

Assignment -4

Assignment Date	8 October 2022
Student Name	M.Shehha
Student Roll Number	E1194034
Maximum Marks	2 Marks

Colab interface showing the initial code execution for 'assignment_4 (1).ipynb'. The code imports necessary libraries and reads a CSV file.

```
import os
os.getcwd()

'C:\Users\pc'

[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.cluster import KMeans

[ ] #2
path='C:\Users\pc\downloads\'
data=pd.read_csv(path+'Mall_Customers.csv')
data.head()
```

The output shows the current directory and the first five rows of the 'Mall_Customers.csv' file.

	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

Colab interface showing the continuation of the code execution. The code displays the shape of the data and creates a Seaborn facet grid for visualization.

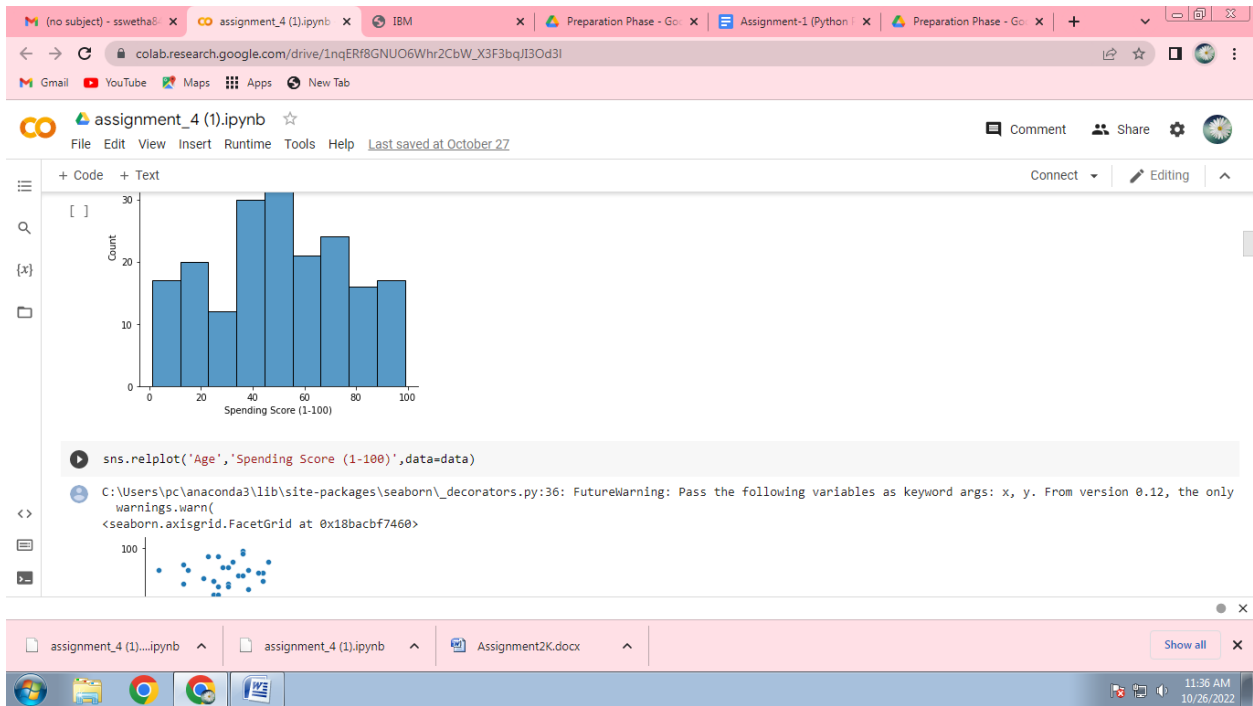
```
[ ] data.head()

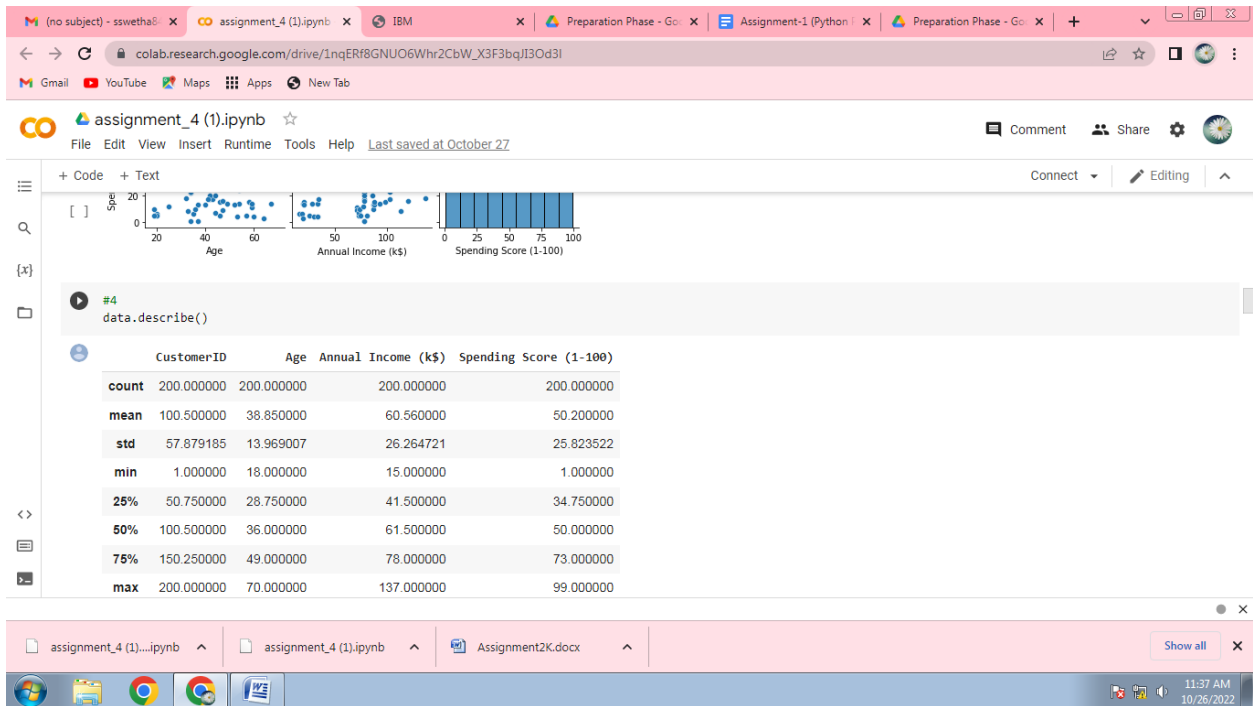
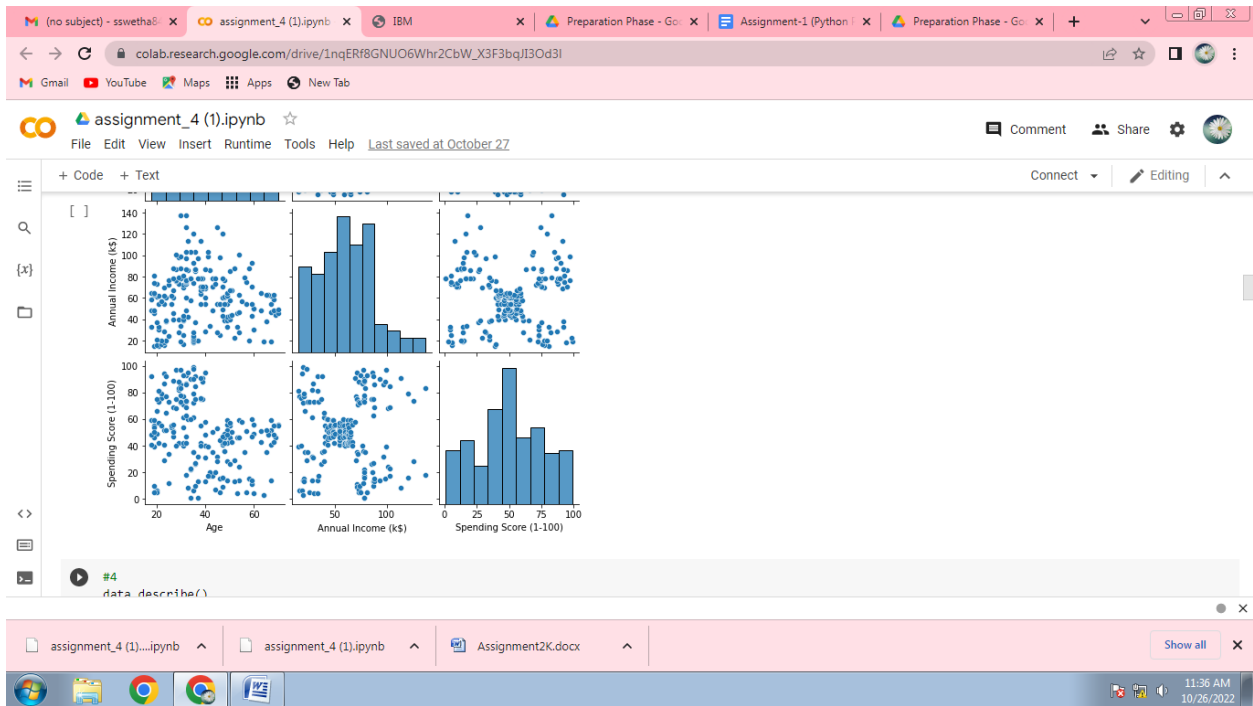
[ ] data.shape

(200, 5)

sns.displot(data['Spending Score (1-100)'])
<seaborn.axisgrid.FacetGrid at 0x18bac797e50>
```

