

```
import pandas as pd
```

```
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
DatasetBaseFolder = '/content/gdrive/MyDrive/ColabNotebooks/PandaTL/Datasets/'
```

```
movies = pd.read_csv(DatasetBaseFolder+"IMDB-Movie-Data.csv", index_col="Title")
```

```
movies.head(3)
```

	Rank		Genre	Description	Director	Actors	Year	Runtime (Minutes)
Title								
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	127	
ng...		X	Mystery,Sci-Fi	Following clues to the origin of	Ridley Scott	Noomi Rapace, Logan Marshall-	2012	124

Saving... X

```
movies.tail(3)
```

	Rank	Genre	Description	Director	Actors	Year	Run (Minut
Title							
Step Up 2: The Streets	998	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	
Search Party	999	Adventure,Comedy	A pair of friends embark on a mission to	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	

```
movies.shape
```

(1000, 11)

```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, Guardians of the Galaxy to Nine Lives
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Rank                   1000 non-null   int64
1   Genre                  1000 non-null   object
2   Description            1000 non-null   object
3   Director               1000 non-null   object
4   Actors                 1000 non-null   object
5   Year                   1000 non-null   int64
6   Runtime (Minutes)      1000 non-null   int64
7   Rating                 1000 non-null   float64
8   Votes                  1000 non-null   int64
9   Revenue (Millions)     872 non-null    float64
10  Metascore              936 non-null    float64
dtypes: float64(3), int64(4), object(4)
memory usage: 133.8+ KB
```

```
movies = movies.drop_duplicates(keep = 'first')
```

```
movies.columns
```

```
Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year',
       'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
       'Metascore'],
      dtype='object')
```

Saving...



```
movies.rename(columns={'Runtime (Minutes)': 'Runtime', 'Revenue (Millions)': 'Revenue_mil'})
movies.columns
```

```
Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year', 'Runtime',
       'Rating', 'Votes', 'Revenue_millions', 'Metascore'],
      dtype='object')
```

```
movies.isnull().sum()
```

```
Rank                0
Genre               0
Description          0
Director            0
Actors              0
Year                0
Runtime (Minutes)   0
Rating              0
Votes               0
Revenue (Millions) 128
```

```
Metascore          64
dtype: int64
```

```
moviesTmp = movies.dropna(axis=0)
moviesTmp.shape #ROWS
```

```
(838, 11)
```

```
moviesTmp = movies.dropna(axis=1)
moviesTmp.shape #COLUMNS
```

```
(1000, 9)
```

```
movies.shape
```

```
(1000, 11)
```

```
revenue = movies['Revenue_millions']
revenue.head(3)
```

```
Title
Guardians of the Galaxy    333.13
Prometheus                 126.46
Split                     138.12
Name: Revenue_millions, dtype: float64
```

```
meanRev = revenue.mean(0)
revenue.fillna(meanRev, inplace=True)
movies.isnull().sum()
```

Saving...



```
Description      0
Director         0
Actors           0
Year             0
Runtime          0
Rating           0
Votes            0
Revenue_millions 0
Metascore        64
dtype: int64
```

```
movies.describe()
```

	Rank	Year	Runtime	Rating	Votes	Revenue_millio
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.0000
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.9563
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	96.4120
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.0000
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	17.4425

movies['Genre'].value\_counts()

```
Action,Adventure,Sci-Fi    50
Drama                      48
Comedy,Drama,Romance       35
Comedy                     32
Drama,Romance              31
..
Comedy,Romance,Western     1
Adventure,Drama            1
Adventure                  1
Comedy,Mystery             1
Action,Horror,Romance      1
Name: Genre, Length: 207, dtype: int64
```

movies.corr()

	Rank	Year	Runtime	Rating	Votes	Revenue_millions
Rank	1.000000	-0.261605	-0.221739	-0.219555	-0.283876	-0.252996
Year	-0.261605	1.000000	-0.164900	-0.211219	-0.411904	-0.117562
Runtime	-0.221739	-0.164900	1.000000	0.392214	0.407062	0.247834
Rating	-0.219555	-0.211219	0.392214	1.000000	0.511537	0.189527
Votes	-0.283876	-0.411904	0.407062	0.511537	1.000000	0.607941
Revenue_millions	-0.252996	-0.117562	0.247834	0.189527	0.607941	1.000000
Metascore	-0.191869	-0.079305	0.211978	0.631897	0.325684	0.133328

movies.loc['Guardians of the Galaxy']  
movies.iloc[0]

```
Rank                                1
Genre                                Action,Adventure,Sci-Fi
Description    A group of intergalactic criminals are forced ...
Director                                James Gunn
Actors    Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...
Year                                2014
Runtime                                121
Rating                                8.1
```

Votes757074

Revenue\_millions333.13

Metascore76

Name: Guardians of the Galaxy, dtype: object

```
movie_subset = movies.iloc[1:5]
movie_subset
```

	Rank	Genre	Description	Director	Actors
Title					
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth...

```
rating = movies['Rating']
rating[rating.lt(8.0)]
```

Title

Saving...X

Sing7.2

Suicide Squad6.2

The Great Wall6.1

...

Secret in Their Eyes6.2

Hostel: Part II5.5

Step Up 2: The Streets6.2

Search Party5.6

Nine Lives5.3

Name: Rating, Length: 922, dtype: float64

```
moviesByRidley = movies[(movies['Director'] == "James Gunn") & movies['Rating'].gt(7.5)]
moviesByRidley.head(3)
```

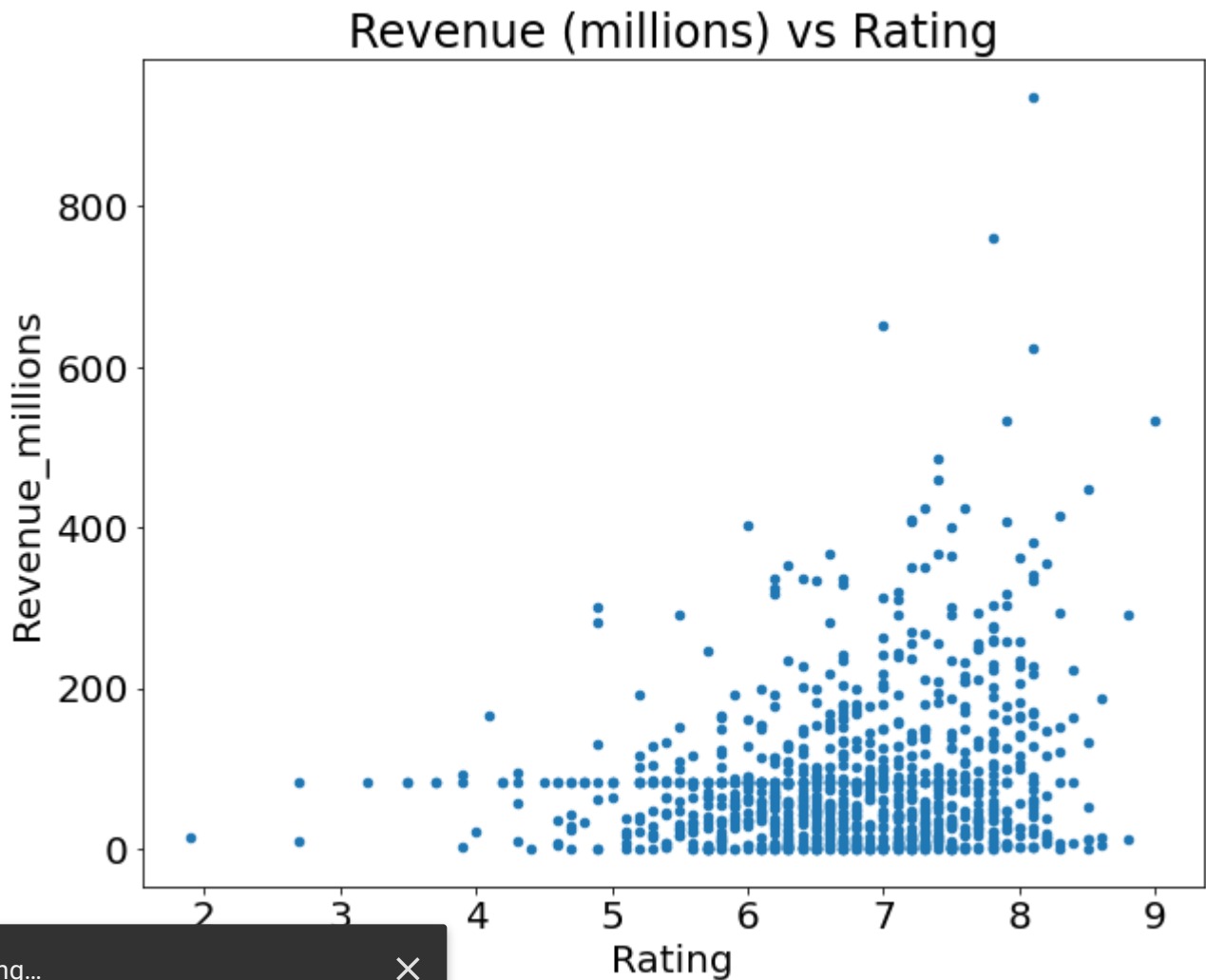
```
Rank      Genre  Description  Director  Actors  Year  Runtime  Ra
movies[
  ((movies['Year'] >= 2002) & (movies['Year'] <= 2015))
  & (movies['Rating'] > 5.0)
  & (movies['Revenue_millions'] < movies['Revenue_millions'].quantile(0.25))
]
```

	Rank		Genre	Description	Director	Actors	Year
Title							
Bahubali: The Beginning	27	Action,Adventure,Drama		In ancient India, an adventurous and daring ma...	S.S. Rajamouli	Prabhas, Rana Daggubati, Anushka Shetty,Tamann...	201
Legend	101	Biography,Crime,Drama		Identical twin gangsters Ronald and Reginald K...	Brian Helgeland	Tom Hardy, Emily Browning, Taron Egerton, Paul...	201
Whiplash	134	Drama,Music		A promising young drummer enrolls at a cut-thr...	Damien Chazelle	Miles Teller, J.K. Simmons, Melissa Benoist, P...	201
Room	144	Drama		A young boy is raised within the confines of a...	Lenny Abrahamson	Brie Larson, Jacob Tremblay, Sean Bridgers,Wen...	201
Warrior	195	Action,Drama,Sport		The youngest son of an alcoholic former boxer ...	Gavin O'Connor	Tom Hardy, Nick Nolte, Joel Edgerton, Jennifer...	201
...	...		...	...	...	...	.
I Spit on Your Grave	964	Crime,Horror,Thriller		A writer who is brutalized during her cabin re...	Steven R. Monroe	Sarah Butler, Jeff Branson, Andrew Howard,Dani...	201
The Walk	968	Adventure,Biography,Crime		In 1974, high-wire artist Philippe Petit recru...	Robert Zemeckis	Joseph Gordon-Levitt, Charlotte Le Bon,Guillau...	201
The Skin I	980	Drama Thriller		A brilliant plastic surgeon	Pedro	Antonio Banderas, Elena	201

```
import matplotlib.pyplot as plt
```

```
plt.rcParams.update({'font.size': 20, 'figure.figsize': (10, 8)})
```

```
movies.plot(kind='scatter', x='Rating', y='Revenue_millions', title='Revenue (millions) vs Ra
```

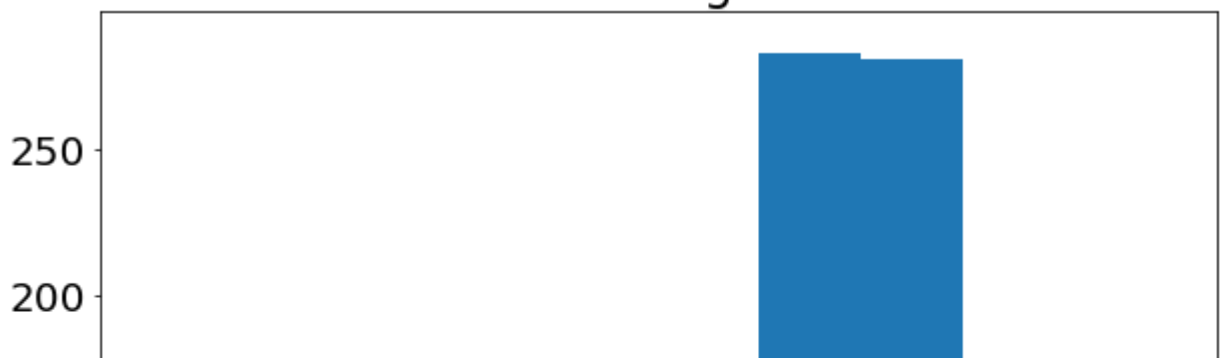


Saving...

