assignment1

April 15, 2024

0.1 ASSIGNMENT # 1

0.2 Humna Khan

[ ]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

df = pd.read_csv("/home/shafeenkhan/Documents/My-all-programs--/Semester-4/
 ↪Aritificial Intelligence/Lab-Assignments/Assignment 01/Humna/total_data_na.
 ↪csv")
# print(df)
df = df.dropna()
# print(df)

# df = df.fillna('-', inplace=True)
print(df)
# df
```

```
               PLAYER  Mat.x  Inns.x  NO  Runs.x  HS  Avg.x   BF    SR.x  \
0         Aaron Finch     10       9   1     134  46  16.75  100  134.00
1      AB de Villiers     12      11   2     480  90  53.33  275  174.54
2     Abhishek Sharma      3       3   2      63  46     63   33  190.90
3      Ajinkya Rahane     15      14   1     370  65  28.46  313  118.21
4          Alex Hales      6       6   0     148  45  24.66  118  125.42
..                …    …     …  ..     …  ..    …  …     …
138     Siddarth Kaul      0       0   0       0   0      0    0    0.00
139       Trent Boult      0       0   0       0   0      0    0    0.00
140       Umesh Yadav      0       0   0       0   0      0    0    0.00
141        Vi0y Kumar      0       0   0       0   0      0    0    0.00
142  Yuzvendra Chahal      0       0   0       0   0      0    0    0.00

     X100  …    Ov  Runs.y  Wkts  BBI  Avg.y   Econ   SR.y  X4w  X5w  y
0       0  …   0.0       0     0    0      0   0.00      0    0    0  0
1       0  …   0.0       0     0    0      0   0.00      0    0    0  0
```

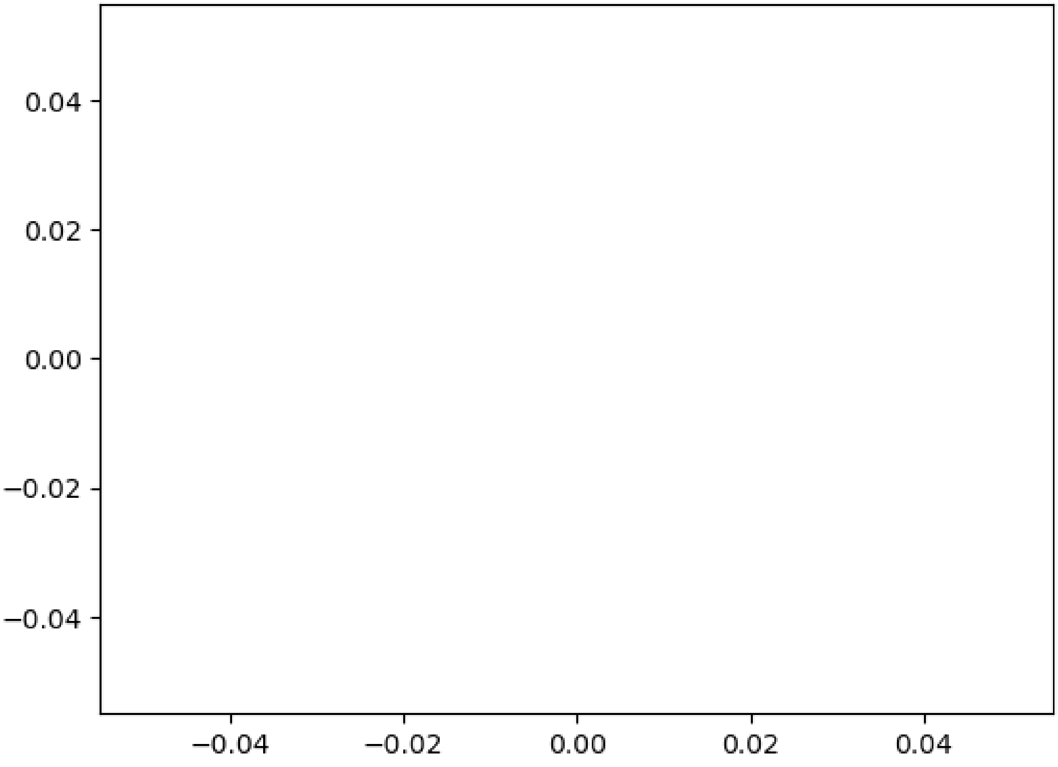
```
2       0  …   0.0       0     0    0      0   0.00      0     0    0  0
3       0  …   0.0       0     0    0      0   0.00      0     0    0  0
4       0  …   0.0       0     0    0      0   0.00      0     0    0  0
..    …  …   …     …    …  …    …    …    …  …   …  ..
138     0  …  66.0     547    21    0  26.04   8.28  18.85    0    0  0
139     0  …  52.4     466    18    0  25.88   8.84  17.55    0    0  0
140     0  …  53.1     418    20    0   20.9   7.86  15.95    0    0  0
141     0  …   3.5      65     2    0   32.5  16.95   11.5    0    0  0
142     0  …  50.0     363    12    0  30.25   7.26     25    0    0  0

[143 rows x 25 columns]
```

```
[ ]: features = df[['Runs.x', 'HS', 'Avg.x', 'BF', 'SR.x', 'X100', 'X50', 'Runs.y',␣
      ↪'Wkts', 'BBI', 'Avg.y', 'Econ', 'SR.y']]
     features = features.replace('-', np.nan)

     imp = SimpleImputer(strategy="mean")
     features = pd.DataFrame(imp.fit_transform(features), columns = features.columns)

     stScaler = StandardScaler()
     scaledFeatures = stScaler.fit_transform(features)
     print(scaledFeatures)

     kmean = KMeans(n_clusters = 3, random_state=  42, n_init = 1000)
     kmean.fit(scaledFeatures)
     centroid = kmean.cluster_centers_
     print("Centroids: ", centroid)

     label = kmean.labels_
     print("Lables:", label)

     iClusters = kmean.fit_predict(scaledFeatures)
```

```
[[ 0.00943771  0.40323538 -0.03937894 … -1.08126244 -1.3206071
  -1.1344344 ]
 [ 1.98807763  1.78437691  2.14671767 … -1.08126244 -1.3206071
  -1.1344344 ]
 [-0.39658378  0.40323538  2.72461691 … -1.08126244 -1.3206071
  -1.1344344 ]
 …
 [-0.75685636 -1.04068531 -1.04039365 … -0.04270068  0.36491249
   0.11820146]
 [-0.75685636 -1.04068531 -1.04039365 …  0.53372595  2.31419659
  -0.23128002]
 [-0.75685636 -1.04068531 -1.04039365 …  0.42191906  0.23624687
   0.82894469]]
Centroids:  [[-0.50046143 -0.54017673 -0.45058281 -0.51949627 -0.36357766
-0.1600461
```

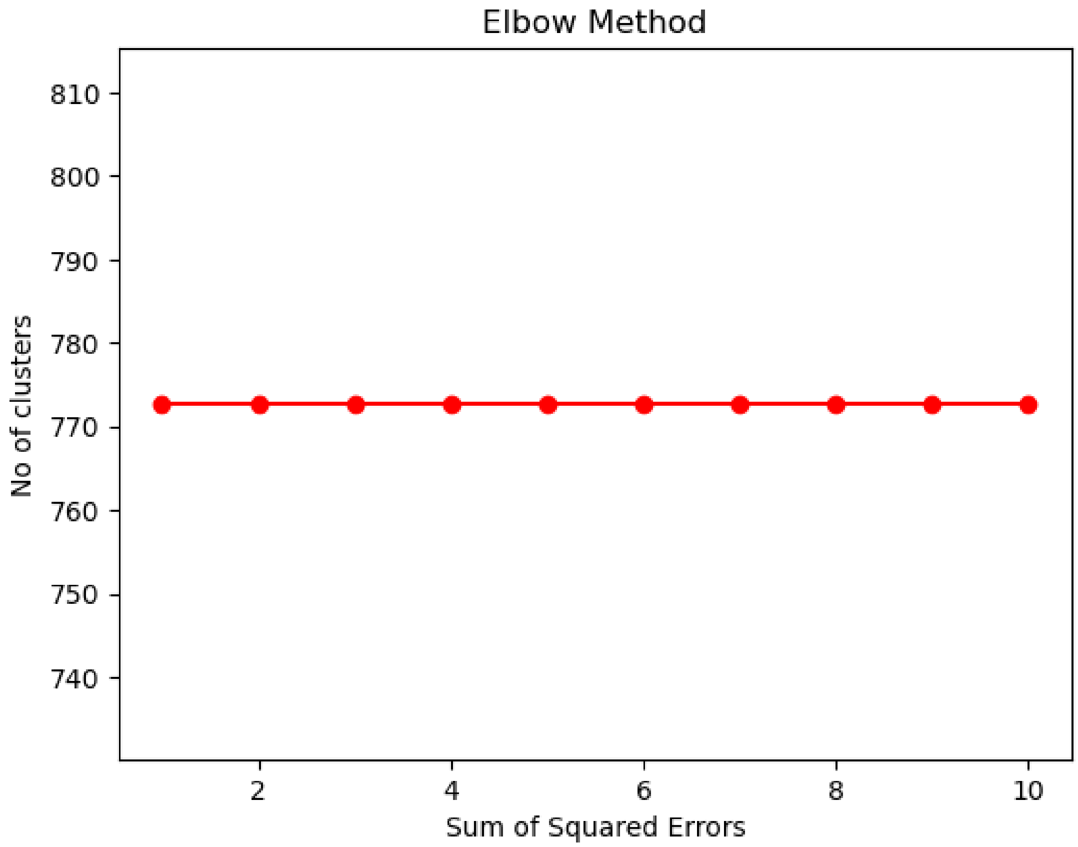
```
  -0.43622554  0.48857749  0.42680737  0.          0.57099546  0.68530401
   0.59020988]
 [ 0.21285693  0.53417477  0.34680174  0.28723146  0.55603584 -0.1600461
   0.00553937 -0.88183644 -0.75698935  0.         -1.01879256 -1.19525003
  -1.04636282]
 [ 2.24070575  1.79176269  1.69793966  2.18947406  0.80143018  1.18622403
   2.32368507 -0.79978095 -0.72616725  0.         -0.95899101 -1.20758319
  -1.00508236]]
Lables: [1 2 1 1 1 2 0 0 0 0 0 0 1 0 2 2 0 0 0 1 0 1 0 1 0 0 2 0 1 1 1 0 0 0 0 1
1
 1 0 0 2 0 2 1 1 1 0 0 2 0 1 1 1 0 0 1 0 0 0 0 0 2 1 1 0 1 1 0 1 0 0 0 1 2
 1 1 1 2 1 0 2 0 2 0 1 0 2 1 0 0 2 2 0 0 0 2 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

```
[ ]: sumofsquarede = []
     for k in range(1, 11):
         plt.plot(numClusters=k, randomState = 0, n_init = 100, max_iterations =␣
      ↪1000)
         kmean.fit(scaledFeatures)
         sumofsquarede.append(kmean.inertia_)
```



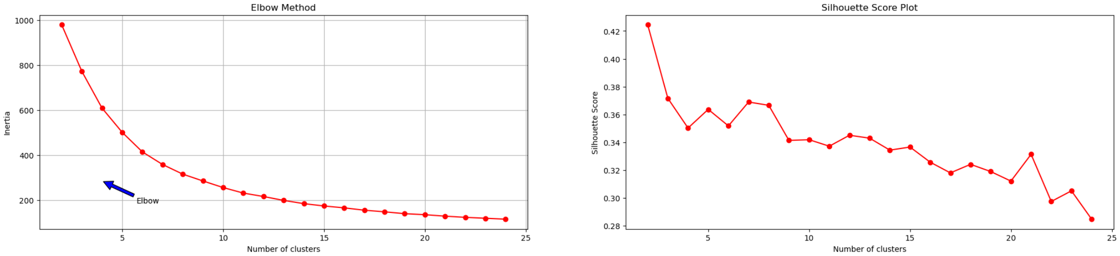
[ ]:

```
num_clusters = range(1, 11)
plt.plot(num_clusters, sumofsquarede, marker = 'o', color = "red")
# plt.annotate('Elbow', xy = (4,200), xytext(4, 300), arrowprops =␣
 ↪{"arrowstyle": "->"})
plt.title("Elbow Method")
plt.xlabel("Sum of Squared Errors")
plt.ylabel("No of clusters")
plt.show()
```



