

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text RAM Disk Editing

2s ✓

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import plotly as py  
import plotly.graph\_objs as go  
from sklearn.cluster import KMeans  
import warnings  
warnings.filterwarnings("ignore")

0s ✓ [2] #We read the csv and print the first 5 rows  
df = pd.read\_csv("/content/Mall\_Customers.csv")  
df.head()

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

0s ✓ [3] #Checking the size of our data  
df.shape

0s completed at 12:15 AM

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text RAM Disk Editing

[3] #Checking the size of our data  
df.shape  
(200, 5)

[4] #Changing the name of some columns  
df = df.rename(columns={'Annual Income (k\$)': 'Annual\_income', 'Spending Score (1-100)': 'Spending\_score'})

[5] #Looking for null values  
df.isna().sum()

CustomerID	0
Gender	0
Age	0
Annual_income	0
Spending_score	0
dtype:	int64

[6] #Checking datatypes  
df.info()

<class 'pandas.core.frame.DataFrame'>			
RangeIndex: 200 entries, 0 to 199			
Data columns (total 5 columns):			
#	Column	Non-Null Count	Dtype
---	---	-----	----
0	CustomerID	200 non-null	int64
1	Gender	200 non-null	object
2	Age	200 non-null	int64
3	Annual_income	200 non-null	int64

0s completed at 12:15 AM

Type here to search

23°C Clear 01:33 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text

RAM Disk

Editing

[6] 1 Gender 200 non-null object  
2 Age 200 non-null int64  
3 Annual\_income 200 non-null int64  
4 Spending\_score 200 non-null int64  
dtypes: int64(4), object(1)  
memory usage: 7.9+ KB

[7] #Replacing objects for numerical values  
df['Gender'].replace(['Female','Male'], [0,1], inplace=True)

[8] #Checking values have been replaced properly  
df.Gender

0 1  
1 1  
2 0  
3 0  
4 0  
..  
195 0  
196 0  
197 1  
198 1  
199 1  
Name: Gender, Length: 200, dtype: int64

[9] #Density estimation of values using distplot  
plt.figure(1 , figsize = (15 , 6))  
feature\_list = ['Age','Annual\_income', "Spending\_score"]  
feature\_listt = ['Age','Annual\_income', "Spending\_score"]  
pos = 1

0s completed at 12:15 AM

Type here to search

23°C Clear 01:34 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

✓ RAM Disk

Editing

[9] #Density estimation of values using distplot

```
1s plt.figure(1 , figsize = (15 , 6))
feature_list = ['Age', 'Annual_income', "Spending_score"]
feature_listt = ['Age', 'Annual_income', "Spending_score"]
pos = 1
for i in feature_list:
    plt.subplot(1 , 3 , pos)
    plt.subplots_adjust(hspace = 0.5 , wspace = 0.5)
    sns.distplot(df[i], bins=20, kde = True)
    pos = pos + 1
plt.show()
```

Density

0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035 0.040

0 20 40 60 80

Density

0.0000 0.0025 0.0050 0.0075 0.0100 0.0125 0.0150 0.0175 0.0200

0 50 100 150

Density

0.0000 0.0025 0.0050 0.0075 0.0100 0.0125 0.0150 0.0175 0.0200

0 50 100

✓ 0s completed at 12:15 AM

Type here to search

O C G M S E

23°C Clear 01:34 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text RAM Disk Editing

0s #Count and plot gender  
sns.countplot(y = 'Gender', data = df, palette="husl", hue = "Gender")  
df["Gender"].value\_counts()

{x} 0 112  
1 88  
Name: Gender, dtype: int64

Gender

0

1

count

Gender

0

1

0 20 40 60 80 100

[11] #Pairplot with variables we want to study  
sns.pairplot(df, vars=["Age", "Annual\_income", "Spending\_score"], kind ="reg", hue = "Gender", palette="husl", markers = ['o', 'D'])

<seaborn.axisgrid.PairGrid at 0x7fb66eca5d0>

70  
60  
50

0s completed at 12:15 AM

Type here to search

23°C Clear 01:34 ENG 23-11-2021



Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text

[11] sns.lmplot(x = "Age", y = "Annual\_income", data = df, hue = "Gender")

{x} <seaborn.axisgrid.FacetGrid at 0x7fbc634ca290>

Annual\_income

Gender

0 1

Age

[13] sns.lmplot(x = "Annual\_income", y = "Spending\_score", data = df, hue = "Gender")

<seaborn.axisgrid.FacetGrid at 0x7fbc6355cb50>

Type here to search

0s completed at 12:15 AM

23°C Clear 01:34 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text RAM Disk Editing

1s [1] sns.lmplot(x = "Annual\_income", y = "Spending\_score", data = df, hue = "Gender")

<seaborn.axisgrid.FacetGrid at 0x7fbc6355cb50>

Scatter plot showing Spending score vs Annual income, colored by Gender. The x-axis ranges from 20 to 140, and the y-axis ranges from 0 to 100. Two regression lines are shown: one for females (blue) and one for males (orange). Shaded regions represent confidence intervals.

{x}

2s [14] sns.lmplot(x = "Age", y = "Spending\_score", data = df, hue = "Gender")

<seaborn.axisgrid.FacetGrid at 0x7fbc66df9a90>

Scatter plot showing Spending score vs Age, colored by Gender. The x-axis ranges from approximately 20 to 100, and the y-axis ranges from 0 to 100. Two regression lines are shown: one for females (blue) and one for males (orange).

0s completed at 12:15 AM

Type here to search

23°C Clear 01:34 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

RAM Disk Editing

[13] Annual\_income

1s

2s sns.lmplot(x = "Age", y = "Spending\_score", data = df, hue = "Gender")

{x}

2s

Scatter plot showing Spending score vs Age, colored by Gender (0: blue, 1: orange). The plot includes two regression lines and shaded confidence intervals.

Spending\_score

Age

Gender

0

1

[15] #Creating values for the elbow

2s X = df.loc[:, ["Age", "Annual\_income", "Spending\_score"]]

inertia = []

k = range(1, 20)

0s completed at 12:15 AM

Type here to search

23°C Clear 01:34 ENG 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk

Comment Share

Editing

#Creating values for the elbow

```
X = df.loc[:, "Age", "Annual_income", "Spending_score"]
inertia = []
k = range(1,20)
for i in k:
    means_k = KMeans(n_clusters=i, random_state=0)
    means_k.fit(X)
    inertia.append(means_k.inertia_)
```

[16] #Plotting the elbow

```
plt.plot(k , inertia , 'bo-')
plt.xlabel('Number of Clusters') , plt.ylabel('Inertia')
plt.show()
```

The figure is a line graph titled 'Elbow Method Plot'. The x-axis is labeled 'Number of Clusters' and ranges from 2.5 to 17.5. The y-axis is labeled 'Inertia' and ranges from 0 to 300,000. The data points show a sharp decrease in inertia as the number of clusters increases from 1 to approximately 5, after which the rate of decrease slows down significantly, forming a distinct 'elbow' shape.

Number of Clusters	Inertia
1	~310,000
2	~215,000
3	~145,000
4	~105,000
5	~80,000
6	~65,000
7	~55,000
8	~50,000
9	~48,000
10	~45,000
11	~43,000
12	~41,000
13	~39,000
14	~37,000
15	~35,000
16	~33,000
17	~31,000
18	~30,000
19	~29,000
20	~28,000

0s completed at 12:15 AM

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKI0D1eYyMx3S7KaH0BX4tlmsonyqKWM#scrollTo=5zFxzGRd8a7B

Apps Gmail YouTube (3) WhatsApp academia-academic... Classes RA1911042010102... Welcome to srmist... All Events Search Snippets an... 1. Introduction to A... Reading list

Mall Customer Segmentation.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[17] #Training kmeans with 5 clusters  
means\_k = KMeans(n\_clusters=5, random\_state=0)  
means\_k.fit(X)  
labels = means\_k.labels\_  
centroids = means\_k.cluster\_centers\_

[18] #Create a 3d plot to view the data separation made by Kmeans  
trace1 = go.Scatter3d(  
 x= X[ 'Spending\_score'],  
 y= X[ 'Annual\_income'],  
 z= X[ 'Age'],  
 mode='markers',  
 marker=dict(  
 color = labels,  
 size= 10,  
 line=dict(  
 color= labels,  
 ),  
 opacity = 0.9  
 )  
)  
layout = go.Layout(  
 title= 'Clusters',  
 scene = dict(  
 xaxis = dict(title = 'Spending\_score'),  
 yaxis = dict(title = 'Annual\_income'),  
 zaxis = dict(title = 'Age')  
 )  
)  
fig = go.Figure(data=trace1, layout=layout)

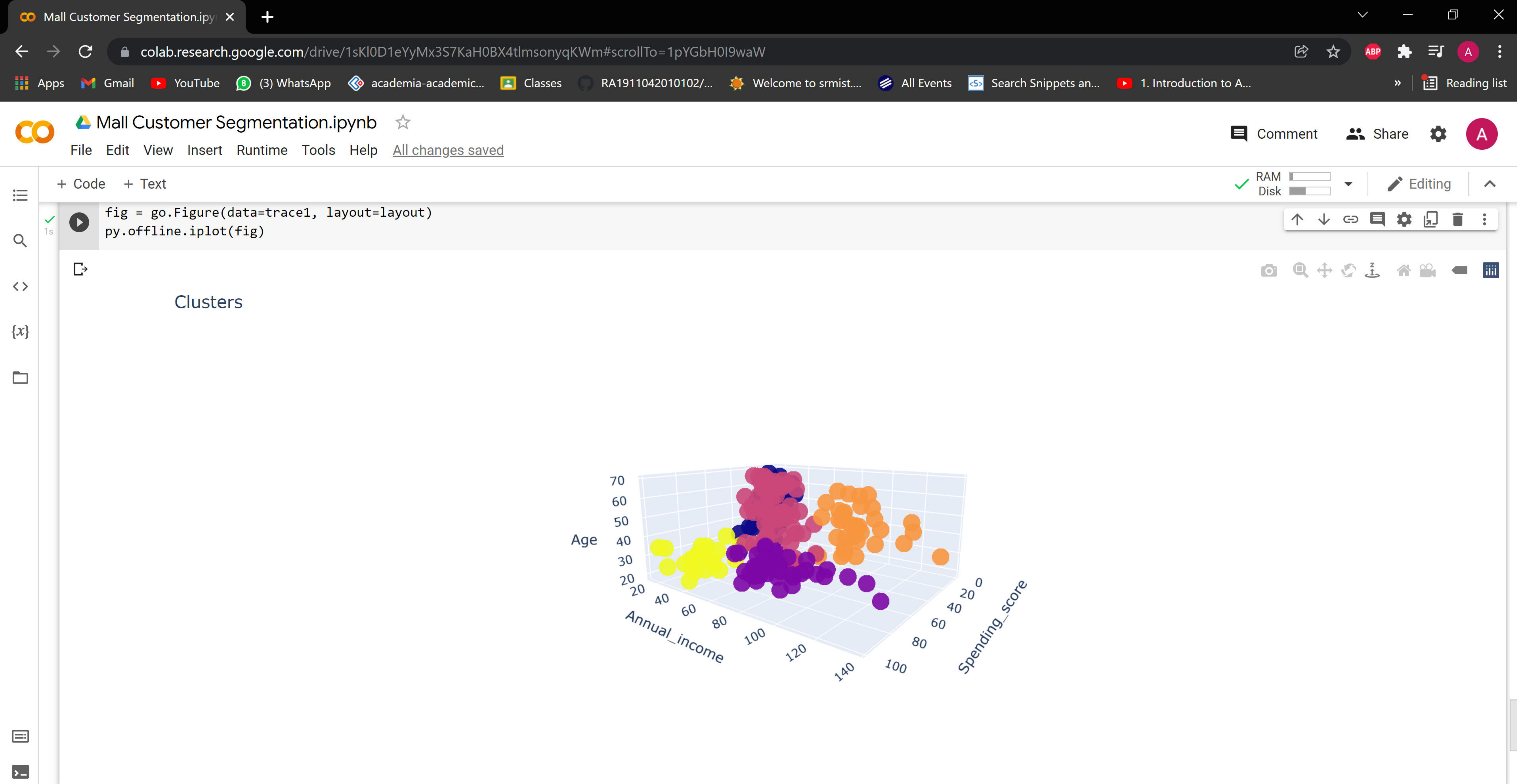
RAM Disk

Comment Share

Editing

Type here to search

23°C Clear 01:34 ENG 23-11-2021



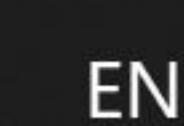
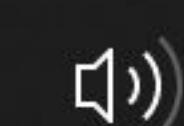
✓ 0s completed at 12:15 AM



Type here to search



23°C Clear



ENG

01:35

23-11-2021

