Code:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv('../input/customer-segmentation-tutorial-in-python/Mall_C
ustomers.csv')
data.head()
data.shape
data.describe()
data.isnull().sum()
plt.figure(figsize=(15,6))
n = 0
for x in ['Age','Annual Income (k$)','Spending Score (1-100)']:
   n+=1
   plt.subplot(1,3,n)
   plt.subplots_adjust(hspace = 0.5,wspace = 0.5)
   sns.distplot(data[x],bins = 20)
plt.figure(figsize=(15,8))
sns.countplot(data=data,x='Age')
sns.countplot(data = data,y ='Gender')
plt.figure(figsize=(8,8))
data['Gender'].value_counts().plot(kind='pie',autopct='%.2f%%')
plt.show()
plt.figure(figsize = (20,8))
sns.heatmap(data.corr(), annot = True)
plt.show()
Age_18_25 = data.Age[(data.Age >= 18)& (data.Age<=25)]
Age_26_35 = data.Age[(data.Age >= 26)& (data.Age<=35)]
Age_36_45 = data.Age[(data.Age >= 36)& (data.Age<=45)]
Age_46_55 = data.Age[(data.Age >= 46)& (data.Age<=55)]
```

```
Age_above_55 = data.Age[data.Age >= 56]
len(Age_18_25.values)
agex = ['Age_18_25','Age_26_35','Age_36_45','Age_46_55','Age_above_55']
agey =
[len(Age_18_25.values),len(Age_26_35.values),len(Age_36_45.values),len(Age_
46_55.values),len(Age_above_55.values)]

plt.figure(figsize=(15,6))
sns.barplot(x = agex,y = agey , palette='mako')
plt.title = (" Range of age ")
plt.xlabl = (" Range of age ")
plt.ylabel = ('No of Customers')
plt.show()
Annual_income_30 = data['Annual Income (k$)'][(data['Annual Income
(k$)']>=0)&(data['Annual Income (k$)']<=30)]
Annual_income_60 = data['Annual Income (k$)'][(data['Annual Income
(k$)']>=31)&(data['Annual Income (k$)']<=60)]
Annual_income_90 = data['Annual Income (k$)'][(data['Annual Income
(k$)']>=61)&(data['Annual Income (k$)']<=90)]
Annual_income_120 = data['Annual Income (k$)'][(data['Annual Income
(k$)']>=91)&(data['Annual Income (k$)']<=120)]
Annual_income_150 = data['Annual Income (k$)'][(data['Annual Income
(k$)']>=121)&(data['Annual Income (k$)']<=150)]
AiX = ['0-30','30-60','60-90','90-120','120-150']
AiY
=[len(Annual_income_30.values),len(Annual_income_60.values),len(Annual_in
come_90.values),len(Annual_income_120.values),len(Annual_income_150.valu
es)]
plt.figure(figsize = (14,6))
sns.barplot(x = AiX,y = AiY,palette = 'Dark2_r')
X1 = data.loc[:,['Age','Annual Income (k$)']].values
from sklearn.cluster import KMeans
acc = []
for k in range(1,10):
```

```
  kmeans = KMeans(n_clusters = k)
  kmeans.fit(X1)
  acc.append(kmeans.inertia_)
plt.figure(figsize = (12,6))
plt.plot(range(1,10),acc,linewidth = 3)
plt.xlabel("no of clusters")

plt.show()
kmean = KMeans(n_clusters = 4)
labels = kmean.fit_predict(X1)
print(labels)
plt.figure(figsize = (12,8))
plt.scatter(X1[:,0],X1[:,1],c = kmean.labels_,cmap = 'rainbow')
plt.scatter(kmean.cluster_centers_[:,0],kmean.cluster_centers_[:,1],color =
'black')
plt.show()
X2 = data.loc[:,['Annual Income (k$)','Spending Score (1-100)']].values
acc = []
for k in range(1,10):
  kmeans = KMeans(n_clusters = k)
  kmeans.fit(X2)
  acc.append(kmeans.inertia_)
plt.figure(figsize = (12,6))
plt.plot(range(1,10),acc,linewidth = 3)
plt.xlabel("no of clusters")
plt.show()
kmean2 = KMeans(n_clusters = 5)
labels = kmean2.fit_predict(X2)
print(labels)
plt.figure(figsize = (14,6))
plt.scatter(X2[:,0],X2[:,1],c = kmean2.labels_,cmap = 'Dark2')
plt.scatter(kmean2.cluster_centers_[:,0],kmean2.cluster_centers_[:,1],color =
'black')
plt.xlabel("Annual Income")
```

```
plt.show()
X3 = data.loc[:,['Age','Annual Income (k$)','Spending Score (1-100)']].values
acc = []
for k in range(1,10):
  kmeans = KMeans(n_clusters = k)
  kmeans.fit(X3)
  acc.append(kmeans.inertia_)
plt.figure(figsize = (12,6))
plt.plot(range(1,10),acc,linewidth = 3)
plt.xlabel("no of clusters")
plt.show()
kmean3 = KMeans(n_clusters = 5)
clusters = kmean3.fit_predict(X2)
data['labels'] = clusters
print(labels)
from mpl_toolkits.mplot3d import Axes3D
fig = plt.figure(figsize=(20,10))
ax = fig.add_subplot(111,projection = '3d')
ax.scatter(data.Age[data.labels==0],data['Annual Income (k$)'][data.labels ==
0],data['Spending Score (1-100)'][data.labels == 0])
ax.scatter(data.Age[data.labels==1],data['Annual Income (k$)'][data.labels ==
1],data['Spending Score (1-100)'][data.labels == 1])
ax.scatter(data.Age[data.labels==2],data['Annual Income (k$)'][data.labels ==
2],data['Spending Score (1-100)'][data.labels == 2])
ax.scatter(data.Age[data.labels==3],data['Annual Income (k$)'][data.labels==
3],data['Spending Score (1-100)'][data.labels == 3])
ax.scatter(data.Age[data.labels==4],data['Annual Income (k$)'][data.labels ==
4],data['Spending Score (1-100)'][data.labels == 4])
ax.view_init(30,185)
plt.show()
```