DATA ANALYTICS WITH R, EXCEL AND TABLAEU

ASSIGNMENT model Deployment 21.1

ANSWERS

By ASHISH S SHANBHAG

ashishshanbhag108@gmail.com

5. Problem Statement

1. Use the below given data set Data Set

2. Perform the below given activities:

a. Apply PCA to the dataset and show proportion of variance

Ans # Taking the numeric part of the Epicurious data

```
data_epi <- iris[1:4]
# Calculating the covariance matrix
Cov_data <- cov(data_epi )
# Find out the eigenvectors and eigenvalues using the covariance matrix
Eigen_data <- eigen(Cov_data)
# Using the inbuilt function
PCA_data <- princomp(data_epi ,cor="False")
# Let’s now compare the output variances
Eigen_data$values
```

```
              Comp.1      Comp.2      Comp.3     Comp.4
Sepal.Length  0.36138659 -0.65658877 -0.58202985  0.3154872
Sepal.Width  -0.08452251 -0.73016143  0.59791083 -0.3197231
Petal.Length  0.85667061  0.17337266  0.07623608 -0.4798390
```

Petal.Width 0.35828920 0.07548102 0.54583143 0.7536574

```
          [,1]       [,2]       [,3]       [,4]
[1,]  0.36138659 -0.65658877 -0.58202985  0.3154872
[2,] -0.08452251 -0.73016143  0.59791083 -0.3197231
[3,]  0.85667061  0.17337266  0.07623608 -0.4798390
[4,]  0.35828920  0.07548102  0.54583143  0.7536574
```

summary(PCA_data)

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	2.0494032	0.49097143	0.27872586	0.153870700
Proportion of Variance	0.9246187	0.05306648	0.01710261	0.005212184
Cumulative Proportion	0.9246187	0.97768521	0.99478782	1.000000000

b. Perform PCA using SVD approach

Ans # Taking the numeric part of the Epicurious data

```
data_epi <- iris[1:4]
# Calculating the covariance matrix
Cov_data <- cov(data_epi )
# Find out the eigenvectors and eigenvalues using the covariance matrix
Eigen_data <- eigen(Cov_data)
# Using the inbuilt function
PCA_data <- princomp(data_epi ,cor="False")
```

Let’s now compare the output variances

```
Eigen_data$values
```

	Comp.1	Comp.2	Comp.3	Comp.4
Sepal.Length	0.36138659	-0.65658877	-0.58202985	0.3154872
Sepal.Width	-0.08452251	-0.73016143	0.59791083	-0.3197231
Petal.Length	0.85667061	0.17337266	0.07623608	-0.4798390
Petal.Width	0.35828920	0.07548102	0.54583143	0.7536574

	[,1]	[,2]	[,3]	[,4]
[1,]	0.36138659	-0.65658877	-0.58202985	0.3154872
[2,]	-0.08452251	-0.73016143	0.59791083	-0.3197231
[3,]	0.85667061	0.17337266	0.07623608	-0.4798390
[4,]	0.35828920	0.07548102	0.54583143	0.7536574

```
summary(PCA_data)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	2.0494032	0.49097143	0.27872586	0.153870700
Proportion of Variance	0.9246187	0.05306648	0.01710261	0.005212184
Cumulative Proportion	0.9246187	0.97768521	0.99478782	1.000000000

c. Show the graphs of PCA components

Ans: