PREDICTING IMDB SCORES

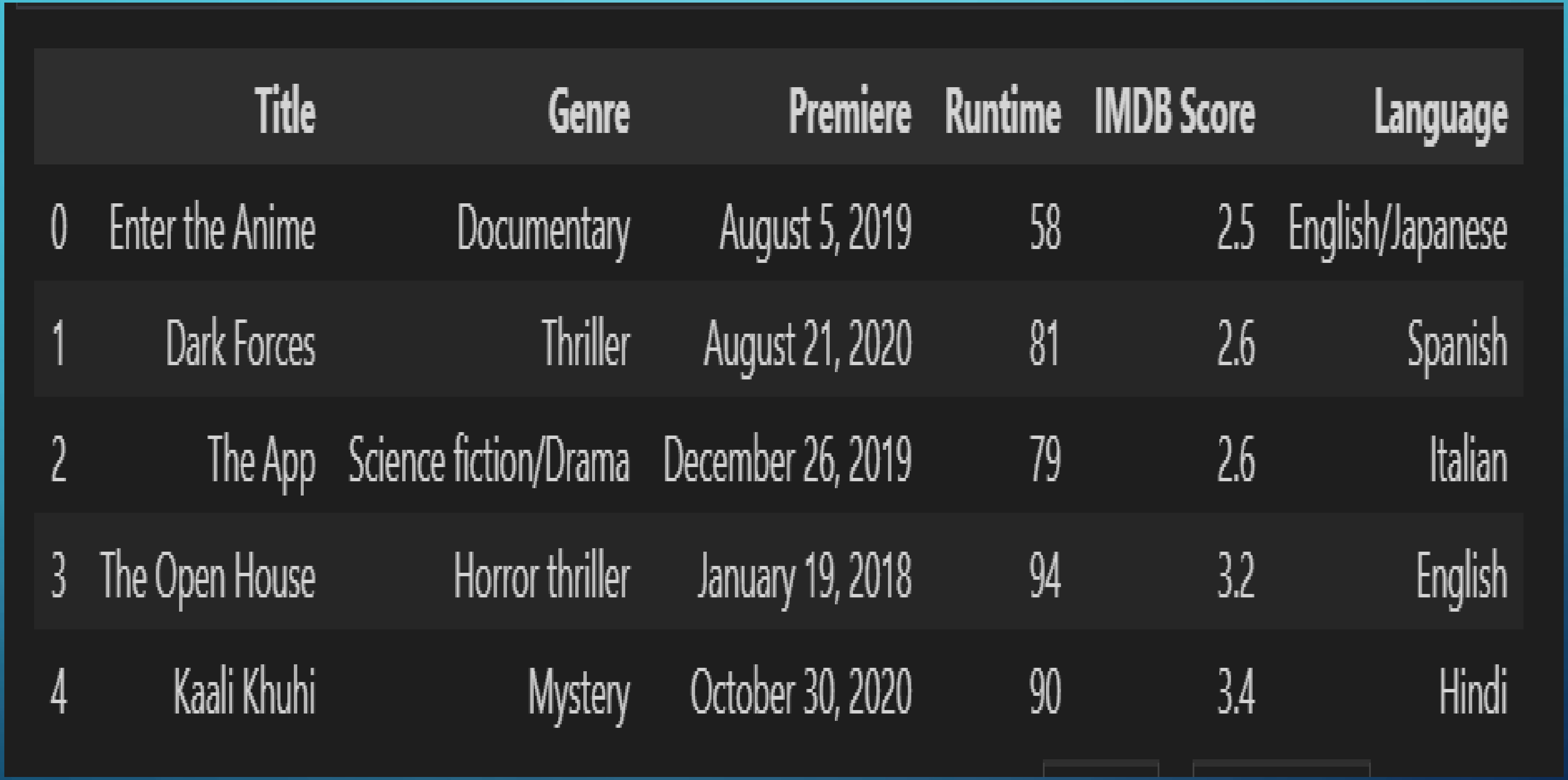
DATASET

- **Download the dataset for Predicting IMDb scores using the following link**
- **Dataset link:**
- https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores

DATASET INSERTION

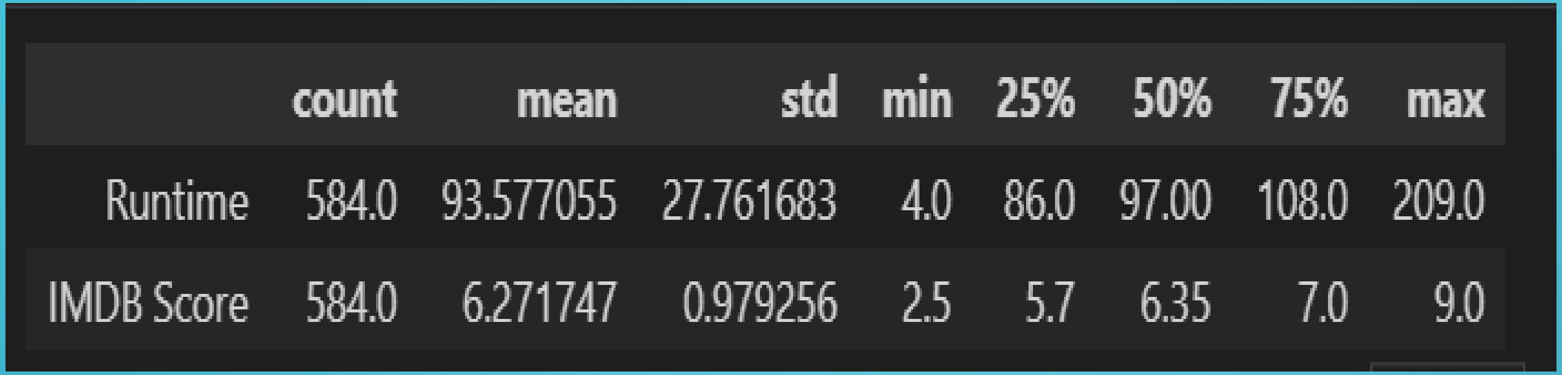
- import numpy as np # linear algebra
- import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
- import matplotlib.pyplot as plt
- import seaborn as sns
- import plotly.express as px
- from datetime import datetime,timedelta
- ds = pd.read_csv(“imdb.csv")
- ds_date = ds.copy()
- ds.head(5)

OUTPUT



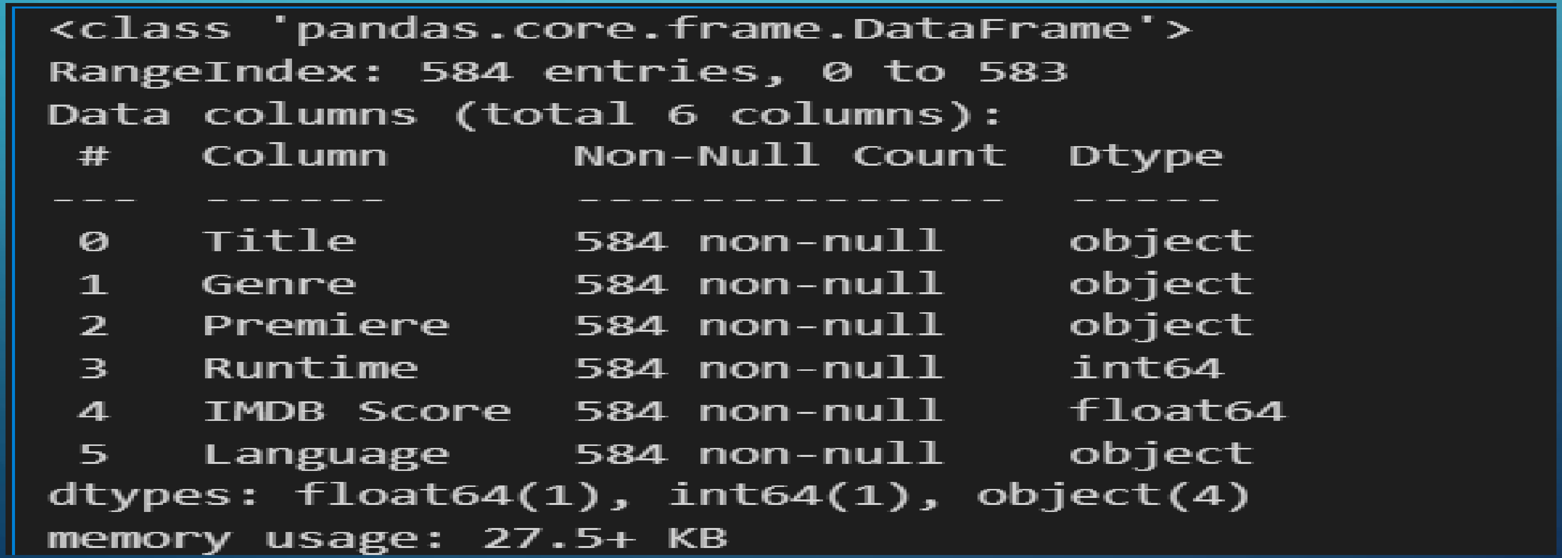
	Title	Genre	Premiere	Runtime	IMDB Score	Language
0	Enter the Anime	Documentary	August 5, 2019	58	2.5	English/Japanese
1	Dark Forces	Thriller	August 21, 2020	81	2.6	Spanish
2	The App	Science fiction/Drama	December 26, 2019	79	2.6	Italian
3	The Open House	Horror thriller	January 19, 2018	94	3.2	English
4	Kaali Khuhi	Mystery	October 30, 2020	90	3.4	Hindi

- ds.describe().T



	count	mean	std	min	25%	50%	75%	max
Runtime	584.0	93.577055	27.761683	4.0	86.0	97.00	108.0	209.0
IMDB Score	584.0	6.271747	0.979256	2.5	5.7	6.35	7.0	9.0

- ds.info(verbose=True,show_counts=True)

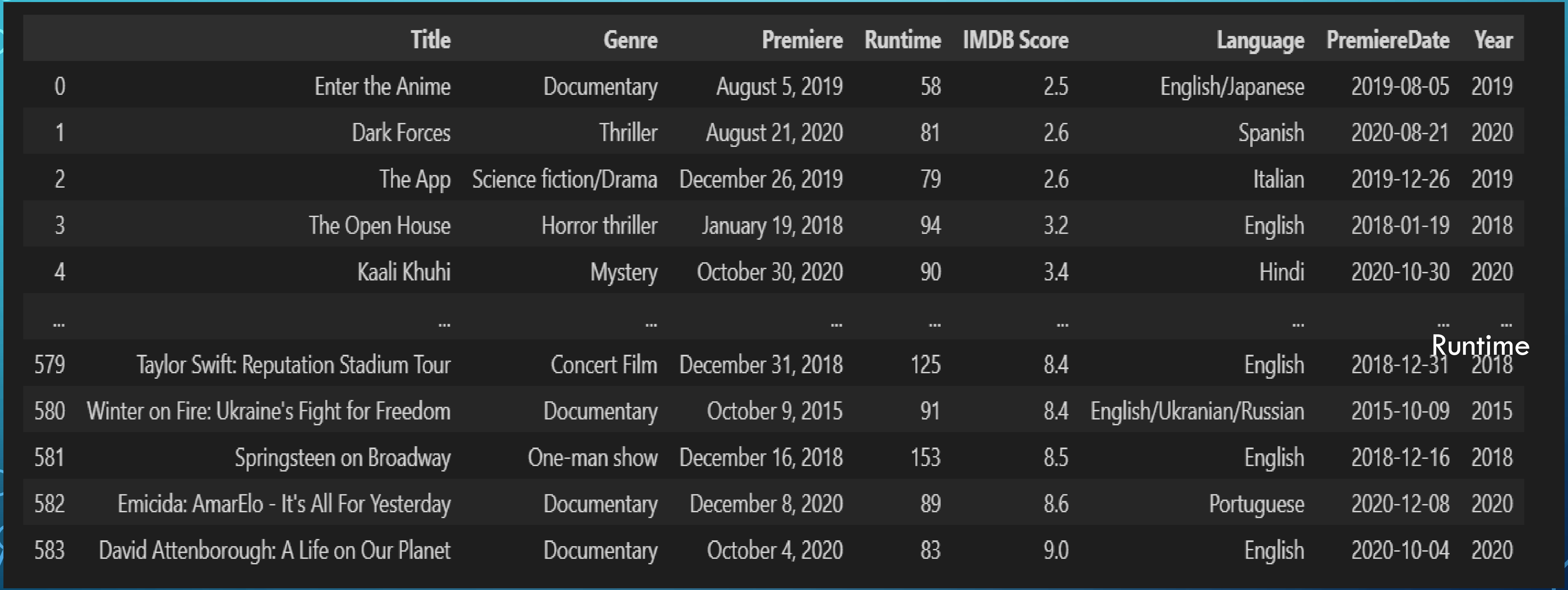


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 584 entries, 0 to 583
Data columns (total 6 columns):
 #    Column       Non-Null Count  Dtype
---   ------       --------------  -----
 0    Title        584 non-null    object
 1    Genre        584 non-null    object
 2    Premiere     584 non-null    object
 3    Runtime      584 non-null    int64
 4    IMDB Score   584 non-null    float64
 5    Language     584 non-null    object
dtypes: float64(1), int64(1), object(4)
memory usage: 27.5+ KB
```

DATASET PREMIERE