[ ]:

```
silhouetteScores = []
sumofsquarede = []

for k in range(2, 25):
    kmean = KMeans(n_clusters=k, random_state=0, n_init=100, max_iter=1000)
    kmean.fit(scaledFeatures)
    score = silhouette_score(scaledFeatures, kmean.labels_)
    silhouetteScores.append(score)
    sumofsquared = kmean.inertia_
    sumofsquarede.append(sumofsquared)

fig, (a1, a2) = plt.subplots(1, 2, figsize=(25, 5))

num_clusters = range(2, 25)
a1.plot(num_clusters, sumofsquarede, marker='o', color="red")
a1.grid(True)
a1.annotate("Elbow", xy=(4, sumofsquarede[7]), xytext=(0.1, 0.2), textcoords="figure fraction", arrowprops=dict(facecolor="blue", shrink=0.04))
a1.set_xlabel("Number of clusters")
a1.set_ylabel("Inertia")
a1.set_title("Elbow Method")

a2.plot(range(2, 25), silhouetteScores, marker="o", color="red")
a2.set_xlabel("Number of clusters")
a2.set_ylabel("Silhouette Score")
a2.set_title("Silhouette Score Plot")
plt.show()
```



[ ]:
```
kmean = KMeans(n_clusters=  3, random_state=  42, n_init = 100)
kmean.fit(scaledFeatures)

centroid = kmean.cluster_centers_
iClusters = kmean.fit_predict(scaledFeatures)
features["cluster"] = iClusters
```

