**Practical1 Apply preprocessing techniques on row data**

import pandas as pd

import numpy as np

pd.read\_csv("titanic.csv")

data = pd.read\_csv('titanic.csv')

data.head()

data.dtypes

data.columns

data["Survived"]==0

print('Total number of passangers in the training data...', len(data))

print('Number of passangers who survived...', len(data[data['Survived'] == 1]))

print("Number of passangers who didn't survived...", len(data[data['Survived'] == 0]))

data['Sex'].value\_counts()

print('% of male who survived', 100\*np. mean(data['Survived'] [data['Sex']=='male'])) print('% of female who survived', 100\*np. mean(data['Survived'][data['Sex']=='female']))

np.mean(data['Survived'] [data['Sex']=='male'])

data['Pclass'].value\_counts()

print('% of passengers who survived in first class', 100\*np.

mean(data['Survived'][data['Pclass'] == 1]))

print('% of passengers who survived in second class', 100\*np.

mean(data['Survived'][data['Pclass'] == 2]))

print('% of passengers who survived in third class', 100\*np.

mean(data['Survived'] [data['Pclass'] == 3]))

#data[["Pelass", "Survived"]] groupby(["Pelass"], as index False).mean()

data[["Pclass","Survived"]].groupby(['Pclass']).mean()

data.shape

data.info()

data['Age'].value\_counts()

data['Cabin']

data['Sex']

df2 = data.copy()

df2['Sex'] = data['Sex'].apply(lambda x: 1 if x == 'male' else 0)

df2 ['Sex']

#Dealing with missing values

df2 = data.copy() #dataframe copy

df2.isnull().sum()

int(data['Age'].mean())

df2['Age'] = df2['Age'].fillna(np.mean(df2['Age'])) df2.isnull().sum()