



Northwind Analysis

Gibran Brahmanta P. -
gibranbrahmanta@gmail.com

TABLE OF CONTENTS

01

Business Overview

About the chosen use case

02

Data Processing

How the data being processed to support the analysis

03

Customer Analysis

Customer behavior analysis using RFM approach

04

Product Analysis

Gross Income and each product category performance

05

Supplier Analysis

Supplier performance and action recommendations related to restock



01. BUSINESS OVERVIEW

About the chosen use case
and their objective

Chosen Use Case

- Customer Analysis
 - **Objective:** Determine how many group that can be made of the Northwind's customers to describe their behavior based on RFM (Recency, Frequency, and Monetary) metrics.
- Product Analysis
 - **Objective:** Obtain the insight regarding Northwind gross income and giving recommendation in order to increasing the gross income
- Supplier Analysis
 - **Objective:** Obtain the insight regarding the relationship between Northwind and it's suppliers and giving recommendation to fulfill demand from the customers

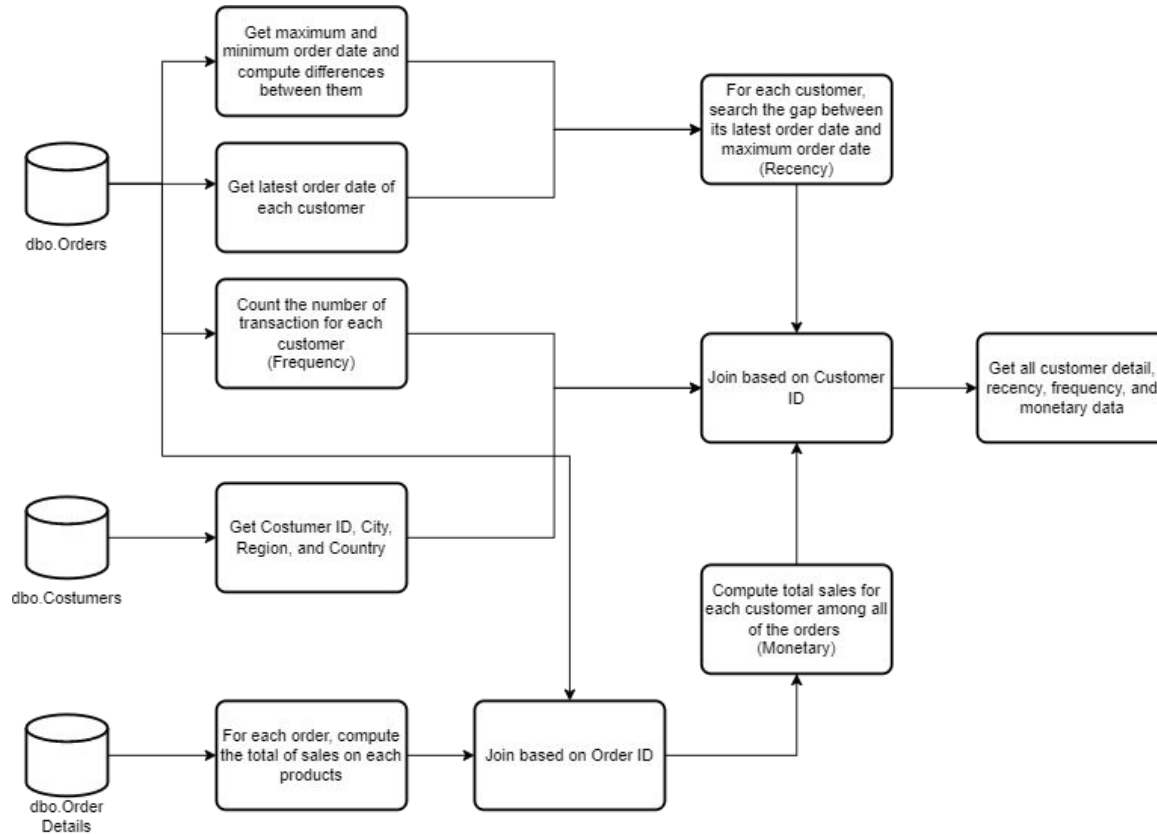
02. Data Processing

How the data being
processed to support the
analysis

Dataset Overview

- Contains **13 tables**, each table represents an entity that related to Northwind business process
- Generally, contains the **order data between 04/07/1996 until 05/06/1998** and other data related to it
- **Several tables were empty**, so those tables won't be used

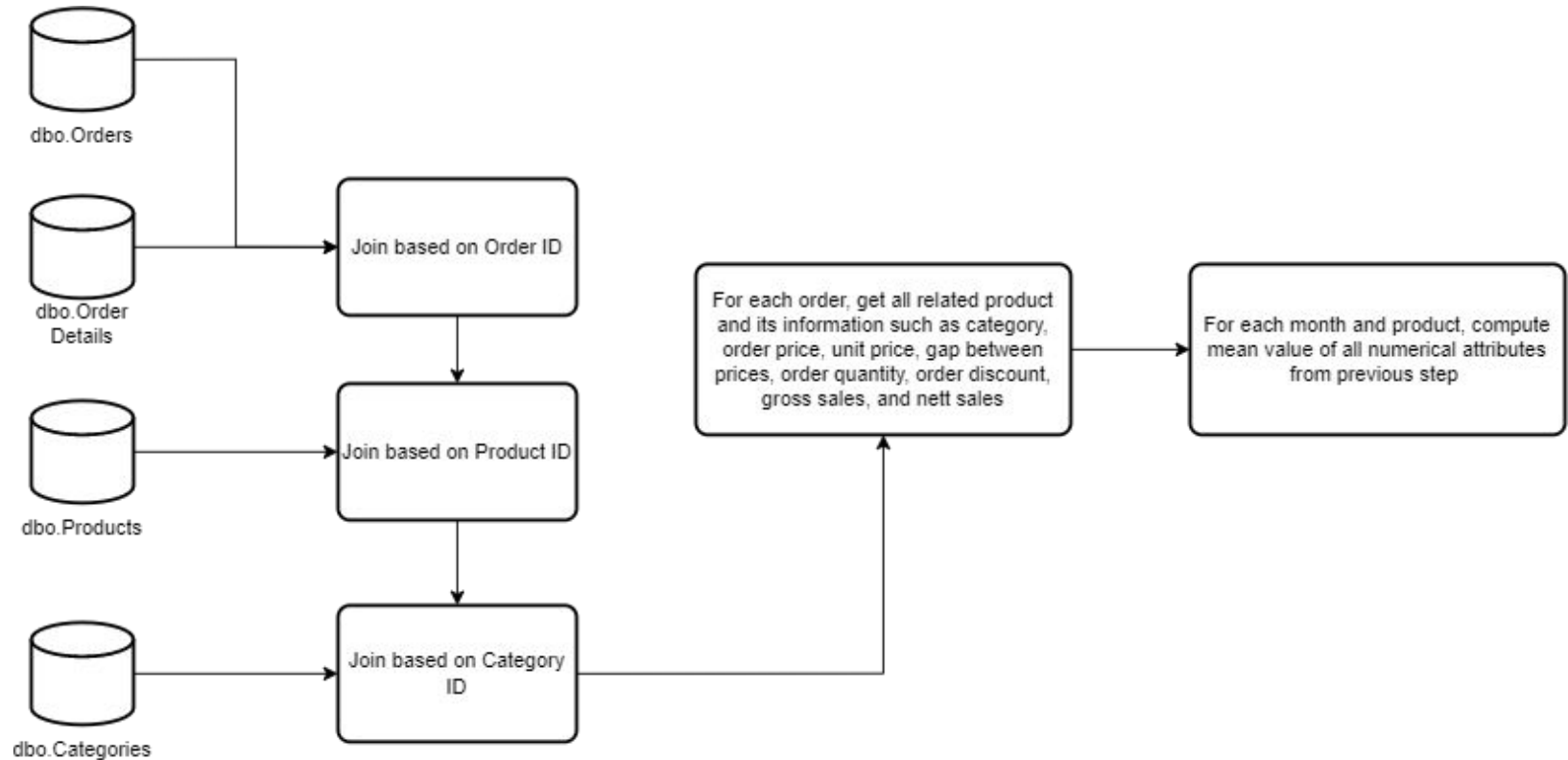
Customer Analysis Flowchart



Customer Analysis Dataset

- Contains 77 rows
- Each row determine a user behavior related to RFM metrics that are being computed like this:
 - Recency: Differences between latest customer order date and the latest order that in the dataset
 - Frequency: How many times that the customer is created an order
 - Monetary: How much the total amount of money that the customer spent
- Source code: [here](#)
- Dataset: [here](#)

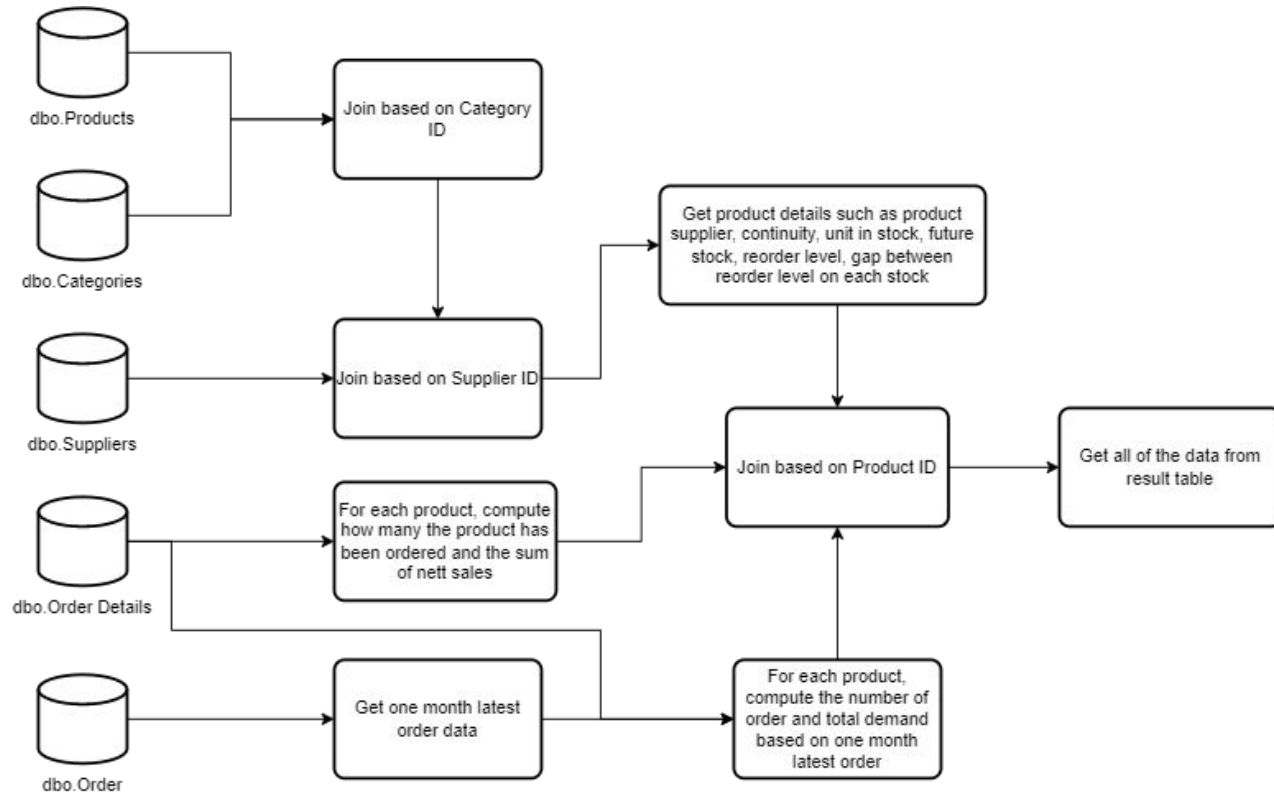
Product Analysis Flowchart



Product Analysis Dataset

- Contains 1.162 rows
- Each row contain information about a product performance (total sales, transaction count, average price, and etc) on a specific month
- Source code: [here](#)
- Dataset: [here](#)

Supplier Analysis Flowchart



Supplier Analysis Dataset

- Contains 77 rows
- Each row contains product data such as stock and demand along with its supplier
- Source code: [here](#)
- Dataset: [here](#)



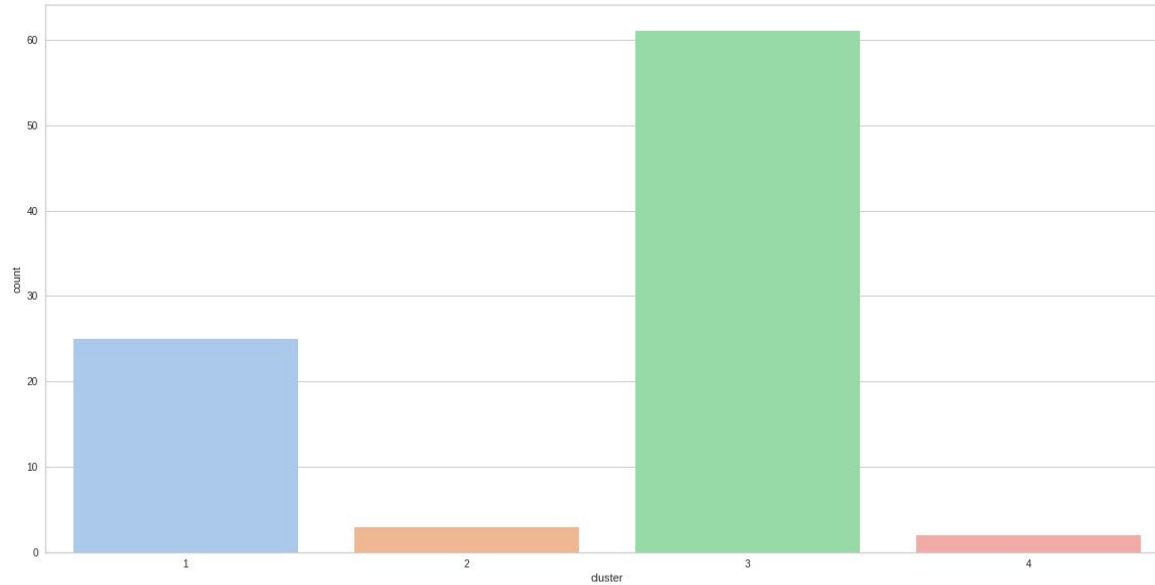
03. Customer Analysis

Customer behavior analysis
using RFM approach

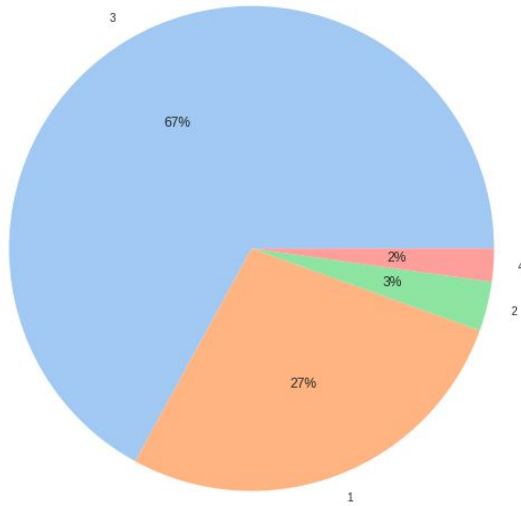
RFM Analysis

- Clustering process being carried out to divide the customers into several group based on its RFM metrics
- Using K-Means algorithm, the customers are being divided into 4 groups
 - To determine how many group that can be made, several experiments were conducted
 - The final result of the experiment is the used algorithm has Silhouette Score equals to 0.689. Therefore, we can conclude that the algorithm can separate those group well apart from each other

Distribution of Customer on Each Cluster



Distribution of Customer on Each Cluster

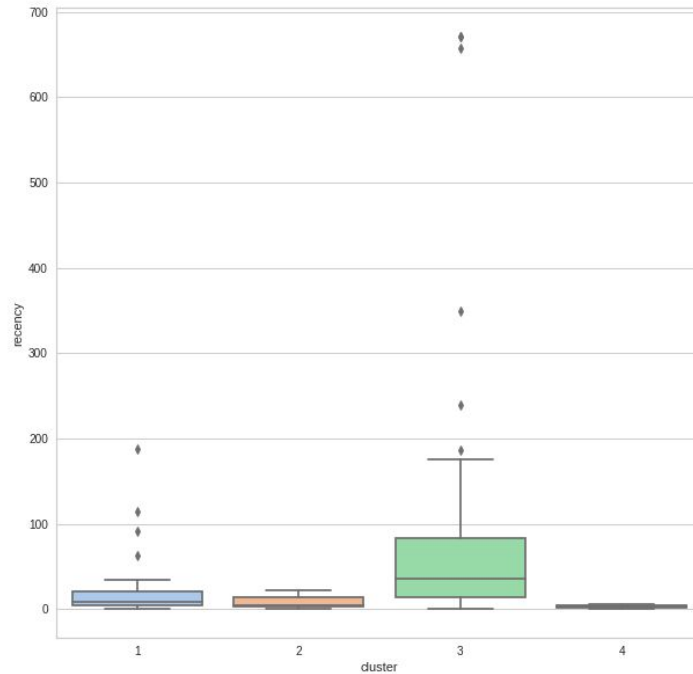


Most of the customers (67%) are being categorized as cluster 3

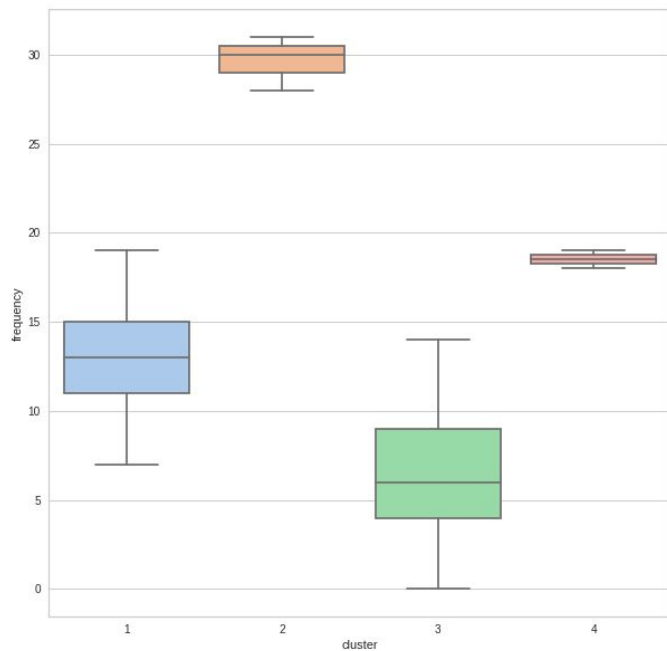
Cluster Interpretation

- In order to interpret the behaviour of customer on each cluster, further analysis must be conducted
- The analysis was carried out using data visualization, especially using box plot. Below is the ideas of the analysis
 - For each metric, visualize boxplot on each cluster. After that, aggregate the information related to each cluster based on the boxplot

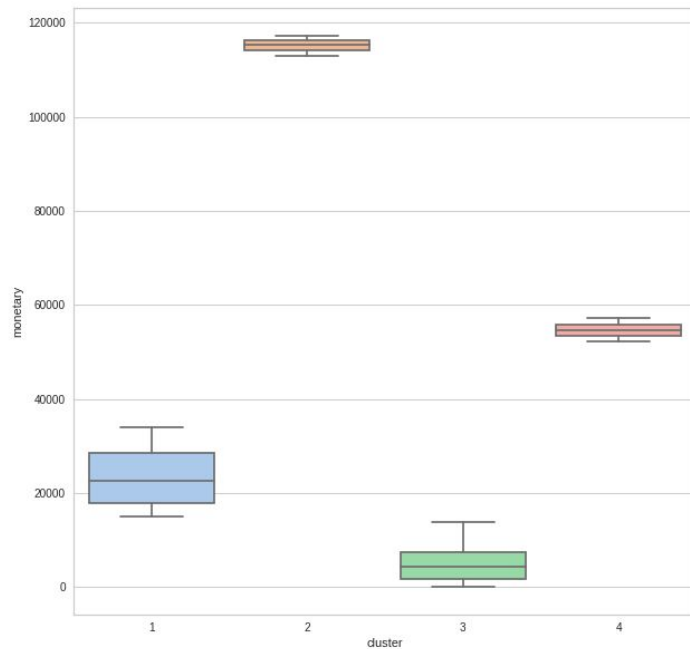
Cluster Interpretation - Recency Boxplot



