



## Data Analysis Competition



Name	ID
Ali Alsaihati	202034820
Hashem Almubarak	202012260

## Table of Content

Introduction .....	3
Category Insights.....	4
Customer Segment Analysis.....	5
Correlation Analysis .....	7
Channel-specific Insights.....	8
Regional Analysis.....	9
Overall Recommendations.....	10
Note .....	11

# Introduction

Montajat, a leading supermarket chain operating across multiple regions in Saudi Arabia, recognizes the importance of gaining deeper customer insights to drive strategic decisions in an evolving retail landscape. To support this objective, Montajat has provided us with a large dataset containing annual spending records of customers categorized by region, channel, and product categories. Our task is to conduct an exploratory data analysis on Montajat's customer transaction records to derive actionable business intelligence.

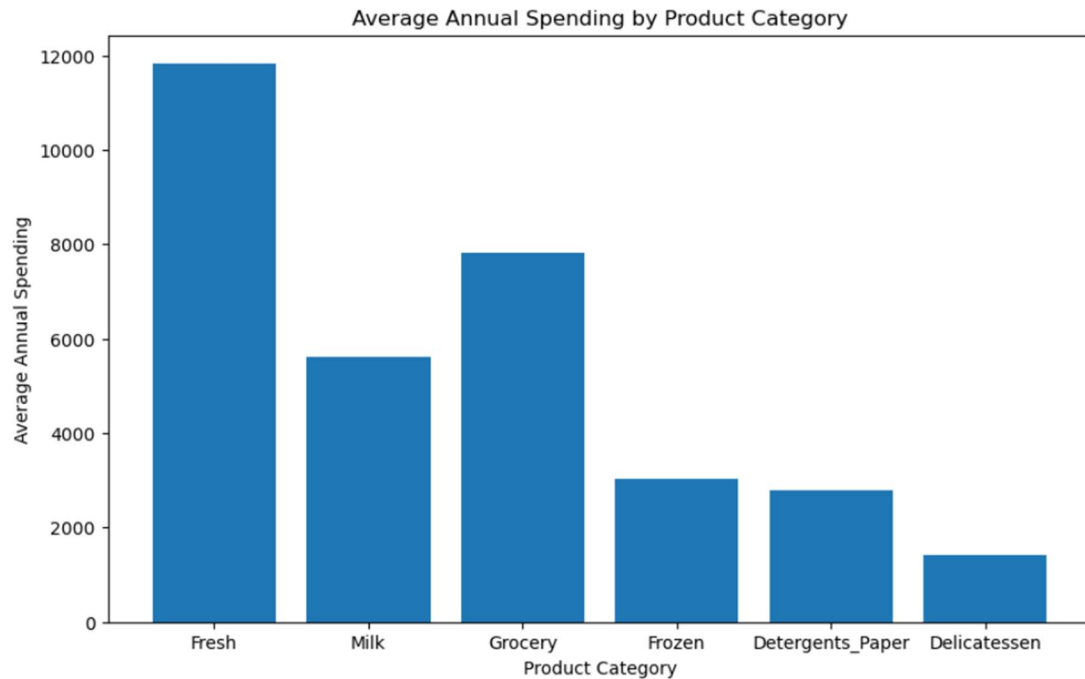
In this case, we will employ various analytical techniques to examine patterns within the dataset and provide Montajat with valuable insights into their customer base. Our analysis will encompass the following key areas:

1. **Data Cleaning:** We will ensure data uniformity and address outliers and missing values to maintain data integrity.
2. **Descriptive Analytics:** By analyzing average spending behaviors and identifying high-priority categories, we will gain a comprehensive understanding of Montajat's customers.
3. **Visualizations:** We will utilize visualizations to uncover spending distributions across different regions and channels, enabling Montajat to identify regional preferences and channel-specific trends.
4. **Segmentation Analysis:** Through segmentation analysis, we will profile top spending groups, allowing Montajat to tailor their marketing strategies and offerings to specific customer segments.
5. **Correlation Assessment:** We will explore correlations between product categories and other attributes, providing insights into cross-category purchasing patterns and potential opportunities for upselling or bundling strategies.

The insights derived from these evaluations will have far-reaching implications for Montajat. They will assist in portfolio optimization, enabling Montajat to refine their product offerings based on customer preferences. Additionally, the findings will inform marketing resource allocation, regional expansion planning, and targeted acquisition strategies, driving Montajat's growth and profitability.

