Mini Project Data Engineering

Project Report

Iswara Jay Jr.

Outline :

1. Objectives
2. Data Preparation
3. Key Insights
4. Conclusion

Project Objectives

Product Analysis

Covering our most popular products traded, and what is our top customers prefer.

RFM Analysis

Helps us understand our consumer-base better, and their classification characteristics.

Shipper Analysis

Covering how our current shippers performance compares with each other.

Data Preparation

A. Dataset Used

Northwind sample database from **LINK,** containing order transaction from
Northwind Traders from 04/July/1996 to 06/May/1998

Relevant tables then went through transformation process described in (C) to be
visualized with PowerBI.

B. Calculated Fields Definition

- **ShippedStatus :** shipped order = 1, not yet shipped = 0 .
- **LateStatus** **:** 1 if shipped date >= required date, 0 if shipped date < required date.
- **Recency** : 07/May/1998 – Order Date (in day)
- **Discounted** : 1 if order done with discount applied, 0 if don't

- **GrossProfit** **:** Total sales – Freight
- **FreightMarkup :** GrossProfit/Freight (**higher is better**)
- All sales value counted by including discount.

C. Data Processing Flow (OrderAgg)

Orders characteristic table aggregated by OrderID
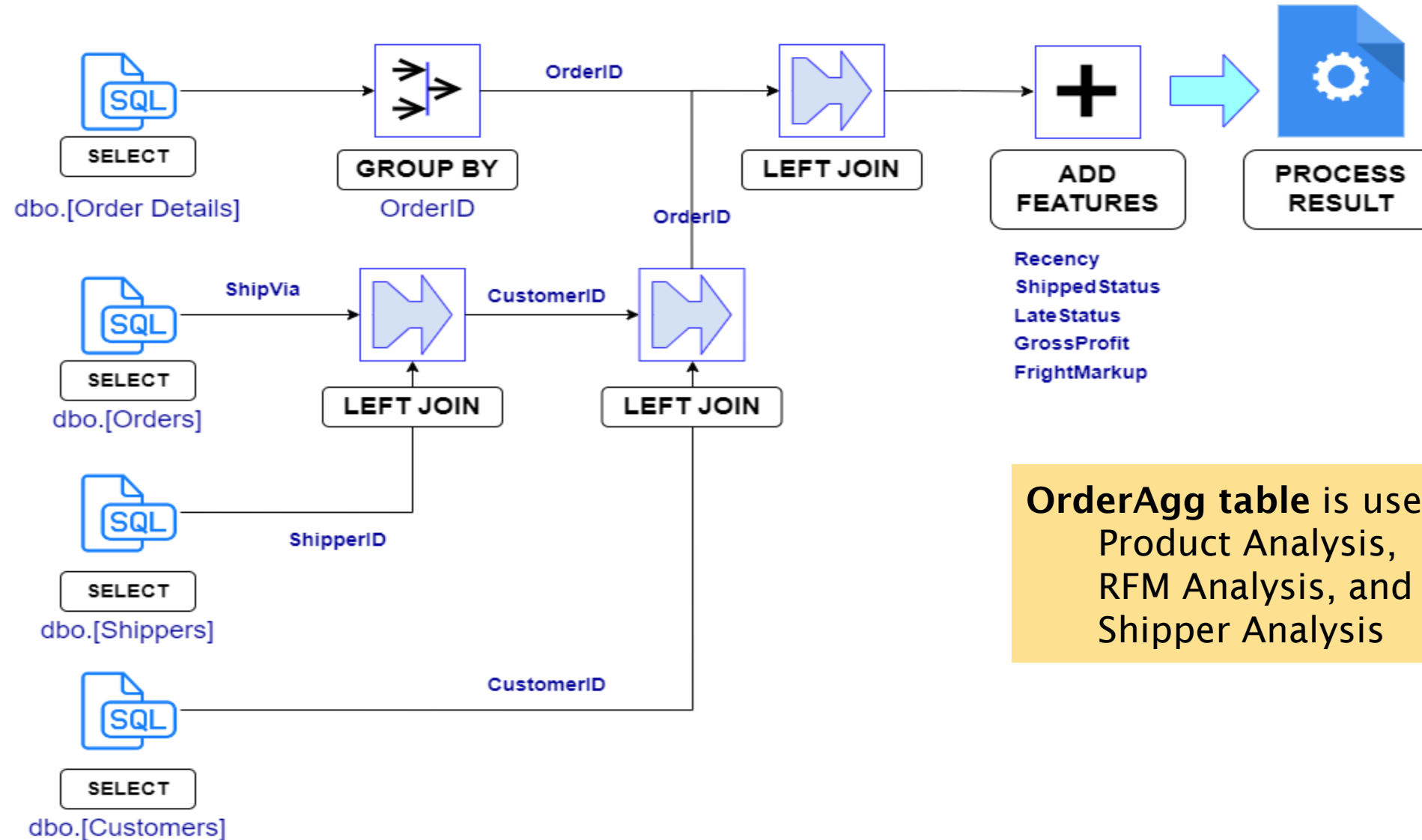
OrderAgg table is used in :
Product Analysis,
RFM Analysis, and
Shipper Analysis

