Online Retail Customer Segmentation

Overview

Welcome to my project brief report, this project was made by Haidar Farras Latief and made in purpose on showcasing my abilities in clustering algorithm. The dataset are Online Retail data, I get the dataset from Kaggle:

https://www.kaggle.com/datasets/mathchi/online-retail-ii-data-set-from-ml-repository

The table below explain the customer segmentation project process:

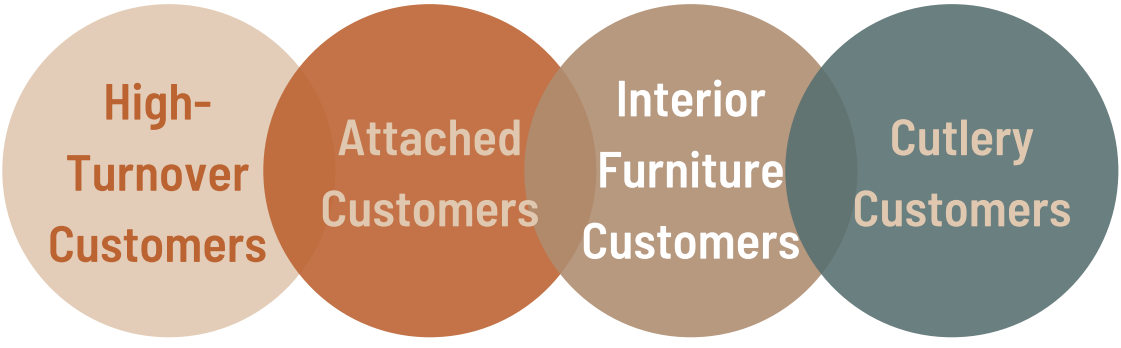
1. Data Loading
Import necessary libraries, Load the dataset, Deciding datetime (1st January 2010 to 1st February 2010)
2. Exploratory Data Analysis: Before Clustering
Feature observation, Missing value observation. Dummy Data Observation, Negative Numerical Value (Cancelled transaction)
3. Data Pre-Processing
Handling Missing Value, Handling cancelled transaction, Handling Dummy Data, Making a retention column, Handling Outlier, Scaling & Encoding
4. Clustering
Model Evaluation, Model Clustering

5. Exploratory Data Analysis: After Clustering

Cluster Comparison, EDA Summary, Suggestion, Overall Conclusion

Customer Segmentation

After clustering process, there are 4 cluster that has been identified, here are the characteristic of each segments:



High-Turnover Customers

This segment can be categorized by the lowest retention in one month. Mostly bought a product that are related to baking equipment. They are having least in total purchase but they are the largest customer compared by other clusters.

Attached Customers

The segment that having higher retention rate compared to High-Turnover Customers but have familiarity in product interest such as baking equipment. The difference is they also like to buy interior furniture products. They are the second of the highest in total price per month.

Interior Furniture Customers

Smallest than other cluster but have the highest retention. They are mostly buying Interior Furniture products.

Cutlery & Glassware Customers

This segments mostly buying cutlery and glassware product. They are the second most low in retention and quantity. But are middle in total price.