

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [2]: movies = pd.read_csv('http://bit.ly/imdbratings')
```

```
In [3]: movies.shape
```

```
Out[3]: (979, 6)
```

```
In [4]: # Knowing the mean and standard deviation of the int columns
movies.describe()
```

```
Out[4]:
```

	star_rating	duration
count	979.000000	979.000000
mean	7.889785	120.979571
std	0.336069	26.218010
min	7.400000	64.000000
25%	7.600000	102.000000
50%	7.800000	117.000000
75%	8.100000	134.000000
max	9.300000	242.000000

```
In [5]: movies.head()
```

```
Out[5]:
```

	star_rating	title	content_rating	genre	duration	actors_list
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...
4	8.9	Pulp Fiction	R	Crime	154	[u'John Travolta', u'Uma Thurman', u'Samuel L....

```
In [6]: # How many movies have duration of 200 and more?

lengthy = movies[movies['duration'] >= 200]
```

```
In [12]: movies.genre.value_counts().head()
```

```
Out[12]: Drama      278
Comedy      156
Action      136
Crime       124
Biography    77
Name: genre, dtype: int64
```

```
In [10]: lengthy[['title', 'genre']].count()
```

```
Out[10]: title      12
genre      12
dtype: int64
```

```
In [9]: # Grouping the movies by genre and sorting it by star rating
```

```
movies.groupby('genre').head().sort_values('star_rating')
```

Out[9]:	star_rating	title	content_rating	genre	duration	actors_list
902	7.5	Frenzy	R	Thriller	116	[u'Jon Finch', u'Barry Foster', u'Alec McCowen']
848	7.5	The Man Who Knew Too Much	PG	Thriller	120	[u'James Stewart', u'Doris Day', u'Brenda de B...
728	7.6	La piel que habito	R	Thriller	120	[u'Antonio Banderas', u'Elena Anaya', u'Jan Co...
687	7.7	Spellbound	UNRATED	Film-Noir	111	[u'Ingrid Bergman', u'Gregory Peck', u'Michael...
638	7.7	The City of Lost Children	R	Fantasy	112	[u'Ron Perlman', u'Daniel Emilfork', u'Judith ...
...
6	8.9	The Good, the Bad and the Ugly	NOT RATED	Western	161	[u'Clint Eastwood', u'Eli Wallach', u'Lee Van ...
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...

67 rows × 6 columns

```
In [8]: # Count of movies whose star rating is below 8.0

movies[movies['star_rating'] < 8.0].count()
```

```
Out[8]: star_rating    585
title              585
content_rating     583
genre              585
duration           585
actors_list        585
dtype: int64
```

```
In [ ]: # Deleting actors_list column
del movies['actors_list']
```

```
In [ ]: # Reloading the url to bring back the deleted column
```

```
In [13]: # Fetching thriller movies with the maximum duration of time

movies[movies['genre'] == 'Thriller'].duration.max()
```

Out[13]: 120

```
In [14]: movies[movies['genre'] == 'Thriller']
```

Out[14]:	star_rating	title	content_rating	genre	duration	actors_list
350	8.0	Shadow of a Doubt	APPROVED	Thriller	108	[u'Teresa Wright', u'Joseph Cotten', u'Macдона...
496	7.8	Misery	R	Thriller	107	[u'James Caan', u'Kathy Bates', u'Richard Farn...
728	7.6	La piel que habito	R	Thriller	120	[u'Antonio Banderas', u'Elena Anaya', u'Jan Co...
848	7.5	The Man Who Knew Too Much	PG	Thriller	120	[u'James Stewart', u'Doris Day', u'Brenda de B...
902	7.5	Frenzy	R	Thriller	116	[u'Jon Finch', u'Barry Foster', u'Alec McCowen']

```
In [16]: movies.sort_values('duration').head()
```

Out[16]:	star_rating	title	content_rating	genre	duration	actors_list
389	8.0	Freaks	UNRATED	Drama	64	[u'Wallace Ford', u'Leila Hyams', u'Olga Bacla...
338	8.0	Battleship Potemkin	UNRATED	History	66	[u'Aleksandr Antonov', u'Vladimir Barsky', u'G...
258	8.1	The Cabinet of Dr. Caligari	UNRATED	Crime	67	[u'Werner Krauss', u'Conrad Veidt', u'Friedric...
293	8.1	Duck Soup	PASSED	Comedy	68	[u'Groucho Marx', u'Harpo Marx', u'Chico Marx']

```
In [17]: movies.columns
```

```
Out[17]: Index(['star_rating', 'title', 'content_rating', 'genre', 'duration',  
             'actors_list'],  
            dtype='object')
```

```
In [18]: movies.isnull().sum()
```

```
Out[18]: star_rating    0  
title                0  
content_rating      3  
genre               0  
duration            0  
actors_list         0  
dtype: int64
```

```
In [19]: # Filtering out not rated movies with at least 8.0 star rating  
  
movies[(movies['content_rating'] == 'NOT RATED') & (movies['star_rating'] >= 8.0)].sort_values('genre').head(4)
```

```
Out[19]:
```

	star_rating	title	content_rating	genre	duration	actors_list
100	8.3	The Treasure of the Sierra Madre	NOT RATED	Action	126	[u'Humphrey Bogart', u'Walter Huston', u'Tim H...]
140	8.3	The Wages of Fear	NOT RATED	Adventure	131	[u'Yves Montand', u'Charles Vanel', u'Peter va...]
328	8.0	Aguirre, the Wrath of God	NOT RATED	Adventure	93	[u'Klaus Kinski', u'Ruy Guerra', u'Helena Rojo']
133	8.3	Barfi!	NOT RATED	Adventure	151	[u'Ranbir Kapoor', u'Priyanka Chopra', u'Ileana']

```
In [21]: movies[(movies['content_rating'] == 'NOT RATED') & (movies['star_rating'] >= 8.0)][['genre', 'duration']].head(8)
```

```
Out[21]:
```

	genre	duration
5	Drama	96
6	Western	161
41	Drama	110
63	Crime	99
66	Comedy	156
72	Drama	157
83	Drama	129
87	Drama	93

```
In [25]: # Changing the columns to a proper case  
movies.columns = [x.title() for x in movies.columns]
```

```
In [26]: movies.head()
```

```
Out[26]:
```

	Star_Rating	Title	Content_Rating	Genre	Duration	Actors_List
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...]
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...]
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...]
4	8.9	Pulp Fiction	R	Crime	154	[u'John Travolta', u'Uma Thurman', u'Samuel L....]

```
In [28]: # Fetching movies in which Morgan Freeman featured
```

```
mov_Freeman = movies['Actors_List'].str.contains('Morgan Freeman', na = False)
```

```
In [29]: movies.loc[mov_Freeman, 'Star_Rating':'Actors_List']
```

Out[29]:

	Star_Rating	Title	Content_Rating	Genre	Duration	Actors_List
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
24	8.7	Se7en	R	Drama	127	[u'Morgan Freeman', u'Brad Pitt', u'Kevin Spac...
119	8.3	Unforgiven	R	Western	131	[u'Clint Eastwood', u'Gene Hackman', u'Morgan ...
227	8.1	Million Dollar Baby	PG-13	Drama	132	[u'Hilary Swank', u'Clint Eastwood', u'Morgan ...
549	7.8	Lucky Number Slevin	R	Crime	110	[u'Josh Hartnett', u'Ben Kingsley', u'Morgan F...
621	7.7	Gone Baby Gone	R	Crime	114	[u'Morgan Freeman', u'Ed Harris', u'Casey Affl...
943	7.4	The Bucket List	PG-13	Adventure	97	[u'Jack Nicholson', u'Morgan Freeman', u'Sean ...
962	7.4	Driving Miss Daisy	PG	Comedy	99	[u'Morgan Freeman', u'Jessica Tandy', u'Dan Ay...

