**CLUSTERING AND FITTING**

**NAME: Tarun Reddy Salevenkayalla**

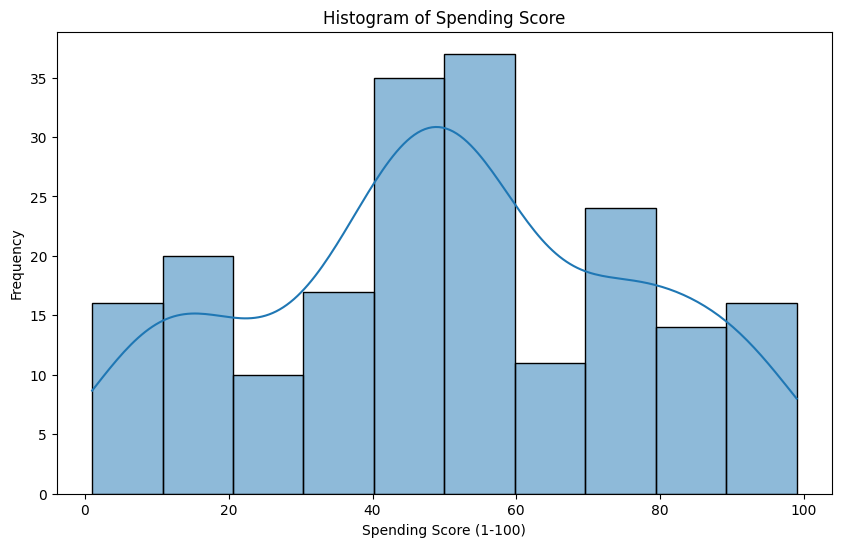
**ID: 23098031**

**Git Hub:**

**Introduction**

Clustering and fitting techniques play an important role in data analysis because they group similar data points and build patterns in those groups. Using Mall Visiting Customer Data, these methods uncover customers' behavior, preferences, and trends, allowing businesses to tailor their marketing approach, optimize service offerings, and improve customer experiences based on key segments and characteristics.

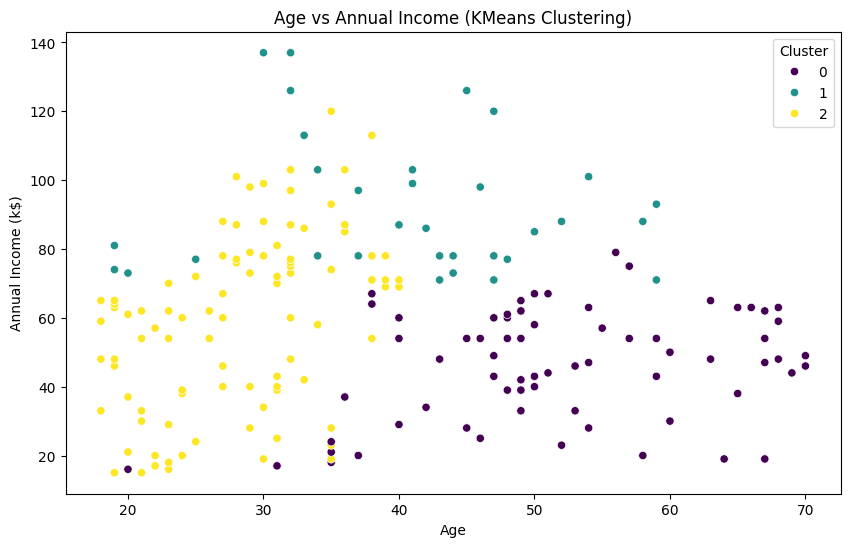
# Histogram of Spending Score

****

**Figure 1:Histogram plot**

This histogram graphs the distribution of spending scores (1-100) on the visitors to malls so that the frequency of all different scores can be established. With 10 bins and a kernel density estimate for a smooth curve, it underlines spending patterns by which businesses can identify predominant groups of spending and tailor strategies in accordance with specific customer groups.

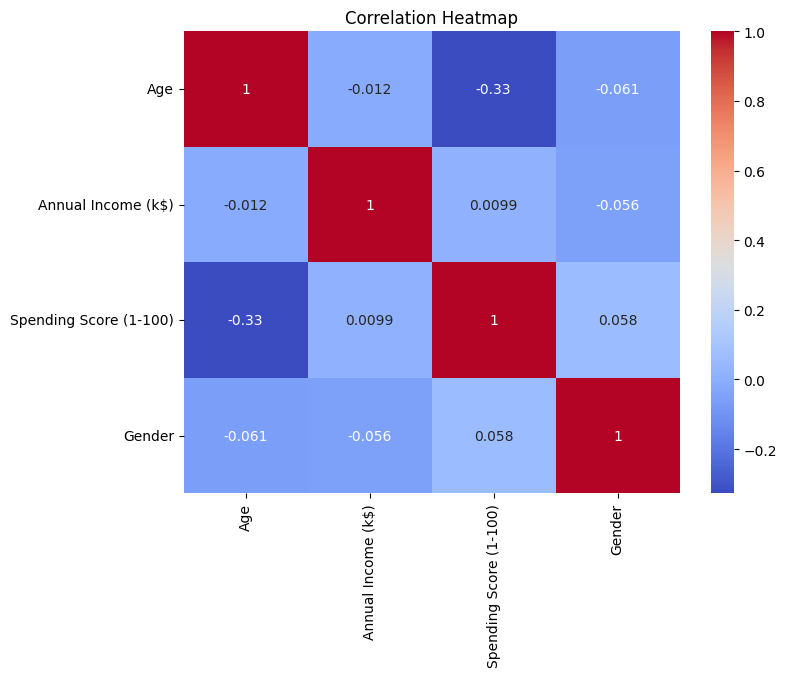
# Age vs Annual Income (KMeans Clustering)

****

**Figure 2: KMeans clustering**

This scatter plot represents the age and annual income relationship, colored according to KMeans clustering results. The application of different cluster colors helps to identify different customer segments using these two variables. The visualization is easy to interpret for quick identification of age and income patterns to assist in targeted marketing strategies.