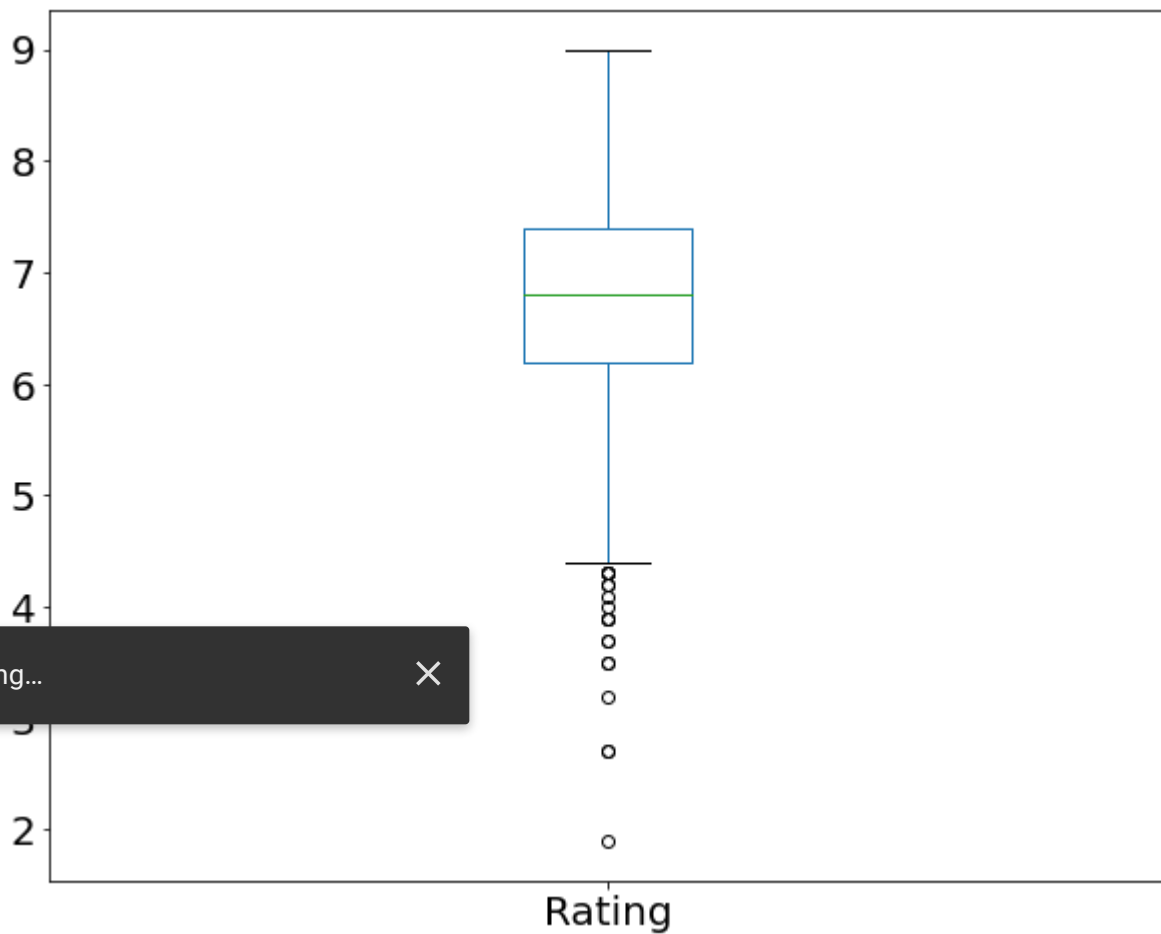


```
movies['Rating'].plot(kind='hist', title='Rating');
```

## Rating



```
movies['Rating'].plot(kind="box");
```



Saving...







Saving...

X