

Created For



Analysis Report

This report was created to fulfil Mini Project Data Engineering in the Data Science Learning Studio Bootcamp 2023

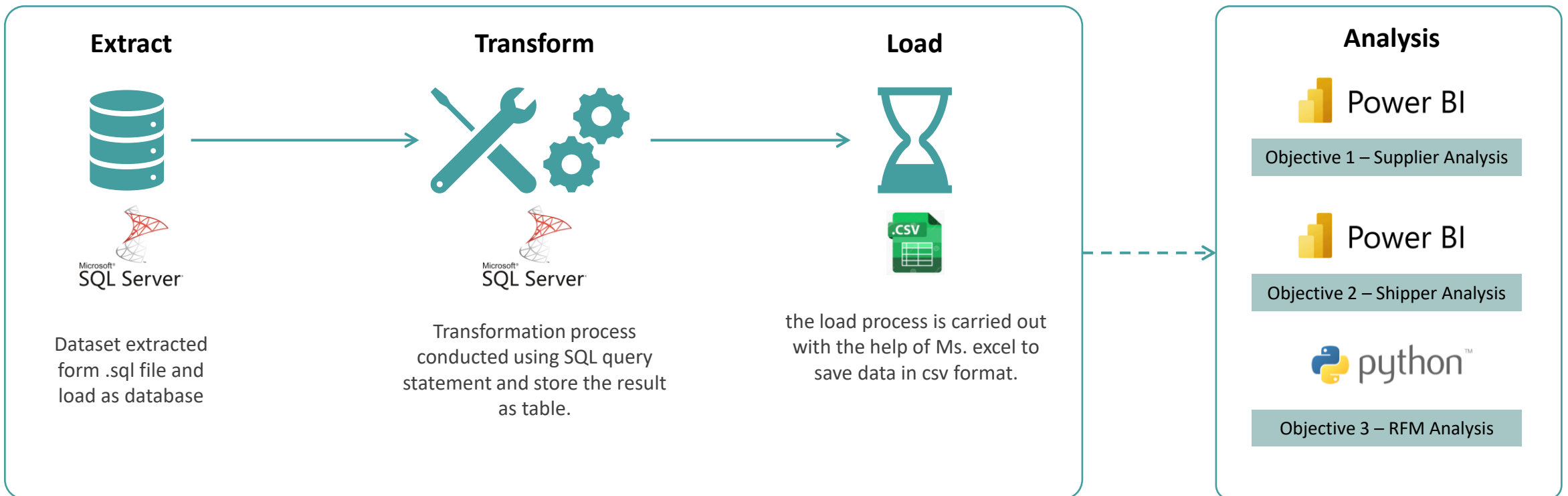
Reported by **Ayub Indra**



Project Overview

As a data consultant is to help companies make data-driven decisions and optimize data usage. To realize this responsibility, the first step that must be mastered by the data consultant is to be able to integrate data using the **ETL (Extract, Load, Transform)** method.

ETL Schema & Tools





Objective 1

Supplier Analysis

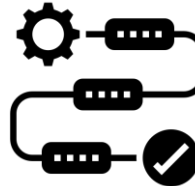
Analysis in Brief



Purpose

This analysis was conducted to gain insight into how each supplier vendor is performing, here are some metrics to look at.

- Which product category with best selling products?
- What is top 5 best selling products?
- Which supplier company has the most number of unit sold? (best selling products)
- Which country with the most number product sold?
- How the stock condition of top 5 best selling products?



Analysis Method

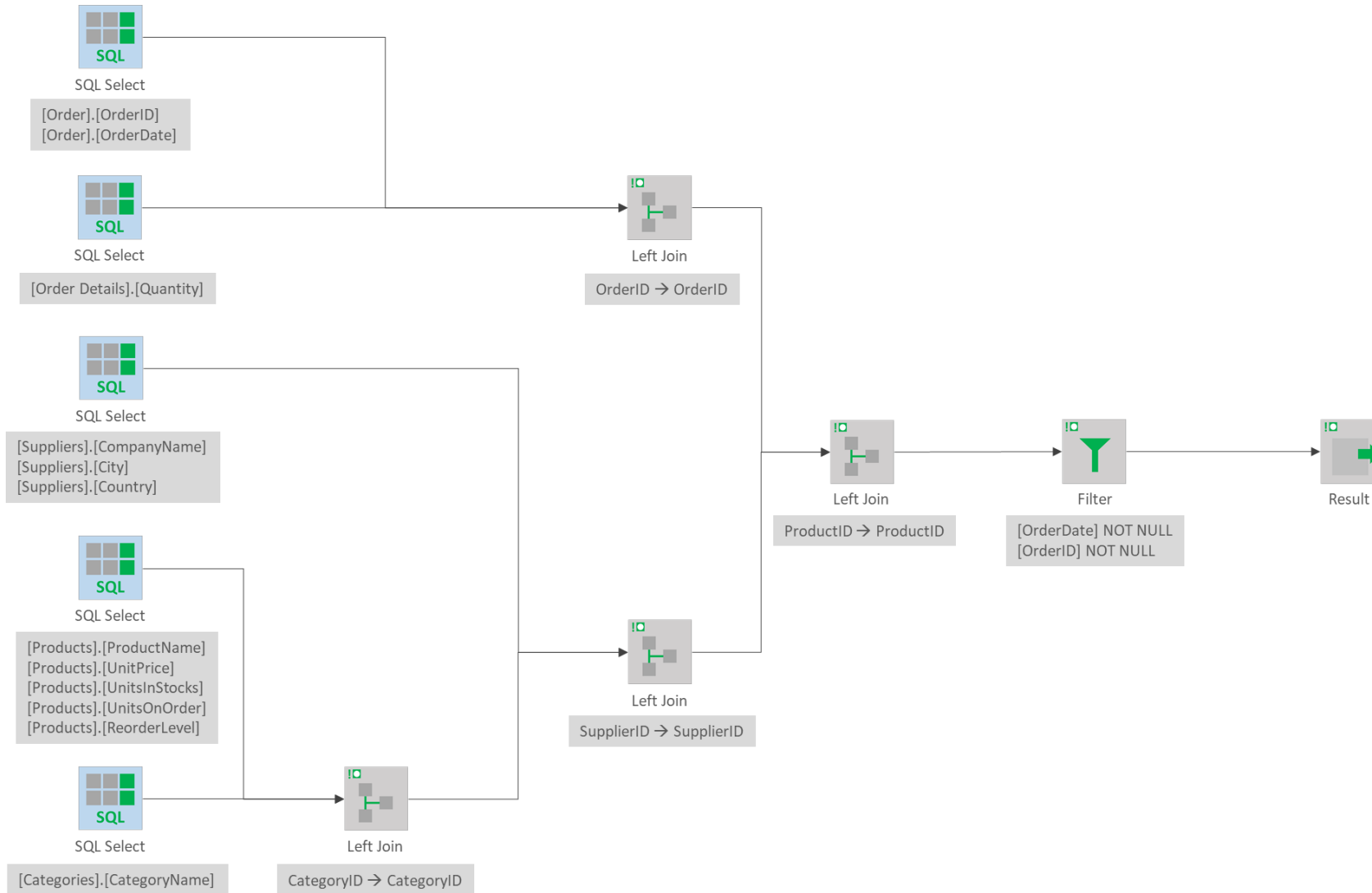
The analysis carried out is based on descriptive analysis, where the datasets that have been processed through the ETL method are explored using Microsoft Power BI Desktop for the process of extracting insights.



Insight Summary

- The beverages, dairy products, and confections contribute about 50% of unit sold in total
- The Gai Paturage (French supplier) has contribute to the top 2 best selling products.
- From the table above, we can gain the insight that 4 of 5 best selling products in the condition Need to Order ASAP, and 1 of them in the empty condition.

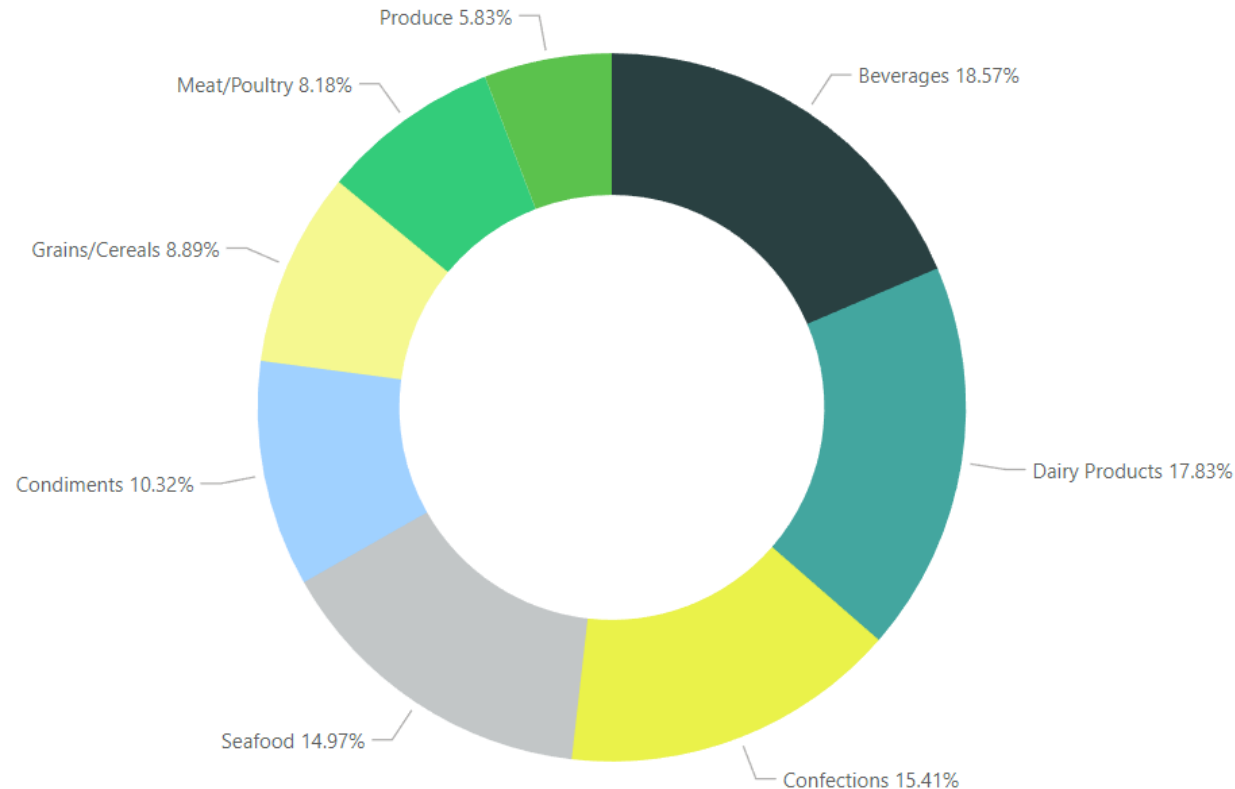
Data Processing Flowchart



Data Dictionary

Columns	Description
OrderID	Number of unique order
Order Date	Date when order occurred
Product	Product name ordered
Unit Sold	Total product ordered in one order (then as re: Quantity)
Unit Price	Price per unit of product
Category	Category of product ordered
Supplier Name	The name of company / vendor / suppliers
Supplier City	The city where supplier located
Supplier Country	The country where supplier located
Units In Stock	Total units in the warehouse
Units On Order	Total units on order to the suppliers

