Customers Segmentation & Clustering Analysis

MACHINE LEARNING

ABSTRACT:

Customer segmentation is one of the key aspects in Sierra Leone that support the business system in this project. We are using python language, clustering Techniques, by which the several segments of the customers are allowing them to target the possible users. In this data science project, it will make use of K-means clustering which is the essential algorithm for clustering the datasets. We are using one way to segment the customer as demographics like age, gender, customer ID, annual income and spending score in Sierra Leone.

```
# This Python 3 environment comes with many helpful analytics libraries i
nstalled
# It is defined by the kaggle/python Docker image: https://github.com/kag
gle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) wil
l list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) th
at gets preserved as output when you create a version using "Save & Run A
```

```
ll"
# You can also write temporary files to /kaggle/temp/, but they won't be
saved outside of the current session
```

```
/kaggle/input/customer-segmentation-tutorial-in-python/Mall_Customer
s.csv
```

What is Customer Segmentation?

Customer segmentation is the process of dividing customers into groups based on common characteristics so companies can market to each group more effectively. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.

Why Customer segmentation?

Customer segmentation has the potential to allow companies to address each customer in the most effective way.

Segmentation helps a company to Create and communicate targeted marketing startegy that will resonate with specific groups of customers, which improve customer service and establish better customer relationships.

How to Segment Customers?

Customer segmentation requires a company to gather specific information – data – about customers and analyze it to identify patterns that can be used to create segments.

Customer segmentation has the potential to allow companies to address each customer in the most effective way.
Segmentation helps a company to Create and communicate targeted marketing startegy that will resonate with specific groups of customers, which improve customer service and establish better customer relationships.

How to Segment Customers?

Customer segmentation requires a company to gather specific information – data – about customers and analyze it to identify patterns that can be used to create segments.
Some of that can be gathered from purchasing information – job title, geography, products purchased etc.

For this demo I am considering a hypothetical Mall Customer
When ever a customer applied for a membership for a Mall Customer. They need to fill the membership form, from which we can get the details like Name, Age, Gender, Annual Income etc. and based on how many times a customer visit the store and how much they spend, I am

computing a spending score between 1-100 for each customer.

+ Code + Markdown

Typical information include:

- CustomerID
- Gender
- Age
- Annual Income (k$)
- Spending Score (1-100)

