# Data Analysis on Spending Patterns of Clients & Recommendations

Executive Summary

Narae Kang , Pranvi Setia, Rebecca Meyer, Shubham Garg, Yun-Chien Yen

## Business Goal: Find meaningful spending patterns of our clients

Our data analysis team analyzed XYZ company's wholesale data to give actionable insights on client spending patterns. These insights will allow XYZ to better meet their client's needs. These are some questions we seek to answer from the data, followed by our recommendations.

**Q. What are spending patterns based on the channels?**

- We see that the Horeca (Channel=1) clients spend on average 2 times more on frozen products than Retail (Channel=2) clients. The mean spending on fresh products by Horeca clients is also 1.5 times higher than Retail.
- Retail channel clients buy on average 3 times more milk products, 4 times more grocery products, and 9 times more detergent & paper products than Horeca channel clients.

**Q. What are spending patterns based on regions: Lisbon, Oporto, and Other Regions?**

- Clients in Lisbon (Region 1) tend to spend a mediocre amount across all categories.
- Clients in Oporto (Region 2) tend to spend more on grocery, frozen, and detergent & paper products.
- And clients in other regions (Region 3) tend to buy more fresh, milk and deli products.

**Q. Is there a correlation between Product Categories?**

- **Grocery and detergent & paper products are highly positively correlated** with a correlation coefficient equal to 0.92. This means that the clients who purchase more on grocery products also spend more on paper & detergent products, and vice versa.
- Milk and grocery products also have a high positive correlation coefficient equal to 0.73, which also indicates similar spending across these two categories.

**Q. What are spending patterns across clients?**

- There are a handful of clients who spend a large amount annually on one or more product categories, as compared to the remaining client base. For example, the client's spending 112,151 on fresh products - this amount spent is the largest (by 1.5 times the second largest amount) in the fresh products category. We can find similar trends across all the product categories.

**Q. Are there natural groups of clients based on spending characteristics?**

- We utilized the K-means clustering technique to cluster our clients into three big groups:

- **Cluster 1** consists of a few clients (n = 41) that belong only to the **Retail channel** and have **higher mean expenditures across Detergent & paper, grocery and milk products**.
- **Cluster 2 has most clients** (n = 338), more than two-thirds of which belongs to the **Horeca channel** and consists of the **low spending group**.
- **Cluster 3** is also dominated by **Horeca clients** and is characterized by a **higher mean spending on fresh and frozen** products.

```
##          Fresh       Milk     Grocery       Frozen Detergents_Paper Delicatessen
## 1  0.07612741 0.26781048 0.31604316 0.03156052       0.34516819   0.04830937
## 2  0.07297318 0.05468315 0.06047771 0.03993169       0.04513443   0.02316561
## 3  0.31612924 0.08086923 0.07042464 0.11874307       0.02650803   0.06815213
```

Mean spending in clusters
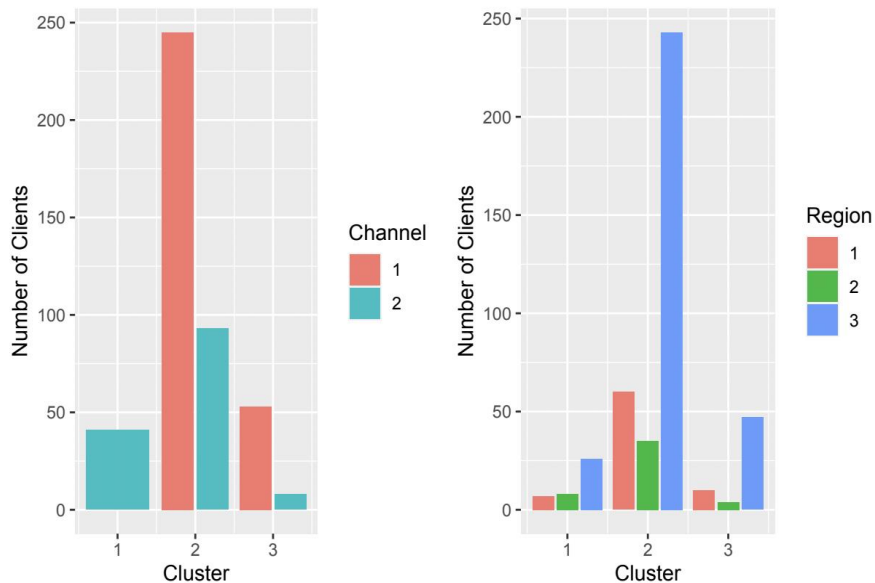


Number of clients in clusters as per region and channel

## Recommendations

**Bundling grocery, milk, paper & detergent products can lead to an increase in spending.**

- XYZ distributors can offer them product bundles consisting of grocery, milk, and paper & detergent products, especially for clients in cluster 1. They can offer discounts on bulk purchases of these bundles, and they can also extend more credit to these clients upon purchase of these products together to improve their customer relations and loyalty. We are not recommending bundling of other products with these categories because all these customers belong to the retail channel, and they might not be interested in investing in other products. XYZ distributors can extend such an offer to all its retail clients.

**Optimizing delivery routes to minimize trips to low-spending client segments can cut costs.**

- For clients belonging to cluster 2, XYZ distributors should focus **primarily on cost reduction,** since this group consists of many low-spending clients, with most clients from region 3. To cut costs they can save on two fronts: **inventory and transportation.**
- They can keep the inventory level low in region 3 (except for fresh and frozen products, which is in high demand among cluster 3 clients) to match demand and reduce inventory carrying costs.
- They can also reduce transportation costs by **reducing the frequency of delivery trips** by pooling clients together that are located nearby and making deliveries to them in the same delivery run.

**Ensure product and delivery capacity to meet high demand for fresh and frozen products**

- Clients belonging to cluster 3 spend the most on the fresh and frozen product categories, which are time-sensitive products and require timely delivery with high product quality. Most clients in cluster 3 come from the Horeca channel and are in other regions. XYZ can reduce transportation costs by optimizing their supply chain for these products.
- XYZ should ensure they have enough inventory in their region 3 (other regions) distribution centers to meet the high demand for fresh and frozen products.
- XYZ should ensure they have enough trucks and drivers for the Horeca channel to consistently provide timely deliveries.

**Customize marketing efforts based on channel specific needs**

- By customizing marketing campaigns to clients based on channel specific needs, XYZ can optimize their marketing efforts, saving money on marketing costs and potentially increasing client purchases.
- XYZ should emphasize detergent & paper, grocery, and milk products in their marketing campaigns to clients in the retail channel, since retail clients spend the highest in these categories.
- XYZ should emphasize fresh and frozen products in their marketing campaigns to clients in the Horeca channel, since clients with the highest annual expenditures on fresh and frozen products mostly come from the Horeca channel.

**Further data analysis on the location of clients can be used to optimize distribution centers**

- Further data analysis can be performed on data pertaining to the location of clients in other regions (Region 3) and the distance between distribution centers to optimize distribution center locations.