Unit Sold by Category

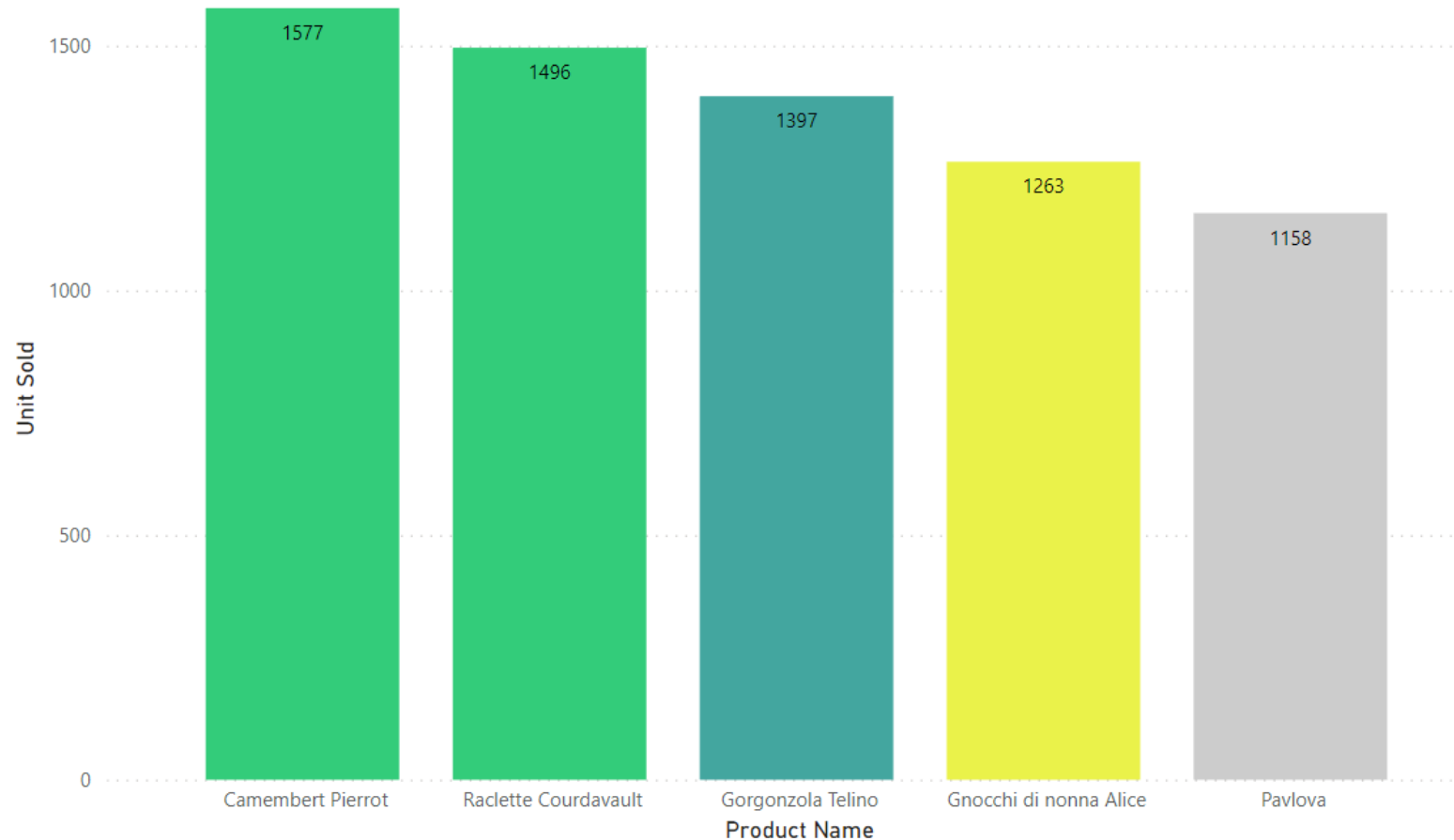


During June 1996 – May 1998, Beverages products is the best selling products by category that contributing 18.57% of total, and then followed by Dairy Products with 17.83% and Confections 15.41%.

The top 3 product category contribute about **50%** of total unit sold, and **50%** contribute by other 5 different categories.

Top 5 Best Selling Products

Supplier Name ● Formaggi Fortini s.r.l. ● Gai p torage ● Pasta Buttini s.r.l. ● Pavlova, Ltd.

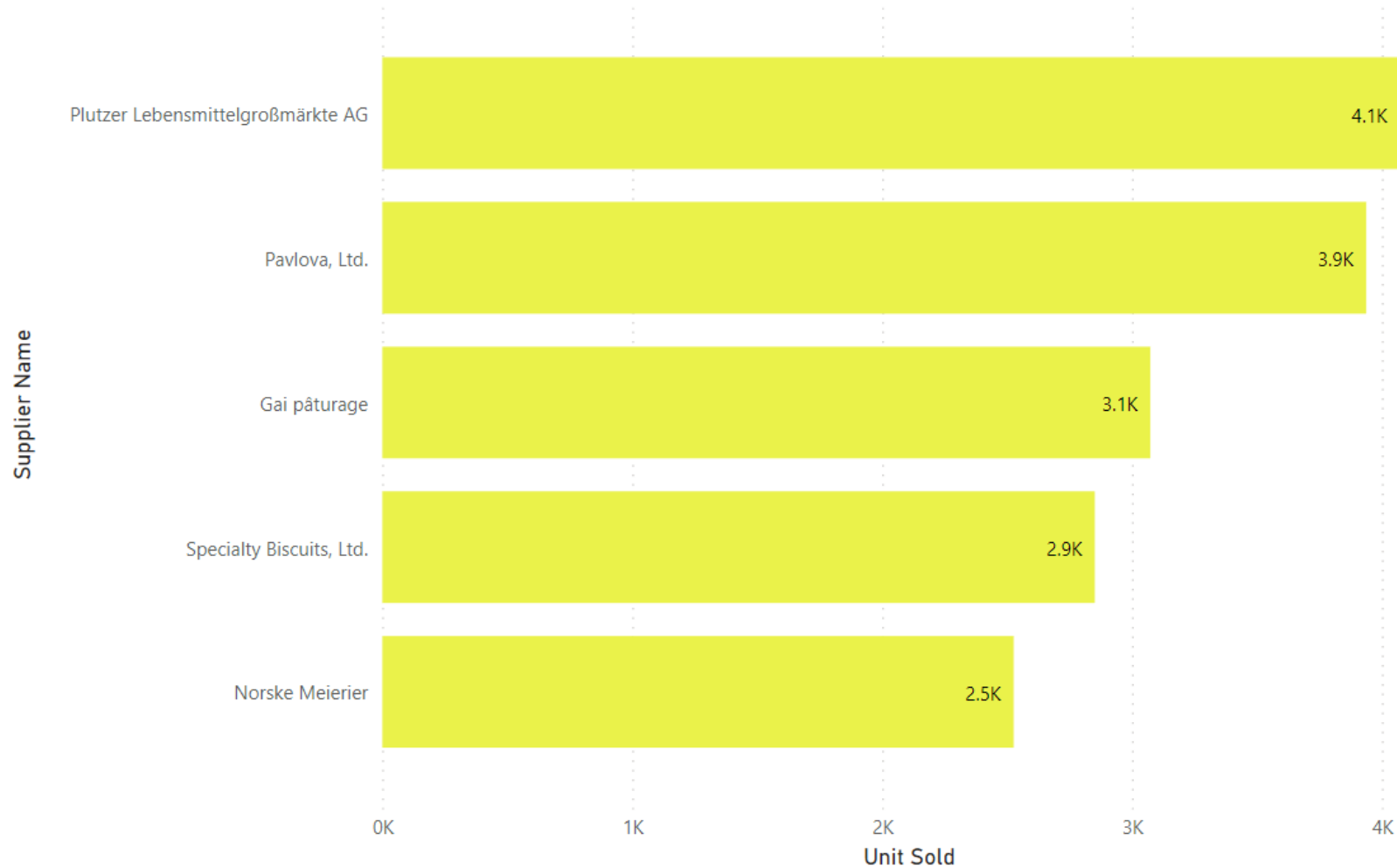


From the chart in the left, top 2 best selling products contributed by 1 supplier namely **Gai p torage** from France with the total about 3,100 unit sold.

And then followed by 3 different products with 3 different suppliers.

The top 3 best selling products dominated by **Dairy Products**, followed by Grains/cereals and confections.

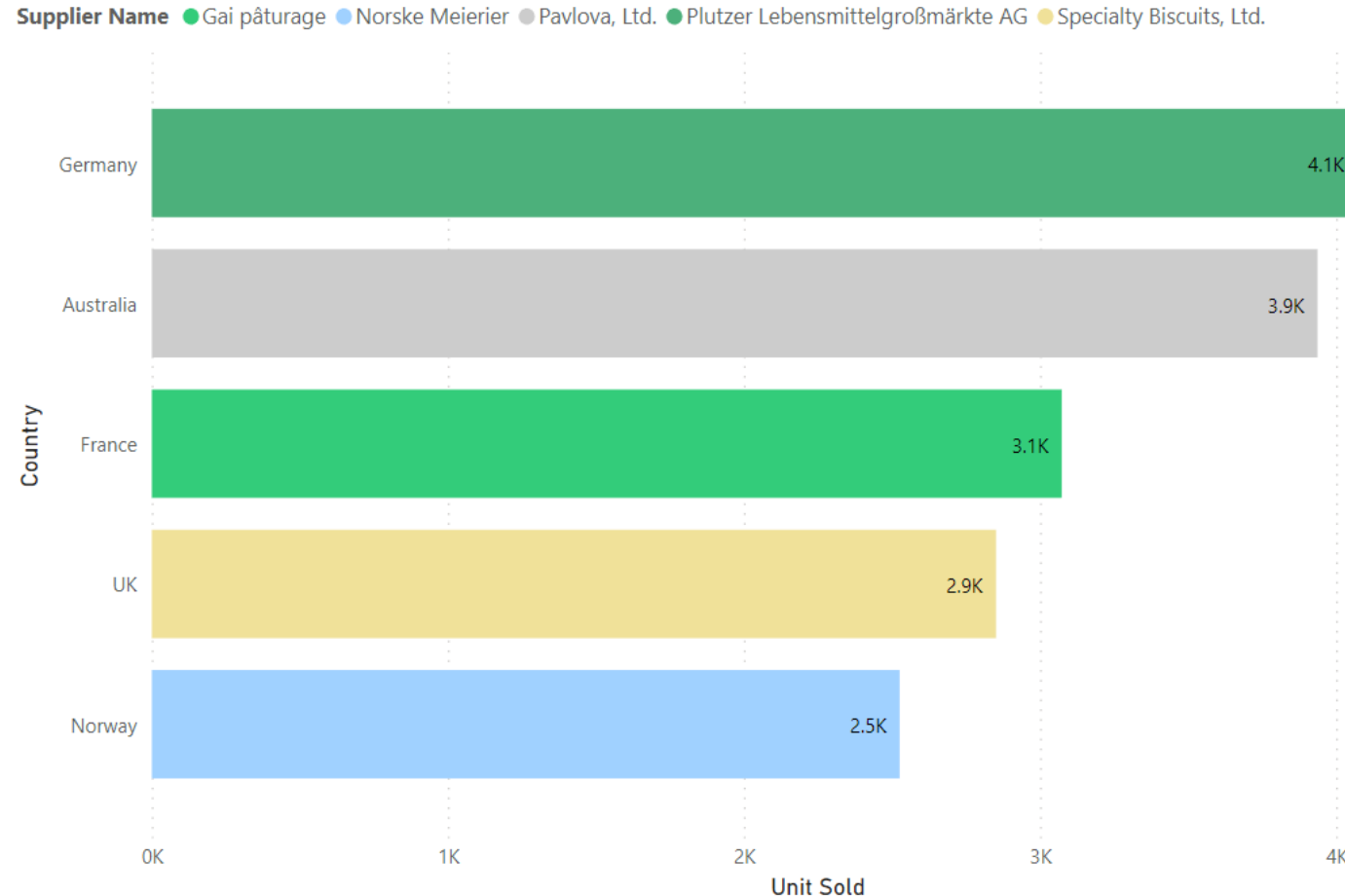
Top 5 Suppliers



In the overall, German suppliers is the top supplier with the most total unit sold.

Pavolva, Ltd. which supply the Pavolva product in the second place, and Gai Paturage as the company with best 2 best selling products following in the third place.

Top 5 Suppliers Based on Total Unit Sold



Continuing from the last slide, we can get the snapshot that the top 5 supplier based on total unit sold only have one supplier for each country.

The **Australian** (Pavolva, Ltd.) and **French** (Gai paturage) in the position 2 and 3 of the rank.

Top 5 Best Selling Products Stock Condition

Product Name	Category	Supplier Name	Supplier Country	Units In Stock	Units On Order	Total Unit Sold
Camembert Pierrot	Dairy Products	Gai pâturage	France	19.00	0.00	1577
Raclette Courdavault	Dairy Products	Gai pâturage	France	79.00	0.00	1496
Gorgonzola Telino	Dairy Products	Formaggi Fortini ...	Italy	0.00	70.00	1397
Gnocchi di nonna Alice	Grains/Cereals	Pasta Buttini s.r.l.	Italy	21.00	10.00	1263
Pavlova	Confections	Pavlova, Ltd.	Australia	29.00	0.00	1158

Stock Empty Need to Order ASAP

Before made the table above, we made the rule to label the stock condition

- Stock empty : when no units in stock and no units on order
- Need to order ASAP : when the units in stock below 200 in total

From the table above, we can gain the insight that **4 of 5 best selling products** in the condition Need to Order ASAP, and only 1 of them just only ordered. And then **only 1 product** (Gorgonzalo Telino) in the condition of empty stock and it just already ordered.

