**Machine Learning Unsupervised Learning 1**

**Individual Assignment – 1**

### Paritosh Sinha

**As mentioned in the problem statement -**

**Objective**: To identify segments of passengers that have similar characteristics for the purpose of targeting different segments for different types of offers.

**Entities being clustered**: Passengers who belong to an airline’s frequent flier program. **Attributes on which clustering is being done on**: Characteristic of the passengers such as mileage history, miles spent etc.

**In order to achieve**: Targeting different segments for different offers in order to increase customer satisfaction such that more and more customer engage with EastWestAirlines for both flight and non-flight services.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

# Answer a.

Before starting our clustering be it Hierarchal or K-Means we need to **standardize** the data because in order to measure similarities between observations and in order to form clusters we use the distance metric.

If we don’t standardize then features with high range of values will have a bigger influence on the clustering. Magnitude of distance can be influenced by one dimension.

Another problem associated which arises is every dimension can be measured in different scales and have different unit(metric).

Therefore, standardization is required before building a clustering model which make sure all dimensions of our data are treated equally. As we want each column to have the same

impact on the distance.

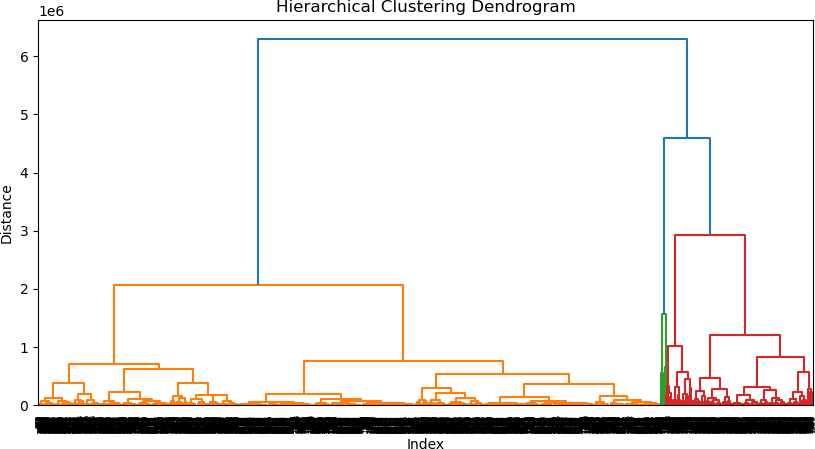
In short normalising and standardising is done to bring all the data on the same scale for capturing the equal effect of all the dimensions in the data.



If we go through the above table, you will find that the values in dimension

Balance is from 0 to 1704838 whereas for Bonus\_trans is from 0 to 86 so the impact or effect of Balance will be high with respect to any other field so we standardize all the data, such that each of the dimension in the dataset will have similar impact.

Below is the Dendogram for the raw data as seen in the image clusters will be very different and will not have proper information.



# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Answer b.

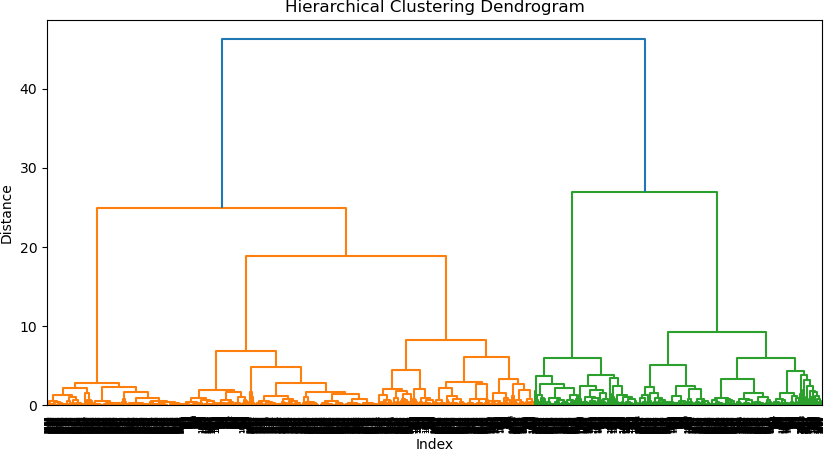
Hierarchical Clustering can be done using different methods like: Single linkage, Complete linkage, Centroid linkage, Average linkage and Wards

Distances which we can use for clustering are Euclidean distance, Manhattan and Mahalanobis

Here as per requirement we are doing clustering on the basis of **Ward’s** method and using

**Euclidean** distance

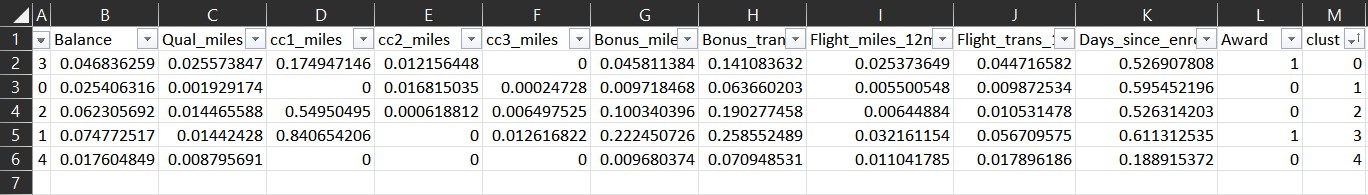
On applying hierarchical clustering with Euclidean distance and Ward’s method we observed following dendogram:



Looking at the dendogram we can go with as minimum as 2 clusters and go as granular as 15 clusters and above, but then I choose to draw the line at distance 10 because as going with 5 clusters looked like an optimum choice. Below image appears which highlights the formation of 5 clusters.

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Answer c.

***Centroid for the clusters which are created as follows:***



Characteristics – On the basis of Cluster characteristics as obtained in the above table using cluster centroids. I’ve assigned names in the level of the preciousness of the gemstones.

***Opal -> Emerald -> Sapphire -> Ruby -> Diamonds***

Cluster 0:

## Fly Ruby

Received Free Flight (Award), Medium Flight miles, Low Bonus Miles no CC3 Usages. Customers in this cluster have received free flight and mostly using our flight services.

Cluster 1:

## Fly Emerald

Not Received Free Flight (Award), Low Flight miles, Low Bonus Miles, no CC1 Usages Customers in this cluster have received free flight and are not using our both flight and non-flight services

Cluster 2:

## Fly Sapphire

Not Received Free Flight (Award), Low Flight miles, High Bonus Miles, High CC1 usage

Customer is an Old customer but they are using our non-flight services.

Cluster 3:

## Fly Diamond

Oldest customers, Received Free Flight (Award), High Flight miles, High Bonus Miles Customers in this Cluster are Old Customers but they are extensively using our flight & non-flight services.

Cluster 4:

## Fly Opal

New Customer, Not Received Free Flight (Award), Medium Flight miles, Low Bonus Miles

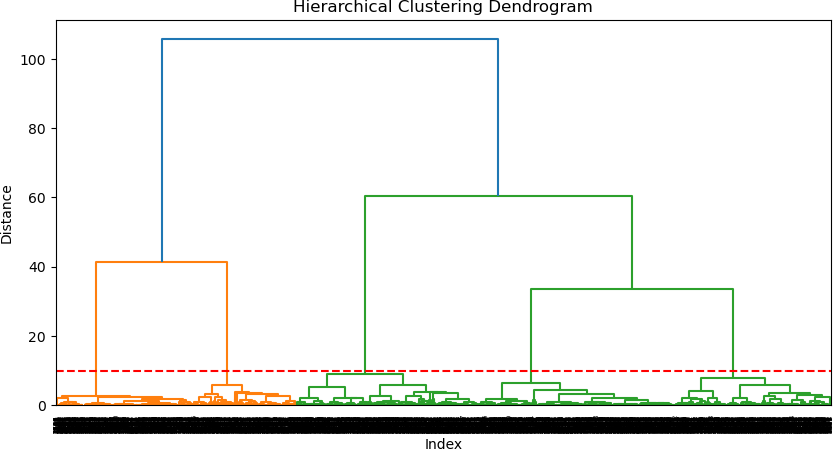
Customers in this cluster are new and are frequently using our flight services.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

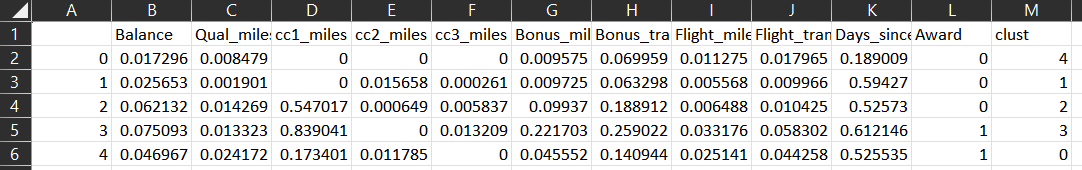
# Answer d.

As suggested in the question we are taking 95% of the data and check for the stability of the cluster which were formed.

On doing the same exercise on the subset data, same number of clusters emerge which is five in my case.



Checking for the same by extracting cluster centroid for the subset data



Clusters centroid are also suggesting that the clusters which are formed with 95% of the entire data are also having similar characteristics.

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Answer e.

#### K-Means Clustering

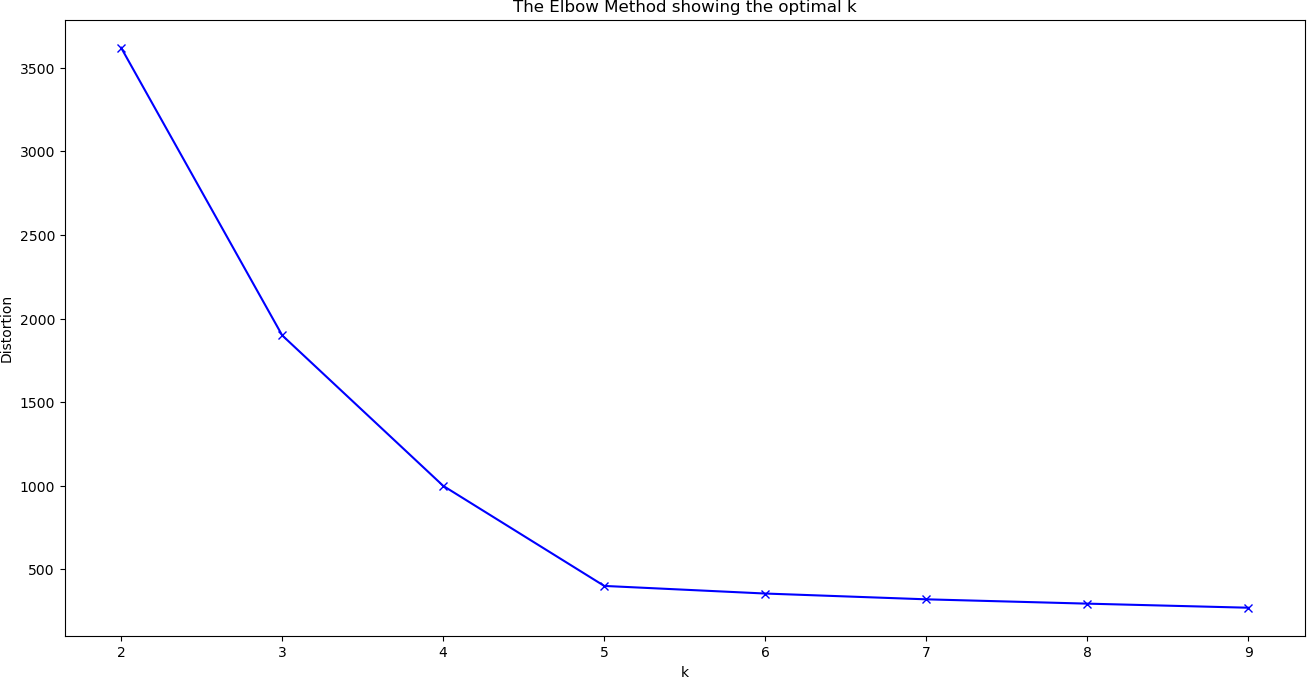
For K-Means Clustering

For finding optimum number of clusters -- **Elbow method**

Observation is finding for that am considering to go with **6 Clusters**

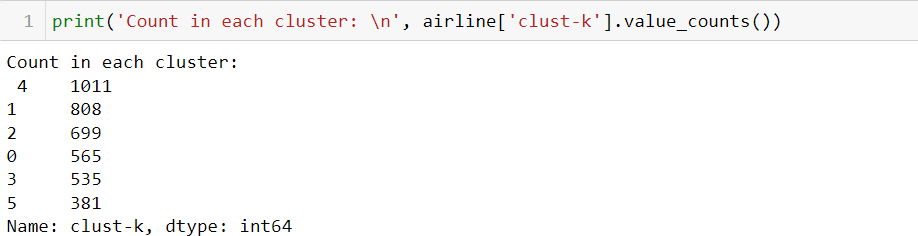
As elbow method is an effective technique for determining the optimal number of clusters. The rate of decrease is showing slow down between 5 and 6. The point at which the rate of decrease slows down is called the elbow point which is point number 5 in my case. The optimal number of clusters is the number that corresponds to the elbow point which can be deduced as point number 6 here.

