Movie Review

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
movies = pd.read_csv("C:\\Users\\Snigdha\\OneDrive\\Desktop\\vysh\\movies.
print(movies.head())
```

```
C:\Users\Snigdha\AppData\Local\Temp\ipykernel_30036\954740050.py:3: Pars
erWarning: Falling back to the 'python' engine because the 'c' engine do
es not support regex separators (separators > 1 char and different from
'\s+' are interpreted as regex); you can avoid this warning by specifyin
g engine='python'.
  movies = pd.read_csv("C:\\Users\\Snigdha\\OneDrive\\Desktop\\vysh\\mov
ies.dat", delimiter='::')

   0000008      Edison Kinetoscopic Record of a Sneeze (1894)  \
0       10                 La sortie des usines Lumière (1895)
1       12                       The Arrival of a Train (1896)
2       25  The Oxford and Cambridge University Boat Race ...
3       91                           Le manoir du diable (1896)
4      131                             Une nuit terrible (1896)

     Documentary|Short
0    Documentary|Short
1    Documentary|Short
2                  NaN
3         Short|Horror
4  Short|Comedy|Horror
```

Defining Columns

In [2]:
```
movies.columns = ["ID", "Title", "Genre"]
print(movies.head())
```

```
    ID                                              Title
Genre
0   10                La sortie des usines Lumière (1895)    Documentary
|Short
1   12                      The Arrival of a Train (1896)    Documentary
|Short
2   25  The Oxford and Cambridge University Boat Race ...
NaN
3   91                          Le manoir du diable (1896)         Short|
Horror
4  131                            Une nuit terrible (1896)  Short|Comedy|
Horror
```

Importing Rating Dataset

In [4]:
```
ratings = pd.read_csv("C:\\Users\\Snigdha\\OneDrive\\Desktop\\vysh\\rating
print(ratings.head())
```

```
C:\Users\Snigdha\AppData\Local\Temp\ipykernel_30036\2497519373.py:1: Par
serWarning: Falling back to the 'python' engine because the 'c' engine d
oes not support regex separators (separators > 1 char and different from
'\s+' are interpreted as regex); you can avoid this warning by specifyin
g engine='python'.
  ratings = pd.read_csv("C:\\Users\\Snigdha\\OneDrive\\Desktop\\vysh\\ra
tings.dat", delimiter='::')

   1  0114508  8  1381006850
0  2   499549  9  1376753198
1  2  1305591  8  1376742507
2  2  1428538  1  1371307089
3  3    75314  1  1595468524
4  3   102926  9  1590148016
```

Defining Columns

In [5]:
```
ratings.columns = ["User", "ID", "Ratings", "Timestamp"]
print(ratings.head())
```

```
   User       ID  Ratings   Timestamp
0     2   499549        9  1376753198
1     2  1305591        8  1376742507
2     2  1428538        1  1371307089
3     3    75314        1  1595468524
4     3   102926        9  1590148016
```

Merging Two Datasets

In [6]:
```
data = pd.merge(movies, ratings, on=["ID", "ID"])
print(data.head())
```

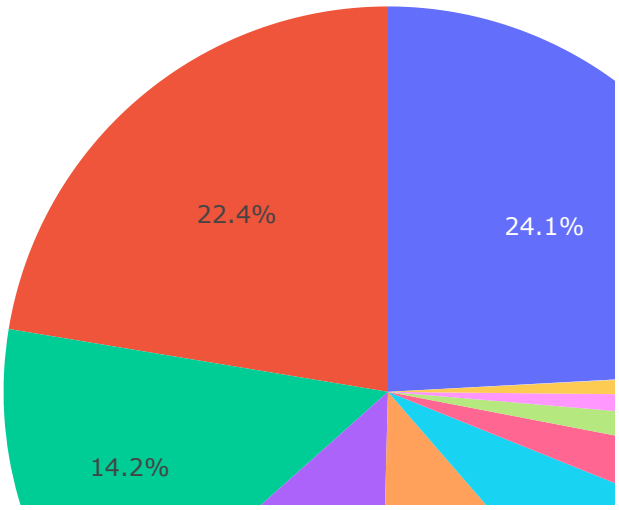
```
   ID                                                Title                Ge
nre  \
0  10                La sortie des usines Lumière (1895)  Documentary|Sh
ort
1  12                      The Arrival of a Train (1896)  Documentary|Sh
ort
2  25  The Oxford and Cambridge University Boat Race ...
NaN
3  91                          Le manoir du diable (1896)       Short|Hor
ror
4  91                          Le manoir du diable (1896)       Short|Hor
ror

    User  Ratings   Timestamp
0  70577       10  1412878553
1  69535       10  1439248579
2  37628        8  1488189899
3   5814        6  1385233195
4  37239        5  1532347349
```

Plots

In [7]:

```
ratings = data["Ratings"].value_counts()
numbers = ratings.index
quantity = ratings.values
import plotly.express as px
fig = px.pie(data, values=quantity, names=numbers)
fig.show()
```



Highest Rated Movies(10)

In [8]:
```
data2 = data.query("Ratings == 10")
print(data2["Title"].value_counts().head(10))
```

```
Joker (2019)                        1479
Interstellar (2014)                 1386
1917 (2019)                          820
Avengers: Endgame (2019)             812
The Shawshank Redemption (1994)      707
Gravity (2013)                       653
The Wolf of Wall Street (2013)       581
Hacksaw Ridge (2016)                 570
Avengers: Infinity War (2018)        535
La La Land (2016)                    510
Name: Title, dtype: int64
```

So, according to this dataset, Joker (2019) got the highest number of 10 ratings from viewers. This is how you can analyze movie ratings using Python as a data science beginner.

In [ ]:
```

```