project1-4

June 19, 2024

[3]:

```
import numpy as np
import pandas as pd
import matplotlib.pylab as plt
import seaborn as sns
df = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')
df.head(10)
```

[3]:

```
          MovieName   Genre  Rating                   Director  \
0        Mouna Guru  Action     7.7               Santha Kumar   
1       7 Aum Arivu  Action     6.2            A.R. Murugadoss   
2  Vaagai Sooda Vaa  Comedy     8.0                A. Sarkunam   
3          Mankatha  Action     7.6              Venkat Prabhu   
4  Kanchana: Muni 2  Comedy     6.5       Lawrence Raghavendra   
5  Deiva Thirumagal   Drama     8.1                 A.L. Vijay   
6            Vaanam  Action     7.2  Radha Krishna Jagarlamudi   
7                Ko  Action     7.8                 K.V. Anand   
8           Payanam  Action     7.3                Radha Mohan   
9        Yutham Sei   Crime     8.0                    Myshkin   

                  Actor  PeopleVote  Year  Hero_Rating  movie_rating  \
0             Arulnithi         746  2011            8             8   
1                Suriya        9479  2011            9             9   
2                 Vimal       14522  2011            8             7   
3           Ajith Kumar       12276  2011            6             8   
4  Lawrence Raghavendra        1044  2011            8             9   
5                Vikram       44517  2011            9             9   
6     T.R. Silambarasan        1307  2011            7             8   
7                 Jiiva        4759  2011            9             7   
8    Nagarjuna Akkineni         677  2011            6             8   
9                Cheran        1678  2011            4             9   

   content_rating  
0        7.900000  
1        8.066667  
2        7.666667  
3        7.200000  
```

```
4        7.833333
5        8.700000
6        7.400000
7        7.933333
8        7.100000
9        7.000000
```

[36]:

```python
import pandas as pd

df = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')

null_values = df.isnull().sum()

print(null_values)
rows_with_null = df[df.isnull().any(axis=1)]
print(rows_with_null)
```

```
MovieName         0
Genre             0
Rating            0
Director          0
Actor             0
PeopleVote        0
Year              0
Hero_Rating       0
movie_rating      0
content_rating    0
dtype: int64
Empty DataFrame
Columns: [MovieName, Genre, Rating, Director, Actor, PeopleVote, Year,
Hero_Rating, movie_rating, content_rating]
Index: []
```

[22]:

```python
import numpy as np
import pandas as pd
import matplotlib.pylab as plt
import seaborn as sns
data = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')

print(data)
```

```
             MovieName   Genre  Rating               Director  \
0           Mouna Guru  Action     7.7           Santha Kumar
1          7 Aum Arivu  Action     6.2        A.R. Murugadoss
2     Vaagai Sooda Vaa  Comedy     8.0            A. Sarkunam
```

```
3             Mankatha  Action     7.6         Venkat Prabhu
4     Kanchana: Muni 2  Comedy     6.5  Lawrence Raghavendra
..                 …       …     …                    …
324  Dhilluku Dhuddu 2  Comedy     5.3               Rambala
325                Dev  Action     4.8    Rajath Ravishankar
326  Charlie Chaplin 2  Comedy     3.8    Sakthi Chidambaram
327              Petta  Action     7.3      Karthik Subbaraj
328           Viswasam  Action     6.7                  Siva

                    Actor  PeopleVote  Year  Hero_Rating  movie_rating  \
0               Arulnithi         746  2011            8             8
1                  Suriya        9479  2011            9             9
2                   Vimal       14522  2011            8             7
3             Ajith Kumar       12276  2011            6             8
4    Lawrence Raghavendra        1044  2011            8             9
..                    …           …   …          …            …
324             Santhanam         497  2019            7             9
325                Karthi         724  2019            5             8
326           Prabhu Deva         215  2019            4             7
327           Rajinikanth        7545  2019            8             8
328           Ajith Kumar        5907  2019            8             9

     content_rating
0          7.900000
1          8.066667
2          7.666667
3          7.200000
4          7.833333
..              …
324        7.100000
325        5.933333
326        4.933333
327        7.766667
328        7.900000

[329 rows x 10 columns]
```

[2]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 329 entries, 0 to 328
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   MovieName       329 non-null    object
 1   Genre           329 non-null    object
 2   Rating          329 non-null    float64
 3   Director        329 non-null    object
 4   Actor           329 non-null    object
 5   PeopleVote      329 non-null    int64
 6   Year            329 non-null    int64
 7   Hero_Rating     329 non-null    int64
 8   movie_rating    329 non-null    int64
 9   content_rating  329 non-null    float64
dtypes: float64(2), int64(4), object(4)
memory usage: 25.8+ KB
```

[23]:

```
df_rt = df[['MovieName','Genre','PeopleVote','movie_rating']]
df_rt
```

[23]:

```
             MovieName   Genre  PeopleVote  movie_rating
0           Mouna Guru  Action         746             8
1          7 Aum Arivu  Action        9479             9
2      Vaagai Sooda Vaa  Comedy       14522             7
3             Mankatha  Action       12276             8
4     Kanchana: Muni 2  Comedy        1044             9
..                 …       …         …            …
324  Dhilluku Dhuddu 2  Comedy         497             9
325                Dev  Action         724             8
326  Charlie Chaplin 2  Comedy         215             7
327              Petta  Action        7545             8
328           Viswasam  Action        5907             9

[329 rows x 4 columns]
```

[6]:

```
df_rt.describe()
```

[6]:

```
         PeopleVote  movie_rating
count    329.000000    329.000000
mean    7372.607903      8.139818
std    14380.829757      0.760232
min        7.000000      6.000000
25%      455.000000      8.000000
50%     1320.000000      8.000000
75%     5907.000000      9.000000
max    71418.000000     10.000000
```

[7]:

```
df_rt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 329 entries, 0 to 328
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   MovieName     329 non-null    object
 1   Genre         329 non-null    object
 2   PeopleVote    329 non-null    int64
 3   movie_rating  329 non-null    int64
dtypes: int64(2), object(2)
memory usage: 10.4+ KB
```

[22]:

```
import pandas as pd

# Load the dataset
df = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')

# Display the data types of each column
print("Data types of each column:")
print(df.dtypes)

# Select only the numeric columns
numeric_df = df.select_dtypes(include=[float, int])

# Calculate the correlation matrix
correlation_matrix = numeric_df.corr()

# Print the correlation matrix
print("Correlation matrix:")
print(correlation_matrix)
```

```
Data types of each column:
MovieName          object
Genre              object
Rating            float64
Director           object
Actor              object
PeopleVote          int64
Year                int64
Hero_Rating         int64
movie_rating        int64
content_rating    float64
dtype: object
Correlation matrix:
                  Rating  PeopleVote      Year  Hero_Rating  movie_rating  \
Rating          1.000000    0.035188 -0.342470     0.070069     -0.046237
PeopleVote      0.035188    1.000000 -0.079931    -0.040341     -0.077999
Year           -0.342470   -0.079931  1.000000     0.048828     -0.005366
Hero_Rating     0.070069   -0.040341  0.048828     1.000000     -0.027608
movie_rating   -0.046237   -0.077999 -0.005366    -0.027608      1.000000
content_rating  0.586455   -0.034643 -0.151539     0.793341      0.249694

                content_rating
Rating                0.586455
```

```
PeopleVote           -0.034643
Year                 -0.151539
Hero_Rating           0.793341
movie_rating          0.249694
content_rating        1.000000
```

[25]:

```python
toprating = pd.DataFrame(df[df['movie_rating']<= 10 ])
toprating
```

[25]:

```
             MovieName   Genre  Rating              Director  \
0           Mouna Guru  Action     7.7          Santha Kumar
1          7 Aum Arivu  Action     6.2       A.R. Murugadoss
2     Vaagai Sooda Vaa  Comedy     8.0           A. Sarkunam
3             Mankatha  Action     7.6         Venkat Prabhu
4     Kanchana: Muni 2  Comedy     6.5  Lawrence Raghavendra
..                  …     …     …                      …
324  Dhilluku Dhuddu 2  Comedy     5.3               Rambala
325                Dev  Action     4.8    Rajath Ravishankar
326  Charlie Chaplin 2  Comedy     3.8    Sakthi Chidambaram
327              Petta  Action     7.3      Karthik Subbaraj
328           Viswasam  Action     6.7                  Siva

                    Actor  PeopleVote  Year  Hero_Rating  movie_rating  \
0               Arulnithi         746  2011            8             8
1                  Suriya        9479  2011            9             9
2                   Vimal       14522  2011            8             7
3             Ajith Kumar       12276  2011            6             8
4    Lawrence Raghavendra        1044  2011            8             9
..                      …       …   …           …           …
324             Santhanam         497  2019            7             9
325                Karthi         724  2019            5             8
326           Prabhu Deva         215  2019            4             7
327           Rajinikanth        7545  2019            8             8
328           Ajith Kumar        5907  2019            8             9

     content_rating
0          7.900000
1          8.066667
2          7.666667
3          7.200000
4          7.833333
..              …
324        7.100000
325        5.933333
326        4.933333
327        7.766667
328        7.900000
```

```
[329 rows x 10 columns]
```

[28]:

```
toprating = pd.DataFrame(df[df['movie_rating']== 10 ])
toprating
```

[28]:

```
      MovieName   Genre  Rating     Director  Actor  PeopleVote  Year  \
288  Kalavani 2  Comedy     4.0  A. Sarkunam  Vimal          68  2019

     Hero_Rating  movie_rating  content_rating
288            5            10        6.333333
```

[29]:

```
df.columns
```

[29]:

```
Index(['MovieName', 'Genre', 'Rating', 'Director', 'Actor', 'PeopleVote',
       'Year', 'Hero_Rating', 'movie_rating', 'content_rating'],
      dtype='object')
```

[ ]:

```

```

[24]:

```
df_rt = df[['MovieName','Genre','PeopleVote','movie_rating','Director','Year']]
df_rt
```

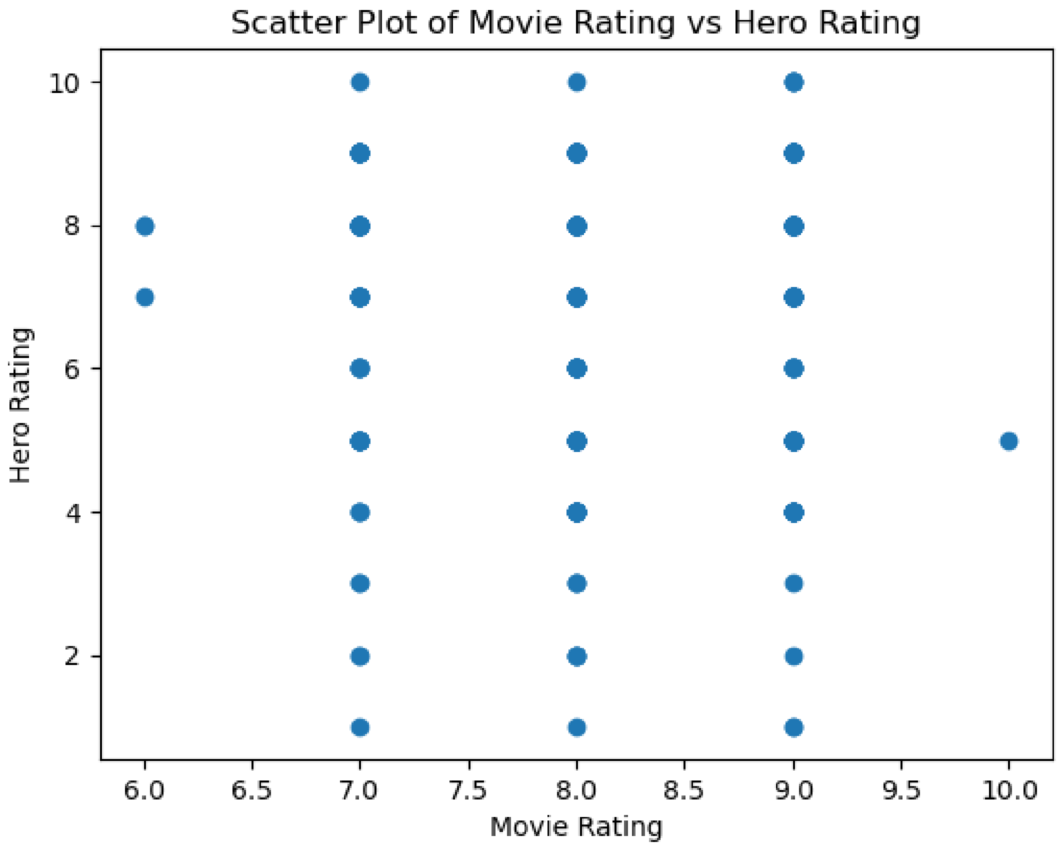
[24]:

```
             MovieName   Genre  PeopleVote  movie_rating  \
0           Mouna Guru  Action         746             8
1          7 Aum Arivu  Action        9479             9
2     Vaagai Sooda Vaa  Comedy       14522             7
3             Mankatha  Action       12276             8
4     Kanchana: Muni 2  Comedy        1044             9
..                 …       …           …            …
324  Dhilluku Dhuddu 2  Comedy         497             9
325                Dev  Action         724             8
326  Charlie Chaplin 2  Comedy         215             7
327              Petta  Action        7545             8
328           Viswasam  Action        5907             9

                Director  Year
0           Santha Kumar  2011
1        A.R. Murugadoss  2011
2            A. Sarkunam  2011
3          Venkat Prabhu  2011
4    Lawrence Raghavendra  2011
..                   …     …
324              Rambala  2019
325    Rajath Ravishankar  2019
326    Sakthi Chidambaram  2019
327      Karthik Subbaraj  2019
```

```
328                   Siva  2019

[329 rows x 6 columns]
```

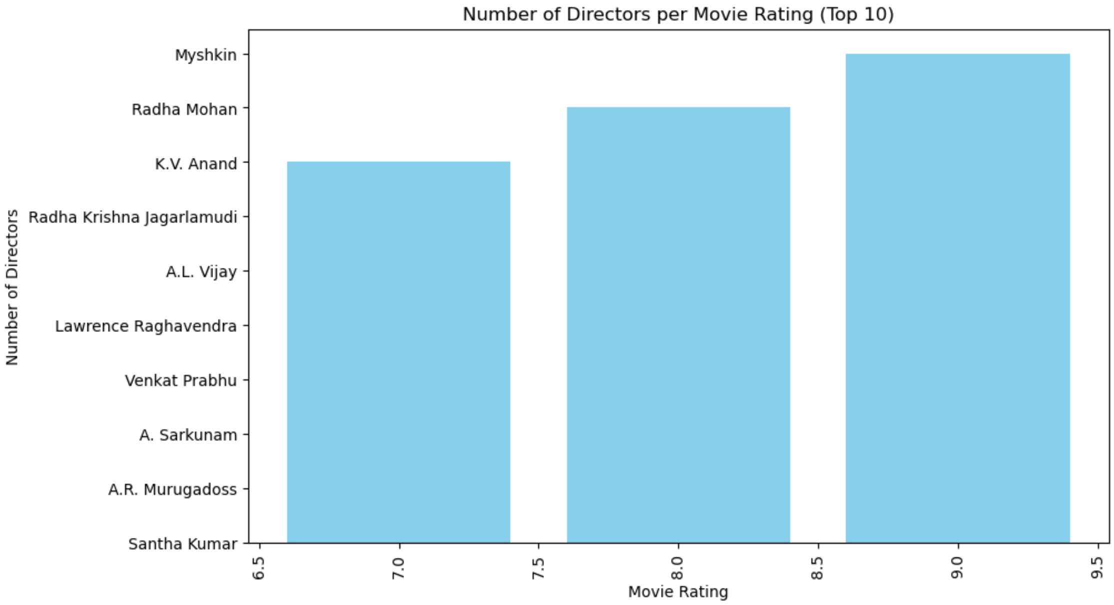
[11]:

```python
# Assuming df is your DataFrame and it has columns 'movie_rating' and 'Hero_Rating'

# Create scatter plot
plt.scatter(df['movie_rating'], df['Hero_Rating'])

# Add labels and title
plt.xlabel('Movie Rating')
plt.ylabel('Hero Rating')
plt.title('Scatter Plot of Movie Rating vs Hero Rating')

# Display the plot
plt.show()
```



[23]:

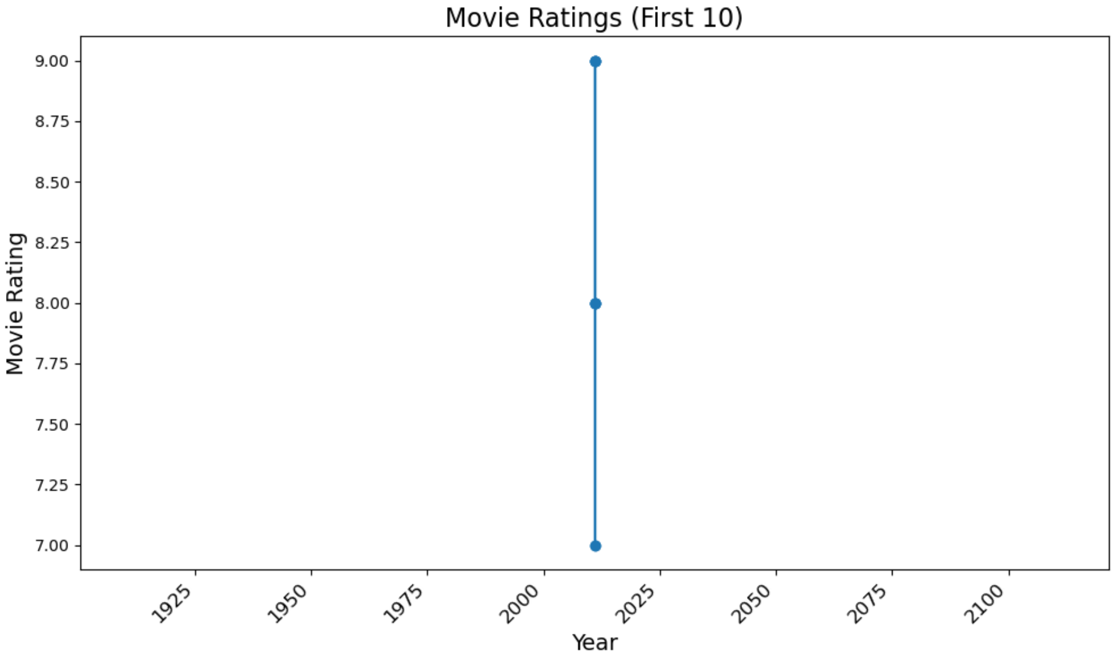
```
import pandas as pd
import matplotlib.pyplot as plt

# Assuming df_rt is your DataFrame

# Select the first 10 data points
top10_data = df_rt.head(10)

# Plot the bar chart for the first 10 data points
plt.figure(figsize=(10, 6))
plt.bar(top10_data['movie_rating'], top10_data['Director'], color='skyblue')
plt.xlabel('Movie Rating')
plt.ylabel('Number of Directors')
plt.title('Number of Directors per Movie Rating (Top 10)')
plt.xticks(rotation=90)  # Rotate x-axis labels if needed
plt.show()
```



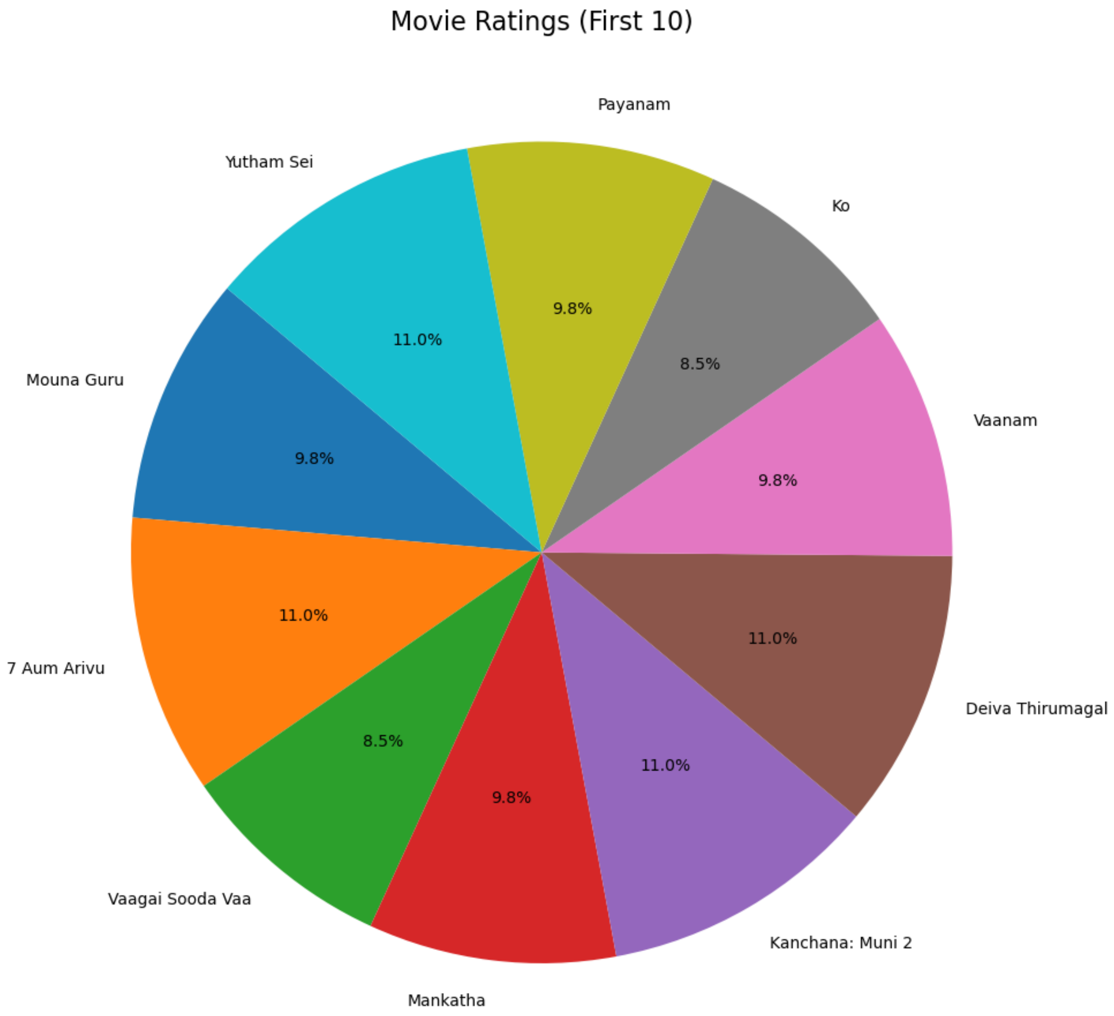
[26]:

```
import matplotlib.pyplot as plt

# Plotting the first 10 data points
plt.figure(figsize=(10, 6))  # Optional: Adjusts the figure size
plt.plot(df_rt['Year'][:10], df_rt['movie_rating'][:10], marker='o')

# Adding labels to the axes
plt.xlabel('Year', fontsize=14)
plt.ylabel('Movie Rating', fontsize=14)
```

```
# Rotating x-axis labels for better readability
plt.xticks(rotation=45, ha='right', fontsize=12)

# Adding a title to the plot
plt.title('Movie Ratings (First 10)', fontsize=16)

# Display the plot
plt.tight_layout()  # Adjusts plot to ensure everything fits without overlapping
plt.show()
```

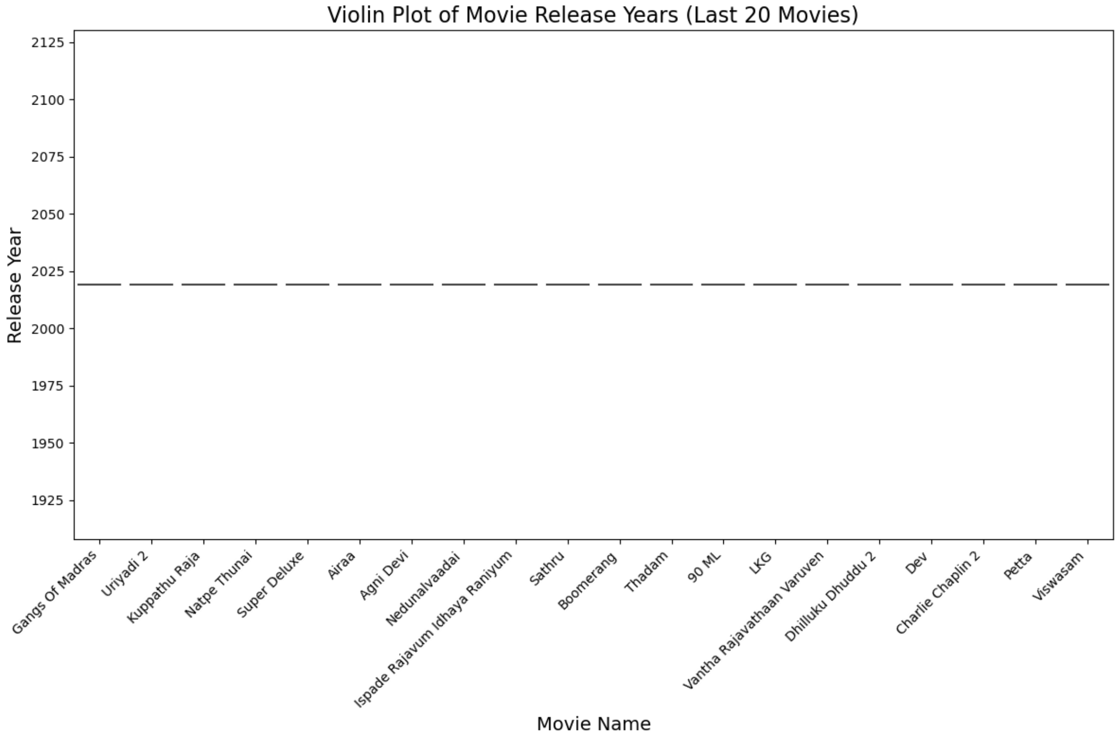


```
[28]: import matplotlib.pyplot as plt

# Plotting the data as a pie chart
plt.figure(figsize=(10, 10))  # Increase the figure size for better readability
plt.pie(df_rt['movie_rating'][:10], labels=df_rt['MovieName'][:10], autopct='%1.
 ↪1f%%', startangle=140)

# Adding a title to the plot (optional)
plt.title('Movie Ratings (First 10)', fontsize=16)

# Display the plot
plt.tight_layout()  # Adjusts plot to ensure everything fits without overlapping
plt.show()
```



[68]:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming 'MovieName' and 'Year' columns exist in df DataFrame
# Slice the DataFrame to get the last 20 rows
df_last_20 = df.tail(20)

plt.figure(figsize=(12, 8))
sns.violinplot(x=df_last_20['MovieName'], y=df_last_20['Year'])
plt.xlabel('Movie Name', fontsize=14)
plt.ylabel('Release Year', fontsize=14)
plt.title('Violin Plot of Movie Release Years (Last 20 Movies)', fontsize=16)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.tight_layout()
plt.show()
```



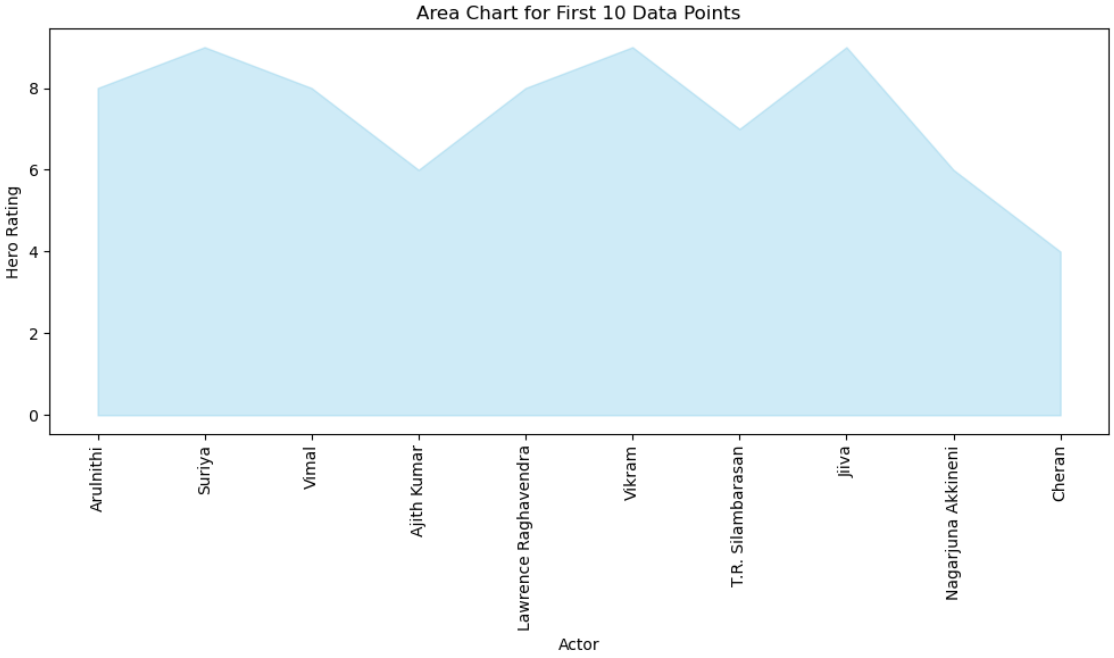
[66]:

```
print(df.columns)
```

```
Index(['MovieName', 'Genre', 'Rating', 'Director', 'Actor', 'PeopleVote',
       'Year', 'Hero_Rating', 'movie_rating', 'content_rating'],
      dtype='object')
```

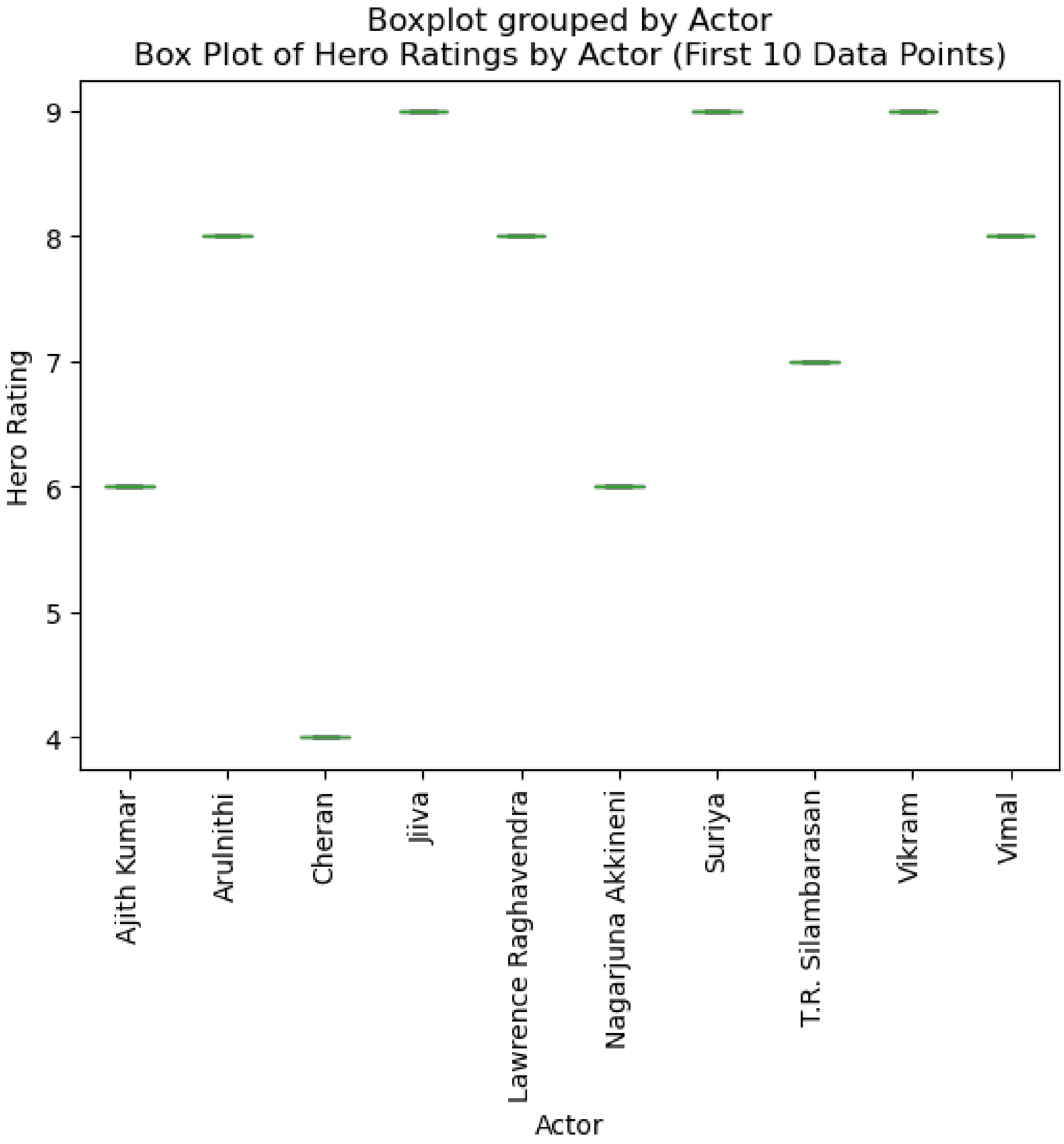
[12]:

```
import matplotlib.pyplot as plt
import pandas as pd


# Select the first 10 rows
df_first_10 = df.head(10)

plt.figure(figsize=(10, 6))  # Optional: Set the figure size
plt.fill_between(df_first_10['Actor'], df_first_10['Hero_Rating'],␣
 ↪color='skyblue', alpha=0.4)
plt.xlabel('Actor')
plt.ylabel('Hero Rating')
plt.title('Area Chart for First 10 Data Points')
plt.xticks(rotation=90)  # Rotate x-tick labels to be vertical
plt.tight_layout()  # Adjust layout to make room for rotated x-axis labels
plt.show()
```



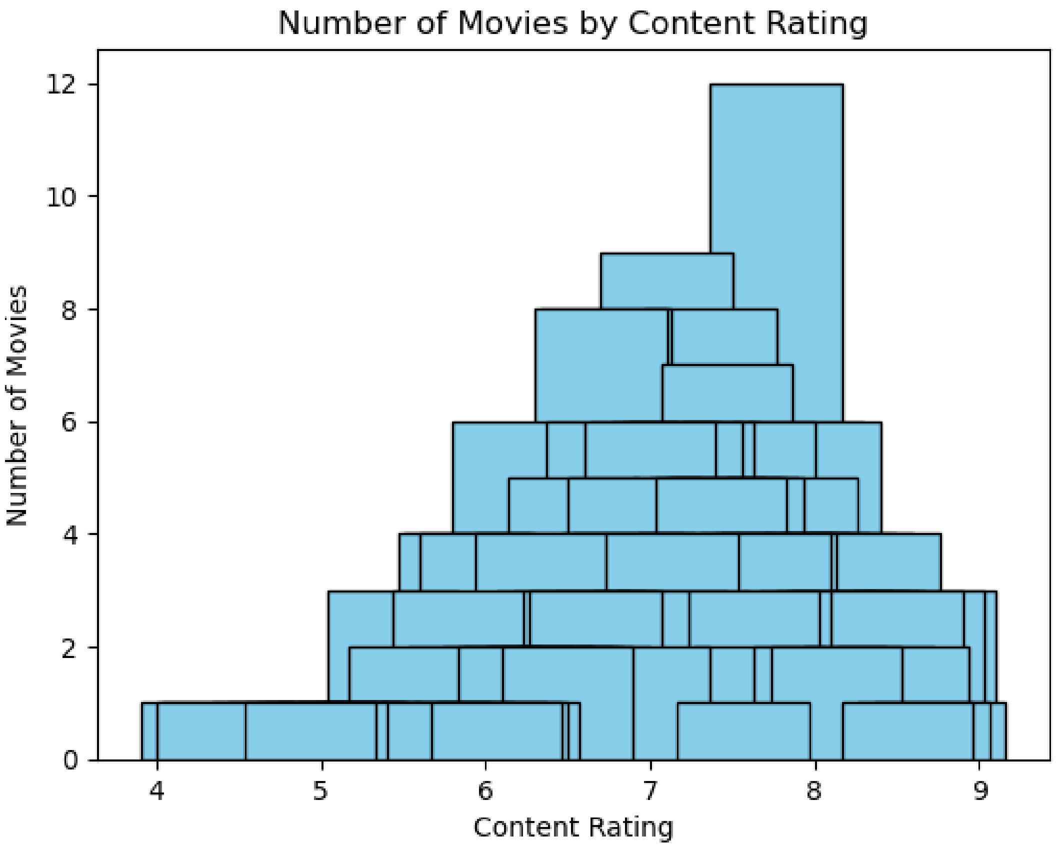
[31]:
```
import matplotlib.pyplot as plt
import pandas as pd


df_first_10 = df.head(10)

# Create a box plot for Hero_Rating grouped by Actor for the first 10 data␣
↪points
plt.figure(figsize=(12, 8))  # Optional: Set the figure size
df_first_10.boxplot(column='Hero_Rating', by='Actor', grid=False,␣
↪patch_artist=True, boxprops=dict(facecolor='skyblue'))

plt.xlabel('Actor')
plt.ylabel('Hero Rating')
plt.title('Box Plot of Hero Ratings by Actor (First 10 Data Points)')
plt.xticks(rotation=90)
plt.show()
```

```
<Figure size 1200x800 with 0 Axes>
```



