

```
In [8]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from scipy.spatial.distance import cdist
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster.hierarchy import cophenet
import scipy.cluster.hierarchy as sch
from scipy.spatial.distance import pdist
from sklearn.cluster import AgglomerativeClustering
```

```
In [9]: var=pd.read_csv('C:/Users/Gopi/Desktop/machine learning/csv files/cars.csv')
var.head(1)
```

Out[9]:

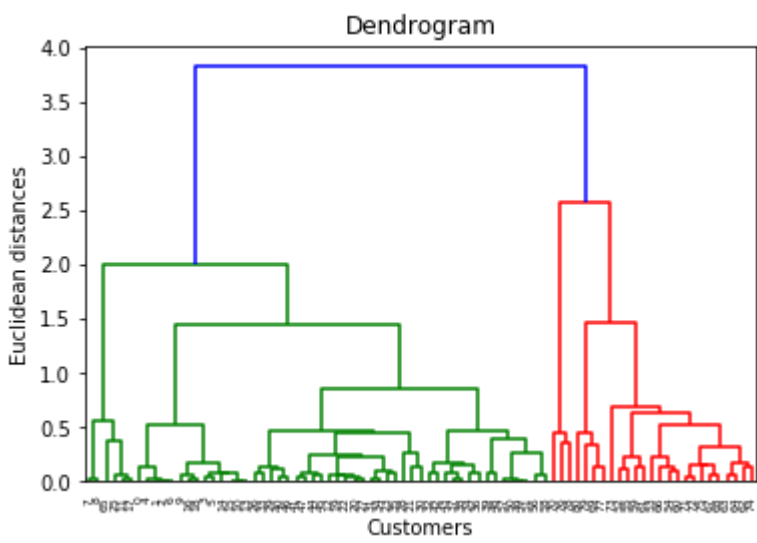
	HP	MPG	VOL	SP	WT
0	49	53.700681	89	104.185353	28.762059

```
In [10]: var1 = (var - var.min()) / (var.max() - var.min())
var1.head(3)
```

Out[10]:

	HP	MPG	VOL	SP	WT
0	0.000000	1.000000	0.354545	0.065975	0.349986
1	0.021978	0.911362	0.381818	0.084193	0.395709
2	0.021978	0.911362	0.381818	0.084193	0.388381

```
In [11]: dendrogram = sch.dendrogram(sch.linkage(var1, method = 'ward'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean distances')
plt.show()
```



```
In [27]: cluster = AgglomerativeClustering(n_clusters=5, affinity='euclidean', linkage='complete')
temp = cluster.fit_predict(var1)
print(temp)
```

```
[0 0 0 0 0 0 0 4 4 0 0 4 0 0 0 0 0 4 0 0 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 1 1 1 1 4 1 1 1 3 2 1 1 1
 1 1 2 3 2 3 3]
```

```
In [28]: var['clusters']=pd.Series(temp)
var.head(3)
```

Out[28]:

	HP	MPG	VOL	SP	WT	clusters
0	49	53.700681	89	104.185353	28.762059	0
1	55	50.013401	92	105.461264	30.466833	0
2	55	50.013401	92	105.461264	30.193597	0

```
In [29]: var['clusters'].value_counts()
```

Out[29]:

```
0    50
1    18
4     6
3     4
2     3
Name: clusters, dtype: int64
```