```
In [37]: movies.loc[mov_Freeman, 'Star_Rating':'Actors_List'].count()
```

Out[37]:

Star_Rating	8
Title	8
Content_Rating	8
Genre	8
Duration	8
Actors_List	8
dtype:	int64

In []:

```
In [39]: # Knowing the distribution of movies by genre

movies.groupby('Genre').describe()
```

Out[39]:

Genre	Star_Rating								Duration							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
Action	136.0	7.884559	0.333084	7.4	7.600	7.80	8.100	9.0	136.0	126.485294	21.982480	80.0	111.00	125.0	137.00	205.0
Adventure	75.0	7.933333	0.361802	7.4	7.650	7.90	8.100	8.9	75.0	134.840000	33.383942	89.0	109.00	127.0	151.50	224.0
Animation	62.0	7.914516	0.292444	7.4	7.700	7.90	8.100	8.6	62.0	96.596774	13.975927	75.0	87.00	94.5	102.00	134.0
Biography	77.0	7.862338	0.310364	7.4	7.600	7.80	8.100	8.9	77.0	131.844156	24.283991	85.0	116.00	127.0	146.00	202.0
Comedy	156.0	7.822436	0.304922	7.4	7.600	7.80	8.000	8.6	156.0	107.602564	20.101574	68.0	94.00	104.0	117.00	187.0
Crime	124.0	7.916935	0.392143	7.4	7.600	7.80	8.100	9.3	124.0	122.298387	25.054406	67.0	104.75	118.0	134.25	229.0
Drama	278.0	7.902518	0.328041	7.4	7.600	7.90	8.100	8.9	278.0	126.539568	25.637086	64.0	108.00	123.0	138.00	242.0
Family	2.0	7.850000	0.070711	7.8	7.825	7.85	7.875	7.9	2.0	107.500000	10.606602	100.0	103.75	107.5	111.25	115.0
Fantasy	1.0	7.700000	NaN	7.7	7.700	7.70	7.700	7.7	1.0	112.000000	NaN	112.0	112.00	112.0	112.00	112.0
Film-Noir	3.0	8.033333	0.305505	7.7	7.900	8.10	8.200	8.3	3.0	97.333333	12.096832	88.0	90.50	93.0	102.00	111.0
History	1.0	8.000000	NaN	8.0	8.000	8.00	8.000	8.0	1.0	66.000000	NaN	66.0	66.00	66.0	66.00	66.0
Horror	29.0	7.806897	0.337989	7.4	7.500	7.70	8.000	8.6	29.0	102.517241	18.908843	70.0	85.00	104.0	116.00	146.0
Mystery	16.0	7.975000	0.387298	7.4	7.575	8.10	8.225	8.6	16.0	115.625000	22.585762	69.0	103.75	115.0	129.25	160.0
Sci-Fi	5.0	7.920000	0.192354	7.7	7.800	7.90	8.000	8.2	5.0	109.000000	17.478558	91.0	92.00	113.0	117.00	132.0
Thriller	5.0	7.680000	0.216795	7.5	7.500	7.60	7.800	8.0	5.0	114.200000	6.340347	107.0	108.00	116.0	120.00	120.0
Western	9.0	8.255556	0.387657	7.6	8.100	8.30	8.500	8.9	9.0	136.666667	28.748913	85.0	131.00	135.0	161.00	175.0

```
In [49]: # Aggregating max and min for duration and star rating

movies.groupby('Genre').agg({'Duration': ['max', 'min'], 'Star_Rating': 'min'}).head()
```

Out[49]:

	Duration	Star_Rating
	max	min

Genre			
Action	205	80	7.4
Adventure	224	89	7.4
Animation	134	75	7.4
Biography	202	85	7.4
Comedy	187	68	7.4

In [47]:

```
# Transposing the result for better visualization
movies.groupby('Genre').agg({'Duration': ['max', 'min'], 'Star_Rating': 'min'}).T
```

Out[47]:

	Genre	Action	Adventure	Animation	Biography	Comedy	Crime	Drama	Family	Fantasy	Film-Noir	History	Horror	Mystery	Sci-Fi
Duration	max	205.0	224.0	134.0	202.0	187.0	229.0	242.0	115.0	112.0	111.0	66.0	146.0	160.0	132.0
	min	80.0	89.0	75.0	85.0	68.0	67.0	64.0	100.0	112.0	88.0	66.0	70.0	69.0	91.0
Star_Rating	min	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.8	7.7	7.7	8.0	7.4	7.4	7.7

In [48]:

```
# Fetching results for Action movie
movies.groupby('Genre').agg({'Duration': ['max', 'min'], 'Star_Rating': 'min'}).loc['Action']
```

Out[48]:

Duration max 205.0
 min 80.0
Star_Rating min 7.4
Name: Action, dtype: float64

In []:

```
movies.dtypes
```

In []: