

▼ GO_STP_5856 - Ashwin S

▼ Assignment-5

```
import pandas as pd
import numpy as np
import missingno as msno
```

▼ 1) Load the IMDb Dataset and read

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

▼ 2) View the dataset

```
df.head()
```

	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, V Diesel, Bradle Cooper, Zoe S.
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapaci Logan Marsha Green, Michael Fa.
2	3	Split	Horror,Thriller	Three girls are kidnapped by	M. Night Shyamalan	James McAvoy, Any Taylor-Joy, Haley L

▼ 3) Understand some basic information about the dataset and Inspect the dataframe Inspect the dataframe's columns, shapes, variable types etc.

```
df.columns
```

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

```
df.shape
```

```
(1000, 12)
```

```
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
```

▼ 4) Data Selection – Indexing and Slicing data

```
df["Title"]
```

```
0      Guardians of the Galaxy
1      Prometheus
2      Split
3      Sing
4      Suicide Squad
...
995    Secret in Their Eyes
996    Hostel: Part II
997    Step Up 2: The Streets
998    Search Party
999    Nine Lives
Name: Title, Length: 1000, dtype: object
```

```
# Slicing columnss in data frame
```

```
df1 = df.iloc[:,0:2]
```

```
# data frame after slicing
```

```
df1
```

	Rank	Title
0	1	Guardians of the Galaxy
1	2	Prometheus
2	3	Split
3	4	Sing
4	5	Suicide Squad

▼ 5) Data Selection – Based on Conditional filtering

```
df2 = df[df["Rating"]>8.0]
df2.head()
```

	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...	2017
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....	2016
16	17	Hacksaw Ridge	Biography,Drama,History	WWII American Army Medic Desmond T.	Mel Gibson	Andrew Garfield, Sam Worthington, Luke Bracey	2016

▼ 6) Groupby operations

```
df = pd.DataFrame({'Animal': ['Falcon', 'Falcon', 'Parrot', 'Parrot'], 'Max Speed': [380., 370., 24., 26.]})
df
```

	Animal	Max Speed
0	Falcon	380.0
1	Falcon	370.0
2	Parrot	24.0
3	Parrot	26.0

```
df.groupby(['Animal']).mean()
```

Max Speed

Animal

▼ 7) Sorting operation

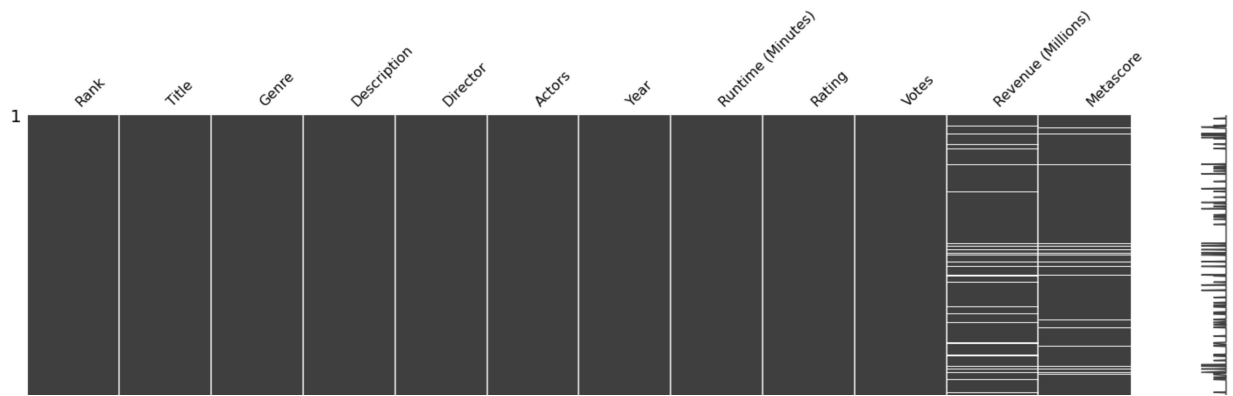
```
import numpy as np
```

```
unsorted_df = pd.DataFrame({'col1':[2,1,1,1], 'col2':[1,3,2,4]})  
sorted_df = unsorted_df.sort_values(by=['col1', 'col2'])  
sorted_df
```

	col1	col2
2	1	2
1	1	3
3	1	4
0	2	1

▼ 8) Dealing with missing values

```
df3 = pd.read_csv("/content/IMDB-Movie-Data.csv")  
_ = msno.matrix(df3)
```



```
mean_value1=df3['Revenue (Millions)'].mean()
mean_value2=df3['Metascore'].mean()
df3['Revenue (Millions)'].fillna(value=mean_value1, inplace=True)
df3['Metascore'].fillna(value=mean_value2, inplace=True)
df3.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) 0
Metascore 0
dtype: int64
```

▼ 9) Dropping columns and null values

```
dict = {'First Score':[100, 90, 34, 95],
        'Second Score': [30, 45, 56, 28],
        'Third Score':[np.nan, 40, 80, 98]}
```

```
# creating a dataframe from dictionary
df = pd.DataFrame(dict)
df.dropna()
```

	First Score	Second Score	Third Score
1	90	45	40.0
2	34	56	80.0
3	95	28	98.0

- 10) Apply() functions

```
import pandas as pd
```

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

def fun(num):

    if num<5:
        return "Low"

    elif num>= 5 and num<8:
        return "Normal"

    else:
        return "High"

new = s["Rating"].apply(fun)
del s["Rating"]
frames = [s,new]
v1 = pd.concat(frames,axis=1)
v1.head()
```

	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, V Diesel, Bradle Cooper, Zoe S.
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapaci Logan Marsha Green, Michael Fa.
2	3	Split	Horror,Thriller	Three girls are kidnapped by	M. Night Shyamalan	James McAvoy, Any Taylor-Joy, Haley L

Double-click (or enter) to edit

✓ 0s completed at 6:29 PM

● ✕