Cluster Interpretation - Frequency Boxplot



Cluster Interpretation - Monetary Boxplot

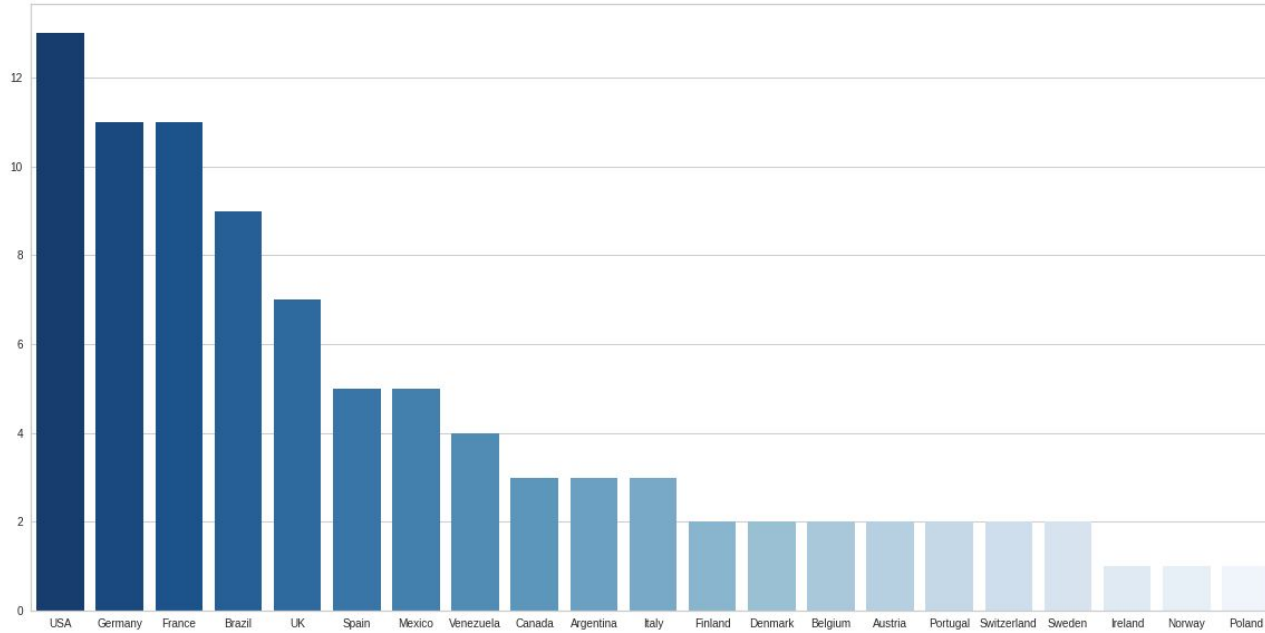


Cluster Interpretation - Result

Based on several boxplot that have been displayed, we can conclude that **attribute that can differentiate** each customer are **frequency** and **monetary**. On those two attributes boxplot, there were such a **huge difference in the distribution of each cluster**. Meanwhile on recency attribute, the difference of the distribution on each cluster is very small. Using the information that are displayed on the boxplot, we can summarize each cluster as:

- **Cluster 1: Medium-Low Customers**, because they have a fairly small number of transactions and spent a few amount of money, but not as few as Cluster 1 (Best Customers)
- **Cluster 2: Best Customers**, because they already done many transaction and have spent a lot of money
- **Cluster 3: Low Customers**, because they have small number of transaction and spent a little amount of money
- **Cluster 4: Medium-High Customers**, because they have a large enough number of transactions and spent a lot of money, but not as much as Cluster 1 (Best Customers)

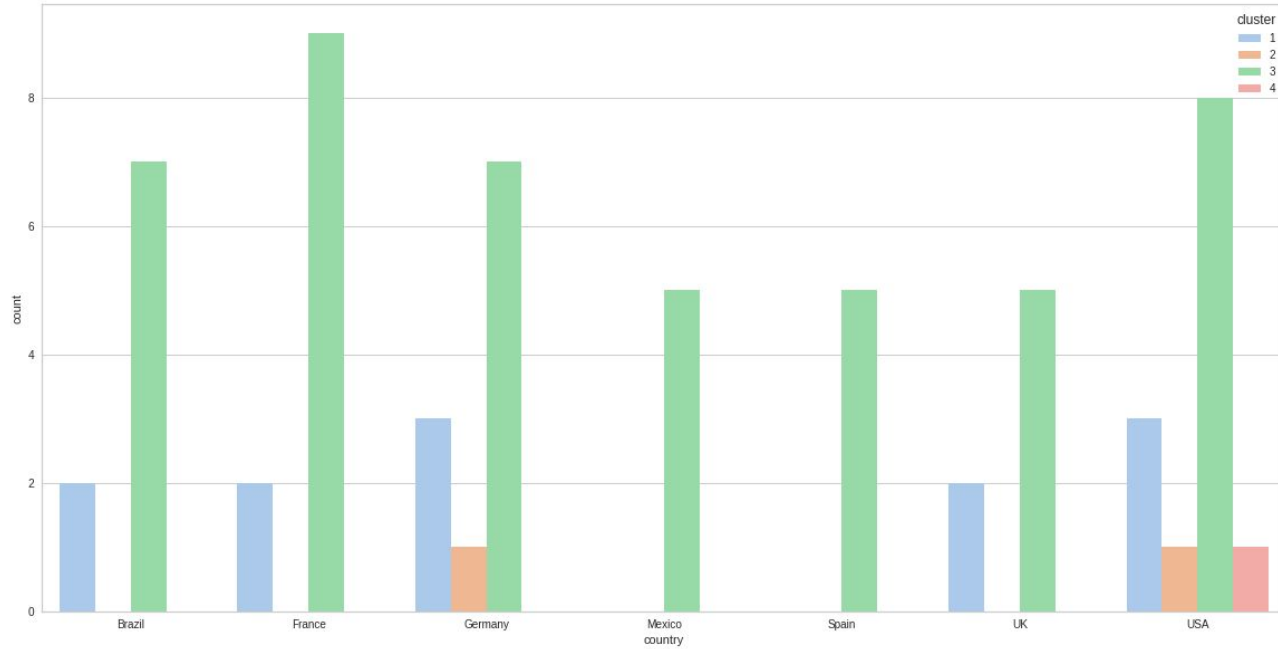
Customer Distribution Based On Country



Customer Distribution Based On Country

- From the graph that has been displayed before, we can conclude that the number of customer on each country are varying. Therefore, we can conduct further analysis based on this several informations:
 - Number of countries that being top 75% countries based on customer count: 7
 - Total customer at top 75% countries based on customer count: 61
 - Ratio (Number of customers in 7 country/number of all customer): 67.03%

Customer Distribution Based On Cluster and Country



Customer Distribution Based On Cluster and Country

- From the graph that has been displayed before, we can conclude that **on every country, most of its customer is on Cluster 3 (Low Customer) group.**
- On **each country beside Mexico and Spain**, they still **have customer on the other cluster** group although still being dominated with Cluster 1 (Medium-Low Customer).
- Therefore, **there must be effort to improve customer behaviour** in term of recency, frequency, and monetary **especially on Mexico and Spain** where all of their customers are in Cluster 3 (Low Customer) group

04. Product Analysis

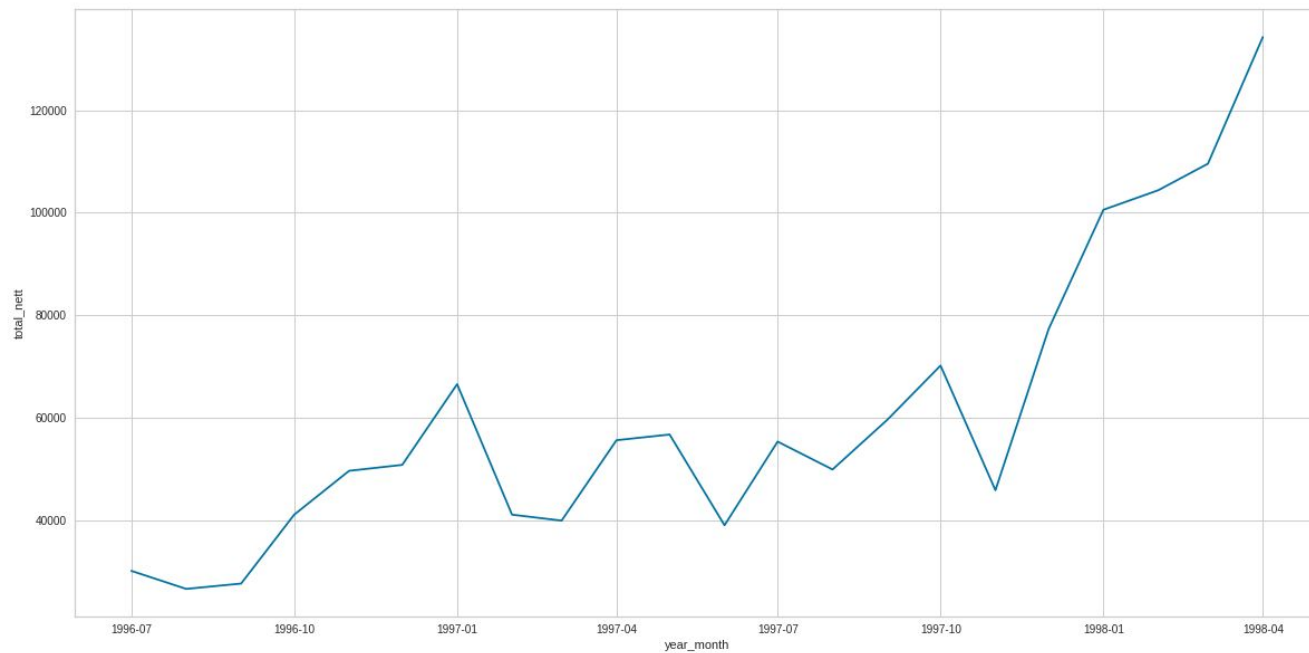
Gross Income and each
product category
performance

Product Analysis

By using the dataset that has been created, several analysis were carried out to in order to achieve the analysis objective such as

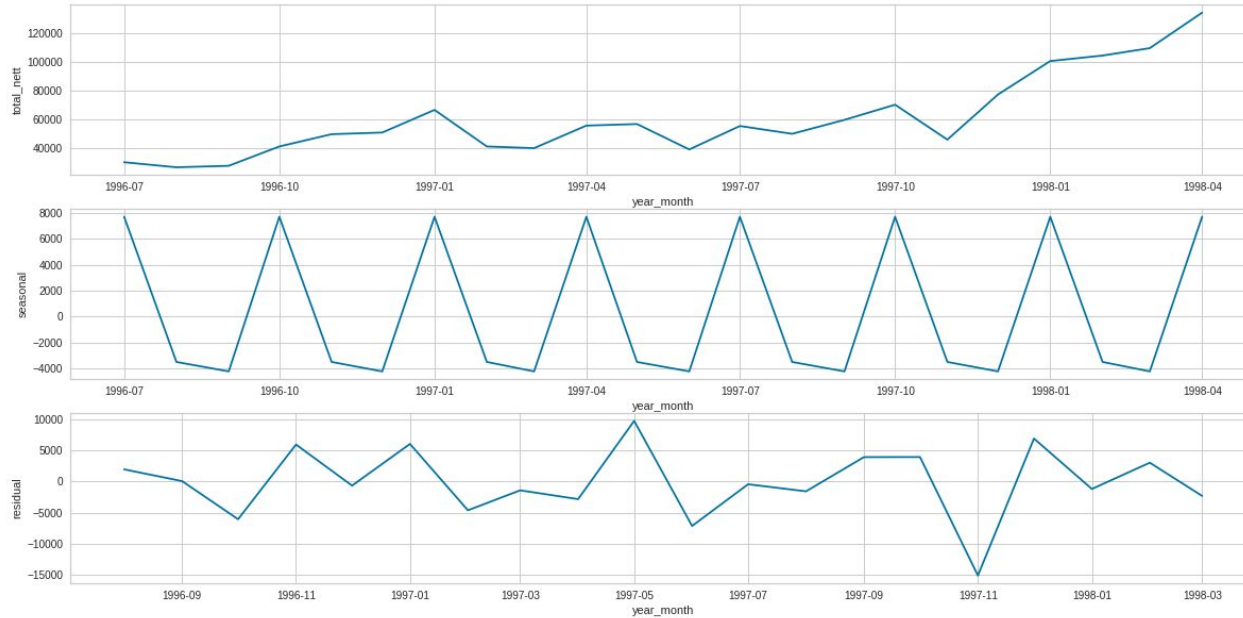
- Northwind's gross income trend that being calculated using sum of total sales on each month
- The amount of loss that Northwind has because differences in product and order price
- Product category popularity based on transaction count

