

Business Case 5

Retail Analysis

**MASTER DEGREE PROGRAM IN DATA SCIENCE
AND ADVANCED ANALYTICS**

GROUP D

Débora Santos : 20200748

Pedro Henrique Medeiros : 20200748

Rebeca Pinheiro: 20201096



Project Objectives



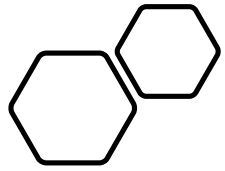
Understand each Point-of-Sale characteristics (top products sold, market share and product co-occurrences)



Points of sales clustering divided by value and product preference



Units' products forecast 6 weeks: one split by product and by point-of-sale and another one total by-product.



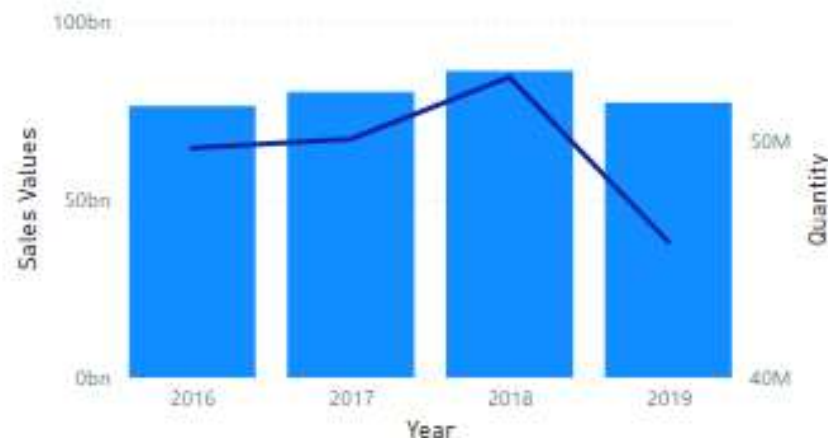
Sales Analysis

- Sales show seasonality/oscillations between months in terms of volumes but in values, it is well aligned
- Quarters 1 and 4 have the highest volumes
- Sunday is the day with the lowest sales volume
- The year with the highest sale were 2018 with more than 86B in sales and quantity more than 52B.
- Product family with a higher percentage in sales is product family 2 (19% in total which correspond to 61B in sales)
- POS_292 is the most representative in sales - 2B in sales (12% of the total)
- ProductName_ID is 2802 has the most purchases - 19B on sales (18,96% in GT)
- ProductCategory_ID 178 is by far the most frequent - 239B in sales (82% in total)

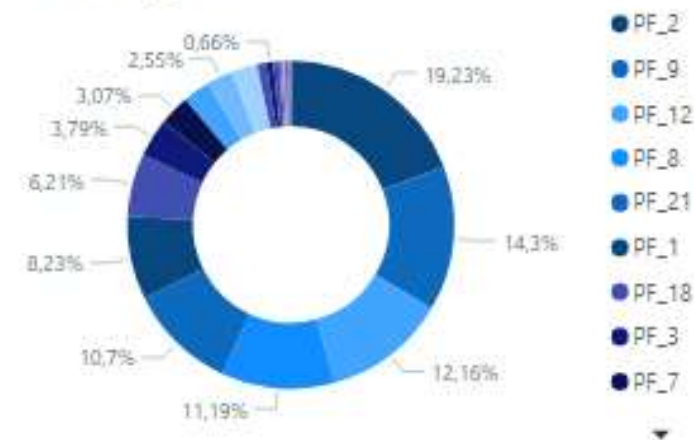
SALES ANALYSIS

Sales Values and Quantity by Year

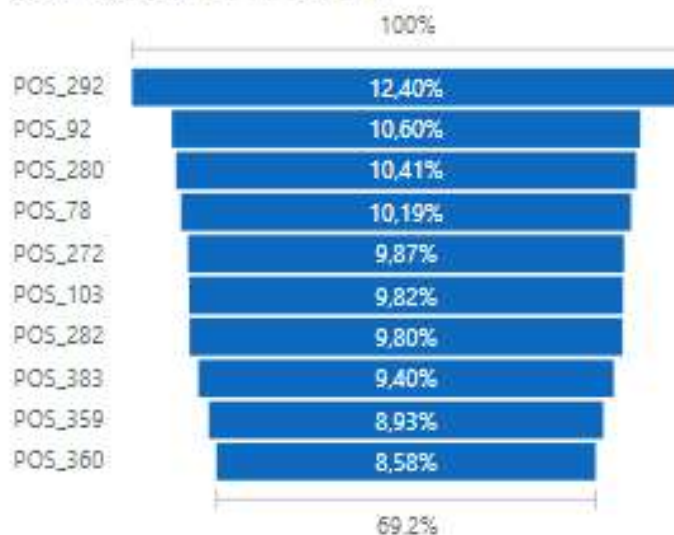
● Sales Values ● Quantity



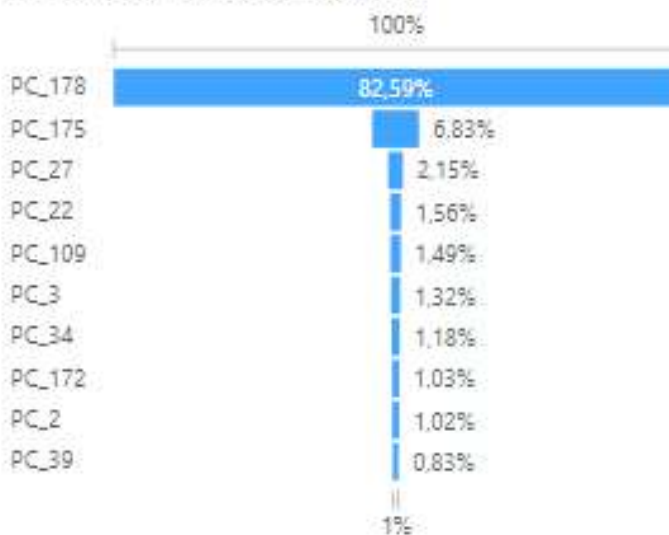
Sales Values, Quantity and %GT Quantity by ProductFamily_ID



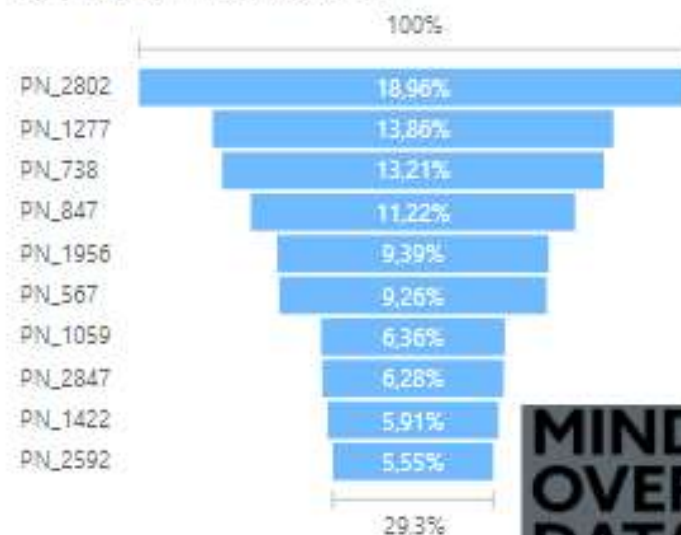
%GT Sales Values, Sales Values, Quantity and %GT Quantity by Point-of-Sale_ID



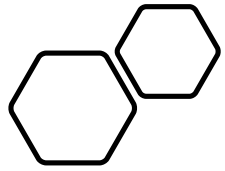
%GT Sales Values, Sales Values, Quantity and %GT Quantity by ProductCategory_ID



