**SET A**

**Q no. 1) Write a python program to find all null values in a given data set and remove them. (Download dataset from github.com)**

import pandas as pd

df = pd.read\_csv('dataset.csv')

print("Original DataFrame:")

print(df)

print("\nRows with null values:")

print(df[df.isnull().any(axis=1)])

df\_cleaned = df.dropna()

print("\nDataFrame after removing rows with null values:")

print(df\_cleaned)

**Q no. 2) Write a python program the Categorical values in numeric format for a given dataset.**

import pandas as pd

from sklearn.preprocessing import LabelEncoder

data = {

'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],

'Gender': ['Female', 'Male', 'Male', 'Male', 'Female'],

'City': ['New York', 'Los Angeles', 'New York', 'Chicago', 'Houston']

}

df = pd.DataFrame(data)

print("Original DataFrame:")

print(df)le = LabelEncoder()

df['Gender\_LabelEncoded'] = le.fit\_transform(df['Gender'])

df['City\_LabelEncoded'] = le.fit\_transform(df['City'])

print("\nDataFrame after Label Encoding:")

print(df)

**SET B**

**Q no. 1) Write a python program to rescale the data between 0 and 1. (use inbuilt dataset)**

import pandas as pd

import numpy as np

from sklearn.preprocessing import MinMaxScaler

data={"feature1":[11,21,31,41,51],

"feature2":[100,200,300,400,500]}

df=pd.DataFrame(data)

scaler=MinMaxScaler()

rescaled\_data=scaler.fit\_transform(df)

rescaled\_df=pd.DataFrame(rescaled\_data,columns=df.columns)

print(rescaled\_df)

**Q no. 2) Write a python program to splitting the dataset into training and testing set.**

import pandas as pd

dataset = pd.read\_csv(“play\_tennis.csv”)

y = dataset.play

x = dataset.drop(‘play’,axis=1)

print(y)

print(x)

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

print(x\_train)

print(x\_test)

**SET C**

**Q no. 1) Write a python program to implement complete data pre-processing in a given data set. (Missing value, encoding categorical value, splitting the dataset into the training and test sets and feature scaling.**

import pandas as pd

df = pd.read\_csv('dataset.csv')

print("Original DataFrame:")

df

print("Missing values:")

missing\_values = data.isnull().sum()

missing\_values

le = LabelEncoder()

for column in df.select\_dtypes(include=['object']).columns:

df[column] = le.fit\_transform(df[column])

print("Categorical values encoded successfully!")

print(df.head())

y = df.Address

x = df.drop('Address',axis=1)

print(y)

print(x)

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

print(x\_train)

print(x\_test)