Northwind Gross Income



Gross Income on Each Month

Northwind Gross Income

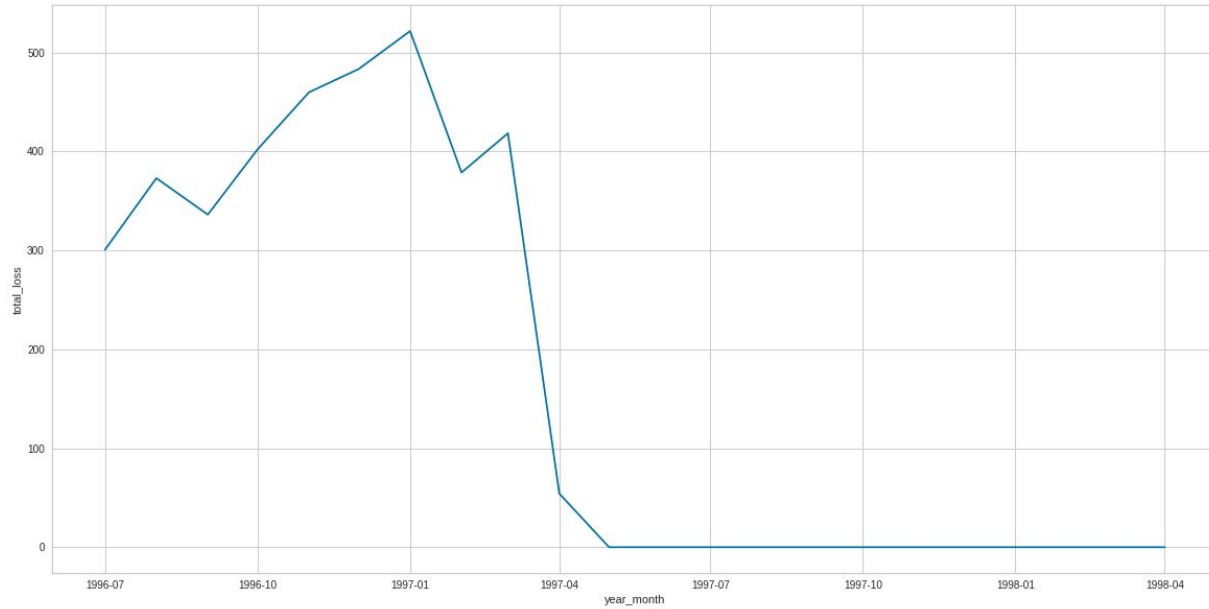


Decomposed Version of Gross Income on Each Month

Northwind Gross Income

- From several graphs that has been displayed, we can conclude that **the gross income that has been obtained is fluctuating** on each month.
- However, in general if we take a look **on the trend, the gross income that has been obtained is increasing**.
- Also, the gross income that has been obtained **have a seasonal pattern which is repeated every 3 months**.
 - At the seasonal pattern, the gross income will have highest growth on the first month and the growth will be decreasing on the other months.

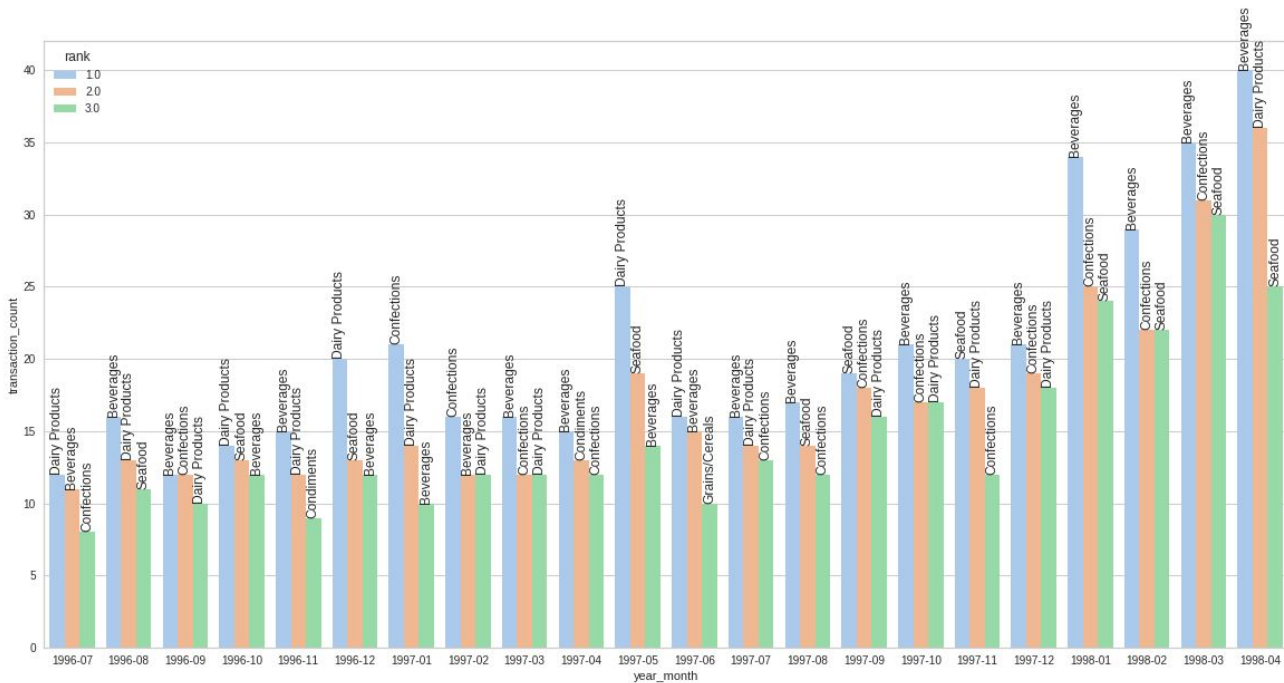
Northwind Amount of Loss



Northwind Amount of Loss

- From the graph that has been displayed before, we can conclude that **in the early period of the data, there were several losses**
- It is being caused by **differences price value on the product data and order data**
- The value is increasing for a short time period but also it remain decreasing for the next time period so that in current data, the loss remains zero

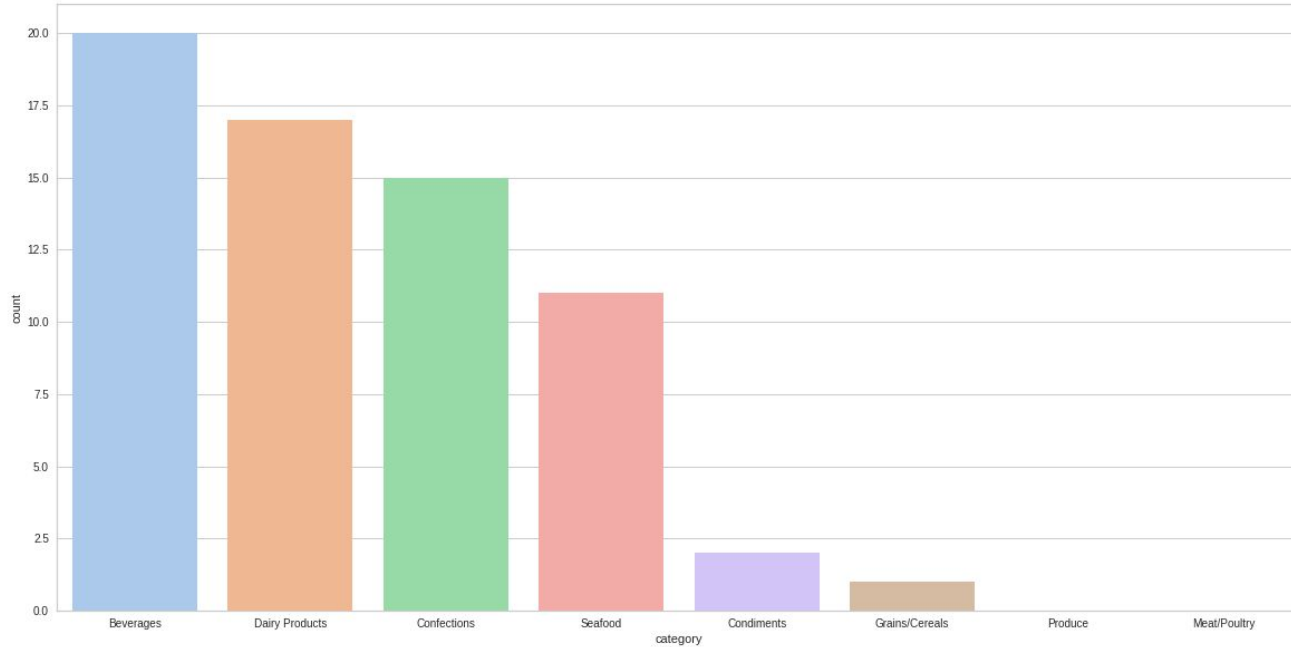
Most Popular Product Category



Most Popular Product Category

- From the graph that has been displayed before, we can conclude that were **differences on the top-3 category that being most purchased on each month.**
- In order to able to give recommendation, further analysis must be concluded by calculating the number of occurrences of each category on the top-3 category on each month data

Most Popular Product Category



Most Popular Product Category

- Based on the computation result and the graph that has been displayed before, we can conclude that '**Beverages**', '**Diary Products**', and '**Confections**' is the **most popular product category** because their number of occurrences is being the most among all of the product category
- **Recommendation:** A marketing strategy that related to those product category could be done every three months in order to improve the gross income

05. Supplier Analysis

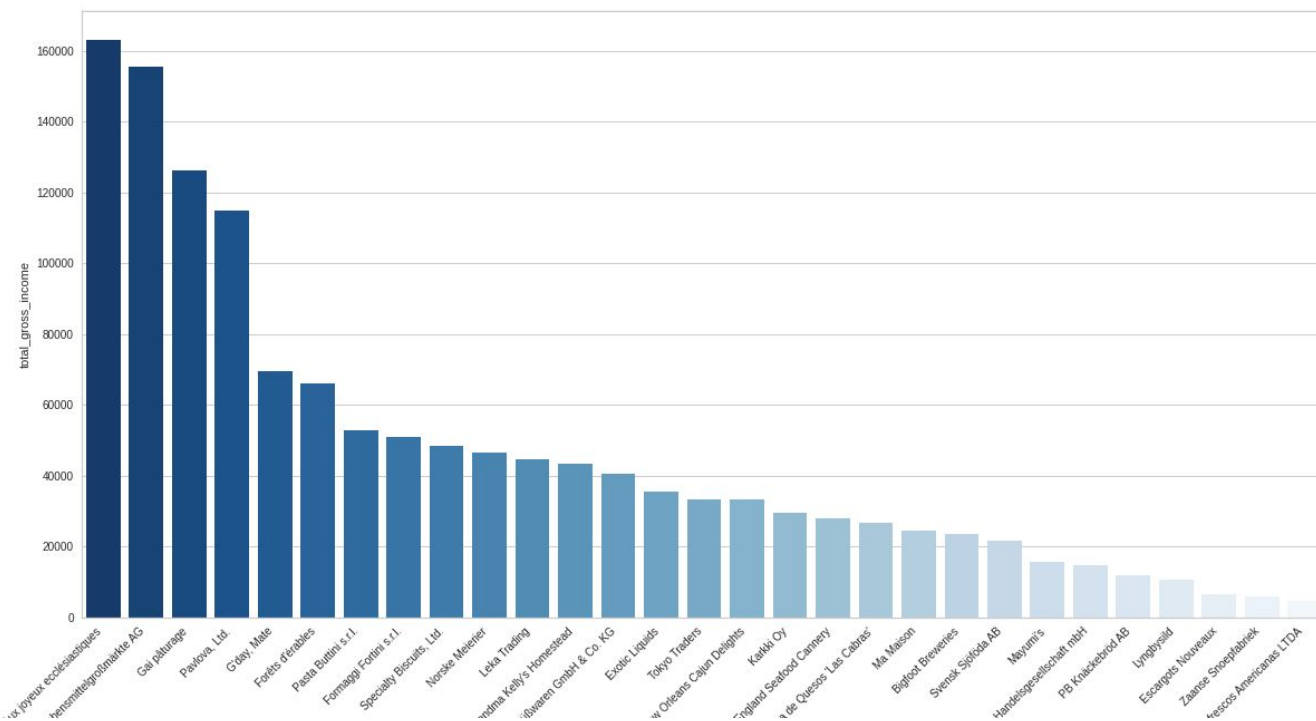
Supplier performance and
action recommendations
related to restock

Supplier Analysis

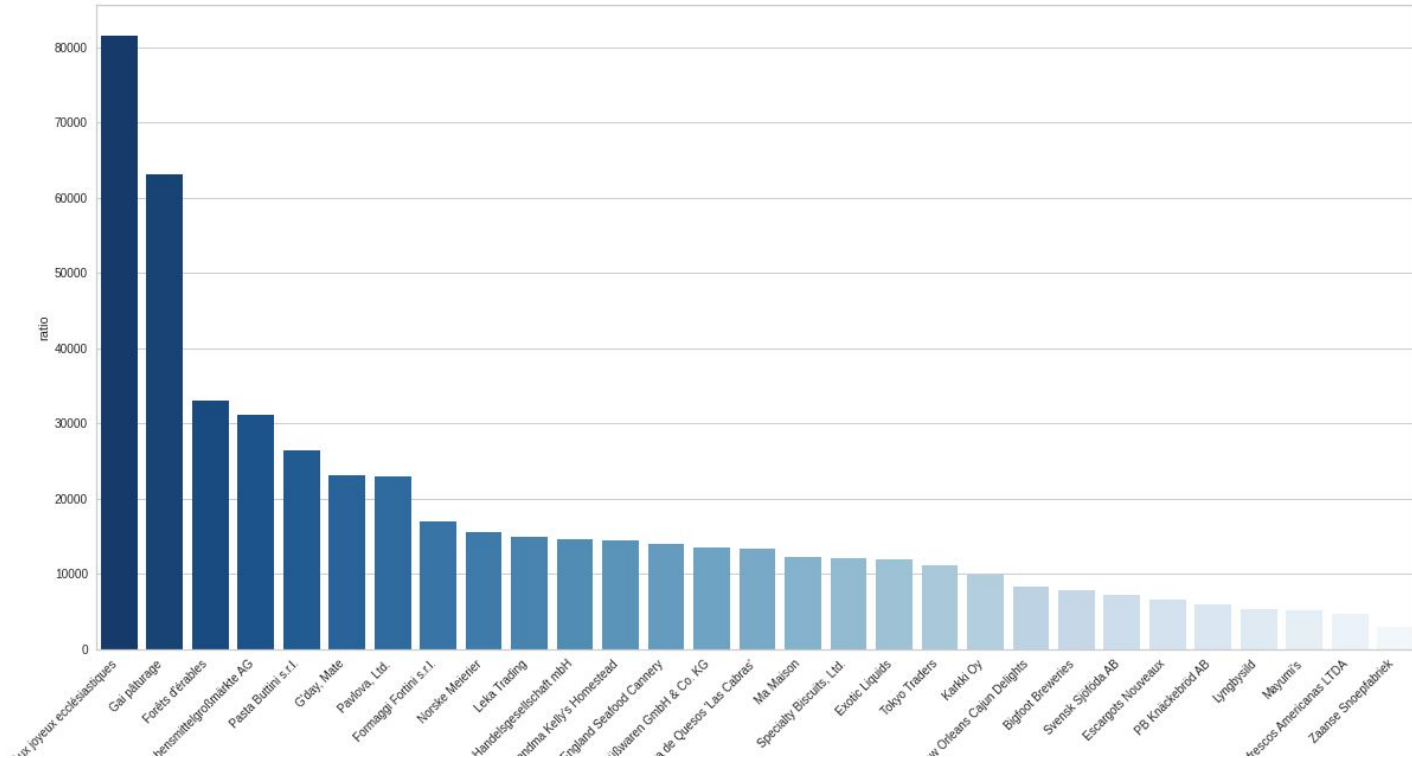
By using the dataset that has been created, several analysis were carried out to in order to achieve the analysis objective such as

- Supplier performance based on two point of view
 - Total gross income that has been obtained
 - Ratio between total gross income and number of product
- Restock amount to fulfill demand
 - Based on order level
 - Based on previous month demand

Supplier Performance - Total Gross Income



Supplier Performance - Ratio



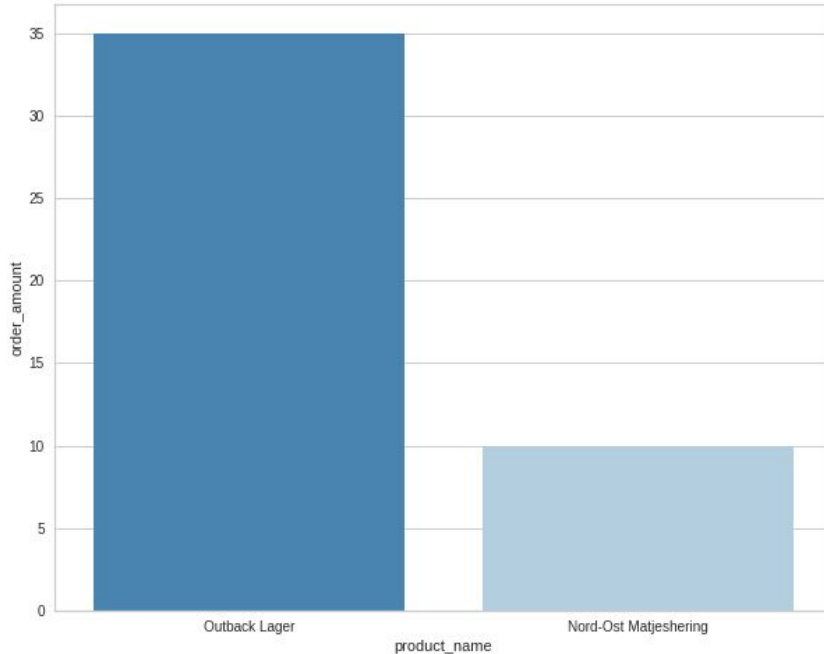
Supplier Performance

- From those two graphs above, we can conclude that **there were differences in the order of suppliers** based on total gross income that has been obtained through their products and the ratio between total gross income and the number of products that they were provided.
 - **Higher ratio** on a supplier means that **every product that are being supplied by it have a higher total gross income value.**
- Therefore, we could know that there were several suppliers have higher total gross income but lower ratio and vice versa, and it can be used to indicated whether supplier has best performance based on several point of view

Restock Analysis

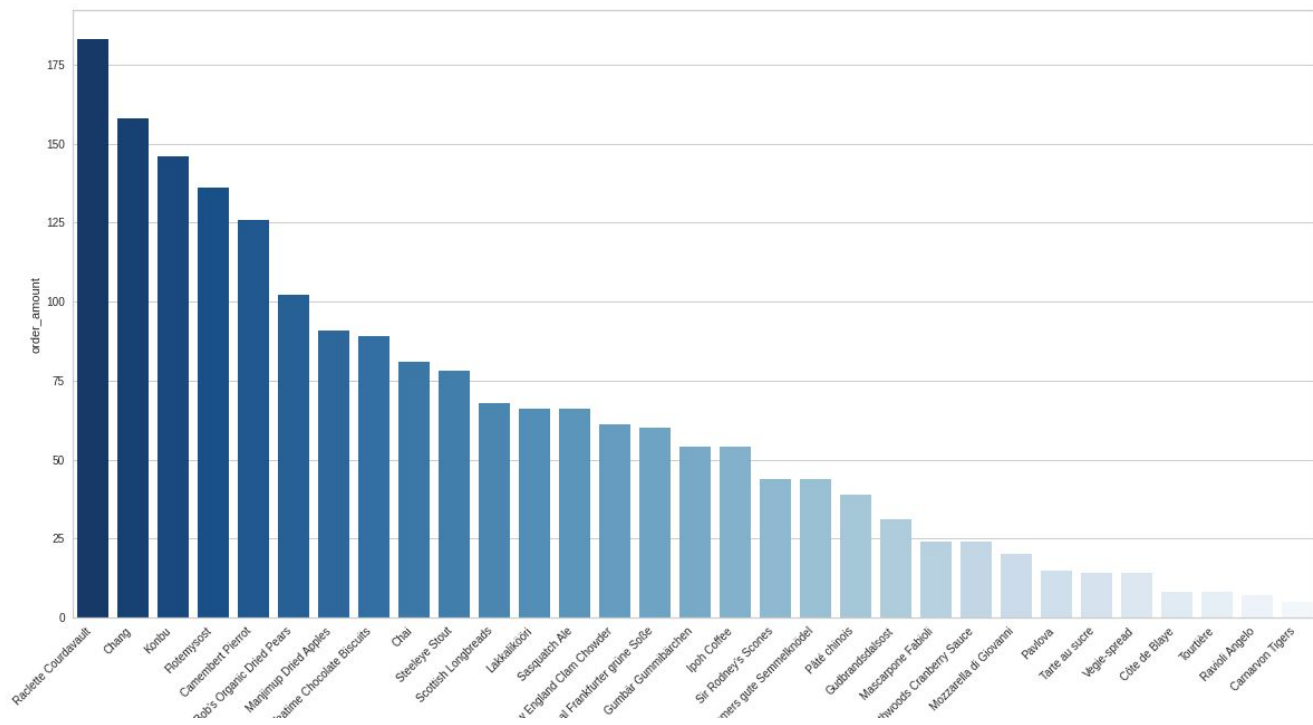
- **Used assumption:** When a product is having a restock, the number of stock must be added until it reach its reorder level if the number of reorder level is higher than the prediction of next month demand (previous month demand multiplied by 0.8). And if not, the number of stock must be added until it reach its previous month demand multiplied by 0.8 and added by 5
- **Note that** 0.8 and 5 is several value to keep the product not being overstock/understock to fulfill the next month demand. Future research can be conducted to determine those value

Restock Analysis - Reorder Level



There were two products that need to be restock as soon as possible because their number of stock are lower than their number of reorder level. The number of stock must be added by a specific amount to fulfill next month demand

Restock Analysis - Previous Month Demand



Restock Analysis - Previous Month Demand

- From the graph that has been displayed before, we could know that **there were several products that need to be restock** because their number of stock are lower than their number of stock that must be achieved in order to fulfill next month demand. The number of stock that must be added to each product are varying.



THANK YOU

Gibran Brahmana P.

gibranbrahmana@gmail.com

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik

Please keep this slide for attribution