**Step 1: Data Import and Initial Exploration**

* **Load the data** into a Python environment using pandas.
* **Inspect columns** like Dt\_Customer (registration date) and Income to ensure they are imported correctly. Convert Dt\_Customer to a datetime format if necessary.
* Use df.info(), df.describe(), and df.head() to understand the structure and characteristics of the dataset.

**Step 2: Handle Missing Data**

* Identify missing values in Income using df.isnull().sum().
* Explore relationships between Income, Education, and Marital\_Status using aggregation (e.g., groupby).
* Impute missing Income values with the average income based on customers' Education and Marital\_Status.
* Clean and standardize categories in Education and Marital\_Status (e.g., address spelling issues or combine similar categories).

**Step 3: Feature Engineering**

* **Total Children**: Add a new column as the sum of children columns (e.g., Kidhome + Teenhome).
* **Age**: Calculate customer age from their Year\_Birth.
* **Total Spending**: Compute the total expenditure by summing all relevant product expenditure columns (e.g., Wine, Fruits, Meat, etc.).
* **Total Purchases**: Sum up the number of transactions across sales channels (e.g., NumWebPurchases + NumCatalogPurchases + NumStorePurchases).

**Step 4: Visualize Distributions and Treat Outliers**

* **Boxplots**: Visualize income, total spending, and age distributions to detect outliers.
* **Histograms**: Examine the spread of age, income, and spending.
* Apply **outlier treatment**:
  + For continuous variables, consider capping at the 1st and 99th percentiles or using the IQR rule.

**Step 5: Categorical Encoding**

* **Ordinal Encoding**: Apply to features with inherent order (e.g., education levels).
* **One-Hot Encoding**: Use for nominal categorical variables (e.g., Marital\_Status).

**Step 6: Correlation Analysis**

* Generate a **correlation heatmap** for numeric variables using seaborn.heatmap() to identify relationships between features.

**Step 7: Hypothesis Testing**

* Perform statistical tests to evaluate hypotheses:
  1. **Older individuals and in-store preferences**:
     + Use a Chi-squared test or ANOVA to test the relationship between age groups and purchase channels.
  2. **Customers with children and online shopping**:
     + Conduct a Chi-squared test to evaluate the relationship between children and NumWebPurchases.
  3. **Cannibalization of physical stores by online sales**:
     + Examine correlations or trends over time in NumStorePurchases vs. NumWebPurchases.
  4. **US vs. Rest of World in total purchases**:
     + Use a t-test or ANOVA to compare the means of total purchases between customers in the US and those elsewhere.

**Step 8: Advanced Visualizations**

1. **Top and bottom-performing products**: Create bar plots of product-wise spending.
2. **Correlation between age and last campaign acceptance**:
   * Use a scatter plot or boxplot to analyze.
3. **Country with the highest campaign acceptance**:
   * Group by country and plot acceptance counts.
4. **Children vs. total expenditure**:
   * Visualize with a scatter plot or boxplot.
5. **Education and complaints**:
   * Create a count plot of complaints by education level.