EDS Theory Activity no.1

DATASET: MOVIE REVIEW

Name: Ragini Dattatray Telange

Roll No: CS2-55

PRN: 202401040067

Google Collab link:

https://colab.research.google.com/drive/1ySm7KAKGy_e7C5kUW2FiHmXYIHn_3AfB

Dataset:

Rank	Title	Genre	Director	Year	Runtime (M	Review	Votes
2	Promethei	Adventure,	Ridley Scot	2012	124	7	485820
3	Split	Horror, Thr	M. Night Sk	2016	117	7.3	157606
4	Sing	Animation,	Christophe	2016	108	7.2	60545
5	Suicide Sq	Action, Adv	David Ayer	2016	123	6.2	393727
6	The Great \	Action, Adv	Yimou Zha	2016	103	6.1	56036
7	La La Land	Comedy, D	Damien Ch	2016	128	8.3	258682
8	Mindhorn	Comedy	Sean Foley	2016	89	6.4	2490
9	The Lost Ci	Action, Adv	James Gra	2016	141	7.1	7188
10	Passenger	Adventure,	Morten Tyl	2016	116	7	192177
12	Hidden Fig	Biography,	Theodore N	2016	127	7.8	93103
13	Rogue One	Action, Adv	Gareth Edv	2016	133	7.9	323118
14	Moana	Animation,	Ron Cleme	2016	107	7.7	118151
15	Colossal	Action, Cor	Nacho Viga	2016	109	6.4	8612
16	The Secret	Animation,	Chris Rena	2016	87	6.6	120259
17	Hacksaw F	Biography,	Mel Gibsor	2016	139	8.2	211760
18	Jason Bour	Action, Thr	Paul Greer	2016	123	6.7	150823
19	Lion	Biography,	Garth Davi	2016	118	8.1	102061
20	Arrival	Drama, My:	Denis Ville	2016	116	8	340798

TWENTY QUESTIONS BASED ON THIS DATASET AND THEIR PYTHON PROGRAM:

1. How to Dispaly Dataset ?

```python
import pandas as pd

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

```
    Rank                    Title                      Genre  \
0      2               Prometheus     Adventure,Mystery,Sci-Fi
1      3                    Split            Horror,Thriller
2      4                     Sing     Animation,Comedy,Family
3      5             Suicide Squad    Action,Adventure,Fantasy
4      6           The Great Wall    Action,Adventure,Fantasy
5      7                La La Land          Comedy,Drama,Music
6      8                 Mindhorn                      Comedy
7      9        The Lost City of Z   Action,Adventure,Biography
8     10               Passengers     Adventure,Drama,Romance
9     12            Hidden Figures      Biography,Drama,History
10    13                Rogue One      Action,Adventure,Sci-Fi
11    14                    Moana    Animation,Adventure,Comedy
12    15                 Colossal         Action,Comedy,Drama
13    16     The Secret Life of Pets  Animation,Adventure,Comedy
14    17             Hacksaw Ridge     Biography,Drama,History
15    18              Jason Bourne            Action,Thriller
16    19                     Lion             Biography,Drama
17    20                   Arrival       Drama,Mystery,Sci-Fi

             Director  Year  Runtime (Minutes)  Review    Votes
0         Ridley Scott  2012                124     7.0   485820
1    M. Night Shyamalan  2016                117     7.3   157606
2   Christophe Lourdelet  2016                108     7.2    60545
3          David Ayer   2016                123     6.2   393727
```

✓ 0s completed at 10:01 PM

---

1. What are the column names?

```python
print(df.columns.tolist())
```

```
['Rank', 'Title', 'Genre', 'Director', 'Year', 'Runtime (Minutes)', 'Review', 'Votes']
```

---

2. How many movies are there?

```python
print(len(df))
```

```
18
```

3. Which movie has the highest rating?

```python
print(df[df['Review'] == df['Review'].max()])
```

```
   Rank      Title             Genre         Director  Year  \
5     7  La La Land  Comedy,Drama,Music  Damien Chazelle  2016

   Runtime (Minutes)  Review   Votes
5               128     8.3  258682
```

4. Which movie has the lowest rating?

```python
print(df[df['Review'] == df['Review'].min()])
```

```
   Rank           Title                    Genre        Director  Year  \
4     6  The Great Wall  Action,Adventure,Fantasy  Yimou Zhang  2016

   Runtime (Minutes)  Review  Votes
4               103     6.1  56036
```

5. What is the average movie rating?

```python
print(df['Review'].mean())
```

```
7.222222222222222
```

6.List all movies directed by 'Ridley Scott'.

```
[7]  print(df[df['Director'] == 'Ridley Scott'])
```

```
   Rank      Title                  Genre     Director  Year  \
0     2  Prometheus  Adventure,Mystery,Sci-Fi  Ridley Scott  2012

   Runtime (Minutes)  Review   Votes
0                124     7.0  485820
```

7. What is the most common genre?

```
from collections import Counter
genres = ','.join(df['Genre']).split(',')
print(Counter(genres).most_common(1))
```

```
[('Adventure', 8)]
```

8. Which year had the most movies?

```
print(df['Year'].value_counts().idxmax())
```

```
2016
```

Files

Analyze your files with code written by Gemini   Upload

..
sample_data
IMDB-Movie-Data.csv

[ ] 9. What is the average runtime of all movies?

```python
print(df['Runtime (Minutes)'].mean())
```

117.11111111111111

Files

Analyze your files with code written by Gemini   Upload

..
sample_data
IMDB-Movie-Data.csv

10. Which movie has the longest runtime?

```python
print(df[df['Runtime (Minutes)'] == df['Runtime (Minutes)'].max()])
```

```
   Rank          Title                  Genre    Director  Year  \
7     9  The Lost City of Z  Action,Adventure,Biography  James Gray  2016

   Runtime (Minutes)  Review  Votes
7                141     7.1   7188
```

Files

Analyze your files with code written by Gemini   Upload

..
sample_data
IMDB-Movie-Data.csv

11. Which movie has the shortest runtime?

```python
print(df[df['Runtime (Minutes)'] == df['Runtime (Minutes)'].min()])
```

```
    Rank               Title                      Genre       Director  \
13    16  The Secret Life of Pets  Animation,Adventure,Comedy  Chris Renaud

    Year  Runtime (Minutes)  Review   Votes
13  2016                 87     6.6  120259
```

Commands   + Code   + Text

Files

Analyze your files with code written by Gemini   Upload

📁 ..
▶ 📁 sample_data
📄 IMDB-Movie-Data.csv

12. How many unique directors are there?

```python
print(df['Director'].nunique())
```

18

Commands   + Code   + Text

Files

Analyze your files with code written by Gemini   Upload

📁 ..
▶ 📁 sample_data
📄 IMDB-Movie-Data.csv

13. What is the name of the movie with the highest number of votes?

```python
print(df[df['Votes'] == df['Votes'].max()]['Title'])
```

```
0    Prometheus
Name: Title, dtype: object
```

Commands   + Code   + Text

Files

Analyze your files with code written by Gemini   Upload

📁 ..
▶ 📁 sample_data
📄 IMDB-Movie-Data.csv

14. How many votes did the most-voted movie get?

```python
print(df['Votes'].max())
```

485820

15. List movies with a rating greater than 8.

```python
print(df[df['Review'] > 8])
```

```
    Rank          Title                     Genre         Director  Year  \
5      7    La La Land       Comedy,Drama,Music  Damien Chazelle  2016
14    17  Hacksaw Ridge  Biography,Drama,History       Mel Gibson  2016
16    19          Lion          Biography,Drama      Garth Davis  2016

    Runtime (Minutes)  Review    Votes
5                 128     8.3   258682
14                139     8.2   211760
16                118     8.1   102061
```

16. List movies that are of genre "Action".

```python
print(df[df['Genre'].str.contains('Action')])
```

```
    Rank             Title                    Genre         Director  \
3      5      Suicide Squad  Action,Adventure,Fantasy       David Ayer
4      6     The Great Wall  Action,Adventure,Fantasy     Yimou Zhang
7      9  The Lost City of Z  Action,Adventure,Biography     James Gray
10    13          Rogue One     Action,Adventure,Sci-Fi  Gareth Edwards
12    15           Colossal       Action,Comedy,Drama  Nacho Vigalondo
15    18        Jason Bourne           Action,Thriller  Paul Greengrass

    Year  Runtime (Minutes)  Review    Votes
3   2016                123     6.2   393727
4   2016                103     6.1    56036
7   2016                141     7.1     7188
10  2016                133     7.9   323118
12  2016                109     6.4     8612
15  2016                123     6.7   150823
```

17. What is the total number of votes across all movies?

```python
print(df['Votes'].sum())
```

```
3082956
```

Files
Analyze your files with code written by Gemini   Upload
..
sample_data
IMDB-Movie-Data.csv

18.What is the average runtime of movies per year?

```python
print(df.groupby('Year')['Runtime (Minutes)'].mean())
```

```
Year
2012    124.000000
2016    116.705882
Name: Runtime (Minutes), dtype: float64
```

Files
Analyze your files with code written by Gemini   Upload
..
sample_data
IMDB-Movie-Data.csv

19.Sort movies by number of votes, lowest first.

```python
print(df.sort_values(by='Votes', ascending=True))
```

```
    Rank                 Title                      Genre  \
6      8              Mindhorn                     Comedy
7      9     The Lost City of Z  Action,Adventure,Biography
12    15              Colossal         Action,Comedy,Drama
4      6        The Great Wall    Action,Adventure,Fantasy
2      4                  Sing      Animation,Comedy,Family
9     12        Hidden Figures      Biography,Drama,History
16    19                  Lion              Biography,Drama
11    14                 Moana    Animation,Adventure,Comedy
13    16   The Secret Life of Pets  Animation,Adventure,Comedy
15    18          Jason Bourne            Action,Thriller
1      3                 Split             Horror,Thriller
8     10            Passengers    Adventure,Drama,Romance
14    17          Hacksaw Ridge      Biography,Drama,History
5      7             La La Land          Comedy,Drama,Music
10    13              Rogue One       Action,Adventure,Sci-Fi
17    20               Arrival        Drama,Mystery,Sci-Fi
3      5          Suicide Squad    Action,Adventure,Fantasy
0      2            Prometheus      Adventure,Mystery,Sci-Fi

              Director  Year  Runtime (Minutes)  Review   Votes
6           Sean Foley  2016                 89     6.4    2490
7           James Gray  2016                141     7.1    7188
12     Nacho Vigalondo  2016                109     6.4    8612
4          Yimou Zhang  2016                103     6.1   56036
2    Christophe Lourdelet  2016            108     7.2   60545
9       Theodore Melfi  2016                127     7.8   93103
```

Disk   70.74 GB available

Files
Analyze your files with code written by Gemini   Upload
..
sample_data
IMDB-Movie-Data.csv

20.Find movies with runtime greater than 2 hours (120 minutes).

```python
print(df[df['Runtime (Minutes)'] > 120])
```

```
    Rank                 Title                      Genre           Director  \
0      2            Prometheus      Adventure,Mystery,Sci-Fi      Ridley Scott
3      5          Suicide Squad    Action,Adventure,Fantasy        David Ayer
5      7             La La Land          Comedy,Drama,Music   Damien Chazelle
7      9     The Lost City of Z  Action,Adventure,Biography        James Gray
9     12        Hidden Figures      Biography,Drama,History    Theodore Melfi
10    13              Rogue One       Action,Adventure,Sci-Fi   Gareth Edwards
14    17          Hacksaw Ridge      Biography,Drama,History        Mel Gibson
15    18          Jason Bourne            Action,Thriller   Paul Greengrass

     Year  Runtime (Minutes)  Review   Votes
0    2012                124     7.0  485820
3    2016                123     6.2  393727
5    2016                128     8.3  258682
7    2016                141     7.1    7188
9    2016                127     7.8   93103
10   2016                133     7.9  323118
14   2016                139     8.2  211760
15   2016                123     6.7  150823
```