

```
In [1]: import pandas as pd
import numpy as np
from IPython.display import display
```

# Data Frame Overview

## (1) Dissecting the Anatomy of a Data Frame

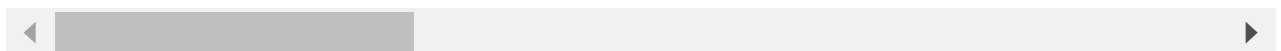
```
In [2]: import pandas as pd

movie = pd.read_csv('movie.csv')
movie
```

```
Out[2]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300
4	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns



```
In [3]: type(movie)
```

```
Out[3]: pandas.core.frame.DataFrame
```

```
In [4]: movie.shape
```

Out[4]: (4916, 28)

In [5]:

```
movie.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4916 entries, 0 to 4915
Data columns (total 28 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   color                                4897 non-null   object
1   director_name                        4814 non-null   object
2   num_critic_for_reviews               4867 non-null   float64
3   duration                             4901 non-null   float64
4   director_facebook_likes              4814 non-null   float64
5   actor_3_facebook_likes               4893 non-null   float64
6   actor_2_name                         4903 non-null   object
7   actor_1_facebook_likes               4909 non-null   float64
8   gross                               4054 non-null   float64
9   genres                               4916 non-null   object
10  actor_1_name                         4909 non-null   object
11  movie_title                          4916 non-null   object
12  num_voted_users                      4916 non-null   int64
13  cast_total_facebook_likes            4916 non-null   int64
14  actor_3_name                         4893 non-null   object
15  facenumber_in_poster                 4903 non-null   float64
16  plot_keywords                        4764 non-null   object
17  movie_imdb_link                      4916 non-null   object
18  num_user_for_reviews                 4895 non-null   float64
19  language                             4904 non-null   object
20  country                              4911 non-null   object
21  content_rating                       4616 non-null   object
22  budget                               4432 non-null   float64
23  title_year                           4810 non-null   float64
24  actor_2_facebook_likes               4903 non-null   float64
25  imdb_score                           4916 non-null   float64
26  aspect_ratio                         4590 non-null   float64
27  movie_facebook_likes                 4916 non-null   int64
dtypes: float64(13), int64(3), object(12)
memory usage: 1.1+ MB
```

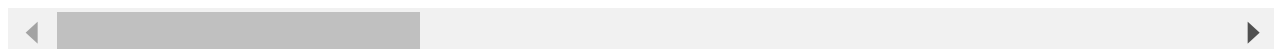
In [6]:

```
movie.head()
```

Out[6]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
0	Color	James Cameron	723.0	178.0	0.0	855.0
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0
4	NaN	Doug Walker	NaN	NaN	131.0	NaN

5 rows × 28 columns

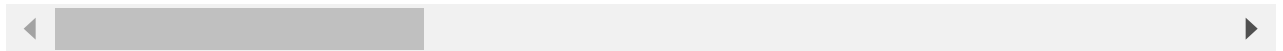


```
In [7]: movie.tail()
```

```
Out[7]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns



## (2) Accessing the Main Data Frame Components

```
In [8]: columns = movie.columns
index = movie.index
data = movie.values
```

```
In [9]: columns
```

```
Out[9]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
              'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
              'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
              'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
              'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
              'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
              'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
              'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],
              dtype='object')
```

```
In [10]: index
```

```
Out[10]: RangeIndex(start=0, stop=4916, step=1)
```

```
In [11]: data
```

```
Out[11]: array([[ 'Color', 'James Cameron', 723.0, ..., 7.9, 1.78, 33000],
                 [ 'Color', 'Gore Verbinski', 302.0, ..., 7.1, 2.35, 0],
                 [ 'Color', 'Sam Mendes', 602.0, ..., 6.8, 2.35, 85000],
                 ...,
                 [ 'Color', 'Benjamin Roberds', 13.0, ..., 6.3, nan, 16],
```

```
['Color', 'Daniel Hsia', 14.0, ..., 6.3, 2.35, 660],
['Color', 'Jon Gunn', 43.0, ..., 6.6, 1.85, 456]], dtype=object)
```

### (3) Understanding Data Types

```
In [12]: movie.dtypes
```

```
Out[12]: color                object
director_name                object
num_critic_for_reviews       float64
duration                     float64
director_facebook_likes      float64
actor_3_facebook_likes       float64
actor_2_name                 object
actor_1_facebook_likes       float64
gross                       float64
genres                      object
actor_1_name                 object
movie_title                  object
num_voted_users              int64
cast_total_facebook_likes    int64
actor_3_name                 object
facenumber_in_poster        float64
plot_keywords                object
movie_imdb_link              object
num_user_for_reviews        float64
language                     object
country                      object
content_rating               object
budget                      float64
title_year                   float64
actor_2_facebook_likes       float64
imdb_score                   float64
aspect_ratio                 float64
movie_facebook_likes         int64
dtype: object
```

```
In [13]: movie.dtypes.value_counts()
```

```
Out[13]: float64    13
object      12
int64       3
dtype: int64
```

```
In [14]: movie
```

```
Out[14]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
4	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns

```
In [15]: movie['title_year']
```

```
Out[15]: 0      2009.0
1      2007.0
2      2015.0
3      2012.0
4         NaN
...
4911    2013.0
4912         NaN
4913    2013.0
4914    2012.0
4915    2004.0
Name: title_year, Length: 4916, dtype: float64
```

```
In [16]: movie['title_year'].dtypes
```

```
Out[16]: dtype('float64')
```

```
In [17]: # Example: Pandas Data Type Conversion

movie['title_year'].astype(object)
```

```
Out[17]: 0      2009.0
1      2007.0
2      2015.0
3      2012.0
4         NaN
...
4911    2013.0
4912         NaN
4913    2013.0
4914    2012.0
4915    2004.0
Name: title_year, Length: 4916, dtype: object
```

```
In [18]: movie['title_year'].dtypes
```

```
Out[18]: dtype('float64')
```

## (4) Series vs. Data Frame

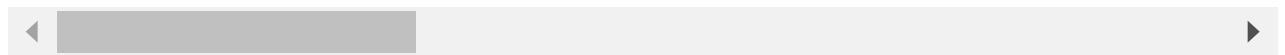
```
In [19]: movie = pd.read_csv('movie.csv')
```

```
In [20]: movie
```

```
Out[20]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300
4	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns



```
In [21]: type(movie)
```

```
Out[21]: pandas.core.frame.DataFrame
```

```
In [22]: movie['director_name']
```

```
Out[22]: 0      James Cameron
1      Gore Verbinski
2      Sam Mendes
```

```
3      Christopher Nolan
4      Doug Walker
...
4911     Scott Smith
4912           NaN
4913     Benjamin Roberds
4914     Daniel Hsia
4915       Jon Gunn
Name: director_name, Length: 4916, dtype: object
```

```
In [23]: type(movie['director_name'])
```

```
Out[23]: pandas.core.series.Series
```

```
In [24]: director = movie['director_name'] # Save to a variable
         director
```

```
Out[24]: 0      James Cameron
         1      Gore Verbinski
         2      Sam Mendes
         3      Christopher Nolan
         4      Doug Walker
...
4911     Scott Smith
4912           NaN
4913     Benjamin Roberds
4914     Daniel Hsia
4915       Jon Gunn
Name: director_name, Length: 4916, dtype: object
```

```
In [25]: type(director)
```

```
Out[25]: pandas.core.series.Series
```

```
In [26]: director.to_frame()
```

```
Out[26]:
```

	director_name
0	James Cameron
1	Gore Verbinski
2	Sam Mendes
3	Christopher Nolan
4	Doug Walker
...	...
4911	Scott Smith
4912	NaN
4913	Benjamin Roberds
4914	Daniel Hsia
4915	Jon Gunn

4916 rows × 1 columns

```
In [27]: type(director.to_frame())
```

```
Out[27]: pandas.core.frame.DataFrame
```

# Data Frame Manipulation

## 1) Columns

### (1) Adding a Column

```
In [28]: movie = pd.read_csv('movie.csv')
```

```
In [29]: movie.columns
```

```
Out[29]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',  
              'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',  
              'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',  
              'movie_title', 'num_voted_users', 'cast_total_facebook_likes',  
              'actor_3_name', 'facenumber_in_poster', 'plot_keywords',  
              'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',  
              'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',  
              'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],  
              dtype='object')
```

```
In [30]: movie['has_seen'] = 0
```

```
In [31]: movie.columns
```

```
Out[31]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',  
              'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',  
              'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',  
              'movie_title', 'num_voted_users', 'cast_total_facebook_likes',  
              'actor_3_name', 'facenumber_in_poster', 'plot_keywords',  
              'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',  
              'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',  
              'imdb_score', 'aspect_ratio', 'movie_facebook_likes', 'has_seen'],  
              dtype='object')
```

```
In [32]: movie['has_seen']
```

```
Out[32]: 0      0  
         1      0  
         2      0  
         3      0  
         4      0  
         ..  
        4911    0  
        4912    0
```



```
4913    0
4914    0
4915    0
Name: has_seen, Length: 4916, dtype: int64
```

## (2) Deleting a Column

```
In [33]: movie.columns
```

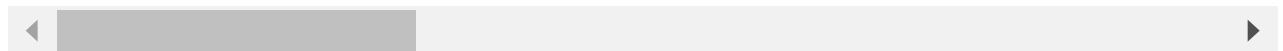
```
Out[33]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
               'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
               'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
               'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
               'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
               'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
               'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
               'imdb_score', 'aspect_ratio', 'movie_facebook_likes', 'has_seen'],
              dtype='object')
```

```
In [34]: movie.drop('has_seen', axis = 'columns')
```

```
Out[34]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300
4	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns



```
In [35]: movie.columns
```

```
Out[35]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
'imdb_score', 'aspect_ratio', 'movie_facebook_likes', 'has_seen'],
dtype='object')
```

```
In [36]: movie = movie.drop('has_seen', axis = 'columns')
```

```
In [37]: movie.columns
```

```
Out[37]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],
dtype='object')
```

### (3) Renaming Column Labels

```
In [38]: movie = pd.read_csv('movie.csv')
```

```
In [39]: movie.columns
```

```
Out[39]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],
dtype='object')
```

```
In [40]: columns_renamed = {'director_name': 'Director Name',
'num_critic_for_reviews': 'Critical Reviews'}
```

```
In [41]: movie.rename(columns = columns_renamed)
```

```
Out[41]:
```

	color	Director Name	Critical Reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom

	color	Director Name	Critical Reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker
...	...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	318.0	Daphne Zuniga
4912	Color	NaN	43.0	43.0	NaN	319.0	Valorie Curry
4913	Color	Benjamin Roberds	13.0	76.0	0.0	0.0	Maxwell Moody
4914	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Danie Henney
4915	Color	Jon Gunn	43.0	90.0	16.0	16.0	Briar Herzlinge

4916 rows × 28 columns

In [42]:

```
movie
```

Out[42]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300
4	NaN	Doug Walker	NaN	NaN	131.0	NaN
...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name
-------	---------------	------------------------	----------	-------------------------	------------------------	--------------

4916 rows × 28 columns

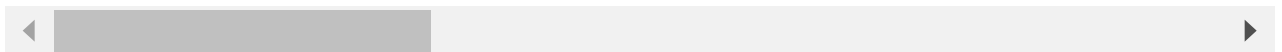
```
In [43]: movie_new = movie.rename(columns = columns_renamed)
```

```
In [44]: movie_new
```

```
Out[44]:
```

	color	Director Name	Critical Reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker
...	...	...	...	...	...	...	...
4911	Color	Scott Smith	1.0	87.0	2.0	318.0	Daphne Zuniga
4912	Color	NaN	43.0	43.0	NaN	319.0	Valorie Curry
4913	Color	Benjamin Roberds	13.0	76.0	0.0	0.0	Maxwell Moody
4914	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Danie Henney
4915	Color	Jon Gunn	43.0	90.0	16.0	16.0	Briar Herzlinge

4916 rows × 28 columns



## 2) Rows

### (1) Adding a Row

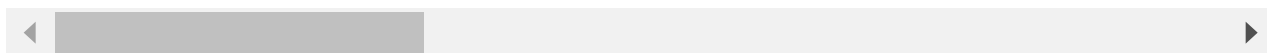
```
In [45]: movie = pd.read_csv('movie.csv')
```

```
In [46]: movie.tail()
```

```
Out[46]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
4911	Color	Scott Smith	1.0	87.0	2.0	31
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns

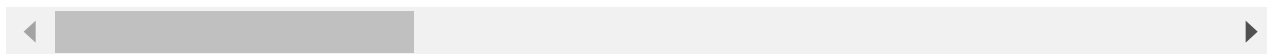


```
In [47]: movie.append({'color': 'Color', 'director_name': 'Blake'}, ignore_index = True)
```

```
Out[47]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
0	Color	James Cameron	723.0	178.0	0.0	85
1	Color	Gore Verbinski	302.0	169.0	563.0	100
2	Color	Sam Mendes	602.0	148.0	0.0	16
3	Color	Christopher Nolan	813.0	164.0	22000.0	2300
4	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	
4912	Color	NaN	43.0	43.0	NaN	31
4913	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Color	Daniel Hsia	14.0	100.0	0.0	48
4915	Color	Jon Gunn	43.0	90.0	16.0	1
4916	Color	Blake	NaN	NaN	NaN	N

4917 rows × 28 columns

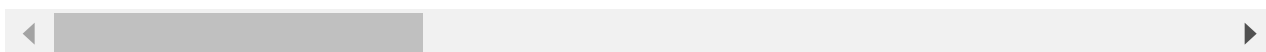


```
In [48]: movie.tail()
```

Out[48]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns



In [49]:

```
movie_new = movie.append({'color': 'Color', 'director_name': 'Blake'}, ignore_index = T
```

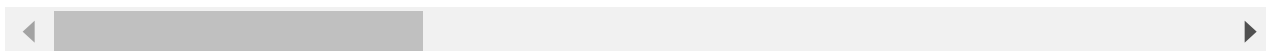
In [50]:

```
movie_new.tail()
```

Out[50]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1
<b>4916</b>	Color	Blake	NaN	NaN	NaN	N

5 rows × 28 columns



## (2) Deleting a Row

In [51]:

```
movie_new.tail()
```

Out[51]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1
<b>4916</b>	Color	Blake	NaN	NaN	NaN	N

5 rows × 28 columns

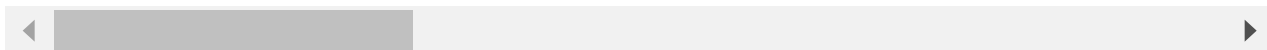
In [52]:

```
movie_new.drop(4916)
```

Out[52]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>0</b>	Color	James Cameron	723.0	178.0	0.0	85
<b>1</b>	Color	Gore Verbinski	302.0	169.0	563.0	100
<b>2</b>	Color	Sam Mendes	602.0	148.0	0.0	16
<b>3</b>	Color	Christopher Nolan	813.0	164.0	22000.0	2300
<b>4</b>	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns



In [53]:

```
movie_new.tail()
```

Out[53]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1
<b>4916</b>	Color	Blake	NaN	NaN	NaN	N

5 rows × 28 columns

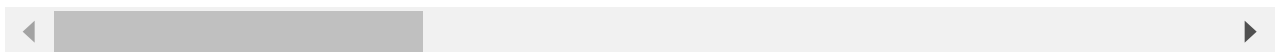
```
In [54]: movie_new = movie_new.drop(4916)
```

