**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: BTI Computer Engineering

**Course: Data Mining**

**Experiment No.03**

PART A

(PART A: TO BE REFFERED BY STUDENTS)

**A.1 Aim:**

Apply Data Cleaning techniques for categorical, numerical attributes and observe the effect on data dispersion

**A.2 Prerequisite:**

**Python packages – matplotlib, pandas, numpy etc.**

**A.3 Outcome:**

**After successful completion of this experiment students will be able to:**

**Answer following questions:**

1. Different types of attribute in dataset
2. Count of Missing Values in Numerical and Categorical attribute
3. Handling the Missing Values
4. Handling Invalid values
5. Give the Count of Unique value for each attribute
6. Converting categorical data to numeric data
7. Visualizing the data after data cleaning using boxplot
8. Comment on the skewness of the data and remove outlier if any

**Task**

Select the dataset of your choice, clean the dataset and answer the above questions

PART B

(PART B: TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Black board access available)***

| **Roll No.** | **Name:** |
| --- | --- |
| **Class :** | **Batch :** |
| **Date of Experiment:** | **Date of Submission** |
| **Grade :** |  |

**B.1 Answers of Task to be written by student:**

***(Paste your answers completed during the 2 hours of practical in the lab here)***

**B.2 Observations and learning:**

***(Students are expected to comment on the output obtained with clear observations and learning for each task/ sub part assigned)***

**B.3 Conclusion:**

*(****Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)***