NAME: DANIYAL SAEED

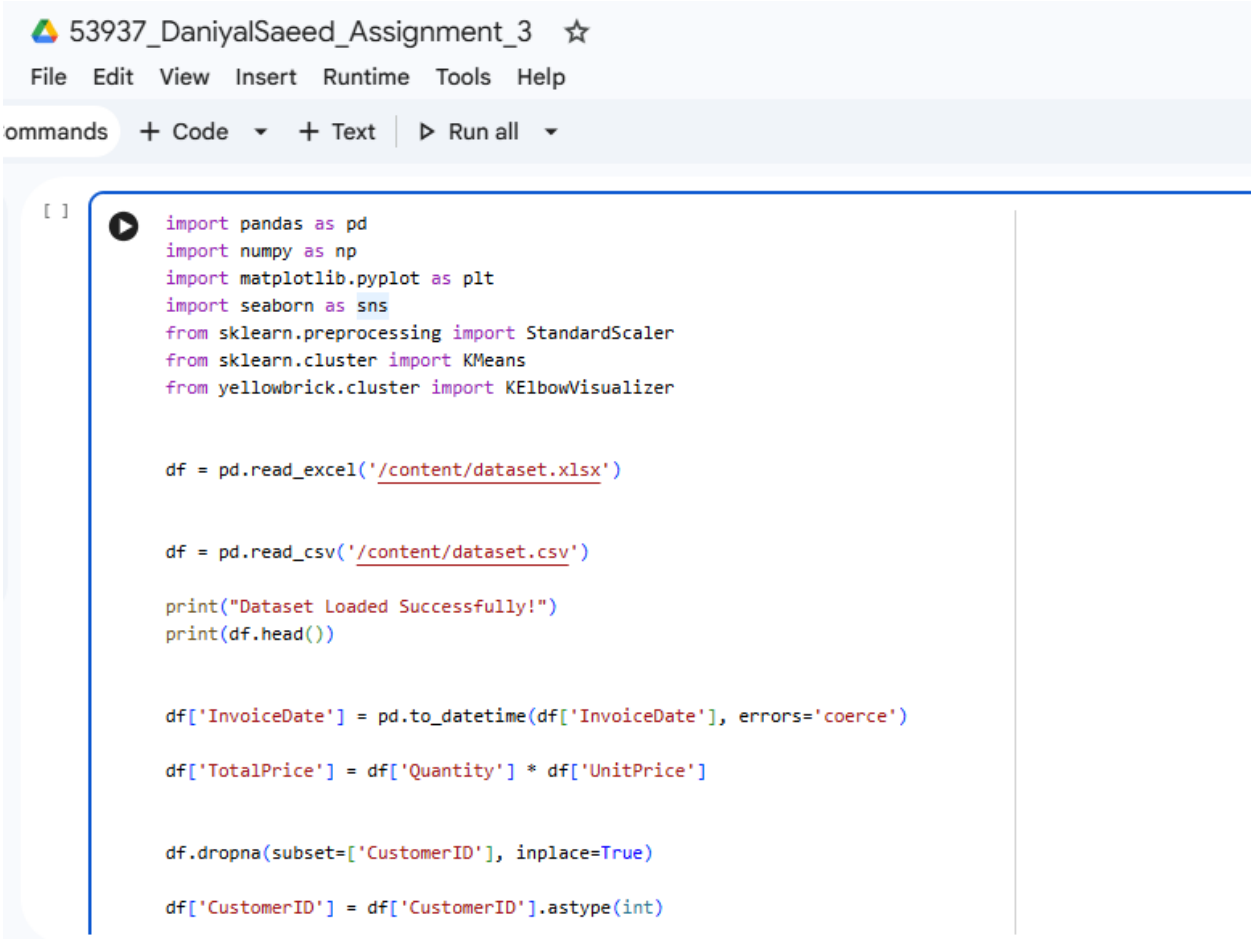
SAP ID: 53937

SECTION: BS Data Science

COURSE: Machine Learning

Assignment 3

K- Means Algorithm:



[ ]

```
df = df[~df['InvoiceNo'].astype(str).str.startswith('C')]

df = df[(df['Quantity'] > 0) & (df['UnitPrice'] > 0)]

print("\nCleaned Data Shape:", df.shape)


snapshot_date = df['InvoiceDate'].max() + pd.Timedelta(days=1)

rfm_df = df.groupby('CustomerID').agg(
    Recency=('InvoiceDate', lambda x: (snapshot_date - x.max()).days),
    Frequency=('InvoiceNo', 'nunique'),
    Monetary=('TotalPrice', 'sum')
).reset_index()

print("\nRFM Calculation Done!")
print(rfm_df.head())


rfm_data = rfm_df[['Recency', 'Frequency', 'Monetary']]
scaler = StandardScaler()
rfm_scaled = scaler.fit_transform(rfm_data)

rfm_scaled_df = pd.DataFrame(rfm_scaled, columns=rfm_data.columns)
```

[ ]

```
print("\nRunning Elbow Method...")
model = KMeans(random_state=42, n_init=10)
visualizer = KElbowVisualizer(model, k=(2, 5), metric='distortion', timings=False)

visualizer.fit(rfm_scaled)
optimal_k = visualizer.elbow_value_

if optimal_k is None:
    print("Warning: No optimal K found by Elbow method. Defaulting to 2 clusters.")
    optimal_k = 2

print(f"Optimal K found: {optimal_k}")
visualizer.show()


kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)
rfm_df['Cluster'] = kmeans.fit_predict(rfm_scaled)

rfm_scaled_df['Cluster'] = rfm_df['Cluster']

print("\nK-Means Clustering Completed!")
```

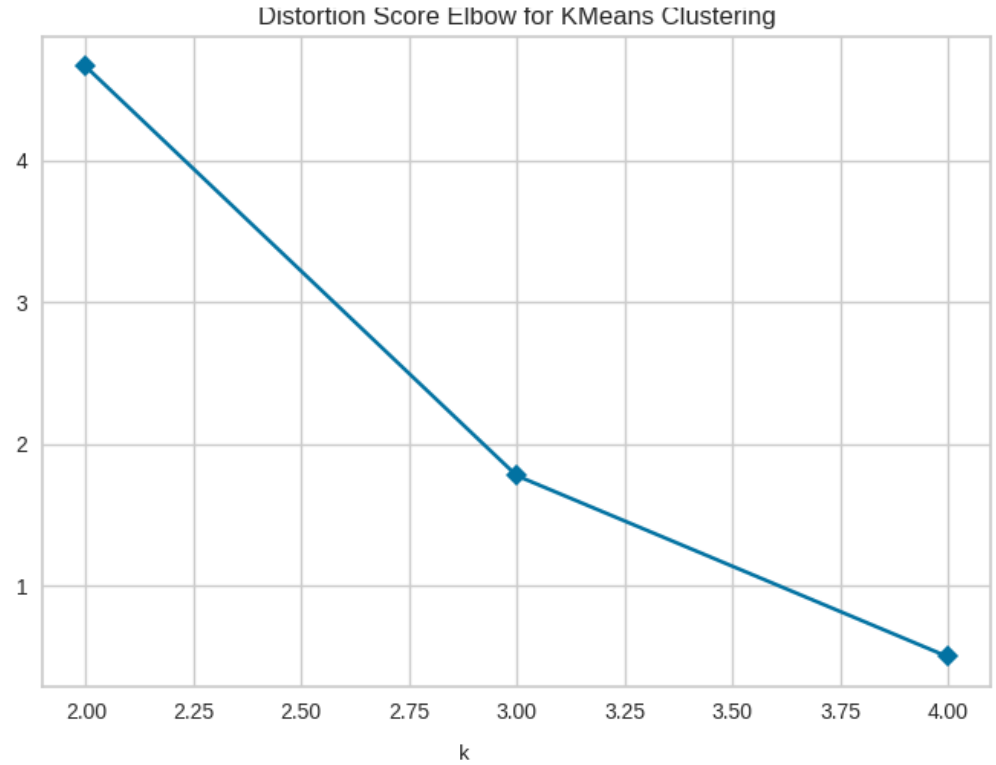
```
cluster_analysis = rfm_df.groupby('Cluster').agg({
    'Recency': 'mean',
    'Frequency': 'mean',
    'Monetary': 'mean',
    'CustomerID': 'count'
}).rename(columns={'CustomerID': 'Count'})

cluster_analysis = cluster_analysis.round(2)

print("\nCluster Analysis:")
print(cluster_analysis)


rfm_df.to_csv('rfm_clusters_output.csv', index=False)
print("\nFile Saved: rfm_clusters_output.csv")


plt.figure(figsize=(15, 5))
plt.suptitle('Cluster Profiles - RFM Comparison', fontsize=16)

plt.subplot(1, 3, 1)
sns.barplot(x=cluster_analysis.index, y='Recency', data=cluster_analysis)
plt.title("Mean Recency")

plt.subplot(1, 3, 2)
sns.barplot(x=cluster_analysis.index, y='Frequency', data=cluster_analysis)
plt.title("Mean Frequency")
```

```
plt.subplot(1, 3, 3)
sns.barplot(x=cluster_analysis.index, y='Monetary', data=cluster_analysis)
plt.title("Mean Monetary Value")

plt.show()



plt.figure(figsize=(15, 6))

plt.subplot(1, 2, 1)
sns.scatterplot(x='Recency', y='Frequency', hue='Cluster', data=rfm_scaled_df)
plt.title("Recency vs Frequency")

plt.subplot(1, 2, 2)
sns.scatterplot(x='Frequency', y='Monetary', hue='Cluster', data=rfm_scaled_df)
plt.title("Frequency vs Monetary")

plt.show()
```

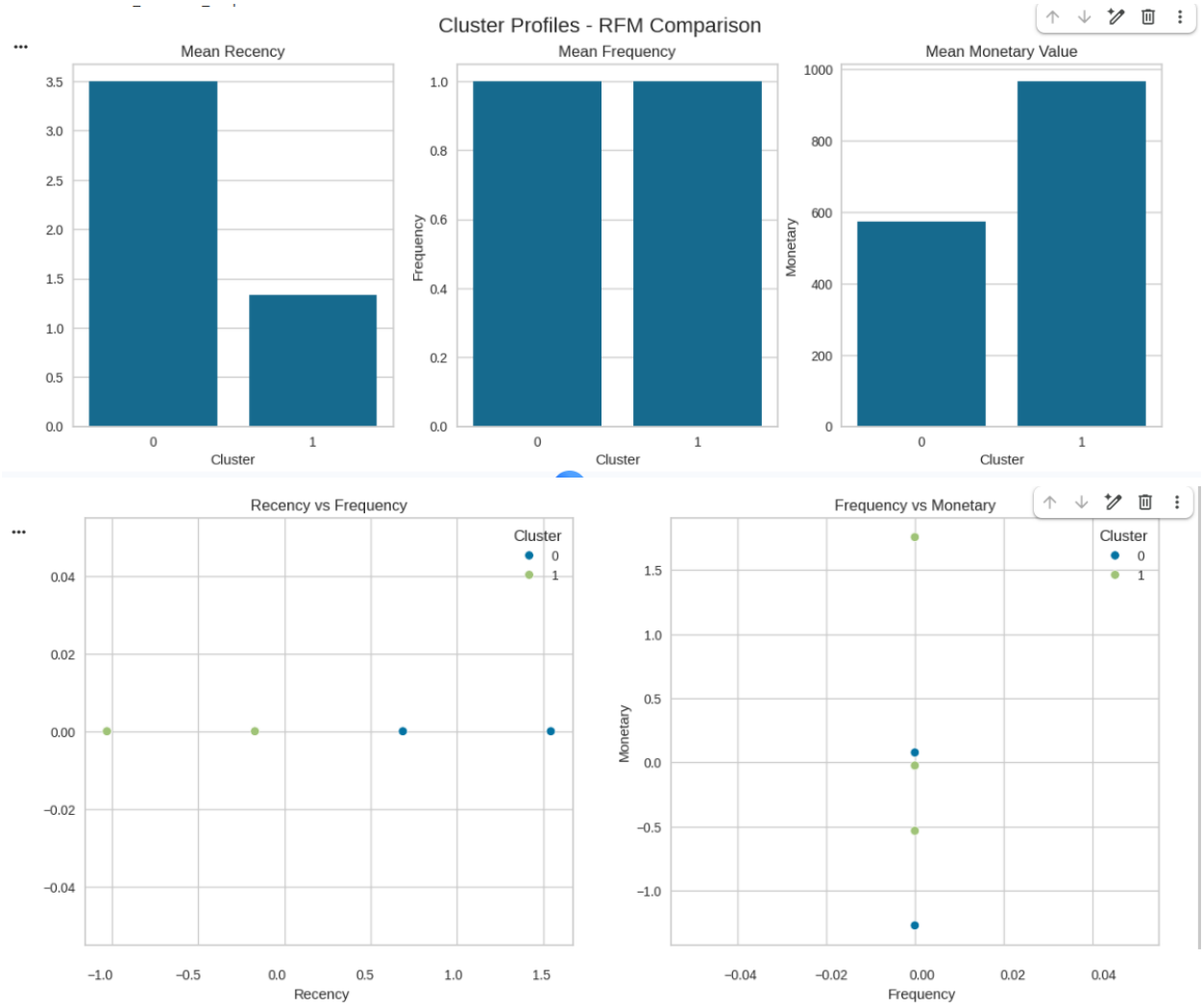
OUTPUT:

```
aset Loaded Successfully!
InvoiceNo StockCode Description  Quantity       InvoiceDate  UnitPrice  \
    10001      A100    Notebook         3  2023-01-01 10:00        200
    10002      A200         Pen         5  2023-01-02 11:30         50
    10003      A300         Bag         1  2023-01-05 09:20       1500
    10001      A400      Marker         2  2023-01-01 10:00        120
    10004      A100    Notebook         4  2023-01-03 14:10        200

CustomerID   Country
       501  Pakistan
       502  Pakistan
       503  Pakistan
       501  Pakistan
       504  Pakistan

aned Data Shape: (7, 9)

 Calculation Done!
CustomerID  Recency  Frequency  Monetary
       501        4          1       840
       502        3          1       310
       503        1          1      1500
       504        2          1       800
       505        1          1       600

ning Elbow Method...
```

```
ning Elbow Method...
ning: No optimal K found by Elbow method. Defaulting to 2 clusters.
imal K found: 2
r/local/lib/python3.12/dist-packages/yellowbrick/utils/kneed.py:156: YellowbrickWarning: No 'knee' or 'elbow point' detected This could be due to bad c
arnings.warn(warning_message, YellowbrickWarning)
r/local/lib/python3.12/dist-packages/yellowbrick/cluster/elbow.py:374: YellowbrickWarning: No 'knee' or 'elbow' point detected, pass `locate_elbow=Fals
arnings.warn(warning_message, YellowbrickWarning)
```




```
eans Clustering Completed!

ster Analysis:
      Recency  Frequency  Monetary  Count
ster
         3.50        1.0    575.00      2
         1.33        1.0    966.67      3

e Saved: rfm_clusters_output.csv
```


Cluster Profiles - RFM Comparison
Mean Recency
Mean Frequency
Mean Monetary Value
Frequency
Monetary
Cluster
Recency vs Frequency
Frequency vs Monetary
Recency
Frequency
Monetary
Cluster
0
1