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

```
movies_df.head(15)
```

```
movies_df.tail(15)
```

Title	Rank	Genre	Description	Director	Actors
Your Highness	986	Adventure,Comedy,Fantasy	When Prince Fabious's bride is kidnapped, he g...	David Gordon Green	Danny McBride, Natalie Portman, James Franco, ...
Final Destination 5	987	Horror,Thriller	Survivors of a suspension-bridge collapse lear...	Steven Quale	Nicholas D'Agosto, Emma Bell, Arlen Escarpeta,...
Endless Love	988	Drama,Romance	The story of a privileged girl and a charismat...	Shana Feste	Gabriella Wilde, Alex Pettyfer, Bruce Greenwoo...
Martyrs	989	Horror	A young woman's quest for revenge against the ...	Pascal Laugier	Morjana Alaoui, Mylène Jampanoï, Catherine Bég...
Selma	990	Biography,Drama,History	A chronicle of Martin Luther King's campaign t...	Ava DuVernay	David Oyelowo, Carmen Ejogo, Tim Roth, Lorrain...
Underworld: Rise of the Lycans	991	Action,Adventure,Fantasy	An origins story centered on the centuries-old...	Patrick Tatopoulos	Rhona Mitra, Michael Sheen, Bill Nighy, Steven...
			An eight-		Darsheel Safary

```
movies_df.shape
```

```
(1000, 11)
```

```
movies_df.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: 93.8+ KB
```

```
movies_df = movies_df.drop_duplicates(keep = 'first') #Drop all instances keep = false inp
```

```
movies_df.columns
```

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

```
movies_df.rename(columns = {'Runtime (Minutes)' : 'Runtime', 'Revenue (Millions)' : 'Reven
movies_df.columns
```

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

```
movies_df.isnull().sum()
```

```
Rank                  0
Genre                 0
Description           0
Director              0
Actors                0
Year                  0
Runtime               0
Rating                0
Votes                 0
Revenue_millions    128
Metascore            64
dtype: int64
```

```
movies_dfTmp = movies_df.dropna(axis=0)
movies_dfTmp.shape
```

```
(838, 11)
```

```
movies_dfTmp = movies_df.dropna(axis=1)
movies_dfTmp.shape
```

```
(1000, 9)
```

```
movies_df.shape
```

```
(1000, 11)
```

```
revenue = movies_df['Revenue_millions']
revenue.head(5)
```

```
Title
Guardians of the Galaxy    333.13
Prometheus                 126.46
Split                      138.12
Sing                       270.32
Suicide Squad              325.02
Name: Revenue_millions, dtype: float64
```

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

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

```
movies_df.describe()
```

	Rank	Year	Runtime	Rating	Votes	Revenue_mil
count	1000.000000	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.0
mean	500.500000	2012.783000	113.172000	6.723200	1.698083e+05	82.9
std	288.819436	3.205962	18.810908	0.945429	1.887626e+05	96.4
min	1.000000	2006.000000	66.000000	1.900000	6.100000e+01	0.0
25%	250.750000	2010.000000	100.000000	6.200000	3.630900e+04	17.4
50%	500.500000	2014.000000	111.000000	6.800000	1.107990e+05	60.3
75%	750.250000	2016.000000	123.000000	7.400000	2.399098e+05	99.1
max	1000.000000	2016.000000	191.000000	9.000000	1.791916e+06	936.6

```
movies_df['Genre'].value_counts()
```

```
Action,Adventure,Sci-Fi     50
Drama                       48
Comedy,Drama,Romance        35
Comedy                      32
```

```
Drama,Romance                31
                             ..
Mystery,Romance,Thriller      1
Mystery,Thriller,Western      1
Comedy,Romance,Western        1
Biography,Drama,Mystery       1
Action,Horror,Romance         1
Name: Genre, Length: 207, dtype: int64
```

```
movies_df.corr()
```

	Rank	Year	Runtime	Rating	Votes	Revenue_milli
Rank	1.000000	-0.261605	-0.221739	-0.219555	-0.283876	-0.252
Year	-0.261605	1.000000	-0.164900	-0.211219	-0.411904	-0.117
Runtime	-0.221739	-0.164900	1.000000	0.392214	0.407062	0.247
Rating	-0.219555	-0.211219	0.392214	1.000000	0.511537	0.189
Votes	-0.283876	-0.411904	0.407062	0.511537	1.000000	0.607
Revenue_millions	-0.252996	-0.117562	0.247834	0.189527	0.607941	1.000
Metascore	-0.191869	-0.079305	0.211978	0.631897	0.325684	0.133

```
movies_df['Rating'].value_counts()
```

```
7.1    52
6.7    48
7.0    46
6.3    44
6.6    42
7.2    42
7.3    42
6.5    40
7.8    40
6.2    37
6.8    37
7.5    35
6.4    35
7.4    33
6.9    31
6.1    31
7.6    27
7.7    27
5.8    26
6.0    26
8.1    26
7.9    23
5.7    21
8.0    19
5.9    19
5.6    17
5.5    14
5.3    12
5.4    12
5.2    11
```

```
8.2    10
4.9     7
8.3     7
4.7     6
8.5     6
4.6     5
5.1     5
5.0     4
4.8     4
4.3     4
8.4     4
3.9     3
8.6     3
8.8     2
2.7     2
4.2     2
3.5     2
3.7     2
9.0     1
3.2     1
4.0     1
4.5     1
4.4     1
4.1     1
1.9     1
Name: Rating, dtype: int64
```

```
subset = movies_df[['Genre', 'Rating']]
type(subset)
```

```
pandas.core.frame.DataFrame
```

```
movies_df.loc['Prometheus']
movies_df.iloc[1]
```

```
Rank                                                          2
Genre                                       Adventure,Mystery,Sci-Fi
Description         Following clues to the origin of mankind, a te...
Director                                               Ridley Scott
Actors              Noomi Rapace, Logan Marshall-Green, Michael Fa...
Year                                                        2012
Runtime                                                      124
Rating                                                         7
Votes                                                     485820
Revenue_millions                                          126.46
Metascore                                                     65
Name: Prometheus, dtype: object
```

```
movie_subset = movies_df.iloc[1:4]
movie_subset
```

Title	Rank	Genre	Description	Director	Acto

```
rating = movies_df['Rating']
rating[rating.gt(8.5)]
```

```
Title
Interstellar        8.6
The Dark Knight     9.0
Inception           8.8
Kimi no na wa       8.6
Dangal              8.8
The Intouchables    8.6
Name: Rating, dtype: float64
```

```
moviesByRidley = movies_df[(movies_df['Director'] == "Ridley Scott") & movies_df['Rating']
moviesByRidley.head(4)
```

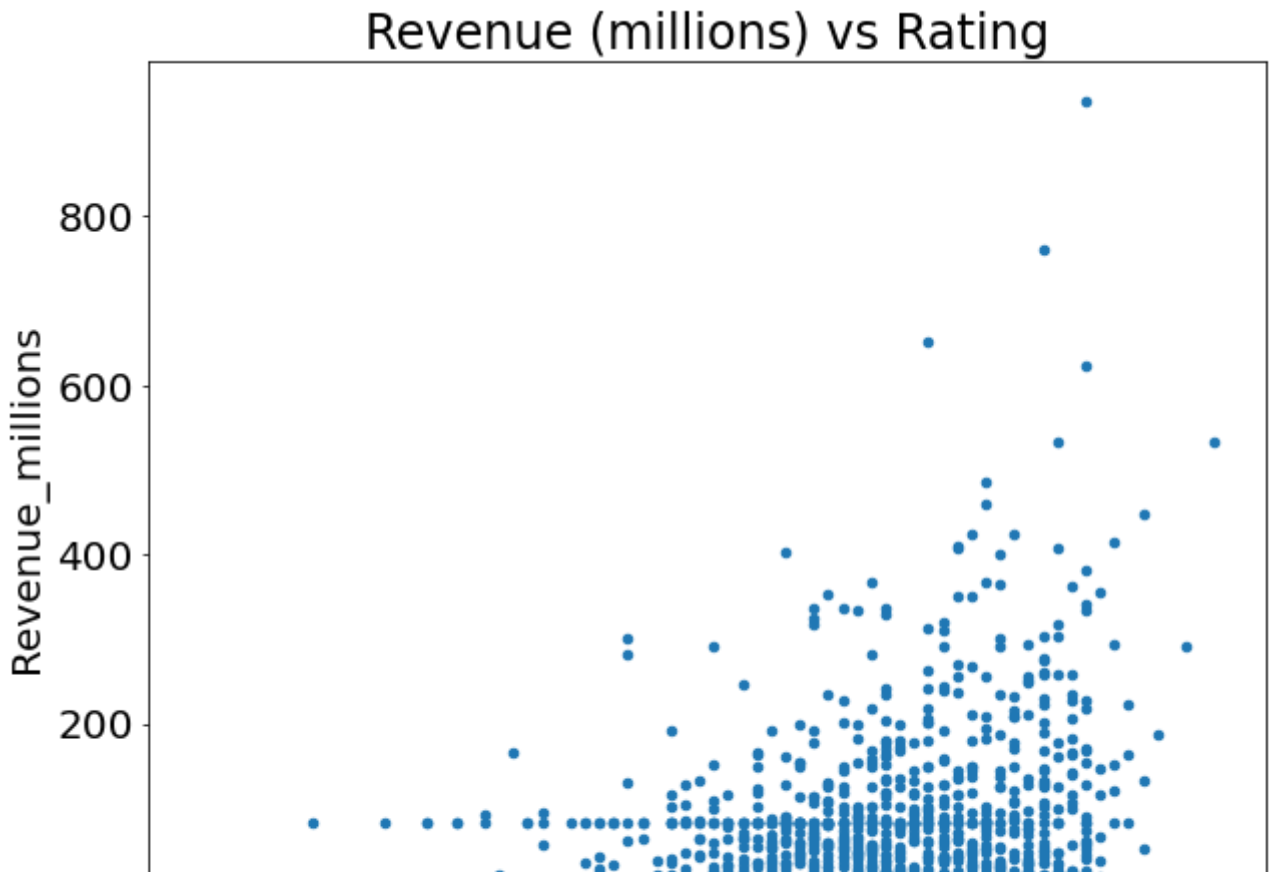
Title	Rank	Genre	Description	Director	Actors	Year	Runtin
The Martian	103	Adventure,Drama,Sci-Fi	An astronaut becomes stranded on Mars after	Ridley Scott	Matt Damon, Jessica Chastain,	2015	14

```
#all movies that were released between 2005 and 2010, have a rating above 8.0, but made be
movies_df[
    ((movies_df['Year'] >= 2005) & (movies_df['Year'] <= 2010))
    & (movies_df['Rating'] > 8.0)
    & (movies_df['Revenue_millions'] < movies_df['Revenue_millions'].quantile(0.25))]
```

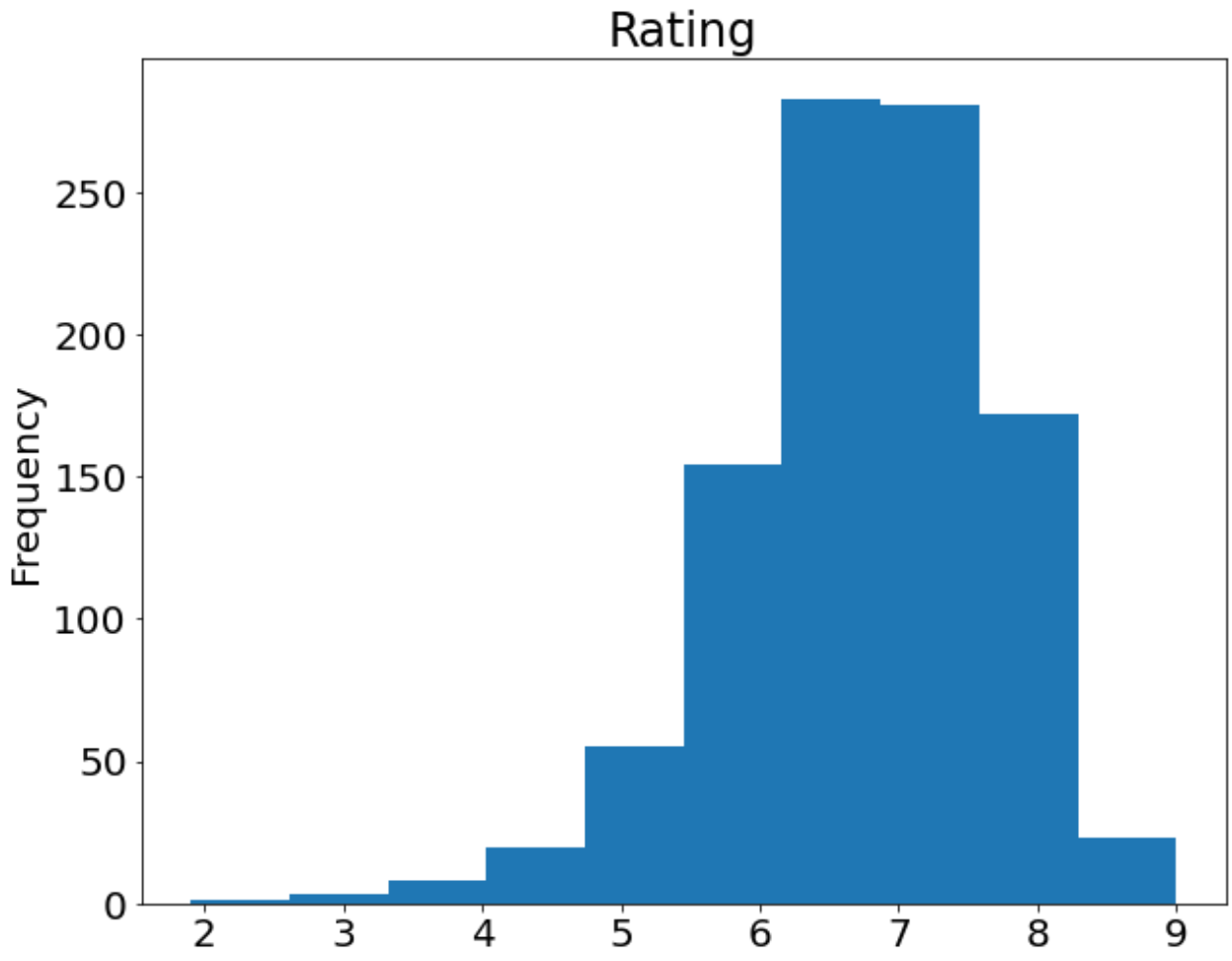
Title	Rank	Genre	Description	Director	Actors	Year
3 Idiots	431	Comedy,Drama	Two friends are searching for their long lost ...	Rajkumar Hirani	Aamir Khan, Madhavan, Mona Singh, Sharman Joshi	2009
The Lives of Others	477	Drama,Thriller	In 1984 East Berlin, an agent of the	Florian Henckel von	Ulrich Mühe, Martina Gedeck,Sebastian	2006

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

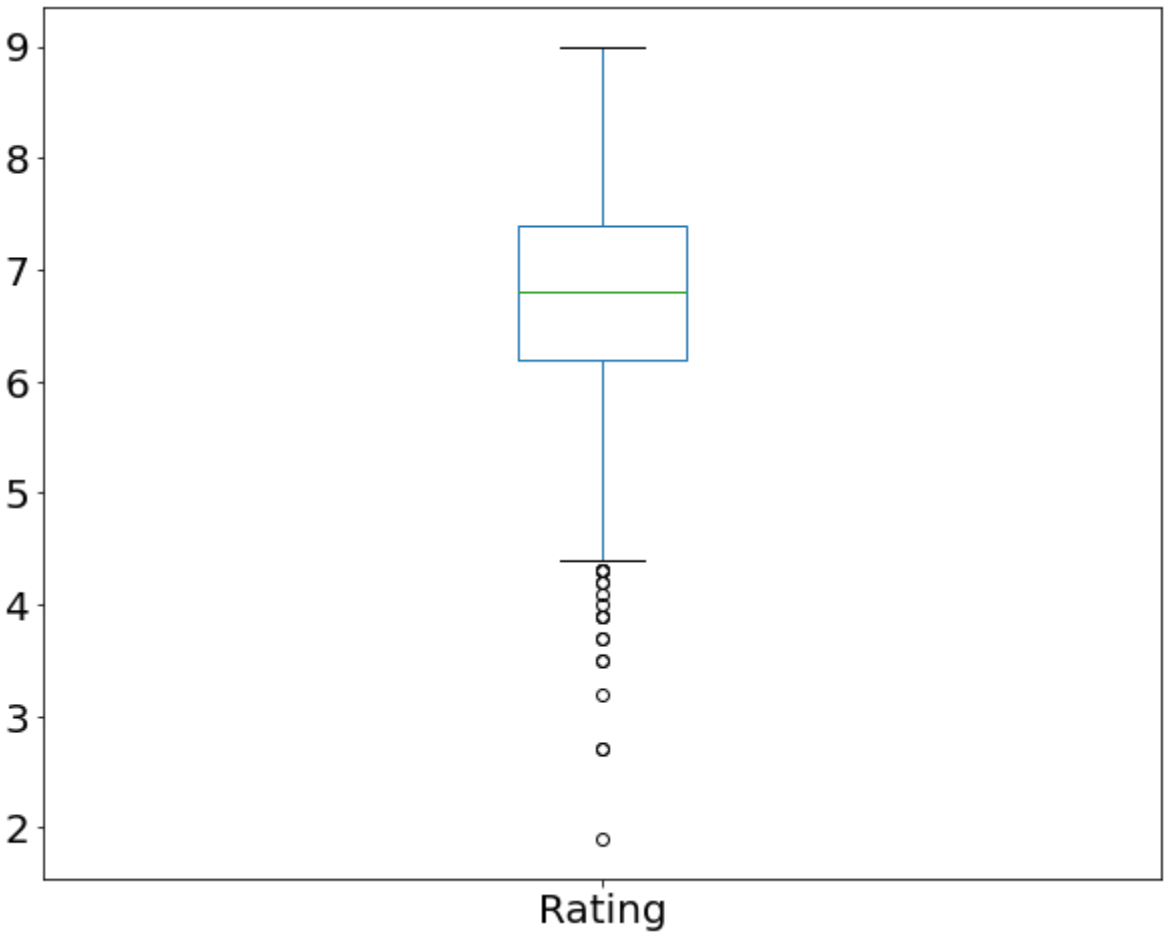
```
movies_df.plot(kind='scatter', x='Rating', y='Revenue_millions', title='Revenue (millions)
```




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




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



✓ 0s completed at 8:21 PM