

Data Mining Model Exam

SET – 5

Name: Akash Sampath

Roll: 18BCS009

1. Download a sample dataset from any Repository. List the attributes and its type in a word Doc.

Dataset/ Repository taken:

(government-finance-statistics-central-government-year-ended-june-2020-csv.csv)

```
In [22]: data=pd.read_csv("government-finance-statistics-central-government-year-ended-june-2020-csv.csv")
```

```
In [23]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2012 entries, 0 to 2011
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   Series_reference      2012 non-null  object 
 1   Period                2012 non-null  float64
 2   Data_value            2012 non-null  int64  
 3   STATUS                2012 non-null  object 
 4   UNITS                 2012 non-null  object 
 5   MAGNTUDE              2012 non-null  int64  
 6   Subject               2012 non-null  object 
 7   Group                 2012 non-null  object 
 8   Series_title_1        2012 non-null  object 
 9   Series_title_2        1160 non-null  object 
10   Series_title_3        560 non-null   object 
11   Series_title_4        366 non-null   object 
12   Series_title_5        0 non-null     float64
dtypes: float64(2), int64(2), object(9)
memory usage: 204.5+ KB
```



In [29]: data=pd.read_csv("government-finance-statistics-central-government-year-ended-june-2020-csv.csv")

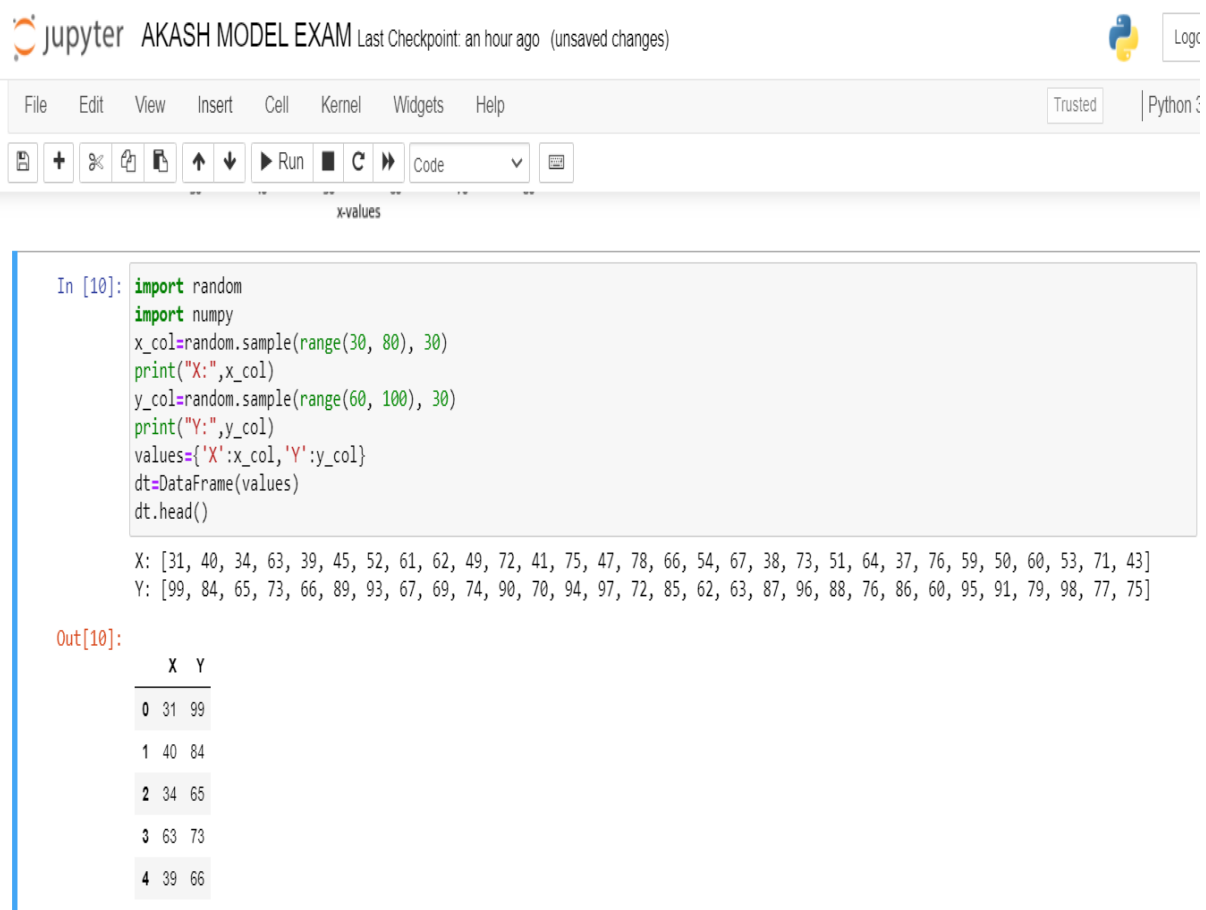
In [27]: data.head()

Out[27]:

	Series_reference	Period	Data_value	STATUS	UNITS	MAGNTUDE	Subject	Group	Series_title_1	Series_title_2	Series_title_3	Series_title_4
0	GFSA.SCS01G01Z90	2009.06	2903	FINAL	Dollars	6	Government Financial Statistics - GFS	Central Government, Operating Statement, Net B...	Net operating balance	NaN	NaN	NaN
1	GFSA.SCS01G01Z90	2010.06	-3564	FINAL	Dollars	6	Government Financial Statistics - GFS	Central Government, Operating Statement, Net B...	Net operating balance	NaN	NaN	NaN
2	GFSA.SCS01G01Z90	2011.06	-13093	FINAL	Dollars	6	Government Financial Statistics - GFS	Central Government, Operating Statement, Net B...	Net operating balance	NaN	NaN	NaN
3	GFSA.SCS01G01Z90	2012.06	-3915	FINAL	Dollars	6	Government Financial Statistics - GFS	Central Government, Operating Statement, Net B...	Net operating balance	NaN	NaN	NaN
4	GFSA.SCS01G01Z90	2013.06	-1191	FINAL	Dollars	6	Government Financial Statistics - GFS	Central Government, Operating Statement, Net B...	Net operating balance	NaN	NaN	NaN

2. Create a random dataset of 30 elements with x and y variables using random function between 30 to 80 integers for x and 60 to 100 integers for y. Apply K- means clustering to cluster the data into 2 clusters. Plot the graph and display the result. Use Tkinter GUI to Display the Results.

- ✓ Creating a random dataset of 30 elements with x and y variables using random function between 30 to 80 integers for x and 60 to 100 integers for y.



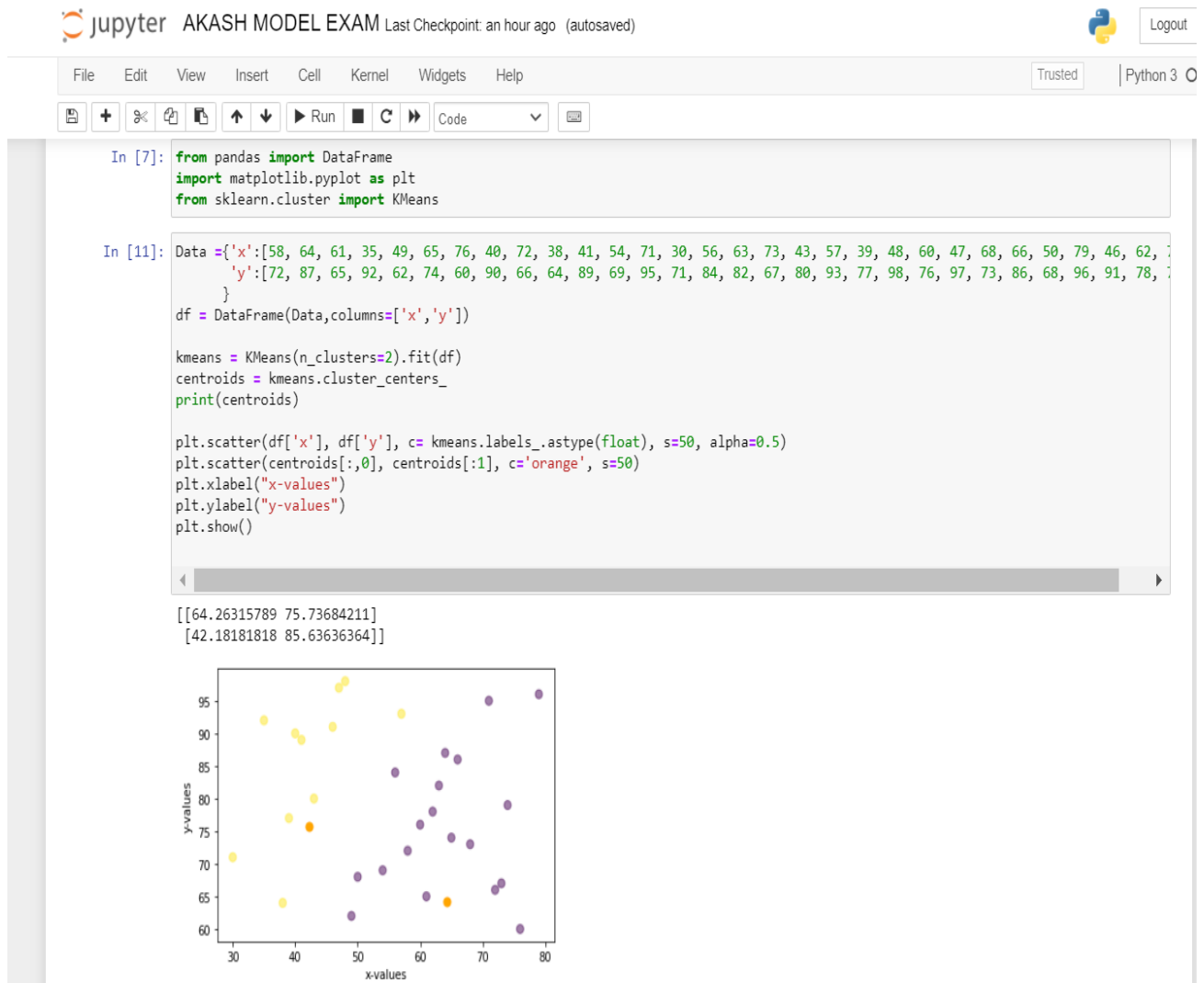
The image shows a Jupyter Notebook interface. At the top, the header includes the Jupyter logo, the text "AKASH MODEL EXAM", and "Last Checkpoint: an hour ago (unsaved changes)". On the right, there are Python and Jupyter logos and a "Log" button. Below the header is a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". To the right of the menu bar are "Trusted" and "Python 3" buttons. Below the menu bar is a toolbar with icons for saving, adding, deleting, and running cells, as well as a "Code" dropdown menu. The main area of the notebook shows a code cell with the following code:

```
In [10]: import random
import numpy
x_col=random.sample(range(30, 80), 30)
print("X:",x_col)
y_col=random.sample(range(60, 100), 30)
print("Y:",y_col)
values={'X':x_col,'Y':y_col}
dt=DataFrame(values)
dt.head()
```

Below the code cell, the output is displayed. It shows the values of X and Y as lists. X is [31, 40, 34, 63, 39, 45, 52, 61, 62, 49, 72, 41, 75, 47, 78, 66, 54, 67, 38, 73, 51, 64, 37, 76, 59, 50, 60, 53, 71, 43] and Y is [99, 84, 65, 73, 66, 89, 93, 67, 69, 74, 90, 70, 94, 97, 72, 85, 62, 63, 87, 96, 88, 76, 86, 60, 95, 91, 79, 98, 77, 75]. Below the output, the text "Out[10]:" is followed by a table with two columns, X and Y, showing the first five rows of the dataset.

	X	Y
0	31	99
1	40	84
2	34	65
3	63	73
4	39	66

- ✓ Apply K- means clustering to cluster the data into 2 clusters. Plot the graph and display the result.



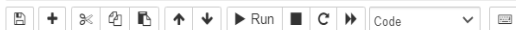
- ✓ Use Tkinter GUI to Display the Results.



```
In [ ]: import tkinter as tk
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg
root = tk.Tk()
canvas1 = tk.Canvas(root, width = 100, height = 100)
canvas1.pack()
label1 = tk.Label(root, text=centroids, justify = 'center')
canvas1.create_window(10, 90, window=label1)
figure1 = plt.Figure(figsize=(6,5), dpi=100)
ax1 = figure1.add_subplot(111)
ax1.scatter(df['x'], df['y'], c= kmeans.labels_.astype(float), s= 40, alpha = 0.5)
ax1.scatter(centroids[:, 0], centroids[:, 1], c='orange', s= 40)
scatter1 = FigureCanvasTkAgg(figure1, root)
scatter1.get_tk_widget().pack(side=tk.LEFT, fill=tk.BOTH)

root.mainloop()
```

Output: - Tinker GUI



```
In [*]: import tkinter as tk
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg
root = tk.Tk()
canvas1 = tk.Canvas(root, width = 100, height = 100)
canvas1.pack()
label1 = tk.Label(root, text=centroids, justify = 'center')
canvas1.create_window(10, 90, window=label1)
figure1 = plt.Figure(figsize=(6,5), dpi=100)
ax1 = figure1.add_subplot(111)
ax1.scatter(df['x'], df['y'], c= kmeans.labels_.astype(float), s= 40, alpha = 0.5)
ax1.scatter(centroids[:, 0], centroids[:, 1], c='orange', s= 40)
scatter1 = FigureCanvasTkAgg(figure1, root)
scatter1.get_tk_widget().pack(side=tk.LEFT, fill=tk.BOTH)

root.mainloop()
```

```
In [21]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [22]: data=pd.read_csv("government-finance-statistics-central-government-year-ended-june-2011")
```

```
In [23]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2012 entries, 0 to 2011
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Series_reference     2012 non-null   object
1   Period              2012 non-null   float64
```