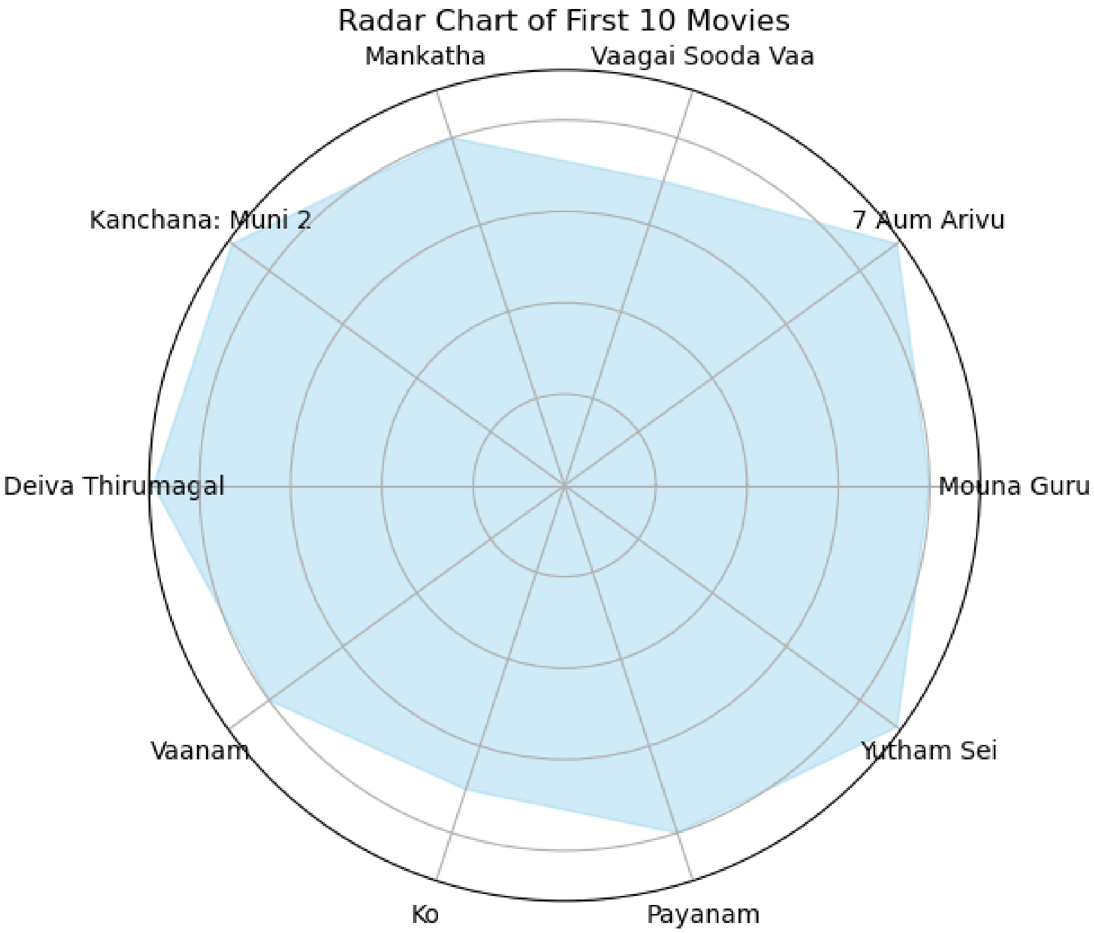
[9]:

```python
import matplotlib.pyplot as plt
import pandas as pd


rating_counts = df['content_rating'].value_counts()

# Plotting the bar chart
plt.bar(rating_counts.index, rating_counts, color='skyblue', edgecolor='black')
plt.xlabel('Content Rating')
plt.ylabel('Number of Movies')
plt.title('Number of Movies by Content Rating')
plt.show()
```



[3]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

# Load data from CSV file
data = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')

# Extract the first 10 rows
data = data.head(10)

# Extract MovieName and movie_rating columns
MovieName = data['MovieName']
movie_rating = data['movie_rating']

# Number of variables
num_vars = len(MovieName)
```

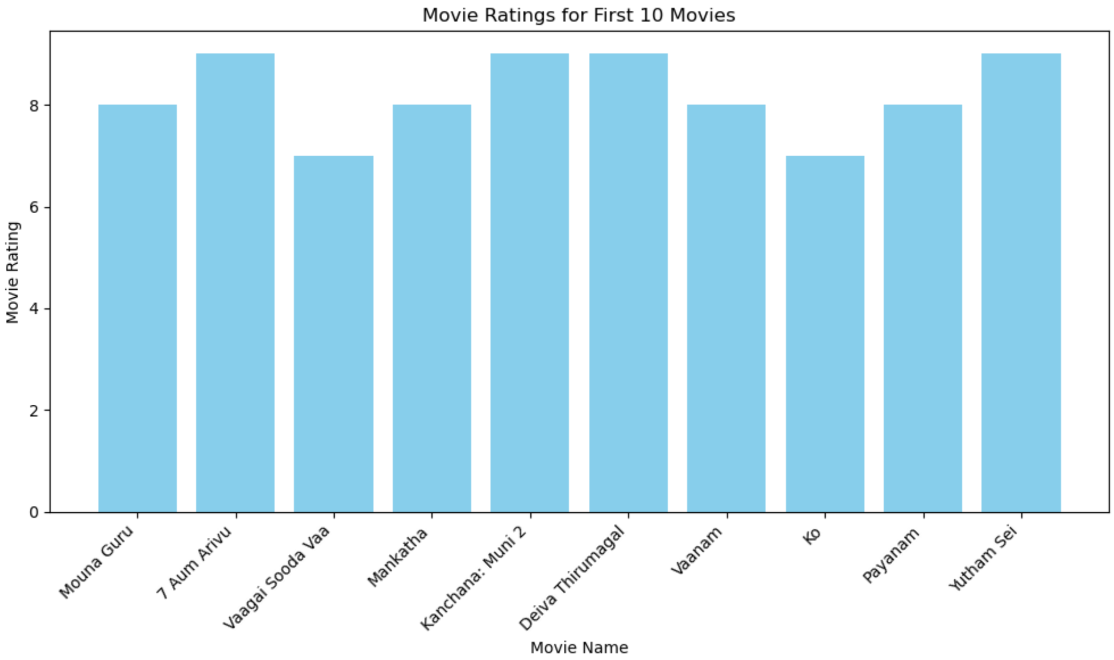
```
# Compute angle for each axis
angles = np.linspace(0, 2 * np.pi, num_vars, endpoint=False).tolist()

# The plot is circular, so we need to "complete the loop" and append the start value to the end.
movie_rating = movie_rating.tolist()  # Convert to list
movie_rating += movie_rating[:1]
angles += angles[:1]

# Plot
fig, ax = plt.subplots(figsize=(6, 6), subplot_kw=dict(polar=True))
ax.fill(angles, movie_rating, color='skyblue', alpha=0.4)

# Add labels
ax.set_yticklabels([])
ax.set_xticks(angles[:-1])
ax.set_xticklabels(MovieName)

# Show the plot
plt.title('Radar Chart of First 10 Movies')
plt.show()
```



[9]:

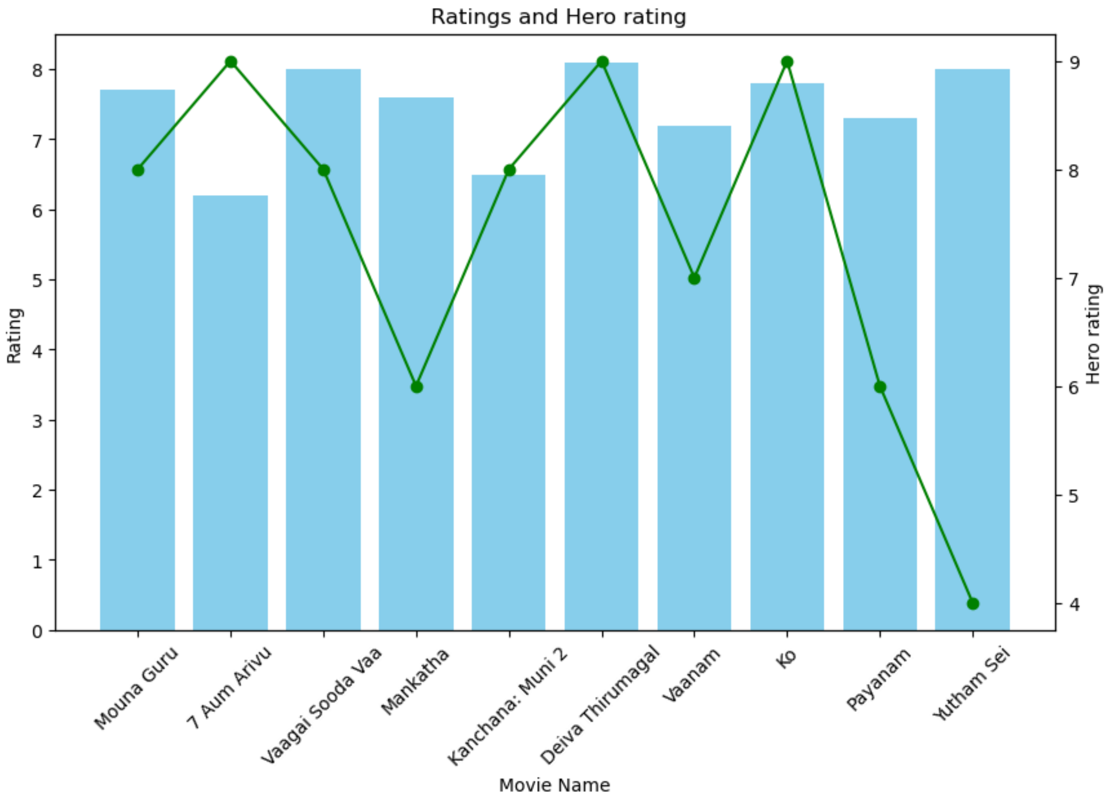
```
import pandas as pd
import matplotlib.pyplot as plt

# Load data from CSV file
data = pd.read_csv('C:\\Users\\subbiah\\OneDrive\\Desktop\\Tamil_movies_dataset.
 ↪csv')

# Extract the first 10 rows
data = data.head(10)

# Extract MovieName and movie_rating columns
MovieName = data['MovieName']
movie_rating = data['movie_rating']

# Plot
plt.figure(figsize=(10, 6))
plt.bar(MovieName, movie_rating, color='skyblue')
```

```
# Add labels and title
plt.xlabel('Movie Name')
plt.ylabel('Movie Rating')
plt.title('Movie Ratings for First 10 Movies')

# Rotate x-axis labels for better readability
plt.xticks(rotation=45, ha='right')

# Show the plot
plt.tight_layout()
plt.show()
```



[13]:
```
import pandas as pd
import matplotlib.pyplot as plt

fig, ax1 = plt.subplots(figsize=(10, 6))

# Plot the bar chart on the primary y-axis
ax1.bar(data['MovieName'], data['Rating'], color='skyblue', label='Rating')
ax1.set_xlabel('Movie Name')
ax1.set_ylabel('Rating')
ax1.set_title(' Ratings and Hero rating')
ax1.tick_params(axis='x', rotation=45)
```

```
# Create a secondary y-axis and plot the line chart
ax2 = ax1.twinx()
ax2.plot(data['MovieName'], data['Hero_Rating'], color='green', marker='o',␣
 ↪label='Votes')
ax2.set_ylabel('Hero rating')


plt.show()
```



[ ]: