TASK 1 PYTHON CODE

import pandas as pd  *#importing library pandas*

data=pd.read\_csv("netflix\_titles.csv")

*#reading a csv file using pd.read\_csv() function*

print(data.head)  *#printing some of the rows of data*

print(data['show\_id'])  *#gives the variable named show\_id*

print(data.show\_id[2])  *#gives the 2nd element of the variable named show\_id*

print(data.isnull().sum()) *#This will show the number of missing values in each column.*

print(data.isnull().sum().sum()) *#gives the total no of missing values in the dataset*

data['director']=data['director'].fillna('Unknown')

data['cast']=data['cast'].fillna('Unknown')

data['country']=data['country'].fillna('Unknown')

data['date\_added']=data['date\_added'].fillna('Unknown')

data['rating']=data['rating'].fillna('Unknown')

data['duration']=data['duration'].fillna('Unknown') *#fill unknown inplace of missing values*

print(data.isnull().sum().sum()) *#rechecking if there is any missing values*

print(data.shape) *# gives the number of cases and the no. of variables*

data=data.drop\_duplicates() *# remove duplicate rows*

print(data.shape) *#rechecking, no of cases remians the same*

data.reset\_index(drop=True, inplace=True) *# reseting the index*

print(data.shape) *# rechecking we can conclude that, there are no duplicate rows in this dataset*

data['show\_id'] = data['show\_id'].str.title()

data['type'] = data['type'].str.title()

data['title'] = data['title'].str.title()

data['director'] = data['director'].str.title()

data['country'] = data['country'].str.upper()

data['date\_added'] = pd.to\_datetime(data['date\_added'], errors='coerce') *# standardising all the values*

print(data.date\_added[2])

data['date\_added'] = data['date\_added'].dt.strftime('%d-%m-%Y')

print(data.date\_added[2]) *# converting date formats to dd-mm-yyyy format. It is only possible to do this, after converting the date to a standard python form*

print(data.columns) *#printing column names*

data.columns = data.columns.str.strip().str.lower().str.replace(' ', '\_')

*#modifying the column names to have no spaces and all lower case*

print(data.columns) *#rechecking*

print(data.dtypes) *#shows the datatype for each column*

print(data['date\_added'].dtype)  *# should be datetime64[ns], but here the data type is object*

data['date\_added'] = pd.to\_datetime(data['date\_added'], format='%d %m, %Y', errors='coerce')

print(data['date\_added'].dtype)

*#used format option to specify the datatype of date\_added*

OUTPUT

[Running] python -u "c:\Users\Varsha\OneDrive\Desktop\Traffic-Sign-Recognition-main\tempCodeRunnerFile.python"

<bound method NDFrame.head of      show\_id  ...                                        description

0         s1  ...  As her father nears the end of his life, filmm...

1         s2  ...  After crossing paths at a party, a Cape Town t...

2         s3  ...  To protect his family from a powerful drug lor...

3         s4  ...  Feuds, flirtations and toilet talk go down amo...

4         s5  ...  In a city of coaching centers known to train I...

...      ...  ...                                                ...

8802   s8803  ...  A political cartoonist, a crime reporter and a...

8803   s8804  ...  While living alone in a spooky town, a young g...

8804   s8805  ...  Looking to survive in a world taken over by zo...

8805   s8806  ...  Dragged from civilian life, a former superhero...

8806   s8807  ...  A scrappy but poor boy worms his way into a ty...

[8807 rows x 12 columns]>

0          s1

1          s2

2          s3

3          s4

4          s5

        ...

8802    s8803

8803    s8804

8804    s8805

8805    s8806

8806    s8807

Name: show\_id, Length: 8807, dtype: object

s3

show\_id            0

type               0

title              0

director        2634

cast             825

country          831

date\_added        10

release\_year       0

rating             4

duration           3

listed\_in          0

description        0

dtype: int64

4307

0

(8807, 12)

(8807, 12)

(8807, 12)

2021-09-24 00:00:00

24-09-2021

Index(['show\_id', 'type', 'title', 'director', 'cast', 'country', 'date\_added',

       'release\_year', 'rating', 'duration', 'listed\_in', 'description'],

      dtype='object')

Index(['show\_id', 'type', 'title', 'director', 'cast', 'country', 'date\_added',

       'release\_year', 'rating', 'duration', 'listed\_in', 'description'],

      dtype='object')

show\_id         object

type            object

title           object

director        object

cast            object

country         object

date\_added      object

release\_year     int64

rating          object

duration        object

listed\_in       object

description     object

dtype: object

object

datetime64[ns]

[Done] exited with code=0 in 1.809 seconds