

+ | X | ☰ | 📄 | Run All | Code ▾

Draft Session (35m)

CPU RAM

...

```
[1]: import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt

import plotly.express as px
import plotly.graph_objects as go

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

+ Code

+ Markdown

```
[2]: movie_data = pd.read_csv("/kaggle/input/imdb-india-movies/IMDb Movies India.csv", encoding='latin1')
```

```
[3]: movie_data.head()
```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
0		NaN	NaN	Drama	NaN	NaN	J.S. Randhawa	Manmaui	Birbal	Rajendra Bhatia
1	#Gadhvi (He thought he was Gandhi)	(2019)	109 min	Drama	7.0	8	Gaurav Bakshi	Rasika Dugal	Vivek Ghambade	Arvind Jangid
2	#Homecoming	(2021)	90 min	Drama, Musical	NaN	NaN	Soumyajit Majumdar	Sayani Gupta	Plabita Borthakur	Roy Angana
3	#Yaaram	(2019)	110 min	Comedy, Romance	4.4	35	Ovais Khan	Prateik	Ishita Raj	Siddhant Kapoor
4	...And Once Again	(2010)	105 min	Drama	NaN	NaN	Amol Palekar	Rajat Kapoor	Rituparna Sengupta	Antara Mali

```
[4]: movie_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15509 entries, 0 to 15508
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   Name        15509 non-null   object 
 1   Year         14981 non-null   object 
 2   Duration    7240 non-null   object 
 3   Genre        13632 non-null   object 
 4   Rating       7919 non-null   float64
 5   Votes        7920 non-null   object 
 6   Director    14984 non-null   object 
 7   Actor 1     13892 non-null   object 
 8   Actor 2     13125 non-null   object 
 9   Actor 3     12365 non-null   object 
dtypes: float64(1), object(9)
memory usage: 1.2+ MB
```

72 min' '110 min' '22 min' '122 min' '101 min' '100 min' '102 min'
'87 min' '132 min' '66 min' '146 min' '112 min' '168 min' '158 min'
'126 min' '94 min' '138 min' '124 min' '144 min' '157 min' '136 min'
'107 min' '113 min' '80 min' '122 min' '149 min' '148 min' '130 min'
'121 min' '188 min' '115 min' '103 min' '114 min' '170 min' '100 min'
'99 min' '140 min' '128 min' '93 min' '125 min' '145 min' '75 min'
'111 min' '134 min' '85 min' '104 min' '92 min' '137 min' '127 min'
'150 min' '119 min' '135 min' '86 min' '76 min' '70 min' '72 min'
'151 min' '95 min' '52 min' '89 min' '143 min' '177 min' '117 min'
'123 min' '154 min' '88 min' '175 min' '153 min' '78 min' '139 min'
'133 min' '101 min' '180 min' '60 min' '46 min' '164 min' '162 min'
'171 min' '160 min' '152 min' '62 min' '163 min' '165 min' '141 min'
'210 min' '129 min' '156 min' '240 min' '172 min' '155 min' '118 min'
'167 min' '106 min' '193 min' '57 min' '108 min' '45 min' '195 min'
'174 min' '81 min' '178 min' '58 min' '184 min' '97 min' '98 min'
'131 min' '176 min' '169 min' '77 min' '91 min' '84 min' '173 min'
'74 min' '67 min' '181 min' '300 min' '79 min' '65 min' '48 min'
'183 min' '159 min' '83 min' '68 min' '49 min' '201 min' '64 min'
'186 min' '50 min' '69 min' '207 min' '55 min' '61 min' '185 min'
'187 min' '216 min' '63 min' '54 min' '198 min' '51 min' '71 min'
'73 min' '218 min' '191 min' '321 min' '199 min' '53 min' '56 min'
'179 min' '47 min' '206 min' '190 min' '211 min' '247 min' '213 min'
'223 min' '2 min' '189 min' '224 min' '202 min' '255 min' '197 min'
'182 min' '214 min' '208 min' '21 min' '200 min' '192 min' '37 min'
'261 min' '238 min' '204 min' '235 min' '298 min' '217 min' '250 min']

Unique values for Genre:

```
[Drama' 'Drama, Musical' 'Comedy, Romance' 'Comedy, Drama, Musical'
'Drama, Romance, War' 'Documentary' 'Horror, Mystery, Thriller'
>Action, Crime, Thriller' 'Horror' 'Horror, Romance, Thriller'
'Comedy, Drama, Romance' 'Thriller' 'Comedy, Drama' 'nan'
'Comedy, Drama, Fantasy, Comedy, Drama, Family' 'Comedy, Drama, Mystery'
'87 min' '132 min' '66 min' '146 min' '112 min' '168 min' '158 min'
'126 min' '94 min' '138 min' '124 min' '144 min' '157 min' '136 min'
'107 min' '113 min' '80 min' '122 min' '149 min' '148 min' '130 min'
'121 min' '188 min' '115 min' '103 min' '114 min' '170 min' '100 min'
```

'99 min' '140 min' '128 min' '93 min' '125 min' '145 min' '75 min'  
'111 min' '134 min' '85 min' '104 min' '92 min' '137 min' '127 min'  
'150 min' '119 min' '135 min' '86 min' '76 min' '70 min' '72 min'  
'151 min' '95 min' '52 min' '89 min' '143 min' '177 min' '117 min'  
'123 min' '154 min' '88 min' '175 min' '153 min' '78 min' '139 min'  
'133 min' '101 min' '180 min' '60 min' '46 min' '164 min' '162 min'  
'171 min' '160 min' '152 min' '62 min' '163 min' '165 min' '141 min'  
'210 min' '129 min' '156 min' '240 min' '172 min' '155 min' '118 min'  
'167 min' '106 min' '193 min' '57 min' '108 min' '45 min' '195 min'  
'174 min' '81 min' '178 min' '58 min' '184 min' '97 min' '98 min'  
'131 min' '176 min' '169 min' '77 min' '91 min' '84 min' '173 min'  
'74 min' '67 min' '181 min' '300 min' '79 min' '65 min' '48 min'  
'183 min' '159 min' '83 min' '68 min' '49 min' '201 min' '64 min'  
'186 min' '50 min' '69 min' '207 min' '55 min' '61 min' '185 min'  
'187 min' '216 min' '63 min' '54 min' '198 min' '51 min' '71 min'  
'73 min' '218 min' '191 min' '321 min' '199 min' '53 min' '56 min'  
'179 min' '47 min' '206 min' '190 min' '211 min' '247 min' '213 min'  
'223 min' '2 min' '189 min' '224 min' '202 min' '255 min' '197 min'  
'182 min' '214 min' '208 min' '21 min' '200 min' '192 min' '37 min'  
'261 min' '238 min' '204 min' '235 min' '298 min' '217 min' '250 min']

Unique values for Genre:

['Drama', 'Drama, Musical', 'Comedy, Romance', 'Comedy, Drama, Musical',  
'Drama, Romance, War', 'Documentary', 'Horror, Mystery, Thriller',  
'Action, Crime, Thriller', 'Horror', 'Horror, Romance, Thriller',  
'Comedy, Drama, Romance', 'Thriller', 'Comedy, Drama', 'nan',  
'Comedy, Drama, Fantasy', 'Comedy, Drama, Family', 'Crime, Drama, Mystery',  
'Horror, Thriller', 'Biography', 'Comedy, Horror', 'Action',  
'Drama, Horror, Mystery', 'Comedy', 'Action, Thriller', 'Drama, History',  
'Drama, History, Sport', 'Horror, Mystery, Romance', 'Horror, Mystery',  
'Drama, Horror, Romance', 'Action, Drama, History', 'Action, Drama, War',  
'Comedy, Family', 'Adventure, Horror, Mystery', 'Action, Sci-Fi',  
'Crime, Mystery, Thriller', 'War', 'Sport', 'Biography, Drama, History',  
'Horror, Romance', 'Crime, Drama', 'Drama, Romance', 'Adventure, Drama',  
'Comedy, Mystery, Thriller', 'Action, Crime, Drama', 'Crime, Thriller',  
'Horror, Sci-Fi, Thriller', 'Crime, Drama, Thriller',  
'Drama, Mystery, Thriller', 'Drama, Sport', 'Drama, Family, Musical',  
'Action, Comedy', 'Thriller', 'Action, Adventure, Fantasy',  
'Drama, Romance, Thriller', 'Action, Drama', 'Drama, Horror, Musical',  
'Action, Biography, Drama', 'Adventure, Comedy, Drama', 'Mystery',  
'Action, Fantasy, Mystery', 'Adventure, Drama, Mystery',  
'Mystery, Thriller', 'Adventure', 'Drama, Musical, Thriller',  
'Comedy, Crime, Drama', 'Musical, Romance', 'Documentary, Music',  
'Documentary, History, Music', 'Drama, Fantasy, Mystery',  
'Drama, Family, Sport', 'Drama, Thriller', 'Documentary, Biography',  
'Action, Adventure, Comedy', 'Romance', 'Comedy, Drama, Music',  
'Comedy, Horror, Mystery', 'Musical', 'Musical, Romance, Drama',  
'Family, Romance', 'Action, Sci-Fi, Thriller', 'Action, Drama, Romance',  
'Mystery, Romance', 'Fantasy', 'Family', 'Drama, Family',  
'Action, Comedy, Drama', 'Action, Drama, Thriller',  
'Drama, Horror, Thriller', 'Drama, Musical, Romance', 'Sci-Fi',  
'Action, Romance', 'Action, Crime', 'Action, Drama, Crime',  
'Drama, Family, Music', 'Action, Mystery, Thriller',  
'Action, Drama, Family', 'Action, Mystery', 'Drama, History, Romance',  
'Crime, Drama, Music', 'Sci-Fi', 'Animation', 'Crime, Mystery, Romance',  
'Action, Adventure, Romance', 'Music', 'Romance', 'Action, Comedy, Crime',  
'Comedy, Family, Fantasy', 'Romance, Drama', 'Drama, Family, Romance',  
'Romance, Drama, Family', 'Musical, Romance, Thriller',  
'Family, Musical, Romance', 'Action, Drama, Fantasy', 'Family, Drama',  
'Crime, Drama, Romance', 'Musical, Drama, Romance', 'Drama, Music, Musical',  
'Drama, Mystery', 'Adventure, Comedy, Romance', 'Crime, Drama, Horror',  
'Family, Music, Musical', 'Action, Musical, Thriller',  
'Action, Romance, Thriller', 'Romance, Thriller', 'Drama, Music',  
'Crime, Drama, Musical', 'Action, Crime, Mystery',  
'Action, Adventure, Thriller', 'Comedy, Romance, Sci-Fi', 'Crime',  
'Action, Drama, Mystery', 'Action, Comedy, Thriller', 'Biography, Drama',  
'Action, Comedy, Fantasy', 'Drama, Family, Horror',  
'Action, Adventure, Family', 'Documentary, Biography, Musical',  
'Action, Drama, Musical', 'Adventure, Thriller', 'Crime, Mystery',  
'Drama, Crime', 'Drama, Fantasy, Romance', 'Comedy, Romance, Thriller',  
'Musical, Comedy, Drama', 'Biography, History, War',  
'Action, Comedy, Romance', 'Drama, History, Musical',  
'Action, Crime, Horror', 'Adventure, Fantasy', 'Adventure, Drama, Fantasy',  
'Adventure, Fantasy, Romance', 'Action, Adventure, Drama',  
'Action, Adventure', 'Comedy, Crime', 'Crime, Drama, Fantasy',  
'Adventure, Drama, Romance', 'History', 'Drama, Fantasy, Thriller',  
'Musical, Fantasy', 'Documentary, Thriller', 'Mystery, Romance, Musical',  
'Family, Drama, Romance', 'History, Musical, Romance',  
'Musical, Drama, Crime', 'Adventure, Crime, Romance',  
'Musical, Thriller, Mystery', 'Drama, Comedy', 'Biography, Drama, Romance',  
'Biography, Music', 'Biography, Drama, Music', 'Drama, Sci-Fi',  
'Drama, Family, Thriller', 'Comedy, Musical, Romance',  
'Drama, Family, Comedy', 'Action, Thriller, Romance',  
'Animation, Adventure', 'Action, Crime, Musical', 'Action, Crime, Romance',  
'Animation, Action, Adventure', 'Action, Drama, Sport', 'Comedy, History',  
'Documentary, History', 'Drama, Comedy, Family', 'Action, Adventure, Crime',  
'Documentary, Biography, Music', 'Comedy, Musical',  
'Biography, Crime, Thriller', 'Adventure, Mystery, Thriller',  
'Biography, Drama, Sport', 'Action, Comedy, Musical',  
'Mystery, Romance, Thriller', 'Action, Adventure, Musical',  
'Crime, Musical, Mystery', 'Action, Thriller, Crime',  
'Adventure, Comedy, Crime', 'Comedy, Horror, Musical', 'Adventure, Family',  
'Family, Thriller', 'Drama, Action, Crime', 'Drama, War',  
'Action, Drama, Adventure', 'Adventure, Fantasy, History',  
'Fantasy, Musical', 'Comedy, Drama, Thriller', 'Drama, Fantasy',  
'Musical, Drama', 'Action, Drama, Horror', 'Biography, Crime, Drama',  
'Action, Drama, Music', 'Adventure, Drama, Family',  
'Drama, Romance, Musical', 'Comedy, Musical, Drama',  
'Adventure, Comedy, Musical', 'Crime, Drama, Family',  
'Thriller, Musical, Mystery', 'Documentary, Adventure, Crime',  
'Drama, Action, Horror', 'Adventure, Crime, Drama',  
'Documentary, Biography, Sport', 'Crime, Fantasy, Mystery',  
'Documentary, Biography, Drama', 'Action, Fantasy, Thriller',  
'Adventure, Drama, History', 'Animation, Drama, History',  
'Comedy, Horror, Thriller', 'Drama, Family, History', 'Animation, History',  
'Biography, Drama, Musical', 'Music', 'Family, Comedy', 'Adventure, Mystery',  
'Family, Fantasy', 'Documentary, History, News', 'Drama, Mystery, Romance',  
'Comedy, Fantasy', 'Action, Crime, Family', 'Drama, Musical, Mystery',  
'Action, Thriller, Mystery', 'Drama, Family, Fantasy', 'Action, Family',  
'Action, Adventure, Mystery', 'Horror, Fantasy', 'Comedy, Action',  
'Adventure, Romance', 'Drama, Adventure', 'Animation, Drama, Romance',  
'Comedy, Crime, Romance', 'Adventure, Comedy', 'Comedy, Drama, Sport']

'Documentary, Crime, History' 'Musical, Mystery, Drama'  
'Adventure, Drama, Sci-Fi' 'Action, Romance, Western'  
'Comedy, Fantasy, Romance' 'Animation, Action, Comedy'  
'Drama, Fantasy, Sci-Fi' 'Drama, Horror' 'Family, Drama, Comedy'  
'Action, Adventure, History' 'Comedy, Family, Romance'  
'Biography, History' 'Animation, Family' 'Drama, Fantasy, History'  
'Animation, Adventure, Fantasy' 'Adventure, Comedy, Family'  
'Drama, History, War' 'Animation, Drama, Fantasy'  
'Action, Musical, Romance' 'Crime, Action, Drama'  
'Comedy, Romance, Musical' 'Fantasy, Drama' 'Musical, Action, Crime'  
'Documentary, Drama' 'Action, Horror, Thriller' 'Action, Horror, Sci-Fi'  
'Mystery, Sci-Fi, Thriller' 'Biography, Family' 'Drama, Action, Comedy'  
'Drama, Music, Romance' 'Action, Biography, Crime'  
'Adventure, Drama, Musical' 'Family, Music, Romance'  
'Fantasy, Mystery, Romance' 'Drama, Crime, Family'  
'Drama, Family, Action' 'Romance, Comedy, Drama'  
'Animation, Adventure, Comedy' 'Sci-Fi, Thriller'  
'Romance, Family, Drama' 'Action, Family, Thriller'  
'Adventure, Crime, Thriller' 'Drama, Romance, Sport'  
'Comedy, Crime, Mystery' 'Adventure, Comedy, Mystery' 'Action, Fantasy'  
'Comedy, Mystery' 'Animation, Adventure, Family'  
'Adventure, Drama, Music' 'Biography, Drama, War'  
'Documentary, Comedy, Drama' 'Musical, Drama, Family'  
'Animation, Comedy, Drama' 'Fantasy, Musical, Drama'  
'Adventure, Crime, Mystery' 'Comedy, Drama, Mystery' 'Documentary, News'  
'Drama, Musical, Family' 'Action, Romance, Drama'  
'Comedy, Crime, Thriller' 'Action, Musical' 'Action, History'  
'Action, Comedy, Mystery' 'Drama, Family, Mystery'  
'Adventure, Drama, Thriller' 'Documentary, Reality-TV'  
'Action, Fantasy, Horror' 'Drama, History, Thriller'  
'Documentary, Family' 'Documentary, Biography, Family' 'Comedy, Sport'  
'Animation, Comedy, Family' 'Crime, Romance, Thriller'  
'Comedy, Musical, Action' 'Action, Mystery, Sci-Fi'  
'Comedy, Crime, Musical' 'Drama, Adventure, Action' 'History, Romance'  
'Reality-TV' 'Fantasy, History' 'Family, Drama, Thriller'  
'Musical, Mystery, Thriller' 'Musical, Comedy, Romance'  
'Musical, Action, Drama' 'Action, Musical, War' 'Romance, Comedy'  
'Horror, Crime, Thriller' 'Crime, Drama, History' 'Comedy, Drama, Horror'  
'Crime, Horror, Thriller' 'Animation, Comedy' 'Romance, Action, Crime'  
'Musical, Thriller' 'Action, Romance, Comedy' 'Comedy, Family, Musical'  
'Horror, Drama, Mystery' 'Thriller, Mystery, Family'  
'Comedy, Drama, Sci-Fi' 'Documentary, Adventure'  
'Documentary, Biography, Crime' 'Musical, Action' 'Musical, Mystery'  
'Action, Crime, Sci-Fi' 'Action, Horror, Mystery' 'Fantasy, Horror'  
'Adventure, Family, Fantasy' 'Fantasy, Sci-Fi' 'Comedy, War'  
'Romance, Action, Drama' 'Musical, Family, Romance'  
'Romance, Drama, Action' 'Family, Comedy, Drama' 'Comedy, Music, Romance'  
'Comedy, Family, Sci-Fi' 'Action, Drama, Western'  
'Adventure, Romance, Thriller' 'Biography, Comedy, Drama'  
'Action, Mystery, Romance' 'Romance, Sport' 'Crime, Romance'  
'Action, Thriller, Western' 'Crime, Musical, Romance'  
'Romance, Thriller, Mystery' 'Drama, Crime, Mystery'  
'Biography, Drama, Family' 'Action, Family, Mystery'  
'Comedy, Mystery, Romance' 'Drama, Thriller, Action' 'Documentary, Short  
Documentary, Western' 'Musical, Family, Drama' 'Action, Family, Musical'  
'Animation, Family, Musical' 'Drama, Fantasy, Horror'  
'Action, Adventure, Sci-Fi' 'Drama, Action, Musical'  
'Drama, Musical, Sport' 'Action, Comedy, Horror'  
'Drama, Fantasy, Musical' 'Action, Fantasy, Musical' 'Animation, Action'  
'Comedy, Music' 'Documentary, Drama, Romance' 'Drama, Music, Thriller'  
'Fantasy, Musical, Mystery' 'Drama, Fantasy, War' 'Action, War'  
'Action, Adventure, War' 'Horror, Musical' 'Fantasy, Mystery, Thriller'  
'Adventure, Biography, Drama' 'Family, Romance, Sci-Fi'  
'Drama, Romance, Family' 'Animation, Adventure, Drama'  
'Family, Romance, Drama' 'Animation, Action, Sci-Fi'  
'Adventure, Comedy, Fantasy' 'Comedy, Crime, Family'  
'Horror, Musical, Thriller' 'Biography, Drama, Thriller' 'Drama, Western'  
'Romance, Sci-Fi, Thriller' 'Comedy, Musical, Family'  
'Comedy, Horror, Romance' 'Thriller, Action' 'Fantasy, Thriller, Action'  
'Fantasy, Romance' 'Action, Drama, Comedy' 'Family, Fantasy, Romance'  
'Comedy, Crime, Horror' 'Horror, Mystery, Sci-Fi'  
'Animation, Action, Drama' 'Family, Mystery'  
'Adventure, Biography, History' 'Fantasy, Horror, Mystery'  
'Family, Musical' 'Drama, Family, Adventure' 'Crime, Horror, Mystery'  
'Documentary, Drama, Fantasy' 'Action, Adventure, Biography'  
'Biography, History, Thriller' 'Action, Family, Drama'  
'Documentary, Drama, Sport' 'Thriller, Mystery' 'Musical, Drama, Comedy'  
'Documentary, History, War' 'Adventure, Horror, Thriller'  
'Action, Adventure, Horror' 'Action, Crime, War'  
'Adventure, Musical, Romance' 'Action, Fantasy, Sci-Fi'  
'Drama, Comedy, Action' 'Documentary, Sport'  
'Documentary, Adventure, Music' 'Drama, Action, Family'  
'Adventure, History, Thriller' 'Adventure, Horror, Romance'  
'Adventure, Crime, Horror' 'Mystery, Musical, Romance'  
'Action, Crime, History' 'Documentary, Musical'  
'Adventure, Fantasy, Musical' 'Documentary, Family, History'  
'Documentary, Drama, Family' 'Drama, Mystery, Sci-Fi'  
'Animation, Drama, Musical' 'Drama, History, Mystery'  
'Drama, Sport, Thriller' 'Action, Crime, Fantasy'  
'Comedy, Musical, Mystery' 'Romance, Musical, Action'  
'Musical, Drama, Fantasy' 'Animation, Family, History'  
'Action, Drama, News' 'Romance, Musical, Comedy'  
'Adventure, Fantasy, Horror' 'Adventure, History'  
'Comedy, Drama, History' 'Mystery, Sci-Fi' 'Action, Thriller, War'  
'Documentary, Drama, News' 'Documentary, Crime, Mystery'  
'Adventure, Horror' 'Animation, Drama, Adventure'  
'Crime, Horror, Romance' 'Documentary, Adventure, Drama'  
'Documentary, Biography, History' 'Fantasy, Horror, Romance'  
'Comedy, Fantasy, Musical' 'Crime, Musical, Thriller' 'Documentary, War'  
'Action, Comedy, War' 'Crime, Drama, Sport' 'Musical, Adventure, Drama'  
'Horror, Romance, Sci-Fi' 'Musical, Mystery, Romance'  
'Romance, Musical, Drama' 'Adventure, Fantasy, Sci-Fi')

Unique values for Rating:

```
[ nan  7.   4.4  4.7  7.4  5.6  4.   6.2  5.9  6.5  5.7  6.3  7.2  6.6
 7.3  7.1  6.9  3.5  5.   4.5  6.4  4.1  4.8  8.1  5.5  6.8  6.1  7.7
 5.1  7.6  3.1  3.3  7.8  8.4  5.2  4.3  5.8  4.6  7.5  6.7  3.6  3.9
 5.4  4.2  5.3  3.4  3.   8.   6.   3.8  7.9  2.7  4.9  2.4  3.7  3.2
 2.5  2.8  2.6  2.9  8.2  8.7  8.3  9.3  8.8  2.1  2.3  8.5  8.6  9.
 9.6  1.7  9.1  2.   1.4  8.9  1.9  9.4  9.7  1.8  9.2  1.6  10.  2.2
 1.1]
```

```

Unique values for Votes:
[nan '8' '35' ... '70,344' '408' '1,496']

Unique values for Director:
['J.S. Randhawa' 'Gaurav Bakshi' 'Soumyajit Majumdar' ... 'Mozez Singh'
 'Ved Prakash' 'Kiran Thej']

Unique values for Actor 1:
['Mannauji' 'Rasika Dugal' 'Sayani Gupta' ... 'Meghan Jadhav'
 'Roohi Berde' 'Sangeeta Tiwari']

Unique values for Actor 2:
['Birbal' 'Vivek Ghamande' 'Plabita Borthakur' ... 'Devan Sanjeev'
 'Prince Daniel' 'Sarah Jane Dias']

Unique values for Actor 3:
['Rajendra Bhatia' 'Arvind Jangid' 'Roy Angana' ... 'Shatakshi Gupta'
 'Valerie Agha' 'Suparna Anand']

```

▶

+ Code + Markdown

[15]:

```

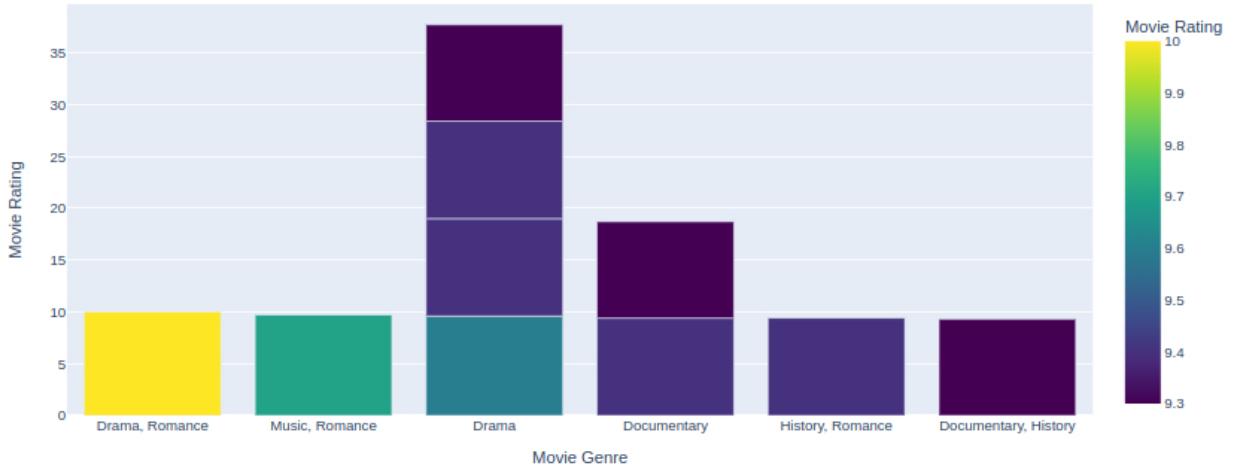
top_10_movies = movie_data.sort_values(by='Rating', ascending=False).head(10)

# Create a bar plot for the top 10 movies
fig = px.bar(top_10_movies, x='Genre', y='Rating', title='Top 10 Movies by Genre and Rating',
             labels={'Rating': 'Movie Rating', 'Genre': 'Movie Genre'},
             color='Rating', color_continuous_scale='Viridis',
             hover_data=['Director', 'Actor 1', 'Actor 2', 'Actor 3'])

fig.show()

```

Top 10 Movies by Genre and Rating



[9]:

```

import plotly.express as px

# Assuming your DataFrame is named 'movie_data'
# Let's focus on the relevant columns for visualization
columns_of_interest = ['Year', 'Duration', 'Genre', 'Rating', 'Votes', 'Director', 'Actor 1', 'Actor 2', 'Actor 3']
subset_data = movie_data[columns_of_interest]

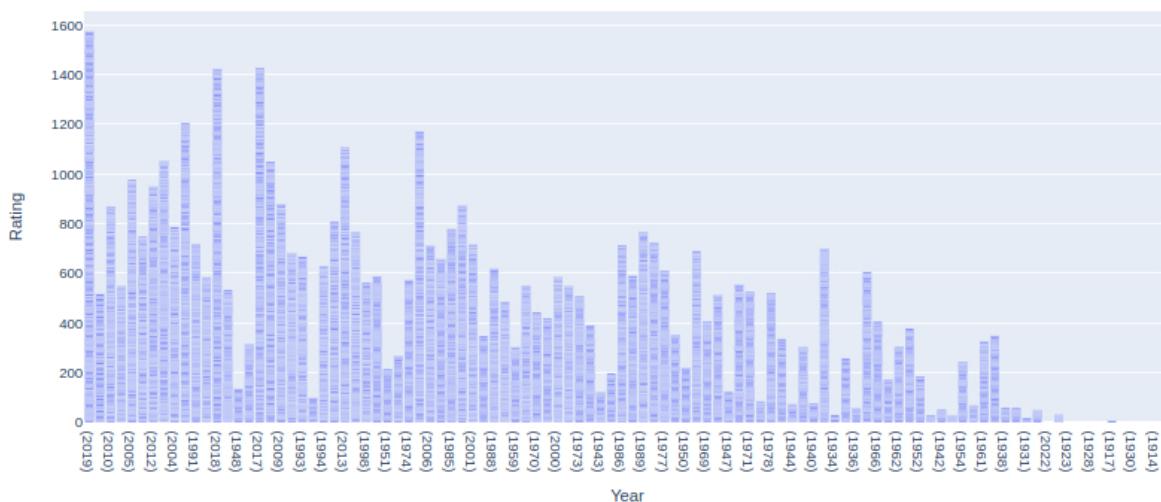
# Pairwise scatter plots for numerical columns
#fig = px.scatter_matrix(subset_data, dimensions=['Duration', 'Rating', 'Votes'], color='Genre')
#fig.update_layout(title='Pairwise Scatter Plots for Numerical Columns')
#fig.show()

# Bar plots for categorical columns
for column in ['Year']:
    fig = px.bar(subset_data, x=column, y='Rating', title=f'Relationship between {column} and Rating')
    fig.show()

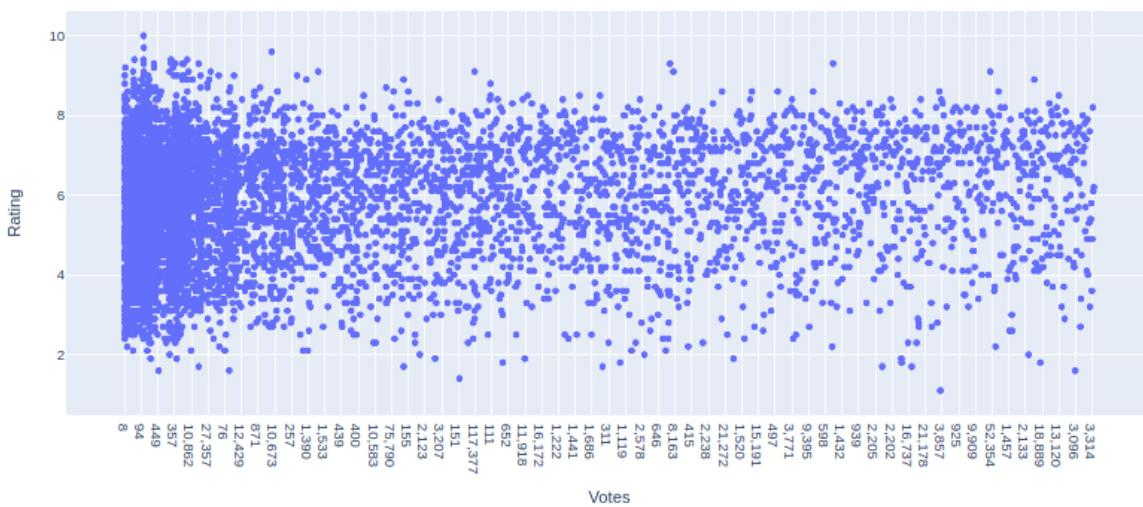
# Scatter plot for 'Votes'
fig = px.scatter(subset_data, x='Votes', y='Rating', title='Scatter Plot: Votes vs Rating')
fig.show()

```

### Relationship between Year and Rating



### Scatter Plot: Votes vs Rating