%GT Sales Values, Sales Values, Quantity and %GT Quantity by ProductName_ID



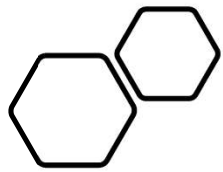
MIND
OVER
DATA



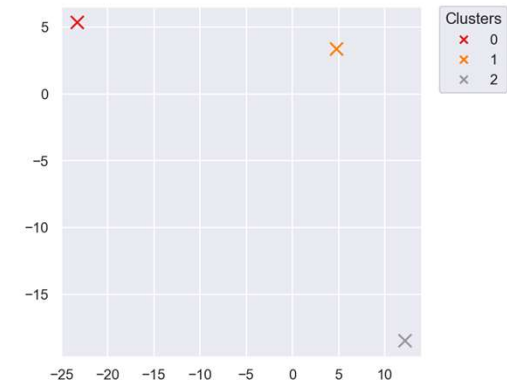
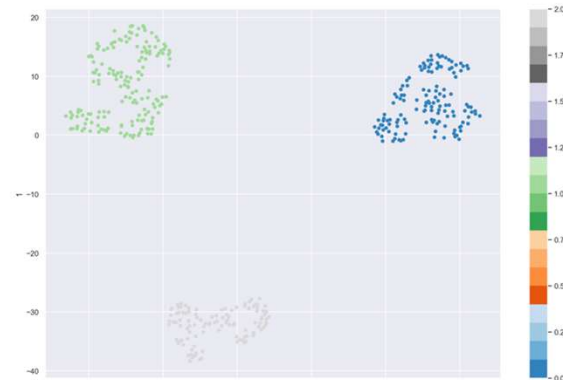
Clusters

Our final solution was able to detect 3 segments:

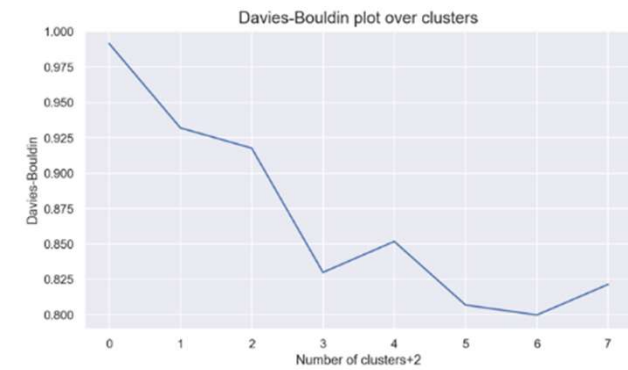
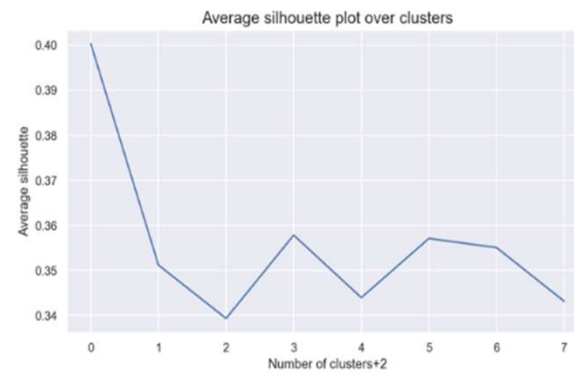
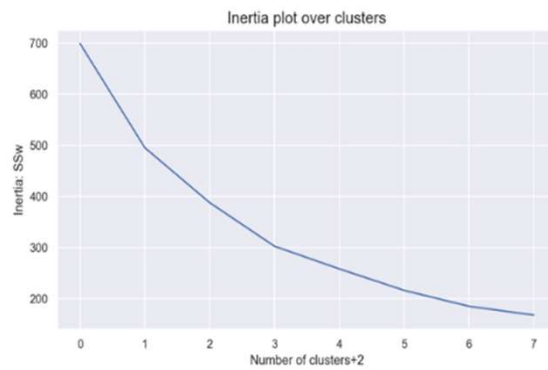
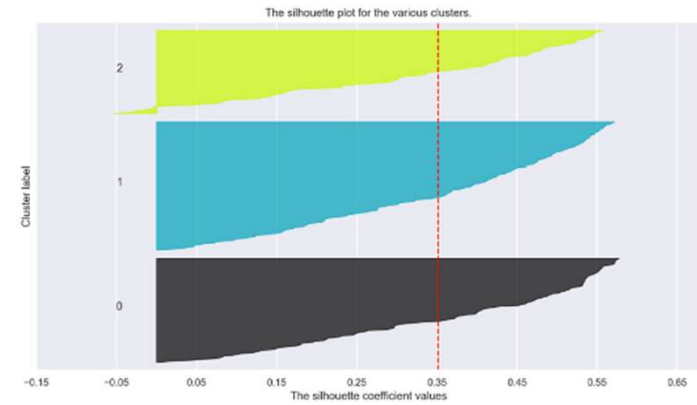
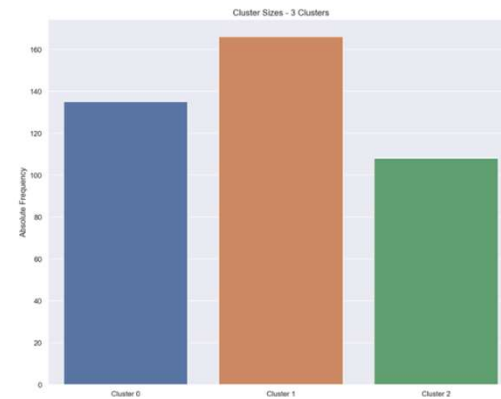
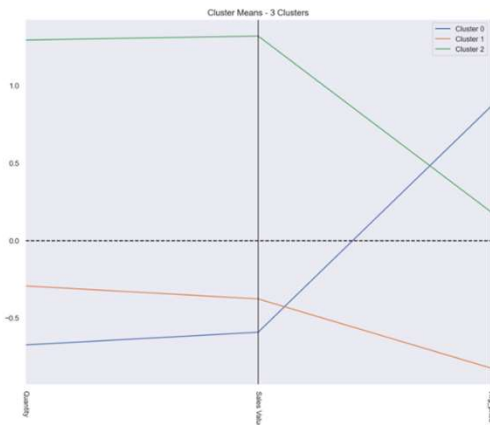
1. **Valuable:** stores with higher sales volume with average price between Bulk and Premium
 2. **Bulk:** stores with average sales volume of products with lower price
 3. **Premium:** stores that sell the products at the highest prices, but with the lowest sales volumes
- K-means algorithm used for clustering
 - Inertia plot, average silhouette plot, Davies-Bouldin plot, Silhouette plot and t-SNE plot to visualize and evaluate the results



Clusters



Kmeans clustering



CLUSTER ANALYSIS

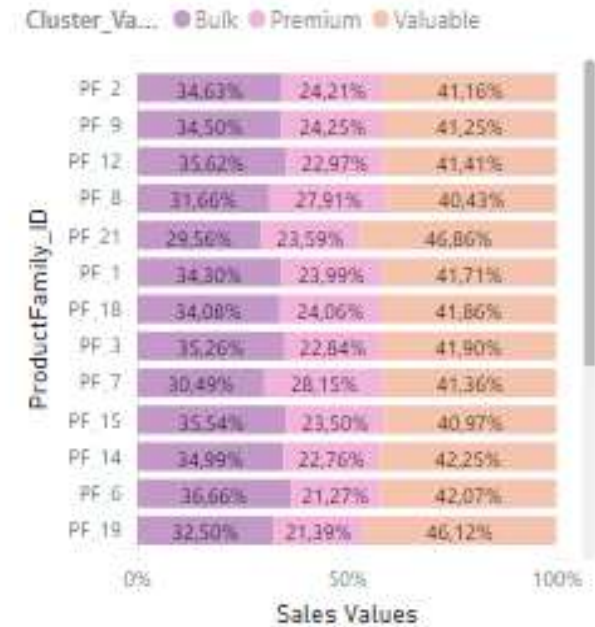
Sales Values and Quantity by Cluster



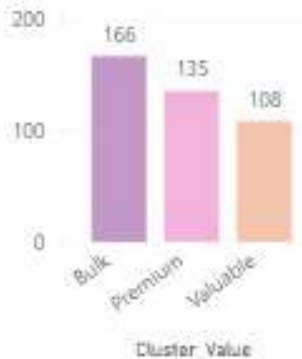
Cluster Analysis



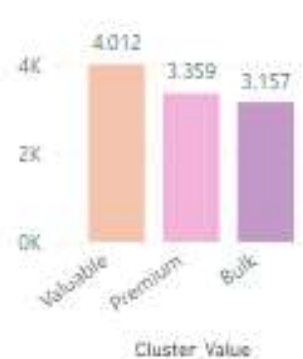
Percentage of Sales values by cluster by Product Family



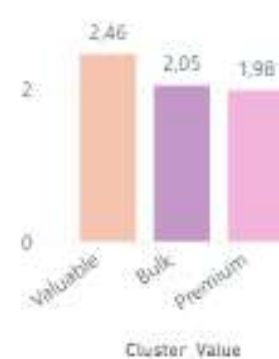
Number of Point of Sales by Cluster



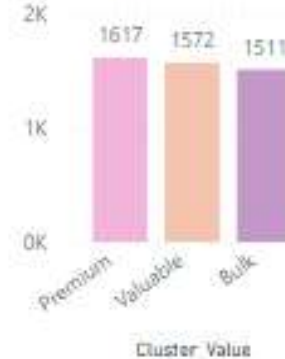
Average of Sales Values by Cluster



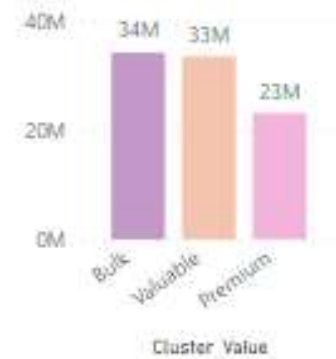
Average of Quantity by Cluster



Average Price by Cluster



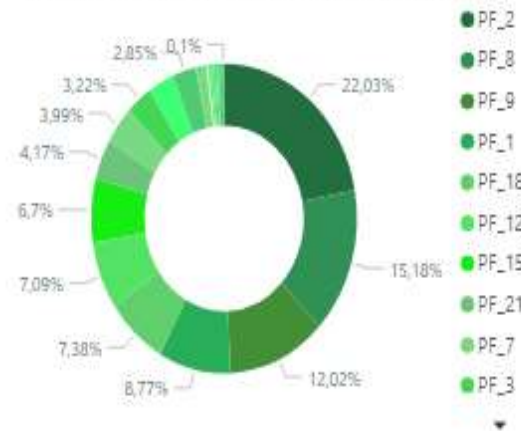
Number of Transactions by Cluster



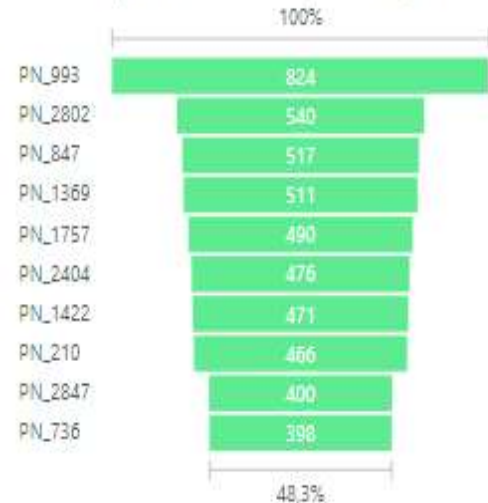
POINT OF SALE'S ANALYSIS



Distribution of Sales Values by Product Family



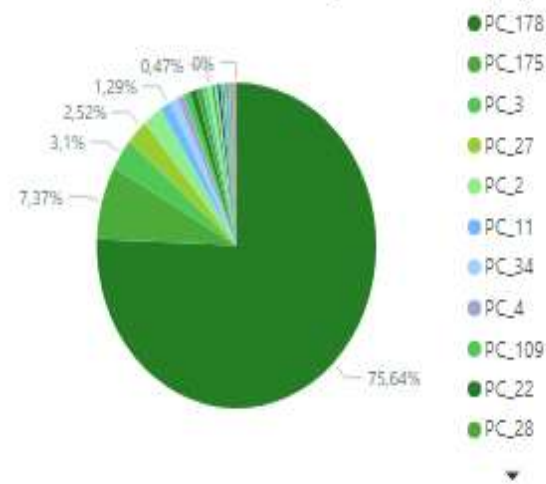
Top 10 Products by Quantity



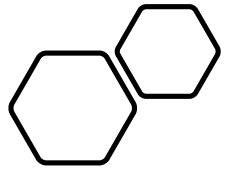
Top 10 Products by Sales Values



Distribution of Sales values by Product Category



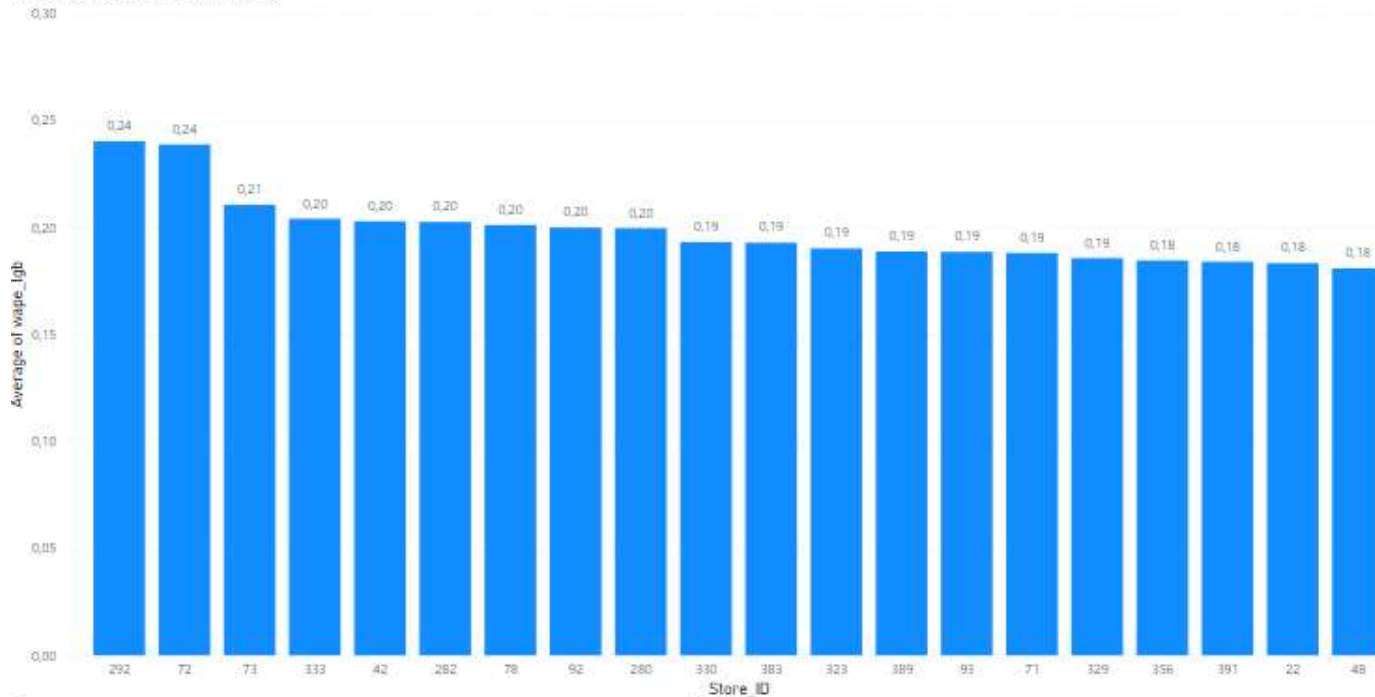
- The category 178 represents more than 75% of sales
- More than 50% of sales are concentrated in 4 families: 2,9, 8 and 1, with family 8 having the highest average selling price in general.
- The best-selling product has more than 15% of sales and the tenth-best selling product represents 50% compared to the first-best selling product



Demand Forecast

- Feature Engineering
- Model LGBMRegressor
- Evaluation: MSE and WAPE

Average of wapec_lgb by Store_ID





Thank You!