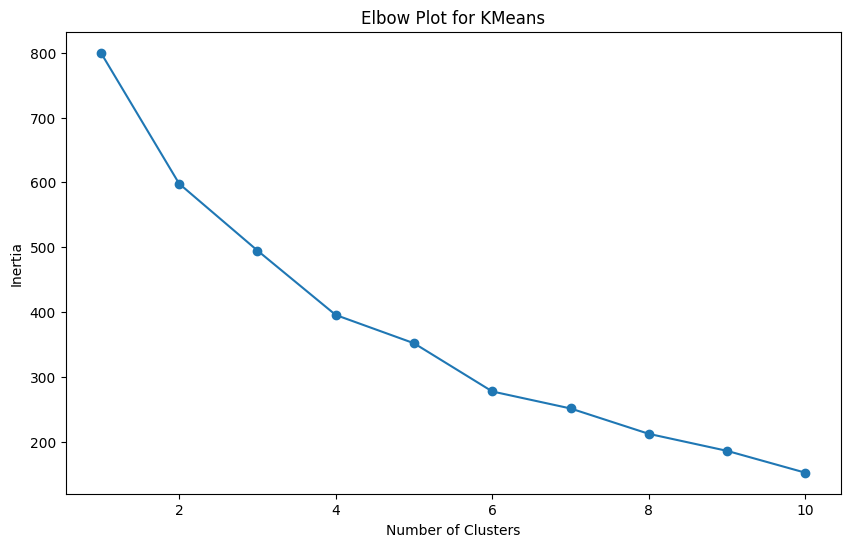
# Correlation Heatmap

****

**Figure 3: Heatmap**

This heatmap shows the correlation matrix for ***“age, annual income, spending score, and gender”***. Annotated values with a 'coolwarm' color scheme help indicate the strength and direction of relationships between variables. For example, the positive correlation of spending score with income and age indicates patterns in customer behavior that may help make data-driven decisions and target marketing efforts.

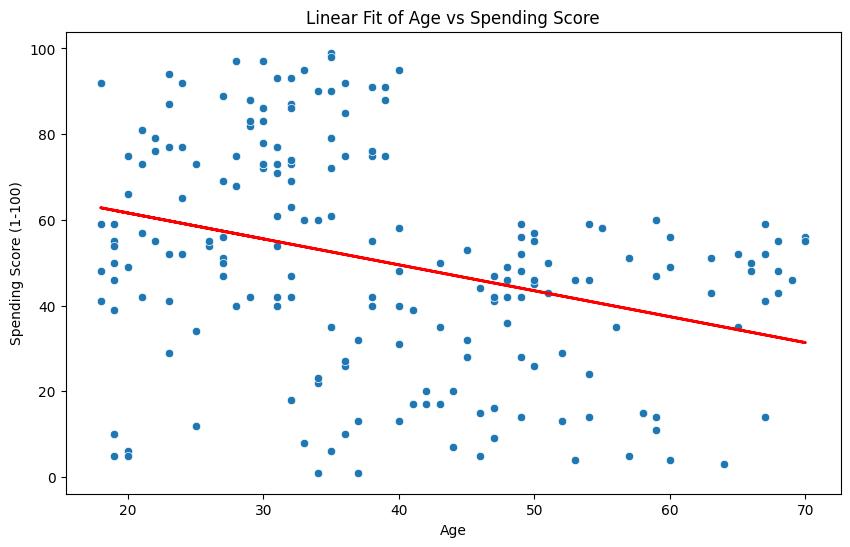
# Elbow Plot for KMeans

****

**Figure 4: Elbow plot**

The elbow represents the finding the ideal number of clusters for KMeans clustering. It plots inertia, or within-cluster sum of squared distances, against the number of clusters. A sharp decrease in inertia is seen until a certain point; after that, the decline slows down. This "elbow" is a good point that can help further data segmentation.

# Linear Fit of Age vs Spending Score

****

**Figure 5: Linear fit**

The linear regression line, fitting the relationship between age and spending score, has shown the trend where the age variable may influence spending behavior. A scatter plot, accompanied by a red regression line, denotes whether a linear relationship prevails. This model analysis aids in exploring possible correlations, thereby helping businesses understand consumer behavior patterns, including age-targeted marketing directions.

# Silhouette Score



**Figure 6: Silhouette score**

The silhouette score of around 0.26 is moderately well-separated with some overlap between clusters. A score closer to 1 would be indicative of well-defined clusters, but a score of around 0.25 leaves room for improvement. This calls for a reconsideration of the clustering approach perhaps with alternative algorithms or feature adjustments that can better group the data into more distinct clusters.