The output shows the shape of the data and a Seaborn facet grid for visualization.





colab.research.google.com/drive/1nqERf8GNUO6Whr2CbW_X3F3bqJI3Od3I

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ]: max 200.000000 49.000000 70.000000 137.000000 99.000000
```

```
[ ]: #S
data.isnull().sum()

CustomerID      0
Gender          0
Age             0
Annual Income (k$) 0
Spending Score (1-100) 0
dtype: int64
```

```
[ ]: data.skew()

CustomerID      0.000000
Age             0.485569
Annual Income (k$) 0.321843
Spending Score (1-100) -0.047220
dtype: float64
```

```
sns.displot(data['Age'], kind='kde')
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:37 AM 10/26/2022

colab.research.google.com/drive/1nqERf8GNUO6Whr2CbW_X3F3bqJI3Od3I

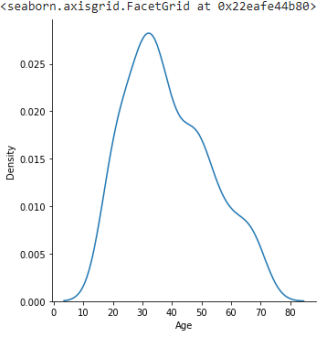
assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

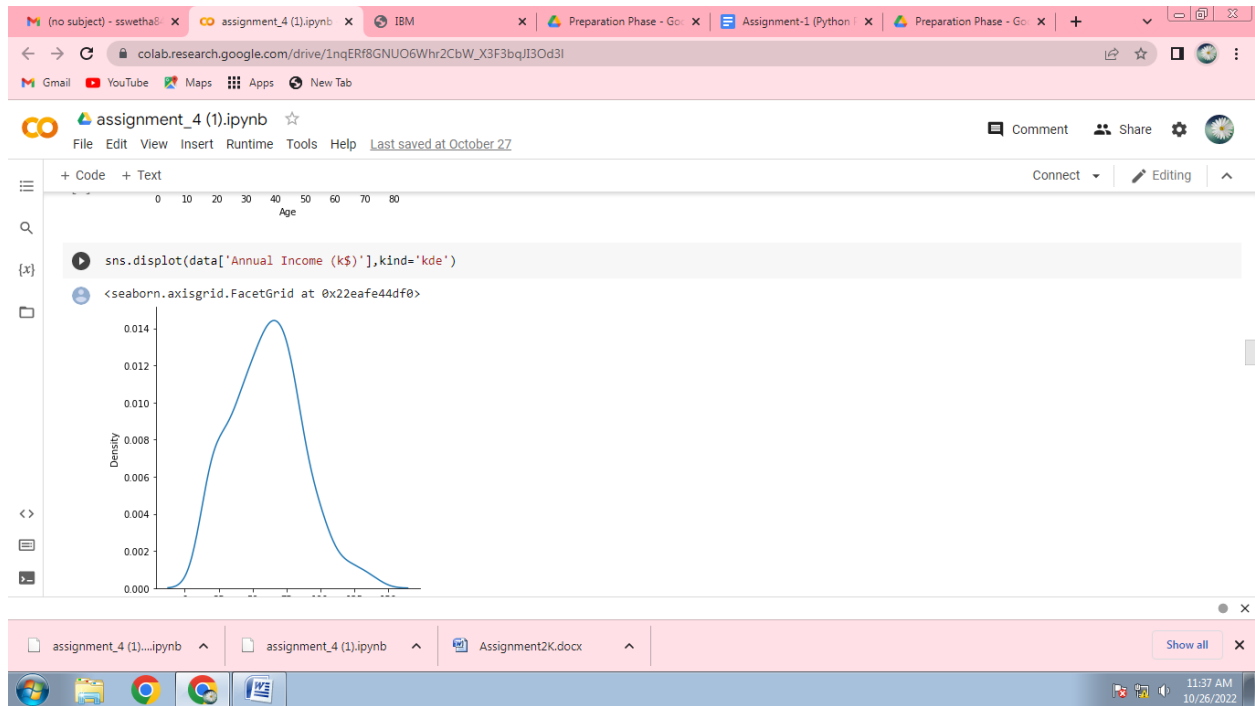
```
sns.displot(data['Age'], kind='kde')
```

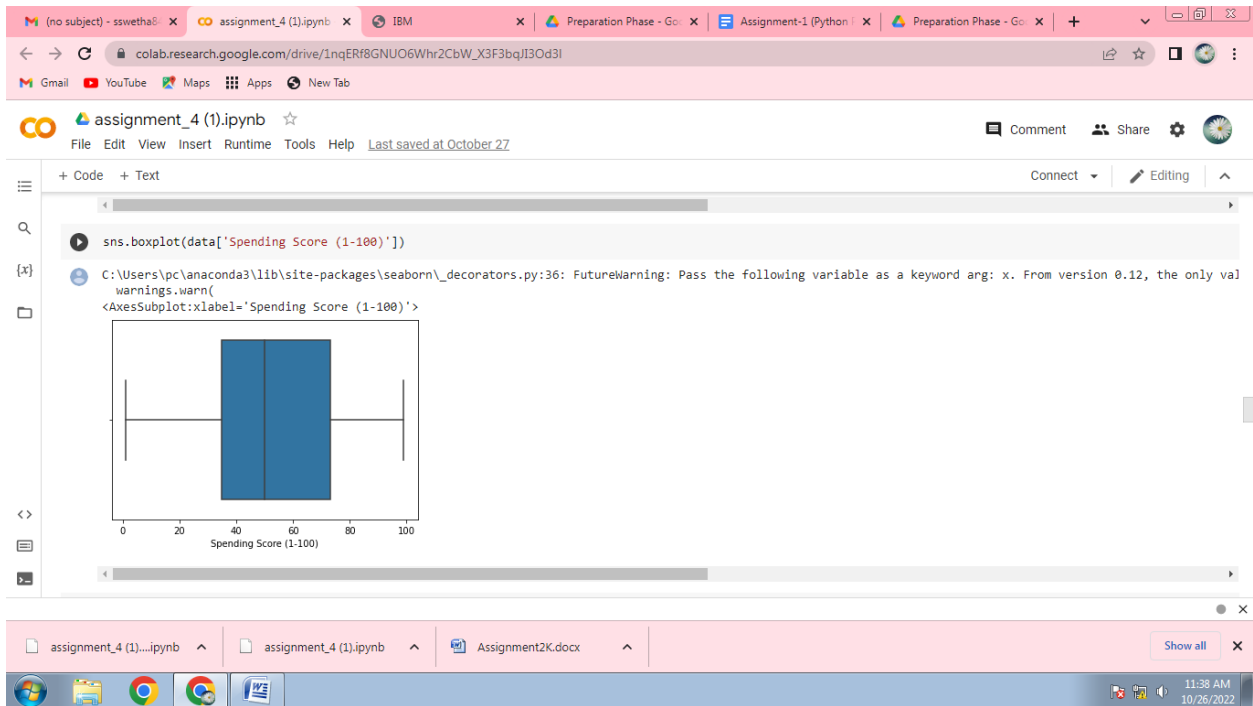
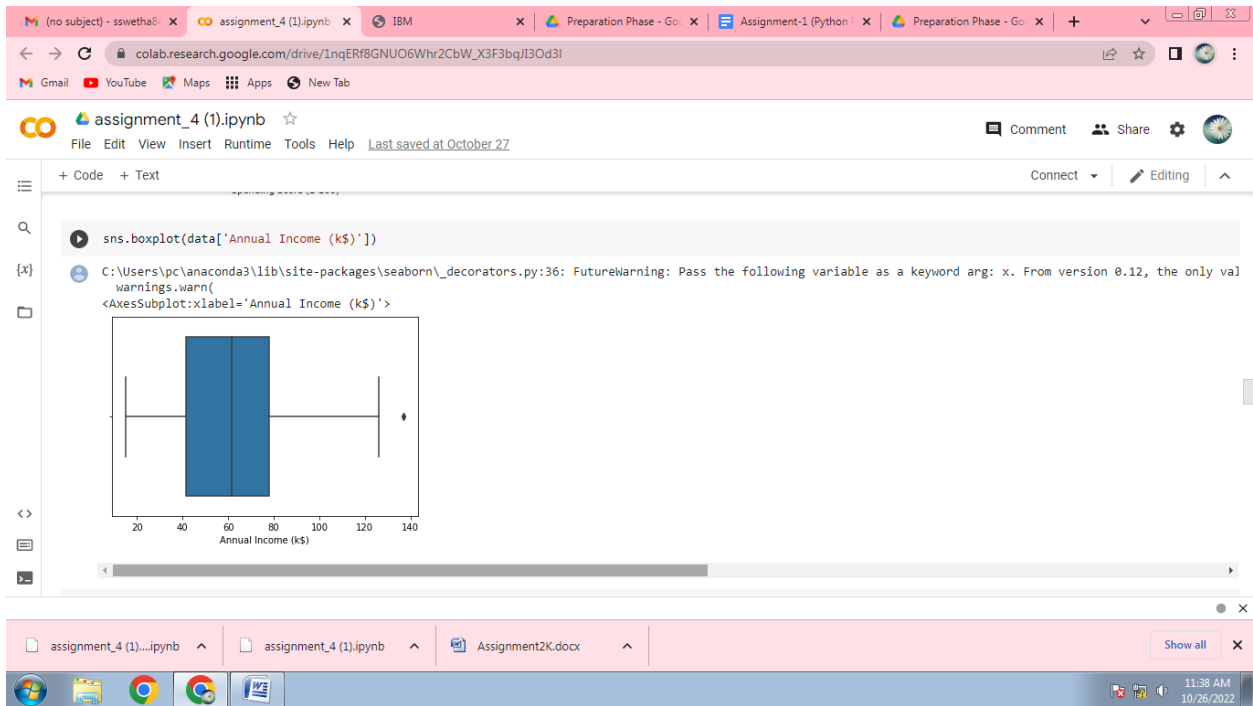
```
<seaborn.axisgrid.FacetGrid at 0x22eafe44b80>
```



assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:37 AM 10/26/2022





assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
#replacing outliers
data['Annual Income (k$)'].replace(mean)
```

0 15
1 15
2 16
3 16
4 17
...
195 120
196 126
197 126
198 137
199 137
Name: Annual Income (k\$), Length: 200, dtype: int64

```
[ ] data['Annual Income (k$)'].isnull().sum()
0
```

```
[ ] #7
data_main=pd.get_dummies(data,columns=['Gender'])
```

assignment_4 (1).ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:38 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] #7
data_main=pd.get_dummies(data,columns=['Gender'])
```

data_main

	CustomerID	Age	Annual Income (k\$)	Spending Score (1-100)	Gender_Female	Gender_Male
0	1	19	15	39	0	1
1	2	21	15	81	0	1
2	3	20	16	6	1	0
3	4	23	16	77	1	0
4	5	31	17	40	1	0
...
195	196	35	120	79	1	0
196	197	45	126	28	1	0
197	198	32	126	74	0	1
198	199	32	137	18	0	1

assignment_4 (1).ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:39 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] 197 198 32 126 74 0 1
    198 199 32 137 18 0 1
    199 200 30 137 83 0 1
```

200 rows x 6 columns

```
type(data_main['Gender_Female'])
pandas.core.series.Series
```

```
[ ] x=data_main[['Age','Annual Income (k$)','Spending Score (1-100)']]
```

```
x
```

	Age	Annual Income (k\$)
0	19	15
1	21	15
2	20	16

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:39 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] ... ...
```

```
195 35 120
196 45 126
197 32 126
198 32 137
199 30 137
```

200 rows x 2 columns

```
#8 scaling
min_max=MinMaxScaler(feature_range=(0,1))
```

```
[ ] norm=min_max.fit_transform(x)
```

```
[ ] norm
```

```
array([[0.01923077, 0. , 0.3877551 ],
       [0.0760334 , 0. , 0.8633653 ]])
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:39 AM 10/26/2022

Colab interface showing a Jupyter Notebook with the following code and output:

```
[ ] norm=min_max.fit_transform(x)
```

Output (norm):

```
array([[0.01923077, 0.03877551],
       [0.05769231, 0.81632653],
       [0.03846154, 0.00819672, 0.05102041],
       [0.09615385, 0.00819672, 0.7755102 ],
       [0.09615385, 0.01639344, 0.39795918],
       [0.25, 0.01639344, 0.76530612],
       [0.07692308, 0.01639344, 0.05102041],
       [0.32692308, 0.02459016, 0.05102041],
       [0.09615385, 0.02459016, 0.94897959],
       [0.88461538, 0.03278689, 0.02040816],
       [0.23076923, 0.03278689, 0.7244898 ],
       [0.94230769, 0.03278689, 0.13265306],
       [0.32692308, 0.03278689, 1.],
       [0.76923077, 0.04098361, 0.14285714],
       [0.11538462, 0.04098361, 0.7755102 ],
       [0.36538462, 0.04098361, 0.12244898],
       [0.07692308, 0.04098361, 0.79591837],
       [0.32692308, 0.04918033, 0.34693878],
       [0.03846154, 0.04918033, 0.66326531],
       [0.65384615, 0.06557377, 0.28571429]])
```

Taskbar shows: assignment_4 (1).ipynb, assignment_4 (1).ipynb, Assignment2K.docx, 11:39 AM 10/26/2022

Colab interface showing a Jupyter Notebook with the following code and output:

```
[ ] norm=min_max.fit_transform(x)
```

Output (norm):

```
[0.03846154, 0.04918033, 0.66326531],
[0.65384615, 0.06557377, 0.28571429],
[0.32692308, 0.06557377, 0.98979592],
[0.32692308, 0.07377049, 0.34693878],
[0.13461538, 0.07377049, 0.73469388],
[0.53846154, 0.08196721, 0.04081633],
[0.25, 0.08196721, 0.73469388],
[0.69230769, 0.10655738, 0.13265306],
[0.21153846, 0.10655738, 0.82653061],
[0.51923077, 0.10655738, 0.31632653],
[0.32692308, 0.10655738, 0.6122449 ],
[0.42307692, 0.1147541, 0.30612245],
[0.09615385, 0.1147541, 0.87755102],
[0.80769231, 0.12295082, 0.03061224],
[0.05769231, 0.12295082, 0.73469388],
[0.67307692, 0.14754098, 0.03061224],
[0., 0.14754098, 0.92857143],
[0.59615385, 0.14754098, 0.13265306],
[0.05769231, 0.14754098, 0.81632653],
```

Taskbar shows: assignment_4 (1).ipynb, assignment_4 (1).ipynb, Assignment2K.docx, 11:40 AM 10/26/2022

Google Colab interface showing a Jupyter Notebook titled "assignment_4 (1).ipynb". The code cell contains the following Python code:

```
[ ] norm=min_max.fit_transform(x)
```

The output cell shows a 2D array of normalized data (min-max scaled) with 15 rows and 3 columns. The values range from approximately 0.05 to 0.59.

Taskbar shows the following open applications: assignment_4 (1).ipynb, assignment_4 (1).ipynb, Assignment2K.docx. The system clock indicates 11:40 AM on 10/26/2022.

Google Colab interface showing a Jupyter Notebook titled "assignment_4 (1).ipynb". The code cell contains the following Python code:

```
[ ] norm=min_max.fit_transform(x)
```

The output cell shows a 2D array of normalized data (min-max scaled) with 15 rows and 3 columns. The values range from approximately 0.05 to 0.59.

Taskbar shows the following open applications: assignment_4 (1).ipynb, assignment_4 (1).ipynb, Assignment2K.docx. The system clock indicates 11:40 AM on 10/26/2022.

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] norm=min_max.fit_transform(x)
```

{x}

norm

```
[0.26923077, 0.2704918, 0.46938776],  
[1., 0.27868852, 0.55102041],  
[0.55769231, 0.27868852, 0.41836735],  
[0.80769231, 0.28688525, 0.48979592],  
[0.80769231, 0.28688525, 0.56122449],  
[0.78846154, 0.31967213, 0.46938776],  
[0.15384615, 0.31967213, 0.54081633],  
[0.51923077, 0.31967213, 0.53061224],  
[0.42307692, 0.31967213, 0.47959184],  
[0.09615385, 0.31967213, 0.52040816],  
[0.59615385, 0.31967213, 0.41836735],  
[0.75, 0.31967213, 0.51020408],  
[0.38461538, 0.31967213, 0.55102041],  
[0.94230769, 0.31967213, 0.40816327],  
[0.53846154, 0.31967213, 0.43877551],  
[0.05769231, 0.31967213, 0.57142857],  
[0.57692308, 0.31967213, 0.45918367],  
[0.71153846, 0.3442623, 0.58163265],  
[0.07692308, 0.3442623, 0.55102041],
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:40 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] norm=min_max.fit_transform(x)
```

{x}

norm

```
[0.57692308, 0.36885246, 0.48979592],  
[0.42307692, 0.36885246, 0.39795918],  
[0.26923077, 0.36885246, 0.41836735],  
[0.11538462, 0.36885246, 0.52040816],  
[0.55769231, 0.36885246, 0.46938776],  
[0.17307692, 0.36885246, 0.5],  
[0.57692308, 0.37704918, 0.41836735],  
[0.03846154, 0.37704918, 0.48979592],  
[0.09615385, 0.3852459, 0.40816327],  
[0.59615385, 0.3852459, 0.47959184],  
[0.94230769, 0.3852459, 0.59183673],  
[0.15384615, 0.3852459, 0.55102041],  
[0.59615385, 0.3852459, 0.56122449],  
[0.05769231, 0.3852459, 0.41836735],  
[0.92307692, 0.39344262, 0.5],  
[0.69230769, 0.39344262, 0.45918367],  
[0.96153846, 0.39344262, 0.42857143],  
[0.92307692, 0.39344262, 0.47959184],  
[0.90384615, 0.39344262, 0.52040816],
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:41 AM 10/26/2022

Google Colab interface showing a Jupyter Notebook titled "assignment_4 (1).ipynb". The code cell contains the following Python code:

```
[ ] norm=min_max.fit_transform(x)
```

The output cell shows a 2D array of normalized data:

```
norm
[[0.92307692, 0.39344262, 0.47959184],
 [0.90384615, 0.39344262, 0.52040816],
 [0.01923077, 0.39344262, 0.54081633],
 [0.38461538, 0.40163934, 0.41836735],
 [0.01923077, 0.40163934, 0.45918367],
 [0.         , 0.40983607, 0.47959184],
 [0.01923077, 0.40983607, 0.5         ],
 [0.86538462, 0.40983607, 0.42857143],
 [0.59615385, 0.40983607, 0.59183673],
 [0.63461538, 0.42622951, 0.42857143],
 [0.61538462, 0.42622951, 0.57142857],
 [0.17307692, 0.42622951, 0.56122449],
 [0.38461538, 0.42622951, 0.39795918],
 [0.42307692, 0.44262295, 0.58163265],
 [0.40384615, 0.44262295, 0.91836735],
 [0.09615385, 0.45081967, 0.28571429],
 [0.25         , 0.45081967, 0.7755102 ],
 [0.48076923, 0.45901639, 0.34693878],
 [0.42307692, 0.45901639, 0.95918367]]
```

The interface includes a file explorer showing "assignment_4 (1).ipynb" and "Assignment2K.docx". The system tray shows the time as 11:41 AM on 10/26/2022.

Google Colab interface showing a Jupyter Notebook titled "assignment_4 (1).ipynb". The code cell contains the following Python code:

```
[ ] norm=min_max.fit_transform(x)
```

The output cell shows a 2D array of normalized data:

```
norm
[[0.48076923, 0.45901639, 0.34693878],
 [0.42307692, 0.45901639, 0.95918367],
 [0.78846154, 0.45901639, 0.10204082],
 [0.38461538, 0.45901639, 0.75510204],
 [0.55769231, 0.45901639, 0.08163265],
 [0.40384615, 0.45901639, 0.75510204],
 [0.13461538, 0.46721311, 0.33673469],
 [0.25         , 0.46721311, 0.71428571],
 [0.03846154, 0.47540984, 0.04081633],
 [0.21153846, 0.47540984, 0.8877551 ],
 [0.5         , 0.47540984, 0.06122449],
 [0.26923077, 0.47540984, 0.73469388],
 [0.01923077, 0.48306056, 0.09183673],
 [0.32692308, 0.48306056, 0.7244898 ],
 [0.75         , 0.49180328, 0.04081633],
 [0.26923077, 0.49180328, 0.93877551],
 [0.19230769, 0.5         , 0.39795918],
 [0.26923077, 0.5         , 0.87755102],
 [0.13461538, 0.50819672, 0.1122449 ],
 [0.42307692, 0.50819672, 0.93877551]]
```

The interface includes a file explorer showing "assignment_4 (1).ipynb" and "Assignment2K.docx". The system tray shows the time as 11:41 AM on 10/26/2022.

Google Chrome browser window showing a Jupyter Notebook titled "assignment_4 (1).ipynb". The notebook is open in "Code" view, displaying a list of 30 data points, each represented as a list of three floating-point numbers. The browser's address bar shows the URL: colab.research.google.com/drive/1nqERi8GNUO6Whr2CbW_X3F3bqJI3Od3l. The notebook interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with options like "Connect", "Editing", and "Comment". The taskbar at the bottom shows the Windows operating system with various application icons and a system clock indicating 11:41 AM on 10/26/2022.

```
[0.19230769, 0.5, 0.39795918],
[0.26923077, 0.5, 0.87755102],
[0.13461538, 0.50819672, 0.1122449 ],
[0.19230769, 0.50819672, 0.97959184],
[0.57692308, 0.50819672, 0.35714286],
[0.26923077, 0.50819672, 0.74489796],
[0.30769231, 0.51639344, 0.21428571],
[0.30769231, 0.51639344, 0.90816327],
[0.48076923, 0.51639344, 0.16326531],
[0.40384615, 0.51639344, 0.8877551 ],
[0.5, 0.51639344, 0.19387755],
[0.38461538, 0.51639344, 0.76530612],
[0.55769231, 0.51639344, 0.15306122],
[0.17307692, 0.51639344, 0.89795918],
[0.36538462, 0.51639344, 0. ],
[0.23076923, 0.51639344, 0.78571429],
[0.30769231, 0.51639344, 0. ],
[0.23076923, 0.51639344, 0.73469388],
[0.73076923, 0.52459016, 0.34693878],
[0.21153846, 0.52459016, 0.83673469],
[0.01923077, 0.54098361, 0.04081633],
[0.25, 0.54098361, 0.93877551],
[0.61538462, 0.57377049, 0.25510204],
[0.34615385, 0.57377049, 0.75510204],
[0.46153846, 0.58196721, 0.19387755],
```

Google Chrome browser window showing the same Jupyter Notebook "assignment_4 (1).ipynb". The notebook is open in "Code" view, displaying a list of 30 data points, each represented as a list of three floating-point numbers. The browser's address bar shows the URL: colab.research.google.com/drive/1nqERi8GNUO6Whr2CbW_X3F3bqJI3Od3l. The notebook interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a toolbar with options like "Connect", "Editing", and "Comment". The taskbar at the bottom shows the Windows operating system with various application icons and a system clock indicating 11:42 AM on 10/26/2022.

```
[0.42307692, 0.59016393, 0.12244898],
[0.19230769, 0.59016393, 0.75510204],
[0.34615385, 0.59016393, 0.09183673],
[0.34615385, 0.59016393, 0.92857143],
[0.65384615, 0.59836066, 0.12244898],
[0.23076923, 0.59836066, 0.86734694],
[0.76923077, 0.59836066, 0.14285714],
[0.17307692, 0.59836066, 0.69387755],
[0.78846154, 0.63934426, 0.13265306],
[0.32692308, 0.63934426, 0.90816327],
[0.36538462, 0.67213115, 0.31632653],
[0.26923077, 0.67213115, 0.86734694],
[0.53846154, 0.68032787, 0.14285714],
[0.21153846, 0.68032787, 0.8877551 ],
[0.44230769, 0.68852459, 0.3877551 ],
[0.23076923, 0.68852459, 0.97959184],
[0.69230769, 0.70491803, 0.23469388],
[0.19230769, 0.70491803, 0.60367347],
[0.44230769, 0.72131148, 0.16326531],
[0.34615385, 0.72131148, 0.85714286],
[0.30769231, 0.72131148, 0.2244898 ],
[0.26923077, 0.72131148, 0.69387755],
[0.28846154, 0.80327869, 0.07142857],
[0.38461538, 0.80327869, 0.91836735],
[0.55769231, 0.86065574, 0.15306122],
[0.32692308, 0.86065574, 0.79591837],
```

assignment_4 (1).ipynb

colab.research.google.com/drive/1nqER8GNUO6Whr2CbW_X3F3bqJI3Od3l

File Edit View Insert Runtime Tools Help Last saved at October 27

+ Code + Text

```
[ ] [0.38461538, 0.80327869, 0.91836735],
      [0.55769231, 0.86065574, 0.15306122],
      [0.32692308, 0.86065574, 0.79591837],
      [0.51923077, 0.90983607, 0.2755102 ],
      [0.26923077, 0.90983607, 0.74489796],
      [0.26923077, 1.      , 0.17346939],
      [0.23076923, 1.      , 0.83673469]]]

y=data_main['Spending Score (1-100)']

[ ] x=data_main.drop(columns=['Spending Score (1-100)', 'CustomerID', 'Gender_Female', 'Gender_Male'], axis=1)

x

Age Annual Income (k$)
0    19             15
1    21             15
2    20             16
3    23             16
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:42 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

Code Text

```
[ ]
```

3	23	16
4	31	17
...
195	35	120
196	45	126
197	32	126
198	32	137
199	30	137

200 rows x 2 columns

```
[ ] x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

```
[ ]
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:42 AM 10/26/2022

assignment_4 (1).ipynb

File Edit View Insert Runtime Tools Help Last saved at October 27

Code Text

```
[ ] x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

```
[ ]
```

```
[ ]
```

```
[ ]
```

assignment_4 (1)...ipynb assignment_4 (1).ipynb Assignment2K.docx Show all

11:42 AM 10/26/2022