K-Means Customer Segmentation Algorithm

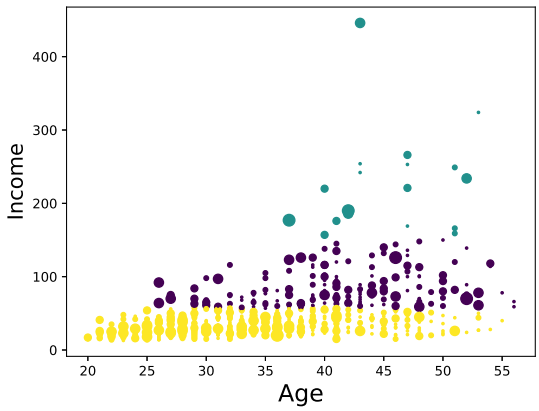
```
import random
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.cluster import KMeans
from sklearn.datasets.samples_generator import make_blobs

## Load data ##

cdf=pd.read_csv('D:\Python\edx\Machine Learning\Clustering\Cust_Segmentation.csv')
with open('Cust_Kmeans.txt','a') as f:
    print(cdf.head(),file=f)

# Preprocessing # - Address column is categorical and is not needed for kmeans algorithm
df=cdf.drop('Address',axis=1)
with open('Cust_Kmeans.txt','a') as f:
    print(df.head(),file=f)

# Normalizing #

from sklearn.preprocessing import StandardScaler
X=df.values[:,1:]
X=np.nan_to_num(X)
d_set=StandardScaler().fit_transform(X)
with open('Cust_Kmeans.txt','a') as f:
    print(d_set,file=f)

#Modeling#
k_means=KMeans(init='k-means++',n_clusters=3,n_init=12)
k_means.fit(X)
k_means_labels=k_means.labels_ #labels for each point in the model using KMeans' .labels
k_means_cluster_centers = k_means.cluster_centers_ #coordinates of the cluster centers us
with open('Cust_Kmeans.txt','a') as f:
    print(k_means_cluster_centers,file=f)
    print(k_means_labels,file=f)

## Insights ##
df['Clus_km']=k_means_labels #Assigning labels to each row in DF
with open('Cust_Kmeans.txt','a') as f:
    print(df.head(),file=f)
```

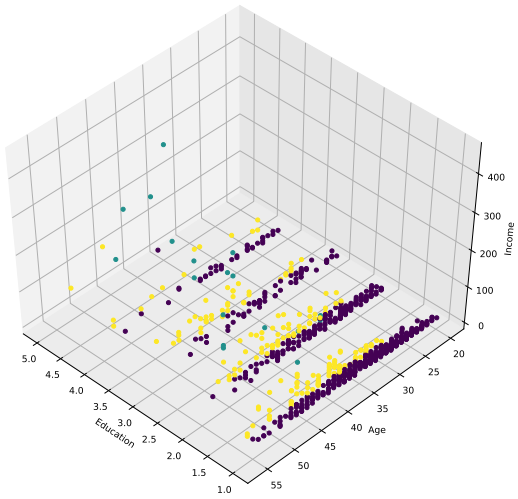
```
df.groupby('Clus_km').mean() # centroid values by avg features in each cluster

# Distribution of customers based on their age and income #
area=np.pi*(X[:,1])**2
plt.scatter(X[:,0],X[:,3],s=area,c=k_means_labels.astype(np.float),alpha=0.5)
plt.xlabel('Age',fontsize=18)
plt.ylabel('Income',fontsize=16)

# 3D plot
from mpl_toolkits.mplot3d import Axes3D
fig=plt.figure(1,figsize=(8,6))
plt.clf()
ax=Axes3D(fig,rect=[0,0,0.95,1],elev=48,azim=134)
plt.cla()
ax.set_xlabel('Education')
ax.set_ylabel('Age')
ax.set_zlabel('Income')
ax.scatter(X[:,1],X[:,0],X[:,3],c=k_means_labels.astype(np.float))

#Display plot
plt.show()
```

Solution:

```
   Customer Id  Age  Edu  Years Employed  Income  Card Debt  Other Debt  Defaulted Addres
DebtIncomeRatio
0            1   41    2               6      19      0.124       1.073        0.0
NBA001               6.3
1            2   47    1              26     100      4.582       8.218        0.0
NBA021              12.8
2            3   33    2              10      57      6.111       5.802        1.0
NBA013              20.9
3            4   29    2               4      19      0.681       0.516        0.0
NBA009               6.3
4            5   47    1              31     253      9.308       8.908        0.0
NBA008               7.2
   Customer Id  Age  Edu  Years Employed  Income  Card Debt  Other Debt  Defaulted
DebtIncomeRatio
0            1   41    2               6      19      0.124       1.073        0.0
6.3
1            2   47    1              26     100      4.582       8.218        0.0
12.8
2            3   33    2              10      57      6.111       5.802        1.0
20.9
3            4   29    2               4      19      0.681       0.516        0.0
```

```
6.3
4             5   47    1               31     253      9.308       8.908         0.0
7.2
[[ 0.74291541  0.31212243 -0.37878978 ... -0.59048916 -0.52379654
  -0.57652509]
 [ 1.48949049 -0.76634938  2.5737211  ...  1.51296181 -0.52379654
   0.39138677]
 [-0.25251804  0.31212243  0.2117124  ...  0.80170393  1.90913822
   1.59755385]
 ...
 [-1.24795149  2.46906604 -1.26454304 ...  0.03863257  1.90913822
   3.45892281]
 [-0.37694723 -0.76634938  0.50696349 ... -0.70147601 -0.52379654
  -1.08281745]
 [ 2.1116364  -0.76634938  1.09746566 ...  0.16463355 -0.52379654
  -0.2340332 ]]
[[4.13333333e+01 1.95628415e+00 1.52568306e+01 8.39289617e+01
  3.10363934e+00 5.76527869e+00 1.36612022e-01 1.07245902e+01]
 [3.29645609e+01 1.61479199e+00 6.37442219e+00 3.11648690e+01
  1.03254083e+00 2.10413251e+00 2.37288136e-01 1.00947612e+01]
 [4.53888889e+01 2.66666667e+00 1.95555556e+01 2.27166667e+02
  5.67844444e+00 1.09071667e+01 2.22222222e-01 7.32222222e+00]]
[1 0 1 1 2 0 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 1 0 1 1 1 1 1 1
 1 1 0 1 0 1 2 1 0 1 1 1 0 0 1 1 0 0 1 1 1 0 1 0 1 0 0 1 1 0 1 1 1 0 0 0 1
 1 1 1 1 0 1 0 0 2 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1
 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 0 1 0 1
 1 1 1 1 1 1 1 0 1 0 0 1 0 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 0 1
 1 1 1 1 0 1 1 0 1 0 1 1 0 2 1 0 1 1 1 1 1 1 2 0 1 1 1 1 0 1 1 0 0 1 0 1 0
 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 2 0 1 1 1 1 1 1 1 0 1 1 1 1
 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 1 0 1 0 0 1 1 1 1 1 1
 1 1 1 0 0 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 0 1 0 0 1
 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 1 2 1 1 1 0 1 0 0 0 1
 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1
 1 0 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 2
 1 1 1 1 1 1 0 1 1 1 2 1 1 1 1 0 1 2 1 1 1 1 0 1 0 0 0 1 1 0 0 1 1 1 1 1 1
 1 0 1 1 1 1 0 1 1 1 0 1 0 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1
 1 0 0 1 1 1 1 1 1 1 1 1 1 1 2 0 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 0 1 1 2 1 2 1
 1 2 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 2 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 0
 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0
 0 1 1 0 1 0 1 1 0 1 0 1 1 2 1 0 1 0 1 1 1 1 1 1 0 0 1 1 1 1 0 1 1 1 0 0 1 1
 0 1 1 1 0 1 2 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1
 1 1 0 1 1 0 1 0 1 0 0 1 1 1 1 0 1 0 1 1 1 1 1 1 0 1 1 1 1 1 0 0 1 1 0 0 1 1 1 1
 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 0 1 0 0 1 1 0 1 1 1 1 1 1 0 0
 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 2 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0]
   Customer Id  Age  Edu  Years Employed  Income  Card Debt  Other Debt  Defaulted
```





DebtIncomeRatio	Clus_km							
0	1	41	2	6	19	0.124	1.073	0.0
6.3	1							
1	2	47	1	26	100	4.582	8.218	0.0
12.8	0							
2	3	33	2	10	57	6.111	5.802	1.0
20.9	1							
3	4	29	2	4	19	0.681	0.516	0.0
6.3	1							
4	5	47	1	31	253	9.308	8.908	0.0
7.2	2							