```
[16]: movie_data['Name'] = movie_data['Name'].str.extract('([a-zA-Z\s]+)', expand=False)
movie_data['Votes'] = pd.to_numeric(movie_data['Votes'], errors='coerce')
movie_data['Duration'] = pd.to_numeric(movie_data['Duration'].str.replace('min', ''), errors='coerce')
movie_data['Year'] = pd.to_datetime(movie_data['Year'].str.strip('(')), errors='coerce')
```

```
[17]: duplicate_rows = movie_data[movie_data.duplicated()]
duplicate_rows
```

	Name	Year	Duration	Genre	Rating	Votes	Director	Actor 1	Actor 2	Actor 3
1250	Arab Ka Sona	1979-01-01	NaN	Action	NaN	NaN	Master Bhagwan	Meena Rai	Dara Singh	NaN
1769	Balidan	1992-01-01	NaN	Drama	NaN	NaN	NaN	NaN	NaN	NaN
4723	First Time	2009-01-01	NaN	NaN	NaN	NaN	Raja Bundela	Zeenat Aman	Nitin Arora	Raj Babbar
8819	Mangal Ho	NaT	NaN	Comedy	NaN	NaN	Pritish Chakraborty	Pritish Chakraborty	NaN	NaN
9713	Musafir	NaT	NaN	Thriller	NaN	NaN	Shiva Dagar	NaN	NaN	NaN
13069	Shivani	2019-01-01	NaN	Crime	NaN	NaN	Ugresh Prasad Ujala	Santosh	NaN	NaN
13308	Slumdog Karodpati	2019-01-01	118.0	Thriller	NaN	NaN	Rajesh Patole	Udhav Garje	Rahul Gavane	Govindrao

```
[18]: # Handle missing values
movie_data['Votes'].fillna(movie_data['Votes'].mean(), inplace=True)
```

```
movie_data['Duration'].fillna(movie_data['Duration'].median(), inplace=True)
```

```
[19]:  
# Calculate mean ratings for each category  
genre_mean_rating = movie_data.groupby('Genre')['Rating'].transform('mean')  
director_mean_rating = movie_data.groupby('Director')['Rating'].transform('mean')  
actor1_mean_rating = movie_data.groupby('Actor 1')['Rating'].transform('mean')  
actor2_mean_rating = movie_data.groupby('Actor 2')['Rating'].transform('mean')  
actor3_mean_rating = movie_data.groupby('Actor 3')['Rating'].transform('mean')  
  
# Encode mean ratings into the original DataFrame with some differences  
movie_data['Genre_mean_rating'] = genre_mean_rating  
movie_data['Director_mean_rating'] = director_mean_rating  
movie_data['Actor1_mean_rating'] = actor1_mean_rating  
movie_data['Actor2_mean_rating'] = actor2_mean_rating  
movie_data['Actor3_mean_rating'] = actor3_mean_rating  
  
# Fill NaN values in new columns with the overall mean rating or any other suitable value  
mean_rating_fill_value = movie_data['Rating'].mean()  
movie_data['Genre_mean_rating'].fillna(mean_rating_fill_value, inplace=True)  
movie_data['Director_mean_rating'].fillna(mean_rating_fill_value, inplace=True)  
movie_data['Actor1_mean_rating'].fillna(mean_rating_fill_value, inplace=True)  
movie_data['Actor2_mean_rating'].fillna(mean_rating_fill_value, inplace=True)  
movie_data['Actor3_mean_rating'].fillna(mean_rating_fill_value, inplace=True)
```

```
[20]:  
movie_data = movie_data.drop(['Name', 'Genre', 'Director', 'Actor 1', 'Actor 2', 'Actor 3'], axis=1)
```

```
[21]:  
movie_data.dropna(subset=['Rating'], inplace=True)
```

```
[22]:  
movie_data.info()  
  
<class 'pandas.core.frame.DataFrame'>  
Index: 7919 entries, 1 to 15508  
Data columns (total 9 columns):  
 #   Column           Non-Null Count  Dtype     
 ---    
 0   Year            7919 non-null    datetime64[ns]  
 1   Duration        7919 non-null    float64  
 2   Rating          7919 non-null    float64  
 3   Votes           7919 non-null    float64  
 4   Genre_mean_rating 7919 non-null    float64  
 5   Director_mean_rating 7919 non-null    float64  
 6   Actor1_mean_rating 7919 non-null    float64  
 7   Actor2_mean_rating 7919 non-null    float64  
 8   Actor3_mean_rating 7919 non-null    float64  
dtypes: datetime64[ns](1), float64(8)  
memory usage: 618.7 KB
```

```
[23]:  
# Assuming your DataFrame is named 'data'  
X = movie_data.drop(['Rating', 'Year'], axis=1) # Remove 'Year' as it is not numeric  
y = movie_data['Rating']  
  
# Train-test split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
≡▶ # Random Forest Regressor  
rf_model = RandomForestRegressor(random_state=42)  
rf_model.fit(X_train, y_train)  
  
# Predictions  
rf_preds = rf_model.predict(X_test)
```

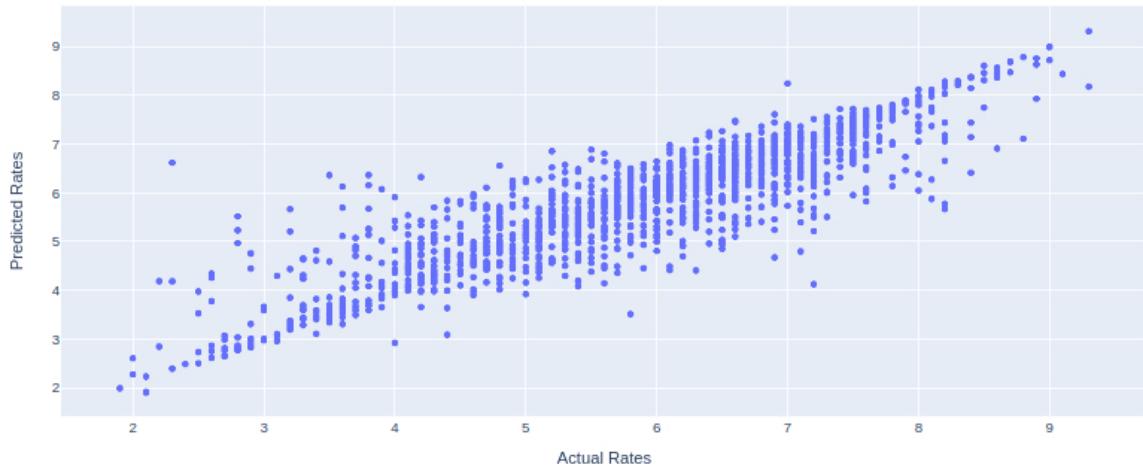
```
[25]:  
# Evaluate Random Forest model  
rf_mse = mean_squared_error(y_test, rf_preds)  
rf_mae = mean_absolute_error(y_test, rf_preds)  
rf_r2 = r2_score(y_test, rf_preds)
```

```
[26]:  
# Display results  
print("Random Forest Metrics:")  
print("MSE:", rf_mse)  
print("MAE:", rf_mae)  
print("R2:", rf_r2)
```

```
Random Forest Metrics:  
MSE: 0.4071709071101641  
MAE: 0.4262225273569024  
R2: 0.7809902528807179
```

```
[29]:  
fig = px.scatter(x=y_test, y=rf_preds, labels={'x': 'Actual Rates', 'y': 'Predicted Rates'},  
                  title='Actual vs Predicted Rates')  
fig.show()
```

Actual vs Predicted Rates



```
[30]:  
# Initialize the linear regression model  
model = LinearRegression()  
  
# Train the model  
model.fit(X_train, y_train)
```

```
[30...]  
└ LinearRegression  
LinearRegression()
```

```
[31]:  
y_pred = model.predict(X_test)
```

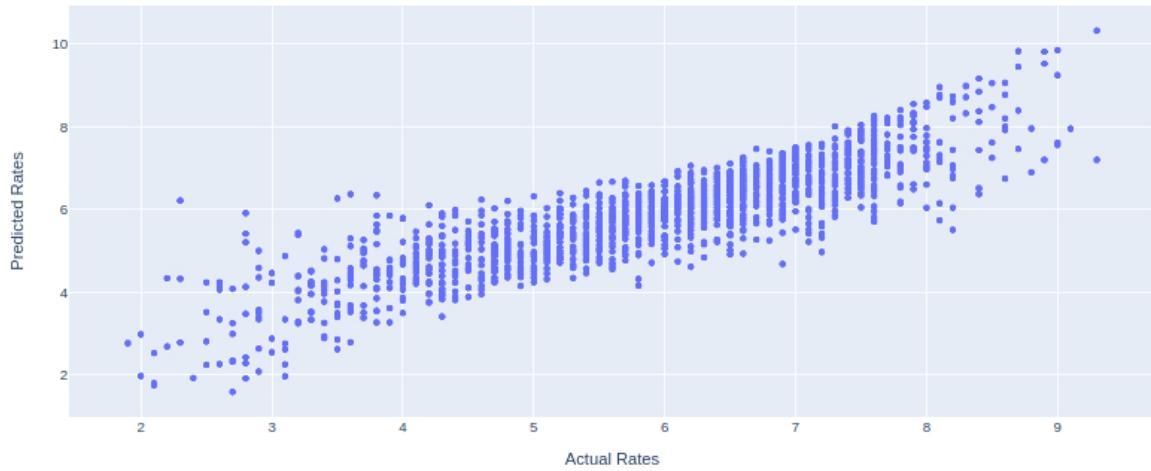
```
[32]:  
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)  
print(f'Mean Squared Error: {mse}')  
print(f'R^2 Score: {r2}')
```

```
Mean Squared Error: 0.457342811364929  
R^2 Score: 0.7540037077434054
```

```
[33]:
```

```
fig = px.scatter(x=y_test, y=y_pred, labels={'x': 'Actual Rates', 'y': 'Predicted Rates'},
                  title='Actual vs Predicted Rates')
fig.show()
```

Actual vs Predicted Rates



[ ]:

