RAMAIAH
Institute of Technology

Ramaiah Institute of Technology
(Autonomous Institute, Affiliated to VTU)
Department of Computer Science & Engineering

Data Visualization with Python Lab(CSL48)

USN: **Week #: 07**

Semester: **Section:** **Date:**

Instructions:

- **Implement the following programs using python language.**

Topic: Introduction to Pandas- DataFrames

Programs:

1. Create a Pandas DataFrame from a Dictionary

Write a Python program to create a DataFrame using a dictionary. The dictionary should contain passenger details such as name, age, and fare. Then, display the created DataFrame.

2. Write a Python program to perform the following operations using the Titanic dataset:

Dataset link: titanic.csv

https://drive.google.com/file/d/1KRqAMwDcJsFazQgaJmMiDaZvnoejjtdJ/view?usp=sharing

1. Read the **Titanic dataset** from a CSV file (titanic.csv) into a Pandas DataFrame and display the first few rows.
2. Filter the dataset to select only the rows where the passenger's age is greater than **30**, and display the filtered results.
3. Replace missing values in the age column with a **default value** (e.g., **30**) and display the updated DataFrame.
4. Remove any rows containing **missing values** from the dataset and display the cleaned DataFrame.
5. Determine and display the **maximum and minimum** values in the age column.
6. Create a new DataFrame with **passenger IDs and ticket discounts**, and perform a **left join** between this DataFrame and the Titanic dataset. Display the merged DataFrame.
7. Check for **duplicate rows** in the dataset, remove them, and display the number of duplicates found along with the cleaned DataFrame.
8. Save the cleaned **Titanic dataset** (after removing duplicates and handling missing values) to an **Excel file** (titanic_cleaned.xlsx).