

School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSEL301
Year- 2022
Date- 08-07-2022

Type- Core
Course Name-AIML
Semester- Odd
Batch- 5th Sem

1 - Lab Assignment No. 1.2

Objective: To use Pandas Python library and perform various pre-processing operations.

1. **Create a Pandas data frame(10)**
2. **Read and load the unclean_data.csv ([unclean_data.csv](#)) file and perform following operations(40)**
 - a) Display first and last 10 rows of the dataset
 - b) Check the detail information of the dataset like number of entries, columns name, datatype.
 - c) Find out the columns which does not have null values entries (if any)
 - d) Find out the total number of null values exist in each column.
 - e) Drop all null values without changing the original dataset.
 - f) Drop all null values in the original dataset itself.
3. **Check and handle wrong format, incorrect and duplicate entries in the dataset.(20)**
 - a) Check if there is there any wrong format entry of 'date' in the dataset. If you find any wrong date format correct it in the form of YYYY-MM-DD.
 - b) Set the maximum time as 120 in the 'time' column of the dataset.
 - c) Check if there is any duplicate row in the dataset. Remove the duplicate rows (if any)
4. **Establish the correlation among the attributes in the dataset.(20)**
5. **Find the mean, standard deviation, minimum, maximum and 50 percentile values of each attribute in the dataset.(20)**

Suggested Platform: Python: Azure Notebook/Google Colab Notebook, packages such as numPy and Pandas.