SUPERMARKET MARKETING STRATEGY

INTRODUCTION :

The objective of this analysis is to provide insights to supermarket owner, which customer is spending more money on shopping at that supermarket based on information Annual_Income , Age and Gender collected through membership card .This will help owner to do better marketing based on customer who is having better spending_Score, which will further help them to improve sales.

DATASET :

This dataset has four numeric variables (Customer_Id, Age, Annual_Income, Spending_Score) and one categorical variable (Gender). Spending_score is the target variable or dependent variable in this dataset.Age and Gender are independent variables .

DATA INSPECTION :

Dataset has five columns and Two-Hundred rows with zero missing values. In this dataset I have renamed two attributes Annual Income (k$), Spending Score (1-100) with Annual_Income, Spending_Score

---

DATA VISUALIZATION :

Box plot represents Age, Annual_Income and Spending_Score on X-axis and count on Y-axis. It has outliers on Annual_Income column represented between 120-140 and median are represented between 20-40, 60-80, and 40-60 for Age, Annual_Income and Spending_Score respectively.

Sub plots of scatter plots are created between Spending Score vs Gender, Age, Annual Income. It is visible that both the genders have almost equal spending_score and for age between 20 - 40 the spending score high compared to other age groups. Those with annual income range between 20 -40k and 80 - 100k USD has the highest spending score compared to other annual income ranges.

MODELING:

OLS using one independent variable

Based on this model, 'Annual_Income' does not appear to be a statistically significant predictor of 'Spending_Score.' The low R-squared and high p-value suggest that the model does not provide a good fit to the data, and the inclusion of 'Annual_Income' does not improve the model's ability to explain the variance in spending scores.

OLS using two independent variable

The model with 'Age' alone had a similar R-squared value, the addition of 'Annual_Income' did not significantly improve the model's explanatory power. The p-value for 'Annual_Income' suggests that it is not a statistically significant predictor in this context. Depending on your goals, you might consider evaluating the model with and without 'Annual_Income' and assessing which model performs better based on relevant criteria.

Random Forest Regressor

The Random Forest Regressor trained on the given features (Age and Annual Income) has a relatively high mean squared error (MSE) of 478.88, indicating a considerable variance between predicted and actual Spending Scores. Additionally, the low R-squared value of 0.03 suggests that the model explains only a small proportion of the variability in the target variable. Further refinement or consideration of additional features may be necessary to improve predictive performance.

A/B Testing

The A/B testing results indicate a statistically significant difference in spending scores between the control and treatment groups, with a t-statistic of -14.7698 and a p-value of 0.0000. The negative t-statistic suggests that the treatment group has a significantly lower mean spending score than the control group. These findings suggest that the marketing campaign may have had a notable impact on customer spending behavior, although practical significance should be considered alongside statistical significance for a comprehensive assessment.

CONCLUSION

In summary, the analysis aimed to assist the supermarket owner in identifying high-spending customers based on attributes such as Annual Income, Age, and Gender.

Modeling efforts, including OLS regression and Random Forest, indicated challenges in predicting Spending Scores. The Random Forest Regressor exhibited a high MSE and low R-squared, suggesting room for improvement. Single variable Annual_Income is not helping much in predicting Spending_score.A/B testing results, however, highlighted a significant impact of a marketing campaign on spending behavior. In conclusion, while predictive models may need refinement, the marketing initiative appears to have positively influenced customer spending.