**EDA – Portfolio Project**

**Project Overview**

You're a marketing analyst and you've been told by the Chief Marketing Officer that recent marketing campaigns have not been as effective as they were expected to be. You need to analyze the data set to understand this problem and propose data-driven solutions.

**Dataset Provided**

1. ifood\_df.csv

The data set consists of 2206 customers of XYZ company with data on:

* Customer profiles
* Product preferences
* Campaign successes/failures
* Channel performance

**Column Details:**

* ID: Customer's Unique Identifier
* Year\_Birth: Customer's Birth Year
* Education: Customer's education level
* Marital\_Status: Customer's marital status
* Income: Customer's yearly household income
* Kidhome: Number of children in customer's household
* Teenhome: Number of teenagers in customer's household
* Dt\_Customer: Date of customer's enrollment with the company
* Recency: Number of days since customer's last purchase
* MntWines: Amount spent on wine in the last 2 years
* MntFruits: Amount spent on fruits in the last 2 years
* MntMeatProducts: Amount spent on meat in the last 2 years
* MntFishProducts: Amount spent on fish in the last 2 years
* MntSweetProducts: Amount spent on sweets in the last 2 years
* MntGoldProds: Amount spent on gold in the last 2 years
* NumDealsPurchases: Number of purchases made with a discount
* NumWebPurchases: Number of purchases made through the company's web site
* NumCatalogPurchases: Number of purchases made using a catalogue
* NumStorePurchases: Number of purchases made directly in stores
* NumWebVisitsMonth: Number of visits to company's web site in the last month
* AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
* AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
* AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
* AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
* AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
* Response: 1 if customer accepted the offer in the last campaign, 0 otherwise
* Complain: 1 if customer complained in the last 2 years, 0 otherwise
* Country: Customer's location

**Objectives:**

**Section 1: Data Analysis and Preprocessing**

* Are there any null values or outliers? How will you handle them?

Example:

1. Removing rows with outliers
2. Remove or impute missing values with a constant value (e.g., mean, median)

* Are there any variables that require type transformations?
* What are the unique values in each column?
* Are there any useful variables that you can engineer with the given data?

Example:

1. Variable Age in replacement of the variable Year\_birth
2. Variable Revenue\_generated as the total sum of the amount spent on the 6 product categories
3. Variable Total\_Purchases by adding all the purchases features
4. Variable TotalAmount\_Spent for each customer by adding all the Mnt\* features
5. Family variable as the sum of Kidhome + Teenhome + Marital\_Status

Marital\_Status can be imputed as {'Divorced':1, 'Single':1, 'Married':2, 'Together':2, 'Widow':1, 'YOLO':1, 'Alone':1, 'Absurd':1}

1. Variable Educationnal\_years as the total number of years of education the individual achieved according to its diploma
2. TotalCampaignsAcc as the total acceptance of advertising campaigns

* Do you notice any patterns or anomalies in the data? Can you plot them?

**Section 02: Exploratory Data Analysis (EDA)**

There are a number of different techniques that can be used in EDA, including:

* Univariate analysis
* Bivariate analysis

**Univariate analysis:** This involves analyzing individual variables to understand their distribution, central tendency, and variability. Common univariate analysis techniques include:

* Descriptive statistics
* Histogram
* Boxplot

**Bivariate analysis:** This involves analyzing the relationship between two variables to identify patterns and trends. Common bivariate analysis techniques include:

* Scatter plot
* Correlation analysis

**Section 03: Data Visualization**  
Plot and visualize the answers to the below questions.

* Which marketing campaign is most successful?
* Display the total amount spent by a customer in each product category.
* What is the average spending in each age group?
* Which products are performing best, and which are performing the least in terms of revenue? Analyze the data and plot a suitable graph to display a report on revenue generated by different products.
* Which Country has the greatest number of customers who accepted the last campaign.

**Bring together everything from Sections 01 to 03 and provide data-driven recommendations/suggestions.**