1. Importing libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
%matplotlib inline
```

2. Loading Dataset

```
dataset=pd.read_csv('../input/Mall_Customers.csv')
dataset.head()
```

	CustomerID	Genre	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

3. Selecting features columns

[3]:
```
df=dataset.iloc[:,3:5]
```

[4]:
```
#Features Matrix
X=df.values
```

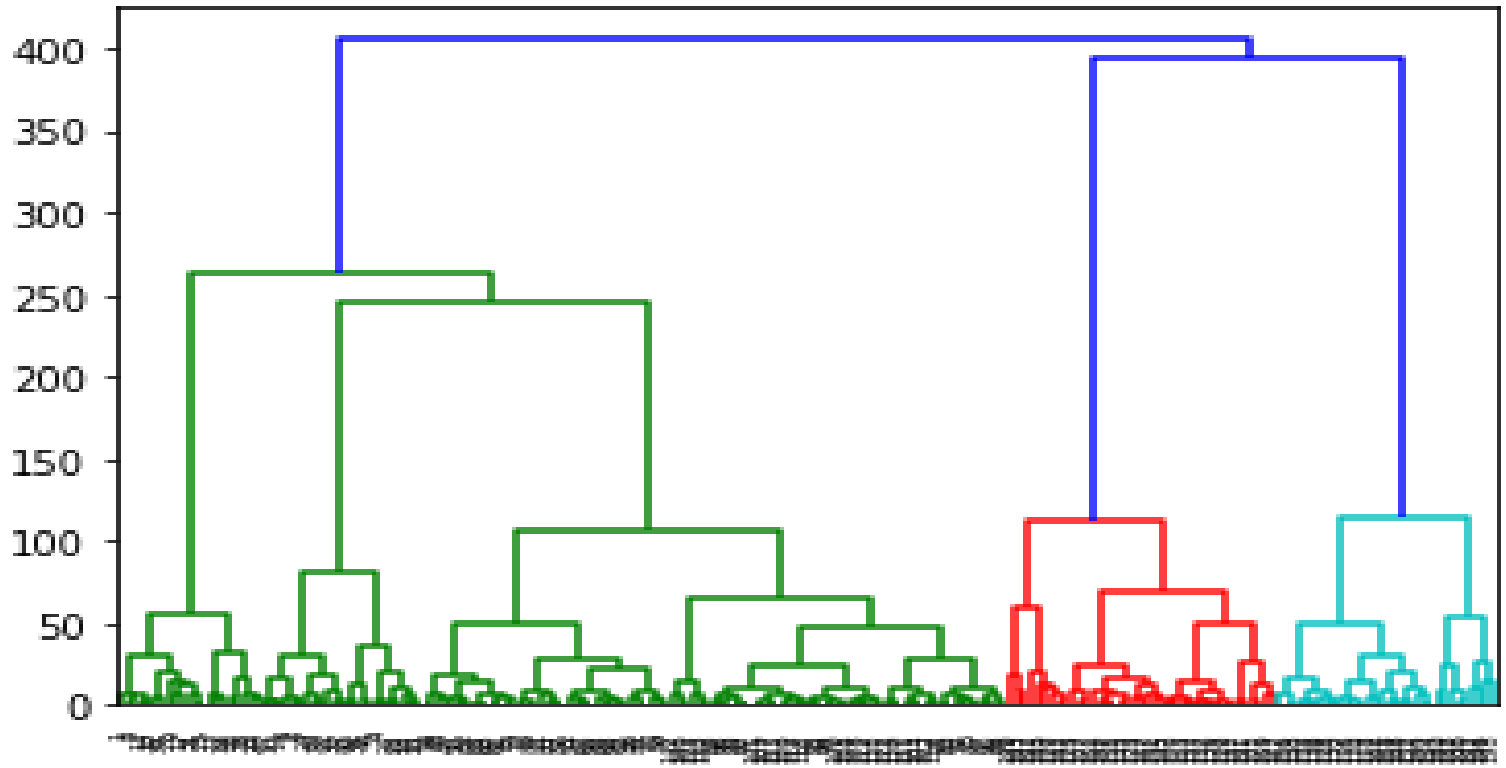
4. Cluster Analysis

+ Code + Markdown

** First Step is to find the Number of Cluster**

Finding optimal number of cluster using dendrogram

```
import scipy
from scipy.cluster import hierarchy
dendro=hierarchy.dendrogram(hierarchy.linkage(X,method='ward'))
```



From the above graph we can see that optimal number of custer will be 5.

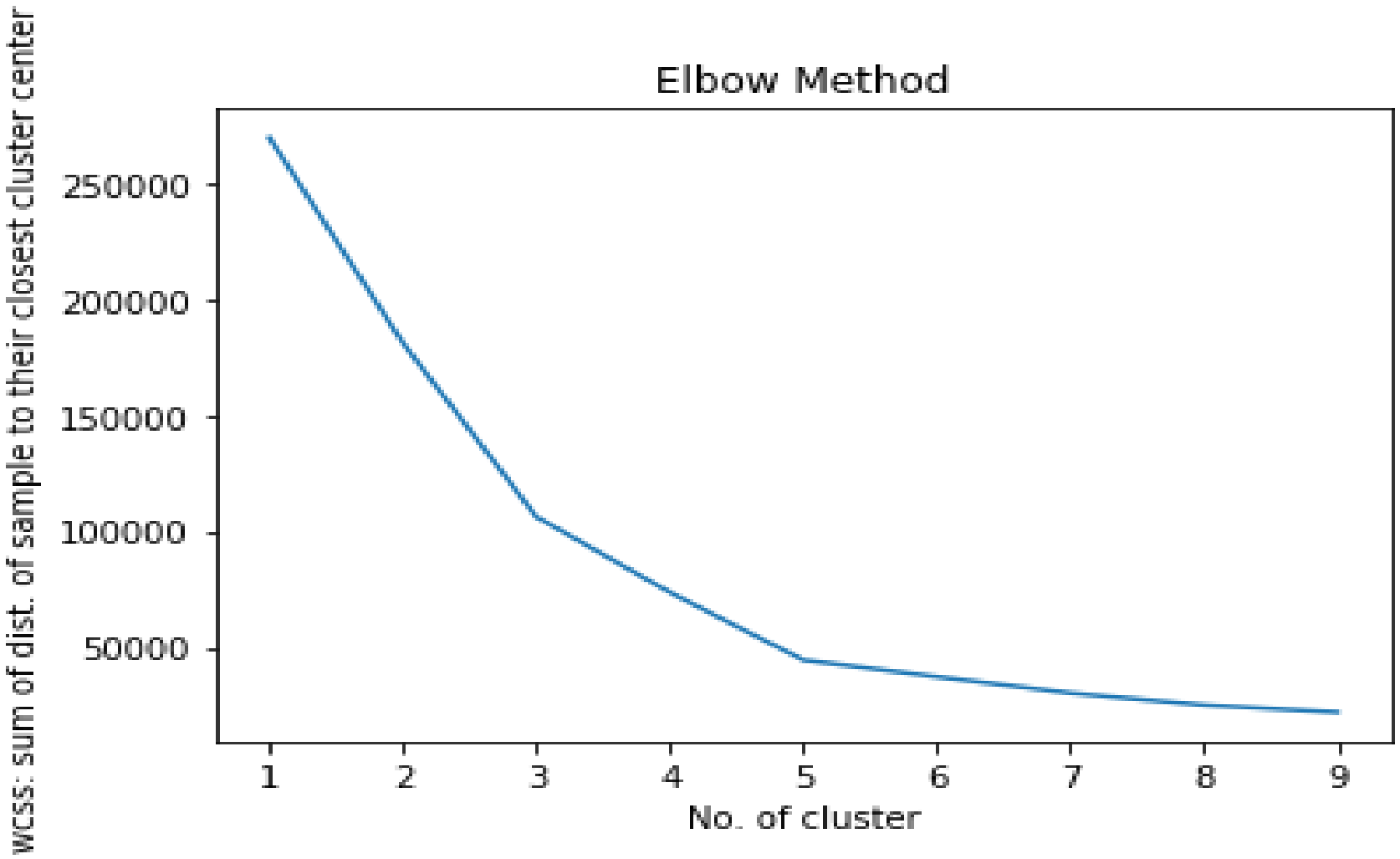
[6]:

```
from sklearn.cluster import KMeans
wcss=[]
for i in range(1,10):
    kmeans=KMeans(n_clusters=i,init='k-means++',)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

plt.plot(range(1,10),wcss)
plt.title('Elbow Method')
plt.xlabel('No. of cluster')
plt.ylabel('wcss: sum of dist. of sample to their closest cluster center' )
```

[6]:

```
Text(0,0.5,'wcss: sum of dist. of sample to their closest cluster center')
```


Elbow Method
wcss: sum of dist. of sample to their closest cluster center
250000
200000
150000
100000
50000
1
2
3
4
5
6
7
8
9
No. of cluster

From elbow method also, the optimal number of cluster will be 5.

5. KMeans Clustering Model with 5 cluster

[7]:

```
kmeans_1=KMeans(n_clusters=5)
kmeans_1.fit(X)
cluster_pred=kmeans_1.predict(X)
cluster_pred_2=kmeans_1.labels_
cluster_center=kmeans_1.cluster_centers_
```

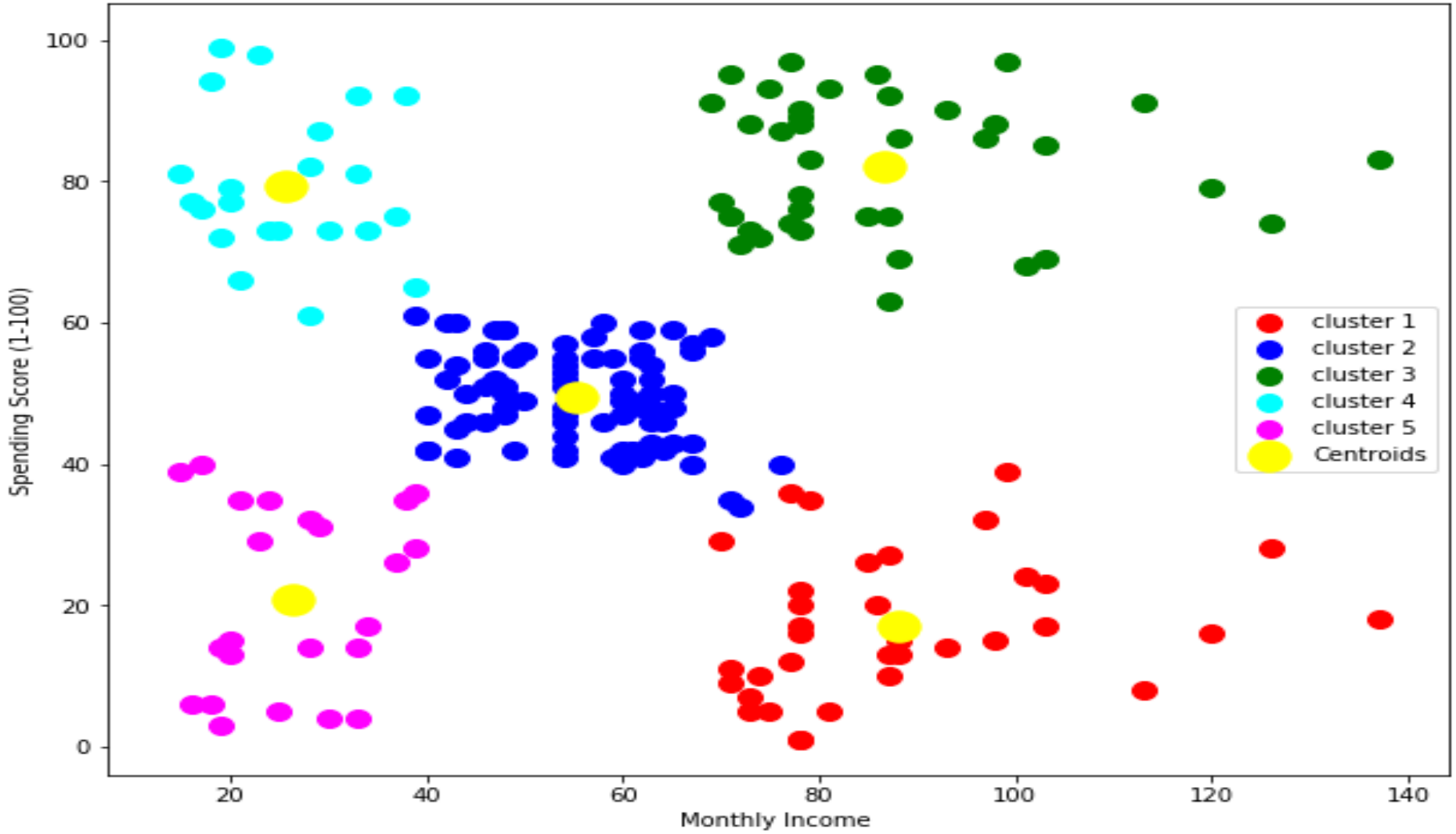
6. Visualization

+ Code + Markdown

[8]:

```
# Visualising the clusters
plt.figure(figsize=(10,8))
plt.scatter(X[cluster_pred==0,0],X[cluster_pred==0,1], s = 100, c = 'red', label ='cluster 1' )
plt.scatter(X[cluster_pred==1,0],X[cluster_pred==1,1], s = 100, c = 'blue', label ='cluster 2' )
plt.scatter(X[cluster_pred==2,0],X[cluster_pred==2,1], s = 100, c = 'green', label ='cluster 3' )
plt.scatter(X[cluster_pred==3,0],X[cluster_pred==3,1], s = 100, c = 'cyan', label = 'cluster 4')
plt.scatter(X[cluster_pred==4,0],X[cluster_pred==4,1], s = 100, c = 'magenta', label = 'cluster 5')
plt.scatter(cluster_center[:,0],cluster_center[:,1], s = 300, c = 'yellow', label = 'Centroids')
plt.title('Clusters of customers')
plt.xlabel('Monthly Income ')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

Clusters of customers



Visualising the clusters

The cluster with high Monthly income and high Spending score can be called as **Target** group
similarly the cluster with high income but low spending score can be called as **Careful** group
and the cluster with low income and high spending score can be called as **Careless** group
and cluster with low income and low spending score can be called as **Sensible** group
and one that is in the middle is **Standard** group.

I have divided the customer into 5 segment namely: 1. Careful 2. Standard 3. Target 4. Sensible 5. Careless

CONLUSION:

The given datasets of this customer segmentation is successfully processed by the using K-means clustering algorithm. And it help to improve customer service.
K- mean algorithm is one of the most popular clustering algorithms and normally the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the datasets. The goal of K- means is to group data points into distinct non- overlapping subgroups. One of the major application of K- means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.

I could not use dataset of sierra Leone, but in the future, since I have build a model, I am planning to use a real dataset from customer segmentation and cluster analysis in Sierra Leone.

MACHINE LEARNING GRADUATION ASSIGNMENT Project

End of Session

DIVE INTO CODE

IBRAHIM TURAY