Dataset

Source: Kaggle - Mall Customers Dataset

(https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python/data)

- **Description**: This dataset contains customer demographic data and their annual spending in a mall, including columns like customer ID, gender, age, income, and spending score.

Business Scenario

MallCo, a bustling shopping hub, ***has noticed a decline in customer retention despite steady footfall. To address this, the management team wants to unlock actionable insights from their customer data to improve loyalty programs, and identify high-value customers.***

The dataset includes the following information about 200 customers:

- **CustomerID**: Unique identifier for each customer.
- **Gender**: Male (0) or Female (1).
- **Age**: Customer's age in years.
- **Annual Income (k$)**: Annual income in thousands of dollars.
- **Spending Score (1-100)**: A score that indicates spending habits and mall engagement.

Your role is to analyze the data, answer critical business questions, and recommend strategies for improving customer retention.

Business Questions to Answer

1. Who are our most valuable customers? What defines them?
2. Are there distinct customer groups with similar spending behaviors? How can we target them effectively?
3. What demographic factors (e.g., age, gender, income) influence spending habits?
4. What specific actions can MallCo take to improve retention and boost spending?

Some guidelines: <u>You can follow these for some hints or structure, but feel free to approach the problem in your own way. There is no one correct method to approach the</u>

problem.

1:Loading the dataset

import numpy as np

 # Load the dataset (replace 'dataset.csv' with the actual path to your file)
 # Assuming the data columns are in the order: CustomerID, Gender, Age, Annual Income, Spending Score

data = np.genfromtxt('dataset.csv', delimiter=',', skip_header=1)

You will notice the gender column is all NaNs. Why is that ? How can you solve for it ?

---

2. Understand Customer Demographics

- Calculate the **average Age, Annual Income, and Spending Score** to understand the typical customer profile.
- Analyze customer distribution by **gender**. Does one gender tend to spend more or earn more on average?
- **Hint:** Use NumPy operations like `mean()`, `min()`, and `max()` for this analysis.

2: Identify High-Value Customers

- Filter customers with a **Spending Score > 80** and calculate their **average Annual Income**.
- Identify the **top 10 customers by Spending Score**. What do they have in common (e.g., age group, gender)?
- **Hint:** Use boolean indexing and NumPy's `argsort()` for ranking.

3: Explore Relationships Between Features

- Compute the **pairwise correlations** between **Age**, **Annual Income**, and **Spending Score** to uncover key drivers of spending.
- Filter young adults (18-25) and calculate their average **Spending Score**. Compare this with older age groups.
- **Hint:** Correlation can be computed using NumPy's `corrcoef()`.

4: Customer Segmentation

- Group customers into three segments based on their **Spending Score**:
    - Low (1-33)
    - Medium (34-66)
    - High (67-100)
- Compute the average **Age** and **Annual Income** for each segment. What are the characteristics of each group?
- Propose marketing strategies for each segment.
- **Hint:** Use NumPy's slicing and aggregation techniques.

5: Find Behavioral Similarities

- Normalize the dataset (subtract the mean and divide by the standard deviation for each column).
- Compute the **Euclidean distance** between each customer and the customer with the highest **Spending Score**. Identify customers who exhibit similar behavior.
- **Hint:** Use NumPy's vectorized operations for distance calculations.

Typical Analyst Deliverable

1. **Analysis Report**
   A concise report addressing the business questions, including:
    - Key findings about customer demographics and spending patterns.
    - Characteristics of high-value customers and identified customer segments.
    - Insights on the relationships between age, income, and spending habits.
2. **Actionable Recommendations**
   Suggest 2-3 personalized marketing strategies for different customer groups, such as:
    - Discounts for low-spending customers to encourage more frequent visits.
    - Exclusive loyalty programs for high-spending customers.
    - Targeted advertising for specific age groups or income brackets