**Task 1 – Prime**

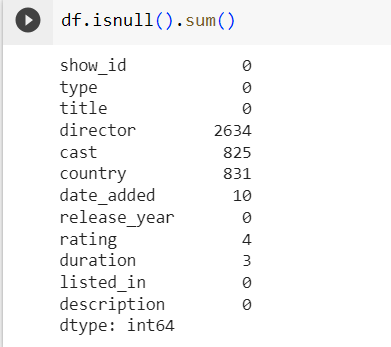
**Introduction**

In today's digital age, the entertainment industry has undergone a significant transformation, with the emergence of online streaming platforms reshaping how we consume movies and television shows. This document aims to provide an overview of the Movie dataset and the insights obtained through the Exploratory Data Analysis (EDA) process.

**EDA Process**

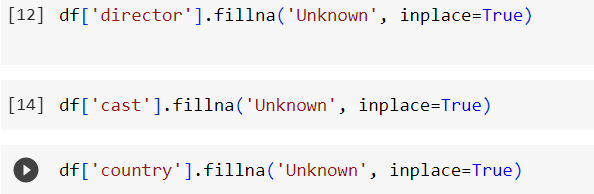
Exploratory Data Analysis (EDA) is a foundational process in data analysis and statistics. It is the initial step in understanding and gaining insights from a dataset. EDA involves the systematic examination of data using various statistical and visualization techniques to identify patterns, relationships, anomalies, and trends within the data.

* Step 1: In the first step we checked the data frame for null values



We have 2634 null values found in the director column following that we have 825 null values found in a cast column, 831 null values in a country column, 10 null values in the date\_added column, and 3 null values in the duration column

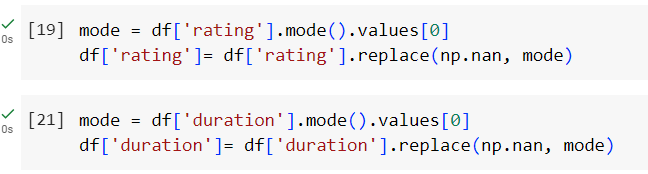
* Step 2: The columns (director, cast, country) have a large number of null values, so we change the null values to unknown because the columns are categorical data we cannot use mean, mode to replace the values.



* Step 3: The column date\_added has a small number of null values so we delete that row because the null values are less so if we drop that It doesn’t affect the insights.



* Step 4: The column (rating, duration) has a small number of null values and it was numerical data so we use the mode to replace the null values with the most repeated value in that column.



Now we have finished the data cleaning process