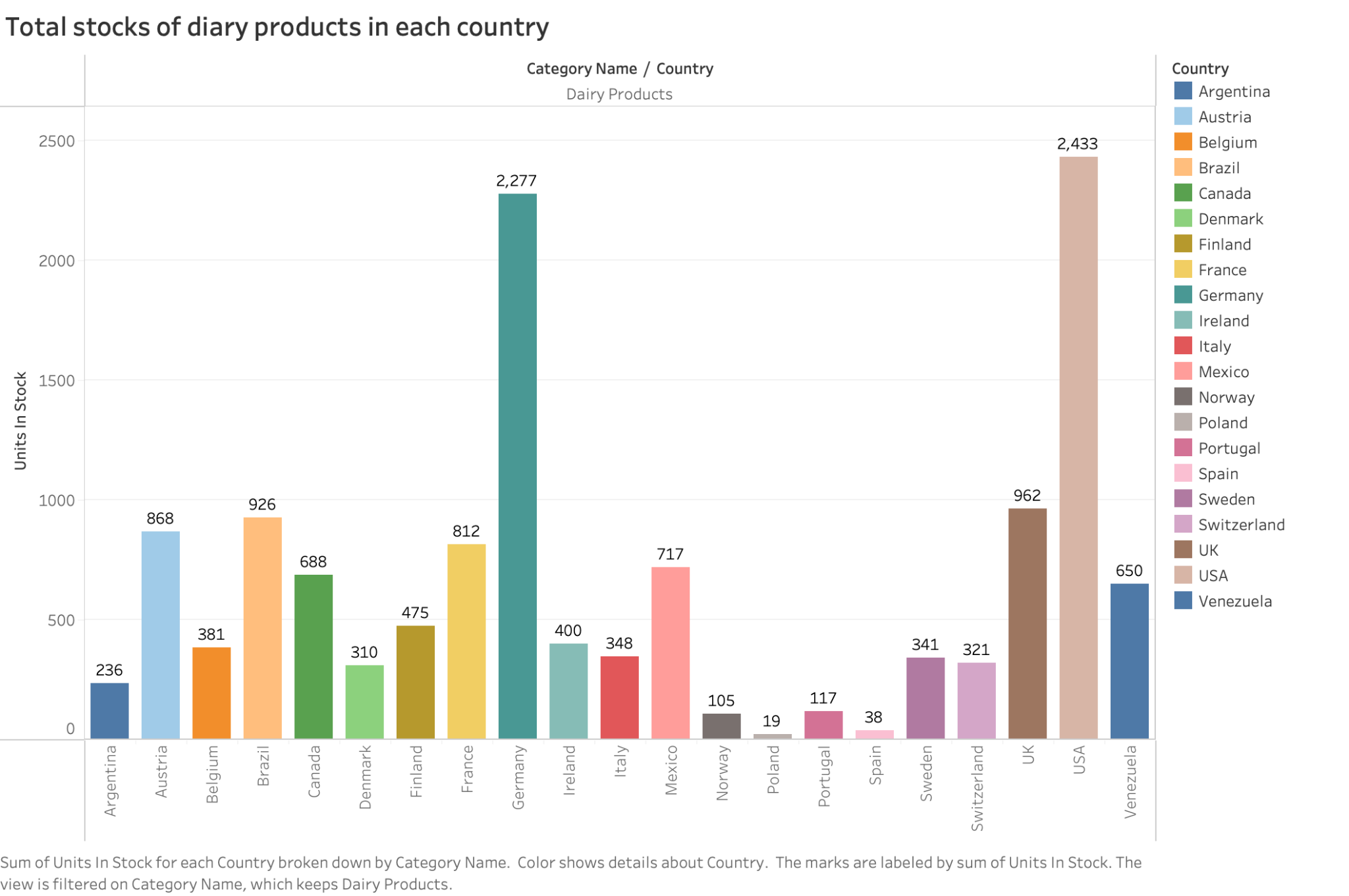
**General grading criteria: Your completed work will be evaluated using the criteria below. I encourage you to use your creativity and other business skills (communication, presentation, critical thinking) in addition to the data management concepts and the SQL and Tableau skills that we have covered in CIS467.**

| **High score** | **Score between high and good** | **Good/medium score** | **Low score** |
| --- | --- | --- | --- |
| All required parts of the final project are complete and technically correct. Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries. Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions. | All required parts of the final project are complete and technically correct (with possibly a few minor errors). Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries (with possibly a few minor errors). Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions (with possibly a few minor errors). | Some required parts of the final project are missing and/or there are more significant errors. Some queries appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple but may still provide interesting useful information based on which senior management of the company can make important decisions. | The final project has large portions missing and/or major conceptual errors. Most/all queries (if any) appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple and **do not** provide interesting useful information based on which senior management of the company can make important decisions. |

**Visualization#1**

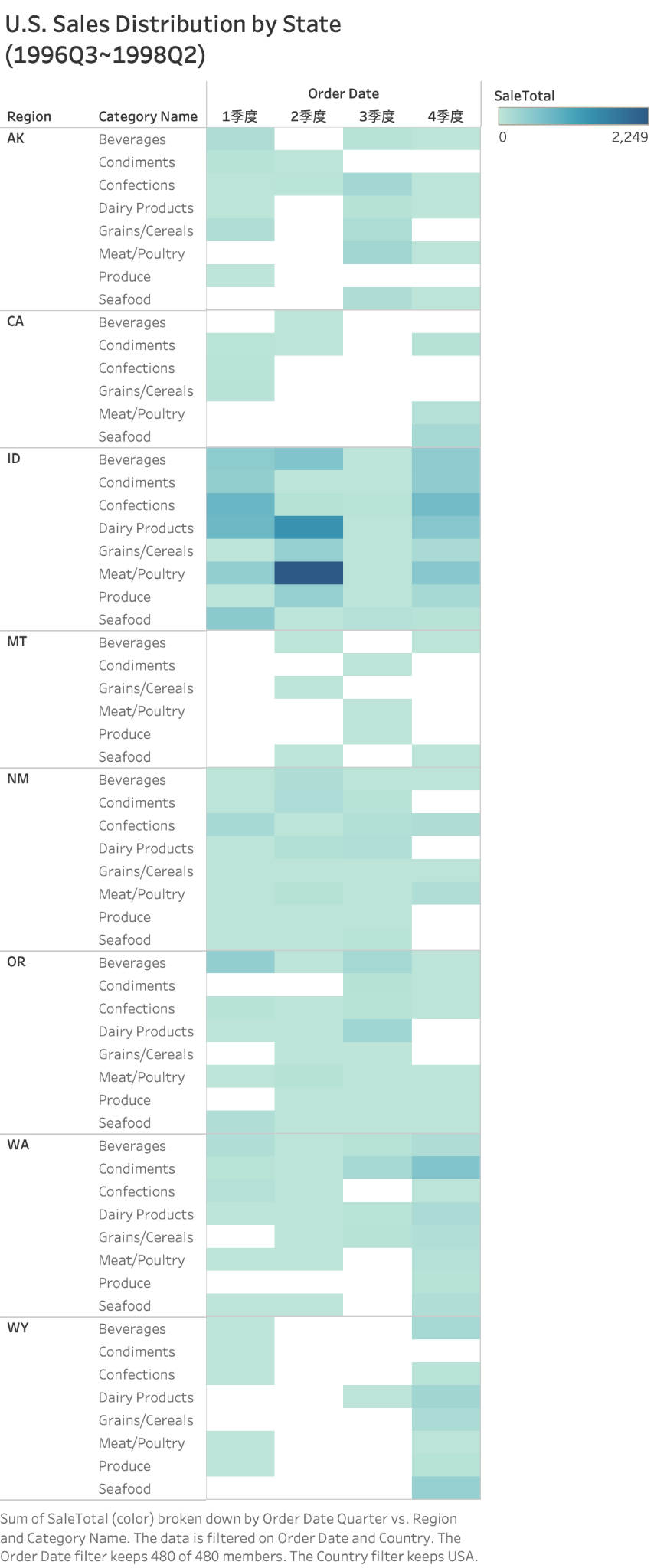
* This table shows the average selling price of dairy products in each country, which helps the company better understand the price positioning in each country and helps executives develop a rational pricing strategy when launching a new product. Based on this graph, the company should set a higher price on dairy products in Portugal and a lower price in Poland.

**Visualization#2**



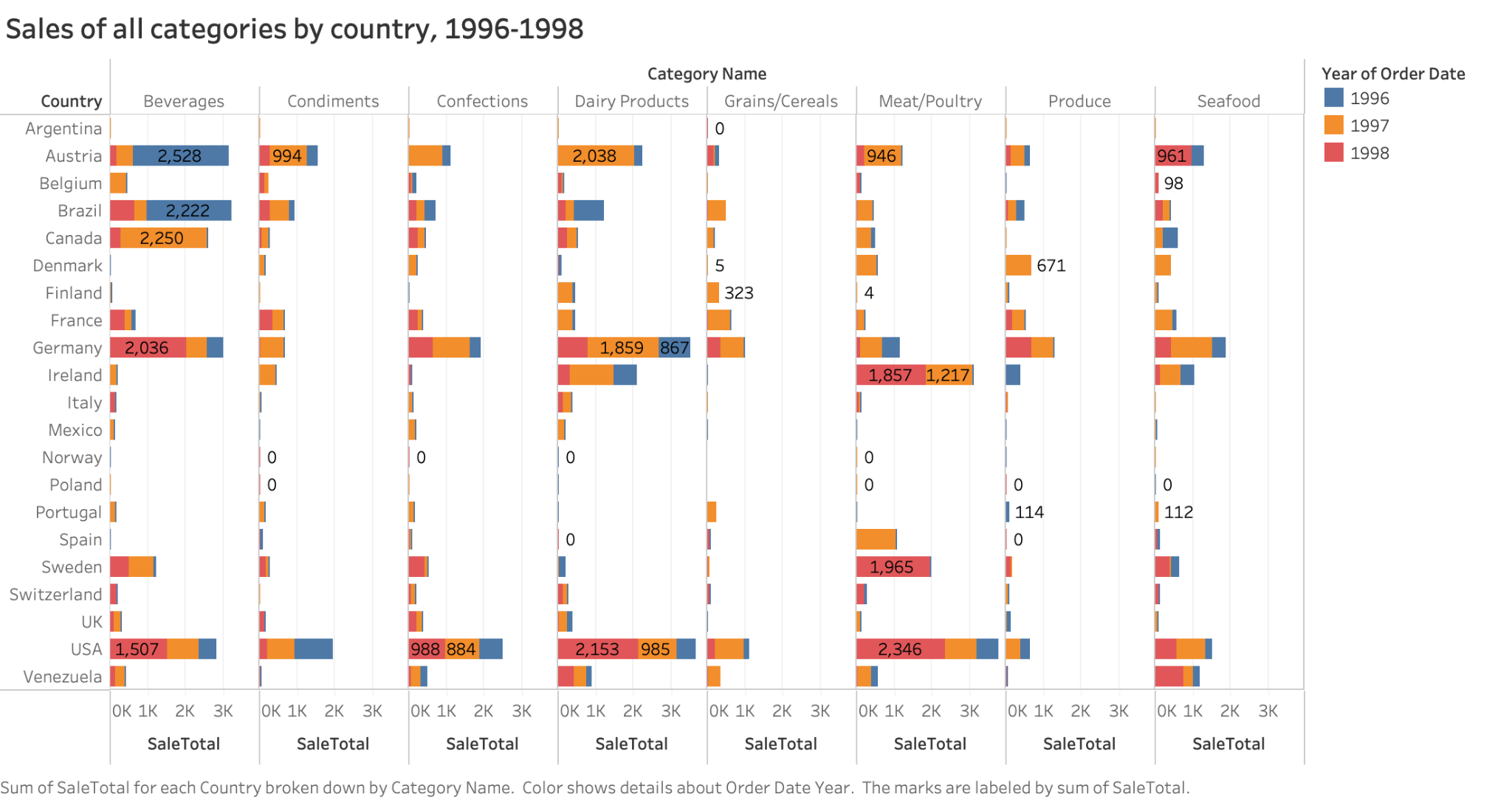
* This chart shows the amount of dairy products in stock in each country. This chart helps companies have a clearer and more visualized picture of how much of their dairy products are in stock in each country. In other words, this chart can help companies manage their inventory more effectively, ensuring that enough stock is available to meet demand while avoiding overstocking. Also, this chart can help companies optimize sales strategies, such as promotions and discounts. Based on this chart, the company should replenish the quantities of dairy products stocked in Spain and Poland in a timely manner to ensure that stock levels meet demand.

**Visualization#3**

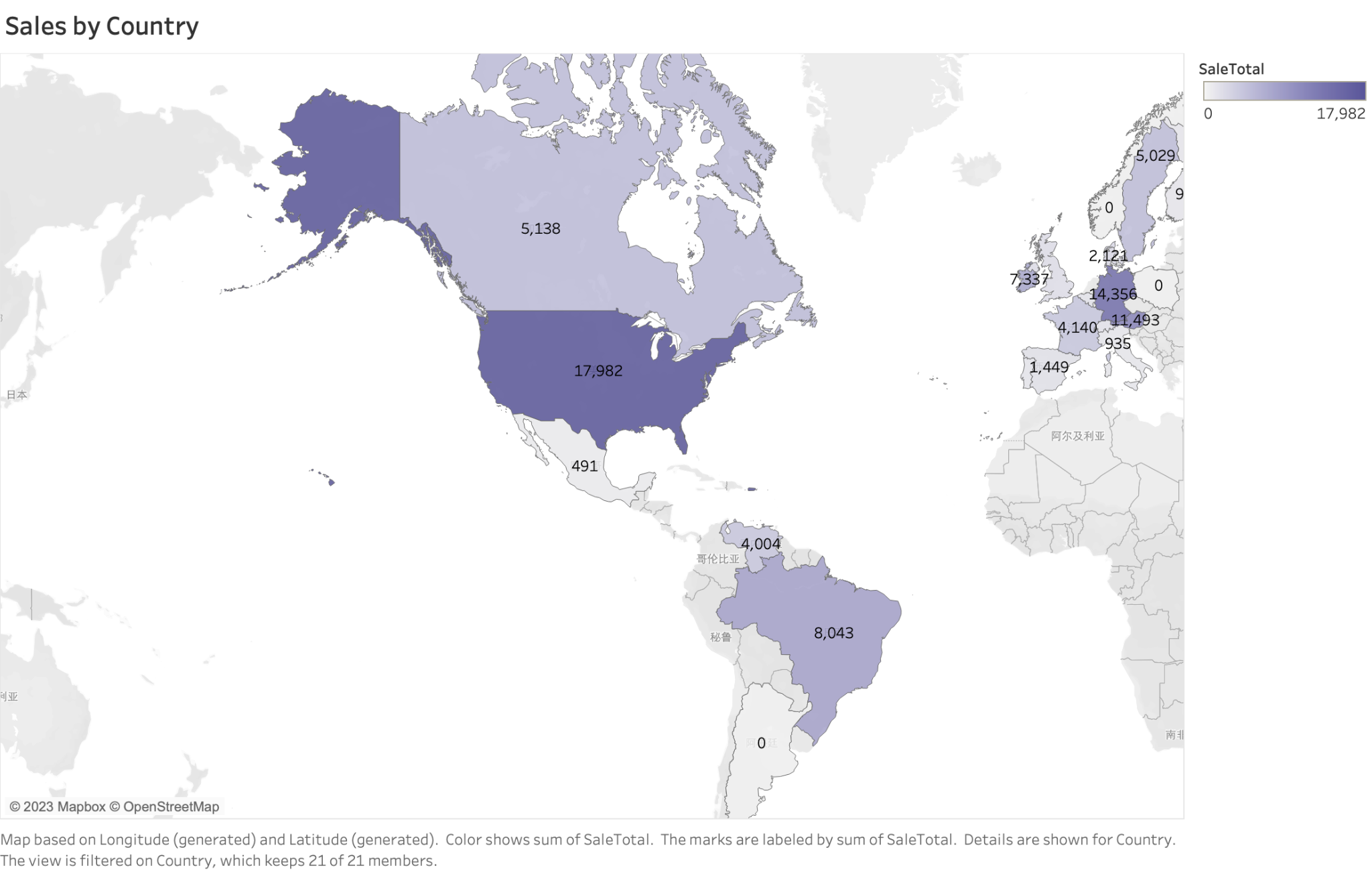


* The chart reflects the distribution of total U.S. sales from the third quarter of 1996 through the second quarter of 1998. The chart shows that Indiana's meat sales were significant in the second quarter. Also, the chart helps companies quickly identify the status of each product category's sales in each state for each quarter. It helps the company to understand the sales growth potential and product performance in each state more quickly, which in turn helps the company formulate a better marketing strategy. For example, with the chart, we can see that Indiana saw a decline in sales of meat products in the third quarter, so the company should consider running promotions from July to September to increase sales.

**Visualization#4**

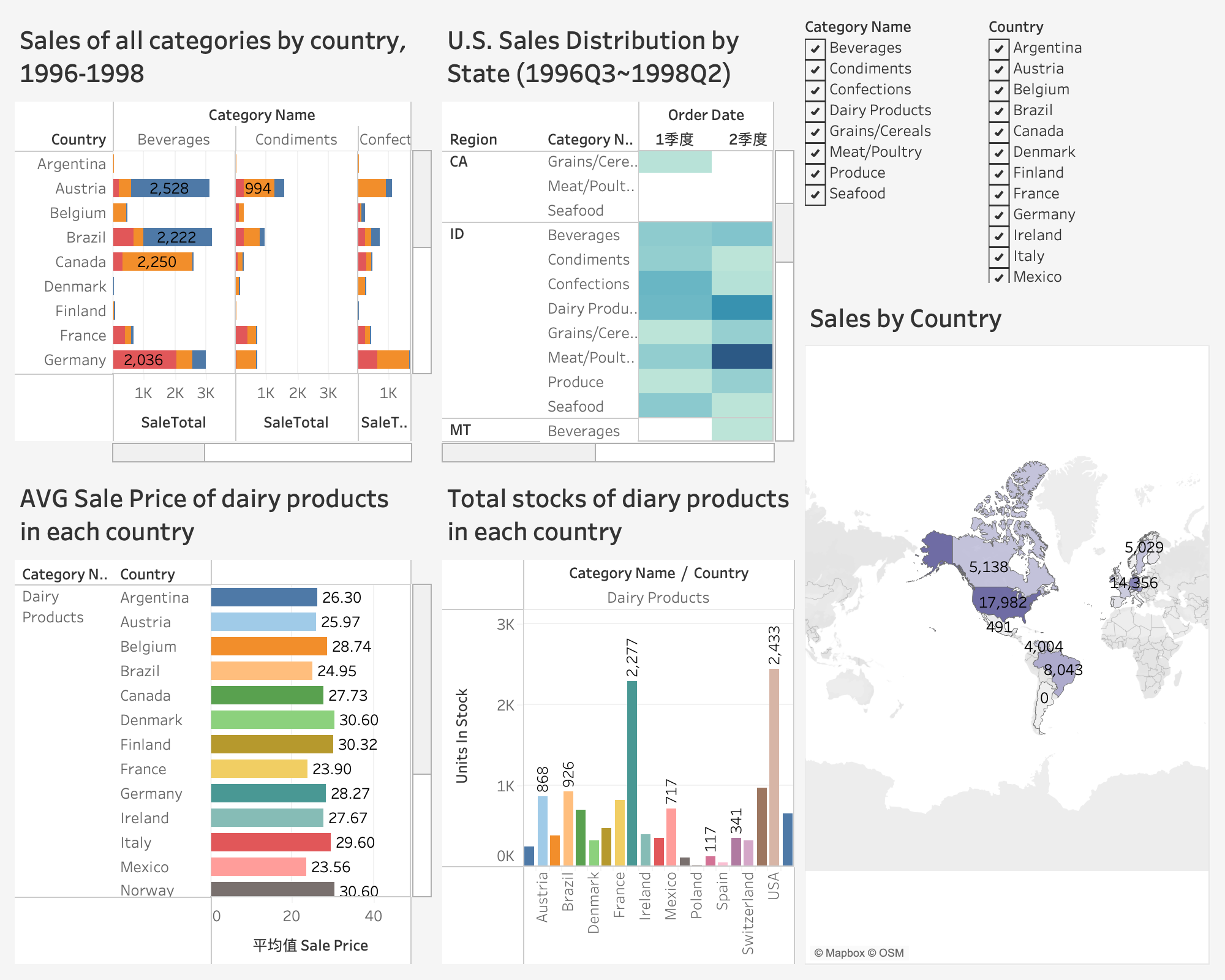


* The chart shows the sales of each category from 1996 to 1998 between countries. The chart gives the company a clear picture of which categories have the highest or lowest sales in which countries, and helps the company identify which categories are contributing the most to sales and which may need to be adjusted or eliminated. For example, the chart shows that Portland's sales of all types of products from 1996 to 1998 were not very promising. Therefore, the company should consider whether to withdraw from the country's market.

**Visulization#5**

* The chart shows the company's sales in each country. This map visualization helps the company to better analyze the market and understand how sales are distributed among different countries. From there, it can be determined which countries contribute the most to the company's sales and which countries may have market potential to be exploited. In addition, it can also help the company understand its competitive position in the international market. Depending on the chart, the company's market sales in some of the Nordic countries are not impressive, and therefore the company should consider withdrawing from certain Nordic markets, such as Norway. We also know that the company has the best sales in the U.S. market, so the company should conduct a more in-depth market study. Explore customer needs to further increase customer satisfaction and maintain customer loyalty. Strengthen its market position in the U.S. to increase sales and remain competitive.

**Dashboard**

****