School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET-214
Year- 2024-25
Date- 23rd -29th August 2024

Type- Specialization Core
Course Name- Data Analysis using Python
Semester- Odd
Batch-

Lab # No. 5

Understanding the concept of Merging, reshaping and missing value handling.

1. Create two dataframes df_karate (column name:Node_A, Node_B) and df_karate_weight (column: Node_1, Node_2, Weight) by reading karate.csv and karate_with_weight_1.csv.

a) Merge both dataframe such that data of df_karate_weight should be in left side.
b) Concatenate both dataframes along the axis = 1 and axis = 0.

c) Create two dataframes df1_students_name(Column name=Student_ID, Name, Mob_no) and df2_students_marks(Column name=Student_ID, Marks, Subject). Fill in the required data.
   i) Merge both dataframes and Returns only the common (on students_ID) rows from both DataFrames.
   ii) Merge and Returns all rows from both DataFrames.

2.

a) Consider the following DataFrame that records sales of different products across different months.

```
data = {
'Month': ['Jan', 'Jan', 'Jan', 'Feb', 'Feb', 'Feb', 'Mar', 'Mar', 'Mar', 'Apr', 'Apr', 'Apr', 'May', 'May', 'May', 'Jun', 'Jun', 'Jun', 'Jul', 'Jul', 'Jul', 'Aug', 'Aug', 'Aug', 'Sep', 'Sep', 'Sep', 'Oct', 'Oct', 'Oct', 'Nov', 'Nov', 'Nov', 'Dec', 'Dec', 'Dec'],

'Product': ['A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C'],

'Sales': [100, 150, 120, 200, 180, 160, 300, 250, 240, 350, 300, 280, 400, 380, 360, 450, 400, 380, 500, 450, 430, 550, 500, 480, 600, 550, 520, 650, 600, 570, 700, 650, 620, 750, 700, 680] }
```

   i) Reshape the dataframe so that each product's sales are in separate columns, with the months as rows.

      Expected Output:

Product	A	B	C
Month			
Jan	100	150	120
Feb	200	180	160
Mar	300	250	240
Apr	350	300	280
May	400	380	360
Jun	450	400	380
Jul	500	450	430
Aug	550	500	480
Sep	600	550	520
Oct	650	600	570
Nov	700	650	620
Dec	750	700	680

ii) Show transpose of previous dataframe.

3. Read "Diabetes Missing Data.csv" dataset and perform following task.

i) Returns a DataFrame where all the values are replaced with a Boolean value True for NULL values, and otherwise False.
ii) Returns a DataFrame where all the missing values of column Serum_Insulin should be replaced by mean of present value in Serum_Insulin.
iii) Returns a DataFrame where all the missing values of column Serum_Insulin should be replaced by 0 of present value in Serum_Insulin.
iv) Remove rows from a DataFrame that contain only missing values.