The objective is to segment the mall customers into different groups(clustering) so that better marketing campaigns and strategies can be adopted and can target the right type of customers.

Firstly, preparation of data was needed as shown in the code:

1. There were no duplicates
2. Gender variable was converted into a numeric binary variable. After converting the data type of "Gender" variable, 0 stands for "Female" and 1 stands for "Male"
3. Checking the data summary suggested the need for normalizing some variables as the age, annual income and spending score:

   Min-Max method was used to normalize the data to get values between 0 and 1 for all variables.

Secondly, 2 methods of clustering were used to be able to segment the mall customers into groups sharing similar features:

1. K-means Clustering:

1.1. Clustering was based upon all variables except for the customer ID which would have no meaning if used in the segmentation process. Different number of clusters were tested to know the number of clusters that will be used: After plotting the total within cluster sum of squares against the number of clusters as shown in figure 1, 6 clusters were found appropriate (The elbow method) since increasing the number of clusters above 6 will not cause a significant improvement in the total within cluster sum of squares.

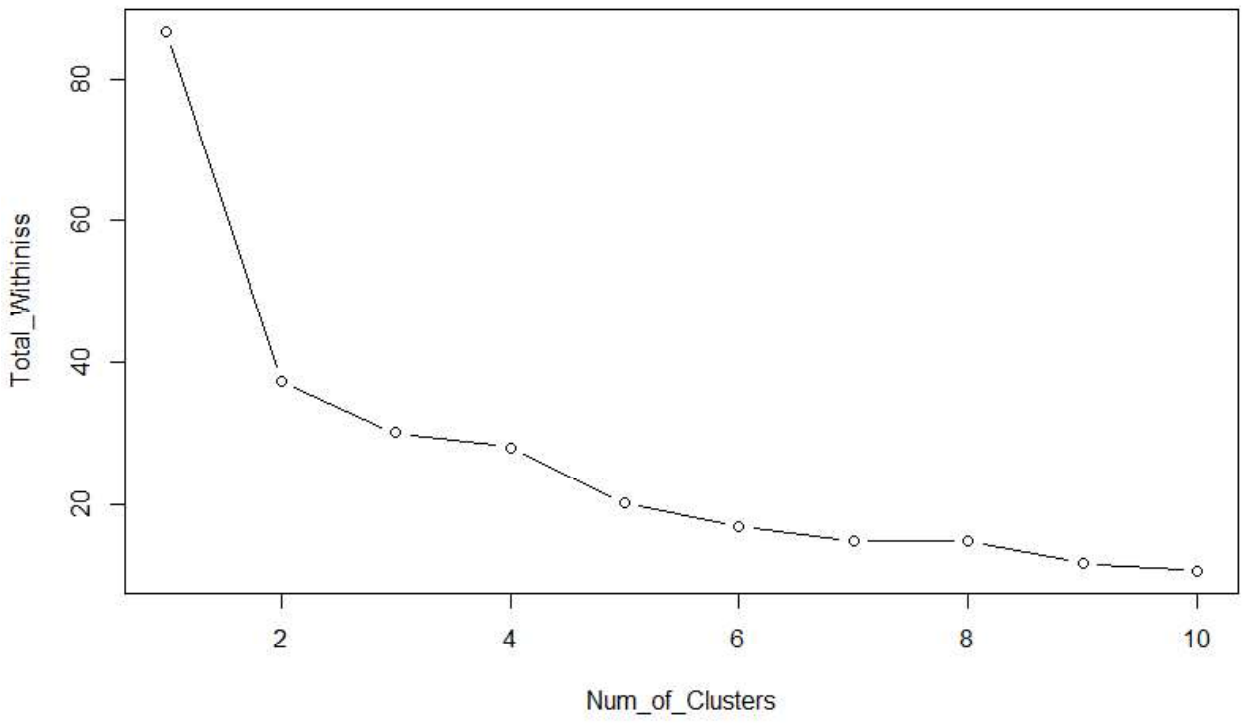

Figure 1: K-means Clustering

1.2. Analyzing the clusters obtained:

Cluster	1	2	3	4	5	6
Gender	Females	Females	Males	Males	Males	Males
Mean Age	28	48	33	59	38	25
Mean Annual Income ($k)	60	59	87	49	83	41
Mean Spending Score	68	35	83	40	14	61
No. of Customers	57	55	18	26	20	24

1.3. Insights:

Segmenting the mall customers into groups knowing which customers belonging to each group (cluster) is very beneficial to marketing team for targeting the right customers:

For example:

1. Advertisements on young Male products with reasonable prices can be targeted to customers in cluster 6, since they are male young customers with the lowest mean annual income (e.g. wallets, male perfumes, fashion products … etc.). On contrast, male products with higher prices can be targeted to cluster 3 since they have high mean annual income.

2. Marketing Team may decide to show more customized advertisements to cluster 5 because they have high mean annual income but low spending score so it possible to use advertisements customized to male-related products suitable for the age group around 38 years old to be able to exploit these high income group of people to increase their spending group

3. If a new cosmetic product protecting against ageing is available, the marketing team may decide to show advertisement to customers in cluster 2 (since they are females and having a mean age of 48 years old). Also, reasonable medium prices will be preferred to this segment since they have medium annual income

4. It is preferable to keep customers in clusters 1, 3 and 6 always updated with new products since they have high spending scores and seems to be frequent customers that are more likely to generate profits to the mall and also building

strong relations with these customers is highly recommended so that they become loyal customers who may bring in more customers to the mall

5. The variability in males segment of customers is higher than that in females segment since males are divided into 4 groups while females were divided only on 2 clusters and thus higher customization of advertisements will be needed for male customers than female customers based upon other features like age group and annual income

2. Hierarchical Clustering:

2.1. Hierarchical Clustering was also based upon all the variables except for the customer ID. 6 Clusters (shown in figure 2) were also used as in k-means clustering.

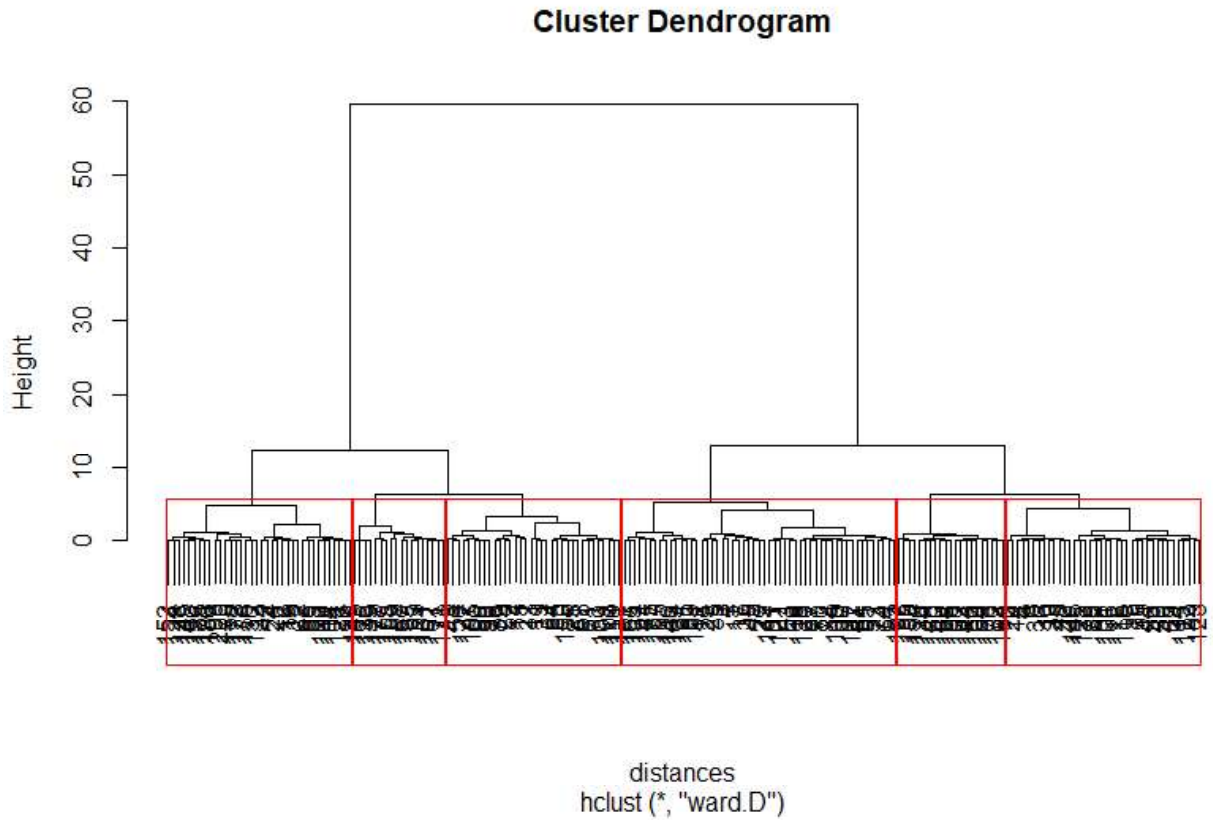

Figure 2: Hierarchical Clustering

2.2. Analyzing the clusters obtained:

Cluster	1	2	3	4	5	6
Gender	Males	Males	Females	Females	Females	Males
Mean Age	53	28	48	27	32	39
Mean Annual Income ($k)	46	65	58	47	86	86
Mean Spending Score	41	74	35	59	82	12
No. of Customers	34	36	53	38	21	18

2.3. Insights:

Segmenting the mall customers into groups knowing which customers belonging to each group (cluster) is very beneficial to marketing team for targeting the right customers:

For example:

1. Marketing Team may decide to show more customized advertisements to cluster 6 because they have high mean annual income but low spending score so it possible to use customized advertisements to male-related products suitable for the age group around 39 years old to be able to exploit these high income group of people and to increase their spending score
2. Customers in cluster 4 are very important since they have high mean annual income and high spending score which means they are more likely to generate profits to the mall than others and should always be kept updated with appropriate customized advertisements.
3. If a new cosmetic product protecting against ageing is available, the marketing team may decide to show advertisement to customers in cluster 3 (since they are females and having a mean age of 48 years)
4. It is preferable to keep customers in clusters 2 and 4 always updated with new products since they have high spending scores and seems to be frequent customers.