PHASE-3 SUBMISSION

APPLIED DATA SCIENCE

IMDB SCORE PREDICTION

Preprocess data in Python – Step by step:

1. Load data in Pandas.

2. Drop columns that aren’t useful.

3. Drop rows with missing values.

4.Finding its tendency

5. Take care of missing values

6.Dimension

1.Load data in Pandas:

To work on the data, you can either load the CSV in Excel or in Pandas. For the purposes of this tutorial, we’ll load the CSV data in Pandas.

import pandas as pd

import numpy as np

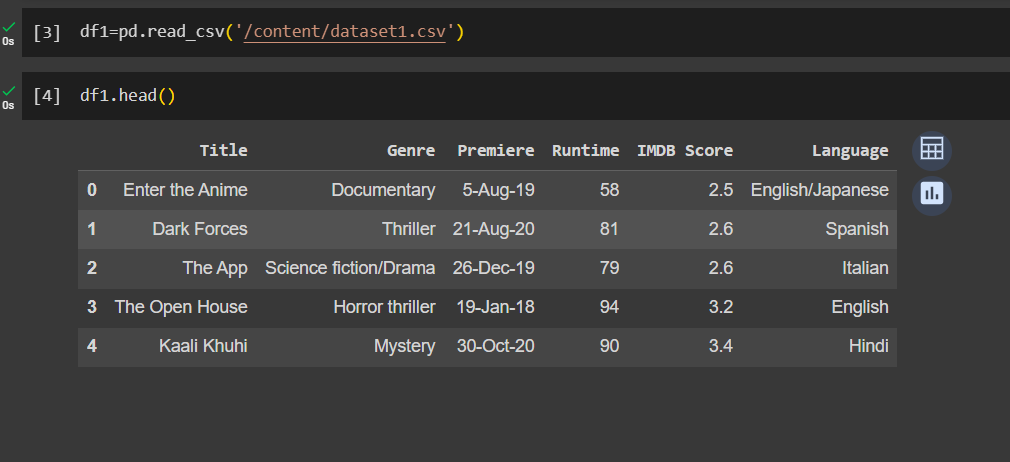
import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

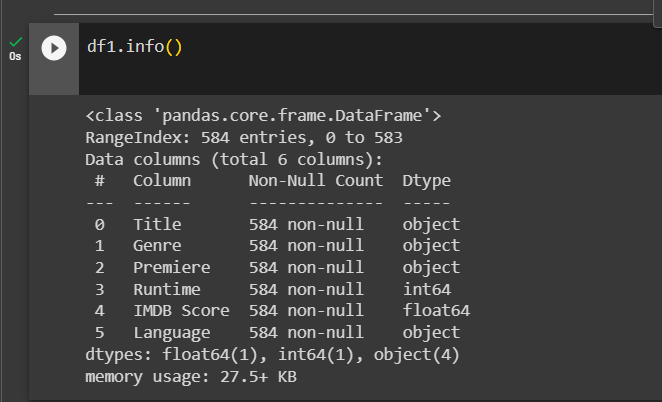
df=pd.read\_csv(‘Data.csv’)

df.head()



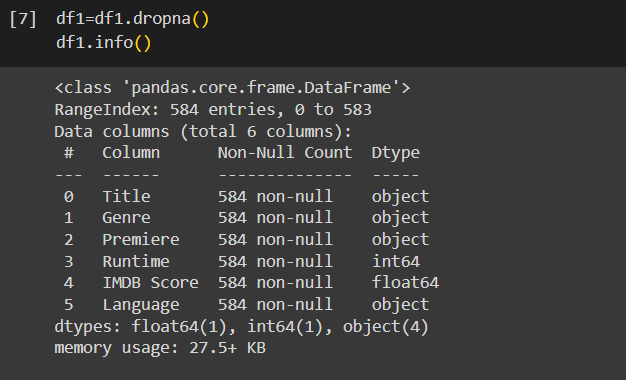
2. Drop Rows With Missing Values:

Next we can drop all rows in the data that have missing values (NaNs). Here’s how:



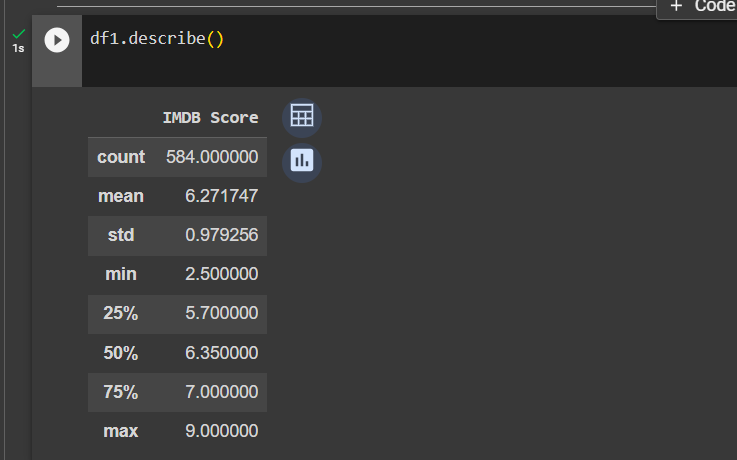
3. Drop Rows With Missing Values:

Next we can drop all rows in the data that have missing values (NaNs). Here’s how:



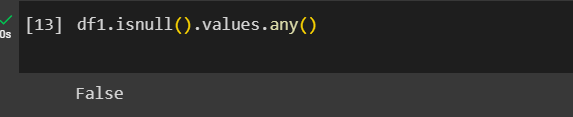
4.describe():

By using you get a overview of the central tendency and spread of the data in each column which can be very helpful for initial data exploration and understanding the characteristics of your dataset



5. isnull() :

This function is used to find the null values .if null value is present then it returns true else false.



6.shape:

It is used to retrieve the dimension of a dataset.it returns a tuple containing the number of rows and columns in the dataset



7.isnull().sum():

By using this we can count the number of missing values in each column

