Technical Report

Utilization of Freezer Assets

Data Storm 4.0 – Semi Final Round

- Team Name: **3_Amigos**
- Team Members:

  Sangaran Thevarasa

  Pairavi Thanancheyan

  Nithursika Kalanantharasa

- GitHub Repository Link: https://github.com/Pairavi/DataStorm-4.0/

Step 1 -Problem Description

- The problem at hand is that the sales department of a retail company XYZ wants to optimally allocate freezers of different volumes and power consumption rates to its 1000+ stores that produce both beverages and ice cream items. They are currently allocating freezers based on the area distributor manager's (ADM) visit to the store to assess outlet size, outlet space availability, outlet sales, and outlet location to determine the freezer type. Company XYZ wants to enhance this process by using advanced analytics to segment their stores and recommend suitable freezer types for each segment, which would improve their sales against the cost invested in freezers.

Examples of problem statements could include:

- The current freezer allocation process is not optimized for maximum capacity utilization, leading to inefficient use of resources and potential waste of valuable samples.
- The current freezer allocation process may cause delays in the testing of new samples.

Objectives:

- Perform a store segmentation to identify the stores with similar characteristics that are potential candidates to share the same freezer type due to the similar nature and behavior of the stores.
- Recommend a suitable freezer type for each identified outlet segment based on area, Weekly_Ave_Total_volume, Weekly_Ave_Total_price to maximize the ROI and sales of freezers.
- Develop an analytical solution that can be developed with simple rule-based systems to machine learning algorithms, providing a well-commented code with clear steps for performing both tasks.
- Evaluate the performance of the segmentation model against various metrics, such as Inertia, Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index.

Tools used.

- Python programming language was used for the entire analysis. The following Python libraries were used:

  Data manipulation and analysis: NumPy, pandas

  Modeling: Scikit-learn

Data visualization – Matplotlib, Seaborn

Step 2(a) - Exploratory Data Analysis

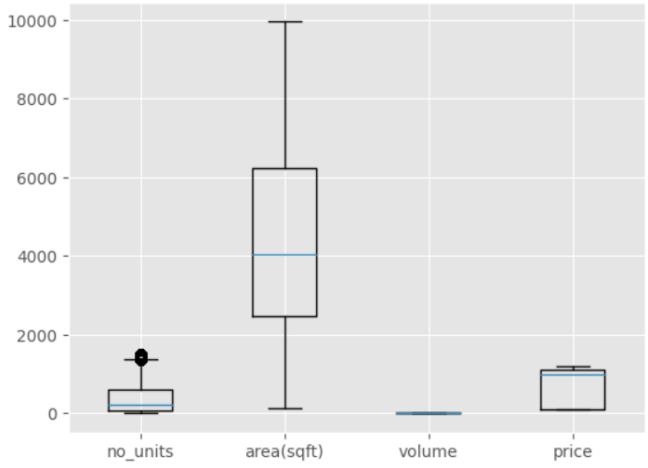
1. Handling missing values (There are no missing values)

2. Removing duplicates (We dealt with this using drop_duplicates code.)

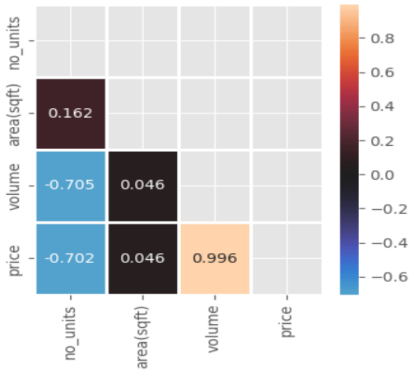
We had some stores with different areas. We handled this by replacing it with its average value(since different area data count was so small).

3. Outliers handling

We had outliers only in no_units category.

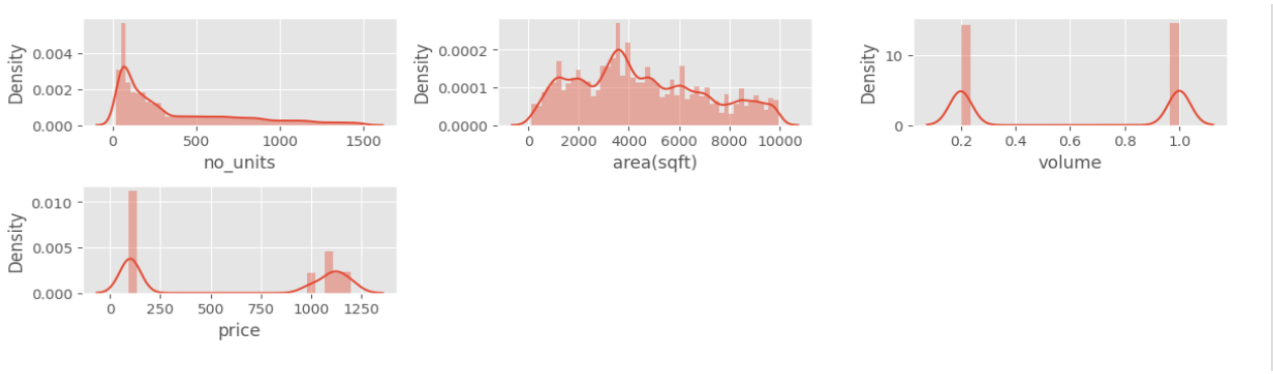


4. Check for correlations



There is only significant correlation between price and volume. But we would not come to any better conclusion.

5. Visualize the data



Step 2(b) - Feature Engineering steps

Feature Transformation

In the merged dataframe, we converted no_units and volume into Weekly_Ave_Total_volume by multiplying both categories and calculate the weekly average amount of volume. we converte d no_units and price into Weekly_Ave_Total_price by multiplying both categories and calculate the weekly average amount of price.

	Outlet_ID	week	pid	no_units	area(sqft)	volume	product_name	price	Total_volume	Total_price	Weekly_Ave_Total_volume	Weekly_Ave_Total_price
0	ID7203	1	IP1	874.0	5660	0.2	Vanilla Mini Cone	100	174.8	87400.0	135.56	69966.0
1	ID7203	2	IP1	644.0	5660	0.2	Vanilla Mini Cone	100	128.8	64400.0	120.10	62152.5
2	ID7203	3	IP1	600.0	5660	0.2	Vanilla Mini Cone	100	120.0	60000.0	112.20	55040.0
3	ID7203	4	IP1	106.0	5660	0.2	Vanilla Mini Cone	100	21.2	10600.0	129.15	64912.5
4	ID7203	5	IP1	775.0	5660	0.2	Vanilla Mini Cone	100	155.0	77500.0	112.08	54366.0

Feature encoding

before, we encode the pid category using label encoding. But we would not come to any better c onclusion. So, we dropped it.

Feature scaling

We used min max scaling on the selected columns Ave_area(sqft), Ave_total_perweek and Ave _volume_perweek.

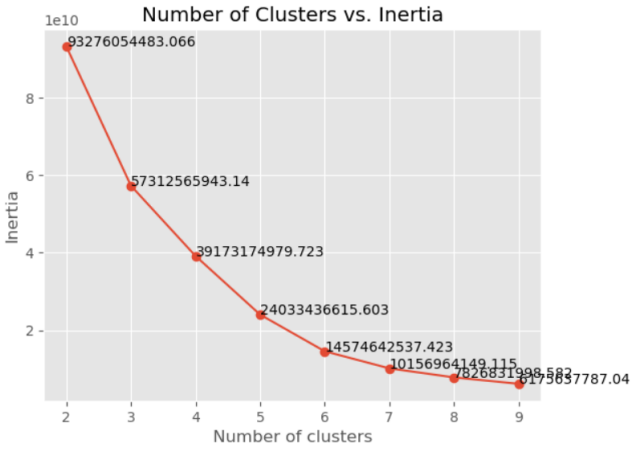
Since some features are not required to categorize stores, we removed them, such as Outlet_ID, week, pid, and product_name.

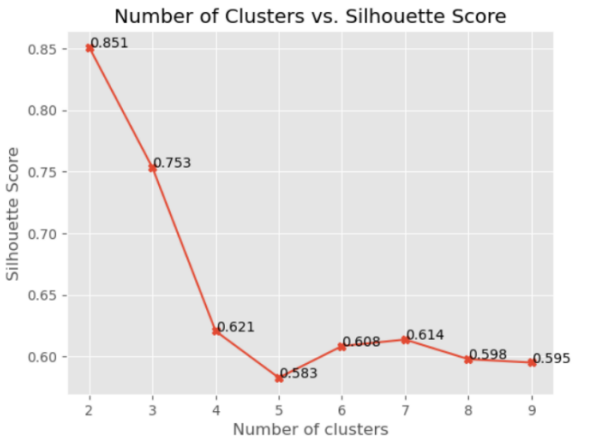
Step 3 - Clustering/ Segmentation Technique

we can use clustering/segmentation techniques to group similar customers together based on the ir similar behaviors. This can help us identify different segments and develop targeted char acteristics for each segment.

We used K-means clustering technique that aims to partition the dataset into 4 clusters. K-means clustering worked well for datasets with numerical data.

To determine elbow point, we used inertia and the silhouette coefficient.

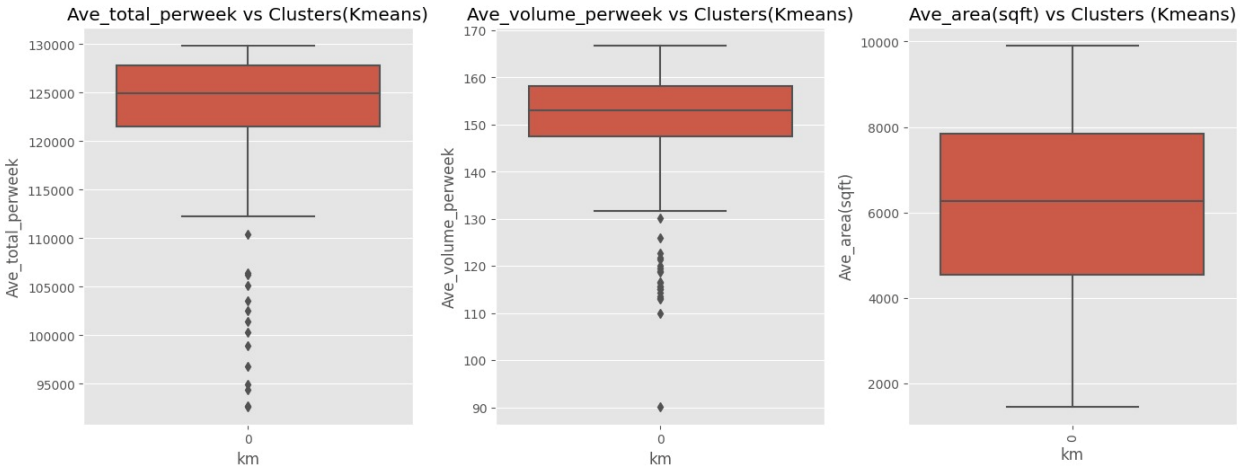


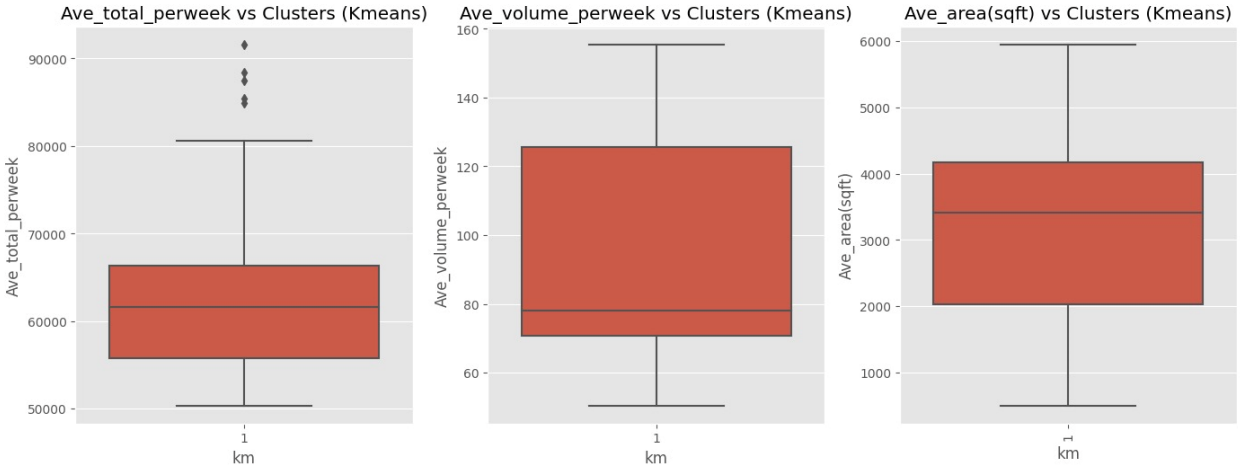


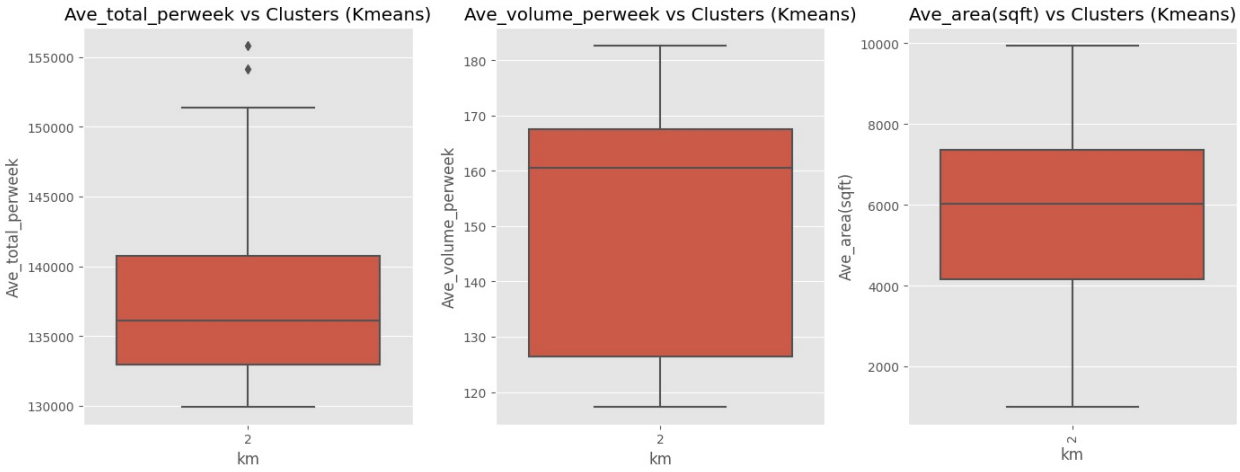
Finally, we decided on 4 clusters for the K-Means clustering.

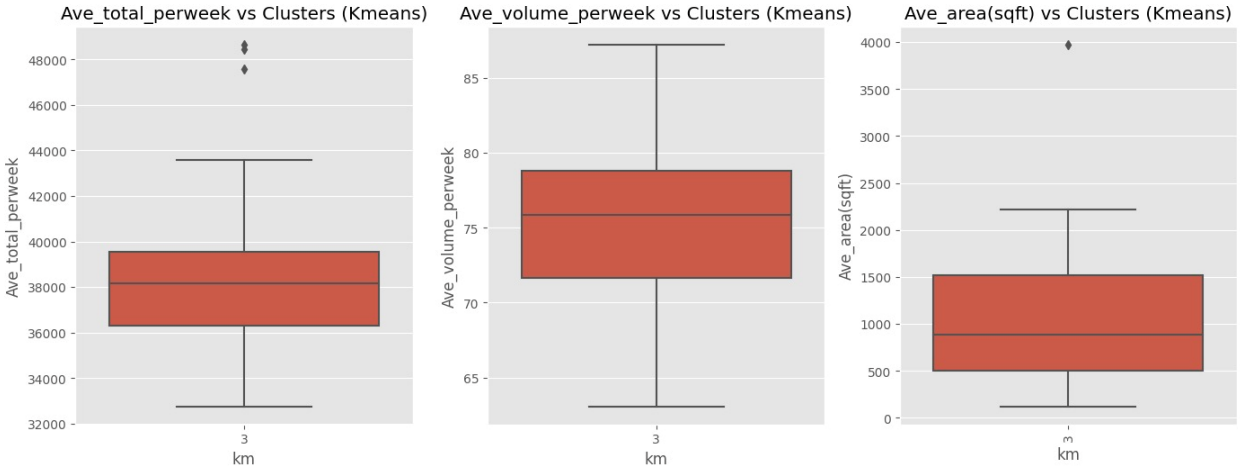
Step 4 - Characteristics of each segment

All of the features in Clusters 1 (Type 2) and 2 (Type 3) had mid- values. The values for all the features were larger in Cluster 0 (type 1), while they were lower in Cluster 3 (type 4) at the same time.



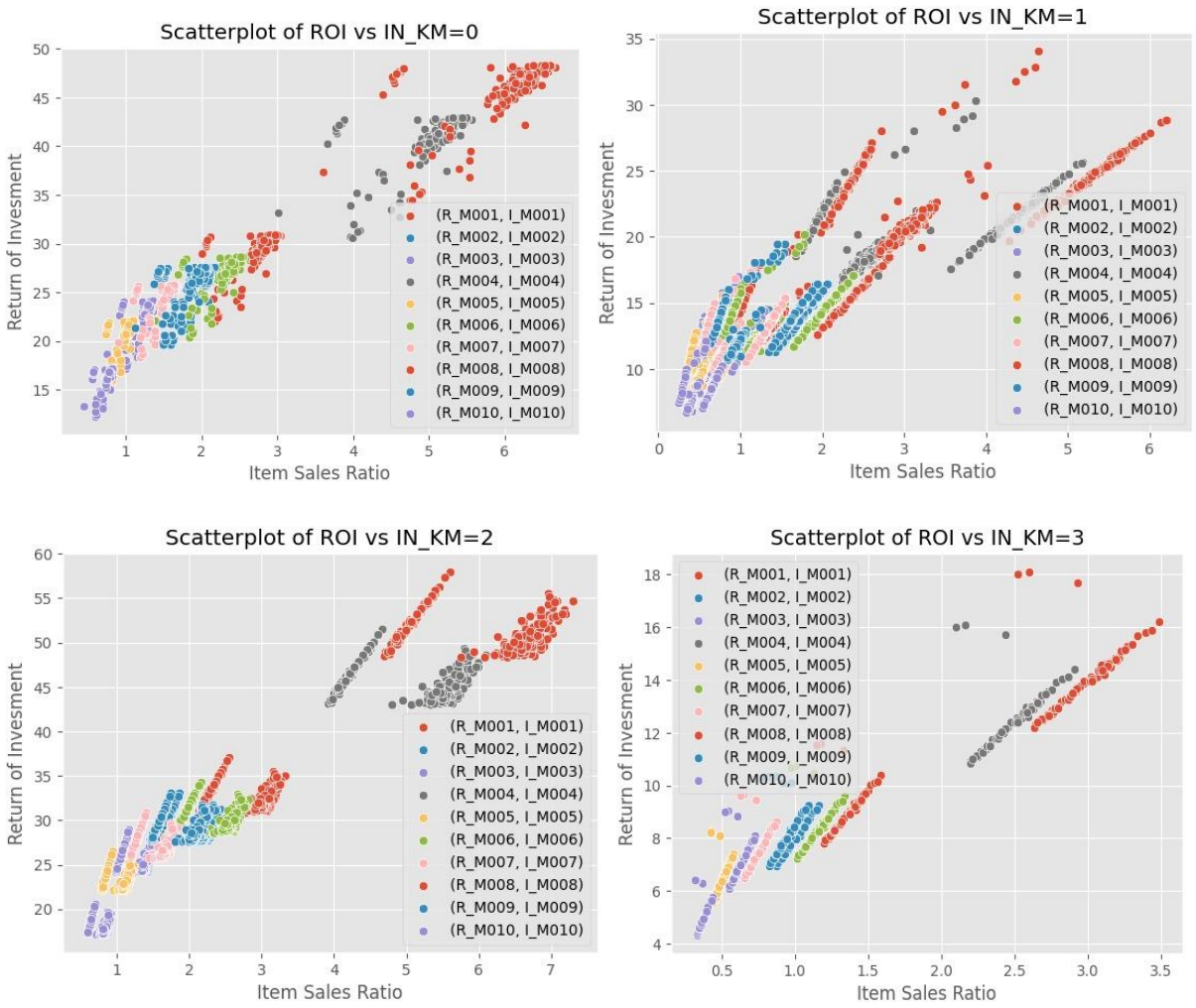

Ave_total_perweek vs Clusters (Kmeans)
Ave_total_perweek
km
Ave_volume_perweek vs Clusters (Kmeans)
Ave_volume_perweek
km
Ave_area(sqft) vs Clusters (Kmeans)
Ave_area(sqft)
km


Ave_total_perweek vs Clusters (Kmeans)
Ave_total_perweek
km
Ave_volume_perweek vs Clusters (Kmeans)
Ave_volume_perweek
km
Ave_area(sqft) vs Clusters (Kmeans)
Ave_area(sqft)
km


Ave_total_perweek vs Clusters (Kmeans)
Ave_total_perweek
km
Ave_volume_perweek vs Clusters (Kmeans)
Ave_volume_perweek
km
Ave_area(sqft) vs Clusters (Kmeans)
Ave_area(sqft)
km

Step 5 - Analytical approach to allocate freezers to stores

First, we tried to optimize this approach by doing linear programming. But it would not work properly. Then, we plotted graph ROI vs ISR for each cluster separately and determined the freezer for allocation.



A high ROI value indicates that the investment in the freezer is profitable, while a low value indicates that it may not be worth the investment. A high ISR value indicates that we are using the freezer capacity efficiently, while a low value may indicate that we are wasting freezer capacity.

In general, an ISR value close to 1 is considered efficient because it means that we are using the freezer capacity optimally without any wastage. If the ISR value is more than 1, it means that we do not have enough capacity to store ice cream and may need to invest in a bigger freezer. On the other hand, if the ISR value is less than 1, it means that we have more free space in the freezer and may incur unnecessary maintenance costs.

Therefore, when considering the investment in a freezer for storing ice cream, it is important to balance the ROI and ISR values to ensure that we are making a profitable investment.

Step 6 - Conclusion

After conducting the analysis on freezer allocation for Company XYZ, the conclusion presents a comprehensive analysis of the results obtained and the implications for the company.

- Our analysis demonstrates that Company XYZ can increase sales and reduce costs by reallocating freezers to high-demand (more sales) stores and reducing the number of freezers in low-demand stores.
- As a result of our analysis, we discovered that Company XYZ's stores' current freezer allocation is not set up to maximize sales. The business can boost sales and cut costs by using a data-driven strategy to assign freezers to stores. We advise the business to purchase a data analytics tool to assist them in optimizing their freezer allocation..

Intervention strategies

- Reassign the freezers: In the context of the analysis, Company XYZ might think about reassigning the freezers to locations with high sales potential and freezer utilization rates.

- Track sales, freezer utilization rates, and customer feedback in regular monitoring and evaluation processes to ensure ongoing success. This would be done by Company XYZ.

- Implement sales promotions: In order to boost customer demand and boost sales, Company XYZ may want to implement specific sales promotions in stores with low sales volumes, according to the analysis.

- Improve inventory management: Company XYZ could enhance its inventory management procedures to further optimize freezer utilization and reduce waste. This could involve automating reordering processes, better inventory tracking, and product mix optimization to guarantee that high-selling items are always available.