

```
!pip install numpy

Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (1.23.5)
```

```
import pandas as pd
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive
```

```
path_to_csv_file = '/content/drive/MyDrive/movies.csv'
df = pd.read_csv(path_to_csv_file)

df.head()
```

| Unnamed: 0 | Movie Name               | Year of Release | Watch Time | Movie Rating | Metascore of movie | Gross  | Votes     | Description                                       |
|------------|--------------------------|-----------------|------------|--------------|--------------------|--------|-----------|---------------------------------------------------|
| 0          | The Shawshank Redemption | 1994            | 142        | 9.3          | 82.0               | 28.34  | 27,77,378 | Over the course of several years, two convicts... |
| 1          | The Godfather            | 1972            | 175        | 9.2          | 100.0              | 134.97 | 19,33,588 | Don Vito Corleone, head of a mafia family, dec... |

```
# Load the dataset into a pandas DataFrame.
import pandas as pd

try:
    path_to_csv_file = '/content/drive/MyDrive/movies.csv'
    df = pd.read_csv(path_to_csv_file)
except FileNotFoundError:
    print("File 'movies.csv' not found.")
    exit(1)

# Display the first few rows of the DataFrame to check the data
print(df.head())
```

| Unnamed: 0 | Movie Name               | Year of Release | Watch Time | Movie Rating | Metascore of movie | Gross  | Votes     | Description                                        |
|------------|--------------------------|-----------------|------------|--------------|--------------------|--------|-----------|----------------------------------------------------|
| 0          | The Shawshank Redemption | 1994            | 142        | 9.3          | 82.0               | 28.34  | 27,77,378 | Over the course of several years, two convicts...  |
| 1          | The Godfather            | 1972            | 175        | 9.2          | 100.0              | 134.97 | 19,33,588 | Don Vito Corleone, head of a mafia family, dec...  |
| 2          | The Dark Knight          | 2008            | 152        | 9.0          | 84.0               | 534.86 | 27,54,087 | When the menace known as the Joker wreaks havoc... |
| 3          | Schindler's List         | 1993            | 195        | 9.0          | 95.0               | 96.9   | 13,97,886 | In German-occupied Poland during World War II,...  |
| 4          | 12 Angry Men             | 1957            | 96         | 9.0          | 97.0               | 4.36   | 8,24,211  | The jury in a New York City murder trial is fr...  |

```
# Preprocess the data by handling any missing values and data types.
missing_values = df.isnull().sum()
print("Missing values:\n", missing_values)
```

|                    |     |
|--------------------|-----|
| Missing values:    |     |
| Unnamed: 0         | 0   |
| Movie Name         | 0   |
| Year of Release    | 0   |
| Watch Time         | 0   |
| Movie Rating       | 0   |
| Metascore of movie | 155 |
| Gross              | 162 |
| Votes              | 0   |
| Description        | 0   |
| dtype: int64       |     |

```
# Calculate the following statistics for the dataset:
# The average movie rating.
```

```
average_rating = df["Movie Rating"].mean()
print("Average Movie Rating:", average_rating)
```

Average Movie Rating: 7.9702

```
import pandas as pd

# Find the year of the highest-grossing movie

if not pd.api.types.is_string_dtype(df["Gross"]):
    # If it's not a string, convert it to string
    df["Gross"] = df["Gross"].astype(str)

df["Gross"] = df["Gross"].str.replace('[\$,]', '', regex=True)
df["Gross"] = pd.to_numeric(df["Gross"], errors='coerce')

highest_grossing_year = df[df["Gross"] == df["Gross"].max()]["Year of Release"].values[0]
print("Year of Highest-Grossing Movie:", highest_grossing_year)
```

Year of Highest-Grossing Movie: 2015

```
# Find the movie with the longest watch time
longest_watch_time_movie = df[df["Watch Time"] == df["Watch Time"].max()]["Movie Name"].values[0]
print("Movie with Longest Watch Time:", longest_watch_time_movie)
```

