Sandra Zaborska-Kucharczyk

Raza Rafique  
MGS\*3101\*01  
11/19/2024

Deliverable 1 (D2): Project Data Acquisition

For the project on customer segmentation and spending behaviors, I chose a dataset from Kaggle called “Analyzing Customer Spending Habits to Improve Sales”.  
(https://www.kaggle.com/datasets/thedevastator/analyzing-customer-spending-habits-to-improve-sa). This dataset contains good information for clustering customers based on their spending behaviors and for showing the best selling products to optimize marketing strategies and improve customer retention.

Description of the Dataset

**Variables:**

The dataset includes 14 variables that represent customer demographics, product categories, spending behaviors, and revenue generated from each sale. Those variables are:

1. **Date:** Date of the sale.
2. **Year:** Year of the sale.
3. **Month:** Month of the sale.
4. **Customer Age:** Age of the customer.
5. **Customer Gender:** Gender of the customer.
6. **Country:** Country of the customer.
7. **State:** State of the customer.
8. **Product Category:** Category of the product.
9. **Sub Category:** Sub-category of the product.
10. **Quantity:** Quantity of the product sold/purchased.
11. **Unit Cost:** Cost of the product per unit.
12. **Unit Price:** Price of the product per unit.
13. **Cost:** Cost
14. **Revenue:** Total revenue generated from the sale.

**Number of Observations:**

The dataset includes 34866 observations. Each observation represents a unique customer and the associated spending data, providing sufficient data points to identify helpful patterns.

The use of this dataset, will allow me to identify patterns and group customers into clusters, which will enable businesses to develop targeted marketing strategies and to better allocate their sales’ resources.