Customer Segmentation Analysis Project

Overview: In this project, I conducted a comprehensive analysis of customer segmentation using the Mall Customers dataset. The goal was to identify distinct customer segments based on their demographic and spending behavior, enabling businesses to tailor marketing strategies and improve customer engagement.

Steps Taken:

1. **Dataset Understanding:**
    - Utilized a dataset containing information about mall customers, including CustomerID, Gender, Age, Annual Income, and Spending Score.
    - Explored the dataset's structure, including the number of records and columns, as well as the data types of each column.
2. **Exploratory Data Analysis (EDA):**
    - Conducted EDA to understand the dataset's structure and distribution.
    - Checked the first few rows of the dataset (`df.head()`) to get a glimpse of the data.
    - Computed summary statistics (`df.describe()`) to gain initial insights into the central tendency and spread of numerical variables.
3. **Univariate Analysis:**
    - Analyzed individual variables such as Age, Annual Income, and Spending Score separately using distributions and KDE plots.
    - Plotted histograms and KDE plots (`sns.distplot()`) to visualize the distribution of each variable.
    - Examined the skewness and kurtosis of each variable to understand their distributional characteristics.
4. **Bivariate Analysis:**
    - Explored relationships between pairs of variables, particularly focusing on potential correlations or patterns between Annual Income and Spending Score using scatterplots and pair plots.
    - Utilized scatterplots (`sns.scatterplot()`) to visualize the relationship between Annual Income and Spending Score.
    - Created pair plots (`sns.pairplot()`) to visualize pairwise relationships between multiple variables simultaneously.

5. **Clustering Analysis:**
    - Employed various clustering techniques to segment customers into distinct groups based on their characteristics.
        - Conducted univariate clustering on Annual Income alone to identify income segments.
        - Performed bivariate clustering on a combination of Annual Income and Spending Score to create clusters based on both income and spending behavior.
        - Executed multivariate clustering using scaled and transformed data including Age, Annual Income, Spending Score, and Gender.
        - Used K-means clustering (KMeans()) on the combination of Annual Income and Spending Score.
        - Utilized the K-means algorithm (KMeans()) to cluster customers based on their Annual Income.
        - Standardized numerical features using StandardScaler() and encoded categorical variable Gender using pd.get_dummies().
6. **Optimal Cluster Selection:**
    - Determined the optimal number of clusters for each clustering method using the elbow method.
    - Plotted inertia scores against the number of clusters and identified the "elbow point" where adding more clusters doesn't significantly reduce inertia.
7. **Visualization and Interpretation:**
    - Visualized the identified clusters and their centroids using scatterplots.
    - Color-coded data points based on their assigned clusters for easy interpretation.
    - Analyzed the characteristics of each cluster, including mean values of different attributes, to understand distinct customer segments.
8. **Documentation and Reporting:**
    - Documented findings, methodologies, and insights obtained from the analysis.
    - Prepared a comprehensive report or presentation summarizing the project's objectives, methodologies, key findings, and recommendations.
    - Included visualizations, tables, and statistical summaries to support the analysis and communicate key insights effectively.

Outcome: This project provides valuable insights into customer segmentation, enabling businesses to tailor their marketing strategies, enhance customer satisfaction, and drive business growth effectively.