# Executive Summary

Customer shopping behavior is crucial for business success. Understanding and leveraging it helps businesses operate more efficiently, especially in inventory control. This summary offers key insights for a wholesale company serving Horeca and Retail clients across Lisbon, Oporto, and other regions. Analyzing shopping patterns across product categories enables data-driven decisions to enhance customer satisfaction and optimize inventory management. To explore the wholesale dataset, which includes customer spending across six product categories, client types, and geographic information, we employed cluster analysis to uncover these insights:

- Clustering

  Before conducting the clustering analysis, we removed outliers to focus on identifying patterns among the majority of clients. However, further analysis will still be conducted on these outliers to better understand their unique behaviors and contributions. This technique identifies natural groupings within the data. In this case, we aimed to group customers based on the following parameters:

  - **Spending per category:** The average amount one customer spends on a product.
  - **Channel:** Customers are divided into two channels: **Horeca** and **Retail**.
  - **Region:** Customers are categorized into three regions: **Lisbon**, **Oporto**, and **Others**.
  - **Spending power:** A newly created metric based on the total revenue and the number of clients within each cluster. It is calculated as: *Spending Power = Total Revenue in a Cluster / Number of Clients in that Cluster*

We segmented the customers into three distinct groups and uncovered the following key insights:
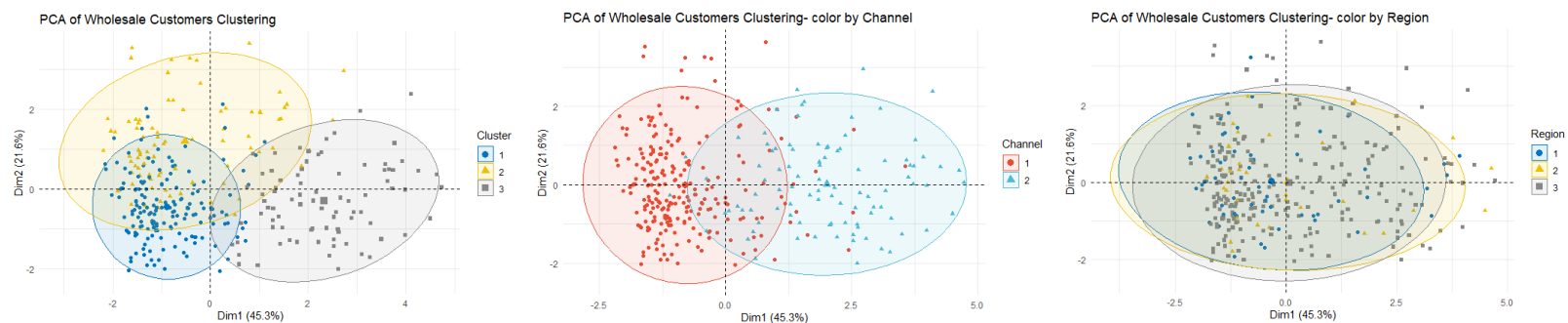
The key differentiator between the clusters is based on **"Channel."** Data points in Cluster 3 are primarily from Channel 2, while Clusters 1 and 2 mainly consist of customers from Channel 1. In contrast, the distribution of regions is spread widely across all clusters, indicating that **"Region"** may not be a significant factor in the clustering. Therefore, we conclude that **"Channel"** is more closely related to the clustering results than **"Region"** or the combination of **"Channel and Region."**
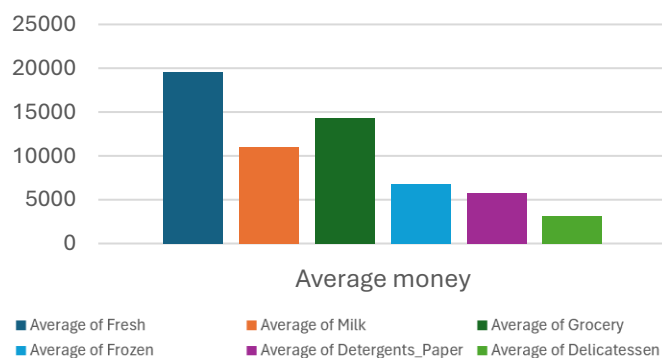
- **Tim Horton (Cluster 1)**: #Clients: 169 / Revenue generated: $2,561,696 / Spending power: $15,158 / Favorite product: Fresh / Majority channel: Horeca (channel 1)

    - Description: These are the highest number of clients with least spending power. Customers in this group tend to spend more on fresh products, with an average expenditure exceeding $5,000, generating 18% of total revenue. Compared to the other two clusters, they generally spend less across other product categories.

- **Starbucks (Cluster 2):** #Clients: 83 / Revenue generated: $2,852,779 / Spending power: $34,371/ Favorite product: Fresh & Grocery / Majority channel: Horeca (channel 1)

    - Description: These clients are less in numbers but have adequate spending power. Customers in this group generally spend on fresh and grocery products with heavy emphasis on fresh with an average expenditure exceeding $5,000, generating 20% of total revenue.

- **Walmart (Cluster 3):** #Clients: 80 / Revenue generated: $2,675,309 / Spending power: $33,441 / Favorite product: Grocery & Milk & detergent_paper / Majority channel: Retail (channel 2)

    - Description: These are retailers and have similar spending power like the Starbucks (cluster 2). They have different preferred categories. Customers in this group typically spend more on grocery, milk, and detergent products, with an average expenditure exceeding $5,000, generating 18% of total revenue.

Average money spent per cluster
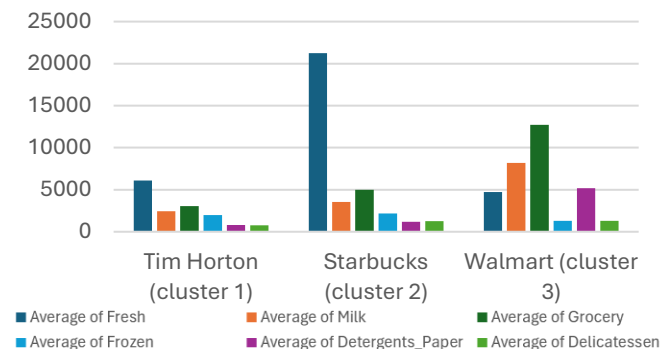
- **Ikea (Outliers)**: #Clients: 108 / Revenue generated: $6,529,716 / Spending power: $60,460 / Favorite product: Fresh & Grocery / Majority channel: Both Horeca & Retail

    - Description: These clients are the revenue-generating "cash cows" with very high spending power, operating in both retail and Horeca channels. They purchase large quantities of products but focus on only a few product categories, contributing roughly 45% of total revenue. Further analysis of this segment

Outlier average money spent

is necessary to get a better understanding of their characteristics. The purchases of the outliers are primarily concentrated in fresh products, groceries, and milk, making them key contributors across both channels.

Recommendations

- Tim Horton & Starbucks (Cluster 1 and 2): Customers in these clusters exhibit similar spending patterns, with the highest expenditure on fresh products and groceries. Both clusters come from Horeca. Therefore, in areas with a higher concentration of Horeca-type customers, the distributor should focus on stocking more fresh and grocery products, particularly fresh items, to meet the heavy demands of the clients.

- Walmart (Cluster 3): Customers in this cluster spend more on grocery, milk and detergent products and this cluster is mainly from Retail. As a result, in areas with more retail-oriented customers, the distributor should prioritize inventory for these product categories to better serve the retail market.

- The company should reduce its resource allocation on deli and detergent products, as these items have relatively low demand, and reallocate funds towards fresh products, which see significant client spending. Since the majority of clients (~70%) fall under the Horeca segment, efforts should focus on strengthening operations and supply chain pipelines to streamline the purchasing experience.

- If the company is considering investing in a new warehouse, Lisbon or Oporto may not be ideal locations, as the majority of revenue is generated from other regions.

Average spending by Regions

- If we are able to convert 50% of Tim Horton (cluster1) clients into Starbucks (cluster2) clients, company XYZ would be able to see an increase of $1.6 Millions in their revenue.

Report Prepared by:
Ko Jen Kang | Anantha Narayanan Balaji | Aastha | Ameer Akhtar