

Applied Data Science – 1

Name : Dhanesh Kanakaraj

Student ID : 23056970

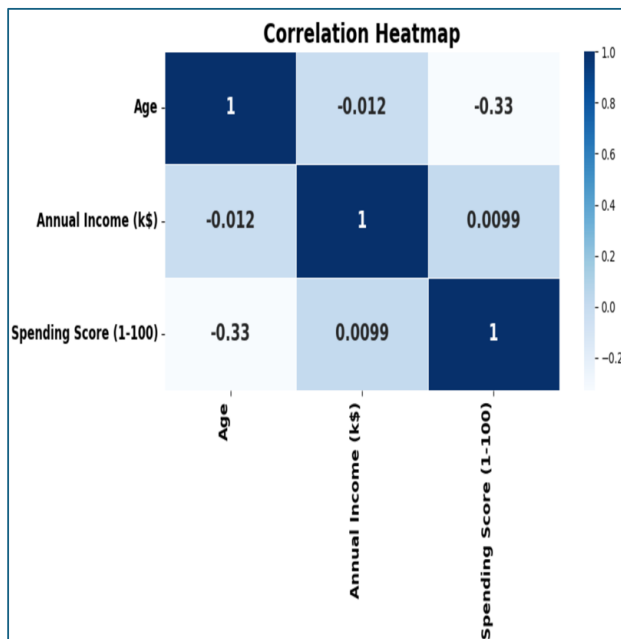
GITHUB : <https://github.com/kdhanesh619/Clustering-and-Fitting-Assignment.git>

Assignment : Clustering and Fitting

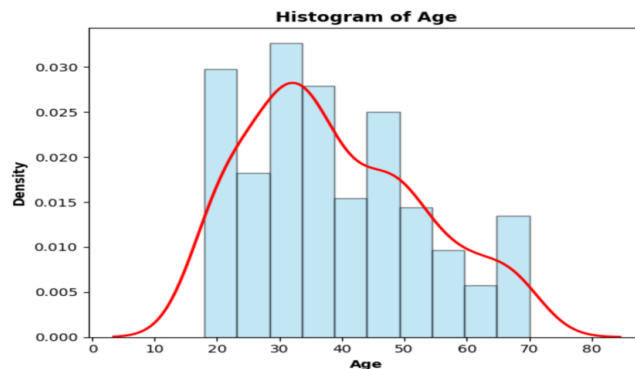
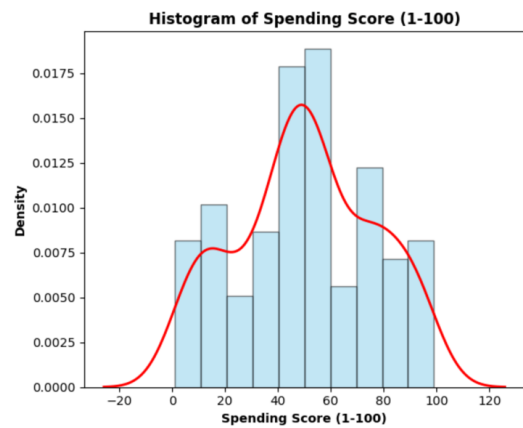
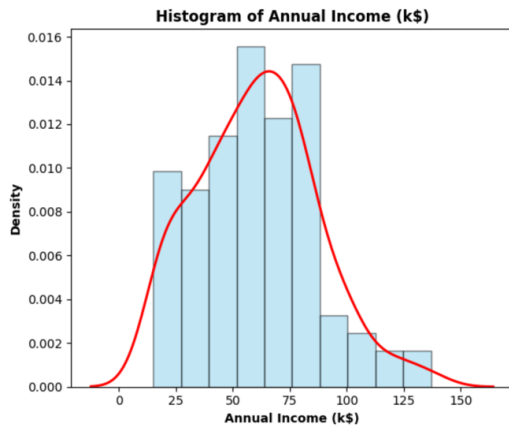
Prediction of Customers Visiting Mall

INTRODUCTION

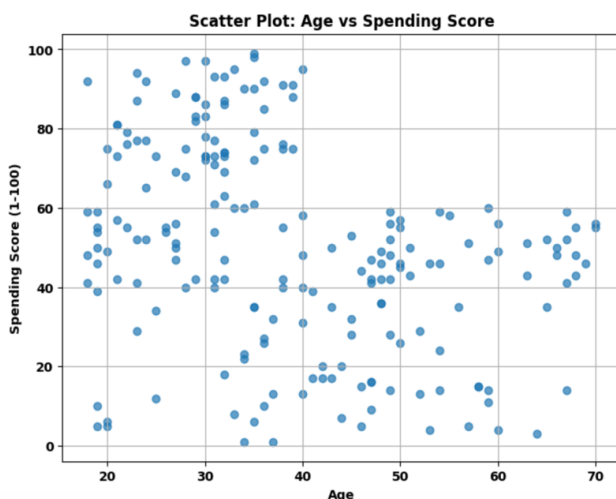
This report is about understanding the different kinds of people who go to malls and what makes them visit. The data includes information like age, gender, income, and spending habits, which are analyzed to uncover patterns in customer behavior. Using these insights, we aim to identify the key groups of customers who are more likely to visit the mall and spend, helping businesses plan better marketing strategies. This analysis also provides a clear picture of who the target customers are and how to engage with them effectively. By looking at statistical trends and associations in the data, this report gives the importance of knowing customer needs to create better shopping experiences and improve business success.



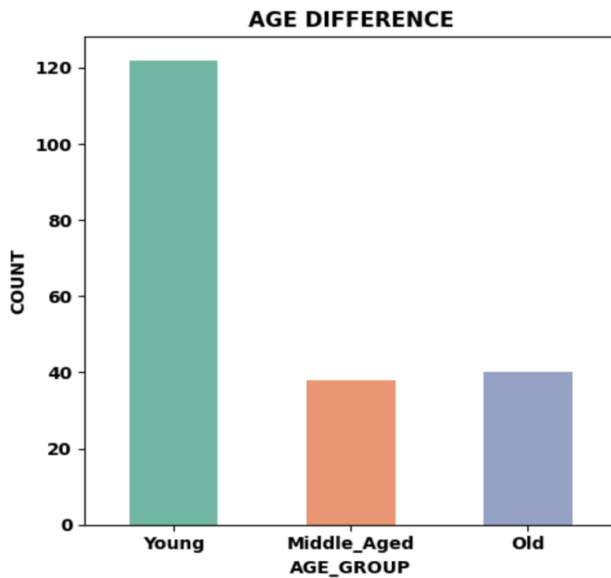
The heatmap helps us understand how age, income, and spending are connected to each other. It shows that older people tend to spend less, as there is a negative relationship between age and spending score (-0.33). This means younger people are more likely to have higher spending scores. Income doesn't seem to have much of an effect on spending, as the correlation is very weak (0.0099). Also, there's almost no connection between age and income (-0.012), which means they are mostly independent of each other. These insights give us a clearer picture of how these factors influence customer behavior, helping us understand the spending habits of different groups.



The histogram plots provide a detailed overview of the data related to age, annual income, and spending scores, with some statistical observations. The Age plot shows that most people are aged between 20 and 50, with the highest point around 30 years, indicating a slight left-skewed distribution. This shows that younger to middle-aged individuals form the main group of customers. The Annual income plot shows that annual income is balanced, with most people earning between 50,000 dollars and 75,000 dollars, showing a symmetric distribution with little skewness, which aligns with a middle-income people. The Spending scores plot focuses on spending scores, which are spread across the range but are peaked around 50, suggesting a normal distribution. These statistical insights, such as skewness and spread, help businesses better understand their target customers, enabling them to create better deals and plans to attract and keep customers happy.

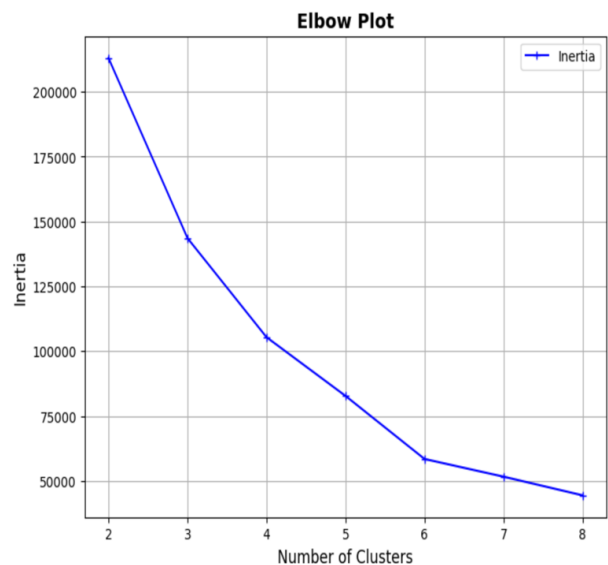


The scatter plot shows how age and spending are connected. Younger people, especially those below 40, have more variety in their spending, with many of them spending a lot (above 80). Older people, particularly those above 50, usually have lower spending scores, mostly below 50. This means that as people get older, they tend to spend less. Younger people show more ups and downs in their spending, while older customers are steadier with their spending, mostly on the lower side. This helps us understand how spending habits change with age.



The bar chart shows the distribution of customers based on their age groups. Young, Middle-aged, and Old. The "Young" group has the highest count, with over 120 individuals, providing the largest share of customers. The "Middle-aged" and "Old" groups have fewer customers, with similar counts below 50. Statistically, this suggests a positively skewed distribution, where the younger age group dominates the customer base. This insight highlights the importance of targeting younger customers in marketing strategies.

The elbow plot helps us to figure out how to group customers into meaningful categories. The curve shows that the best balance happens around 4 to 6 clusters, where the inertia drops but doesn't reduce too much with more clusters. The silhouette scores, which tell us how well the clusters are separated, start low at 0.29 for 2 clusters and improve, reaching the highest score of 0.45 for 6 clusters. This means 6 clusters would be best to group customers. When we combine this with earlier insights like younger people spending more and makes the biggest customer group it becomes clear how these groups can be used. Businesses can use this information to create personalized offers and plans for each group, making customers happy and improving their experience.



CONCLUSION

In conclusion, this report highlights key insights into mall customers based on their age, income, and spending habits. Younger customers, especially those under 40, dominate the customer base and tend to spend more, while older customers spend less and make up a smaller portion of the audience. The analysis showed clear patterns, such as the connection between age and spending and the independence of income from spending habits. Using clustering, we identified 6 meaningful groups of customers, helping businesses better understand and target their audience. These findings provide a strong foundation for creating personalized marketing strategies, improving customer satisfaction, and enhancing the overall shopping experience at the mall.