

Observations and Analysis for East-West-Airlines Data

1. Introduction

This document presents the results of K-means and Hierarchical cluster analysis, based on a dataset containing information on 3999 passengers who belong to an airline's frequent flier program. The goal is to try to identify clusters of passengers that have similar characteristics for the purpose of targeting different segments for different types of mileage offers.

2. Clustering analysis:

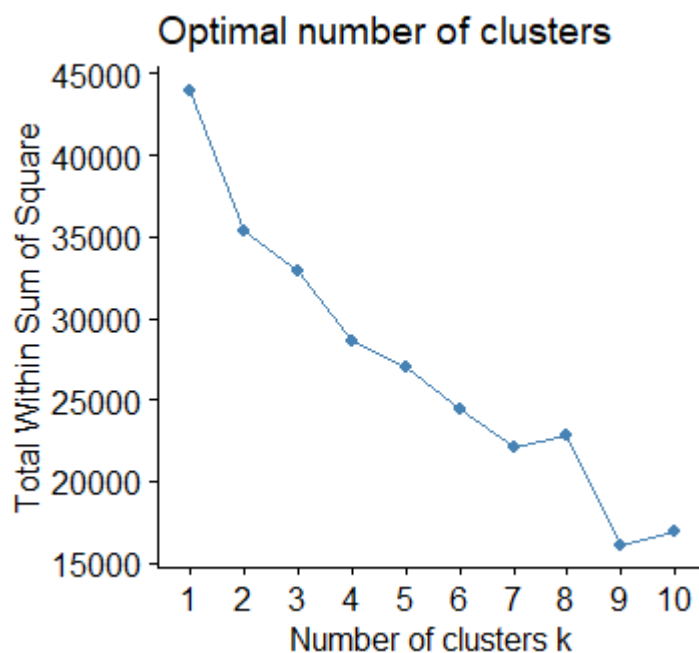
There are various methods to perform clustering analysis. In this case, K-means and Hierarchical method is used.

2.1. K means Clustering:

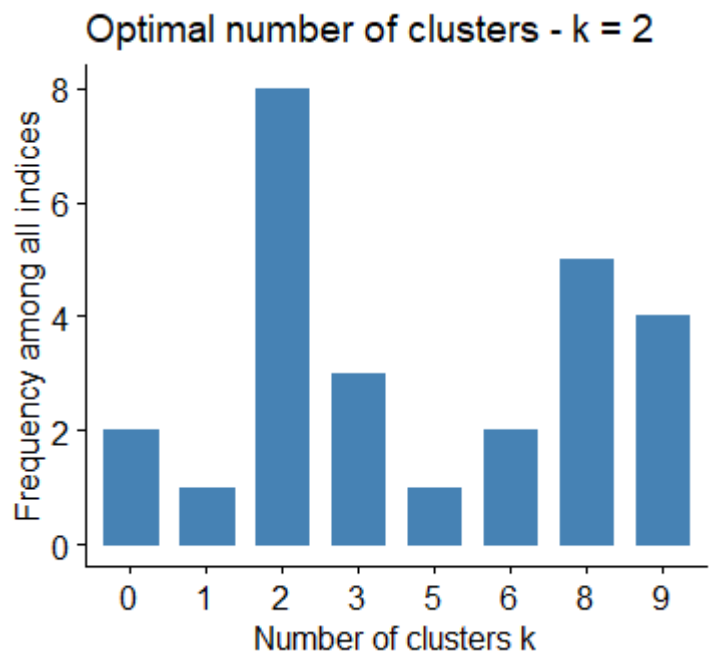
- **Determining number of clusters:**

We need to know number of clusters needed as an input to the K-means algorithm. This can be done using 2 approaches:

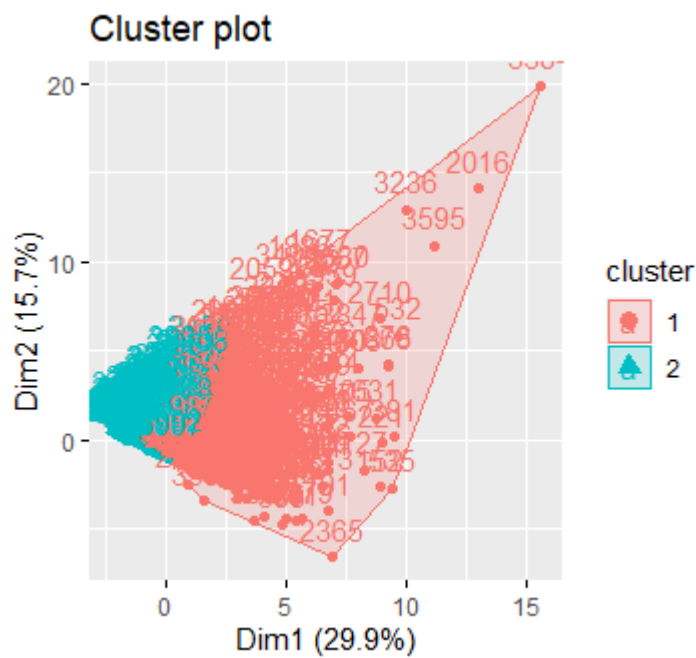
- Scree Plot (elbow method) – The scree plot obtained from the analysis, as shown below, does not provide a conclusive result to determine how many clusters we need.



- NbClust() – This is a comprehensive approach that can be used to determine number of clusters required. On using this method, we obtain a histogram, as shown below, that gives an optimal number of clusters based on tallies between various approaches. In this case, the result was 2, hence 2 clusters would be used as an input for K-means clustering analysis.



