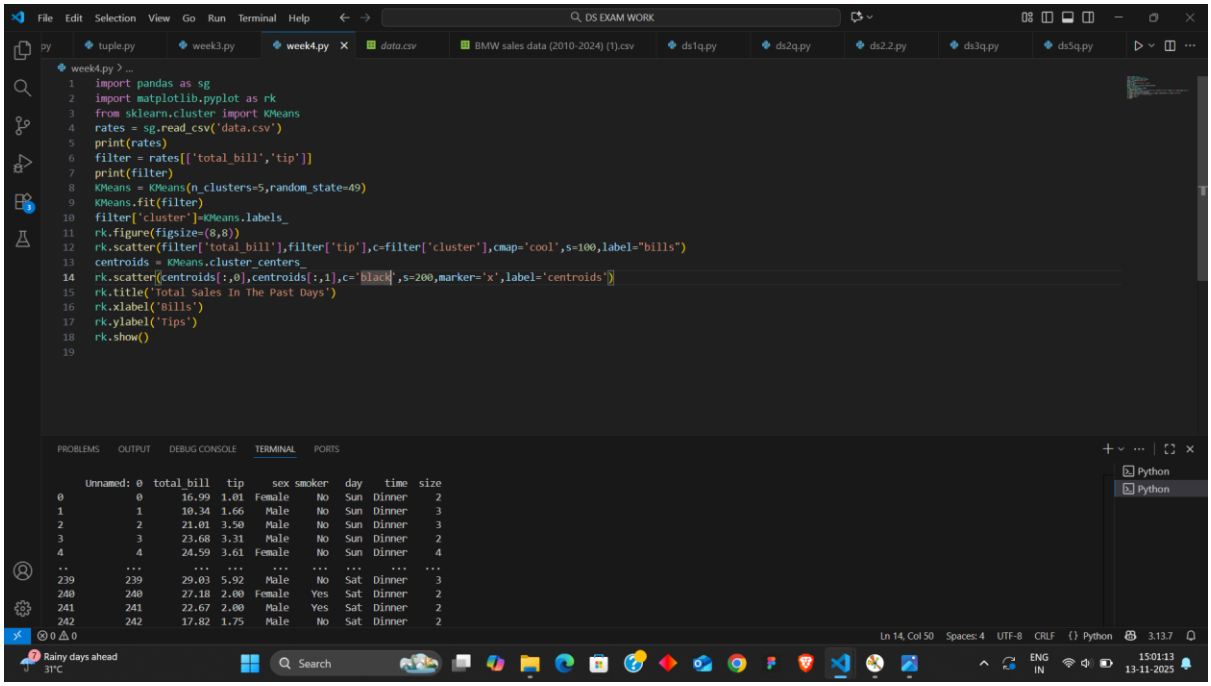
# DATA SCIENCE WEEK IV

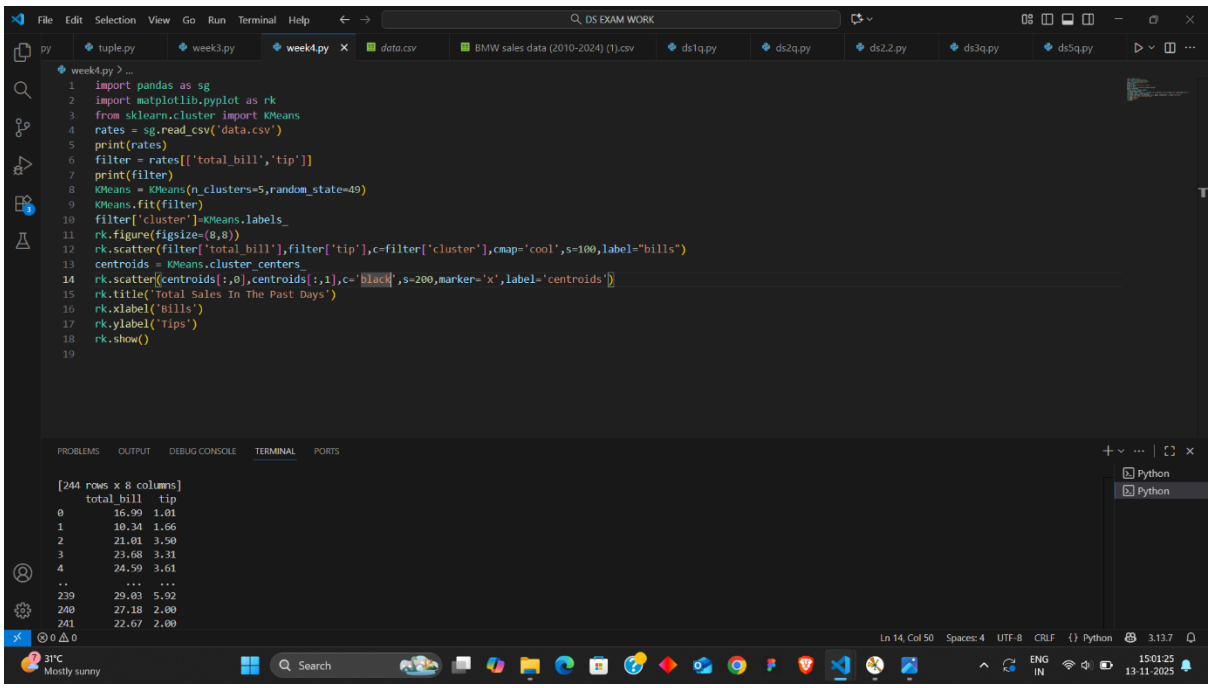
**BY N GOWTHAM KARTHIK**

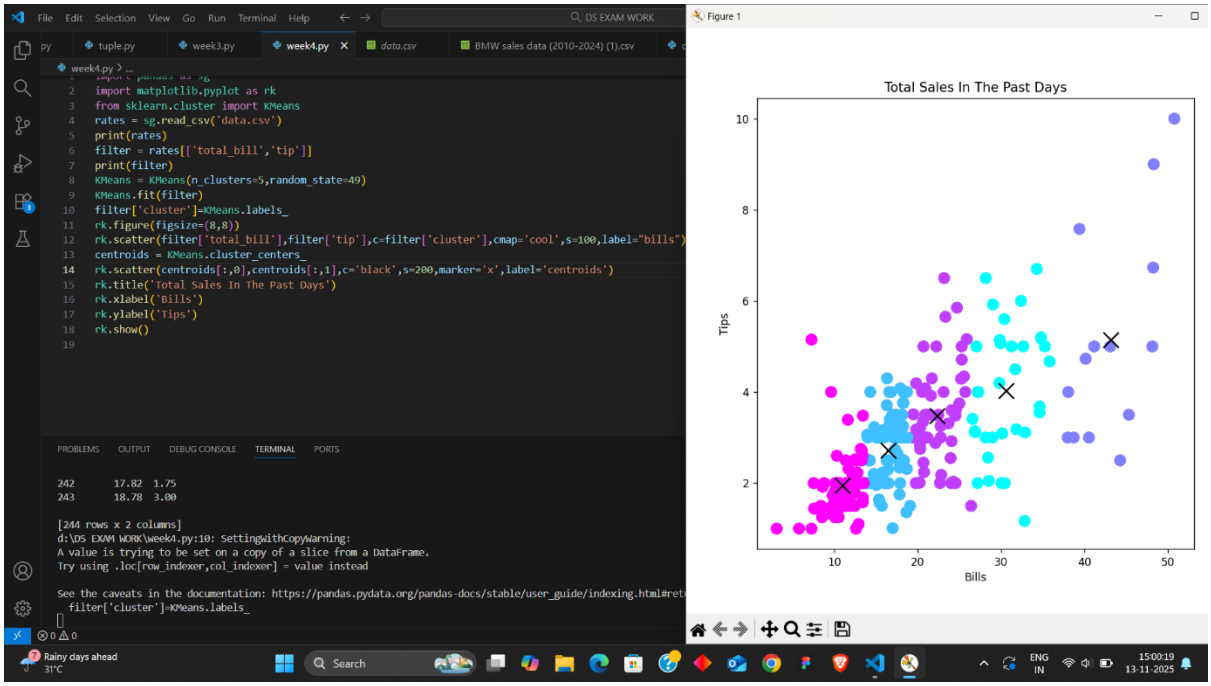
**PROGRAM:**

```
import pandas as sg
import matplotlib.pyplot as rk
from sklearn.cluster import KMeans
rates = sg.read_csv('data.csv')
print(rates)
filter = rates[['total_bill','tip']]
print(filter)
KMeans = KMeans(n_clusters=5,random_state=49)
KMeans.fit(filter)
filter['cluster']=KMeans.labels_
rk.figure(figsize=(8,8))
rk.scatter(filter['total_bill'],filter['tip'],c=filter['cluster'],cmap='cool',s=
100,label="bills")
centroids = KMeans.cluster_centers_
rk.scatter(centroids[:,0],centroids[:,1],c='black',s=200,marker='x',label
='centroids')
rk.title('Total Sales In The Past Days')
rk.xlabel('Bills')
rk.ylabel('Tips')
rk.show()
```

**SCREENSHOTS:**


File Edit Selection View Go Run Terminal Help
DS EXAM WORK
tuple.py
week3.py
week4.py
data.csv
BMW sales data (2010-2024) (1).csv
ds1q.py
ds2q.py
ds2.2.py
ds3q.py
ds5q.py
week4.py > ...
1 import pandas as sg
2 import matplotlib.pyplot as rk
3 from sklearn.cluster import KMeans
4 rates = sg.read_csv('data.csv')
5 print(rates)
6 filter = rates[['total_bill','tip']]
7 print(filter)
8 KMeans = KMeans(n_clusters=5,random_state=49)
9 KMeans.fit(filter)
10 filter['cluster']=KMeans.labels_
11 rk.figure(figsize=(8,8))
12 rk.scatter(filter['total_bill'],filter['tip'],c=filter['cluster'],cmap='cool',s=100,label="bills")
13 centroids = KMeans.cluster_centers_
14 rk.scatter(centroids[:,0],centroids[:,1],c='black',s=200,marker='x',label='centroids')
15 rk.title('Total Sales In The Past Days')
16 rk.xlabel('Bills')
17 rk.ylabel('Tips')
18 rk.show()
19
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Unnamed: 0 total_bill tip sex smoker day time size
0 0 16.99 1.01 Female No Sun Dinner 2
1 1 10.34 1.66 Male No Sun Dinner 3
2 2 21.01 3.50 Male No Sun Dinner 3
3 3 23.68 3.31 Male No Sun Dinner 2
4 4 24.59 3.61 Female No Sun Dinner 4
.. ... ... ... ... ... ... ... ...
239 239 29.03 5.92 Male No Sat Dinner 3
240 240 27.18 2.00 Female Yes Sat Dinner 2
241 241 22.67 2.00 Male Yes Sat Dinner 2
242 242 17.82 1.75 Male No Sat Dinner 2
Python
Python
Ln 14, Col 50 Spaces: 4 UTF-8 CRLF {} Python 3.13.7
Rainy days ahead
31°C
Search
ENG
IN
15:01:13
13-11-2025


File Edit Selection View Go Run Terminal Help
DS EXAM WORK
week4.py > ...
1 import pandas as sg
2 import matplotlib.pyplot as rk
3 from sklearn.cluster import KMeans
4 rates = sg.read_csv('data.csv')
5 print(rates)
6 filter = rates[['total_bill','tip']]
7 print(filter)
8 KMeans = KMeans(n_clusters=5,random_state=49)
9 KMeans.fit(filter)
10 filter['cluster']=KMeans.labels_
11 rk.figure(figsize=(8,8))
12 rk.scatter(filter['total_bill'],filter['tip'],c=filter['cluster'],cmap='cool',s=100,label="bills")
13 centroids = KMeans.cluster_centers_
14 rk.scatter(centroids[:,0],centroids[:,1],c='black',s=200,marker='x',label='centroids')
15 rk.title('Total Sales In The Past Days')
16 rk.xlabel('Bills')
17 rk.ylabel('Tips')
18 rk.show()
19
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
[244 rows x 8 columns]
total_bill tip
0 16.99 1.01
1 10.34 1.66
2 21.01 3.50
3 23.68 3.31
4 24.59 3.61
.. ... ...
239 29.03 5.92
240 27.18 2.00
241 22.67 2.00
Python
Python
Ln 14, Col 50 Spaces: 4 UTF-8 CRLF {} Python 3.13.7
31°C
Mostly sunny
Search
ENG
IN
15:01:25
13-11-2025


File Edit Selection View Go Run Terminal Help
DS EXAM WORK
tuple.py
week3.py
week4.py
data.csv
BMW sales data (2010-2024) (1).csv
week4.py > ...
import matplotlib.pyplot as rk
from sklearn.cluster import KMeans
rates = sg.read_csv('data.csv')
print(rates)
filter = rates[['total_bill','tip']]
print(filter)
KMeans = KMeans(n_clusters=5,random_state=49)
KMeans.fit(filter)
filter['cluster']=KMeans.labels_
rk.figure(figsize=(8,8))
rk.scatter(filter['total_bill'],filter['tip'],c=filter['cluster'],cmap='cool',s=100,label="bills")
centroids = KMeans.cluster_centers_
rk.scatter(centroids[:,0],centroids[:,1],c='black',s=200,marker='x',label='centroids')
rk.title('Total Sales In The Past Days')
rk.xlabel('Bills')
rk.ylabel('Tips')
rk.show()
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
242 17.82 1.75
243 18.78 3.00
[244 rows x 2 columns]
d:\DS EXAM WORK\week4.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ret
filter['cluster']=KMeans.labels_
Figure 1
Total Sales In The Past Days
Tips
Bills
10
8
6
4
2
10
20
30
40
50
Rainy days ahead
31°C
Search
ENG
IN
15:00:19
13-11-2025