import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from matplotlib.colors import ListedColormap

from sklearn.cluster import KMeans

# importing our csv dataset

mydata=pd.read\_csv('DM\_PR8\_Full.csv')

X=mydata.iloc[:,[5,9]].values

# Find out the best number of clusters

Array=[] # to store sum of squares within the groups

for i in range(1,14):

kmeans=KMeans(n\_clusters=i,init='k-means++',random\_state=0)

kmeans.fit(X)

Array.append(kmeans.inertia\_) # inertia --> Sum of squared distances of samples to their closest cluster center

plt.plot(range(1,14),Array)

plt.title('Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('Sum of Squares within groups')

plt.show()

# K-Means clustering algorithm on Rider age and Gender using Uber cabs

kmeans=KMeans(n\_clusters=4,init='k-means++',random\_state=0)

# fit function will give output of kmeans but fit\_predict will give the cluster index for each sample

Y=kmeans.fit\_predict(X)

plt.scatter(X[Y == 0,0], X[Y == 0,1],s=25,c='red',label='cluster 1') #s --> zoom level

plt.scatter(X[Y == 1,0], X[Y == 1,1],s=25,c='blue',label='cluster 2')

plt.scatter(X[Y == 2,0], X[Y == 2,1],s=25,c='magenta',label='cluster 3')

plt.scatter(X[Y == 3,0], X[Y == 3,1],s=25,c='cyan',label='cluster 4')

plt.scatter(kmeans.cluster\_centers\_[:,0],kmeans.cluster\_centers\_[:,1],s=25,c='yellow',label='Centroid')

plt.title('K-Means Clustering')

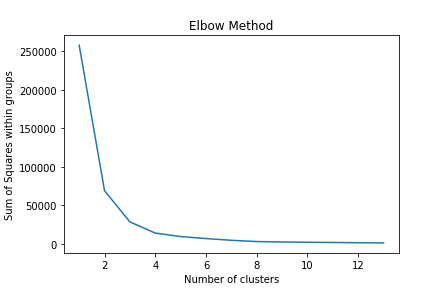
plt.xlabel('Gender')

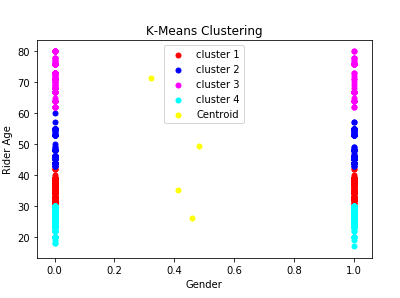
plt.ylabel('Rider Age')

plt.legend()

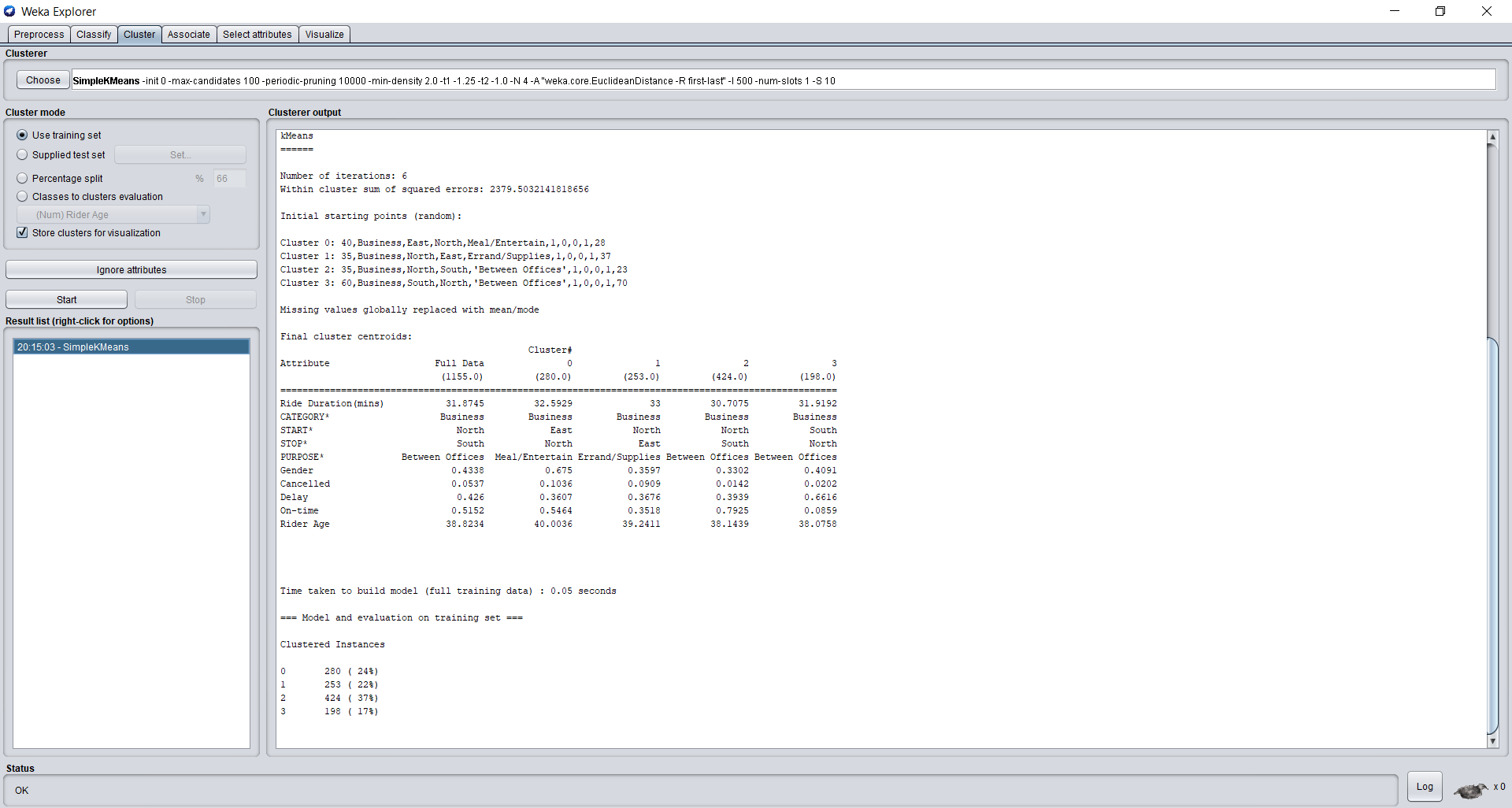
plt.show()