Furthermore, the regular analysis of updated customer information will empower Montajat to promptly recognize market shifts and adjust long-term strategies accordingly. By optimizing their segmentation approach, distribution network design, and customer experience roadmap based on these insights, Montajat will enhance their competitive advantage and ensure customer satisfaction.

## Category Insights



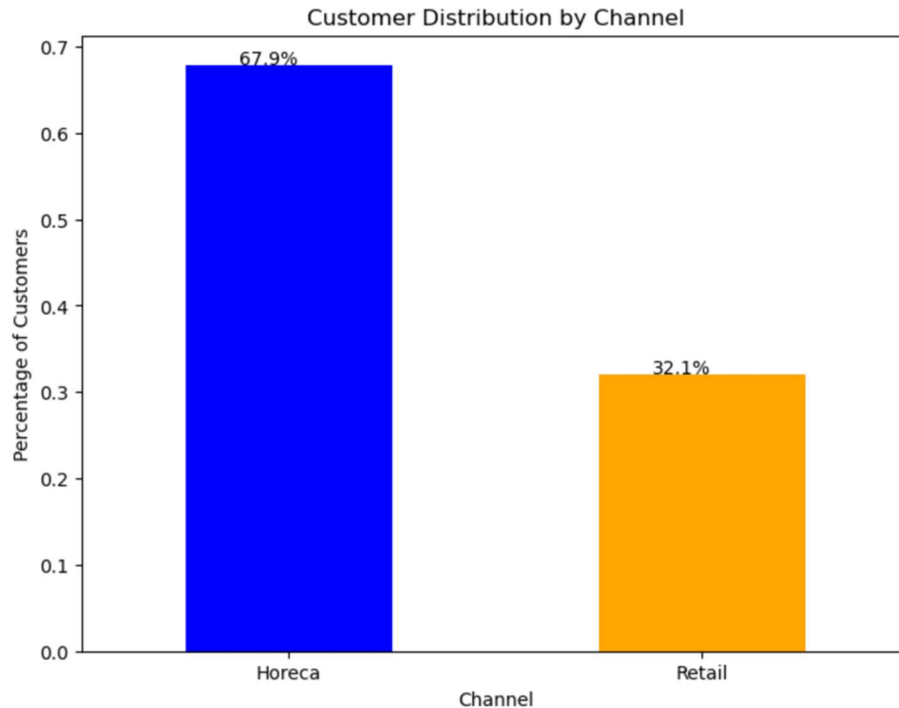
### Which product category has the highest average annual spending?

Based on the data analyzed from the Average Annual Spending by Product Category figure, it is evident that the 'Fresh' category has the highest average annual spending among the analyzed product categories. This indicates that customers tend to allocate a significant portion of their budget to fresh products when purchasing from Montajat's Wholesale Customer Data.

### Are there any noticeable trends or patterns in spending across different product categories?

The data reveals that customers prioritize spending more on categories such as 'Fresh', 'Milk', and 'Grocery' compared to categories like 'Frozen', 'Detergents\_Paper', and 'Delicatessen'. This suggests a trend of higher average spending on staple foods and dairy products. Montajat can leverage these insights to focus on these high-demand categories and consider strategies to promote and expand offerings in lower spending categories like 'Frozen'.

