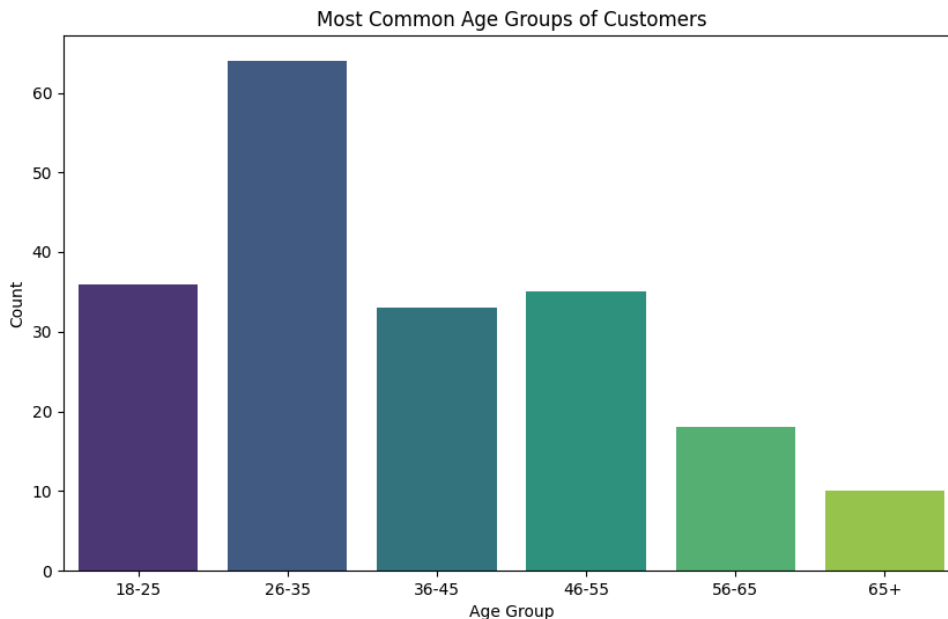


# Customer Segmentation - EDA and Clustering Report

---

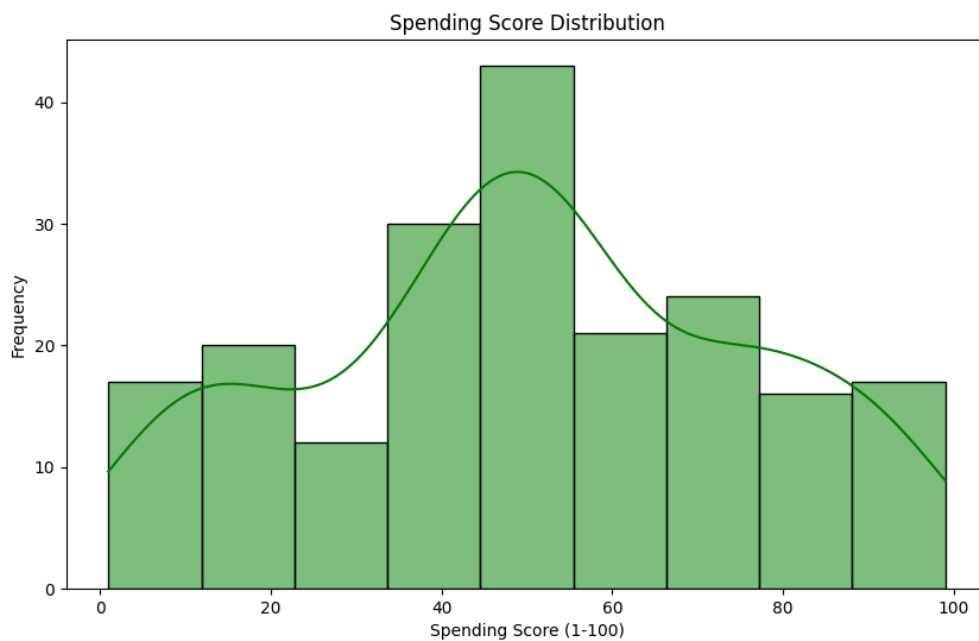
## 1. Most Common Age Groups of Customers :

- This bar chart displays the distribution of customers across various age groups. The "26-35" age group is the most dominant, comprising the highest number of customers, followed by the "18-25" age group. The "65+" group represents the least number of customers.
- Understanding the distribution of customers by age helps businesses tailor their marketing strategies. Knowing the most dominant age groups can inform decisions on product offerings, promotions, and engagement tactics targeted toward different age ranges.



