## FUNDAMENTALS OF MACHINE LEARNING

### FINAL PROJECT REPORT

### Professor Chaojiang (CJ) Wu, Ph.D.

## Kent State University

Submitted By

Marneni Abhinay Chiranjeeth

ANALYSIS OF CUSTOMER’S PURCHASE

BEHAVIOR­

**INTRODUCTION**

Nowadays, to satisfy the customers, the shopping mall business facing most intense competition in the modern age. The purpose of consumers choose to shop at malls rather than traditional markets while looking to buy satisfactory goods or services. Customers are allowed to choose and select based on their opinions of a certain good or service and their post-purchase experiences at particular locations. For the purpose of gathering data and conducting a thorough analysis of the study, we selected a sample of Mall customers data. A systematic questionnaire was used to gather primary data. The results were drawn using the K-means and linear regression approach. To ascertain the association between the variables, demographic and correlation analysis were also utilized. All of the variables were significantly connected, according to my analysis. I observed that customers to visit shopping malls for enjoyment include the shopping environment, comfort of shopping, availability of various products, entertainment provided at malls, good product quality, discount, and sales promotion

**STATEMENT OF THE PROBLEM**

After analyzing given data, I approach a particular problem that I would solve with k means clustering. The Shopping Mall mainly depend on customer, based on the preferences of the general population, which encompasses all groups that span both genders and a range of ages. Numerous shopping centers have popped up as a result of rising demand. Increasing customer activity and sales are always important for shops, whether they are organized or not. But given the way that online retail and e-commerce are evolving right now, both customers and sales are trending down, giving retailers especially large mall stores options. The modern customer wants the best value for their money and is more aware, informed, and prepared with good research. To understand how customers shop and what affects them and which age group gender mostly spends on shopping. The primary focus of the project is on how well we can understand and offer solutions to the Analysis of Mall Customers' Purchase Behavior after completing certain analyses and data segmentation.

**DATA DESCRIPTION**

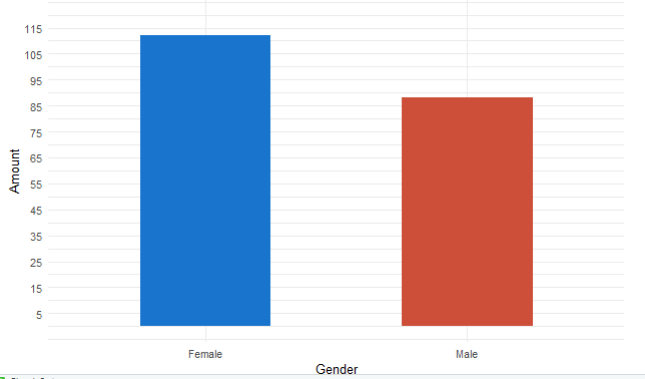
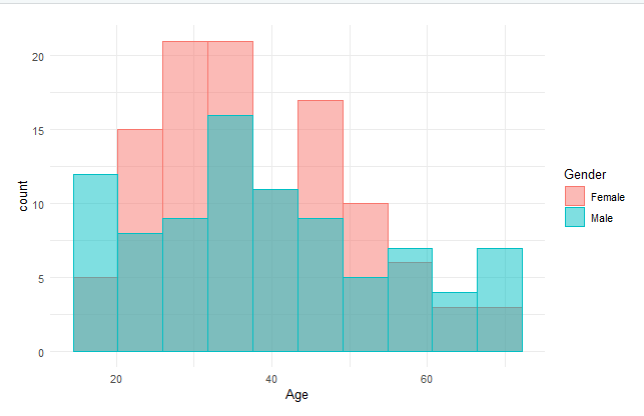
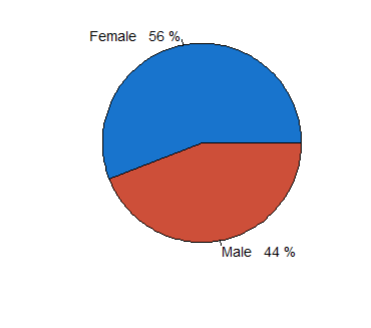
## The data was taken from the Kaggle website which consists of fundamental data about the customers (ID, age, gender, income, spending score).