C. Data Processing Flow (OrderDetailed)

Orders characteristic table without aggregation, used to view the details for each orders (products composition)

OrderDetailed table is used in :
Product Analysis
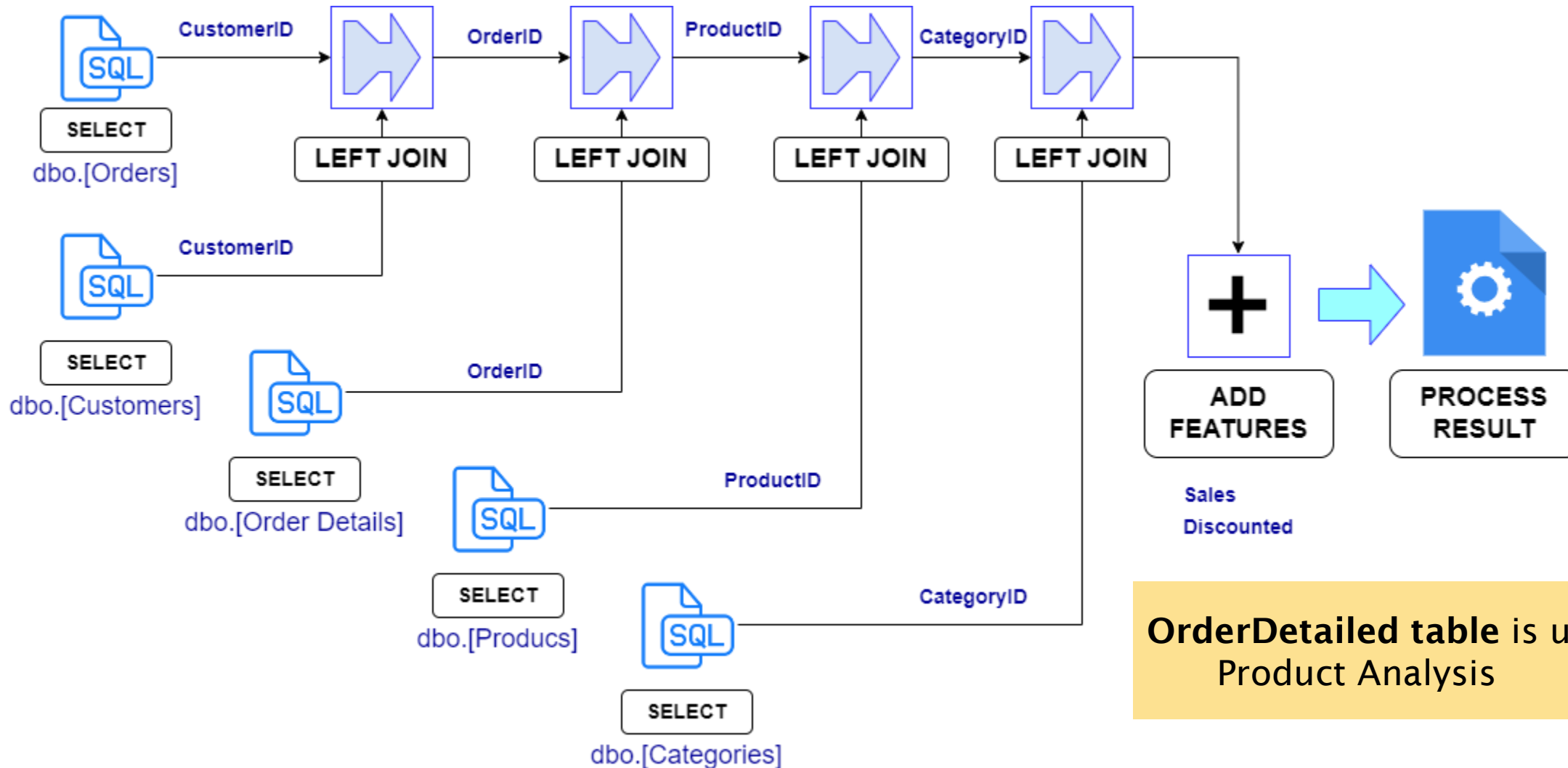
Key Takeaways from Product Analysis

Total Quantity by Product Category

Total Sales by Product Category
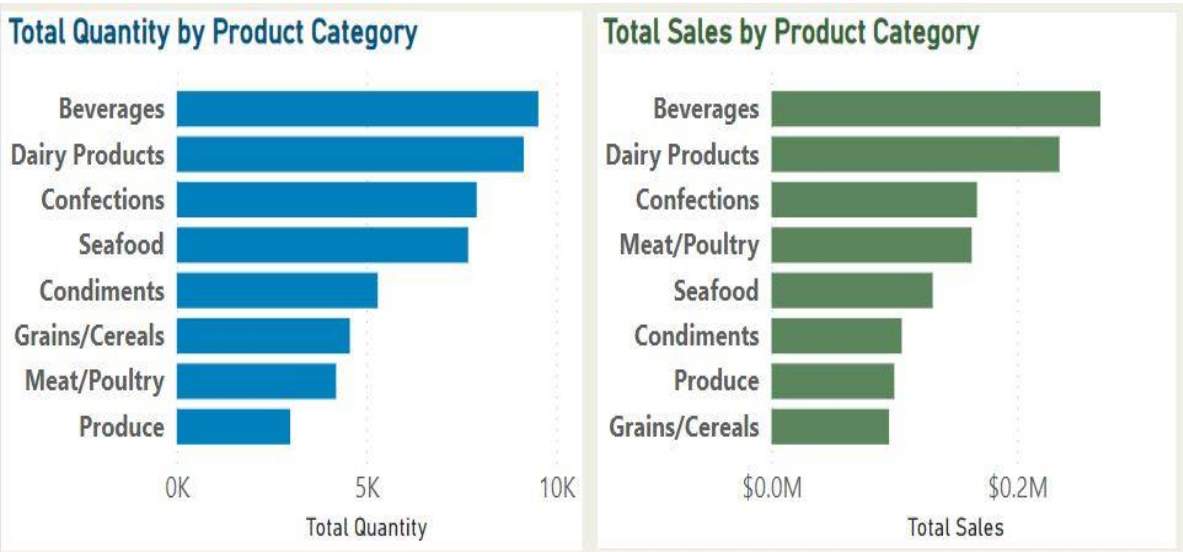
Beverages and Dairy Products create the most sales value, followed by Confections and Meat/Poultry in similar standing.
**This presumably supported by our most valued customers**

Discounts mostly don't incentivized our current customers to prefers one product compared to another.

Several products in each categories went way above their competitions (Sales-wise) :

Product Name	Category	%Sales of Product Category
Cote de Blaye	**Beverages**	**54.18%**
Manjimup Dried Apple	Produce	41,83%
Raclette Courdavault	Dairy Products	30,34%
Gnocchi di nonna Alice	**Grains/Cereals**	**44,49%**
Thuringer Rostbratwurst	**Meat/Poultry**	**49,3%**

Key Takeaways from RFM Analysis
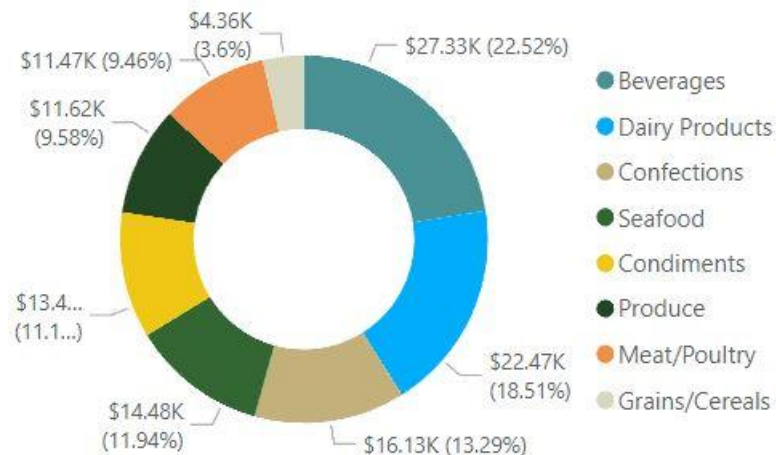
Number of Customer by RFM Segment
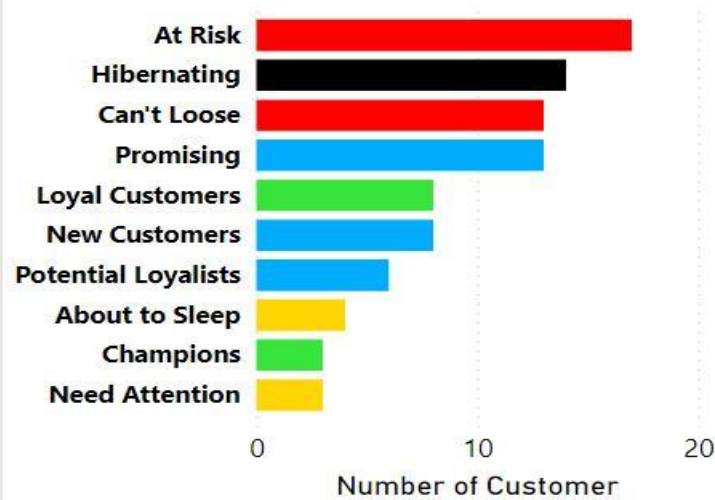
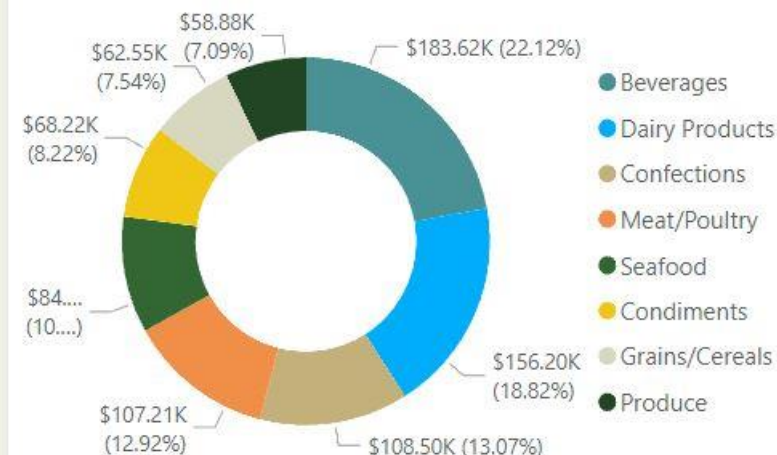**Significant amount of our key customers are in Inactive/near Inactive zone.**

Red/black shade

However a significant amount also can be classified as Promising

Blue/green shade

Sales Distribution of Promising Customer

- Beverages — $27.33K (22.52%)
- Dairy Products — $22.47K (18.51%)
- Confections — $16.13K (13.29%)
- Seafood
- Condiments — $14.48K (11.94%)
- Produce — $13.4... (11.1...)
- Meat/Poultry — $11.62K (9.58%)
- Grains/Cereals — $11.47K (9.46%)
- $4.36K (3.6%)

Sales Distribution of High Value Customer

- Beverages — $183.62K (22.12%)
- Dairy Products — $156.20K (18.82%)
- Confections — $108.50K (13.07%)
- Meat/Poultry — $107.21K (12.92%)
- Seafood — $84... (10....)
- Condiments — $68.22K (8.22%)
- Grains/Cereals — $62.55K (7.54%)
- Produce — $58.88K (7.09%)

Our Promising and Highly Valued Customer-base have similar taste in the types product they want to buy

Key Takeaways from Shipper Analysis

There's quite a tradeoff going on :

Our **best Shipper (United Package)** in terms of markup (highest bang for the buck) also has the highest Late Shipment Rate, however **it's markup is 3 times better than Federal Shipping.**

Average Freight Markup by Shipper

Shipper	Freight Markup
United Package	158
Speedy Express	99
Federal Shipping	53

Freight Markup

Proportion of Late Deliveries by Shipper

Shipper	Proportion of Late Shipment
United Package	4.91%
Speedy Express	4.82%
Federal Shipping	3.53%

However a significant amount of shipping to USA and Germany still uses Federal Shipping.

So being able to convert good amount of these trades to be shipped with better Shipper will contribute a lot to our profit.

Gross Profit by Shipped Country

● Federal Shipping ● Speedy Express ● United Package

USA
Germany
Austria
Brazil
France
UK
Venezuela
Sweden
Canada
Ireland

$0.0M $0.1M $0.2M

Gross Profit

Conclusion

1. Beverages and Dairy is out top product categories, and for some categories there are "Champion Items" which at least constitute 40% of total sales of their respective category

2. Discounts mostly don't affect our consumer-base product preference.

3. Significant amount of key customer being inactive/near-inactive needs to be tackled quickly.

4. Our promising and key customers have similar taste in product categories they want to buy.

5. Federal Shipping is highly underperforming in terms of markup compared to the other shipper, while still being used significantly in our most shipped regions.