

CLUSTER ANALYSIS

```
dt<-read.csv(file.choose())  
head(dt)
```

salary	age group	occasions	preferences
3	2	1	2
4	1	3	4
2	2	2	4
2	2	4	1
2	4	3	3

code	salary	age group	occasions	preferences
1	below 50000	15-25	party	formal
2	above 50000	26-35	date	casual
3	equal to 100000	36-45	festival	trending
4	above 100000	above 46	wedding	innovation

Used hierarchical clustering to find maximum number of clusters can be formed

```
dt1<-dt[c(1:4)]
```

```
NbClust(dt1,distance = 'euclidean',method = 'ward.D')
```

* Among all indices:

* 5 proposed 2 as the best number of clusters

* 2 proposed 3 as the best number of clusters

* 4 proposed 4 as the best number of clusters

* 1 proposed 5 as the best number of clusters

This output shows 2 clusters are best fit for this data

Used k modes clustering to make it into clusters

```
kmodes(dt1,2)
```

K-modes clustering with 2 clusters of sizes 47, 53

Cluster modes:

sno	salary	age.group	occasions	preferences
1	above 100000	36-45	festival	trending
2	above 50000	26-35	date	casual

K modes clustering made 100 responses in to 2 clusters with cluster sizes 47,53 based on age group, salary,occasion,preferenece.