



Lab Digital Assignment 3

Programme Name & Branch : **MSc Data Science**

Slot(s) : **L27+L28,L33+L34**

Class Number : **VL2024250504652,VL2024250504655**

Course Code : **PMDS604P**

Course Name : **Exploratory Data Analysis Lab**

General Instruction(s): *Prepare your Lab Digital Assignment in Python using Jupyter Notebook, and convert it to a PDF file for upload to VTOP.*

Task: Perform outlier analysis using the following methods: Z-score, Interquartile Range (IQR), K-Nearest Neighbors (KNN), and Mahalanobis Distance on a synthetic dataset and one of the provided datasets. For each method, experiment with different thresholds and parameters, and explain your findings based on the variation in detected outliers. Additionally, using an example dataset with more than 3 dimensions, explain why it is more effective to apply dimensionality reduction techniques (such as PCA) before conducting outlier analysis, rather than performing outlier detection directly on high-dimensional data. Support your explanation with plots and examples from your analysis.
