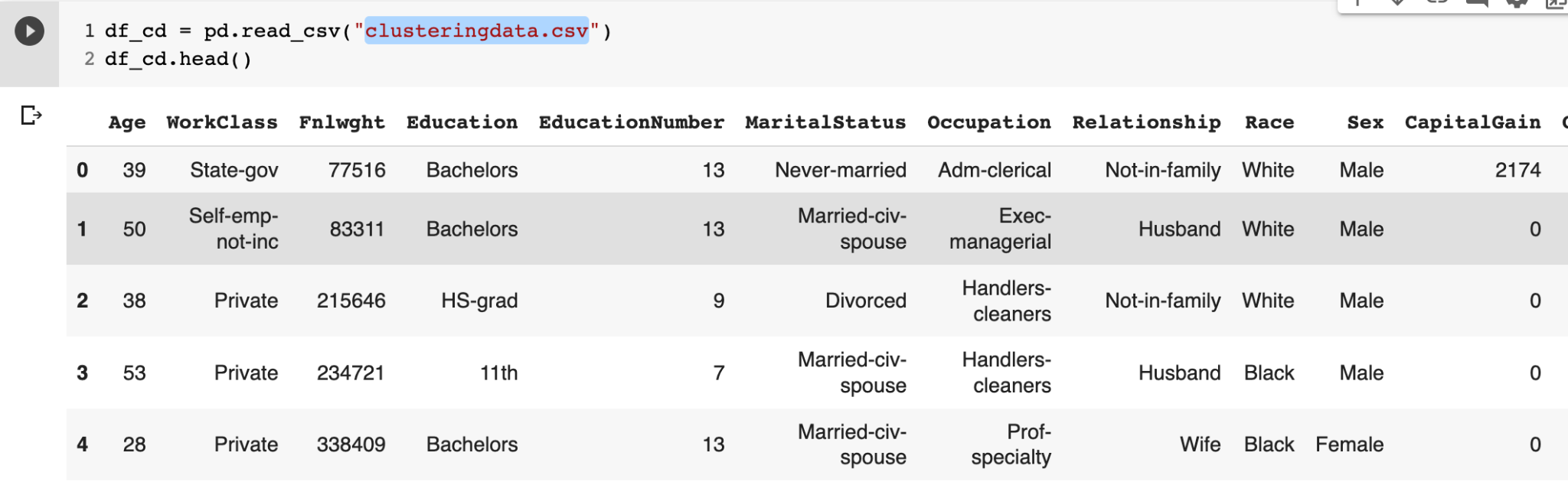
Report - Assignment 4 - Clustering

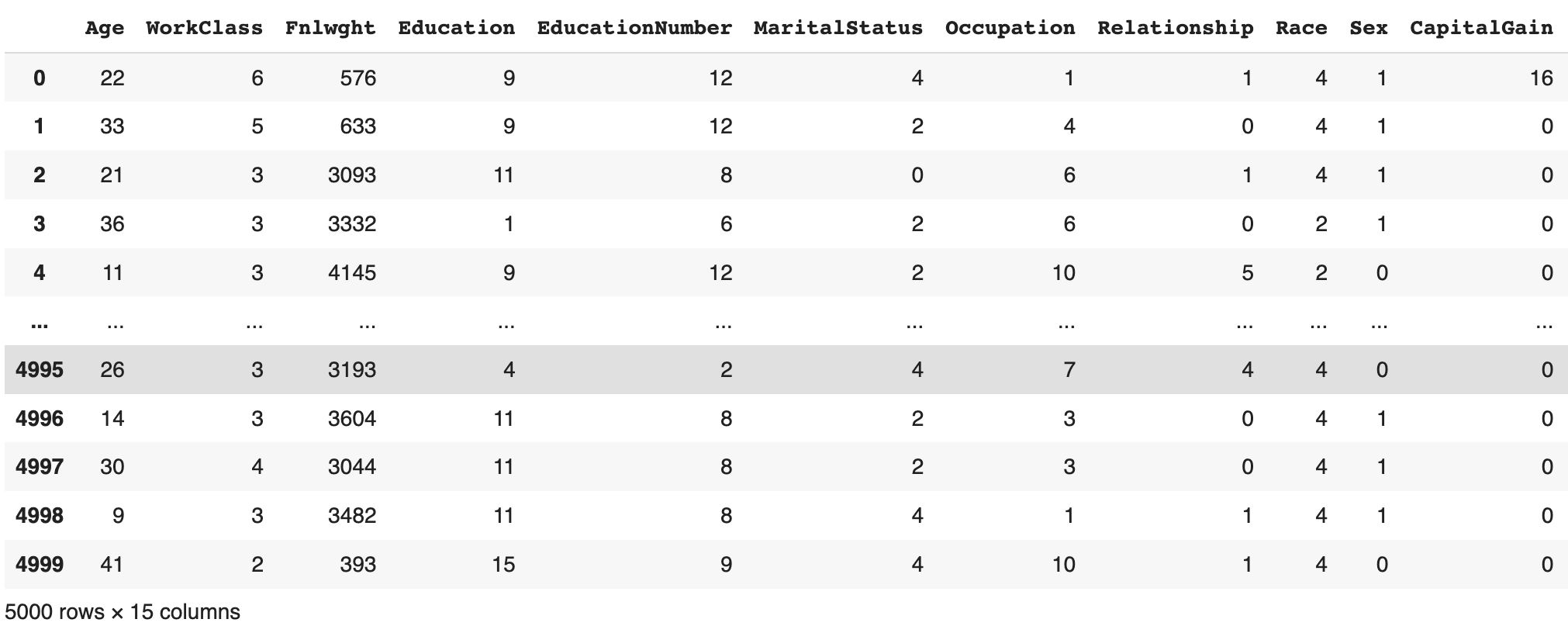
(Using “**clusteringdata.csv**” for this analysis)