[ ]:
```
topOrder = 2
middleOrder = 4
allRounders = 2
ballers = 4

tOrder = features[features["cluster"]==0]
mOrder = features[features["cluster"]==1]
all_rounders = features[features["cluster"]==2]
```

```
baller = features[features["cluster"]==3]

print("Top Order Batsmen:")
print(df.loc[tOrder.index])
print("Middle Order Batsmen:")
print()
print(df.loc[mOrder.index])
print("All Rounders:")
print()
print(df.loc[all_rounders.index])
print("Bowlers:")
print(df.loc[baller.index])
print()
```

```
Top Order Batsmen:
               PLAYER  Mat.x  Inns.x  NO  Runs.x  HS  Avg.x   BF    SR.x  \
6       Andre Russell     16      14   3     316  88  28.72  171  184.79
7          Andrew Tye     14       8   2      32  14   5.33   38   84.21
8          Axar Patel      9       8   2      80  19  13.33   69  115.94
9         Ben Cutting      9       6   2      96  37     24   58  165.51
10         Ben Stokes     13      13   1     196  45  16.33  161  121.73
..                 …    …     …  ..     …  ..    …  …     …
138     Siddarth Kaul      0       0   0       0   0      0    0    0.00
139       Trent Boult      0       0   0       0   0      0    0    0.00
140       Umesh Yadav      0       0   0       0   0      0    0    0.00
141        Vi0y Kumar      0       0   0       0   0      0    0    0.00
142  Yuzvendra Chahal      0       0   0       0   0      0    0    0.00

     X100  …    Ov  Runs.y  Wkts  BBI  Avg.y   Econ   SR.y  X4w  X5w  y
6       0  …  37.5     355    13    0   27.3   9.38  17.46    0    0  0
7       0  …  56.0     448    24    0  18.66   8.00     14    3    0  0
8       0  …  26.0     218     3    0  72.66   8.38     52    0    0  0
9       0  …  17.0     168     2    0     84   9.88     51    0    0  0
10      0  …  37.0     303     8    0  37.87   8.18  27.75    0    0  0
..   …  …   …     …   …  …    …    …    …  …  …  ..
138     0  …  66.0     547    21    0  26.04   8.28  18.85    0    0  0
139     0  …  52.4     466    18    0  25.88   8.84  17.55    0    0  0
140     0  …  53.1     418    20    0   20.9   7.86  15.95    0    0  0
141     0  …   3.5      65     2    0   32.5  16.95   11.5    0    0  0
142     0  …  50.0     363    12    0  30.25   7.26     25    0    0  0

[91 rows x 25 columns]
Middle Order Batsmen:

               PLAYER  Mat.x  Inns.x  NO  Runs.x  HS  Avg.x   BF    SR.x  \
0         Aaron Finch     10       9   1     134  46  16.75  100  134.00
2     Abhishek Sharma      3       3   2      63  46     63   33  190.90
3      Ajinkya Rahane     15      14   1     370  65  28.46  313  118.21
```

```
4          Alex Hales      6       6   0     148  45  24.66  118  125.42
12   Brendon McCullum      6       6   0     127  43  21.16   88  144.31
19        Colin Munro      5       5   0      63  33   12.6   41  153.65
21       D'Arcy Short      7       7   0     115  44  16.42   99  116.16
23       David Miller      3       3   1      74  26     37   64  115.62
28         Evin Lewis     13      13   0     382  65  29.38  276  138.40
29     Faf du Plessis      6       6   1     162  67   32.4  129  125.58
30     Gautam Gambhir      6       5   0      85  55     17   88   96.59
35   Heinrich Klaasen      4       4   1      57  32     19   47  121.27
36       Ishan Kishan     14      12   0     275  62  22.91  184  149.45
37          Jason Roy      5       5   1     120  91     30   94  127.65
43          Karun 0ir     13      12   0     301  54  25.08  221  136.19
44       Kedar Jadhav      1       1   1      24  24      -   22  109.09
45     Kieron Pollard      9       8   1     133  50     19  100  133.00
50         Ma0n Vohra      4       4   0      55  45  13.75   47  117.02
51      Mandeep Singh     14      13   3     252  47   25.2  186  135.48
52      Manish Pandey     15      13   2     284  62  25.81  246  115.44
55     Mayank Agarwal     11      11   1     120  30     12   94  127.65
62         Nitish Ra0     15      15   2     304  59  23.38  232  131.03
63      Parthiv Patel      6       6   1     153  53   30.6  109  140.36
65       Prithvi Shaw      9       9   0     245  65  27.22  160  153.12
66    Quinton de Kock      8       8   0     201  53  25.12  162  124.07
68     Rahul Tripathi     12      12   3     226  80  25.11  167  135.32
72        Rinku Singh      4       4   0      29  16   7.25   31   93.54
74      Robin Uthappa     16      16   0     351  54  21.93  265  132.45
75       Rohit Sharma     14      14   2     286  94  23.83  215  133.02
76       Sam Billings     10       8   0     108  56   13.5   78  138.46
78      Sarfaraz Khan      7       6   1      51  22   10.2   41  124.39
84  Shreevats Goswami      6       3   0      52  35  17.33   40  130.00
87       Shubman Gill     13      11   5     203  57  33.83  139  146.04
97    Wriddhiman Saha     11      10   2     122  35  15.25  102  119.60
98       Yusuf Pathan     15      13   4     260  45  28.88  200  130.00

    X100  …    Ov  Runs.y  Wkts  BBI  Avg.y  Econ  SR.y  X4w  X5w  y
0      0  …   0.0       0     0    0      0  0.00     0    0    0  0
2      0  …   0.0       0     0    0      0  0.00     0    0    0  0
3      0  …   0.0       0     0    0      0  0.00     0    0    0  0
4      0  …   0.0       0     0    0      0  0.00     0    0    0  0
12     0  …   0.0       0     0    0      0  0.00     0    0    0  0
19     0  …   0.0       0     0    0      0  0.00     0    0    0  0
21     0  …   3.0      19     1    0     19  6.33    18    0    0  0
23     0  …   0.0       0     0    0      0  0.00     0    0    0  0
28     0  …   0.0       0     0    0      0  0.00     0    0    0  0
29     0  …   0.0       0     0    0      0  0.00     0    0    0  0
30     0  …   0.0       0     0    0      0  0.00     0    0    0  0
35     0  …   0.0       0     0    0      0  0.00     0    0    0  0
36     0  …   0.0       0     0    0      0  0.00     0    0    0  0
37     0  …   0.0       0     0    0      0  0.00     0    0    0  0
```

```
43      0  …  0.0       0     0    0       0  0.00     0    0    0  0
44      0  …  0.0       0     0    0       0  0.00     0    0    0  0
45      0  …  0.0       0     0    0       0  0.00     0    0    0  0
50      0  …  0.0       0     0    0       0  0.00     0    0    0  0
51      0  …  0.0       0     0    0       0  0.00     0    0    0  0
52      0  …  0.0       0     0    0       0  0.00     0    0    0  0
55      0  …  0.0       0     0    0       0  0.00     0    0    0  0
62      0  …  6.1      44     4    0      11  7.13  9.25    0    0  0
63      0  …  0.0       0     0    0       0  0.00     0    0    0  0
65      0  …  0.0       0     0    0       0  0.00     0    0    0  0
66      0  …  0.0       0     0    0       0  0.00     0    0    0  0
68      0  …  0.0       0     0    0       0  0.00     0    0    0  0
72      0  …  0.0       0     0    0       0  0.00     0    0    0  0
74      0  …  0.0       0     0    0       0  0.00     0    0    0  0
75      0  …  0.0       0     0    0       0  0.00     0    0    0  0
76      0  …  0.0       0     0    0       0  0.00     0    0    0  0
78      0  …  0.0       0     0    0       0  0.00     0    0    0  0
84      0  …  0.0       0     0    0       0  0.00     0    0    0  0
87      0  …  0.0       0     0    0       0  0.00     0    0    0  0
97      0  …  0.0       0     0    0       0  0.00     0    0    0  0
98      0  …  2.0      14     1    0      14  7.00    12    0    0  0

[35 rows x 25 columns]
All Rounders:

              PLAYER  Mat.x  Inns.x  NO  Runs.x   HS  Avg.x   BF    SR.x  \
1     AB de Villiers     12      11   2     480   90  53.33  275  174.54
5      Ambati Rayudu     16      16   2     602  100     43  402  149.75
14       Chris Gayle     11      11   2     368  104  40.88  252  146.03
15        Chris Lynn     16      16   1     491   74  32.73  377  130.23
26    Dinesh Karthik     16      16   6     498   52   49.8  337  147.77
40       Jos Buttler     13      13   3     548   95   54.8  353  155.24
42  Kane Williamson      17      17   3     735   84   52.5  516  142.44
48      Lokesh Rahul     14      14   2     659   95  54.91  416  158.41
61          MS Dhoni     16      15   9     455   79  75.83  302  150.66
73      Rishabh Pant     14      14   1     684  128  52.61  394  173.60
77      Sanju Samson     15      15   1     441   92   31.5  320  137.81
80      Shane Watson     15      15   1     555  117  39.64  359  154.59
82    Shikhar Dhawan     16      16   3     497   92  38.23  363  136.91
86      Shreyas Iyer     14      14   3     411   93  37.36  310  132.58
90       Suresh Rai0     15      15   3     445   75  37.08  336  132.44
91  Suryakumar Yadav     14      14   0     512   72  36.57  384  133.33
95       Virat Kohli     14      14   3     530   92  48.18  381  139.10

    X100  …     Ov  Runs.y  Wkts  BBI  Avg.y  Econ  SR.y  X4w  X5w  y
1      0  …    0.0       0     0    0      0  0.00     0    0    0  0
5      1  …    0.0       0     0    0      0  0.00     0    0    0  0
14     1  …    0.0       0     0    0      0  0.00     0    0    0  0
```

```
15     0  …   0.0       0     0    0      0  0.00     0     0    0  0
26     0  …   0.0       0     0    0      0  0.00     0     0    0  0
40     0  …   0.0       0     0    0      0  0.00     0     0    0  0
42     0  …   0.0       0     0    0      0  0.00     0     0    0  0
48     0  …   0.0       0     0    0      0  0.00     0     0    0  0
61     0  …   0.0       0     0    0      0  0.00     0     0    0  0
73     1  …   0.0       0     0    0      0  0.00     0     0    0  0
77     0  …   0.0       0     0    0      0  0.00     0     0    0  0
80     2  …  28.0     251     6    0  41.83  8.96    28     0    0  0
82     0  …   0.0       0     0    0      0  0.00     0     0    0  0
86     0  …   0.0       0     0    0      0  0.00     0     0    0  0
90     0  …   0.0       0     0    0      0  0.00     0     0    0  0
91     0  …   0.0       0     0    0      0  0.00     0     0    0  0
95     0  …   0.0       0     0    0      0  0.00     0     0    0  0

[17 rows x 25 columns]
Bowlers:
Empty DataFrame
Columns: [PLAYER, Mat.x, Inns.x, NO, Runs.x, HS, Avg.x, BF, SR.x, X100, X50,
X4s, X6s, Mat.y, Inns.y, Ov, Runs.y, Wkts, BBI, Avg.y, Econ, SR.y, X4w, X5w, y]
Index: []

[0 rows x 25 columns]
```

[ ]:

```python
tOrder = features[features["cluster"]==0].nlargest(3, ["Runs.x", "SR.x"])
mOrder = features[features["cluster"]==1].nlargest(3, ["Runs.x"])
all_rounders = features[features["cluster"]==2].nlargest(3, ["Wkts", "Runs.x"])
baller = features[features["cluster"]==3].nlargest(3, ["Wkts", "Econ"])
```

[ ]:

```python
print("Top Order Batsmen:")
print(df.loc[tOrder.index])
print()
print("Middle Order Batsmen:")
print(df.loc[mOrder.index])
print()
print("All Rounders:")
print(df.loc[all_rounders.index])
print()
print("Bowlers:")
print(df.loc[baller.index])
```

```
Top Order Batsmen:
           PLAYER  Mat.x  Inns.x  NO  Runs.x  HS  Avg.x   BF    SR.x  X100  \
89    Sunil Orine     16      16   0     357  75  22.31  188  189.89     0
6   Andre Russell     16      14   3     316  88  28.72  171  184.79     0
33  Hardik Pandya     13      13   4     260  50  28.88  195  133.33     0
```

```
    …    Ov  Runs.y  Wkts  BBI  Avg.y  Econ   SR.y  X4w  X5w  y
89  …  61.0     467    17    0  27.47  7.65  21.52    0    0  0
6   …  37.5     355    13    0   27.3  9.38  17.46    0    0  0
33  …  42.4     381    18    0  21.16  8.92  14.22    0    0  0

[3 rows x 25 columns]

Middle Order Batsmen:
            PLAYER  Mat.x  Inns.x  NO  Runs.x  HS  Avg.x   BF    SR.x  X100  \
28      Evin Lewis     13      13   0     382  65  29.38  276  138.40     0
3   Ajinkya Rahane     15      14   1     370  65  28.46  313  118.21     0
74   Robin Uthappa     16      16   0     351  54  21.93  265  132.45     0

    …   Ov  Runs.y  Wkts  BBI  Avg.y  Econ  SR.y  X4w  X5w  y
28  …  0.0       0     0    0      0   0.0     0    0    0  0
3   …  0.0       0     0    0      0   0.0     0    0    0  0
74  …  0.0       0     0    0      0   0.0     0    0    0  0

[3 rows x 25 columns]

All Rounders:
              PLAYER  Mat.x  Inns.x  NO  Runs.x   HS  Avg.x   BF    SR.x  \
80      Shane Watson     15      15   1     555  117  39.64  359  154.59
42  Kane Williamson      17      17   3     735   84   52.5  516  142.44
73      Rishabh Pant     14      14   1     684  128  52.61  394  173.60

    X100  …    Ov  Runs.y  Wkts  BBI  Avg.y  Econ  SR.y  X4w  X5w  y
80     2  …  28.0     251     6    0  41.83  8.96    28    0    0  0
42     0  …   0.0       0     0    0      0  0.00     0    0    0  0
73     1  …   0.0       0     0    0      0  0.00     0    0    0  0

[3 rows x 25 columns]

Bowlers:
Empty DataFrame
Columns: [PLAYER, Mat.x, Inns.x, NO, Runs.x, HS, Avg.x, BF, SR.x, X100, X50,
X4s, X6s, Mat.y, Inns.y, Ov, Runs.y, Wkts, BBI, Avg.y, Econ, SR.y, X4w, X5w, y]
Index: []

[0 rows x 25 columns]
```