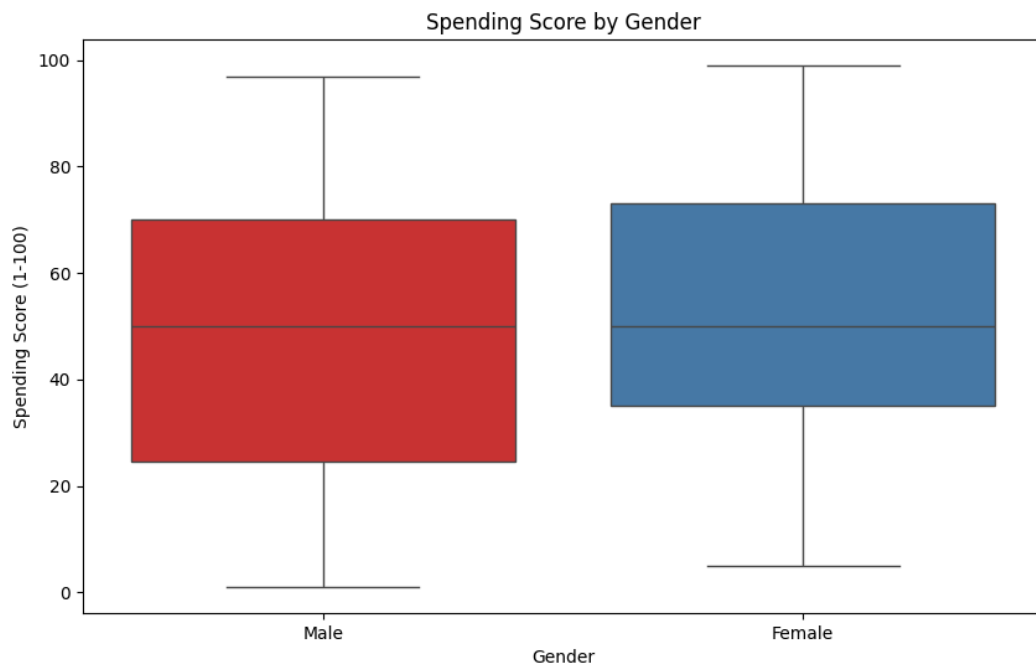
## 2. Spending Score Distribution (Histogram):

- This histogram represents the frequency distribution of customers' spending scores, which range from 1 to 100. The plot shows a fairly even distribution of spending scores across the population, with a slight peak observed in the 30-50 range.
- This distribution is helpful for understanding customer behavior in terms of their spending tendencies. It identifies a balanced segment of customers across all spending levels, which can be useful for designing customized marketing campaigns or identifying high-value customers.



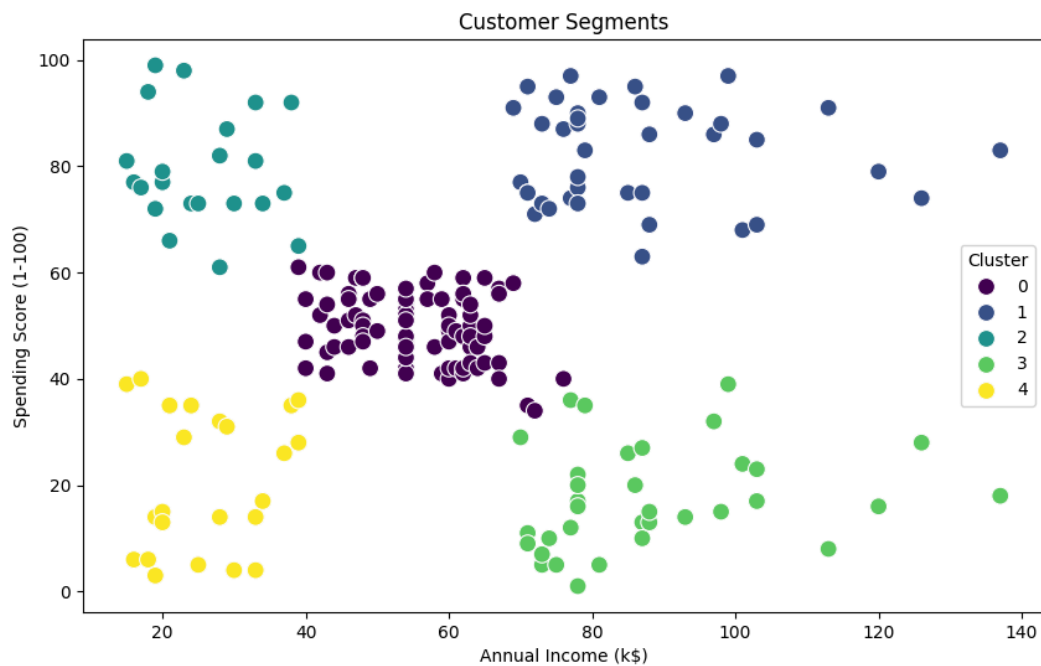
### 3. Spending Score by Gender (Box Plot):

- This box plot visually compares the distribution of spending scores between male and female customers. Males generally exhibit lower spending scores on average compared to females, but both genders display similar interquartile ranges.
- The box plot highlights that both genders show a similar spread of spending behavior, with no significant outliers present. This can help understand gender-based differences in spending patterns, guiding targeted campaigns or promotions designed for specific genders.



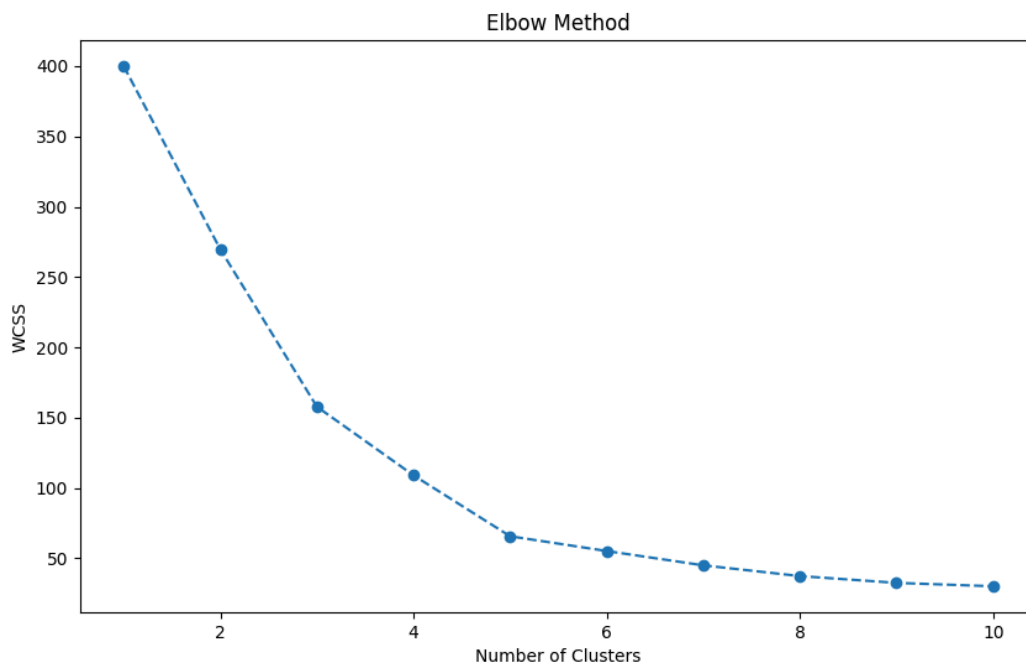
#### 4. Customer Segments (K-means Clustering):

- This scatter plot illustrates the segmentation of customers based on annual income and spending score using the K-means clustering technique. Different customer clusters are colored distinctly to showcase how customers are grouped according to their financial behaviors.
- The separation of clusters is visually clear, indicating the K-means algorithm's effectiveness in segmenting customers into meaningful groups. Each cluster represents customers with similar income and spending scores, which can help businesses personalize their marketing strategies based on distinct customer segments.



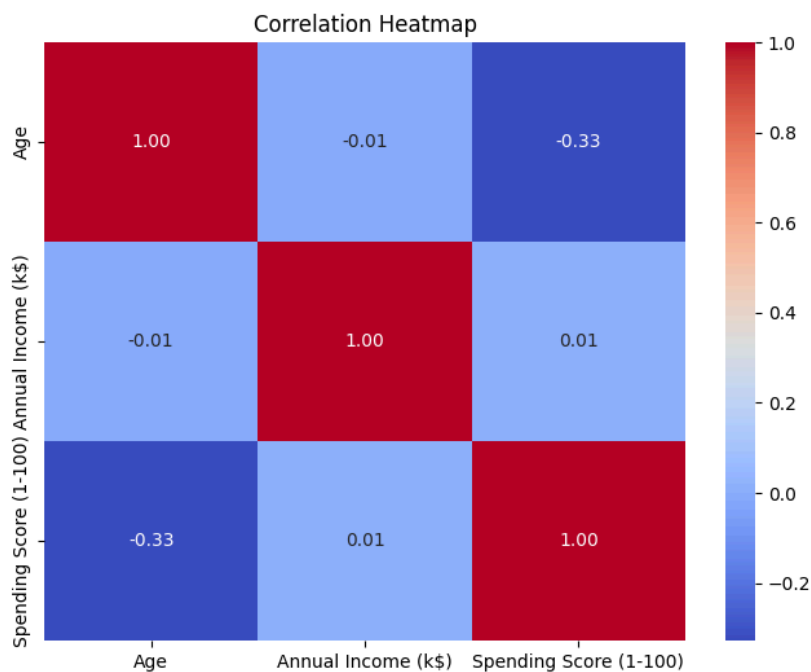
## 5. Elbow Method (K-means Clustering):

- The elbow method plot assists in determining the optimal number of clusters for K-means. The Within-Cluster Sum of Squares (WCSS) is plotted for different cluster numbers. A sharp decrease in WCSS is observed as the number of clusters increases, but the rate of decrease slows after 5 clusters, suggesting that 5 clusters is the optimal choice.
- Identifying the optimal number of clusters is crucial for creating well-defined customer segments. A well-chosen number of clusters helps ensure the model's predictive power and relevance to real-world applications, where too many or too few clusters may not yield actionable insights.



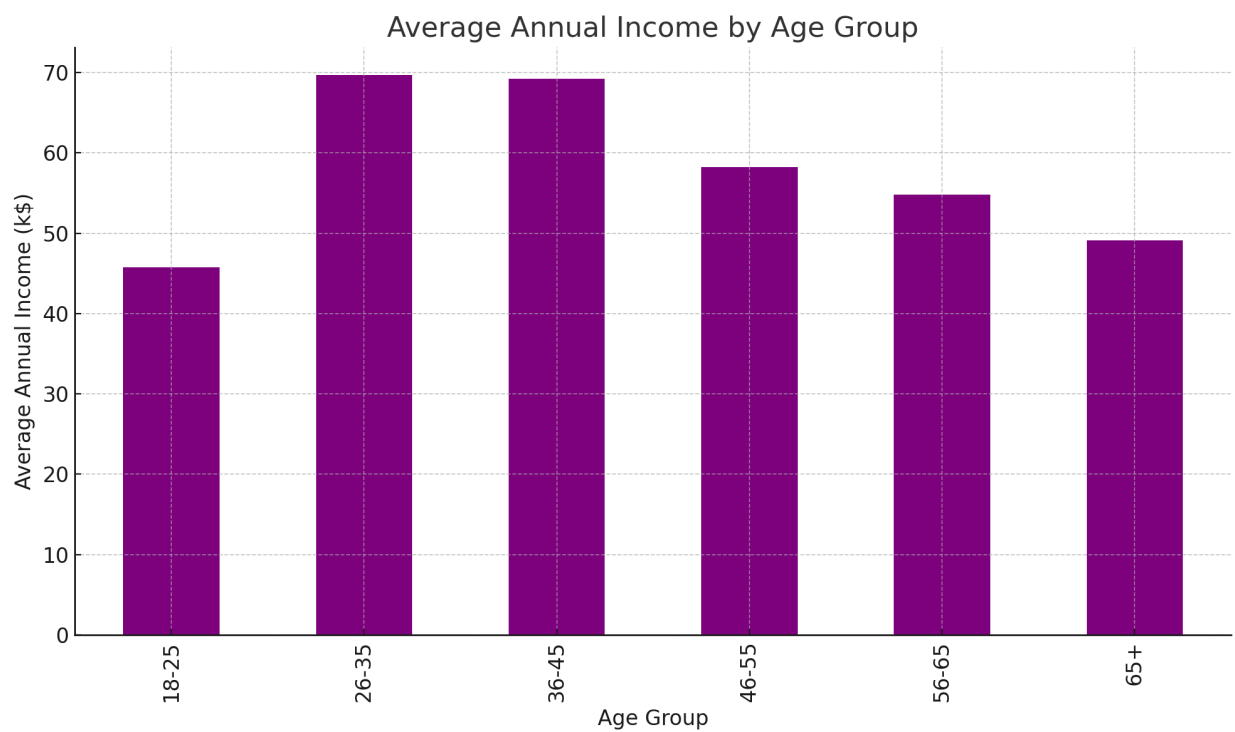
## 6. Correlation Heatmap:

- The correlation heatmap shows the relationships between key variables in the dataset, including Age, Annual Income, and Spending Score. The correlations between these variables are weak to very weak, with the highest correlation being between Annual Income and Age.
- This insight indicates that there is minimal relationship between spending scores and other variables such as age and income. Understanding the lack of strong correlations helps businesses focus on other variables or customer behaviors that may drive spending behavior more effectively.



7. Average Annual Income by Age Group (Bar Chart):

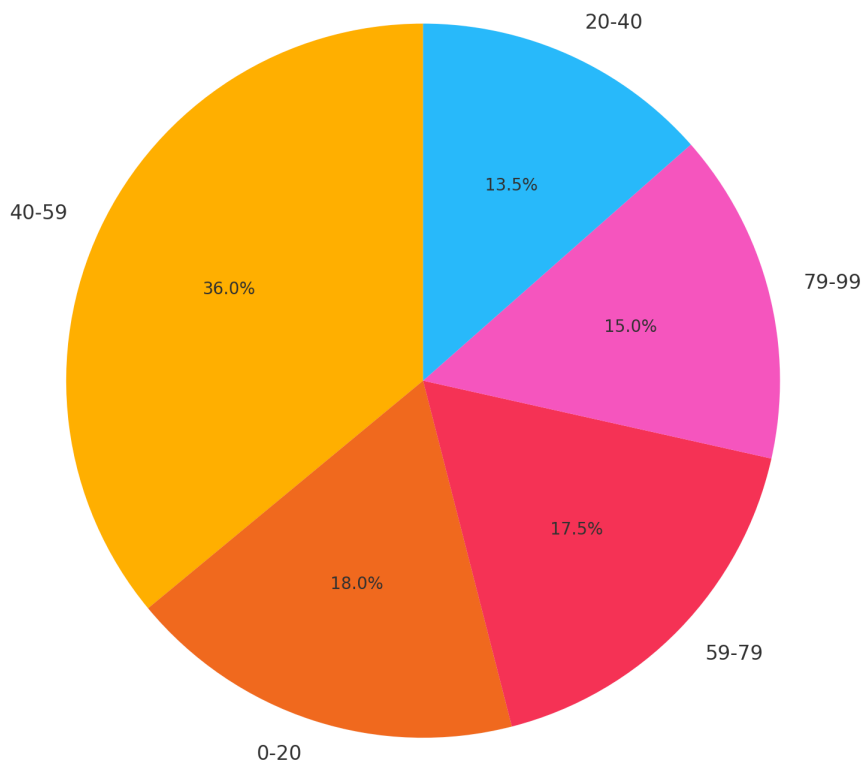
- This bar chart displays the average annual income for each age group. The "26-35" and "36-45" age groups have the highest average incomes, while the "18-25" and "65+" groups have comparatively lower incomes.
- This analysis can provide businesses with valuable insights into the income distribution among different age segments. Understanding income differences can help refine targeting strategies and product pricing for different customer groups.



## 8. Customer Distribution by Spending Score (Pie Chart):

- The pie chart categorizes customers into different spending score ranges (0-20, 20-40, etc.), with the largest share in the "40-59" range. Other ranges like "0-20" and "59-79" also represent notable portions of the customer base.
- This chart gives a clear picture of how customers are distributed based on their spending tendencies. It helps businesses understand the overall spending behavior and identify which customers are more likely to spend higher or lower, enabling more targeted promotional strategies.

Customer Distribution by Spending Score





## 9. Dendrogram (Hierarchical Clustering):

- The dendrogram is used to visualize the results of hierarchical clustering. It groups the data points based on their similarity, which is represented by vertical lines. Customers that share similar features are grouped together in clusters, and these clusters are visually shown in different colors.
- The hierarchical clustering method is helpful for understanding how customers naturally cluster together based on multiple features. The dendrogram allows businesses to explore relationships among different customer segments and decide on how to group customers for more personalized marketing strategies.