## Data Exploration and Data preparation

Import the data into R studio, the given data contain a gender column so for that, so I convert the gender column into dummy variables. And I removed the Customer ID column from the dataset. I and use Standard Scale to standardize the data by normalizing it. I obtained a dataset of mall customers. Later, I analyzed the data using a K means cluster.

**PLOTS AND PER CHART OF DATASET**

Before discussing K means I want to show a few plots and per chart based on the given dataset containing a more majority of female customers than male customers. and creating the plot base on gender and range of Age. So that the K mean method may easily evaluate the dataset.

. 

**ANALYSIS AND DISCUSSION**

**K means Clustering:**

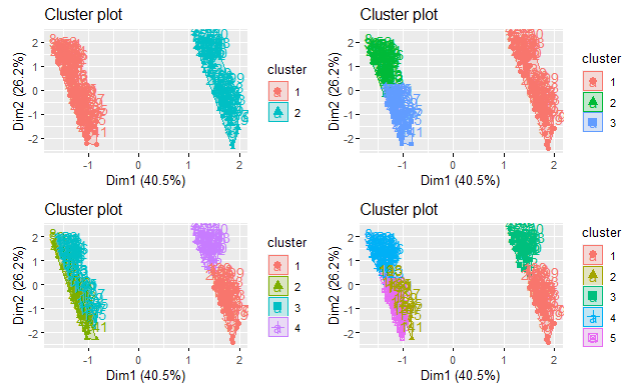
The collection of data sets is separated by K-means into various subgroups (clusters), while I earlier had to manually adjust K numbers to see how the dataset clusters. k values were randomly selected from 2, 3, 4, and 5. 

Fig: The dataset clusters of randomly values 2, 3, 4, and 5.

To Determine the optimal number of cluster present in the given dataset I used the elbow method.

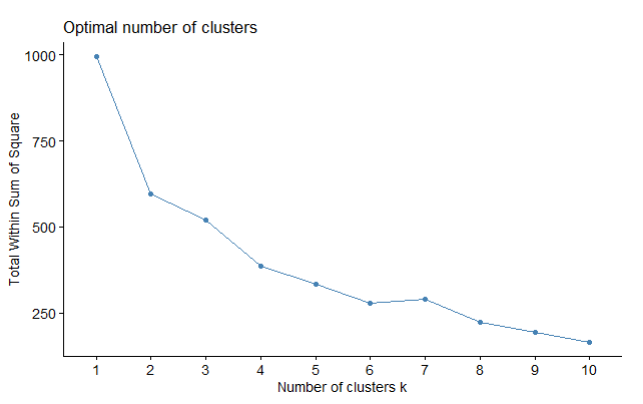


Fig: Graph of Elbow method

For our analysis, I decided to use k-means clustering. To identify the most suitable clusters for our task, I used elbow approaches. From the graph above, I determined that k=2. As you can see in the graph below, I have decided to create 2 clusters for my project.

Fig: Cluster of K value of 2

The clusters divide the data uniformly and each cluster contains the Customer details which use such as Gender, Age, Annual Income, and Spending Score.

Finding the cluster with the highest mean in each column when the data is grouped by clusters is the major goal.

Fig: Means values of clusters

From the above tables, we can say that

Cluster 1:

\*Customers with an average age of 28 years have been grouped into the first cluster.

\*This group of customers has the maximum spending score of 69 percent, which indicates that young people tend to spend more in malls than aged people.

\*Among all the customers of this age group, females are 58% and males are 42%. This indicates that females spend the most at malls compared to men.

\*Surprisingly, the average annual income of both clusters is close to 60k dollars per annum.

Cluster 2:

\*This group consists of customers whose average age is 48 years.

\*The spending score of this group of customers is very low compared to the first cluster.

\*Even in this cluster, females take up a majority of the percentage.

\*The average percentage of females who spend at malls is 53% whereas males are 47%.

\*The average income of this group of customers is also close to 60k dollars per annum.

For better interpretation, check for any correlation between the spending score variable and other variables using basic linear regression.

Using a Simple Linear regression model on spending score variables. Using Simple Linear regression model on spending score variables.

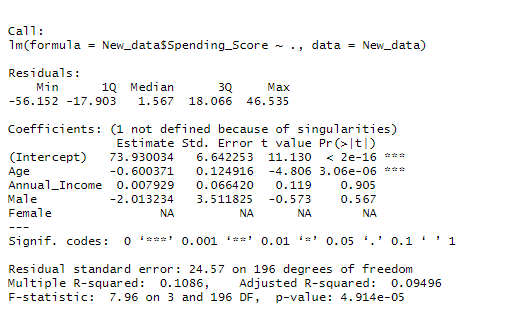


Fig: Simple Linear regression model-1

As can be shown, the spending score and age have a substantial correlation (as the P value is too small). We can infer that there is a negative correlation because we have a negative slope. The spending score and all other variables have relatively low variation, as shown by the R square value of 10.86%.

Implementing a simple linear regression model on spending score and age variable based on the results of the above model.

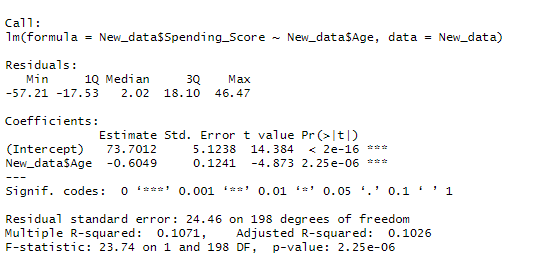


Fig: Simple Linear regression model-2

After implementing the model results show a significant correlation between spending score and age. Concerning the above model, the P value and R square values are rather similar to the values observed.

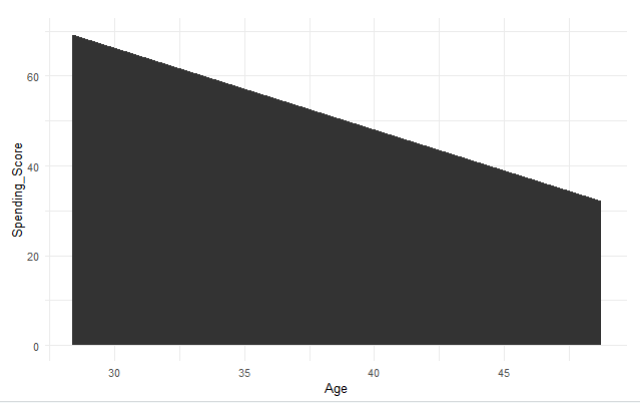
showing the correlation between age and the spending score variable by using Plot. 

Fig: Graph of Age vs Spending Score

According to the graph above, mall customers' expenditure is dropping as their age increases.

**CONCLUSION**

Consumer perceptions on shopping in malls are the focus of the present analysis, For finding consumer purchasing behaviors that differ according to opinions, Customers under the age of 30 are the ones who are spending the most, according to an analysis of the mall's customers' purchasing behavior.

Additionally, more women than men spend money at the mall across all age categories.

Therefore, it would be wise to have stores linked to ladies like apparel and beauty products in malls to get more customers to spend money there. This would increase the mall's profitability.

An overview of the connections between customer behavior and Mall management. Managing the target market and the organization as a whole exploration an understanding of consumer behavior. It is the responsibility of a marketer to understand customer behavior in order to change it through the development of goods, services, and product promotion that are in line with the customer's goals and preferences. In this approach, shopping malls enhance the value of the firms for which they serve, increase customer satisfaction. individuals who purchase.

**Dataset Link**: <https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python>