## Customer Segment Analysis

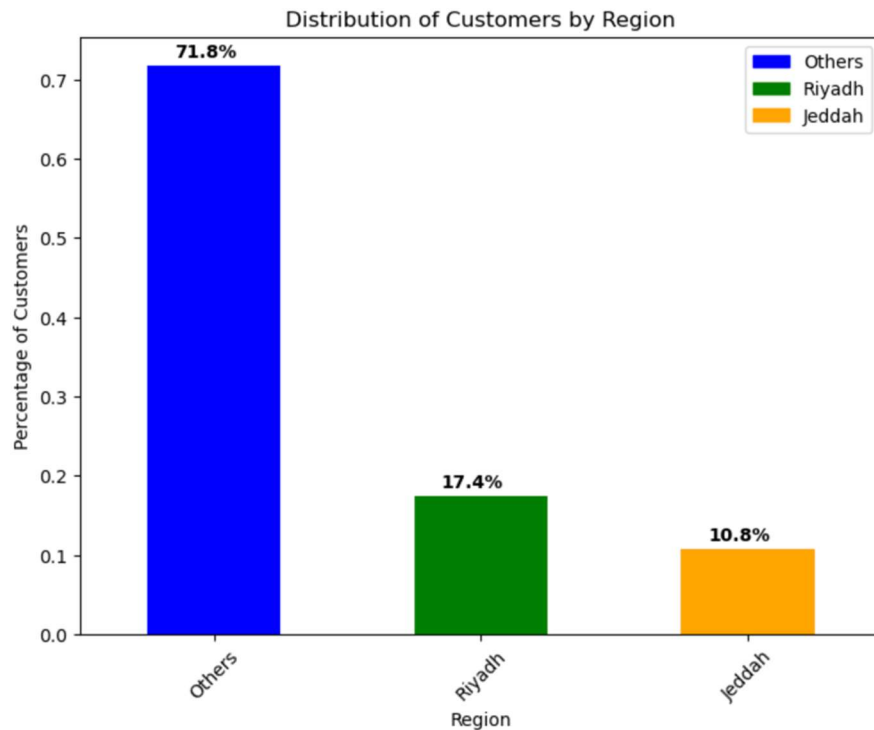


### How is spending distributed between the Horeca (Hotel/Restaurant/Café) and Retail channels?

Analyzing the spending distribution by channel provides insights into how spending is distributed between the Horeca and Retail channels. This information allows Montajat to understand the market segment distribution and tailor their product offerings, marketing strategies, and operational efficiency to meet the diverse needs of their clientele more effectively.

From the graph, it is evident that a significant majority of customers, accounting for 67.9% of the distribution, are in the Horeca channel, which includes hotels, restaurants, and cafes. On the other hand, the Retail channel represents 32.1% of the customer distribution.

This data suggests that Montajat has a larger customer base in the Horeca channel, indicating a higher demand for their products from businesses in the hospitality and foodservice industry. Understanding this distribution can help Montajat focus their efforts on serving Horeca customers effectively while also considering strategies to attract and retain customers in the Retail channel.



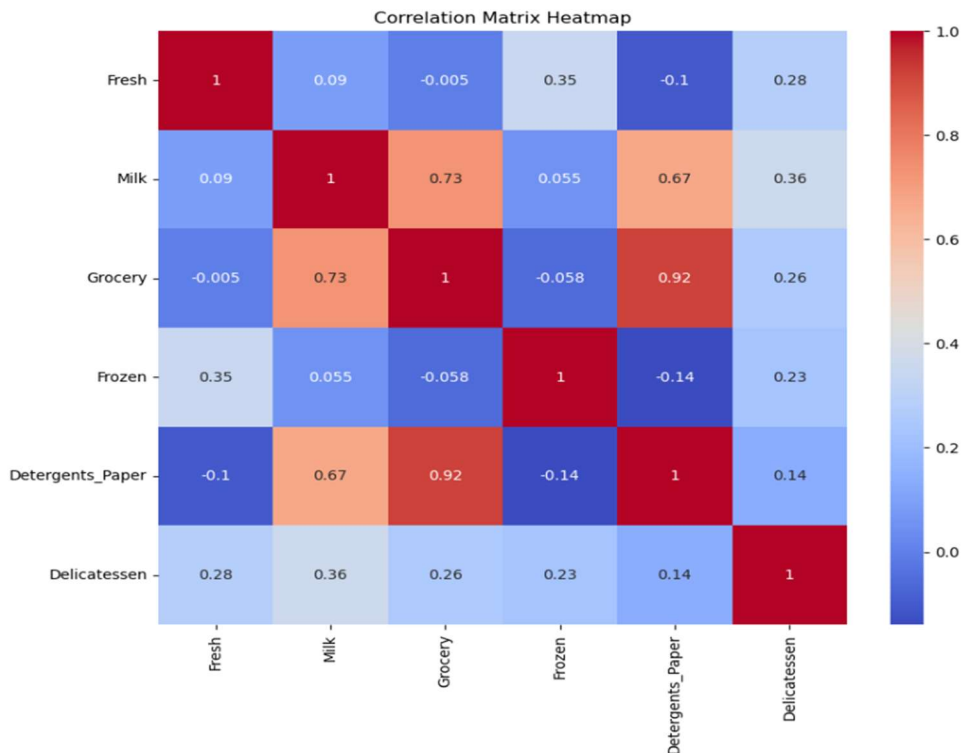
Is there a significant difference in spending patterns between the regions (Riyadh, Jeddah, and Other)?

Yes, there is a significant difference in spending patterns between the regions of Riyadh, Jeddah, and Other. From the data, it can be observed that Riyadh represents a lower customer concentration at 17.4% of the total distribution. Jeddah, on the other hand, has a minimum distribution of nearly 11%. The remaining regions collectively make up 71.8% of the customer distribution.

This discrepancy in customer concentrations suggests that spending patterns vary across these regions. Montajat can leverage this information to tailor their strategies and operations to meet the specific spending behaviors and preferences of each region. By understanding the distinct spending patterns, Montajat can allocate resources effectively, implement targeted marketing efforts, and identify growth opportunities in different regions.

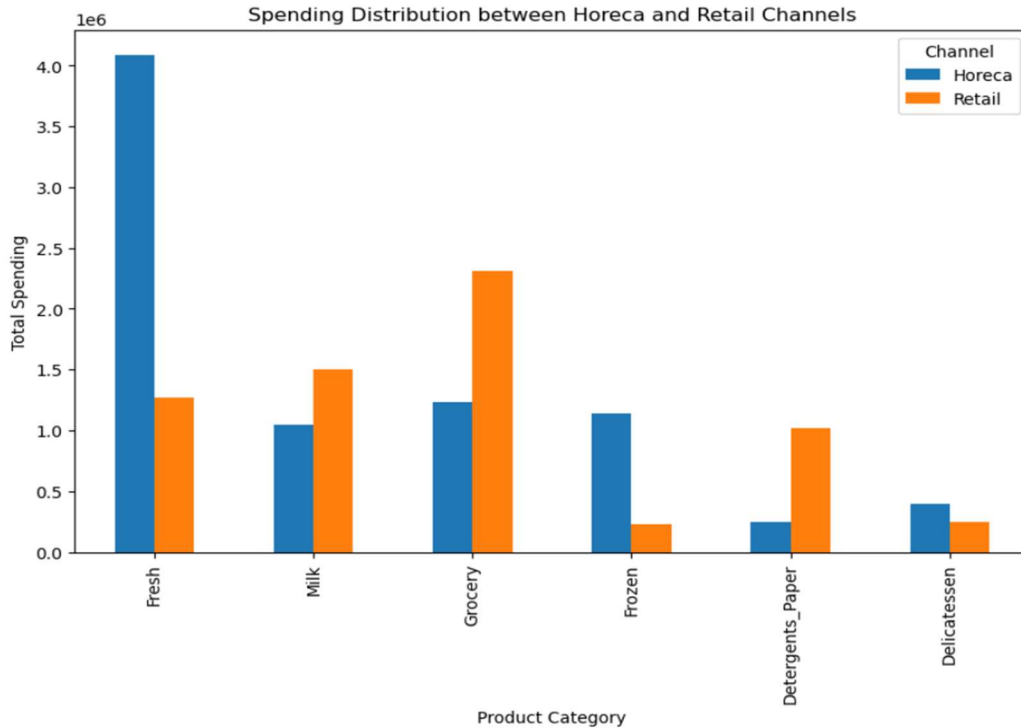
Analyzing spending distribution by region provides Montajat with valuable insights to optimize their business operations and better serve their customers across Riyadh, Jeddah, and other regions.

## Correlation Analysis



- Are there any significant correlations between different product categories?
  - As shown in the figure above, there is a Strong Positive Relationship between the categories “Grocery” and “Milk”. These two categories could be used as a combination in some recipes, so they need to be provided together.
  - Moreover, “Grocery” and “Detergents\_Paper” have a Strong Positive Relationship, and “Milk” and “Detergents\_Paper” have a Strong Positive Relationship as well.
  - Therefore, if a particular category demonstrates a high spending rate in specific channels and regions, it is likely that the other correlated category will also exhibit a similar elevated spending pattern.
- How does spending on detergents and paper correlate with other categories?
  - “Detergents\_Paper” is highly positively correlated with “Grocery” and “Milk”. As the spending of “Detergents\_Paper” increases, the other variable tends to highly increase as well.

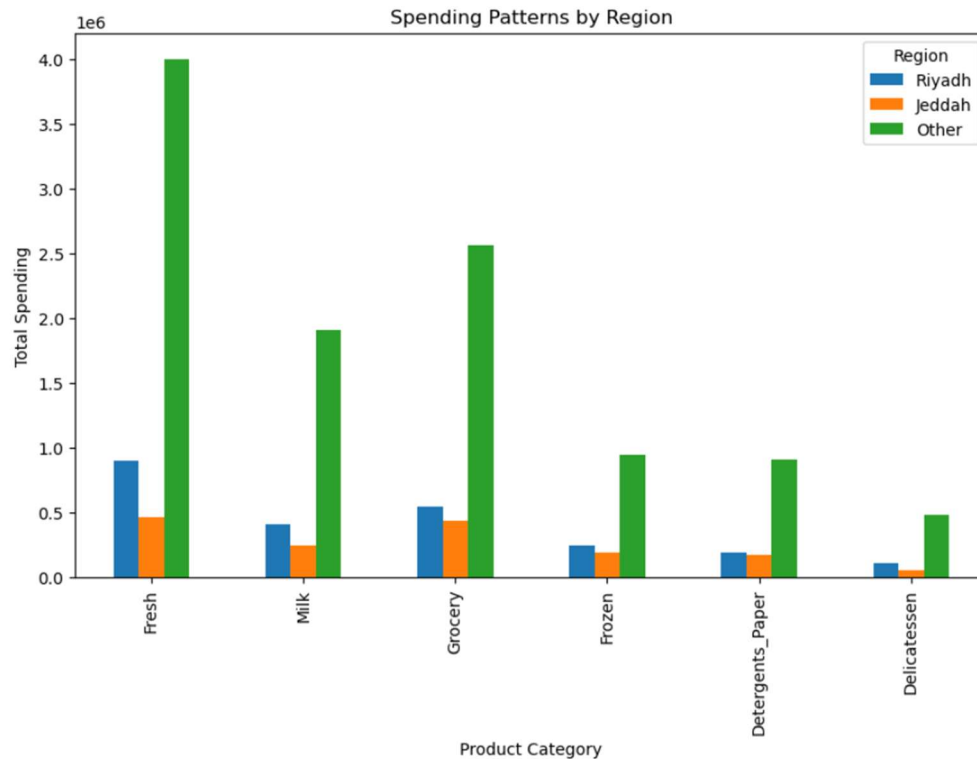
## Channel-specific Insights



- For each channel (Horeca and Retail), identify the top three product categories in terms of spending.
  - As shown in the figure above, the top three product categories in terms of spending in the “Horeca” channel are “Fresh”, “Grocery”, and “Frozen”.
  - As shown in the figure above, the top three product categories in terms of spending in the “Retail” channel are “Grocery”, “Milk”, and “Fresh”.
- Are there specific product categories that perform exceptionally well in a particular channel?
  - The category product category “Fresh” is performing exceptionally well in the “Horeca” channel.
  - The category product category “Grocery” is performing exceptionally well in the “Retail” channel.



## Regional Analysis



### Which region has the highest overall spending?

Based on the available data and the provided graph, it is evident that the "Other" region exhibits the highest overall spending. The green bar representing the "Other" region consistently demonstrates the highest spending across all product categories. The accompanying graph depicting customer distribution in regions further supports this finding, as regions with higher customer concentrations are likely to have higher spending as well. Thus, it can be inferred that the "Other" region surpasses Riyadh and Jeddah in terms of overall spending.

### Are there specific product categories that dominate in each region?

As can be seen from the graph, the fresh category dominates in all regions. We can see this from the fact that the bars representing the fresh product, colored blue, orange, and green, are the longest among all categories in each region. It is essential for Montajat to take note of this and continue to prioritize providing fresh products. At the same time, they can focus on promoting other product categories to attract customers and increase their overall revenue. By maintaining a strong emphasis on fresh products while actively promoting the rest of their offerings, Montajat can effectively capture customer attention and drive growth in their business.

## Overall Recommendations

Based on your analysis, provide Montajat with actionable recommendations to improve its product offerings and enhance customer satisfaction.

- Montajat should prioritize and invest in expanding its presence in retail channels due to the low customer base.
- As the percentage of customers is low in Jeddah, Montajat should optimize its distribution channel and invest in targeted marketing to raise awareness about its products and to increase customer engagement.
- Montajat should place a strong emphasis on prioritizing and allocating resources towards enhancing the quality of its Delicatessen products, taking into consideration the fact that these products currently face a challenge of low annual spending.
- Montajat should expand both domestically and internationally, increasing product distribution in more regions within Saudi Arabia and exploring new markets worldwide.

## Note

There is a python notebook containing all the code generated to implement the analysis with visualization in addition to explanation after every cell to make the work clear and concise.