|  |
| --- |
|  |
| '''Python for Data Science - Perform Data Science on Titanic Dataset |
|  | a)Load the Titanic dataset into one of the data structures (NumPy or Pandas). |
|  | b)Display header rows and description of the loaded dataset. |
|  | c) Remove unnecessary features (E.g. drop unwanted columns) from the dataset. |
|  | d) Manipulate data by replacing empty column values with a default value. |
|  | e) Perform the following visualizations on the loaded dataset: |
|  | i) Passenger status (Survived/Died) against Passenger Class |
|  | ii) Survival rate of male vs female |
|  | iii) No of passengers in each age group |
|  |  |
|  | ''' |
|  |  |
|  | import numpy as np |
|  |  |
|  | import pandas as pd |
|  |  |
|  | titanic\_df = pd.read\_csv('train.csv') |
|  |  |
|  | # Convert the survived column to strings for easier reading |
|  | titanic\_df ['Survived'] = titanic\_df ['Survived'].map({ |
|  | 0: 'Died', |
|  | 1: 'Survived' |
|  | }) |
|  |  |
|  |  |
|  | print("======Data Headers Before Dropping Columns=======") |
|  | print(titanic\_df.head(5)) |
|  |  |
|  | print("\*\*\*\* \n\nDATA TRANSFORMATION \*\*\*\*\*\n") |
|  |  |
|  | print("======Data Headers After Dropping Columns - First Way=======") |
|  | titanic\_df.drop(['Parch','PassengerId','Name','Ticket'], axis=1, inplace=True) |
|  | print(titanic\_df.head(5)) |
|  | print("======Data Headers After Dropping Columns - Second Way =======") |
|  | titanic\_df = titanic\_df.drop(['SibSp','Fare'], axis=1) |
|  | print(titanic\_df.head(5)) |
|  | # Convert the Class column to strings for easier reading |
|  | titanic\_df ['Pclass'] = titanic\_df ['Pclass'].map({ |
|  | 1: 'Luxury Class', |
|  | 2: 'Economy Class', |
|  | 3: 'Lower Class' |
|  | }) |
|  |  |
|  | print("======Data Headers After Transforming Class Column =======") |
|  | print(titanic\_df.head(5)) |
|  |  |
|  | titanic\_df["Embarked"] = titanic\_df["Embarked"].fillna("S") |
|  | print("======Data Headers After Filling with default value for Embarked Column =======") |
|  | print(titanic\_df.head(5)) |
|  |  |
|  | # Convert the Embarked column to strings for easier reading |
|  | titanic\_df ['Embarked'] = titanic\_df ['Embarked'].map({ |
|  | 'C':'Cherbourg', |
|  | 'Q':'Queenstown', |
|  | 'S':'Southampton' |
|  | }) |
|  | print("======Data Headers After Transforming Embarked Column =======") |
|  | print(titanic\_df.head(5)) |
|  |  |
|  |  |