# Task 1: K-Means Clustering

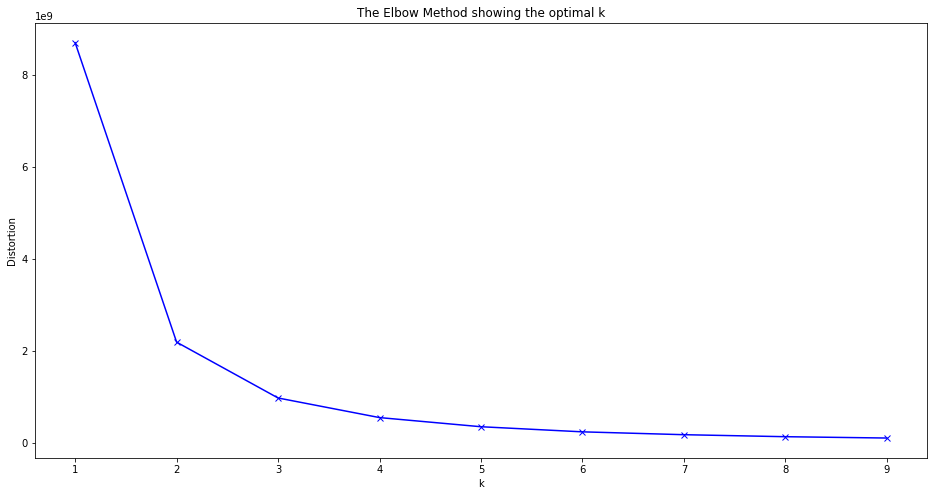
### Loading Dataset:



### Label Encoding the string columns to apply k-mean clustering:

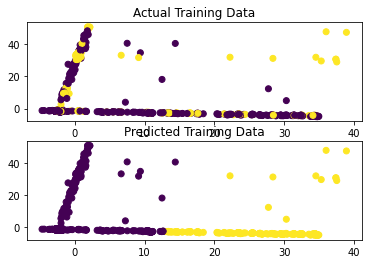


### Elbow Method to get optimal value of K:

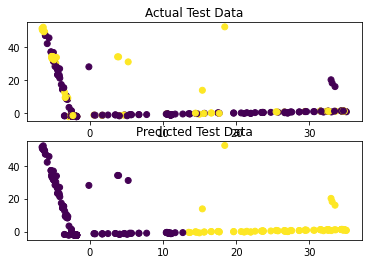


From the figure, we get the elbow at k=2, So the optimal value of k is 2 in k-means clustering.

### Visualizing the k-means clustering for training data:



### Visualizing the k-means clustering for Test data:



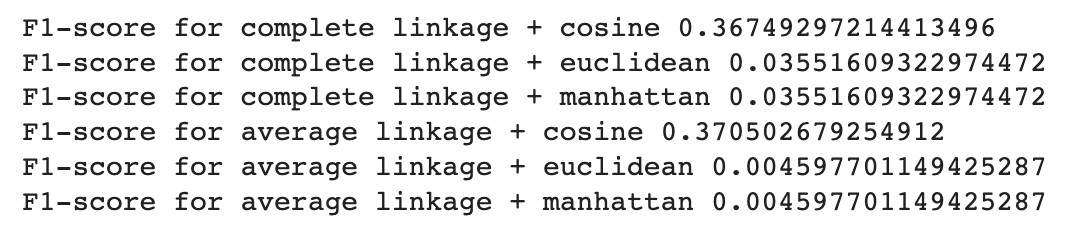
### Confusion Matrix for Test data prediction:

**[[1078 69]**

**[ 322 31]]**

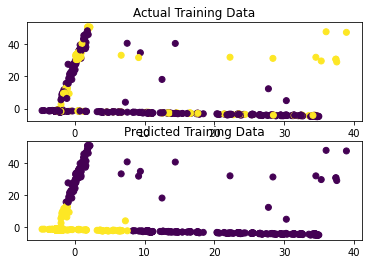
# Task 2: Hierarchical Agglomerative Clustering

### Finding the best Hierarchical Agglomerative Clustering Model



> "average linkage + cosine" is best among all.

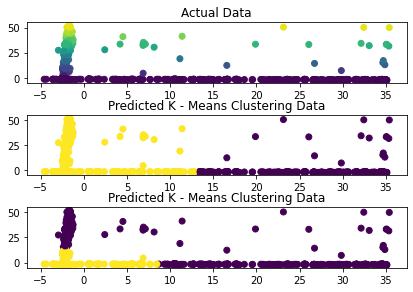
### Visualization for Hierarchical Agglomerative Clustering



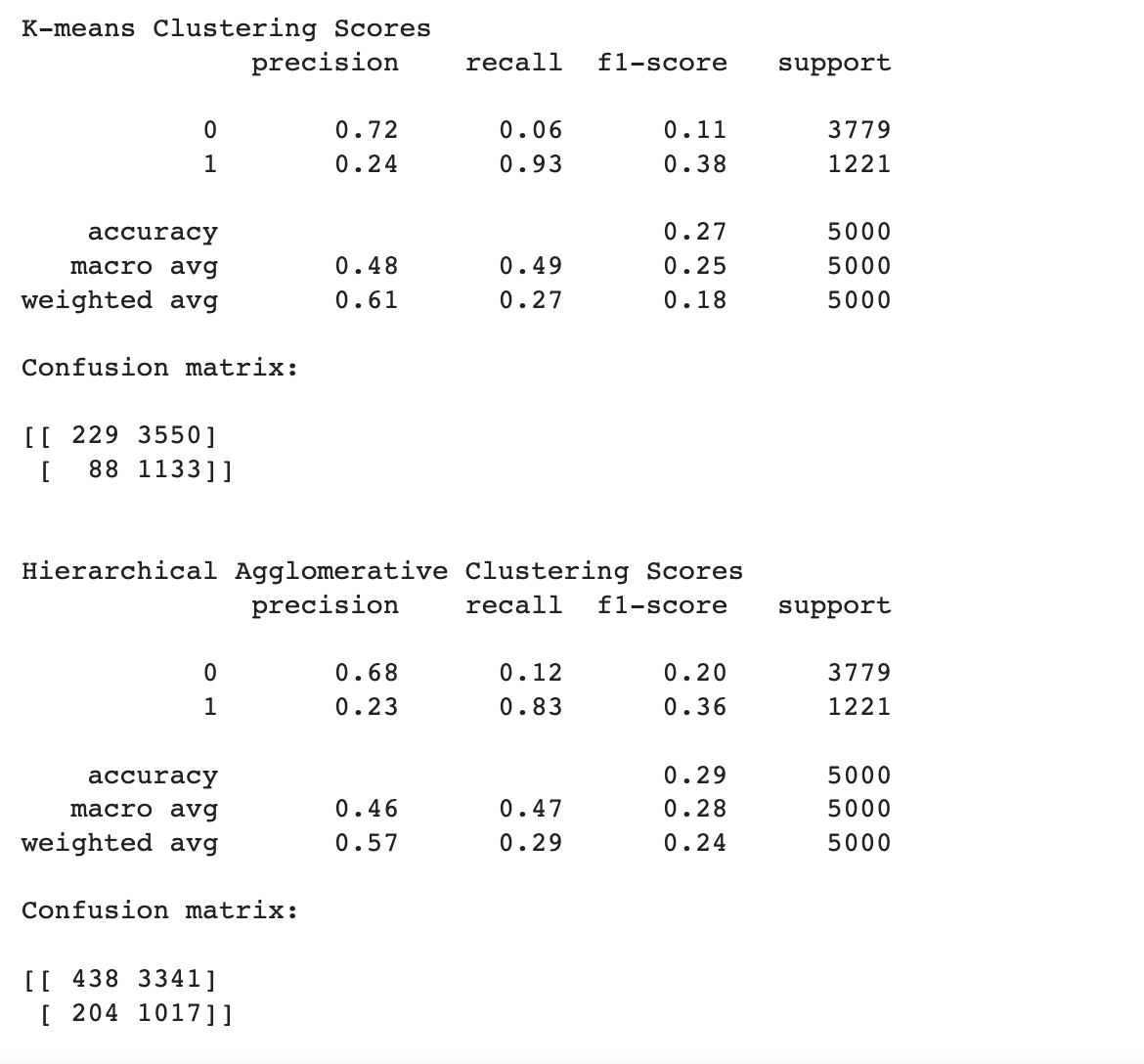
# Task 3: Compare K-Means Clustering and Hierarchical Agglomerative Clustering

### Visualize Clusters

### 



### Comparing precision, recall, and F1-score for both model



#### Reasoning:

K-mean Clustering performance is much better than heirarchical Agglomerative Clustering.

### 

### 