**Wholesale Data Analysis**

Data

A wholesale distributor operating in different regions of Portugal has information on annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers’ annual spending on 6 different varieties of productsin 3 different regions (Lisbon, Oporto, Other) and across different sales channel (Hotel, Retail).

Data Description

1. **Buyer/Spender**- ID's of customers
2. **Region**- Region of the distributor
3. **Fresh**- spending on Fresh Vegetables
4. **Milk**- spending on milk
5. **Grocery**- spending on grocery
6. **Frozen**- spending on frozen food
7. **Detergents\_paper**- spending on detergents and toilet paper
8. **Delicatessen**- spending on instant foods

**Basic Steps:**

* + 1. Display the top 5 rows.
    2. Display the last 5 rows
    3. Check the shape of dataset.
    4. Check the datatypes of each feature.
    5. Check the Statistical summary
    6. Check the null values
    7. Check the duplicate values
    8. Check the anomalies or wrong entries.
    9. Check the outliers and their authenticity.
    10. Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

1. **Spending Analysis**
   * What is the total number of buyers in the dataset?
   * What is the average spending on each category (Fresh, Milk, Grocery, Frozen, Detergents\_paper, Delicatessen)?
   * Which category has the highest average spending?
   * How many buyers spend above the average on Fresh Vegetables?
2. **Regional Demand**
   * What is the total spending in each region?
   * Which region has the highest spending on Milk?
   * How does the average spending on Grocery vary across different regions?
   * Which region has the highest average spending per buyer?
3. **Category Preferences**
   * What percentage of buyers spend more on Frozen food compared to Delicatessen?
   * Which category shows the most variation in spending among buyers?
   * Are there any regions where spending on Detergents\_paper is significantly higher than others?
   * What is the correlation between spending on Fresh and Frozen food?
4. **Customer Segmentation**
   * Can buyers be grouped into segments based on their spending patterns? (e.g., using clustering analysis)
   * What are the characteristics of the top 10% spenders in each category?
   * How do spending patterns differ between high spenders and low spenders?
5. **Cross-Category Analysis**
   * Is there a correlation between spending on Milk and Grocery?
   * Do buyers who spend more on Delicatessen also spend more on Frozen food?
   * What is the combined average spending on Fresh and Milk for each region?
6. **Demand Trends**
   * Which region has the fastest growing spending on Fresh Vegetables?
   * How does the total spending on Grocery change across regions over time (if time data is available)?
   * What is the average spending per buyer in each category over a specified time period (if time data is available)?
7. **Buyer Insights**
   * What is the repeat purchase rate for buyers who spend above the average in at least three categories?
   * How many buyers spend consistently (i.e., similar amounts) across all categories?
   * Which region has the most diverse spending patterns (i.e., high variance in spending across categories)?