**Department of Computer Science & Engineering**

**Professional Skills Lab -2**

**Sixth Sem B Tech 2022-23**

Write Programs in python

**P1: DESCRIPTIVE STATISTICS**

a. Write an program to find basic descriptive statistics using summary, str, quartile function on mtcars& cars datasets.

b. Write an program to find subset of dataset by using

subset (), aggregate () functions on iris dataset.

P2: **READING AND WRITING DIFFERENT TYPES OF DATASETS**

1. Reading different types of data sets (.txt, .csv) from Web and disk and writing in file in specific disk location.
2. Reading Excel data sheet in Python.

P3: **VISUALIZATIONS**

**a.** Find the data distributions using box and scatter plot. b. Find the outliers using plot. c. Plot the histogram, bar chart and pie chart on sample data.

P4: **CORRELATION AND COVARIANCE**

a. Find the correlation matrix.

b. Plot the correlation plot on dataset and visualize giving an overview of relationships

among data on iris data.

c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.

P5: **REGRESSION MODEL** Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. Require (foreign), require (MASS).

P6: MULTIPLE REGRESSION MODEL

Apply multiple regressions, if data have a continuous Independent variable. Apply on above dataset.

P7: CLASSIFICATION MODEL

a. Install relevant package for classification.

b. Choose classifier for classification problem.

c. Evaluate the performance of classifier.

P8: CLUSTERING MODEL

a. Clustering algorithms for unsupervised classification.

b. Plot the cluster data using matplotlib visualizations.

P9: Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and incorrect predictions.