# **MLUL1- Group Assignment**

This deliverable has 50% weightage in the Consolidated Total Score.

# Dataset Overview:

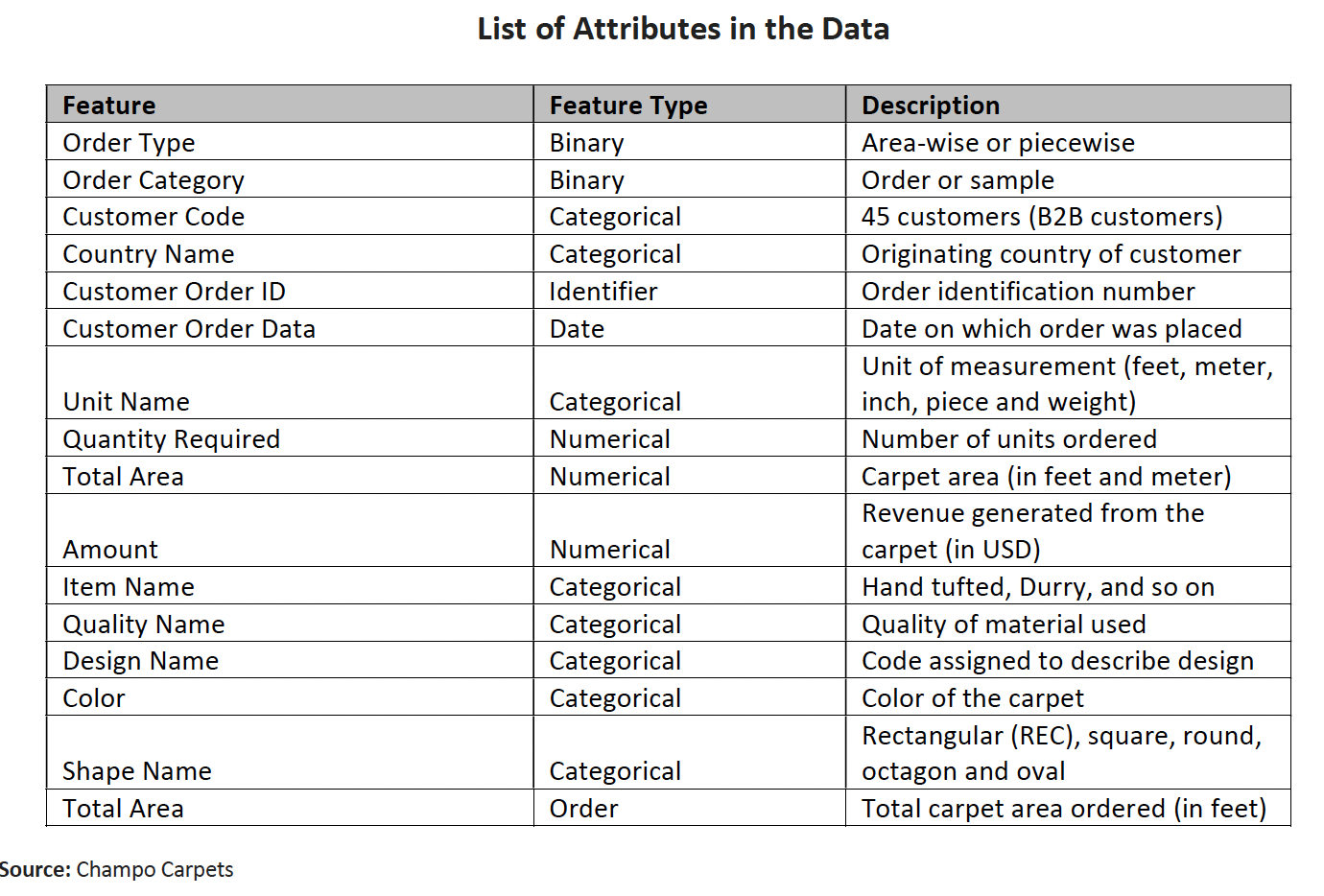
This assignment is based on

<https://store.hbr.org/product/champo-carpets-improving-business-to-business-sales-using-machine-learning-algorithms/IMB879>

The dataset file name is: Champo Carpets V1.csv

The dataset contains about 45 customers data. Each customer has one or more orders, and each order contains one or more items.

These customers are B2B customers. Some of the item’s orders are expensive (or high value items) and some are not.



# Objective:

Objectives of this assignment are:

* Understand customers purchase behavior and segment them to understand their characteristics better.
* Segment the products based on how customers buy them and segment them.
* Understand how products are being bought together and cluster them together so that the clusters can become basis for recommendations.

# Problem Description:

## Exploratory Data Analysis

1. Participants are expected to complete basic exploratory data analysis and find out at least 5 key insights from the data.
2. These could be univariate or bivariate analysis.
3. The insights should be depicted using appropriate charts.

## Customer Segmentation

### Stage 1:

Generate the following features based on customers purchase behavior.

* frequency (number of orders)
* monetary value from the customers.

Create clusters to segment customers based on above two features. Use appropriate methods to understand number of clusters and verify the quality of clusters.

Analyze each cluster and interpret them. Also, name each segment based on your understanding.

### Stage 2:

Repeat the above exercise for all the three following algorithms and compare and interpret the results.

1. **K-Means**
2. **Agglomerative**
3. **HDBSCAN**.

### Stage 3:

Add some more features the above features using feature engineering for example kind of products customers are buying, high value or low value items bought, diversity of items (number of unique items bought) in orders.

Choose the right algorithms based on your understanding and complete the clustering exercises and infer if the clusters created have better understanding the customer segments or not.

## Products Segmentation

Create the following features for each of the products or items (item\_id).

* total number quantities ordered.
* total revenue generated.
* number of customers orders the item.
* average number of quantities per order and so on.
* Add two more features based on your intuition or understanding.

Create clusters to segment items or products based on above two features. Use appropriate methods to understand number of clusters and verify the quality of clusters.

Analyze each cluster and interpret them. Also, name each segment based on your understanding.

## Product Cluster Based on Purchase Patterns

Cluster the products based on how they are being bought together in orders. Choose the right metrics to cluster the products.

Cluster the products or items in smaller sized clusters (min cluster size to be 2).

Display some of the clusters and discuss the results.

# Assignment Instructions

1. Use Jupyter notebook or Google Colab for developing these solutions.
2. Code should follow PEP8 coding guidelines for code formatting (basics of coding standards)
3. The code should be segregated into separate sections with proper headings. Proper sections should be created for clarity (like heading 2/3).
4. The project name, team members (with IDs) and problem description at the top.
5. Each section should be explained clearly (objective and approach). Each design decision should be documented clearly in the appropriate sections.
6. **Two COLAB Notebook (.ipynb) should be created and submitted.** 
   1. One for EDA and Customer segmentation
   2. One for Product Segmentation and Clustering.
7. The name of the notebook should be
   1. **GR\_n\_MLUL1\_<name>\_Customers.ipynb** and **GR\_n\_MLUL1\_<name>\_Products.ipynb**
   * Where n is the group number.
   * Group Name is name of the group. Choose a name by which your group want to be known or called.

# Submission Instructions

1. A **.pdf report** (not more than 5 pages) with only the project finding and your inferences. It can contain the charts, table and results discussion. It should also contain 3 key challenges or lessons you learnt while doing the assignment. ***It should not contain any code.***
2. Both the notebooks containing the solutions.
3. Both the notebooks must be fully executed with outputs before submission**. It should not contain any warning or errors**. Suppress the warning if necessary.
4. The notebooks should also be fully executable without any error, if we intend to run it for evaluation.

# Assessment Weightage

75% weightage will be given for accomplishing the above tasks.  Another 25% will be given for the following.

1. Code clarity
2. Documentation in the notebooks (use markdown)
3. Clearly creating sections for each task.
4. Document 3 keys Lessons learnt in the assignment.

**General Instructions:**

1. This is a group assignment with 4-5 member teams.
2. Do NOT submit .zip files otherwise, the submission will not be considered.
3. Please include your team member names and PGIDs in the submission.
4. **One person should submit on behalf of all the team members.**
5. Any late submission will attract a penalty, as mentioned in the course outline.
6. The honor code for this submission is **3N-a.**
7. **Please look through the honor code restrictions carefully before attempting the assignment, as there will be strong consequences for breaking them.**
8. **Please adhere to the given instructions; otherwise your submission will not be accepted, or a severe penalty will be applied**

**Deadline: 30th December 2023, 11:55 pm**