- ds_date["Premiere"] = ds_date["Premiere"].apply(lambda x: "".join(x for x in x.replace(".",",")))
- ds_date["PremiereDate"] = ds_date["Premiere"].apply(lambda x: datetime.strptime(x, "%B %d, %Y").date())
- ds_date["Year"] = ds_date["Premiere"].apply(lambda x: "".join(x for x in x.replace(",","").split()[-1]))
- ds_date["PremiereDate"] = pd.to_datetime(ds_date["PremiereDate"])
- ds_date

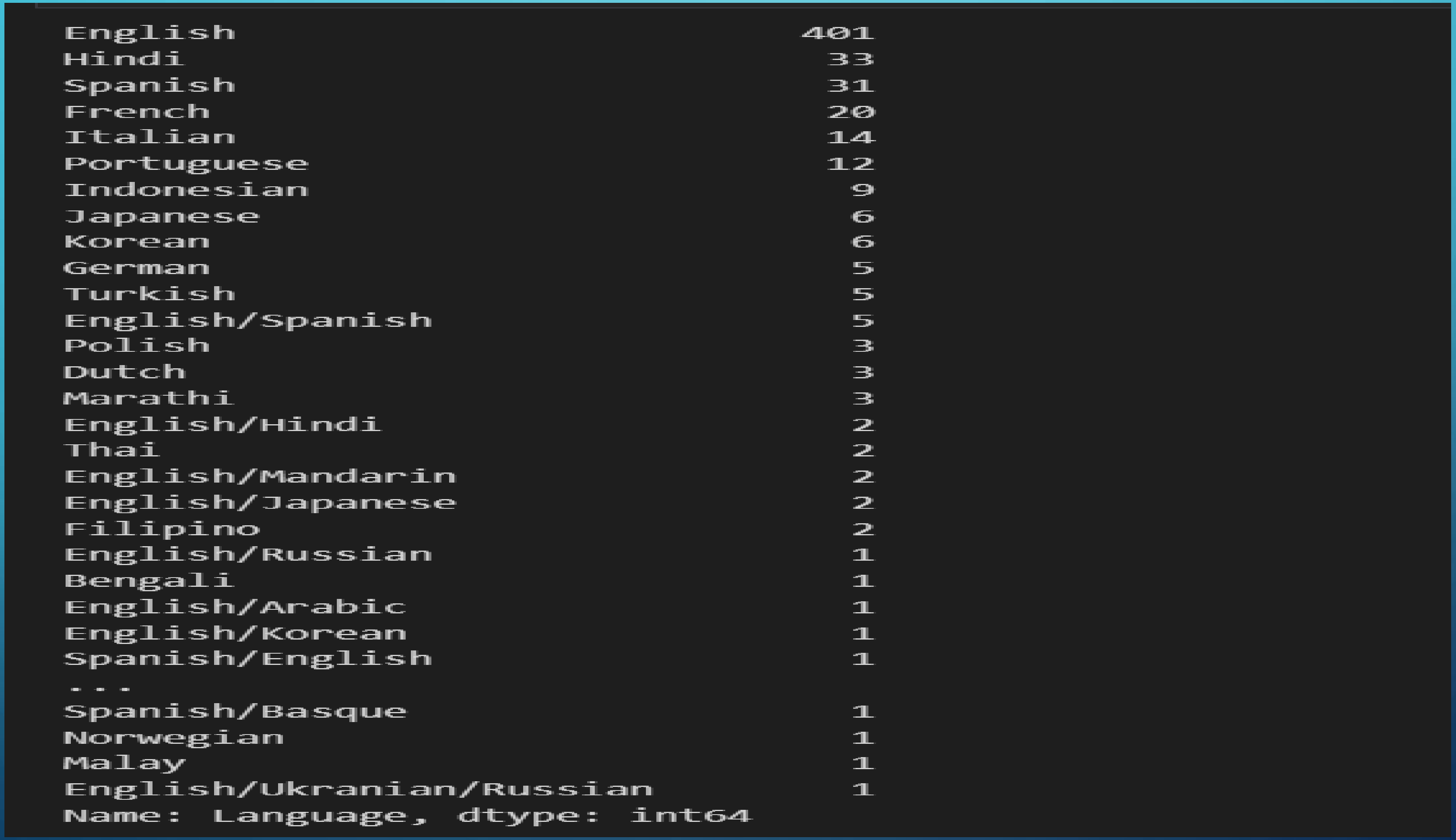
OUTPUT



	Title	Genre	Premiere	Runtime	IMDB Score	Language	PremiereDate	Year
0	Enter the Anime	Documentary	August 5, 2019	58	2.5	English/Japanese	2019-08-05	2019
1	Dark Forces	Thriller	August 21, 2020	81	2.6	Spanish	2020-08-21	2020
2	The App	Science fiction/Drama	December 26, 2019	79	2.6	Italian	2019-12-26	2019
3	The Open House	Horror thriller	January 19, 2018	94	3.2	English	2018-01-19	2018
4	Kaali Khuhi	Mystery	October 30, 2020	90	3.4	Hindi	2020-10-30	2020
...	...	...	...	...	...	...	...	...
579	Taylor Swift: Reputation Stadium Tour	Concert Film	December 31, 2018	125	8.4	English	2018-12-31	2018
580	Winter on Fire: Ukraine's Fight for Freedom	Documentary	October 9, 2015	91	8.4	English/Ukranian/Russian	2015-10-09	2015
581	Springsteen on Broadway	One-man show	December 16, 2018	153	8.5	English	2018-12-16	2018
582	Emicida: AmarElo - It's All For Yesterday	Documentary	December 8, 2020	89	8.6	Portuguese	2020-12-08	2020
583	David Attenborough: A Life on Our Planet	Documentary	October 4, 2020	83	9.0	English	2020-10-04	2020

Runtime

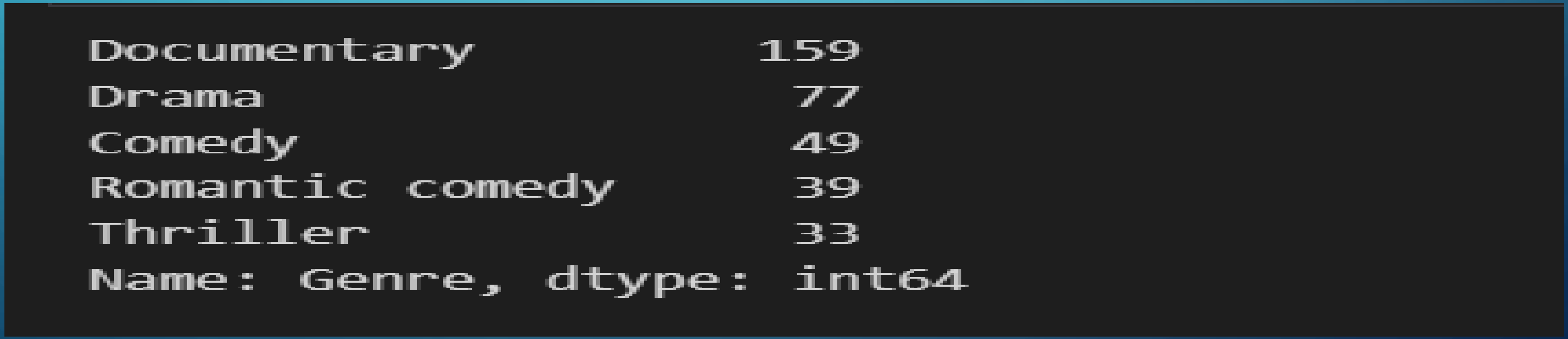
- ds['Language'].value_counts()



```
English                       401
Hindi                          33
Spanish                        31
French                         20
Italian                        14
Portuguese                     12
Indonesian                      9
Japanese                        6
Korean                          6
German                          5
Turkish                         5
English/Spanish                 5
Polish                          3
Dutch                           3
Marathi                         3
English/Hindi                   2
Thai                            2
English/Mandarin                2
English/Japanese                2
Filipino                        2
English/Russian                 1
Bengali                         1
English/Arabic                  1
English/Korean                  1
Spanish/English                 1
...
Spanish/Basque                  1
Norwegian                       1
Malay                           1
English/Ukranian/Russian        1
Name: Language, dtype: int64
```

- ds['Genre'].value_counts()
- genre = ds['Genre'].value_counts()
- genre.head()

OUTPUT

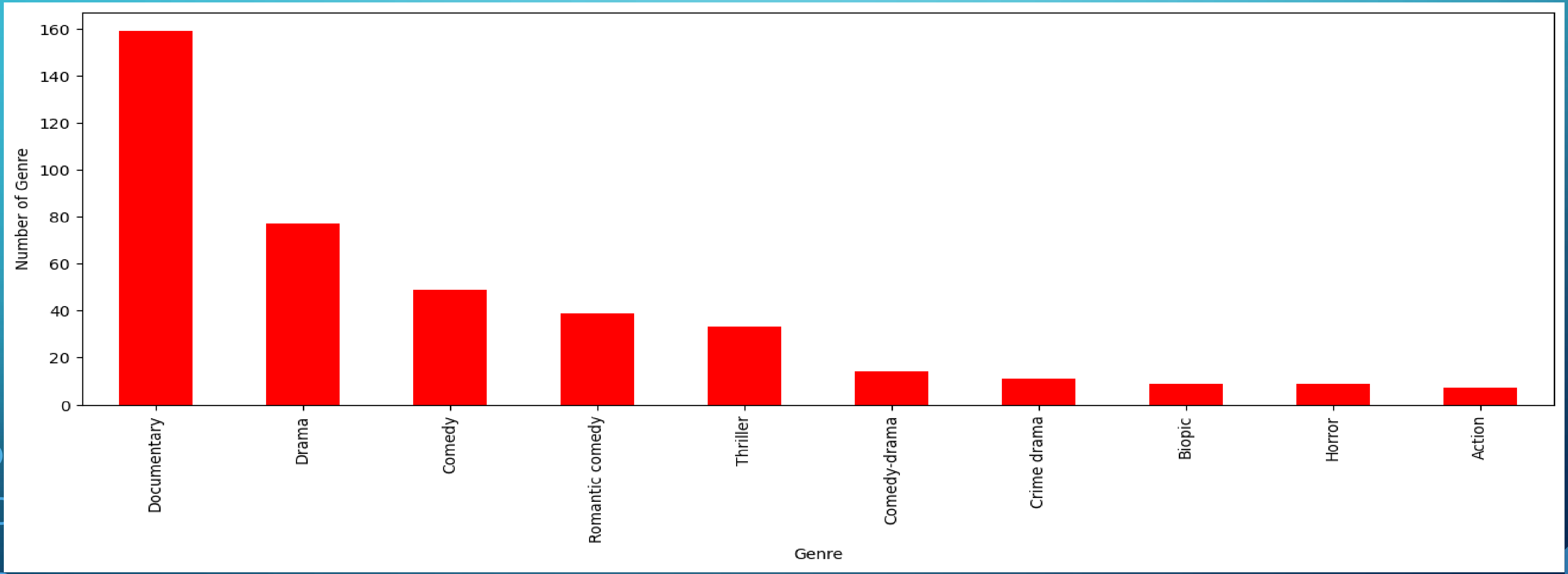