**Output:**

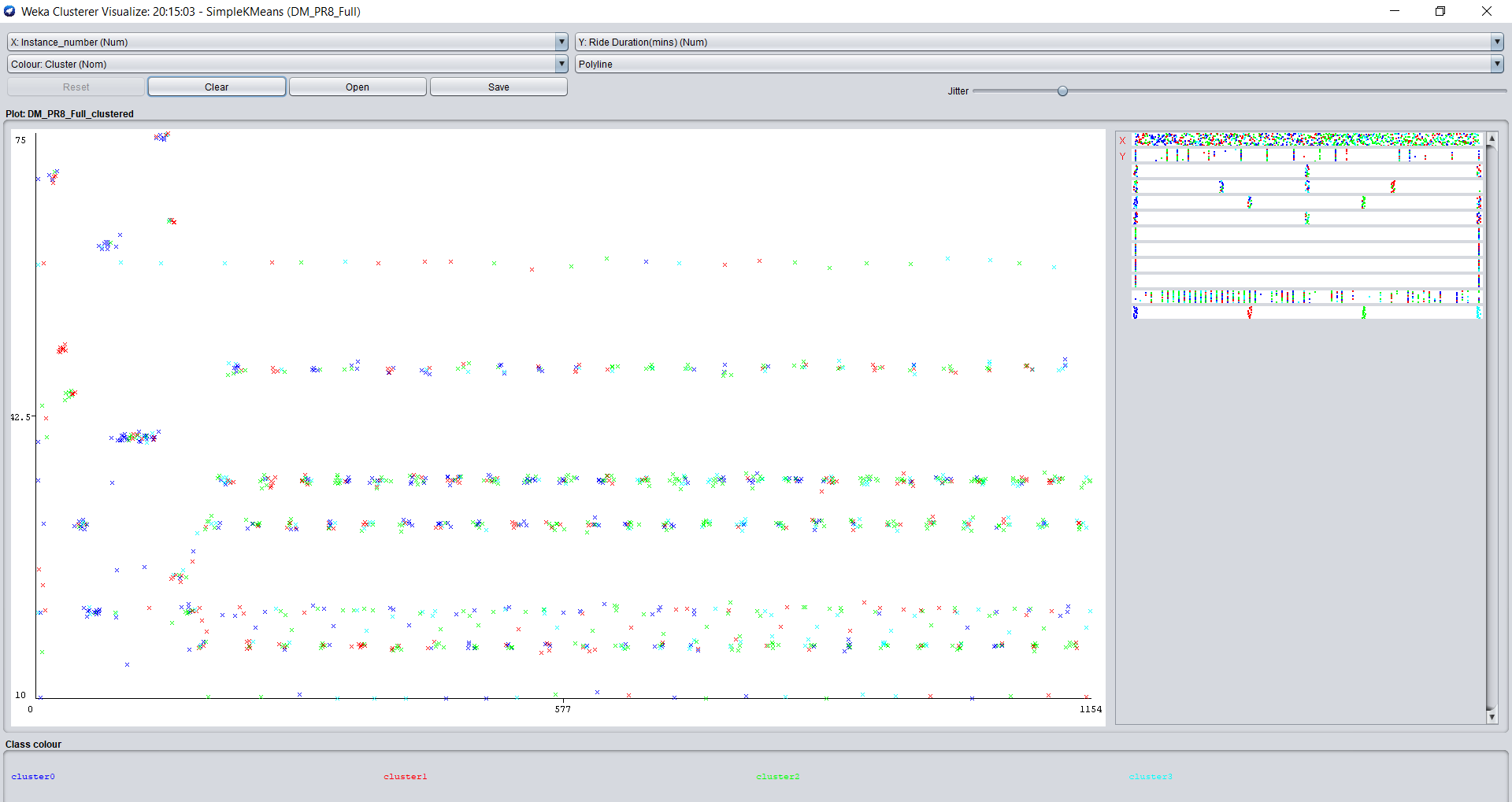




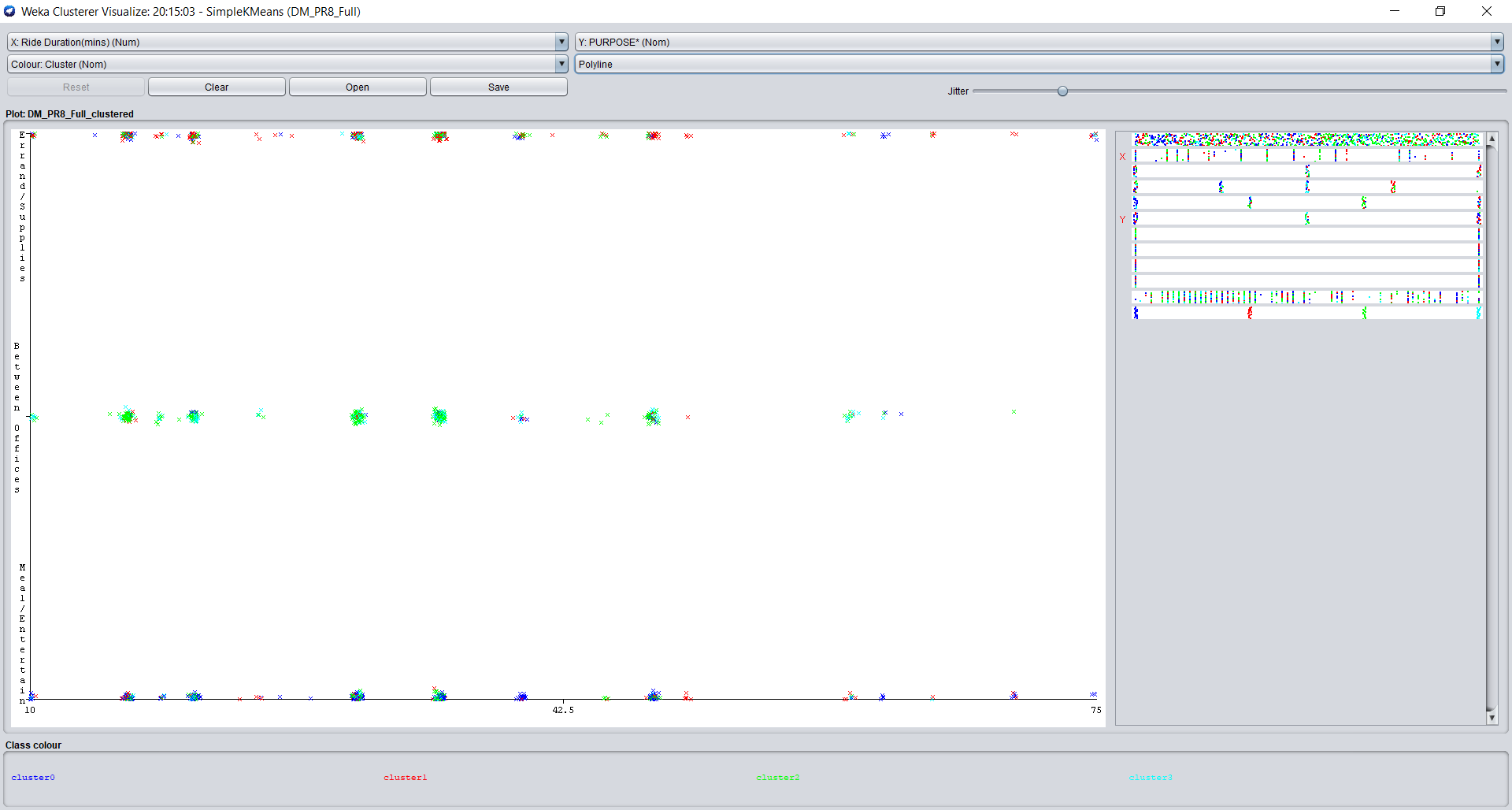
**By Weka:**



**Visualization of clusters in K means clustering Algo with the help of Weka**



**Ride duration vs Purpose graph with the help of clusters:**



**Category vs Purpose:**