Recommendations

1

For more beneficial and focused trading, we need to maximize the partnership with Gai paturage, because they contribute to our top 2 best selling products.

2

From the product category, for more focused trading we can optimize our product stock with beverages, dairy, and confection products because of they are contribute to the 50% of unit sold.

3

The top 5 best selling product is in dangerous stock condition, lets update the stock ASAP with ordering to the suppliers.

The Dashboard



51.32K
Total Unit Sold

77
Total Products

29
Total Suppliers

STOCK ITEM & SUPPLIER DASHBOARDS

Product Name

All

Category

All

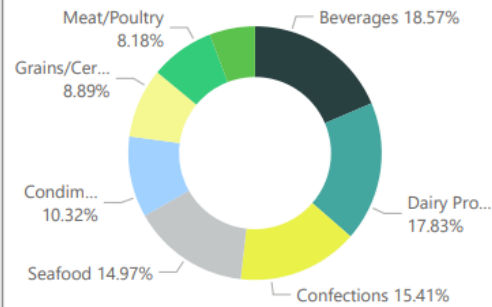
Supplier Name

All

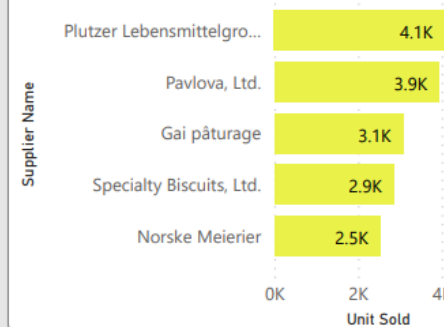
Year

All

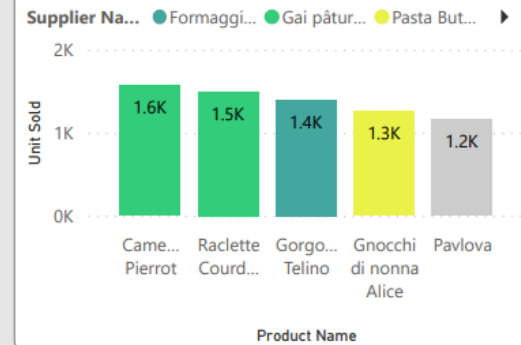
Unit Sold by Category



Top 5 Suppliers



Top 5 Best Selling Products

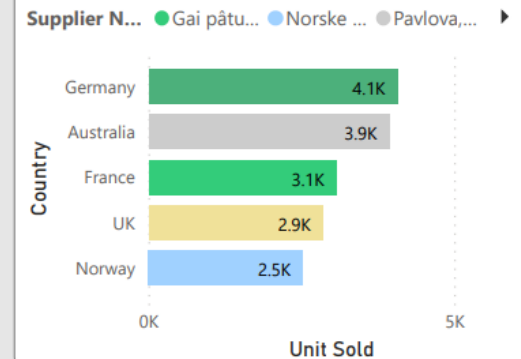


Stocks Update

Product Name	Category	Supplier Name	Supplier Country	Units In Stock	Units On Order
Camembert Pi...	Dairy Products	Gai pâturage	France	19.00	0.00
Raclette Courd...	Dairy Products	Gai pâturage	France	79.00	0.00
Gorgonzola Tel...	Dairy Products	Formaggi Fortini ...	Italy	0.00	70.00
Gnocchi di non...	Grains/Cereals	Pasta Buttini s.r.l.	Italy	21.00	10.00
Pavlova	Confections	Pavlova, Ltd.	Australia	29.00	0.00
Rhönbräu Klos...	Beverages	Plutzer Lebensmit...	Germany	125.00	0.00
Guaraná Fantá...	Beverages	Refrescos Americ...	Brazil	20.00	0.00
Boston Crab M...	Seafood	New England Sea...	USA	123.00	0.00
Tarte au sucre	Confections	Forêts d'érables	Canada	17.00	0.00

Stock Empty Need to Order ASAP

Top 5 Unit Sold by Supplier Country





Objective 2

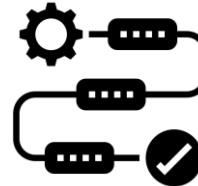
Shipper Analysis



Purpose

This analysis was conducted to gain insight into how each shipping vendor is performing, here are some metrics to look at.

- Which shipper company has the most number of unit shipped?
- Which shipper company has faster shipping time before required date?
- Which country with most product shipped? And how the shipper company distribution in it?
- How the shipping trends over time?
- Which shipping service is used the most?



Analysis Method

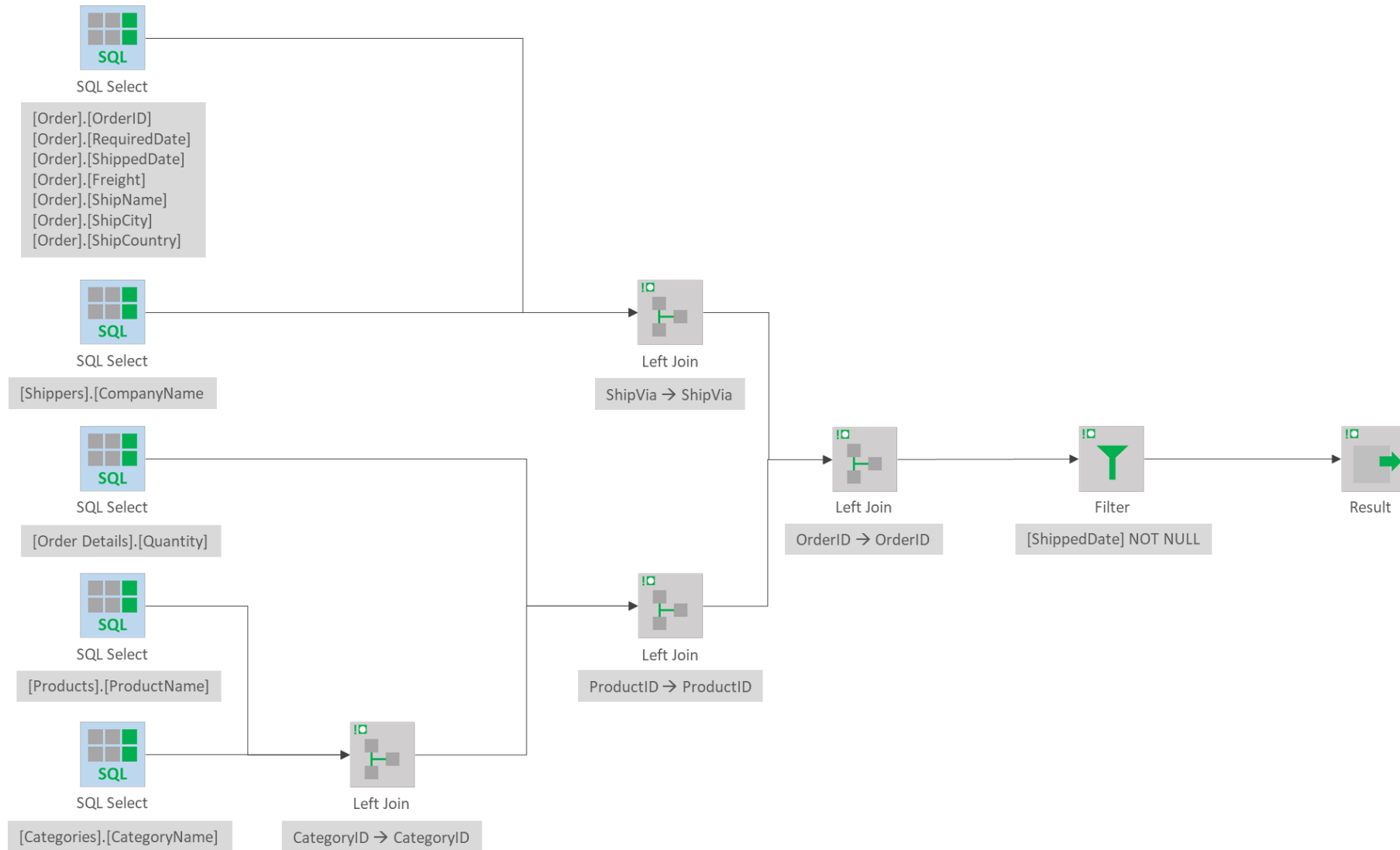
The analysis carried out is based on descriptive analysis, where the datasets that have been processed through the ETL method are explored using Microsoft Power BI Desktop for the process of extracting insights.



Insight Summary

- About 2 years, United Package shipper company with the most number unit shipped (19,195) and the lowest average day difference between required date and shipped date.
- The 2 of 3 shipper company have the most dominant unit shipped were beverages product.
- The USA and Germany are the top 2 countries where the most units are shipped, the three shipper companies tend to have the same contribution in all customer countries

Data Processing Flowchart



Data Dictionary



Columns	Description
OrderID	Number of unique order
Product	Product name ordered
Category	Category of product ordered
Quantity	Total product ordered in one order (then as re: Unit Shipped)
Required Date	Maximum date units have to arrived to the customers
Shipped Date	Date when units shipped to the customers
Day Diff	Number of day difference between required date and shipped date
Freight	Freight (\$) that have to pay to the shippers
Shipper Company	The name of company that giving shipping services
Ship Name	Name of ship service
Ship City	City of shipper company
Ship Country	Country of shipper company

General Information

50.12K

Total Unit Shipped

This metric tell how many unit (quantity of goods ordered) shipped to the customers between 1996 – mid of 1998.

-19.47

Average of Day Difference
Before Required Date

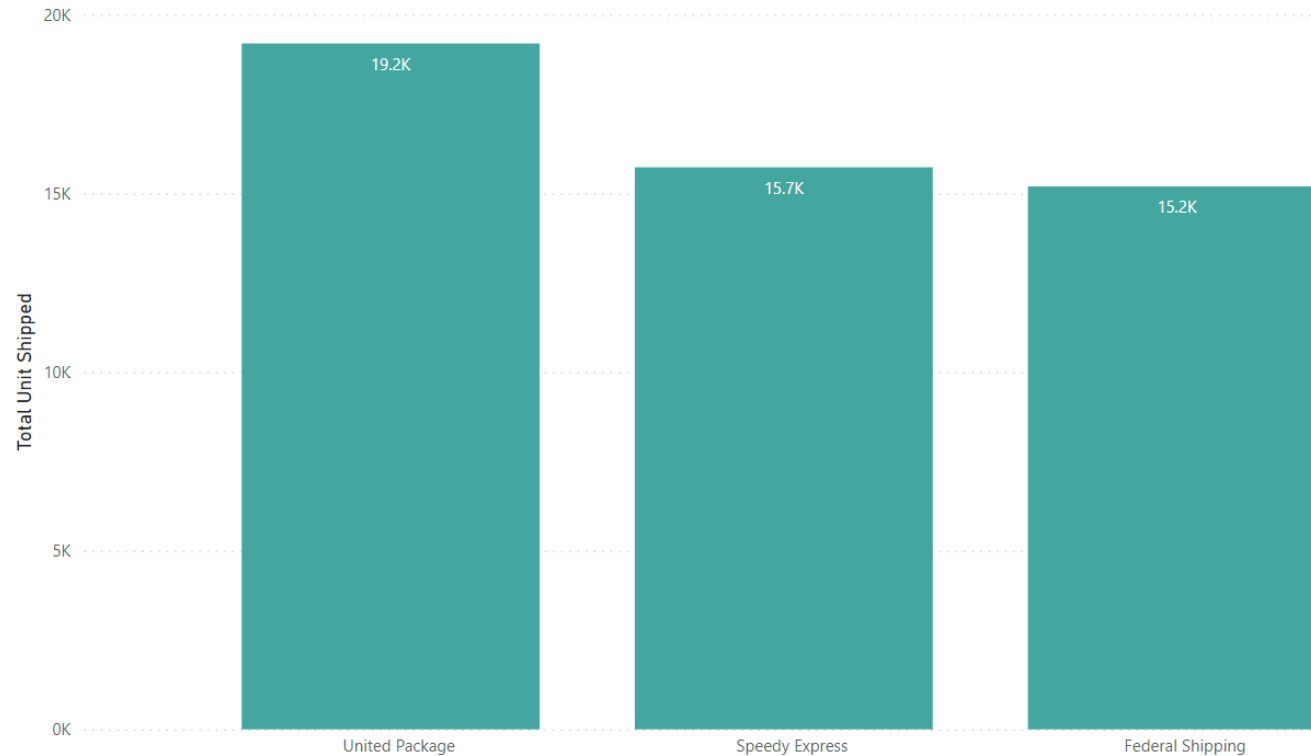
This metric tell the average day unit shipped between shipped date and required date that overall Northwind made.

\$97.96

Average of Freight per
Shipping

This metric tell how many fee that have to Northwind pay to the Shipper per shipping.

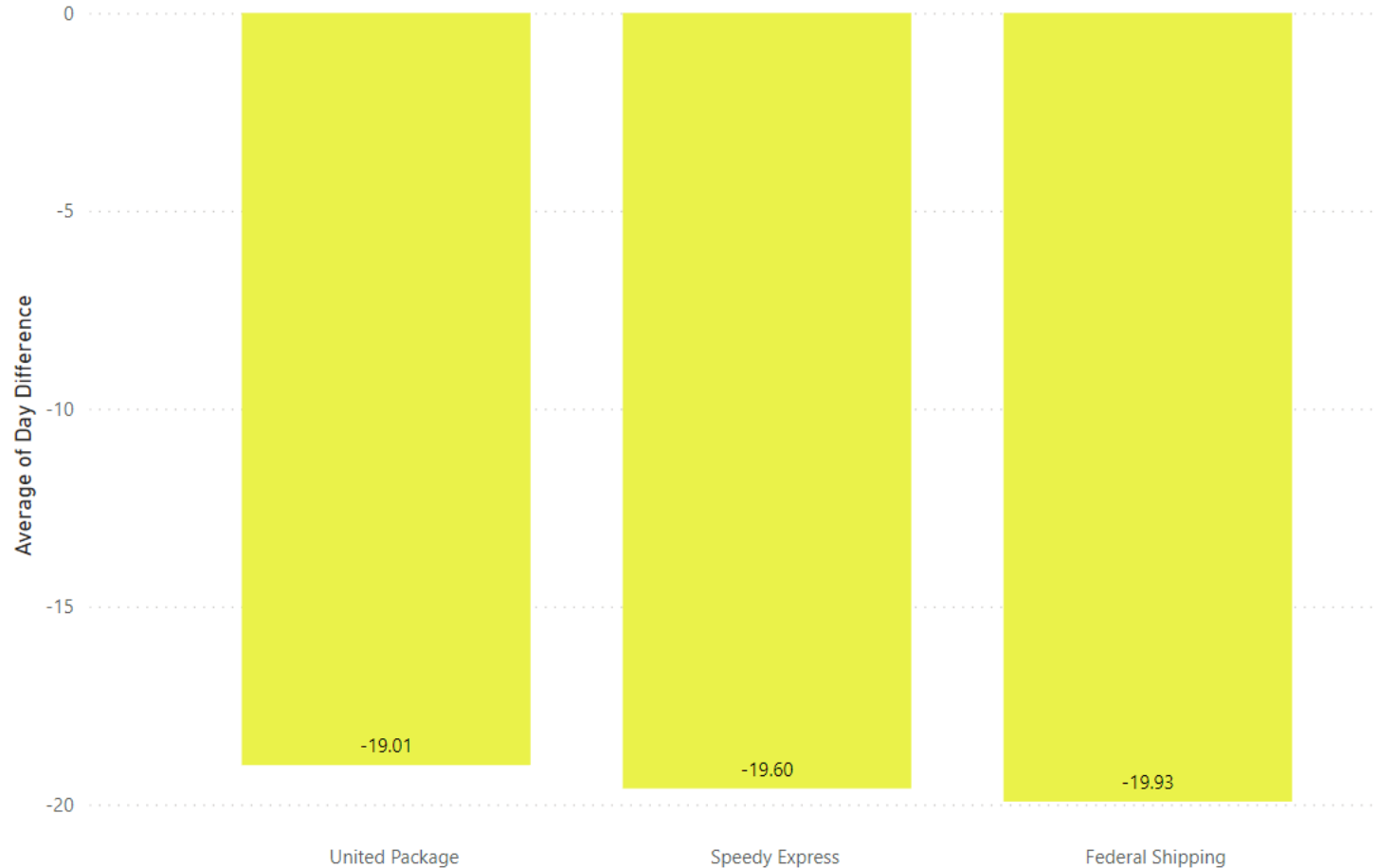
Unit Sold by Shipper Company



Metrics in the left shown how many total unit shipped by Shipper Company. We want to see which shipper company that dominantly send the order to the customers in the number of quantity of unit.

For 2.5 years (1996 – mid of 1998), United Package has the advantage than other shipper company with the **19,195** unit shipped. Both of Speedy Express and Federal Shipping have the similarity number were **15,730** and **15,194** unit shipped

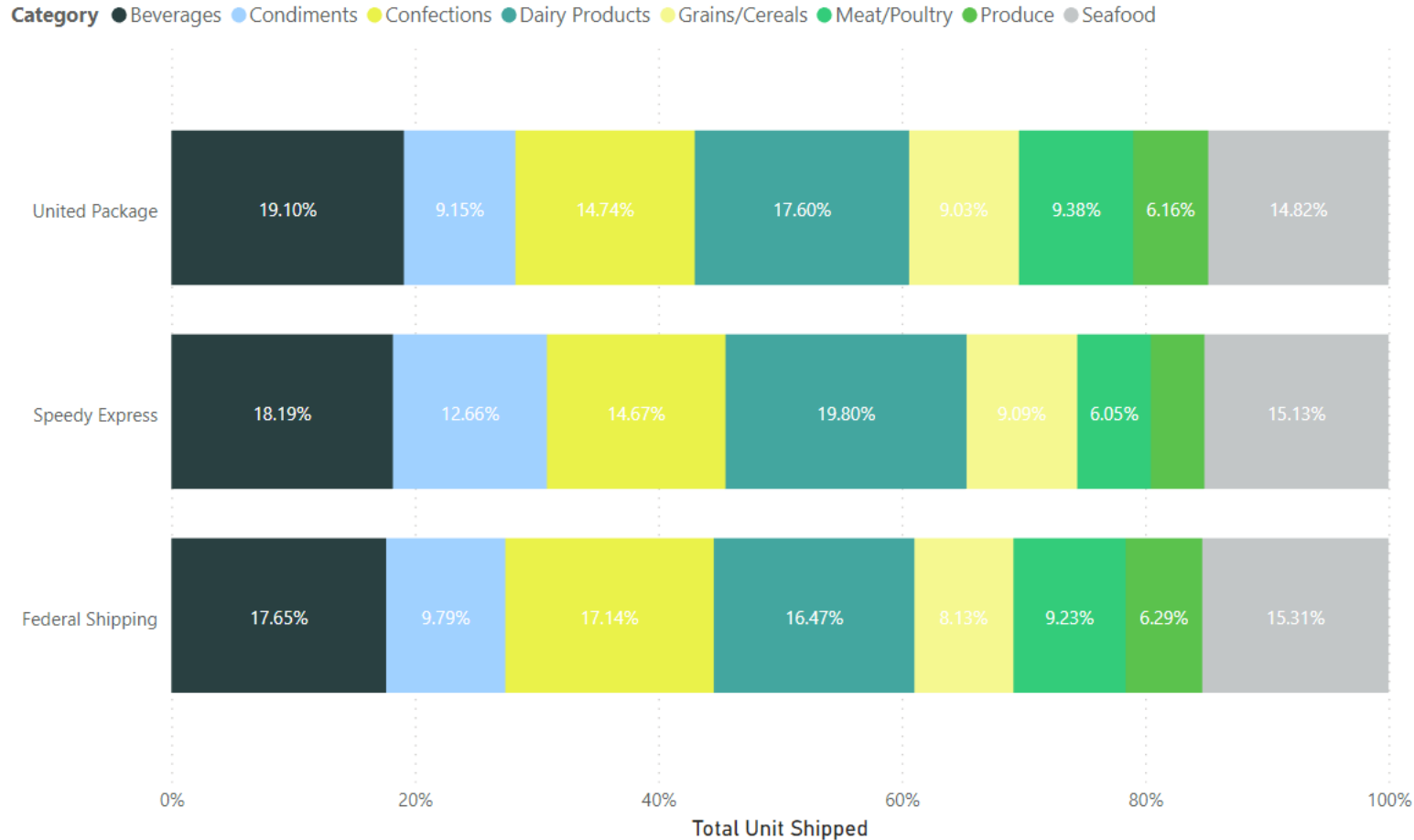
Average Order Shipped Before Required Date



This slide will explain which the shipper company that have earlier average D-required date.

For 2.5 years (1996 – mid of 1998), United Package ship the order to customer in **Day-19.01** before required date in average, this is faster than Speedy Express and Federal Shipping were on average ship the order in **Day-19.60** and **Day-19.93** before required date.

Unit Shipped by Product Category



The 3 (three) shipper company have similar distribution of product category which the products their shipped to the customers. Beverages and Dairy products is the top 2 product category shipped overall.

- **United Package** dominantly ship beverages products.
- **Speedy Express** dominantly ship the dairy products.
- **Federal Shipping** dominantly ship beverages products.

Shipping Maps by Shipper Company



We want to see geographical insight where country Northwind product shipped by the shipper company.

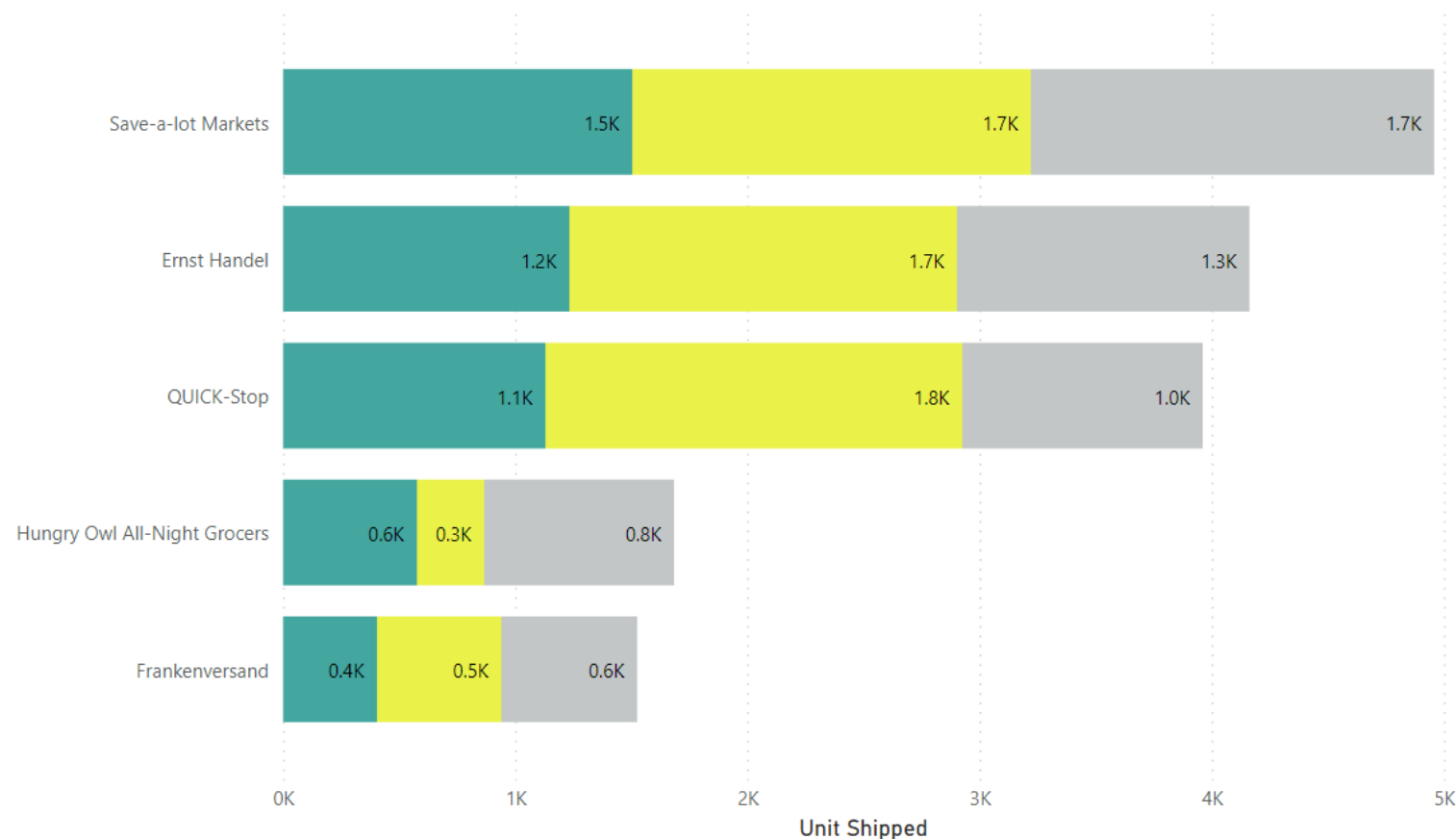
Overall, **USA** and **Germany** is the top 2 country where product shipped.

For the 3 (three) shipper companies, they have **similar distributions** per country by number of unit shipped.

It's means the shipper companies covered all of customer country base.

Top 5 Unit Shipped by Ship Name

Shipper Company ● Federal Shipping ● Speedy Express ● United Package



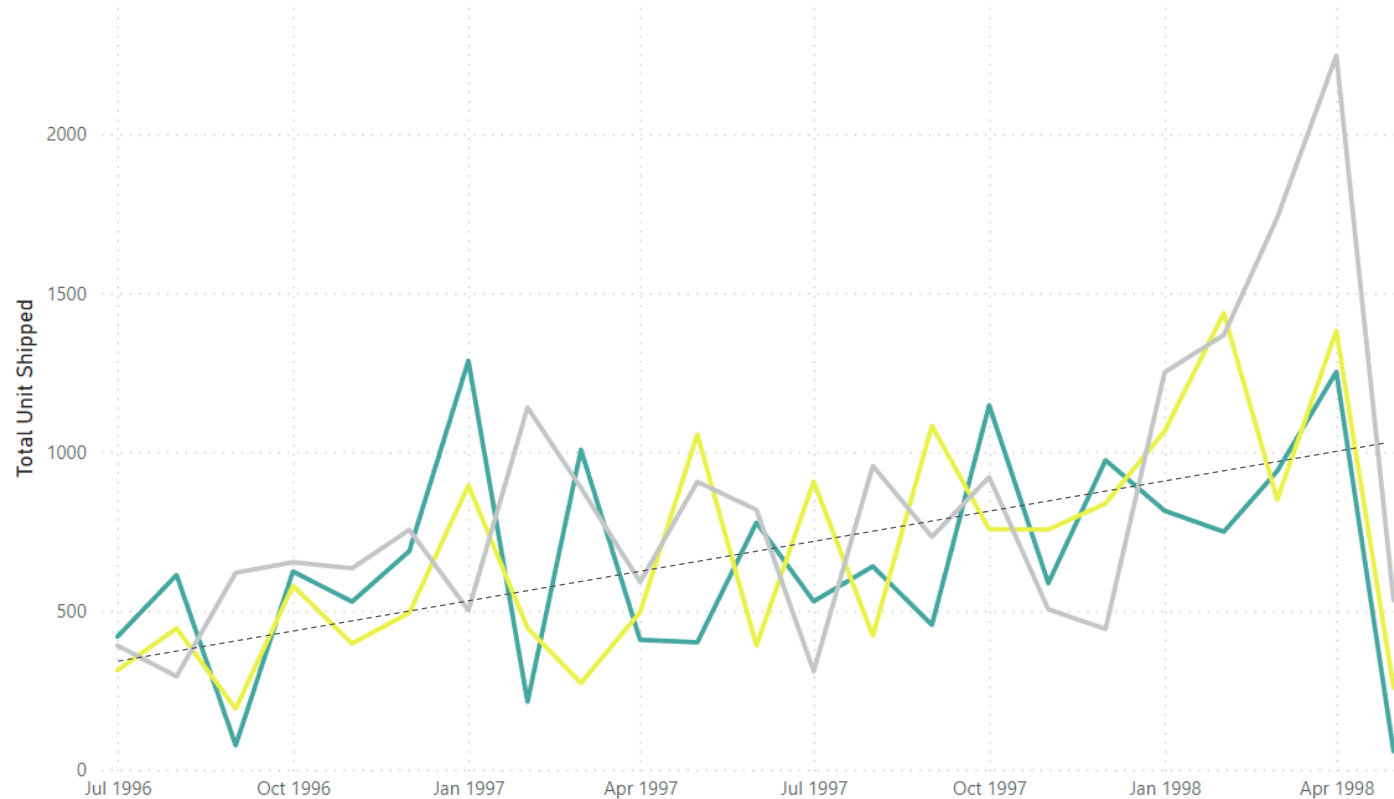
We want to see which shipping service that mostly used to sending the ordered goods.

Save-a-lot markets is the top service with the total about 5,000 times shipping, it is have similar portion of each shipper company. The same patter also applied to **Ernst Handel** and **QUICK-Stop** service.

In this top 5 total unit shipped, **Speedy Express** tend to dominant than other shipper company by portion for each service.

Unit Sold by Shipper Company

Shipper Name ● Federal Shipping ● Speedy Express ● United Package



Overall, from July 1996 – April 1998 Northwind has **uptrend pattern** of the number of total unit shipped. But the trend broken and falling down deeper after April 1998.

For the 3 (three) shipper company, they have similarity pattern in the chart left. No significant difference between them.

United Package and Speedy Express gaining **their peak after 1998** begin.

Recommendations

1

Because of their performance (number of shipping and low number day difference between required date and shipped date), we need to optimize the United Package to ship more products to our customers.

2

The drop of shipping number trend after April 1998 need to investigate, but we recommend the marketing department to improving their strategy to get more orders and increasing our trends again.

The Dashboard



50.12K

Total Unit Shipped

-19.47

Average of Day Difference
Before Required Date

\$97.96

Average of Freight per
Shipping

SHIPPING DASHBOARD

[Shipping Summary]

Product Name

All

Category

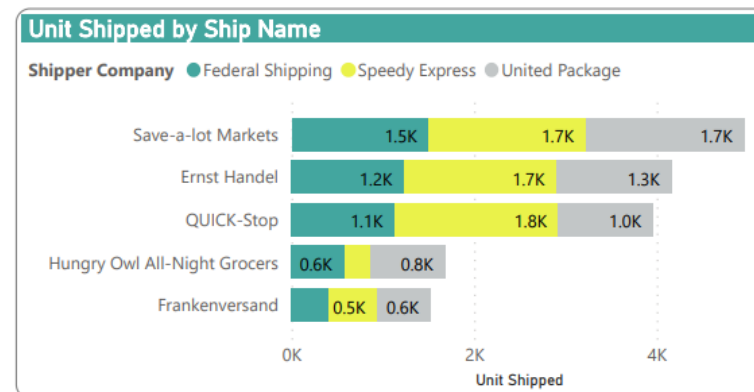
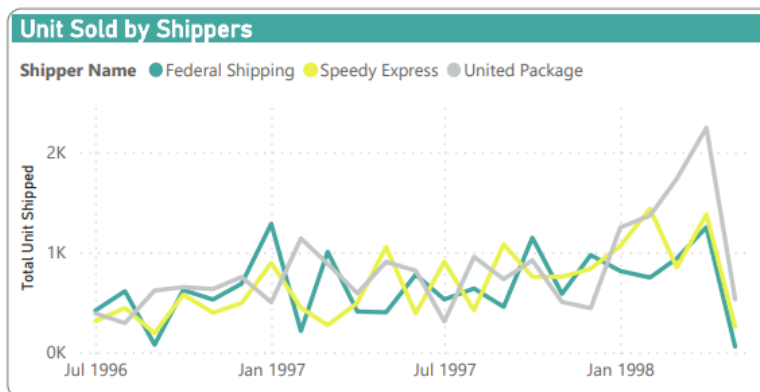
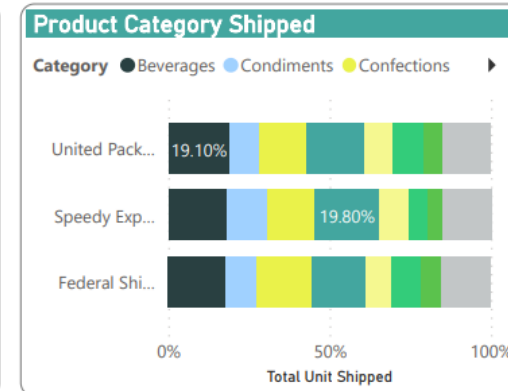
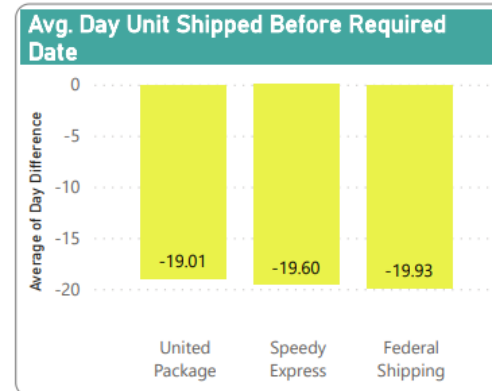
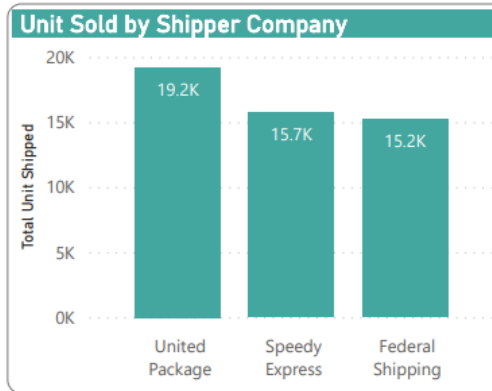
All

Ship Country

All

Ship Name

All



Objective 3

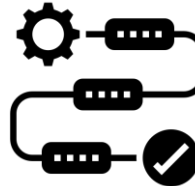
RFM Analysis



Purpose

RFM (Recency-Frequency-Monetary) analysis is a simple technique for behaviour based customer segmentation. It groups customers based on their transaction history – how recently, how often and how much did they buy. It is a handy method to find the best customers, understand their behaviour and then run targeted marketing campaigns to increase sales, satisfaction and customer lifetime value.

The goal of this analysis is to identify customer segments for global retail company using RFM analysis and to understand how those groups differ from each other.



Analysis Method

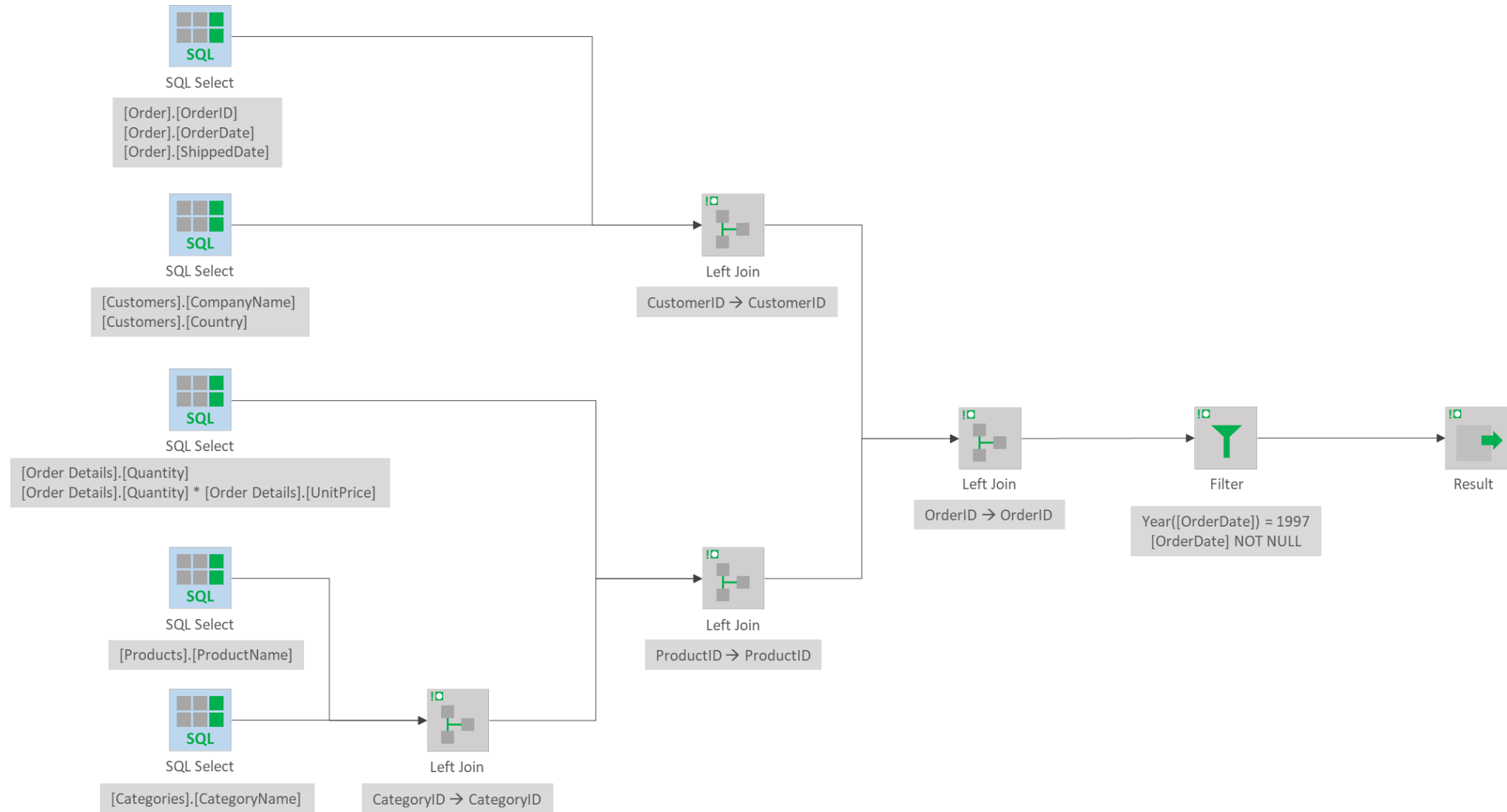
The analysis carried out is based on descriptive analysis, where the datasets that have been processed through the ETL method are explored using Python programming language (IDE: Jupyter notebook) for the process of extracting insights.



Insight Summary

- *Hibernating* segment dominating the customer that northwind's have with the total 20 customers followed by *Loyal customers* with 17 customers.
- *Champions* and *Loyal Customers* dominating the monetary share with the proportion 29.4% and 39.1% share. The other segments only have maximum share below 10%.
- The top 3 frequency is Champions, loyal customers, and can't loose them segment.
- The top 3 recency is champions, potential loyalist, and new customers.

Data Processing Flowchart



Data Dictionary

Columns	Description
OrderID	Number of unique order
Order Date	Date when order occurred
Shipped Date	Date when units shipped to the customers
Customer Name	The name of company that ordered the products
Country	Country of customers
Product Name	Product name ordered
Product Category	Category of product ordered
Unit Sold	Total product ordered in one order (then as re: Quantity)
Sales	Calculation of Quantity (Unit Sold) x Unit Price

RFM Category Segments

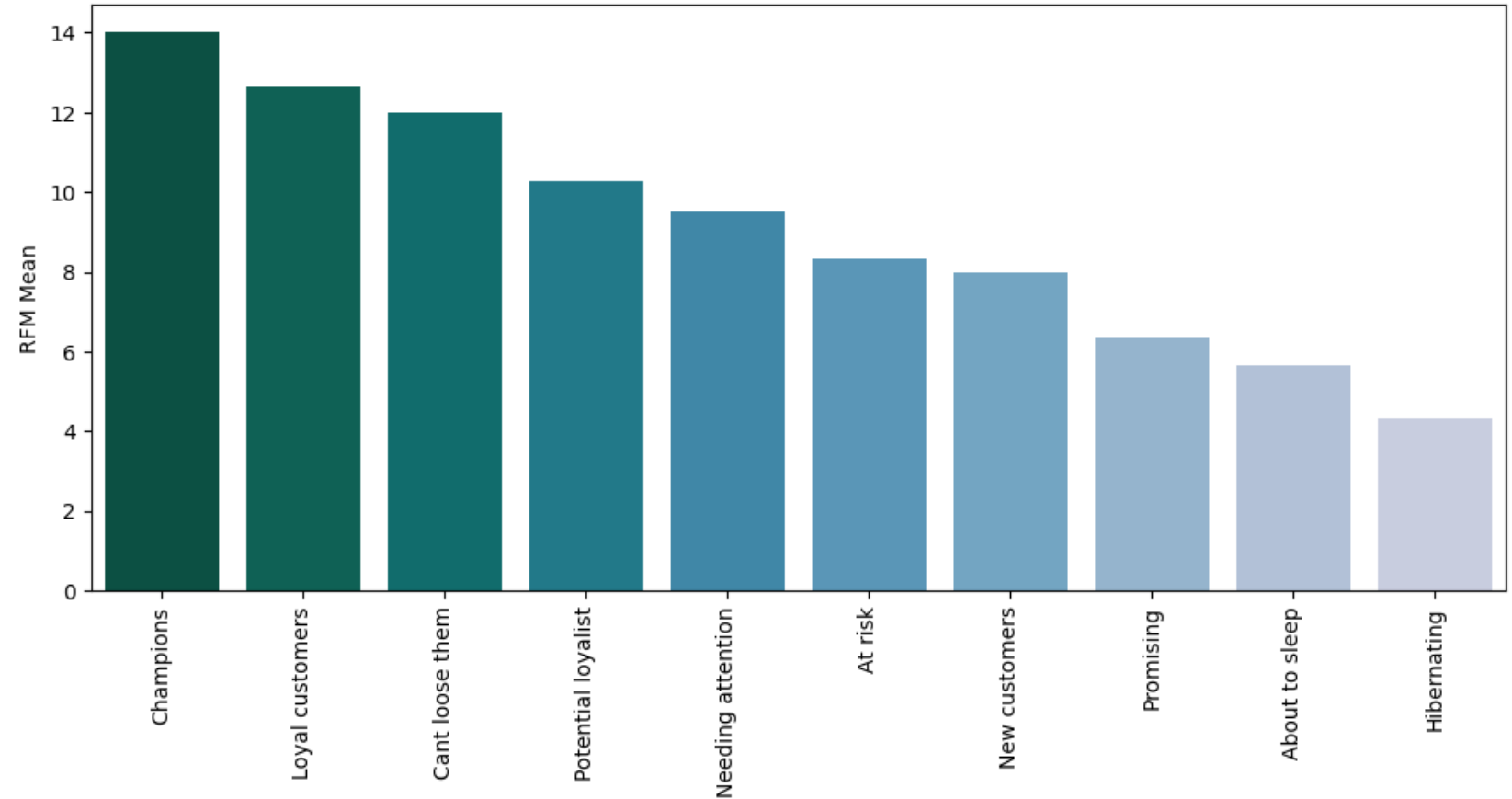
Rule : Before segmented, the value of recency, frequency, and monetary normalized (using for each R, F, and M score with the range number of 1 – 5).

Segment	Description
Champions	bought recently, buy often and spend the most
Loyal Customers	spend good money and often, responsive to promotions
Potential Loyalist	recent customers, but spent a good amount and bought more than once
New Customers	bought most recently, but not often
Promising	recent shoppers, but haven't spent much
Needing Attention	above average recency, frequency and monetary values; may not have bought very recently though
About To Sleep	below average recency, frequency and monetary values; will lose them if not reactivated
At Risk	<ul style="list-style-type: none">spent big money and purchased often but long time ago; need to bring them back
Can't Loose Them	<ul style="list-style-type: none">made biggest purchases, and often but haven't returned for a long time
Hibernating	last purchase was long back, low spenders and low number of orders

RFM Mean in Each RFM Segments

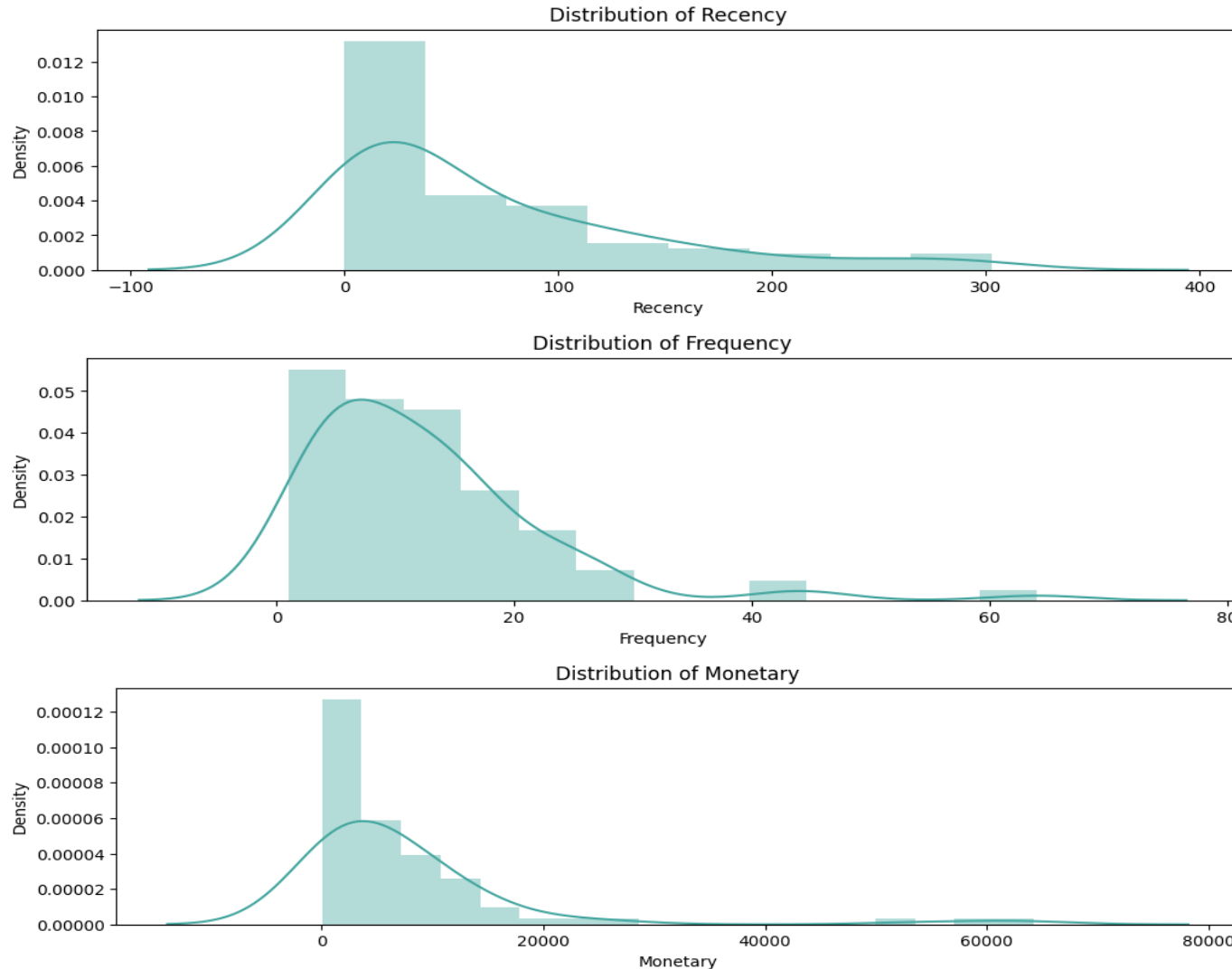
After the segmentation process and normalization that explained in the previous slide, in the right shown the chart that how many average **RFM_sum** for each segments.

Champions is the segment with the highest average RFM_sum and **Hibernating** is the lowest average RFM_sum.



$$\text{RFM_sum} = (\text{recency score} + \text{frequency score} + \text{monetary score})$$

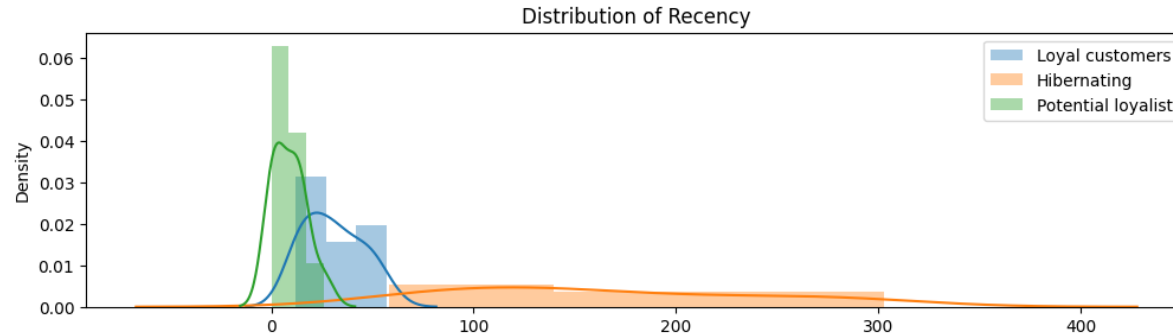
RFM Distribution Profile



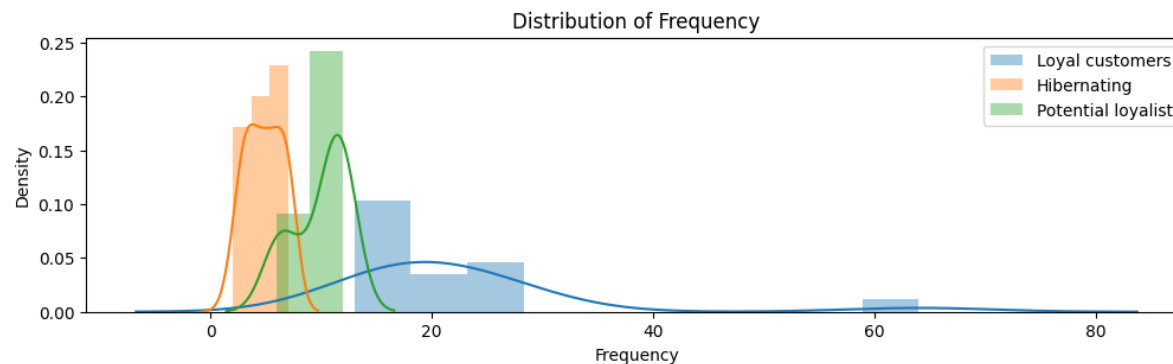
Recency distribution is right-skewed, showing that the majority of customers made a purchase in the recent past. It has a long tail containing single customers who made their order long time ago.

Both **frequency** and **monetary** distributions are bimodal (with two different peaks - modes). The chart reveals that one group of customers has low frequency (small number of orders, around 10 per customer) and another group is much **more "loyal"**. When it comes to monetary, it follows similar pattern. Part of customers have only spent little amount of money but there is another (significant in size) group that have spent much more.

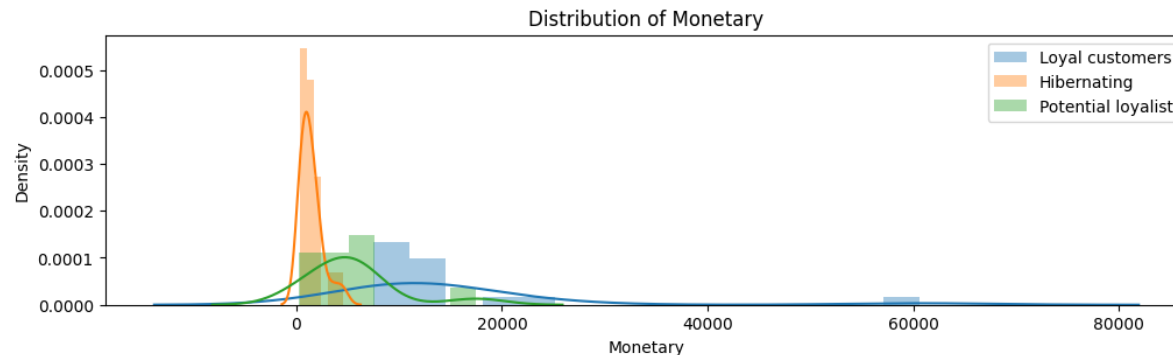
RFM Segment Distribution



Recency distribution give the insight that that potential loyalists have a low recency rate, this indicates that this segment has the potential to increase its level to loyal customers because in terms of recency it is lower than loyal customers.

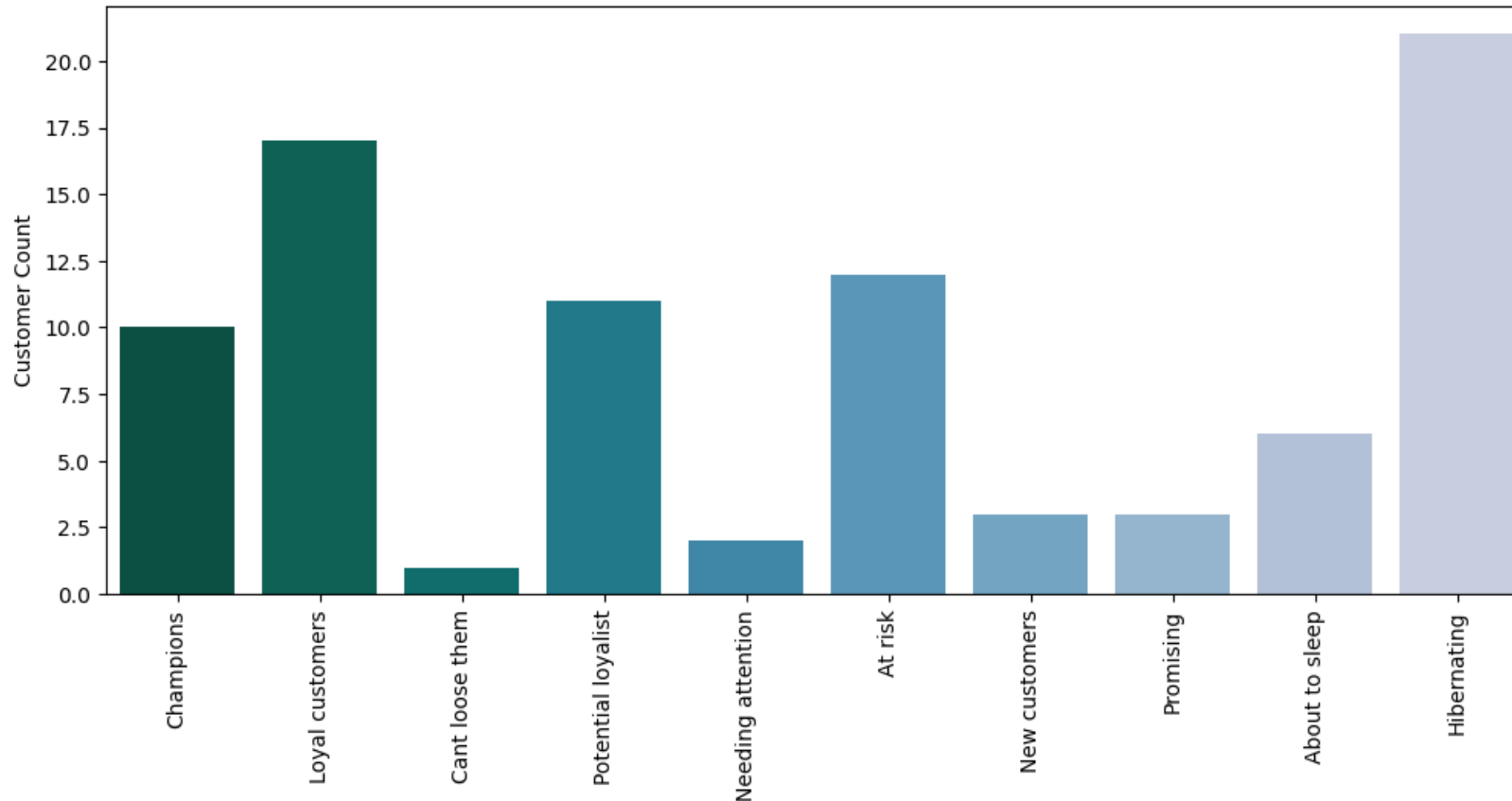


Frequency distribution give the insight that that sequentially hibernating < potential loyalist < loyal customer in the number of frequencies. It was also explained that the more loyal the customer, the lower the peak generated.



Monetary distribution give the picture that *hibernating* customer segment is dominant (shown by the peak), followed by *potential loyalist* and *loyal customers*. It's means that *hibernating* customer spent lower than other 2 segments

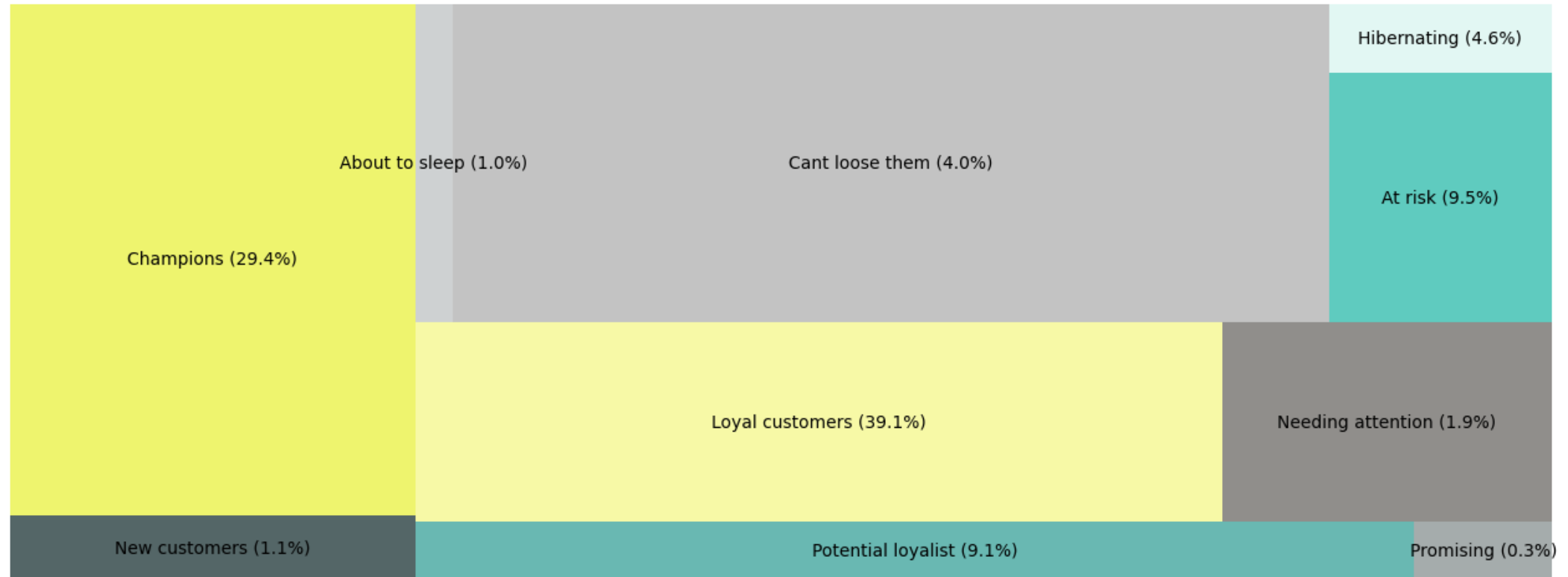
Customer Number for Each RFM Label (Segment)



From the northwind's customer that made purchase in 1997, it is dominating by the *Hibernating* segment with the total 20 customers followed by *Loyal customers* with 17 customers.

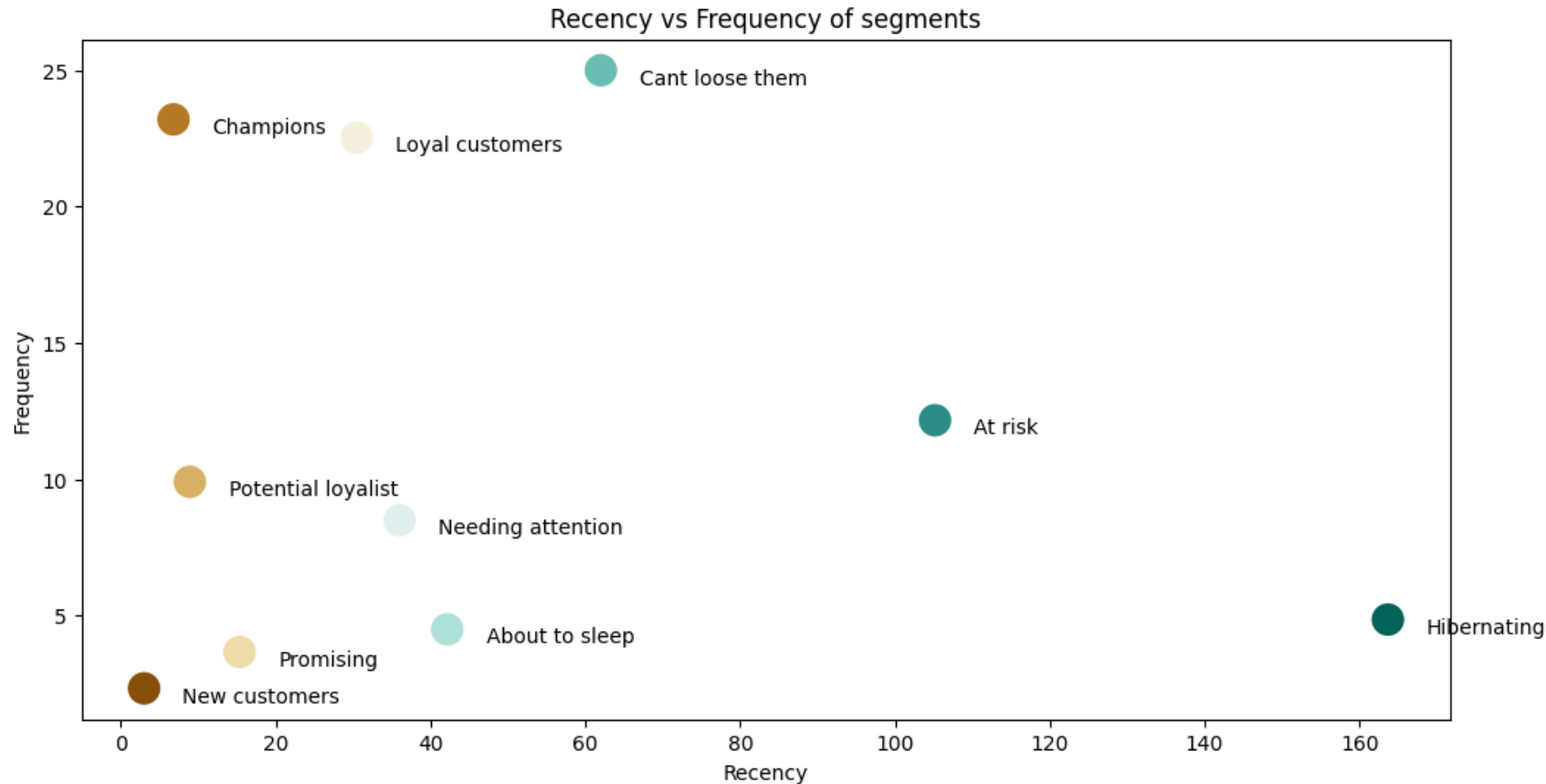
And the segments with the lowest number of customers is *Can't loose them* segments that only have 1 customers.

Monetary Share in Segments



Champions and **Loyal Customers** dominating the monetary share with the proportion 29.4% and 39.1% share.
The other segments only have maximum share below 10%.

Unit Sold by Shipper Company

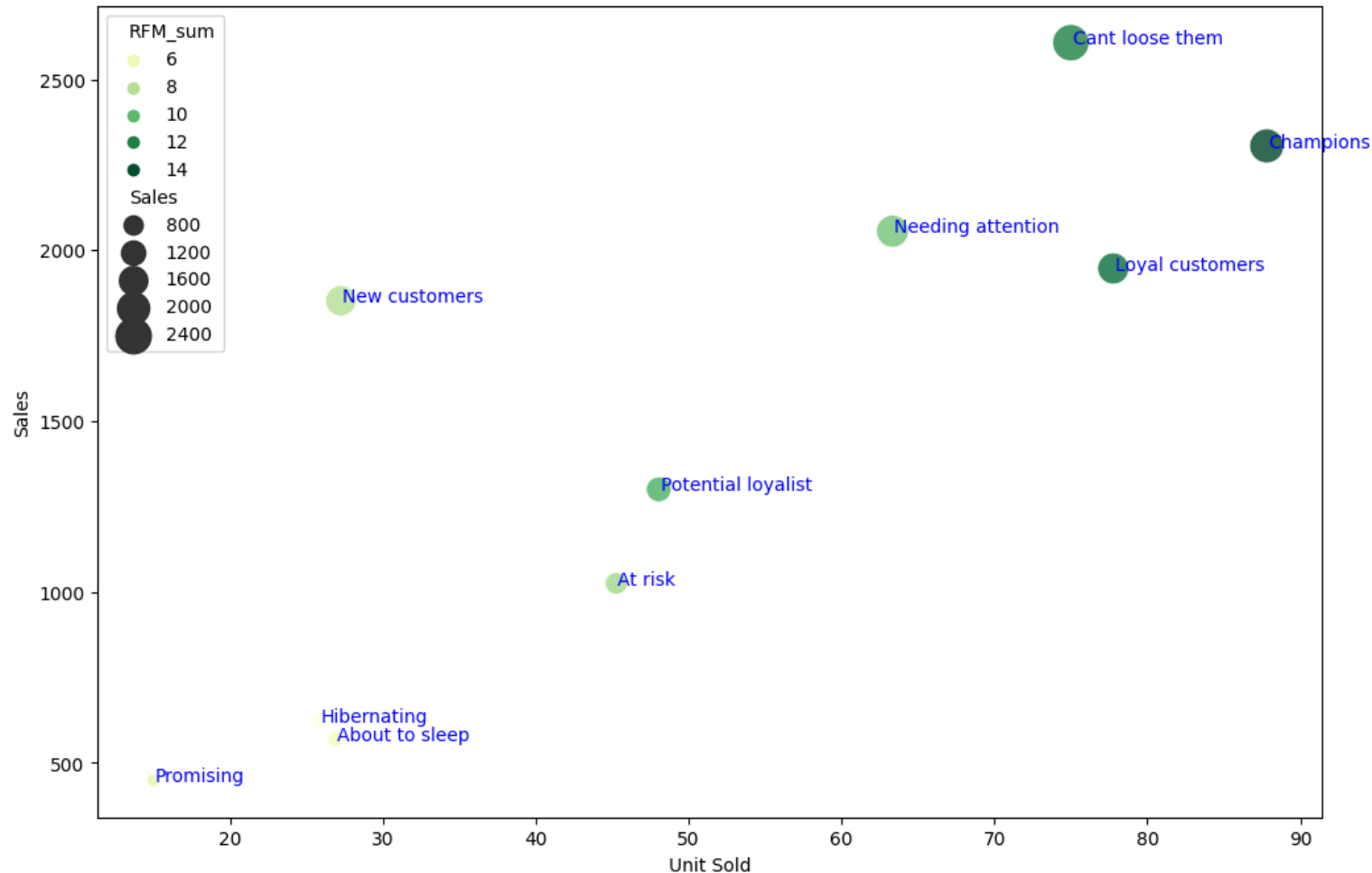


Through this chart, we can see how northwind's customer mapping is in the segment, the **higher the frequency number and the lower the recency number**, the better the segment will be.

The top 3 frequency is Champions, loyal customers, and can't loose them segment.

The top 3 recency is champions, potential loyalist, and new customers.

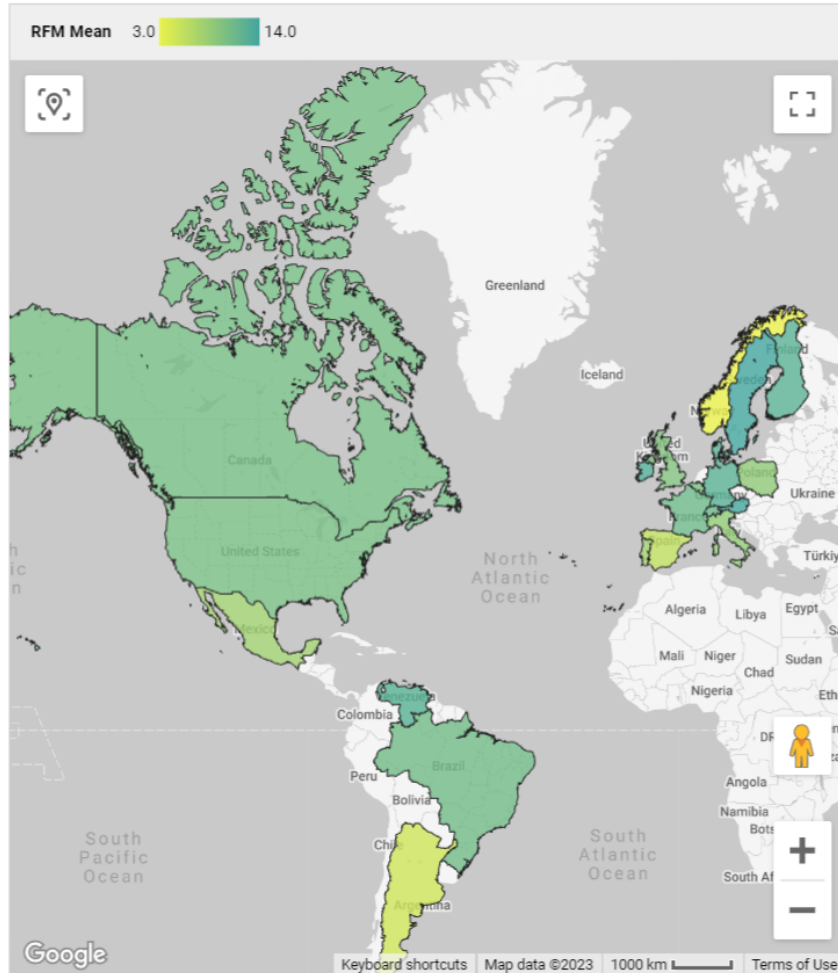
Order Size vs Sales in Segments



Through this scatter chart, we want to find out which segments can become priority customers for a specific marketing strategy, while only segments have the **potential to be prioritized in the future**.

Based on order size vs total sales (purchase) of the Champions segment, can't loose them, and loyal customers have a high intersection of total RFM and sales. This list of segments will become priority customers in the future.

RFM Score and Order Count by Country



	Country	RFM Mean ▾	Order Count
1.	Sweden	14	17
2.	Austria	13.93	21
3.	Ireland	13	10
4.	Venezuela	12.59	20
5.	Germany	12.46	64
6.	Denmark	12.41	11
7.	Finland	12.4	13
8.	Switzerland	11.28	8
9.	France	11.14	39
10.	Brazil	10.95	42
11.	Canada	10.65	17
12.	United States of America	10.52	60
13.	Belgium	10.16	7
14.	United Kingdom	9.88	30
15.	Poland	9	2
16.	Italy	8.86	15
17.	Portugal	7.89	7
18.	Mexico	7.85	12
19.	Spain	5.73	5
20.	Argentina	4.83	6
21.	Norway	3	2

- Based on geographical insight, Sweden is the top country with the highest RFM mean (low recency, high frequency and monetary).
- From the analysis in page 19 (Shipper analysis), those analysis correlated with this insight, Germany and United States is the top 2 countries order count.
- In the RFM mean ranking Germany and USA were ranked in 5 and 12, both of them can be prioritize in the future marketing strategy.
- United Kingdom has high order count but lower RFM mean, this country need consideration to up their RFM ranking.

Recommendations

1

For top 3 segments with highest RFM Mean (champions, loyal customers, can't lose them), to maintain them and raise their RFM level we need a loyalty program to increase retain orders.

2

Hibernating segments dominating the number of northwind customers, we need the specialized program such as order discount, give away, or promo to get their attention so they are not churned.

3

We need to give more attention to the United Kingdom's (UK) market, because they have the number 3 country with the most order count but they only have RFM mean below 10, special targeted marketing need to be carried out because UK is very potential to increasing their level

Thank You

Contact me



ayubindra8@gmail.com



[LinkedIn Profile](#)



[Github Link](#)