Customer Analytics Project Documentation Report

1. Introduction

The **Customer Analytics Project** is aimed at analysing customer behaviour, spending patterns, and engagement with promotional campaigns. The project combines exploratory data analysis (EDA) using Python and an interactive Tableau dashboard to provide insights into customer demographics, spending, purchasing behaviour across various product categories, and customer retention.

Objective

The main objective of the project is to:

1. **Explore** customer demographics and purchasing behaviour.
2. **Visualise** trends and key metrics like average spending, customer retention, and promotional engagement.
3. **Build an interactive Tableau dashboard** to allow users to filter and view the data from different perspectives.
4. **Analyse** customer segmentation, recency of purchases, and purchasing patterns across channels (web, catalog, store).

Tools and Technologies Used

- **Data Analysis**: Python (Pandas, NumPy, Matplotlib, Seaborn)
- **Data Visualisation**: Tableau
- **Notebook Environment**: Jupyter Notebook
- **Data Sources**: Customer demographics, purchasing behaviour, marketing engagement, and purchase channel data.

---

2. Data Description

The dataset used in this project consists of the following key areas:

People (Customer Demographics)

- **ID**: Unique identifier for each customer.
- **Year_Birth**: Year of birth.
- **Education**: Highest level of education attained.
- **Marital_Status**: Current marital status.
- **Income**: Annual household income.
- **Kidhome**: Number of children in the household.
- **Teenhome**: Number of teenagers in the household.
- **Dt_Customer**: Date the customer enrolled with the company.

- **Recency**: Number of days since the customer's last purchase.
- **Complain**: Whether the customer has made a complaint in the past two years (1 = Yes, 0 = No).

Products (Purchasing Behaviour)

- **MntWines**: Amount spent on wine in the last two years.
- **MntFruits**: Amount spent on fruits in the last two years.
- **MntMeatProducts**: Amount spent on meat in the last two years.
- **MntFishProducts**: Amount spent on fish in the last two years.
- **MntSweetProducts**: Amount spent on sweets in the last two years.
- **MntGoldProds**: Amount spent on gold-related products in the last two years.

Promotion (Marketing Engagement)

- **NumDealsPurchases**: Number of purchases made with a discount.
- **AcceptedCmp1-5**: Whether the customer accepted promotional offers in five different campaigns (1 = Yes, 0 = No).
- **Response**: Whether the customer accepted the offer in the most recent campaign (1 = Yes, 0 = No).

Place (Purchase Channels)

- **NumWebPurchases**: Number of purchases made through the company's website.
- **NumCatalogPurchases**: Number of purchases made using a catalog.
- **NumStorePurchases**: Number of purchases made directly in physical stores.
- **NumWebVisitsMonth**: Number of visits to the company's website in the last month.

---

3. Data Preprocessing

Loading and Cleaning the Data

The data was loaded into a Pandas DataFrame, and several steps were taken to clean the dataset:

- **Handling Missing Values**: Missing or null values were checked and handled. In most cases, rows with missing values were dropped, or simple imputation methods were used where appropriate.
- **Date Parsing**: Dates, like the customer enrolment date (`Dt_Customer`), were parsed and transformed to calculate the tenure of customers.
- **Recency Normalisation**: The `Recency` variable was used to determine how recently a customer made a purchase. We calculated the number of days since their last purchase and normalised the values accordingly.
- **Feature Engineering**: New features, such as total spending (`Total_Spending`), were created by summing up the different spending categories (e.g., `MntWines`, `MntFruits`, etc.).

Exploratory Data Analysis (EDA)

Various EDA techniques were used to uncover insights from the data:

- **Descriptive Statistics**: Calculated basic statistics like mean, median, and standard deviation for numerical columns.
- **Correlation Analysis**: Checked correlations between spending across different product categories.
- **Customer Segmentation**: Based on spending patterns and demographic features like income and family size.

---

4. Key Metrics and Calculations

Several key metrics were calculated for the analysis:

1. Average Spending Per Customer

This metric provides an insight into the average amount customers spent across all product categories.

```python
df['Average_Spending_Per_Customer'] = df[['MntWines', 'MntFruits',
'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
'MntGoldProds']].mean(axis=1)
```

2. Percentage of Customers Engaging with Discounts

Calculated as the proportion of customers who have made at least one purchase with a discount.

```python
df['Discount_Engagement'] = df['NumDealsPurchases'] / df['NumPurchases']
```

3. Customer Retention Rate

The retention rate measures the proportion of customers who return for subsequent purchases. This was calculated by examining the `Recency` feature and segmenting customers based on their purchase frequency.

4. Recency Analysis

Analysed the recency of customer purchases and how it impacts future spending behaviour. A binning approach was used to segment customers by recency, and insights were drawn based on customer activity patterns.

---

5. Tableau Dashboard

The final Tableau dashboard was created to provide an interactive view of customer insights:

Features of the Dashboard:

- **Customer Demographics**: Filters based on age, income, and family structure.
- **Spending Analysis**: Visualisation of spending across different product categories.
- **Engagement with Promotional Campaigns**: Pie charts and bar graphs showing the success rates of various marketing campaigns.
- **Customer Retention Trends**: A line chart tracking retention rates over time.
- **Purchase Channels**: Breakdown of purchases made through different channels (website, catalogue, in-store).
- **Filters**: Multiple filter options, including customer demographics, recency of purchases, product categories, and marketing campaign engagement.

The interactive elements of the dashboard allow users to view specific metrics for various customer segments and explore trends across multiple dimensions.

---

6. Insights & Findings

- **Customer Segmentation**: Customers can be grouped based on their spending across different product categories. High spenders typically engage more with promotional campaigns and tend to shop across multiple channels.
- **Promotional Engagement**: Customers who engage with discounts tend to spend more over time. Campaigns offering higher discounts show higher response rates.
- **Customer Retention**: The retention rate tends to decrease with increasing recency. Customers who made a purchase recently are more likely to return.
- **Purchasing Channels**: Web purchases are the most popular channel, but customers who engage with both online and offline channels show higher retention rates.

---

7. Conclusion

This project provided a comprehensive analysis of customer data, including their spending habits, response to promotional campaigns, and retention over time. By leveraging Python for data analysis and Tableau for visualisation, we were able to provide actionable insights that could inform marketing strategies, customer segmentation, and overall business decisions.

Future Work

- **Predictive Modelling**: Adding predictive analytics to anticipate customer churn or forecast future spending.
- **More Granular Segmentation**: Investigating specific customer segments (e.g., based on geographic location or age group).

- **Real-time Dashboard**: Integrating real-time data into the dashboard to monitor live customer behaviour and marketing campaign performance.