# nrcm-hierarchical-clustering-1

August 28, 2023

NAME : A.SUMANTH

ROLL NO : 21X05A6704

BRANCH : CSE(DS)

NRCM

## 0.1 project title

Analysis and prediction of "mall_customers.csv" of American mall market called as Phonix Mall,to findout requirements of dendrogram using seipy library with the help of "seipy.cluster.hirearchy ,to ace the no of linkage of the clustering to predict

## 0.2 PROJECT TITLE :The american finance market client as per the rate of gdp of 2011 found as highest number of growth in there business market.

As a data science engineer findout which hirerachy gives upcoming feature

## 0.3 TASk

1. with the help of seipy import the libraries and datasets

2. using the dendogram to find the optimal number of the clusters

3. create the hirearchy model and visualize the cluster with the help of matplot librarie

4.

```
[3]: #Import the numpy, pandas , matplotlib, seaborn libery's
     import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```
[10]: #Assign variable name "dataset" and the input variable as "X" indcludong select␣
      ↪all the row and index columns which you want [colum_index, Column_index].
      dataset=pd.read_csv("Mall_Customers.csv")
      X=dataset.iloc[:,[3,4]].values
```

```
[11]: #import scipy  cluster using attribute "scipy.cluster.hierarchy" as sch alias
      import scipy.cluster.hierarchy as sch
```

```
[13]: #Using the dendrogram to find the optimal number of clusters
      # Assign a variable as dendograph and declers the "sch.dendrogram(sch.
        ↪linkage(X, method = 'ward'))"
      dendrogram = sch.dendrogram(sch.linkage(X,method='ward'))
```
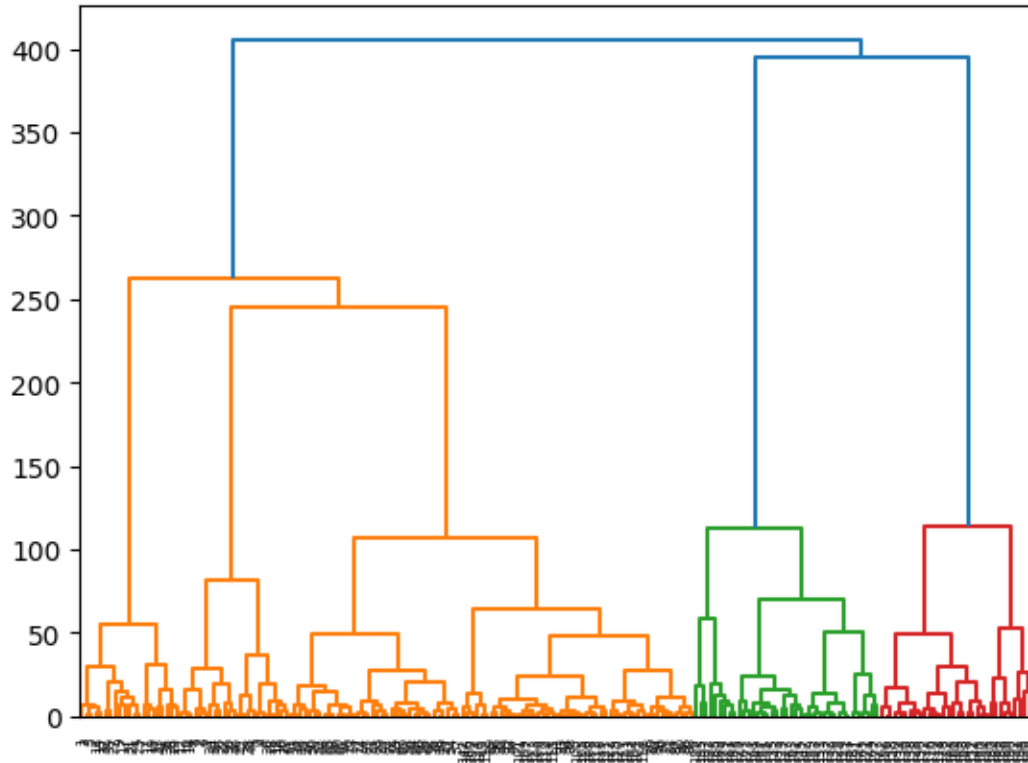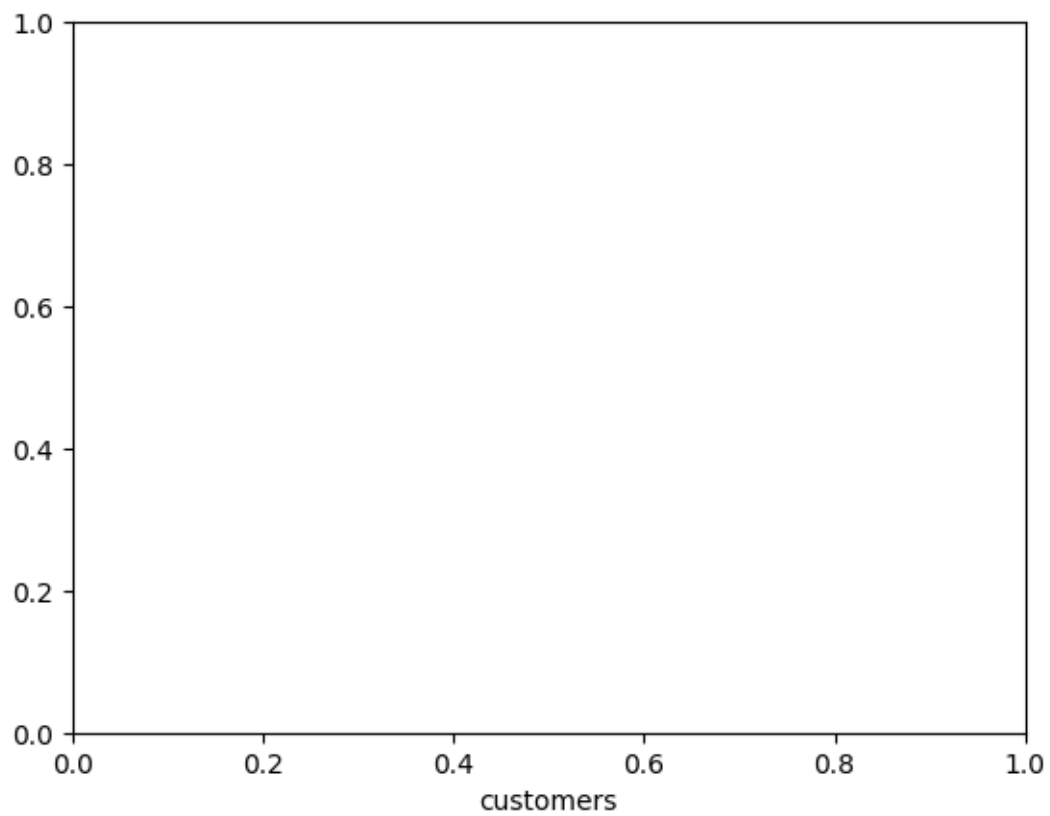


```
[14]: #Assign the title as "Dendograms""
      plt.title("Dendograms")
```
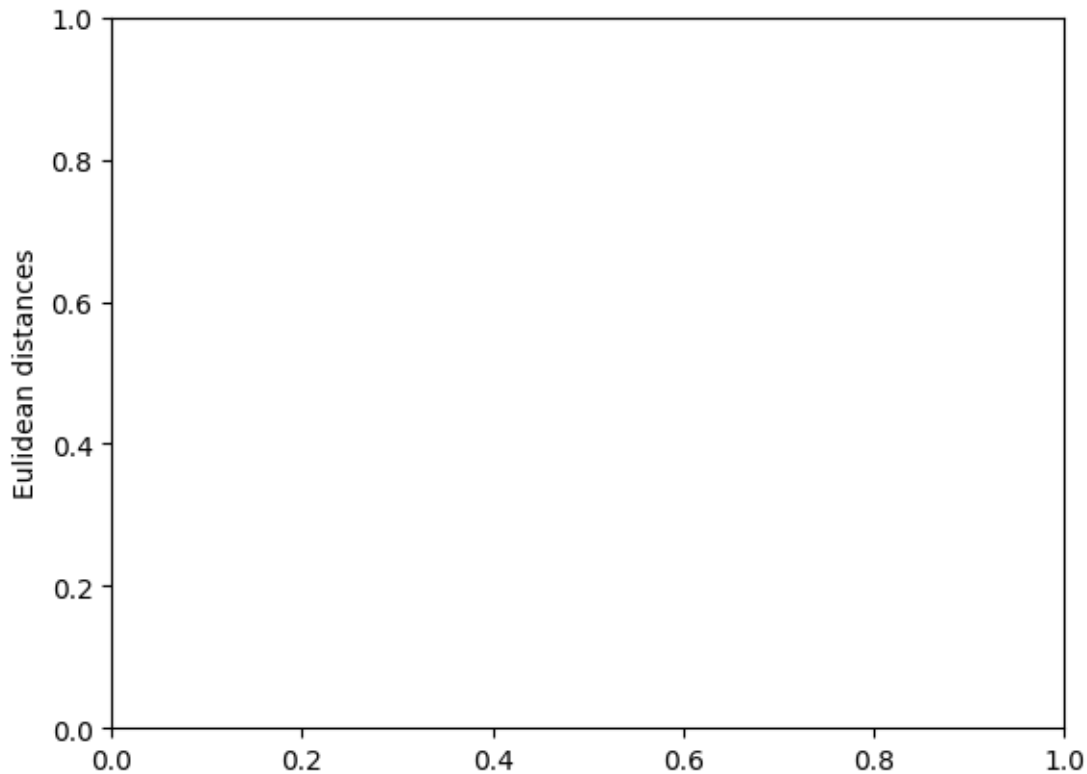
```
[14]: Text(0.5, 1.0, 'Dendograms')
```

Dendograms
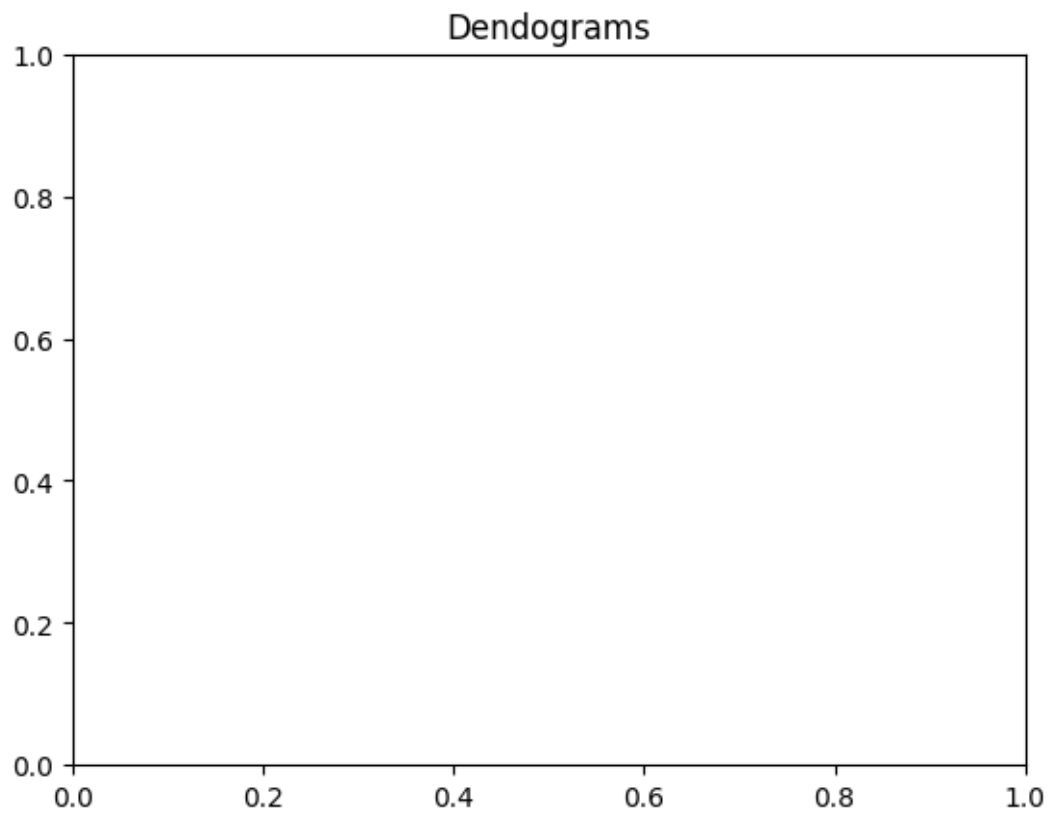
[15]: #Label X axis as "Customers"
      plt.xlabel("customers")

[15]: Text(0.5, 0, 'customers')

customers

[16]: ```python
#Label Y axis as 'Euclidean distances'
plt.ylabel("Eulidean distances")
```

[16]: Text(0, 0.5, 'Eulidean distances')

[18]: ```python
# from "sklearn.cluster" attribute import "AgglomerativeClustering" default
 argument.
from sklearn.cluster import AgglomerativeClustering
```

[19]: ```python
#Create a cluster for  five or nth cluster which you want.
hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage =
 'ward')
y_hc = hc.fit_predict(X)
```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_agglomerative.py:983:
FutureWarning: Attribute `affinity` was deprecated in version 1.2 and will be
removed in 1.4. Use `metric` instead
  warnings.warn(

[23]: ```python
#Plot the scatter plot for scatter  visualization.
plt.scatter(X[y_hc ==0,0 ], X[y_hc == 0,1 ], s = 100, c = 'red', label =
 'Cluster 1')
plt.scatter(X[y_hc == 1,0], X[y_hc ==1,1], s = 100, c = 'blue', label =
 'Cluster 2')
plt.scatter(X[y_hc == 2,0], X[y_hc ==2,0 ], s = 100, c = 'green', label =
 'Cluster 3')
```

```
plt.scatter(X[y_hc == 3,0], X[y_hc ==3,0 ], s = 100, c = 'cyan', label =␣
  ↪'Cluster 4')
plt.scatter(X[y_hc ==4,0 ], X[y_hc == 4,0], s = 100, c = 'magenta', label =␣
  ↪'Cluster 5')
```

[23]: <matplotlib.collections.PathCollection at 0x7927c4c34b80>
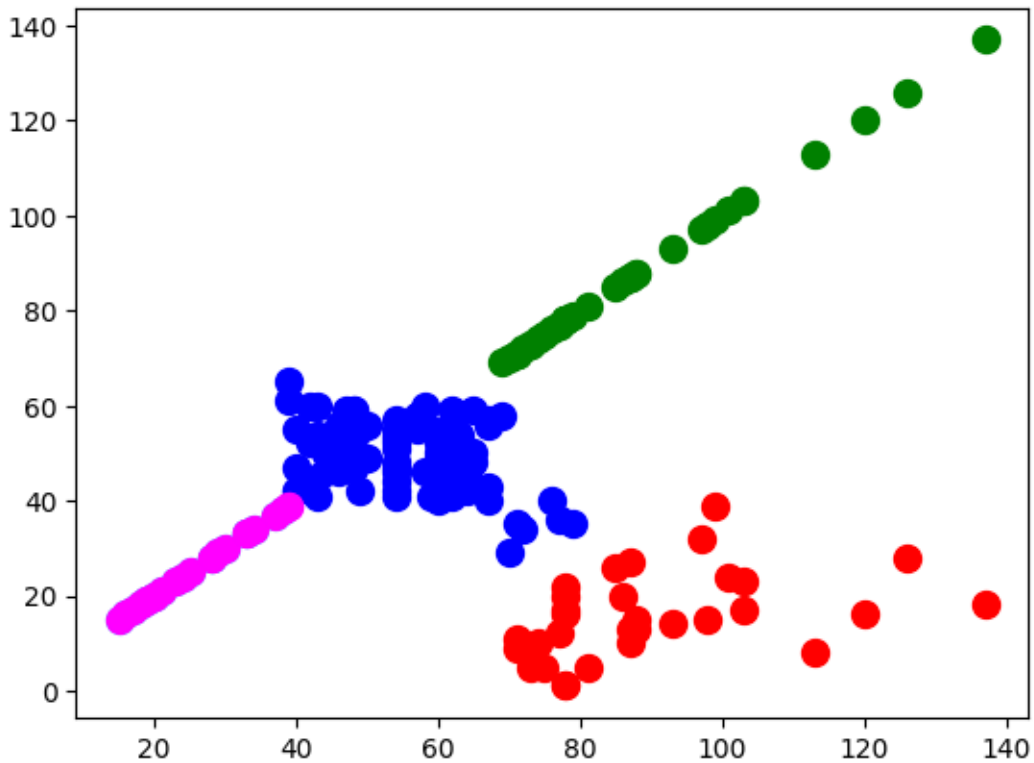


[25]:
```
plt.title('Clusters of customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

WARNING:matplotlib.legend:No artists with labels found to put in legend.  Note
that artists whose label start with an underscore are ignored when legend() is
called with no argument.

Clusters of customers
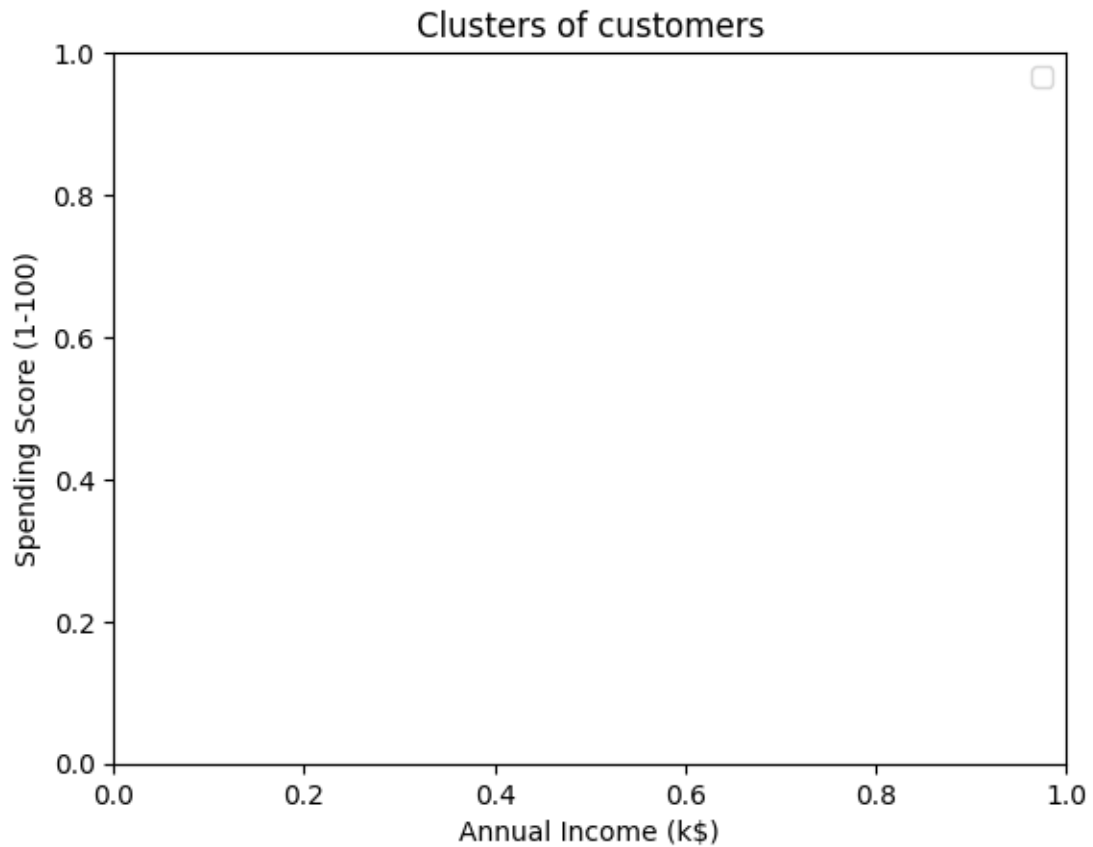
## 0.4 CONCLUSION

According to the model building as a enginner my prediction is cluster number 3 heighest number linkage.

[ ]: