Customer Segmentation Report

1. Project Overview

This project focuses on segmenting customers based on their demographics and purchasing behavior using K-Means clustering. The data originates from a marketing campaign dataset containing features such as income, age, spending on various products, and campaign responses.

2. Data Preprocessing & Feature Engineering

- Missing values in 'Income' were handled using KNN Imputer.
- A new column 'Total_Amount_spent' was created as a sum of all product-related expenditures.
- 'Age' was calculated using the 'Year_Birth' feature.

3. Suggestions for Improvement

- Normalize/standardize numerical data for better clustering.
- Use One-Hot Encoding for categorical features.
- Perform PCA for dimensionality reduction.
- Use silhouette score, elbow method for determining optimal number of clusters.
- Add visualizations of clusters using PCA or t-SNE.

4. Key Insights

- High-spending customers fall in the 30-50 age range.
- Wine and Meat products drive the highest spending.
- Income and Age are likely top contributors to customer segmentation.

5. Recommendations

- Focus marketing on younger, higher-income clusters for premium product upsell.
- Revamp strategies for low-engagement clusters.
- Introduce targeted offers and re-engagement emails for underperforming groups.

6. Tools Used

Customer Segmentation Report

Python, Pandas, Seaborn, Matplotlib, Scikit-learn