الغاية من الجلسة :

بداية بناء و عرض بعض توابع التفعيل

التعرف على مكتبة pandas التي تتعامل مع البيانات المهيكلة (مثال الجلسة csv)

Python implementation of Activation function for Neural Network:

An activation function in a neural network is a mathematical function applied to the output of a neuron or a layer of neurons. The purpose of an activation function is to introduce non-linear transformations to the network's computations. Without activation functions, the network would be limited to performing only linear transformations.

Commonly used activation functions include the sigmoid, tanh, ReLU (Rectified Linear Unit), Leaky ReLU, and softmax. Each activation function has its own properties and characteristics, making it suitable for different types of problems and architectures.

Let's go thorugh the definition of each Activation functions and implement them in Python code.

Sigmoid Function:

The sigmoid activation function is often used in ***binary classification*** tasks to produce ***probabilities*** that an input belongs to a certain class.

it has an S-shaped curve and is defined as **f(x) = 1 / (1 + e^(-x))**

In [90]:
```
import numpy as np

def sigmoid(x):
    return 1 / (1 + np.exp(-x))
```

In [91]:
```
print(sigmoid(2))
```

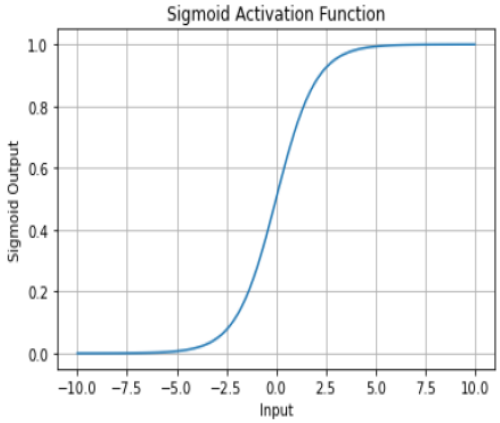
In [92]:
```
#plot Sigmoid:

import matplotlib.pyplot as plt

def plot_sigmoid():
    x = np.linspace(-10, 10, 100)  # Generate 100 equally spaced values from -10
    y = sigmoid(x) # Compute the sigmoid function values

    plt.plot(x, y)
    plt.xlabel('Input')
    plt.ylabel('Sigmoid Output')
    plt.title('Sigmoid Activation Function')
    plt.grid(True)
    plt.show()
```

In [93]:
```
plot_sigmoid()
```



Tanh Function:

It is useful for capturing nonlinear relationships and is often used in ***recurrent neural networks (RNNs)*** and ***hidden layers*** of feedforward neural networks.

it has an S-shaped curve and is defined as **f(x) = (e^(x) - e^(-x)) / (e^(x) + e^(-x))**

In [94]:
```
def tanh(x):
    return (np.exp(x) - np.exp(-x)) / (np.exp(x) + np.exp(-x))
```

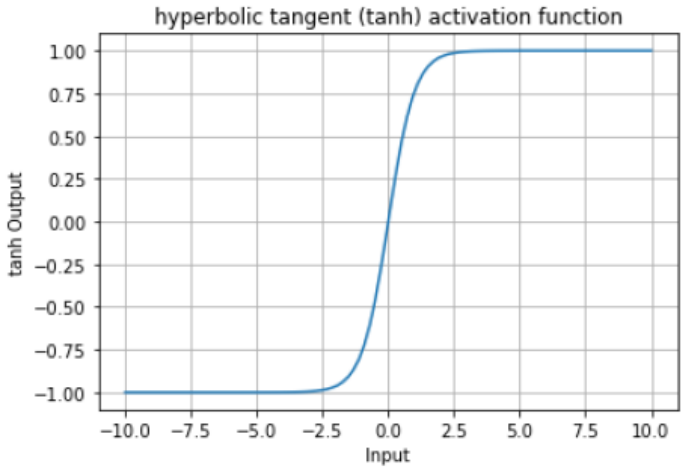

In [95]:
```
tanh(-60)
```

Out[95]: -1.0

[98]:
```
def plot_tanh():
    x = np.linspace(-10, 10, 100)  # Generate 100 equally spaced values from -10
    y = tanh(x) # Compute the tanh function values

    plt.plot(x, y)
    plt.xlabel('Input')
    plt.ylabel('tanh Output')
    plt.title('hyperbolic tangent (tanh) activation function')
    plt.grid(True)
    plt.show()
```

[99]:
```
plot_tanh()
```



Relu Function:

The Rectified Linear Unit (ReLU) activation function is a commonly used mathematical function in neural networks, It is often used in ***deep learning architectures***.

The ReLU function can be defined as **f(x) = max(0, x)**

In [100]:
```
def relu(x):
    return max(0, x)
```

In [101]:
```
relu(-10)
```

Out[101]: 0

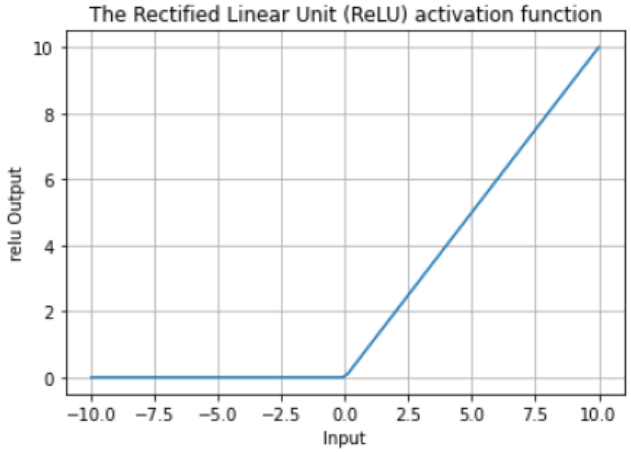
In [102]:
```
relu(9)
```

Out[102]: 9

[n [103]:
```
def plot_relu():
    x = np.linspace(-10, 10, 100)  # Generate 100 equally spaced values from -10
    y = []
    for i in x:
        y.append(relu(i)) # Compute the relu function values

    plt.plot(x, y)
    plt.xlabel('Input')
    plt.ylabel('relu Output')
    plt.title('The Rectified Linear Unit (ReLU) activation function')
    plt.grid(True)
    plt.show()
```

[n [104]:
```
plot_relu()
```



Data Analysis using Pandas:

Pandas library is a popular open-source data analysis and manipulation tool in python. It provides data structures and functions for efficiently handling and analyzing structured data, such as numerical tables and time series data.

Pandas is extremely useful and comes handy when want to load data from various file formats like *csv, excel, json* etc.

Analysis of IMDB-Movie-Dataset:

We will use 'IMDB-Movie-Data' opensource data to learn more about pandas. This dataset

contains **7787 rows** of data in **12 columns**. It contains data about movies and shows on netflix.

For this analysis we will load data from a csv file.

1. understanding the data:

1.1. Import pandas package

In [105]:
```
import pandas as pd
import numpy as np
```

1.2. Read dataset

In [106]:
```
# Read data from .csv file
data = pd.read_csv('IMDB-Movie-Data.csv')

# Read data with specified explicit index.
# We will use this later in our analysis
data_indexed = pd.read_csv('IMDB-Movie-Data.csv', index_col="Title")
```

1.3. Explore the data

In [107]:
```
#Lets first understand the basic information about this data

data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Rank                1000 non-null   int64
 1   Title               1000 non-null   object
 2   Genre               1000 non-null   object
 3   Description         1000 non-null   object
 4   Director            1000 non-null   object
 5   Actors              1000 non-null   object
 6   Year                1000 non-null   int64
 7   Runtime (Minutes)   1000 non-null   int64
 8   Rating              1000 non-null   float64
 9   Votes               1000 non-null   int64
 10  Revenue (Millions)  872 non-null    float64
 11  Metascore           936 non-null    float64
dtypes: float64(3), int64(4), object(5)
memory usage: 93.9+ KB
```

As you can see, the info( ) method gives all the details about this dataframe like columns, number of observations and the datatype of these columns.

info( ) method comes in handy when we first explore the data

1.4. Preview of the data

Let's do a quick preview of the data by using head( ) and tail( ) methods

- head( ) ---> returns the top 5 rows in the dataset by default head( ) can also take the number of rows to be viewed as a parameter
- tail( ) ---> returns the bottom 5 rows in the dataset by default tail( ) can also take the number of rows as an optional parameter

In [108]:
```
# Preview top 5 rows using head()
data.head()
```

Out[108]:

	Rank	Title	Genre	Description	Director	Actors	Year	(
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	

In [109]:
```
# preview bottom 5 rows using tail()
data.tail()
```

Out[109]:

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015	111
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	98
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	93
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87

The column names in the data can be viewed by using 'columns'

In [110]:
```
data.columns
```

Out[110]:
```
Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
       'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
       'Metascore'],
      dtype='object')
```

The shape of the dataset can be viewed by using 'shape'

In [111]:
```
data.shape
```

Out[111]:
```
(1000, 12)
```

. shape function tells us that there are 1000 rows and 12 columns in the dataset

and the maximum rating is 9.0

- The maximum revenue earned by a movie is 936.6 millions

```
In [113]: # check null values
          data.isnull().sum()
```

```
Out[113]: Rank                    0
          Title                   0
          Genre                   0
          Description             0
          Director                0
          Actors                  0
          Year                    0
          Runtime (Minutes)       0
          Rating                  0
          Votes                   0
          Revenue (Millions)    128
          Metascore              64
          dtype: int64
```

isnull() function shows how many missing values are in the dataset with respect to different

2. Data Selection - Indexing and Slicing

Extracting data from Pandas column is similar to Series. The column label is used to extract data from the column

2.1. Extracting data by columns

```
In [114]: # Extract data as series
          genre = data['Genre']
```

```
In [115]: genre
```

```
Out[115]: 0       Action,Adventure,Sci-Fi
          1      Adventure,Mystery,Sci-Fi
          2               Horror,Thriller
          3       Animation,Comedy,Family
          4      Action,Adventure,Fantasy
                          ...
          995         Crime,Drama,Mystery
          996                      Horror
          997         Drama,Music,Romance
          998           Adventure,Comedy
          999       Comedy,Family,Fantasy
          Name: Genre, Length: 1000, dtype: object
```

This will return the 'genre' column as a series.

```
In [116]: type(genre)
```

```
Out[116]: pandas.core.series.Series
```

If we want to retrive it as a dataframe, then indexing must be done using double square brackets

In [117]:
```
# Extract data as dataframe
d = data[['Genre']]
```

In [118]:
```
d
```

Out[118]:

	Genre
0	Action,Adventure,Sci-Fi
1	Adventure,Mystery,Sci-Fi
2	Horror,Thriller
3	Animation,Comedy,Family
4	Action,Adventure,Fantasy
...	...
995	Crime,Drama,Mystery
996	Horror
997	Drama,Music,Romance
998	Adventure,Comedy
999	Comedy,Family,Fantasy

1000 rows × 1 columns

In [119]:
```
type(d)
```

Out[119]:
```
pandas.core.frame.DataFrame
```

In [120]:
```
d.shape  #Dataframe is a 2D array
```

Out[120]:
```
(1000, 1)
```

Extract coulumns from data

If we want to extract multiple columns from the data, simply add the column names to the list

In [122]:
```
some_cols = data[['Title','Genre','Actors','Director','Rating']]
```

In [123]:
```
some_cols.head()
```

Out[123]:

	Title	Genre	Actors	Director	Rating
0	Guardians of the Galaxy	Action,Adventure,Sci-Fi	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	James Gunn	8.1
1	Prometheus	Adventure,Mystery,Sci-Fi	Noomi Rapace, Logan Marshall-Green, Michael Fa...	Ridley Scott	7.0
2	Split	Horror,Thriller	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	M. Night Shyamalan	7.3
3	Sing	Animation,Comedy,Family	Matthew McConaughey,Reese Witherspoon, Seth Ma...	Christophe Lourdelet	7.2
4	Suicide Squad	Action,Adventure,Fantasy	Will Smith, Jared Leto, Margot Robbie, Viola D...	David Ayer	6.2

2.2. Extract data by row index

We already have a dataset with 'title' as index and we can subset that dataframe using loc

If we want to slice data from specific rows, **loc** and **iloc** can be used

loc

- Here indexing is done based on explicit index - locates by name.
- It can take string indexes to retrive data from the specified rows

iloc

- Here indexing is done based on Python's numerical index - locates by integer index.
- This works only with integer indexes to retrieve data from specified rows

In [124]:
```
data_indexed.head()
```

Out[124]:

Title	Rank	Genre	Description	Director	Actors	Year	Ru (Mi
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	
Suicide Squad	5	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	

In [125]:
```
data_indexed.columns
```

Out[125]:
```
Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year',
       'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
       'Metascore'],
      dtype='object')
```

In [126]:
```
data_indexed.shape
```

Out[126]:
```
(1000, 11)
```

In [127]:
```
# Retrieving movie with title 'Suicide Squad'
data_indexed.loc['Suicide Squad']
```

Out[127]:
```
Rank                                                              5
Genre                                          Action,Adventure,Fantasy
Description           A secret government agency recruits some of th...
Director                                                     David Ayer
Actors                Will Smith, Jared Leto, Margot Robbie, Viola D...
Year                                                             2016
Runtime (Minutes)                                                 123
Rating                                                            6.2
Votes                                                          393727
Revenue (Millions)                                             325.02
Metascore                                                          40
Name: Suicide Squad, dtype: object
```

In [128]:
```
data_indexed.loc[['Suicide Squad']][['Genre','Actors','Director','Rating','Revenu
```

Out[128]:

Title	Genre	Actors	Director	Rating	Revenue (Millions)
Suicide Squad	Action,Adventure,Fantasy	Will Smith, Jared Leto, Margot Robbie, Viola D...	David Ayer	6.2	325.02

Retrieving data with row indices

In [129]:
```
# Pick rows from indexes 10-15
data.iloc[10:15]
```

Out[129]:

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)
10	11	Fantastic Beasts and Where to Find Them	Adventure,Family,Fantasy	The adventures of writer Newt Scamander in New...	David Yates	Eddie Redmayne, Katherine Waterston, Alison Su...	2016	133
11	12	Hidden Figures	Biography,Drama,History	The story of a team of female African-American...	Theodore Melfi	Taraji P. Henson, Octavia Spencer, Janelle Mon...	2016	127
12	13	Rogue One	Action,Adventure,Sci-Fi	The Rebel Alliance makes a risky move to steal...	Gareth Edwards	Felicity Jones, Diego Luna, Alan Tudyk, Donnie...	2016	133
13	14	Moana	Animation,Adventure,Comedy	In Ancient Polynesia, when a terrible curse in...	Ron Clements	Auli'i Cravalho, Dwayne Johnson, Rachel House,...	2016	107
14	15	Colossal	Action,Comedy,Drama	Gloria is an out-of-work party girl forced to ...	Nacho Vigalondo	Anne Hathaway, Jason Sudeikis, Austin Stowell,...	2016	109

In [130]:
```
data.iloc[10:15][['Title','Rating','Revenue (Millions)']]
```

Out[130]:

	Title	Rating	Revenue (Millions)
10	Fantastic Beasts and Where to Find Them	7.5	234.02
11	Hidden Figures	7.8	169.27
12	Rogue One	7.9	532.17
13	Moana	7.7	248.75
14	Colossal	6.4	2.87

3. Data Selection - based on Conditional Filtering

3.1. select data based on condition

Let's try to pick the 'Title','Actors','Director' for movies where the rating is minimum

In [131]:
```
rating_min = np.where(data['Rating'] == min(data['Rating']))
data.iloc[rating_min][['Title','Actors','Director','Year','Rating']]
```

Out[131]:

	Title	Actors	Director	Year	Rating
829	Disaster Movie	Carmen Electra, Vanessa Lachey,Nicole Parker, ...	Jason Friedberg	2008	1.9

3.2.Get unique values

In [132]:
```
# Get unique values in Director column
data['Director'].unique()
```

```
'Christopher Nolan', 'Scott Derrickson', 'Antoine Fuqua',
'Patrick Read Johnson', 'Greg Tiernan', 'Barry Jenkins',
'Shawn Burkett', 'John Lee Hancock', 'Ricardo de Montreuil',
'Rob Marshall', 'John Madden', 'Scott Hicks', 'Justin Lin',
'Sean Penn', 'J.J. Abrams', 'Anna Foerster', 'Garry Marshall',
'Chad Stahelski', 'Martin Scorsese', 'Fede Alvarez',
'Thea Sharrock', 'Lone Scherfig', 'Clint Eastwood', 'Zack Snyder',
'André Øvredal', 'Tate Taylor', 'Sam Taylor-Johnson',
'Matthew Vaughn', 'Peter Berg', 'George Miller', 'Robin Swicord',
'Terry George', 'Robert Zemeckis', 'J.A. Bayona', 'David Frankel',
'Byron Howard', 'Gore Verbinski', 'Joss Whedon',
```

In [133]:
```
data['Director'].value_counts()
```

Out[133]:
```
Ridley Scott          8
Paul W.S. Anderson    6
Michael Bay           6
M. Night Shyamalan    6
David Yates           6
                     ..
Gus Van Sant          1
Josh Gordon           1
Chad Stahelski        1
Kimberly Peirce       1
Breck Eisner          1
Name: Director, Length: 644, dtype: int64
```

==========

4. Dealing with missing values

Pandas has isnull( ) and notnull( ) for detecting null values in a dataframe

Lets see how to use these

In [134]:
```
# To see if there are null values in the whole data
data.isnull()
```

Out[134]:

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False

In [135]:
```
# To check null values row-wise
data.isnull().sum()
```

Out[135]:
```
Rank                    0
Title                   0
Genre                   0
Description             0
Director                0
Actors                  0
Year                    0
Runtime (Minutes)       0
Rating                  0
Votes                   0
Revenue (Millions)    128
Metascore              64
dtype: int64
```

Here we know that 'Revenue (Millions)' and 'Metascore' are two columns where there are null values

As we have seen null values in data, we can either choose to drop those or impute these values

Dropping null values

1. Dropping can be done either by rows or by columns depending on our need
2. drop( ), dropna( ) are some functions used to drop null values
3. When we drop values, either the complete column or row is dropped and not the single values

Let's say we want to drop the rows where Metascore is null

In [136]:
```
# Drops all rows containing missing data
data.dropna()
```

Out[136]:

	Rank	Title	Genre	Description	Director	Actors	Year
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016
...	...	...	...	...	...	...	...
993	994	Resident Evil: Afterlife	Action,Adventure,Horror	While still out to destroy the evil Umbrella C...	Paul W.S. Anderson	Milla Jovovich, Ali Larter, Wentworth Miller,K...	2010
				3 high school seniors		Thomas Mann, Oliver	

In [137]:
```
# Drop all columns containing missing data
data.dropna(axis=1)
```

Out[137]:

	Rank	Title	Genre	Description	Director	Actors	Year
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016
...	...	...	...	...	...	...	...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015

Here we can clearly see that 2 columns with missing data are dropped

we can use **thresh** parameter to specify the minimum number of non null values for the column/row to be held without dropping

In [139]:
```
# Use drop function to drop columns
data.drop('Metascore', axis=1).head()
```

Out[139]:

	Rank	Title	Genre	Description	Director	Actors	Year	(
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	

Here we can see that 'Metascore' column is dropped completely.

We know that there are some records where the Revenue is null.

We can impute these null values with mean Revenue,

fillna( ) --> function used to fill null values with specified values

In [140]:
```
revenue_mean = data['Revenue (Millions)'].mean()
print("The mean revenue is: ", revenue_mean)
```

```
The mean revenue is:  82.95637614678897
```

In [141]:
```
# We can fill the null values with this mean revenue
data['Revenue (Millions)'].fillna(revenue_mean, inplace=True)
```

inplace = True signifies that the changes be made permanently in the dataset.

In [142]:
```
data.isnull().sum()
```

Out[142]:
```
Rank                   0
Title                  0
Genre                  0
Description            0
Director               0
Actors                 0
Year                   0
Runtime (Minutes)      0
Rating                 0
Votes                  0
Revenue (Millions)     0
Metascore             64
dtype: int64
```