Buyer Persona and Clustering Analysis - Documentation

# Introduction

This project performs two main tasks:  
1. \*\*General Buyer Persona Calculation\*\*: It calculates the general profile of the buyer using summary statistics such as average age, most common gender, profession, salary range, and nationality.  
2. \*\*Customer Segmentation via K-Means Clustering\*\*: It uses K-Means clustering to segment customers based on numerical features like age, transaction amount, and total transactions, then analyzes each segment.

# 1. General Buyer Persona Calculation

The following buyer attributes are calculated from the dataset:  
- \*\*Average Age\*\*  
- \*\*Most Common Gender\*\*  
- \*\*Most Common Salary Range\*\*  
- \*\*Most Common Province\*\*  
- \*\*Most Common Profession\*\*  
- \*\*Most Common Nationality\*\*  
- \*\*Average Transaction Amount\*\*  
- \*\*Total Number of Transactions\*\*  
  
These statistics give a general overview of the most common traits of buyers in the dataset.

# 2. K-Means Clustering for Customer Segmentation

K-Means clustering is applied to the dataset to segment customers into distinct groups based on their age, transaction amount, and total number of transactions. The following steps are followed:  
  
- \*\*Data Preparation\*\*: Select the relevant numerical features (Age, Transaction Amount, Total Transactions) and handle missing values.  
- \*\*Standardization\*\*: Use `StandardScaler` to standardize the data before applying K-Means clustering.  
- \*\*Clustering\*\*: Apply K-Means clustering to group customers into 3 clusters (adjustable).  
- \*\*Cluster Analysis\*\*: For each cluster, the average age, most common gender, average transaction amount, and total number of transactions are analyzed.

# Usage

1. Load the dataset from an Excel file.  
2. Run the persona calculation script to get the general profile of buyers.  
3. Run the K-Means clustering script to segment customers and analyze each segment based on the provided features.