Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

```
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")
```

2s

[2]
```
#We read the csv and print the first 5 rows
df = pd.read_csv("/content/Mall_Customers.csv")
df.head()
```

0s

	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

[3]
```
#Checking the size of our data
df.shape
```

0s

0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:33 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

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

```
(200, 5)
```

[4] 0s
```
#Changing the name of some columns
df = df.rename(columns={'Annual Income (k$)': 'Annual_income', 'Spending Score (1-100)': 'Spending_score'})
```

[5] 0s
```
#Looking for null values
df.isna().sum()
```

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

[6] 0s
```
#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 ENG 01:33 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

```
 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
```

[6]

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

[7]

```
#Checking values have been replaced properly
df.Gender
```

[8]

```
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
```

```
#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
```

[9]

0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tImsonyqKWm#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

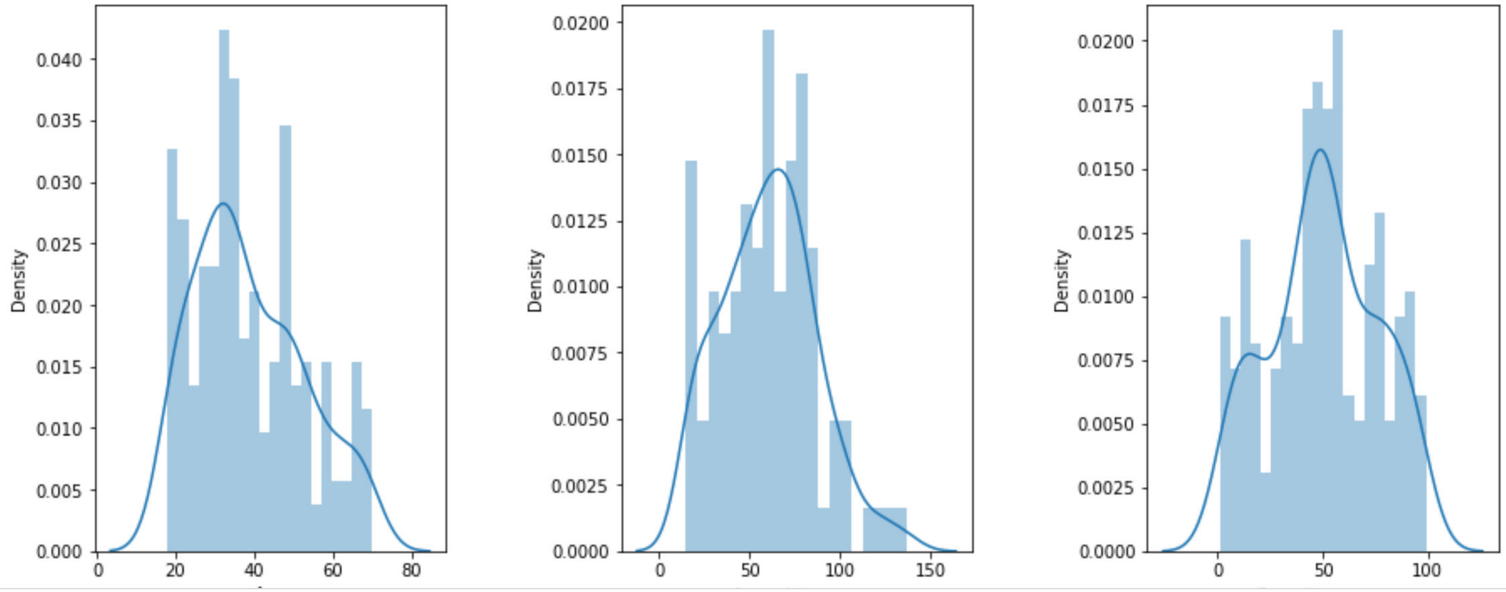
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Editing

```
[9]  #Density estimation of values using distplot
     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()
```



0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

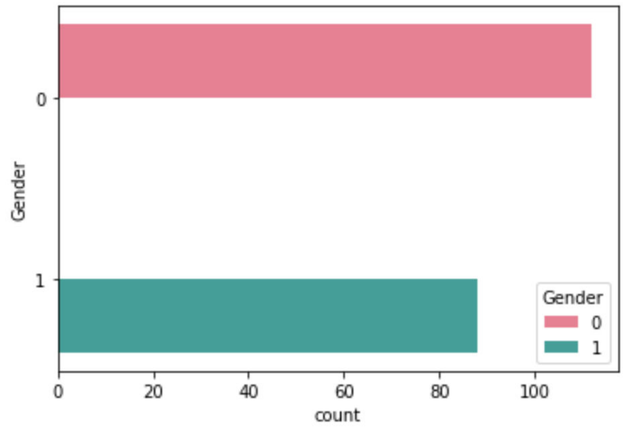
Comment Share

+ Code + Text

RAM Disk Editing

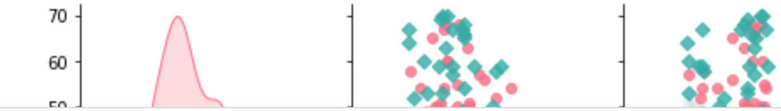
```
#Count and plot gender
sns.countplot(y = 'Gender', data = df, palette="husl", hue = "Gender")
df["Gender"].value_counts()
```

```
0    112
1     88
Name: Gender, dtype: int64
```




```
[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 0x7fbc66eca5d0>
```



0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipynb

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

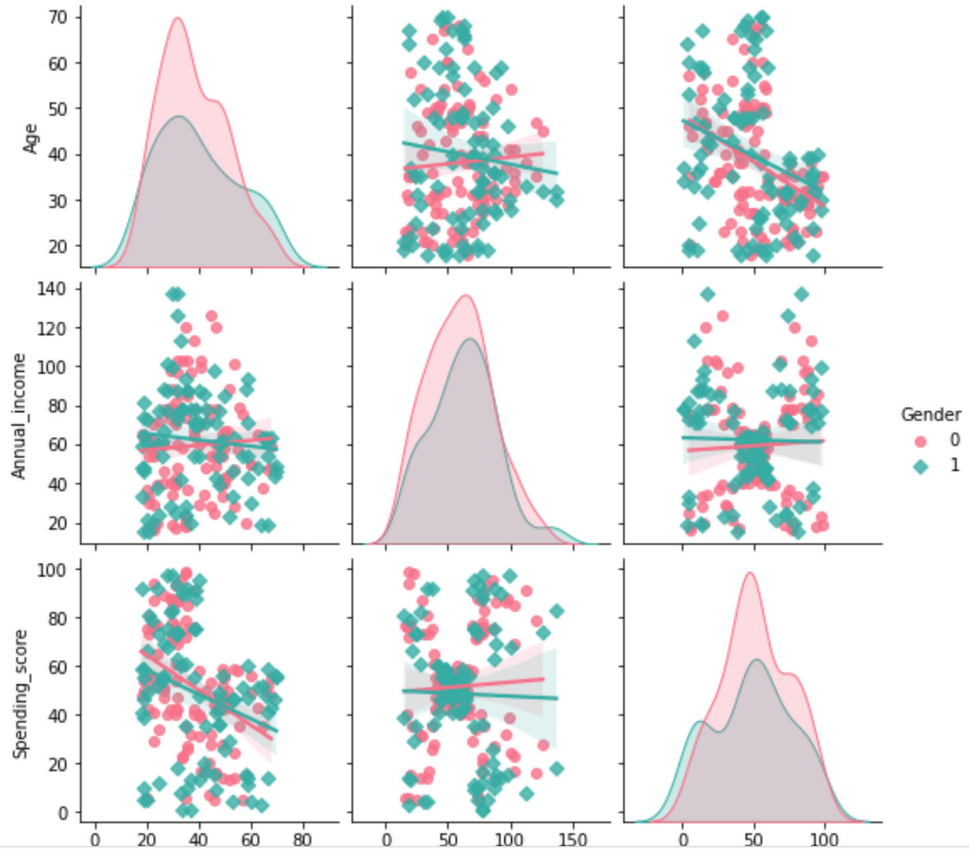
Comment Share

+ Code + Text

RAM Disk Editing

```
#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 0x7fbc66eca5d0>
```



0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

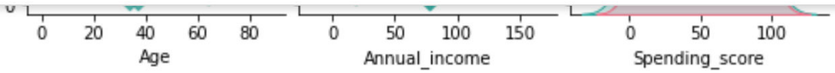
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

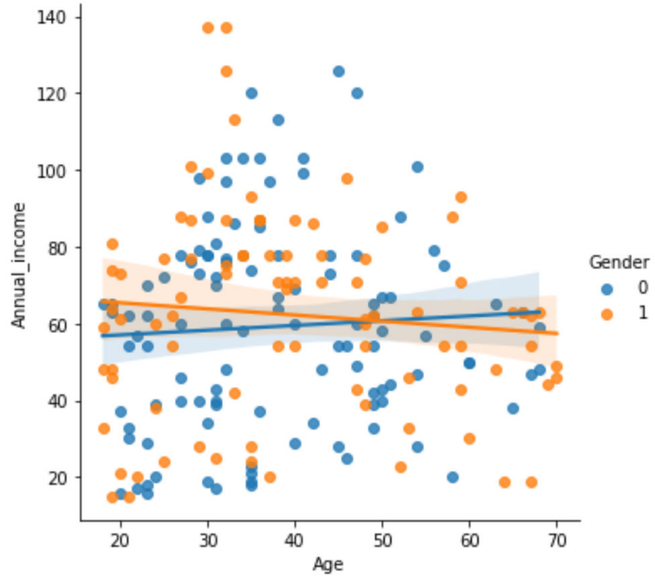
RAM Disk Editing

[11]




```
sns.lmplot(x = "Age", y = "Annual_income", data = df, hue = "Gender")
```

```
<seaborn.axisgrid.FacetGrid at 0x7fbc634ca290>
```




```
[13] sns.lmplot(x = "Annual_income", y = "Spending_score", data = df, hue = "Gender")
```

```
<seaborn.axisgrid.FacetGrid at 0x7fbc6355cb50>
```

0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

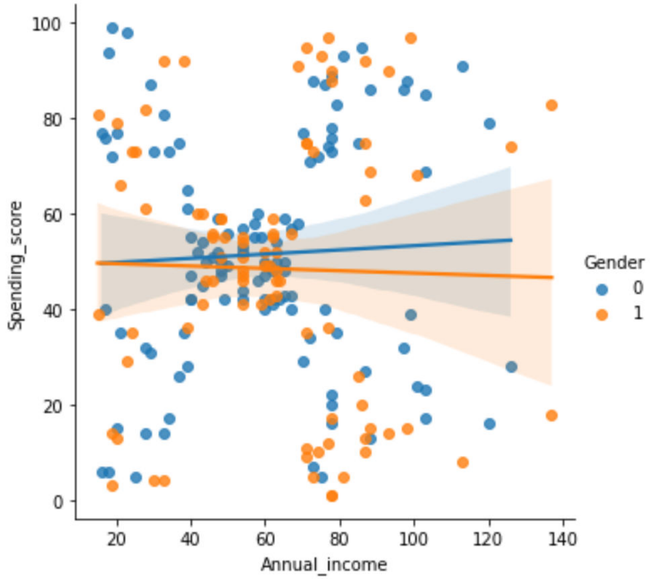
+ Code + Text

RAM Disk Editing

```
sns.lmplot(x = "Annual_income", y = "Spending_score", data = df, hue = "Gender")
```

1s

```
<seaborn.axisgrid.FacetGrid at 0x7fbc6355cb50>
```




```
[14] sns.lmplot(x = "Age", y = "Spending_score", data = df, hue = "Gender")
```

2s

```
<seaborn.axisgrid.FacetGrid at 0x7fbc66df9a90>
```

0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

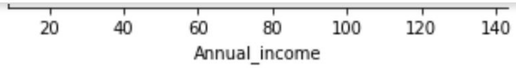
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

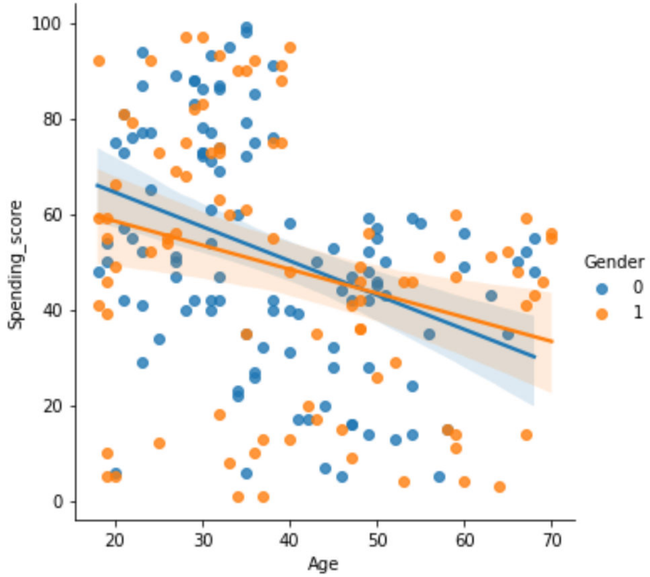
RAM Disk Editing

[13]




```
sns.lmplot(x = "Age", y = "Spending_score", data = df, hue = "Gender")
```

```
<seaborn.axisgrid.FacetGrid at 0x7fbc66df9a90>
```




```
[15] #Creating values for the elbow
     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 ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

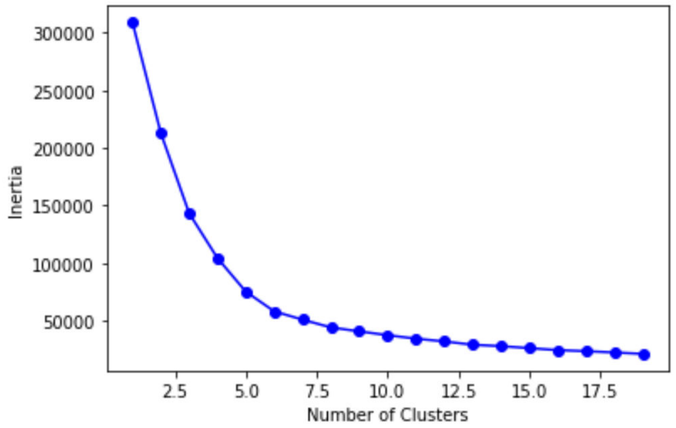
Age

```
#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_)
```

2s

```
[16] #Plotting the elbow
plt.plot(k , inertia , 'bo-')
plt.xlabel('Number of Clusters') , plt.ylabel('Inertia')
plt.show()
```

0s



0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#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

[17] 0s
```
#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] 1s
```
#Create a 3d plot to view the data sepparation 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)
```

0s completed at 12:15 AM

Type here to search 23°C Clear ENG 01:34 23-11-2021

Mall Customer Segmentation.ipy

colab.research.google.com/drive/1sKl0D1eYyMx3S7KaH0BX4tlmsonyqKWm#scrollTo=1pYGbH0l9waW

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

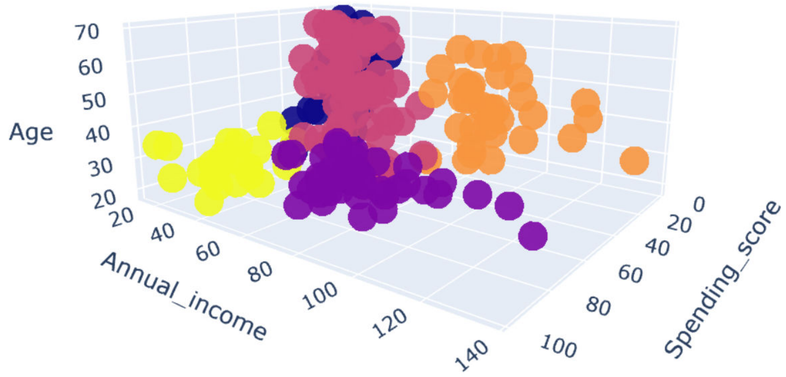
Comment Share

+ Code + Text

RAM Disk Editing

```
fig = go.Figure(data=trace1, layout=layout)
py.offline.iplot(fig)
```

Clusters



0s completed at 12:15 AM

Type here to search

23°C Clear ENG 01:35 23-11-2021