- Cluster structure:**
Below is the cluster plot generated. It is not very clean, due to overlaps between the 2 clusters.



- Cluster distribution table:**
Below is the cluster distribution table that shows the number of records, for the clusters 'HighSpenders' and 'LowSpenders'.

Clusters	HighSpenders	LowSpenders
Number of records	1299	2700

- **Summary table of cluster centers (i.e., cluster means):**

The below table shows summary of center values determined of both clusters, which gives a summary of the entire dataset.

	Kmeans distribution										
Cluster	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award.
HighSpenders	0.587921	0.1477122	1.127513	-0.03045	0.1265025	1.0076855	0.9154157	0.363811	0.4048241	0.3985725	0.7028251
LowSpenders	-0.282855	-0.071066	-0.54246	0.01465	-0.060862	-0.4848087	-0.4404167	-0.1750335	-0.1947654	-0.1917576	-0.338137

2.2. Hierarchical clustering:

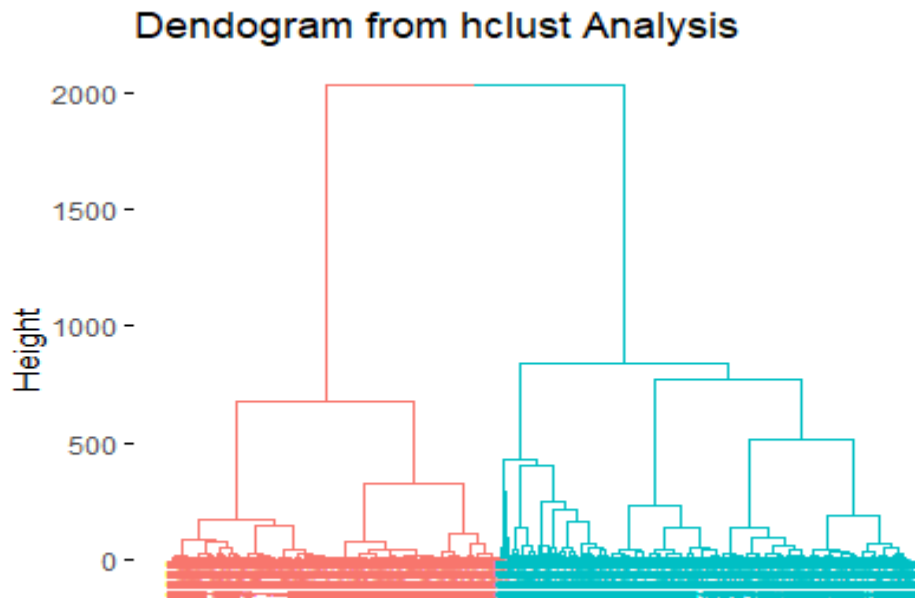
- **Determining number of clusters:**

Hierarchical cluster analysis was performed using hclust() hierarchical clustering algorithm with Euclidian distance and Ward's method.

The number of clusters is 2, as we obtained earlier using NbClust().

- **Dendrogram for hierarchical clustering results:**

The tree was cut using a 2 cluster solution. The resultant dendrogram is as shown below.



- **Cluster distribution table:**

Below is the cluster distribution table that shows the number of records, for the clusters 'HighSpenders' and 'LowSpenders'.

Clusters	HighSpenders	LowSpenders
Number of records	2207	1792

- **Summary table of cluster centers (i.e., cluster means):**

	Hierarchical distribution										
Cluster	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award.
HighSpenders	0.2717898	0.1421304	0.527714	0.079759	0.0509577	0.4527601	0.4401669	0.197203	0.2113687	0.2437211	0.6226314
LowSpenders	-0.334732	-0.1750456	-0.64993	-0.09823	-0.062759	-0.55761245	-0.54210287	-0.24287225	-0.26031843	-0.30016324	-0.7668234

3. Clustering analysis observations and results:

3.1. Summary:

- It is clear from the summary tables of the 2 cluster analysis results, that we can consider 2 groups of passengers namely: “HighSpenders” and “LowSpenders”.
- The HighSpenders spend relatively more on flights or other transactions, as compared to the LowSpenders.
- As a result, the HighSpenders have more Balance, which is the number of miles eligible for award travel
- A peculiar observation can be seen in the feature ‘Number of miles earned with Rewards credit card in the past 12 months’ (cc2_miles) – HighSpenders show a lower value for this feature as compared to LowSpenders – which is converse to the behavior of other features.

3.2. K - means vs Hierarchical clustering results:

- There is a high contrast observed between K-means and Hierarchical clustering results with respect to HighSpenders and LowSpenders.
- For example, consider the feature Balance – The centroid of HighSpenders in K-means distribution is 0.58, whereas for Hierarchical distribution, it is 0.27.
- This can be observed in almost all features, except Qualifying miles where both centroids are almost same – 0.14.

3.3. Conclusion:

Based on the above analysis, following conclusions can be drawn:

- LowSpenders: This cluster shows lesser non-flight transactions (as compared to HighSpenders) as well as flight transactions.
However, the miles earned, are higher with Rewards credit cards.
This cluster can be targeted with more non-flight transaction offers.
Rewards credit card offers can especially be beneficial.
- HighSpenders: This cluster shows more transactions as compared to LowSpenders. Customers falling under this cluster, hence, should be retained.
Thus, this cluster should be targeted with customer-retention offers – such as ‘Loyalty discounts’, based on the history of transactions with the airline.
Other discounts, such as ‘Active Customer’ discounts can be offered to customers who participate in frequent transactions.