Task One (Data preprocessing and cleaning)

Executive Summary

This project involves data cleaning and preprocessing on the "Customer Personality Analysis" dataset to prepare it for further analysis and clustering. The primary goal was to structure, enhance, and clean the data to ensure quality and relevance for downstream analytics.

Key Steps in Data Cleaning and Preprocessing

1. **Data Loading and Inspection:**
    - Loaded the dataset from a CSV file using `pandas`, with tab-separated values.
    - Initial inspection was performed to understand the data structure and identify missing values.
2. **Missing Value Treatment:**
    - Identified missing values in the `Income` column.
    - Rows with missing `Income` values were removed to maintain data integrity.
3. **Date Conversion and Tenure Calculation:**
    - Converted the `Dt_Customer` column to datetime format.
    - Calculated `Customer_Tenure` as the number of days since each customer joined, relative to the most recent join date.

4. **Feature Engineering:**
    - **Age:** Derived from `Year_Birth` by subtracting it from the current year.
    - **Total_Children:** Sum of `Kidhome` and `Teenhome` to indicate total children at home.
    - **Total_Spending:** Aggregated spending across six product categories (`Wines`, `Fruits`, `MeatProducts`, `FishProducts`, `SweetProducts`, `GoldProds`).
    - **Total_Purchases:** Combined total of purchases across `Web`, `Catalog`, and `Store`.
    - **Total_Accepted_Cmp:** Sum of responses to all five marketing campaigns plus final response.
5. **Column Elimination:**
    - Removed less informative columns `Z_CostContact` and `Z_Revenue` to streamline the dataset.
6. **Descriptive Analysis:**
    - Performed statistical summary of the cleaned dataset to understand feature distributions and check for anomalies.
    - Initiated exploratory plots such as count plots of categorical variables to visualize distribution.