

## ASSIGNMENT/TASK 5

---

**NAME - SHWETA JHA**

**REG. ID.:GO\_STP\_102574**

---

### **Pandas - Data Analysis of IMDB movies data**

"As we have a basic understanding of the different data structures in Pandas, let's explore the fun and interesting 'IMDB-movies-dataset' and get our hands dirty by performing practical data analysis on real data."

It is an open-source dataset and you can download it from this [link](#).

**(Q.)Load the IMDb Dataset and read**

**(ANS.)**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("/content/IMDB-Movie-Data.csv")
```

**(Q.)View the dataset**

**ANS.**

```
df
```



```
df.columns
```

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

following

```
df.corr()
```

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Rank	1.000000	-0.261605	-0.221739	-0.219555	-0.283876	-0.271592	-0.1918
Year	-0.261605	1.000000	-0.164900	-0.211219	-0.411904	-0.126790	-0.0793
Runtime (Minutes)	-0.221739	-0.164900	1.000000	0.392214	0.407062	0.267953	0.2119
Rating	-0.219555	-0.211219	0.392214	1.000000	0.511537	0.217654	0.6318
Votes	-0.283876	-0.411904	0.407062	0.511537	1.000000	0.639661	0.3256
Revenue (Millions)	-0.271592	-0.126790	0.267953	0.217654	0.639661	1.000000	0.1423

```
df.dtypes
```

```
Rank          int64
Title         object
Genre         object
Description    object
Director      object
Actors        object
Year          int64
Runtime (Minutes)  int64
Rating        float64
Votes         int64
Revenue (Millions) float64
Metascore     float64
dtype: object
```

```
df.axes
```

```
[RangeIndex(start=0, stop=1000, step=1),
 Index(['Rank', 'Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
      'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
      'Metascore'],
      dtype='object')]
```

```
df.ndim
```

2

df.shape

(1000, 12)

df.describe()

	Rank	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)
<b>count</b>	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	872.000000
<b>mean</b>	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.956376
<b>std</b>	288.819436	3.205962	18.810908	0.945429	1.887626e+05	103.253540
<b>min</b>	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.000000
<b>25%</b>	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	13.270000
<b>50%</b>	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	47.985000
<b>75%</b>	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	113.715000
<b>max</b>	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.630000

df.head(50).mean()

```

Rank                25.500000
Year                2015.540000
Runtime (Minutes)   117.100000
Rating              6.872000
Votes              170516.520000
Revenue (Millions)  139.630465
Metascore           62.545455
dtype: float64

```

df.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

```

```
df.Year.value_counts()
```

```

2016    297
2015    127
2014     98
2013     91
2012     64
2011     63
2010     60
2007     53
2008     52
2009     51
2006     44
Name: Year, dtype: int64

```

```
df.Genre.value_counts()
```

```

Action,Adventure,Sci-Fi    50
Drama                     48
Comedy,Drama,Romance       35
Comedy                     32
Drama,Romance              31
..
Romance,Sci-Fi,Thriller     1
Adventure,Horror,Mystery    1
Fantasy,Horror,Thriller     1
Thriller,War                1
Horror,Mystery,Sci-Fi       1
Name: Genre, Length: 207, dtype: int64

```

#### (Q.4)Data Selection – Indexing and Slicing data.

**ANS.**

```
df['Votes']
```

```

0      757074
1      485820
2      157606
3       60545
4      393727
...
995     27585
996     73152
997     70699
998      4881
999     12435
Name: Votes, Length: 1000, dtype: int64

```

```
df[['Rank', 'Votes']]
```

	Rank	Votes
0	1	757074
1	2	485820
2	3	157606
3	4	60545
4	5	393727
...	...	...
995	996	27585
996	997	73152
997	998	70699
998	999	4881
999	1000	12435

1000 rows × 2 columns

```
df[20:40]
```

Rank		Title		Genre	Description	Director	
20	21	Gold	Adventure,Drama,Thriller		Kenny Wells, a prospector desperate for a luck...	Stephen Gaghan	McCo Edgar Bry
	22	Manchester by the Sea		Drama	A depressed uncle is asked to take care of his...	Kenneth Lonergan	Case Michelle Kyle C
	23	Hounds of Love	Crime,Drama,Horror		A cold-blooded predatory couple while cruising...	Ben Young	Emn Cu
	24	Trolls		Animation,Adventure,Comedy	After the Bergens invade Troll Village, Poppy,...	Walt Dohrn	Anna Timberla D
	25	Independence Day: Resurgence	Action,Adventure,Sci-Fi		Two decades after the first Independence Day i...	Roland Emmerich	Liam He Jeff C Bill Pullr
25	26	Paris pieds nus		Comedy	Fiona visits Paris for the first time to	Dominique Abel	Fiona D Abel Em

df.loc[:, "Rank" : "Director"]

df.iloc[:, 0:5]

Rank		Title	Genre	Description	Director
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer
...	...	...	...	...	...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth
		Star Wars: The Force Awakens	Adventure,Sci-Fi	Resistance leader Rey joins forces with Han Solo...	J.J. Abrams

Data Selection – Based on Conditional filtering

df.iloc[10:20,[0, 3, 4, 5]]

Rank		Description		Director	Actors
10	11	The adventures of writer Newt Scamander in New...		David Yates	Eddie Redmayne, Katherine Waterston, Alison Su...
11	12	The story of a team of female African-American...		Theodore Melfi	Taraji P. Henson, Octavia Spencer, Janelle Mon...
12	13	The Rebel Alliance makes a risky move to steal...		Gareth Edwards	Felicity Jones, Diego Luna, Alan Tudyk, Donnie...
13	14	In Ancient Polynesia, when a terrible curse in...		Ron Clements	Auli'i Cravalho, Dwayne Johnson, Rachel House,...
14	15	Gloria is an out-of-work party girl forced to ...		Nacho Vigalondo	Anne Hathaway, Jason Sudeikis, Austin Stowell,...
15	16	The quiet life of a terrier named Max is upend...		Chris Renaud	Louis C.K., Eric Stonestreet, Kevin Hart, Lake...
16	17	WWII American Army Medic Desmond T. Doss, who...		Mel Gibson	Andrew Garfield, Sam Worthington, Luke Bracey



```
df.query('Rating>7')
```

	Rank	Title	Genre	Description	Director	Actors
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vir Diesel, Bradley Cooper, Zoe S...
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..
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma..
6	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress i...	Damien Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J...
8	9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering on British explor...	James Gray	Charlie Hunnam Robert Pattinson Sienna Mille..
...	...	...	...	...	...	..
979	980	The Skin I Live In	Drama,Thriller	A brilliant plastic surgeon, haunted by past t...	Pedro Almodóvar	Antonio Banderas Elena Anaya, Jar Cornet,Mari..
982	983	Across the Universe	Drama,Fantasy,Musical	The music of the Beatles and the Vietnam War	Julie Taymor	Evan Rachel Wood Jim Sturgess, Joe Anderson, ..

```
df.where(df.Metascore>60.0)
```

	Rank	Title	Genre	Description	Director	Actors	Year	R (Mi
0	1.0	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...	2014.0	
1	2.0	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012.0	
2	3.0	Split	Horror Thriller	Three girls are kidnapped by	M. Night	James McAvoy, Anya Taylor	2016.0	

```
df['Votes']>400000
df[df['Votes']>400000]
```

	Rank	Title	Genre	Description	Director	Actor
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vi Diesel, Bradle Cooper, Zoe S.
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace Logan Marshal Green, Michael Fa.

(Q.6)Groupby operations

ANS.

```
df_Rank =df.groupby('Rank')
df_Rank.mean()
```

	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Rank						
1	2014	121	8.1	757074	333.13	76.0
2	2012	124	7.0	485820	126.46	65.0
3	2016	117	7.3	157606	138.12	62.0
4	2016	108	7.2	60545	270.32	59.0
5	2016	123	6.2	393727	325.02	40.0
...	...	...	...	...	...	...
996	2015	111	6.2	27585	NaN	45.0
997	2007	94	5.5	73152	17.54	46.0
998	2008	98	6.2	70699	58.01	50.0
999	2014	93	5.6	4881	NaN	22.0
1000	2016	87	5.3	12435	19.64	11.0

1000 rows × 6 columns

```
gp = df.groupby('Actors')
gp[['Rating']].mean()
```

	Rating
Actors	
Aamir Khan, Anushka Sharma, Sanjay Dutt,Boman Irani	8.2
Aamir Khan, Madhavan, Mona Singh, Sharman Joshi	8.4
Aamir Khan, Sakshi Tanwar, Fatima Sana Shaikh,Sanya Malhotra	8.8
Aaron Paul, Dominic Cooper, Imogen Poots, Scott Mescudi	6.5
Aaron Poole, Kenneth Welsh,Daniel Fathers, Kathleen Munroe	5.8
...	...
Zachary Gordon, Robert Capron, Rachael Harris,Steve Zahn	6.2
Zoey Deutch, James Franco, Tangie Ambrose,Cedric the Entertainer	6.3
Zoey Deutch, Lucy Fry, Danila Kozlovsky, Gabriel Byrne	5.6
Zoey Deutch, Nicholas Braun, Mateo Arias, Israel Broussard	6.1

(Q.7)Sorting operation

996 rows x 7 columns

```
df_sorted=df.sort_values(by='Director')
df_sorted.head()
```

Rank	Title	Genre	Description	Director	Actors	Year
991	Taare Zameen Par	Drama,Family,Music	An eight-year-old boy is thought to be a lazy ...	Aamir Khan	Darsheel Safary, Aamir Khan, Tanay Chheda, Sac...	2007
311	La vie d'Adèle	Drama,Romance	Adèle's life is changed when she meets Emma, a...	Abdellatif Kechiche	Léa Seydoux, Adèle Exarchopoulos, Salim Kechio...	2013
783	Tramps	Comedy,Romance	A young man and woman find love in an unlikely...	Adam Leon	Callum Turner, Grace Van Patten, Michal Vondel...	2016
			Four		Christian Bale	

```
df.groupby('Rank',sort=False)[['Votes']].mean()
```

	Votes
Rank	
1	757074
2	485820
3	157606
4	60545
5	393727
...	...
996	27585
997	73152
998	70699
999	4881

```
df[df['Votes']>40000].groupby('Director')[['Votes']].count()
```

	Votes
Director	
Aamir Khan	1
Abdellatif Kechiche	1
Adam McKay	4
Adam Shankman	2
Adam Wingard	1
...	...
William Brent Bell	1
Woody Allen	5
Yimou Zhang	1
Yorgos Lanthimos	2
Zack Snyder	5

430 rows × 1 columns

### (Q.8) Dealing with missing values

ANS>

```
df.isnull().sum()
```

```

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

```

```
df.dropna().sum()
```

```

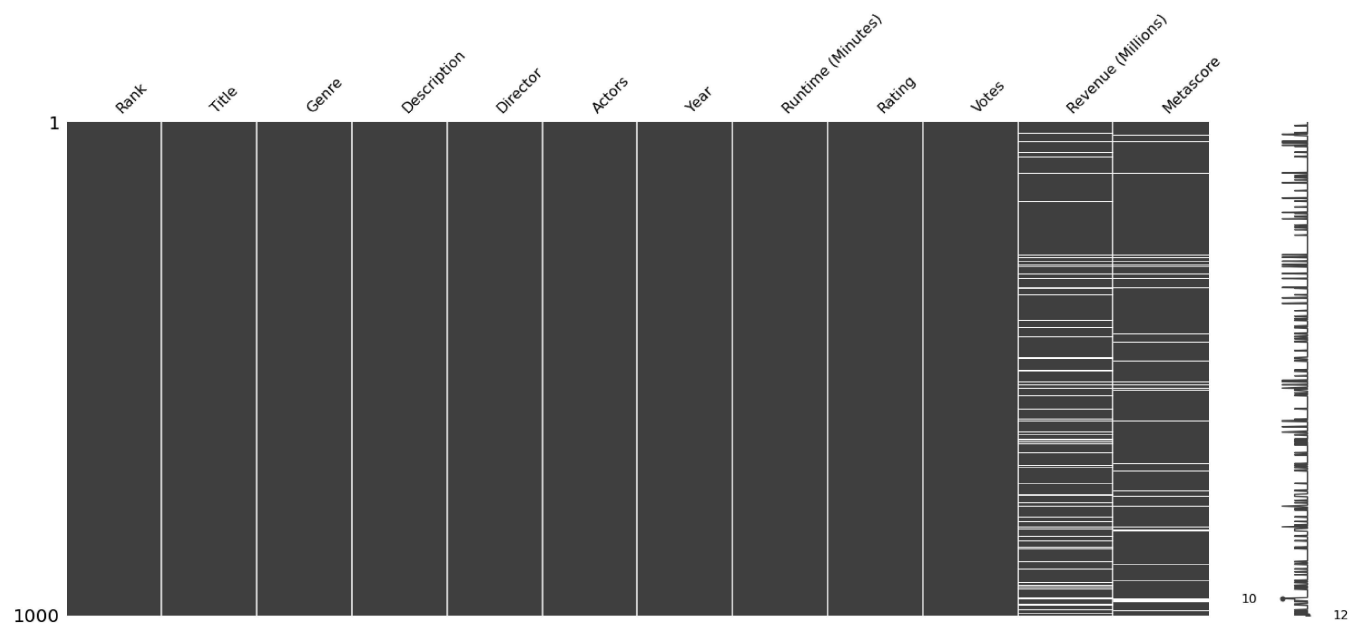
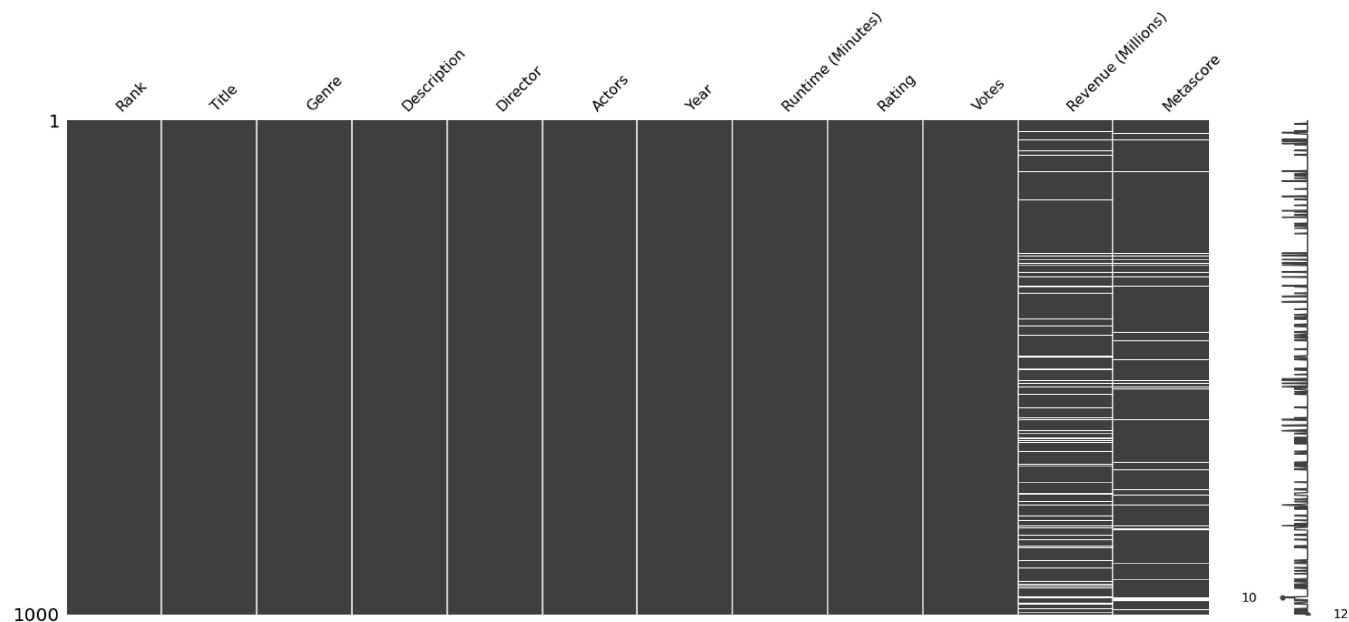
Rank          406637
Title         Guardians of the GalaxyPrometheusSplitSingSuic...
Genre         Action,Adventure,Sci-FiAdventure,Mystery,Sci-F...
Description    A group of intergalactic criminals are forced ...
Director      James GunnRidley ScottM. Night ShyamalanChrist...
Actors        Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...
Year          1686481
Runtime (Minutes)  96067
Rating        5710.4
Votes         161926951
Revenue (Millions)  70865.1
Metascore     49924
dtype: object

```

```

import missingno as msno
msno.matrix(df)
plt.show()

```



### (Q-9) Dropping columns and null values

ANS.

```
df.drop('Description',axis=1,inplace=True)
df.head()
```

	Rank	Title	Genre	Director	Actors	Year	Runtime (Minute)
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	1
					Neomi Ran...		

Dropping Null values

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

Rank	0
Title	0
Genre	0
Director	0
Actors	0
Year	0
Runtime (Minutes)	0
Rating	0
Votes	0
Revenue (Millions)	0
Metascore	0
dtype: int64	

```
msno.matrix(df)
plt.show()
```



	Rank	Title	Genre	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1											

(Q>10)Apply( ) functions

ANS.

```
def sub9(value):
    return value - 9
df['Revenue (Millions)'].apply(sub9)
```

0	324.13
1	117.46
2	129.12
3	261.32
4	316.02
	...
993	51.13
994	45.72
996	8.54
997	49.01
999	10.64

Name: Revenue (Millions), Length: 838, dtype: float64