```
In [55]: movie_new.tail()
```

```
Out[55]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns



### (3) Updating a Single Value In a Row

```
In [56]: movie = pd.read_csv('movie.csv')
```

```
In [57]: movie.tail()
```

```
Out[57]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1



	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
--	-------	---------------	------------------------	----------	-------------------------	----------------------

5 rows × 28 columns

```
In [58]: movie['director_name']
```

```
Out[58]: 0      James Cameron
1      Gore Verbinski
2      Sam Mendes
3      Christopher Nolan
4      Doug Walker
...
4911    Scott Smith
4912         NaN
4913    Benjamin Roberds
4914    Daniel Hsia
4915      Jon Gunn
Name: director_name, Length: 4916, dtype: object
```

```
In [59]: movie['director_name'][4912]
```

```
Out[59]: nan
```

```
In [60]: movie['director_name'][4912] = 'Blake'
```

<ipython-input-60-69abe7859ac3>:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

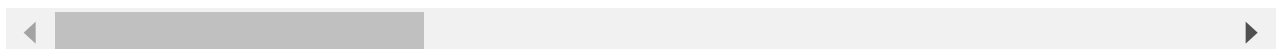
See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
movie['director\_name'][4912] = 'Blake'

```
In [61]: movie.tail()
```

```
Out[61]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	Blake	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns



```
In [62]: movie.at[4912, 'director_name']
```

```
Out[62]: 'Blake'
```

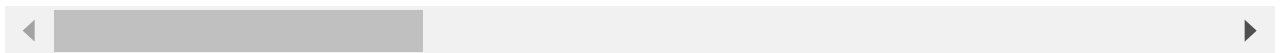
```
In [65]: movie.at[4912, 'director_name'] = 'Blake Pappas'
```

```
In [66]: movie.tail()
```

```
Out[66]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	Blake Pappas	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

5 rows × 28 columns



## 3) Index

### (1) Making a Meaningful Index

```
In [67]: movie = pd.read_csv('movie.csv')
```

```
In [68]: movie
```

```
Out[68]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_lil
<b>0</b>	Color	James Cameron	723.0	178.0	0.0	85
<b>1</b>	Color	Gore Verbinski	302.0	169.0	563.0	100
<b>2</b>	Color	Sam Mendes	602.0	148.0	0.0	16
<b>3</b>	Color	Christopher Nolan	813.0	164.0	22000.0	2300
<b>4</b>	NaN	Doug Walker	NaN	NaN	131.0	N
...	...	...	...	...	...	...

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
<b>4911</b>	Color	Scott Smith	1.0	87.0	2.0	31
<b>4912</b>	Color	NaN	43.0	43.0	NaN	31
<b>4913</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>4914</b>	Color	Daniel Hsia	14.0	100.0	0.0	48
<b>4915</b>	Color	Jon Gunn	43.0	90.0	16.0	1

4916 rows × 28 columns

```
In [69]: movie2 = movie.set_index('movie_title')
movie2
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>Avatar</b>	Color	James Cameron	723.0	178.0	0.0	
<b>Pirates of the Caribbean: At World's End</b>	Color	Gore Verbinski	302.0	169.0	563.0	
<b>Spectre</b>	Color	Sam Mendes	602.0	148.0	0.0	
<b>The Dark Knight Rises</b>	Color	Christopher Nolan	813.0	164.0	22000.0	
<b>Star Wars: Episode VII - The Force Awakens</b>	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	...
<b>Signed Sealed Delivered</b>	Color	Scott Smith	1.0	87.0	2.0	
<b>The Following</b>	Color	NaN	43.0	43.0	NaN	
<b>A Plague So Pleasant</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>Shanghai</b>	Color	Daniel Hsia	14.0	100.0	0.0	

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
Calling						
My Date with Drew	Color	Jon Gunn	43.0	90.0		16.0

4916 rows × 27 columns

```
In [70]: movie.columns
```

```
Out[70]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
               'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
               'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
               'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
               'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
               'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
               'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
               'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],
              dtype='object')
```

```
In [71]: len(movie.columns)
```

```
Out[71]: 28
```

```
In [72]: movie2.columns
```

```
Out[72]: Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
               'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
               'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
               'num_voted_users', 'cast_total_facebook_likes', 'actor_3_name',
               'facenumber_in_poster', 'plot_keywords', 'movie_imdb_link',
               'num_user_for_reviews', 'language', 'country', 'content_rating',
               'budget', 'title_year', 'actor_2_facebook_likes', 'imdb_score',
               'aspect_ratio', 'movie_facebook_likes'],
              dtype='object')
```

```
In [73]: len(movie2.columns)
```

```
Out[73]: 27
```

```
In [74]: movie2.reset_index()
```

```
Out[74]:
```

	movie_title	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
0	Avatar	Color	James Cameron	723.0	178.0		0.0
1	Pirates of the Caribbean: The Curse of the Black Pearl	Color	Gore Verbinski	302.0	169.0	563.0	

	movie_title	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3
	At World's End						
2	Spectre	Color	Sam Mendes	602.0	148.0	0.0	
3	The Dark Knight Rises	Color	Christopher Nolan	813.0	164.0	22000.0	
4	Star Wars: Episode VII - The Force Awakens	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	...	...
4911	Signed Sealed Delivered	Color	Scott Smith	1.0	87.0	2.0	
4912	The Following	Color	NaN	43.0	43.0	NaN	
4913	A Plague So Pleasant	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Shanghai Calling	Color	Daniel Hsia	14.0	100.0	0.0	
4915	My Date with Drew	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 28 columns

```
In [75]: movie2
```

Out[75]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
movie_title						
Avatar	Color	James Cameron	723.0	178.0	0.0	
Pirates of the Caribbean: At World's End	Color	Gore Verbinski	302.0	169.0	563.0	
Spectre	Color	Sam Mendes	602.0	148.0	0.0	
The Dark Knight Rises	Color	Christopher Nolan	813.0	164.0	22000.0	
Star Wars: Episode VII - The	NaN	Doug Walker	NaN	NaN	131.0	

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>Force Awakens</b>	...	...	...	...	...	...
<b>Signed Sealed Delivered</b>	Color	Scott Smith	1.0	87.0	2.0	
<b>The Following</b>	Color	NaN	43.0	43.0	NaN	
<b>A Plague So Pleasant</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>Shanghai Calling</b>	Color	Daniel Hsia	14.0	100.0	0.0	
<b>My Date with Drew</b>	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 7 columns

```
In [76]: movie2 = movie2.reset_index()
```

```
In [77]: movie2
```

```
Out[77]:
```

	movie_title	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
<b>0</b>	Avatar	Color	James Cameron	723.0	178.0	0.0	
<b>1</b>	Pirates of the Caribbean: At World's End	Color	Gore Verbinski	302.0	169.0	563.0	
<b>2</b>	Spectre	Color	Sam Mendes	602.0	148.0	0.0	
<b>3</b>	The Dark Knight Rises	Color	Christopher Nolan	813.0	164.0	22000.0	
<b>4</b>	Star Wars: Episode VII - The Force Awakens	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	...	
<b>4911</b>	Signed Sealed Delivered	Color	Scott Smith	1.0	87.0	2.0	

	movie_title	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3
4912	The Following	Color	NaN	43.0	43.0	NaN	
4913	A Plague So Pleasant	Color	Benjamin Roberds	13.0	76.0	0.0	
4914	Shanghai Calling	Color	Daniel Hsia	14.0	100.0	0.0	
4915	My Date with Drew	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 28 columns

## (2) Renaming Column Labels (i.e., Index)

```
In [78]: movie_new = pd.read_csv('movie.csv', index_col = 'movie_title')
```

```
In [79]: movie_new
```

```
Out[79]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
movie_title						
Avatar	Color	James Cameron	723.0	178.0	0.0	
Pirates of the Caribbean: At World's End	Color	Gore Verbinski	302.0	169.0	563.0	
Spectre	Color	Sam Mendes	602.0	148.0	0.0	
The Dark Knight Rises	Color	Christopher Nolan	813.0	164.0	22000.0	
Star Wars: Episode VII - The Force Awakens	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	
Signed Sealed Delivered	Color	Scott Smith	1.0	87.0	2.0	
The Following	Color	NaN	43.0	43.0	NaN	

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>A Plague So Pleasant</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>Shanghai Calling</b>	Color	Daniel Hsia	14.0	100.0	0.0	
<b>My Date with Drew</b>	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 7 columns

```
In [80]: indexes_renamed = {'Avatar': 'Ratava', 'Spectre': 'Ertceps'}
```

```
In [81]: movie_new.rename(index = indexes_renamed)
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>Ratava</b>	Color	James Cameron	723.0	178.0	0.0	
<b>Pirates of the Caribbean: At World's End</b>	Color	Gore Verbinski	302.0	169.0	563.0	
<b>Ertceps</b>	Color	Sam Mendes	602.0	148.0	0.0	
<b>The Dark Knight Rises</b>	Color	Christopher Nolan	813.0	164.0	22000.0	
<b>Star Wars: Episode VII - The Force Awakens</b>	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	
<b>Signed Sealed Delivered</b>	Color	Scott Smith	1.0	87.0	2.0	
<b>The Following</b>	Color	NaN	43.0	43.0	NaN	
<b>A Plague So Pleasant</b>	Color	Benjamin Roberds	13.0	76.0	0.0	



	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>Shanghai Calling</b>	Color	Daniel Hsia	14.0	100.0	0.0	
<b>My Date with Drew</b>	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 7 columns

In [82]:

```
movie_new
```

Out[82]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
<b>Avatar</b>	Color	James Cameron	723.0	178.0	0.0	
<b>Pirates of the Caribbean: At World's End</b>	Color	Gore Verbinski	302.0	169.0	563.0	
<b>Spectre</b>	Color	Sam Mendes	602.0	148.0	0.0	
<b>The Dark Knight Rises</b>	Color	Christopher Nolan	813.0	164.0	22000.0	
<b>Star Wars: Episode VII - The Force Awakens</b>	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	
<b>Signed Sealed Delivered</b>	Color	Scott Smith	1.0	87.0	2.0	
<b>The Following</b>	Color	NaN	43.0	43.0	NaN	
<b>A Plague So Pleasant</b>	Color	Benjamin Roberds	13.0	76.0	0.0	
<b>Shanghai Calling</b>	Color	Daniel Hsia	14.0	100.0	0.0	
<b>My Date with Drew</b>	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 27 columns

```
In [83]: movie_new = movie_new.rename(index = indexes_renamed)
```

```
In [84]: movie_new
```

```
Out[84]:
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
Ratava	Color	James Cameron	723.0	178.0	0.0	
Pirates of the Caribbean: At World's End	Color	Gore Verbinski	302.0	169.0	563.0	
Ertceps	Color	Sam Mendes	602.0	148.0	0.0	
The Dark Knight Rises	Color	Christopher Nolan	813.0	164.0	22000.0	
Star Wars: Episode VII - The Force Awakens	NaN	Doug Walker	NaN	NaN	131.0	
...	...	...	...	...	...	
Signed Sealed Delivered	Color	Scott Smith	1.0	87.0	2.0	
The Following	Color	NaN	43.0	43.0	NaN	
A Plague So Pleasant	Color	Benjamin Roberds	13.0	76.0	0.0	
Shanghai Calling	Color	Daniel Hsia	14.0	100.0	0.0	
My Date with Drew	Color	Jon Gunn	43.0	90.0	16.0	

4916 rows × 27 columns

