task-1

June 28, 2024

```
[1]: from google.colab import drive
     drive.mount('/content/drive')
```

    Mounted at /content/drive

```
[2]: import pandas as pd
     import numpy as nm
     from sklearn.model_selection import train_test_split as ttp
     from sklearn.metrics import classification_report
     from sklearn.linear_model import LinearRegression
     LR= LinearRegression()
     import re
     import string
     import matplotlib.pyplot as plt
```

```
[3]: df = pd.read_csv("/content/drive/MyDrive/imdb dataset/imdb_kaggle.csv")
```

```
[5]: df.head(5)
```

```
[5]:    rank  year duration age_limit  rating numberof_ratings  Metascore  \
     0     1  1994    2h 22m        15     9.3           (2.9M)       82.0
     1     2  1972    2h 55m        15     9.2             (2M)      100.0
     2     3  2008    2h 32m       12A     9.0           (2.9M)       84.0
     3     4  2003    3h 21m       12A     9.0             (2M)       94.0
     4     5  1993    3h 15m        15     9.0           (1.5M)       95.0

                                           description  \
     0  Over the course of several years, two convicts…
     1  The aging patriarch of an organized crime dyna…
     2  When the menace known as the Joker wreaks havo…
     3  Gandalf and Aragorn lead the World of Men agai…
     4  In German-occupied Poland during World War II,…

                                          name
     0              The Shawshank Redemption
     1                         The Godfather
     2                       The Dark Knight
     3   The Lord of the Rings: The Return of the King
```

```
4                              Schindler's List
```

[6]: `df.shape`

[6]: `(1000, 9)`

[13]: `df=df.drop(['description','Metascore'],axis=1)`

[14]: `df.head(5)`

[14]:
```
   rank  year duration age_limit  rating numberof_ratings  \
0     1  1994   2h 22m        15     9.3           (2.9M)
1     2  1972   2h 55m        15     9.2             (2M)
2     3  2008   2h 32m       12A     9.0           (2.9M)
3     4  2003   3h 21m       12A     9.0             (2M)
4     5  1993   3h 15m        15     9.0           (1.5M)


                                           name
0                     The Shawshank Redemption
1                               The Godfather
2                             The Dark Knight
3   The Lord of the Rings: The Return of the King
4                            Schindler's List
```

[15]: `df1=df.copy()`

[16]: `df1=df1.drop(['rank','duration','numberof_ratings'],axis=1)`

[18]:
```
df1=df1.dropna()
df1.shape
df1.head(25)
```

[18]:
```
    year  age_limit  rating                                            name
0   1994         15     9.3                      The Shawshank Redemption
1   1972         15     9.2                                 The Godfather
2   2008        12A     9.0                               The Dark Knight
3   2003        12A     9.0     The Lord of the Rings: The Return of the King
4   1993         15     9.0                              Schindler's List
5   1957          U     9.0                                  12 Angry Men
6   1974          X     9.0                          The Godfather Part II
7   2001         PG     8.9    The Lord of the Rings: The Fellowship of the …
8   1994         18     8.9                                   Pulp Fiction
9   2023        12A     8.9                                     12th Fail
10  2010        12A     8.8                                     Inception
11  1999         15     8.8                                    Fight Club
12  1994         12     8.8                                  Forrest Gump
13  2002        12A     8.8            The Lord of the Rings: The Two Towers
```

```
14  1966          18  8.8                  The Good, the Bad and the Ugly
15  2014         12A  8.7                                     Interstellar
16  1999          15  8.7                                       The Matrix
17  1990          18  8.7                                        GoodFellas
18  1975          15  8.7                    One Flew Over the Cuckoo's Nest
19  1980           U  8.7  Star Wars: Episode V - The Empire Strikes Back
20  2021   Not Rated  8.7                                          Jai Bhim
21  2020          15  8.7                                   Soorarai Pottru
22  2022         12A  8.7                                        777 Charlie
23  2022         12A  8.7                          Rocketry: The Nambi Effect
24  2024         12A  8.6                                    Dune: Part Two
```

[ ]: