Name: Allan Rodrigues
Class: TE IT A
Roll no: 59
Pid:191104

St. Francis Institute of Technology, Mumbai-400 103
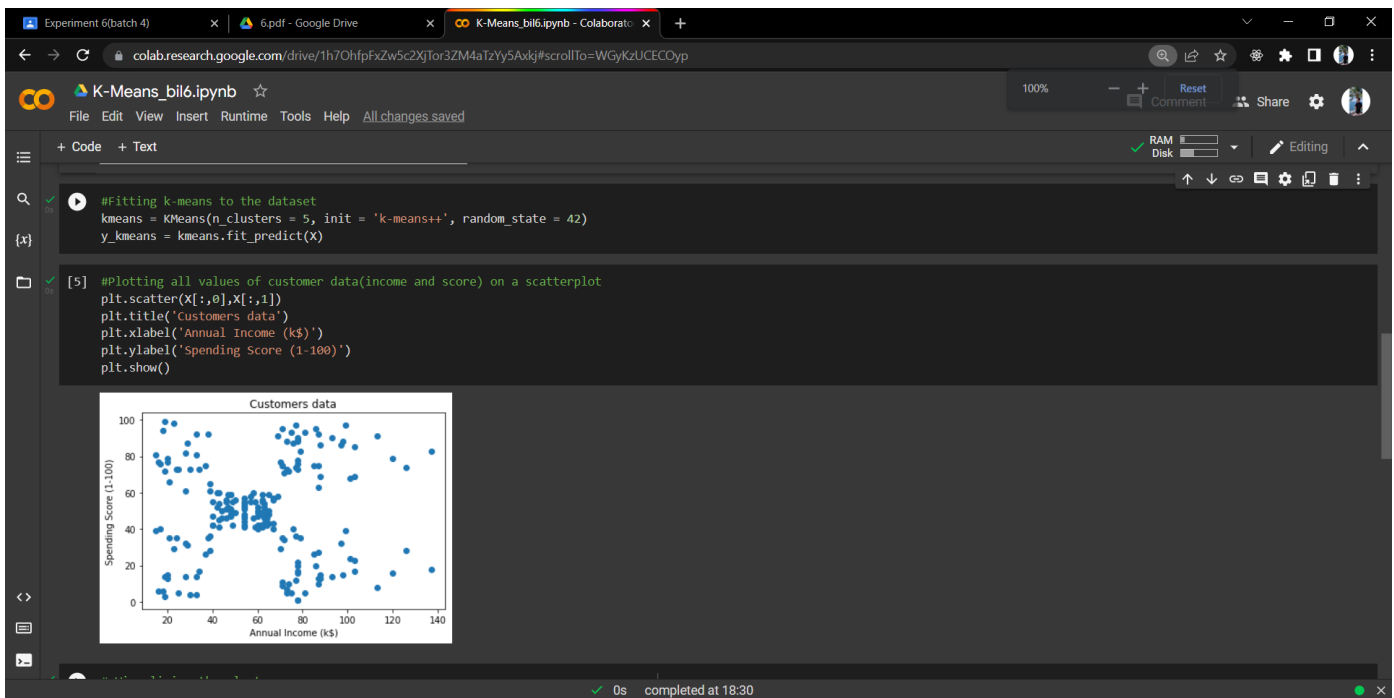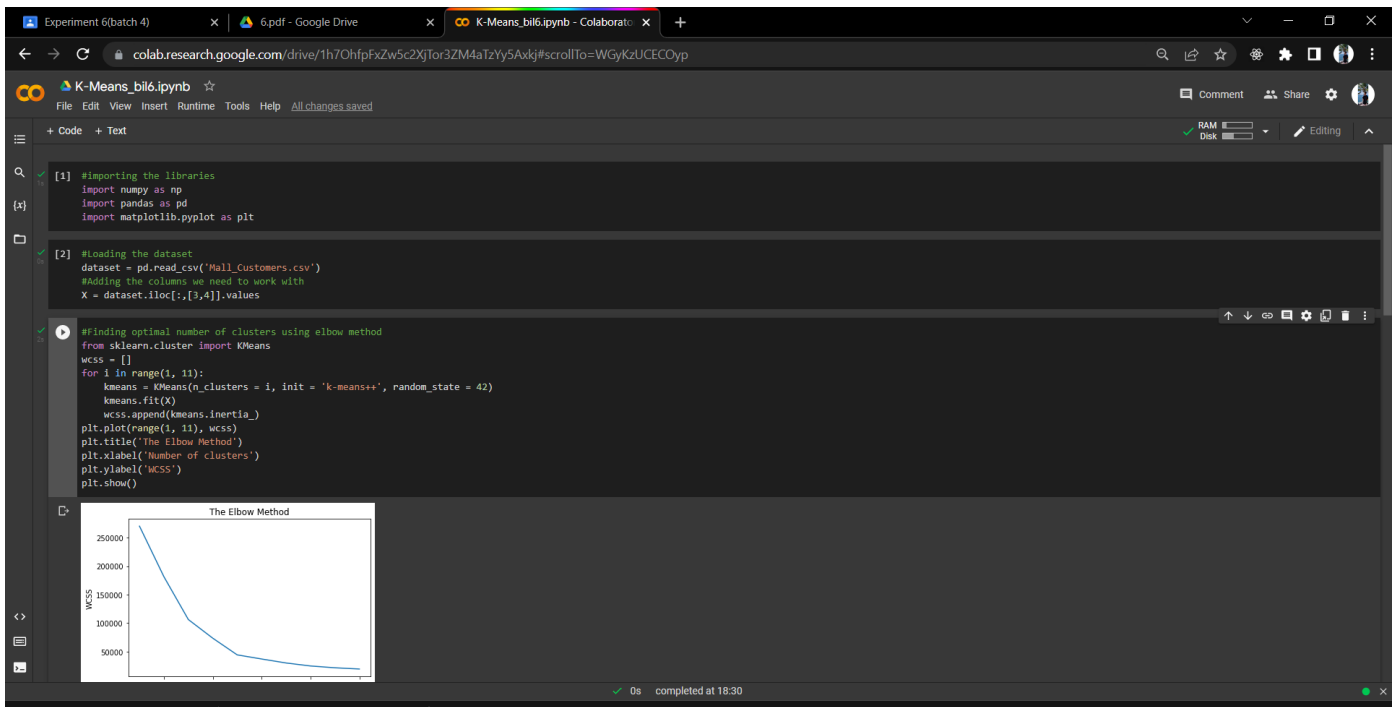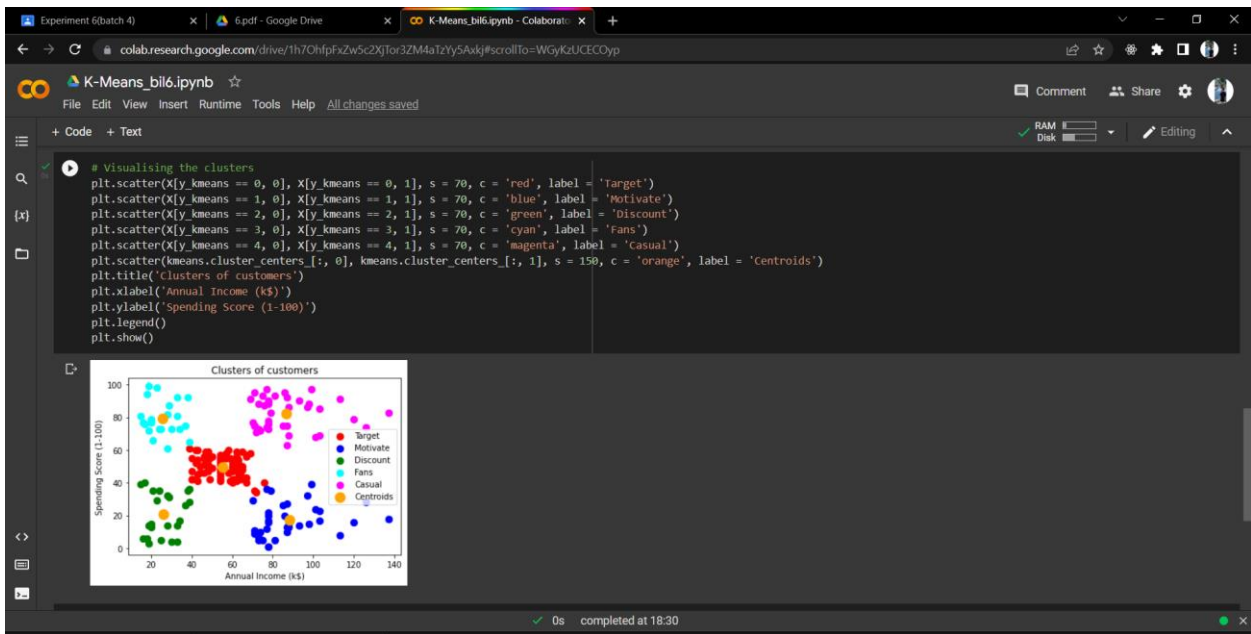## Department Of Information Technology
Subject: **Business Intelligence Lab**

