Progressive Education Society's

Modern College of Engineering, Pune

**MCA Department A.Y.2024-25**

# (410907) Big Data Analytics Laboratory

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Class: SY-MCA Shift / Div : S3 / A Roll Number : 52062

Name: Laxman Shinde Assignment No: 2 Date of Implementation: 06 / 09 / 2024

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Q3. i. Using Python's Scikit-learn and Matplotlib, apply the K-Means clustering algorithm to**

**a dataset (e.g., customer segmentation data) and visualize the clusters using a scatter**

**plot.**

**Code :**

if (!require("ggplot2")) install.packages("ggplot2", dependencies=TRUE)

library(ggplot2)

data("iris")

iris\_data <- iris[, 1:4]

set.seed(123)

kmeans\_result <- kmeans(iris\_data, centers = 3, nstart = 20)

iris$Cluster <- as.factor(kmeans\_result$cluster)

ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = Cluster)) +

geom\_point(size = 3) +

labs(title = "K-Means Clustering of Iris Data", x = "Petal Length", y = "Petal Width") +

theme\_minimal()

**Output :**