Movie with Longest Watch Time: Gangs of Wasseypur

```
# Create a new DataFrame containing movies released after the year 2000
import pandas as pd

df["Year of Release"] = df["Year of Release"].str.extract(r'(\d{4})', expand=False)

df["Year of Release"] = pd.to_numeric(df["Year of Release"], errors='coerce')

movies_after_2000 = df[df["Year of Release"] > 2000].copy()
```

```
# Calculate the correlation between "Movie Rating" and "Metascore of Movie"

print(df.columns)
```

```
Index(['Unnamed: 0', 'Movie Name', 'Year of Release', 'Watch Time',
      'Movie Rating', 'Metascore of movie', 'Gross', 'Votes', 'Description'],
      dtype='object')
```

```
# Calculate the correlation between "Movie Rating" and "Metascore of Movie"
print(df.columns)

correlation = df["Movie Rating"].corr(df["Metascore of movie"])
print("Correlation between Movie Rating and Metascore:", correlation)
```

```
Index(['Unnamed: 0', 'Movie Name', 'Year of Release', 'Watch Time',
      'Movie Rating', 'Metascore of movie', 'Gross', 'Votes', 'Description'],
      dtype='object')
Correlation between Movie Rating and Metascore: 0.25696983698437564
```

```
# Find the top 5 most popular movies based on the number of votes
top_5_popular_movies = df.sort_values(by="Votes", ascending=False).head(5)
print("Top 5 Most Popular Movies:\n", top_5_popular_movies)
```

↗

| Top 5 Most Popular Movies: |                           |                 |            |   |  |
|----------------------------|---------------------------|-----------------|------------|---|--|
| Unnamed: 0                 | Movie Name                | Year of Release | Watch Time | \ |  |
| 835                        | 835 The Magnificent Seven | 1960            | 128        |   |  |
| 696                        | 696 My Fair Lady          | 1964            | 170        |   |  |
| 849                        | 849 Malcolm X             | 1992            | 202        |   |  |
| 194                        | 194 Andhadhun             | 2018            | 139        |   |  |
| 632                        | 632 The Sandlot           | 1993            | 101        |   |  |

  

| Movie Rating | Metascore of movie | Gross | Votes | \      |  |
|--------------|--------------------|-------|-------|--------|--|
| 835          | 7.7                | 74.0  | 4.91  | 99,132 |  |
| 696          | 7.8                | 95.0  | 72.00 | 98,938 |  |
| 849          | 7.7                | 73.0  | 48.17 | 98,857 |  |
| 194          | 8.2                | NaN   | 1.37  | 98,180 |  |
| 632          | 7.8                | 55.0  | 32.42 | 98,129 |  |

```

                                Description
835 Seven gunfighters are hired by Mexican peasant...
696 In 1910s London, snobbish phonetics professor ...
849 Biographical epic of the controversial and inf...
194 A series of mysterious events change the life ...
632 In the summer of 1962, a new kid in town is ta...
```

```
# Create a new column "Earnings per Minute"
df["Earnings per Minute"] = df["Gross"] / df["Watch Time"]
```

```
print(df)
```

|     | Unnamed: 0 | Movie Name                      | Year of Release | Watch Time | \ |
|-----|------------|---------------------------------|-----------------|------------|---|
| 0   | 0          | The Shawshank Redemption        | 1994            | 142        |   |
| 1   | 1          | The Godfather                   | 1972            | 175        |   |
| 2   | 2          | The Dark Knight                 | 2008            | 152        |   |
| 3   | 3          | Schindler's List                | 1993            | 195        |   |
| 4   | 4          | 12 Angry Men                    | 1957            | 96         |   |
| ..  | ...        | ...                             | ...             | ...        |   |
| 995 | 995        | Philomena                       | 2013            | 98         |   |
| 996 | 996        | Un long dimanche de fiançailles | 2004            | 133        |   |
| 997 | 997        | Shine                           | 1996            | 105        |   |
| 998 | 998        | The Invisible Man               | 1933            | 71         |   |
| 999 | 999        | Celda 211                       | 2009            | 113        |   |

|     | Movie Rating | Metascore of movie | Gross  | Votes     | \ |
|-----|--------------|--------------------|--------|-----------|---|
| 0   | 9.3          | 82.0               | 28.34  | 27,77,378 |   |
| 1   | 9.2          | 100.0              | 134.97 | 19,33,588 |   |
| 2   | 9.0          | 84.0               | 534.86 | 27,54,087 |   |
| 3   | 9.0          | 95.0               | 96.90  | 13,97,886 |   |
| 4   | 9.0          | 97.0               | 4.36   | 8,24,211  |   |
| ..  | ...          | ...                | ...    | ...       |   |
| 995 | 7.6          | 77.0               | 37.71  | 1,02,336  |   |
| 996 | 7.6          | 76.0               | 6.17   | 75,004    |   |
| 997 | 7.6          | 87.0               | 35.81  | 55,589    |   |
| 998 | 7.6          | 87.0               | NaN    | 37,822    |   |
| 999 | 7.6          | NaN                | NaN    | 69,464    |   |

|     | Description                                       | Earnings per Minute |
|-----|---------------------------------------------------|---------------------|
| 0   | Over the course of several years, two convicts... | 0.199577            |
| 1   | Don Vito Corleone, head of a mafia family, dec... | 0.771257            |
| 2   | When the menace known as the Joker wreaks havo... | 3.518816            |
| 3   | In German-occupied Poland during World War II,... | 0.496923            |
| 4   | The jury in a New York City murder trial is fr... | 0.045417            |
| ..  | ...                                               | ...                 |
| 995 | A world-weary political journalist picks up th... | 0.384796            |
| 996 | Tells the story of a young woman's relentless ... | 0.046391            |
| 997 | Pianist David Helfgott, driven by his father a... | 0.341048            |
| 998 | A scientist finds a way of becoming invisible,... | NaN                 |
| 999 | The story of two men on different sides of a p... | NaN                 |

[1000 rows x 10 columns]

```
# Sort the dataset by "Earnings per Minute" in descending order
sorted_df = df.sort_values(by="Earnings per Minute", ascending=False)
```

```
print("Sorted DataFrame:")
print(sorted_df)
```

| Sorted DataFrame: |            |                                            |                    |        |           |
|-------------------|------------|--------------------------------------------|--------------------|--------|-----------|
|                   | Unnamed: 0 | Movie Name                                 | Year of Release    |        | \         |
| 635               | 635        | Star Wars: Episode VII - The Force Awakens | 2015               |        |           |
| 91                | 91         | Top Gun: Maverick                          | 2022               |        |           |
| 142               | 142        | Spider-Man: No Way Home                    | 2021               |        |           |
| 48                | 48         | The Lion King                              | 1994               |        |           |
| 66                | 66         | Avengers: Endgame                          | 2019               |        |           |
| ..                | ...        | ...                                        | ...                |        |           |
| 979               | 979        | Dark Waters                                | 2019               |        |           |
| 983               | 983        | Rebel Without a Cause                      | 1955               |        |           |
| 987               | 987        | Sabrina                                    | 1954               |        |           |
| 998               | 998        | The Invisible Man                          | 1933               |        |           |
| 999               | 999        | Celda 211                                  | 2009               |        |           |
|                   | Watch Time | Movie Rating                               | Metascore of movie | Gross  | Votes \   |
| 635               | 138        | 7.8                                        | 80.0               | 936.66 | 9,52,718  |
| 91                | 130        | 8.3                                        | 78.0               | 718.73 | 6,16,792  |
| 142               | 148        | 8.2                                        | 71.0               | 804.75 | 8,18,919  |
| 48                | 88         | 8.5                                        | 88.0               | 422.78 | 10,97,285 |
| 66                | 181        | 8.4                                        | 78.0               | 858.37 | 12,00,821 |
| ..                | ...        | ...                                        | ...                | ...    | ...       |
| 979               | 126        | 7.6                                        | 73.0               | NaN    | 94,877    |
| 983               | 111        | 7.6                                        | 89.0               | NaN    | 94,956    |
| 987               | 113        | 7.6                                        | 72.0               | NaN    | 68,022    |
| 998               | 71         | 7.6                                        | 87.0               | NaN    | 37,822    |
| 999               | 113        | 7.6                                        | NaN                | NaN    | 69,464    |

|     | Description                                        | Earnings per Minute |
|-----|----------------------------------------------------|---------------------|
| 635 | As a new threat to the galaxy rises, Rey, a de...  | 6.787391            |
| 91  | After thirty years, Maverick is still pushing ...  | 5.528692            |
| 142 | With Spider-Man's identity now revealed, Peter...  | 5.437500            |
| 48  | Lion prince Simba and his father are targeted ...  | 4.804318            |
| 66  | After the devastating events of Avengers: Infi...  | 4.742376            |
| ..  | ...                                                | ...                 |
| 979 | A corporate defense attorney takes on an enviro... | NaN                 |
| 983 | A rebellious young man with a troubled past co...  | NaN                 |
| 987 | A playboy becomes interested in the daughter o...  | NaN                 |
| 998 | A scientist finds a way of becoming invisible,...  | NaN                 |
| 999 | The story of two men on different sides of a p...  | NaN                 |

[1000 rows x 10 columns]

```
# Save the sorted DataFrame to a new CSV file
sorted_df.to_csv("sorted_movies.csv", index=False)
```

```
# Read the sorted DataFrame from the CSV file
sorted_movies_df = pd.read_csv("sorted_movies.csv")
```

```
# Print the sorted DataFrame
print("Sorted Movies DataFrame:")
print(sorted_movies_df)
```

Sorted Movies DataFrame:

| Unnamed: 0 | Movie Name                                 | Year of Release |
|------------|--------------------------------------------|-----------------|
| 0          | Star Wars: Episode VII - The Force Awakens | 2015            |
| 1          | Top Gun: Maverick                          | 2022            |
| 2          | Spider-Man: No Way Home                    | 2021            |
| 3          | The Lion King                              | 1994            |
| 4          | Avengers: Endgame                          | 2019            |
| ..         | ...                                        | ...             |
| 995        | Dark Waters                                | 2019            |
| 996        | Rebel Without a Cause                      | 1955            |
| 997        | Sabrina                                    | 1954            |
| 998        | The Invisible Man                          | 1933            |
| 999        | Celda 211                                  | 2009            |

|     | Watch Time | Movie Rating | Metascore of movie | Gross  | Votes     |
|-----|------------|--------------|--------------------|--------|-----------|
| 0   | 138        | 7.8          | 80.0               | 936.66 | 9,52,718  |
| 1   | 130        | 8.3          | 78.0               | 718.73 | 6,16,792  |
| 2   | 148        | 8.2          | 71.0               | 804.75 | 8,18,919  |
| 3   | 88         | 8.5          | 88.0               | 422.78 | 10,97,285 |
| 4   | 181        | 8.4          | 78.0               | 858.37 | 12,00,821 |
| ..  | ...        | ...          | ...                | ...    | ...       |
| 995 | 126        | 7.6          | 73.0               | NaN    | 94,877    |
| 996 | 111        | 7.6          | 89.0               | NaN    | 94,956    |
| 997 | 113        | 7.6          | 72.0               | NaN    | 68,022    |
| 998 | 71         | 7.6          | 87.0               | NaN    | 37,822    |
| 999 | 113        | 7.6          | NaN                | NaN    | 69,464    |

|     | Description                                        | Earnings per Minute |
|-----|----------------------------------------------------|---------------------|
| 0   | As a new threat to the galaxy rises, Rey, a de...  | 6.787391            |
| 1   | After thirty years, Maverick is still pushing ...  | 5.528692            |
| 2   | With Spider-Man's identity now revealed, Peter...  | 5.437500            |
| 3   | Lion prince Simba and his father are targeted ...  | 4.804318            |
| 4   | After the devastating events of Avengers: Infi...  | 4.742376            |
| ..  | ...                                                | ...                 |
| 995 | A corporate defense attorney takes on an enviro... | NaN                 |
| 996 | A rebellious young man with a troubled past co...  | NaN                 |
| 997 | A playboy becomes interested in the daughter o...  | NaN                 |
| 998 | A scientist finds a way of becoming invisible,...  | NaN                 |
| 999 | The story of two men on different sides of a p...  | NaN                 |

[1000 rows x 10 columns]

✓ 0s completed at 7:44 PM

● ×