Experiment – 6: **To Implement any one clustering algorithm using Java (K Means, Agglomerative and Divisive)**

1. **Aim: :** To Implement any one clustering algorithm using Java (K Means, Agglomerative and Divisive)

2. **Objectives:** After study of this experiment, the students will be able to implement K Means Algorithm

3. **Outcomes:** After study of this experiment, the students will be able to
   CO 4: Design and Implement various clustering data mining techniques such as Partitioning methods, Hierarchical Methods, Density - Based methods along with identification and analysis of outlier.

4. **Prerequisite:** Introduction to all the three clustering algorithms & Problem solving approach.

5. **Requirements:** Personal Computer, Windows XP operating system/Windows 7, Internet Connection, Microsoft Word, WEKA tool.

6. **Theory:**
   a. Explain Data Mining Algorithm
   b. Give detailed example.Solve numerical
   c. Implementation of Clustering Algorithms in Java/R/Python

7. **Laboratory Exercise:** Implementation of Classification Algorithm using Java/R/Python, Printout of implementation along with coding and snapshot

8. **Post-Experiments Exercise**
   a. **Questions:**
      ● In form of MCQ type test
   b. **Conclusion:**
      ● Summary of Experiment
      ● Importance of Experiment
      ● Application of Experiment

9. **Reference:** Data Mining: Concept & Techniques, 3rd Edition, Jiawei Han, Micheline   Kamber, Jian Pei, Elsevier.

Q7)


```python
[1] #importing the libraries
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
```

```python
[2] #Loading the dataset
    dataset = pd.read_csv('Mall_Customers.csv')
    #Adding the columns we need to work with
    X = dataset.iloc[:,[3,4]].values
```

```python
#Finding optimal number of clusters using elbow method
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



```python
#Fitting k-means to the dataset
kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(X)
```

```python
[5] #Plotting all values of customer data(income and score) on a scatterplot
    plt.scatter(X[:,0],X[:,1])
    plt.title('Customers data')
    plt.xlabel('Annual Income (k$)')
    plt.ylabel('Spending Score (1-100)')
    plt.show()
```

colab.research.google.com/drive/1h7OhfpFxZw5c2XjTor3ZM4aTzYy5Axkj#scrollTo=WGyKzUCECOyp

K-Means_bil6.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```python
# Visualising the clusters
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 70, c = 'red', label = 'Target')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 70, c = 'blue', label = 'Motivate')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 70, c = 'green', label = 'Discount')
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 70, c = 'cyan', label = 'Fans')
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 70, c = 'magenta', label = 'Casual')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 150, c = 'orange', label = 'Centroids')
plt.title('Clusters of customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```


✓ 0s   completed at 18:30

Q6)

Allan Rodrigues   TE IT A-59

B1L : Exp 6

b) 2, 4, 10, 12, 3, 20, 30, 11, 2, 5

$k_1 = \{2, 3\}$
$k_2 = \{4, 10, 12, 20, 30, 11, 25\}$

New mean , $m_1 = \dfrac{2+3}{2} = 2.5$

$m_2 = \dfrac{4+10+12+20+30+11+25}{7} = 16$

i) New cluster
$k_1 = \{2, 3, 4\}$ $k_2 = \{10, 12, 20, 30, 11, 25\}$
New mean,
∴ $m_1 = 3$, $m_2 = 18$

iii) New clusters
$k_1 = \{2, 3, 4, 10\}$
$k_2 = \{12, 20, 30, 11, 25\}$
New mean,
$m_1 = 4.75$
$m_2 = 196$

iv) New cluster
$k_1 = \{2, 3, 4, 10, 11, 12\}$
$k_2 = \{20, 30, 25\}$ $m_1 = 7$ $m_2 = 25$

Since the objects are same
∴ The mean will also not change
Hence these are the final clusters

Q8)

Q.8

b) Conclusion.

In this experiment we have learned about clustering algorithm. we have learned k-means clustering algorithm and implement it using python.

Clustering methods helps in grouping the data points into clusters.

Companies can cluster different customer① types into one group based on different factors.