Project 1: Data Manipulation, Handling Missing Data, Data Visualization

1. Load the dataset file, **'titanic.csv'** from the following link:
   https://raw.githubusercontent.com/PulockDas/pd-12-resources/refs/heads/master/titanic.csv

2. Find out all the feature names where Null values exist.

3. Fill the null values of the 'Age' column with the mean of the other values.
   And also fill the Null values of the 'Cabin' column with 'Unknown'.

4. Plot the dataset with 2 bars; Survived, Dead.
   And include Female and Male survivors' measurement in each bar.

5. Follow the step 4 and include the Survived, and Dead bars amongst the **'Pclass'.**

6. Create a column named **'AgeClass'** and insert values;
   * 0 if age <= 16
   * 1 if age <= 26
   * 2 if age <= 36
   * 3 if age <= 62
   * 4 otherwise

7. Now drop the column **'Age'.**

8. Follow step 4 and create a bar plot of Survived and Dead using the value counts amongst 'AgeClass' groups.

Assignments

Divide your identification number with 6.
Suppose, the remainder is x. You're assigned to complete the x^{th} task.
If x is 0, then go for **task-6**.

1. Create a Scatter plot of Male and Female Survivors.
   Display Passengers' **Age** on the X-axis, and **Fare** on the Y-axis.
   Note: Use color = 'green' for male and color = 'red' for female.

2. Create a Scatter plot of the Male, Female passengers who were dead.
   Display Passengers' **Age** on the X-axis, and **Fare** on the Y-axis.
   Note: Use color = 'blue' for male and color = 'yellow' for female.

3. Create a new column in the pandas dataframe with the name **'Number of Relatives'**.
   Assign the value like this: Number of Relatives = SibSp + Parch

   Create a Scatter plot of the passengers who were **dead**.
   Display Passengers' 'Number of Relatives' on the X-axis, and **Fare** on the Y-axis.

4. Create a new column in the pandas dataframe with the name **'Number of Relatives'**.
   Assign the value like this: Number of Relatives = SibSp + Parch

   Create a Scatter plot of the passengers who were **alive**.
   Display Passengers' 'Number of Relatives' on the X-axis, and **Fare** on the Y-axis.

5. Create a new column in the pandas dataframe with the name **'Number of Relatives'**.
   Assign the value like this: Number of Relatives = SibSp + Parch

   Create a Scatter plot of the passengers who were **dead**.
   Display Passengers' 'Number of Relatives' on the X-axis, and **Age** on the Y-axis.

6. Create a new column in the pandas dataframe with the name **'Number of Relatives'**.
   Assign the value like this: Number of Relatives = SibSp + Parch

   Create a Scatter plot of the passengers who were **alive**.
   Display Passengers' 'Number of Relatives' on the X-axis, and **Age** on the Y-axis.

Project 2: Dataset Merging, Data Manipulation, K-Means Clustering

1. Create a **CSV** file with the name, **'term-test-1-result.csv'**.
   There'll be three columns: 'Registration Number', 'Name', and 'TT-1 Marks'.
   You must include **50** students with their respective values for each column in the file.

2. Follow step 1 and create another **CSV** file with the name, **'term-test-2-result.csv'**.
   'Registration Number', 'Name' columns will have the same values as mentioned in the '**term-test-1-result.csv**' file. The value of the 'TT-2 Marks' column is most likely to be changed.

   Note: Don't make both files identical, and you mustn't copy each other's files.
   The full marks of each term test is **20**.

3. Load both the files in different pandas dataframes.
   Make a new merged pandas dataframe on their 'Registration Number' column.

4. Make a new column with the best term test marks for each student.
   Make a new column with the average term test marks for each student.

5. Drop both the columns named 'TT-1 Marks' and 'TT-2 Marks'.

6. Now, make a **CSV** file which will have the attendance and term final marks of every student.
   - Attendance full marks = 10
   - Term Final full marks = 100

   Load the CSV file as a pandas dataframe and merge it with the latest term test dataframe.
   Create a new column named Final Marks.
   - Final Marks = Term Final marks * 0.7 + Average Term Test marks + Attendance marks

7. Write the content of the latest pandas dataframe to a new **CSV** file named 'final result.csv'.

8. Cluster the final marks of each student using K-Means clustering algorithm. (**1<K<6**)
   Visualize the final clusters.