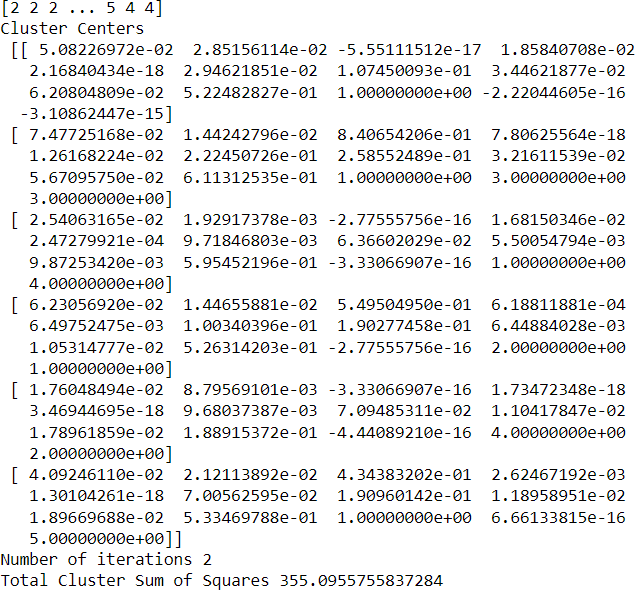
So Optimal number of clusters which we should go with is 6.



Flattening of the elbow curve suggests that by adding one more cluster will not increase homogeneity.

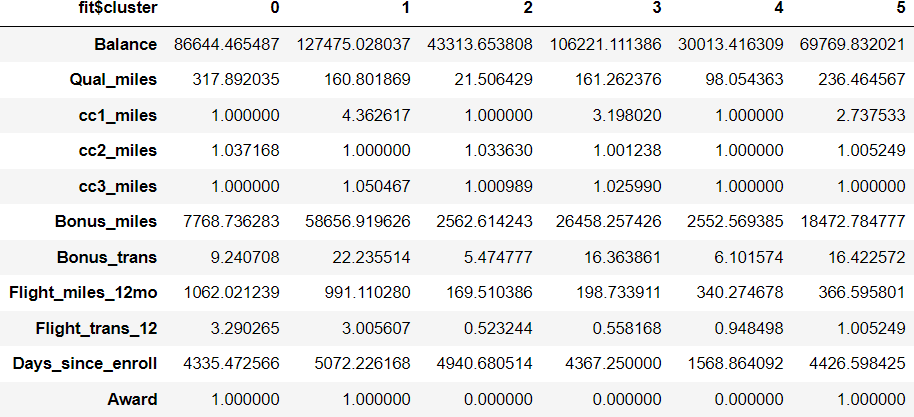
Clusters which are formed and number of observations in each cluster





***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

# Answer f.

Characteristics of cluster formed by K-Means Clustering –

***Pearl->Opal -> Emerald -> Sapphire -> Ruby -> Diamonds***

Cluster 0:

## Fly Ruby

Received Free Flight (Award), High Flight miles, Low Bonus Miles.

Customers in this cluster have received free flight and mostly using our flight services.

Cluster 1:

## Fly Diamond

Received Free Flight (Award), High Flight miles, High Bonus Miles

Customers in this cluster have received free flight and mostly using our both flight and non-flight services.

Cluster 2:

## Fly Opal

Not Received Free Flight (Award), Low Flight miles, Low Bonus Miles

Customer is an Old customer but they are not using any flight or non-flight services.

Cluster 3:

## Fly Emerald

Not Received Free Flight (Award), Low Flight miles, High Bonus Miles

Customers in this Cluster are Old Customers but they are using our Cards non-flight services.

Cluster 4:

## Fly Pearl

New Customer, Not Received Free Flight (Award), High Flight miles, Lowest Bonus Miles

Customers in this cluster are new and are frequently using our flight services.

Cluster 5:

## Fly Sapphire

Received Free Flight (Award), Low Flight miles, Medium Bonus Miles

Customers in this cluster have received free flight and mostly using our non-flight services

#### Comparative Analysis of Cluster formed by both the methods K-Means(e) and Hierarchal(c)

On Analysis of Clustered which are formed by both the both the techniques, clusters which are formed are majorly influenced by the following dimensions:

1. *Bonus Miles*
2. *Flight Miles in last 12 months*
3. *Days since enrol*
4. *Awards*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kmeans Clusters | | | | |
| Bonus\_miles | Flight\_miles\_12mo | Days\_since\_enroll | Awards | Clusters Name |
| low | high | old | 1 | Cluster 0 (Ruby) |
| high | high | old | 1 | Cluster 1 (Diamond) |
| Low | low | old | 0 | Cluster 2 (Opal) |
| high | low | old | 0 | Cluster 3 (Emerald) |
| low | high | New | 0 | Cluster 4 (Pearl) |
| medium | low | old | 1 | Cluster 5 (Sapphire) |
| Hierarchal Clusters | | | | |
| low | medium | old | 1 | Cluster 0 (Ruby) |
| low | low | old | 0 | Cluster 1 (Emerald) |
| high | low | old | 0 | Cluster 2 (Sapphire) |
| high | high | old | 1 | Cluster 3 (Diamond) |
| low | medium | new | 0 | Cluster 4 (Opal) |

Above table represents what is the influencing factor in the formation of the clusters. So, I have tried relating the cluster formed by both the techniques with each other like which cluster from Hierarchal Clustering corresponds to which cluster from K-Means Clustering

|  |  |
| --- | --- |
| Hierarchal Clusters | K-Means Clusters |
| Cluster 0 | Cluster 0 or Cluster 5 data got converged into  this |
| Cluster 1 | Cluster 2 |

|  |  |
| --- | --- |
| Cluster 2 | Cluster 3 |
| Cluster 3 | Cluster 1 also there might be some customers  from Cluster 0 and Cluster 5 of K-means |
| Cluster 4 | Cluster 4 |

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Answer g.

#### Offers to the Cluster which can be provided with respect to the Cluster characteristics are For this answer I am considering the clusters which are created by K-Means Technique.

Cluster 0: Fly Ruby Class

Received Free Flight (Award), High Flight miles, Low Bonus Miles.

Customers in this cluster have received free flight and mostly using our flight services.

### Offers –

* 1. As their Bonus Miles are low, we can offer additional bonus miles to boost their usage.
  2. Provide perks or rewards for their High Flight miles.
  3. Provide customised discounts or class upgrades in flights.
  4. Reward their loyalty by giving offers on their birthdays or anniversaries.

Cluster 1: Fly Diamond Class

Received Free Flight (Award), High Flight miles, High Bonus Miles

Customers in this cluster have received free flight and mostly using our both flight and non-flight services

### Offers –

1. Customised offer for them as they are using both flight and non-flight services.
2. Provide them a personalized recommendations and promotions to reward their usage.
3. Assign them a relationship manager for making sure they can avail all the services hassle free.
4. Provide them with unlimited lounge access to retain their trust and appreciate their loyalty.

Cluster 2: Fly Opal Class

Not Received Free Flight (Award), Low Flight miles, Low Bonus Miles Customer is an old customer but they are not availing any flight or non-flight services.

Offers –

1. Incentivise their non-flight expenditure better for retaining them.
2. Provide them recommendations of offers and provide discounted travel plans to lure them use or services for booking flights as well.
3. Reward their loyalty by providing seasonal or quarterly discounts on the various services.

Cluster 3: Fly Emerald Class

Not Received Free Flight (Award), Low Flight miles, High Bonus Miles

Customers in this Cluster are Old Customers but they are using our Cards non-flight services.

Offers –

1. Incentivise their non-flight expenditure better for retaining them.
2. Give 1% of Customer (like a lottery) of this Cluster a round trip and advertise it on our social media platform with personal story for increasing the usage and motivating others to increase the usage of our flight services.
3. For others provide recommendations for flight services that is inline with their area of interests or preferences.

Cluster 4: Fly Pearl Class

New Customer, Not Received Free Flight (Award), High Flight miles, Lowest Bonus Miles

Customers in this cluster are new and are frequently using our flight services.

Offers –

1. Provide newest Customers with a welcome kit of variety coupons for flight and non-flight services. This will result in better referrals and more customer onboarding.
2. Incentivise their non-flights expenses like promotional offers, discounted services to increase their non-flight expenses.
3. Give them free upgrades on booking flights services for using our services rigorously.
4. Easy onboarding on flights.
5. Assign them a relationship manager for making sure they can avail flight related services hassle free.

Cluster 5: Fly Sapphire Class

Received Free Flight (Award), Low Flight miles, Medium Bonus Miles

Customers in this cluster have received free flight and mostly using our non-flight services

Offers –

1. Consider this cluster’s customer for feedback to understand why they are not preferring our cards for flight and even non-flight services. Draw insights from their feedback and provide them assistance to improve their experience with us.
2. Offer bonus miles or rewards for using services occasionally.
3. Lure them by giving cashbacks and fewer discount coupons.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****