OBJECTIVE

In this lab, we will use the K-Means clustering algorithm to group customers based on their sales and profit data. The goal is to segment customers into clusters to identify distinct purchasing behaviors.

PROBLEM STATEMENT

A retail company wants to segment its customers into groups based on their total sales and profit. This clustering can help in identifying various groups.

SOURCE CODE

```
# Setting up the Python Environment
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Reading the file
df = pd.read_csv("./Lab_3_Data.txt", sep=",")
display(df.head(10))
```

...

	CustomerID	Sales	Profit
0	C001	15000	4500
1	C002	20000	7000
2	C003	12000	4000
3	C004	30000	9000
4	C005	5000	1300
5	C006	8000	2500
6	C007	18000	6500
7	C008	5000	1200
8	C009	22000	18000
9	C010	9000	3000

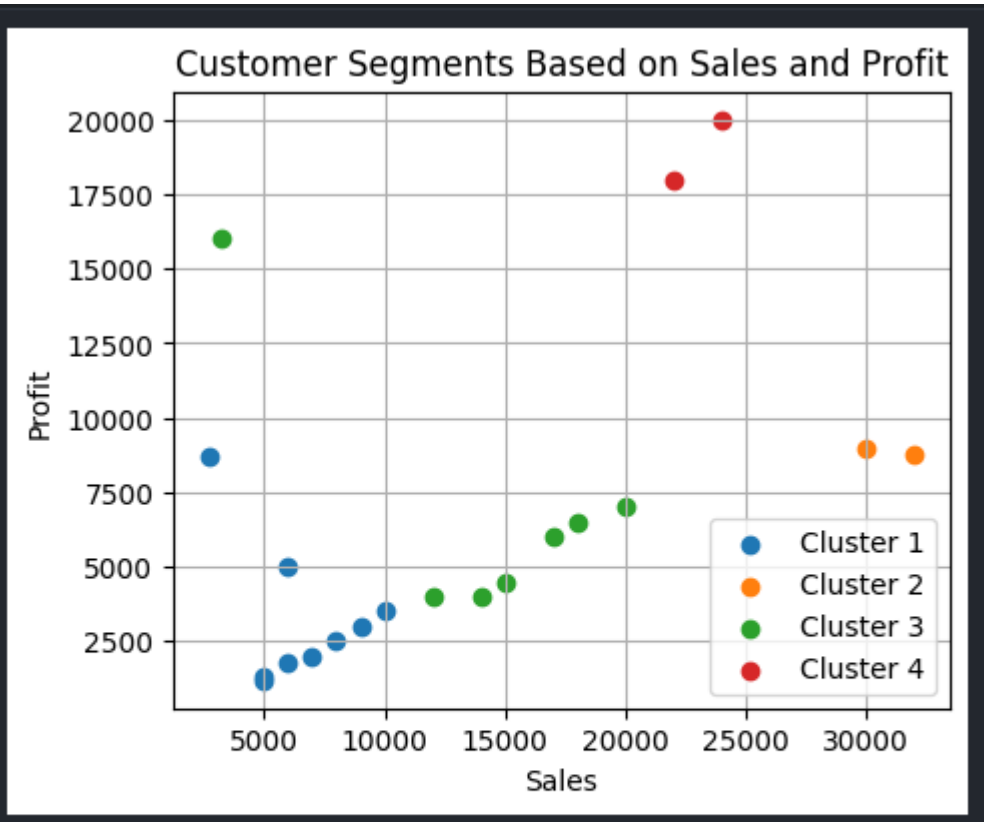
PROBLEMS 7 OUTPUT DEBUG CONSOLE TERMINAL PO

```
Microsoft Windows [Version 10.0.22631.4541]
(c) Microsoft Corporation. All rights reserved.

C:\Users\sayuj\OneDrive\Desktop\BIM054\00_BIM_Repo>
```

```
# Pre processing
# Check for missing values
df.isnull().sum()
```

```
...   CustomerID    0
      Sales         0
      Profit        0
      dtype: int64

PROBLEMS 7   OUTPUT   DEBUG CONSOLE   TERMINAL   PO

Microsoft Windows [Version 10.0.22631.4541]
(c) Microsoft Corporation. All rights reserved.

C:\Users\sayuj\OneDrive\Desktop\BIM054\00_BIM_Repo>
```

```
# Standardize the data
x = df[['Sales', 'Profit']]
scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)

# Applying the Kmeans algorithm
no_of_cluster = 4
kmeans = KMeans(n_clusters= no_of_cluster, random_state=42)
df['Cluster'] = kmeans.fit_predict(x_scaled)

# Visualize the clusters
plt.figure(figsize=(5, 4))
for cluster in range(no_of_cluster):
    cluster_data = df[df['Cluster'] == cluster]
    plt.scatter(cluster_data['Sales'], cluster_data['Profit'], label=f'Cluster {cluster + 1}')

plt.title('Customer Segments Based on Sales and Profit')
plt.xlabel('Sales')
plt.ylabel('Profit')
plt.legend()
plt.grid()
plt.show()
```



PROBLEMS 7 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER SPELL CHEC

```
Microsoft Windows [Version 10.0.22631.4541]
(c) Microsoft Corporation. All rights reserved.

C:\Users\sayuj\OneDrive\Desktop\BIM054\00_BIM_Repo>
```

DISCUSSION QUESTIONS

1. How can the company use these clusters to improve customer engagement?

The company can use the clusters in the following ways to improve customer engagement:

- **Targeted Marketing:** Design specific campaigns for each cluster, e.g., promotions for low-sales customers to increase their purchases or loyalty rewards for high-profit customers.
- **Resource Allocation:** Allocate customer service resources strategically, prioritizing high-value clusters.
- **Product Recommendations:** Use cluster insights to offer tailored product recommendations to different customer groups.

2. What additional features could improve clustering accuracy?

Clustering Accuracy can be improved by considering the following factors

- **Customer Demographics:** Adding features such as age, gender, location, or income could provide more refined segmentation.
- **Purchase Frequency:** Including how often a customer purchases could differentiate between frequent buyers and occasional shoppers.
- **Product Categories:** Features like the type of products purchased can refine clustering.

3. What would happen if the number of clusters (k) increased or decreased?

If the clusters (k) are increased, it leads to smaller, more specific clusters. While it may uncover subtle patterns, it risks overfitting and losing generalizability.

If the clusters (k) are decreased, results are broader, and less specific clusters are formed. It simplifies analysis but might mask meaningful differences between customers.