```
Documentary           159
Drama                  77
Comedy                 49
Romantic comedy        39
Thriller               33
Name: Genre, dtype: int64
```

MOST POPULAR MOVIES FROM GENRE

- plt.figure(figsize=(16, 5))
- ds['Genre'].value_counts().head(10).plot(kind='bar', color='red')
- plt.xlabel('Genre')
- plt.ylabel('Number of Genre')
- plt.xticks(rotation=90)
- plt.show(block=True)

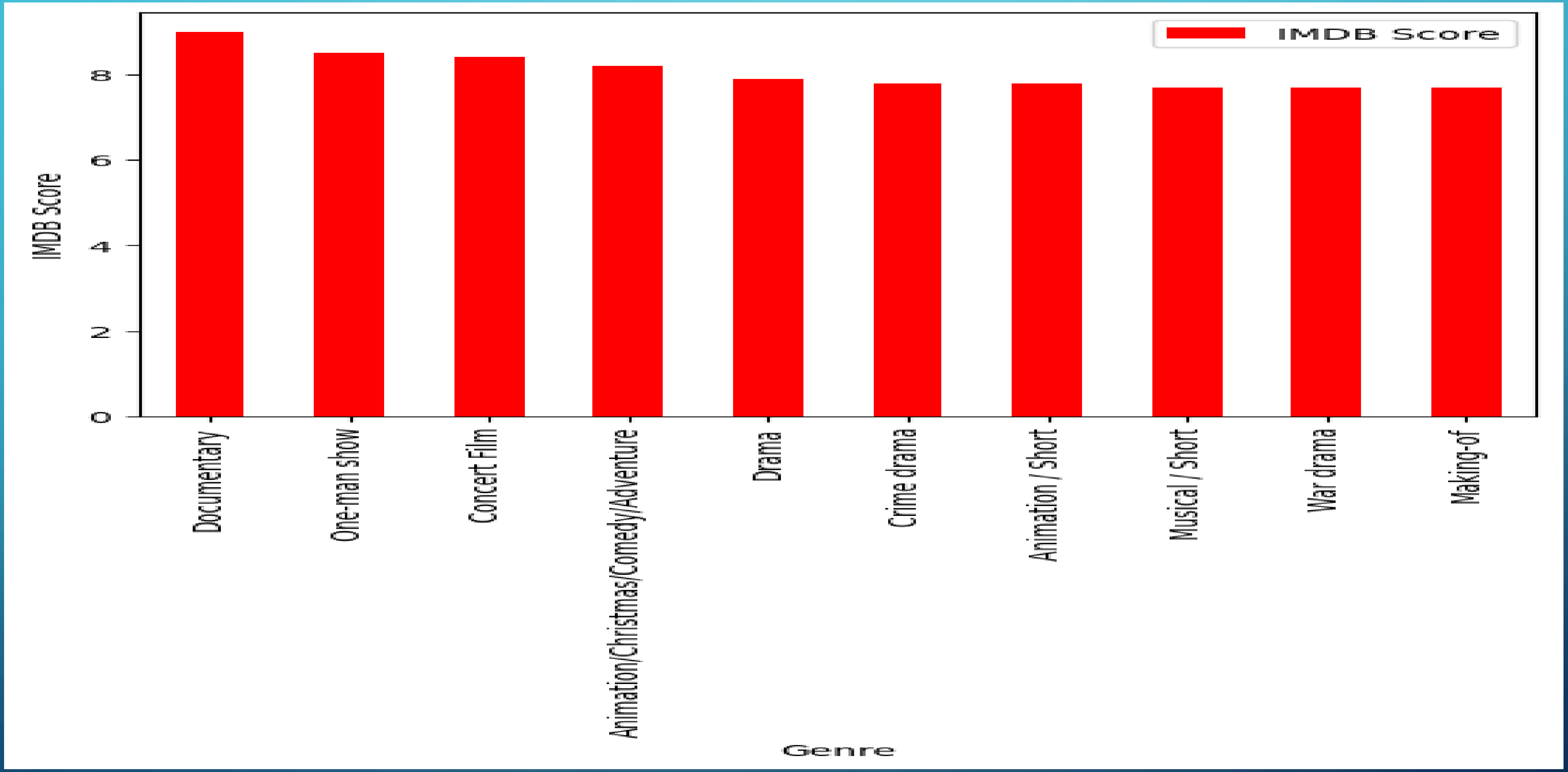
OUTPUT


160
140
120
100
80
60
40
20
0
Number of Genre
Documentary
Drama
Comedy
Romantic comedy
Thriller
Comedy-drama
Crime drama
Biopic
Horror
Action
Genre

IMDB SCORES

- ds[['Genre', 'IMDB Score']].sort_values('IMDB Score', ascending=False).drop_duplicates('Genre').head(10).plot(x='Genre',y='IMDB Score', kind='bar', color='red')
- plt.xlabel('Genre')
- plt.ylabel('IMDB Score')
- plt.show(block=True)

OUTPUT


IMDB Score
IMDB Score
0
2
4
6
8
Documentary
One-man show
Concert Film
Animation/Christmas/Comedy/Adventure
Drama
Crime drama
Animation / Short
Musical / Short
War drama
Making-of
Genre

CONCLUSION:

- Predicting IMDb Scores were classified and various processings were done using the given dataset