

Forward School

Program Code: J620-002-4:2020

Program Name: FRONT-END SOFTWARE DEVELOPMENT

Title : Exercise 6 - Pandas

Name: Justin Chong

IC Number: 960327-07-5097

Date : 26/6/2023

Introduction : Practicing on more exercises with Pandas Dataframes.

Conclusion : Doing a lot better on extracting data from Dataframes.

EXERCISE 6

Pandas

```
In [12]: ▶ import pandas as pd
```

Question 1

UFO (Unidentified Flying Objects) data is from <http://www.nuforc.org/webreports.html> (<http://www.nuforc.org/webreports.html>).

Given the UFO sightings data and the path for csv file, read the csv file from the URL.

```
In [13]: ▶ path = 'http://bit.ly/uforeports'
```

use the `read_csv` to read the csv and assign to a variable called 'ufo'

```
In [39]: ▶ #read csv
ufo = pd.read_csv(path)
pd.set_option('display.max_rows', None)
```

```
In [15]: ▶ #show the top 5 rows of data
ufo.head()
```

Out[15]:

| | City | Colors Reported | Shape Reported | State | Time |
|---|----------------------|-----------------|----------------|-------|-----------------|
| 0 | Ithaca | NaN | TRIANGLE | NY | 6/1/1930 22:00 |
| 1 | Willingboro | NaN | OTHER | NJ | 6/30/1930 20:00 |
| 2 | Holyoke | NaN | OVAL | CO | 2/15/1931 14:00 |
| 3 | Abilene | NaN | DISK | KS | 6/1/1931 13:00 |
| 4 | New York Worlds Fair | NaN | LIGHT | NY | 4/18/1933 19:00 |

```
In [16]: ▶ #show the last 10 rows of data
ufo.tail(10)
```

Out[16]:

| | City | Colors Reported | Shape Reported | State | Time |
|-------|----------------|-----------------|----------------|-------|------------------|
| 18231 | Pismo Beach | NaN | OVAL | CA | 12/31/2000 20:00 |
| 18232 | Lodi | NaN | NaN | WI | 12/31/2000 20:30 |
| 18233 | Anchorage | RED | VARIOUS | AK | 12/31/2000 21:00 |
| 18234 | Capitola | NaN | TRIANGLE | CA | 12/31/2000 22:00 |
| 18235 | Fountain Hills | NaN | NaN | AZ | 12/31/2000 23:00 |
| 18236 | Grant Park | NaN | TRIANGLE | IL | 12/31/2000 23:00 |
| 18237 | Spirit Lake | NaN | DISK | IA | 12/31/2000 23:00 |
| 18238 | Eagle River | NaN | NaN | WI | 12/31/2000 23:45 |
| 18239 | Eagle River | RED | LIGHT | WI | 12/31/2000 23:45 |
| 18240 | Ybor | NaN | OVAL | FL | 12/31/2000 23:59 |

```
In [17]: ▶ #check the data type
type(ufo)
```

Out[17]: pandas.core.frame.DataFrame

```
In [18]: ▶ #check the data type for 'Time' column
ufo['Time'].dtypes
```


Out[18]: dtype('O')

In [30]:  *#show all rows for the column 'City'*

```
temp_ufo = ufo.copy()
temp_ufo.columns = [col.lower() for col in temp_ufo.columns.str.replace(' ', '_')]

cities = temp_ufo['city']
len(cities.unique())
```

Out[30]: 6476

In [31]:  *#determine the shape of the data*
ufo.shape

Out[31]: (18241, 5)

```
In [65]: #show all data for 'City' that starts with 'E'  
temp_ufo[temp_ufo['city'].str.startswith('E', na = False)]['city']
```

```
Out[65]: 8      Eklutna
          55      Espanola
          109     Excelsior
          140     East Palestine
          179     Evergreen
          200     Evergreen
          270     El Cerrito
          276     Estes Park
          280     Ellington AFB
          340     Eagle Pass
          355     Elmore
          359     Edroy
          377     East Hampton
          387     Edna
          400     El Paso
          401     E. Los Angeles
          444     Exeter
          452     El Paso
          461     England
          529     El Sobrante
          530     Evanston
          609     East Bernstadt
          683     Evendale
          712     Ellwood City
          738     East Windsor
          785     Emporia
          823     East Vassalboro
          824     Evanston
          852     Engelwood Cliffs
          854     El Cerrito
          933     East Hartford
          1027    Elkins
          1140    Enfield
          1141    Enfield
          1142    Enfield
          1143    Enfield
          1164    Elizabeth
          1166    El Paso
          1186    East Liverpool
          1380    El Paso
          1385    Enid
          1394    El Paso
          1400    Edgemoor
          1404    El Cerrito
          1421    El Paso
          1429    Edison
          1458    East Hanover
          1496    Edgartown
          1706    Eugene
          1717    East Palestine
          1722    Elwood
          1732    Edison
          1733    Edison
          1734    Edison
          1754    Enid
          1790    Estill
          1856    Enid
```

| | |
|------|------------------|
| 1881 | El Cajon |
| 1882 | East Barre |
| 1887 | Enfield |
| 1902 | Exeter |
| 1904 | Erie |
| 1941 | East Aurora |
| 1942 | Evansville |
| 1968 | Eureka |
| 2005 | Englewood |
| 2016 | Everton |
| 2064 | Everglades |
| 2271 | East Brunswick |
| 2280 | El Paso |
| 2294 | Ellensburg |
| 2322 | Eatonton |
| 2336 | Eddington |
| 2385 | Eddington |
| 2419 | East Liverpool |
| 2426 | Erie |
| 2491 | Elyria |
| 2521 | El Centro |
| 2531 | Englewood |
| 2603 | Esopus |
| 2615 | East Liverpool |
| 2645 | Erie |
| 2693 | Evanston |
| 2713 | East Berkshire |
| 2728 | East Haven |
| 2797 | Eastland |
| 2940 | East Grand Forks |
| 2966 | El Monte |
| 2998 | Estes Park |
| 3098 | Enterprise |
| 3125 | Edgewater |
| 3167 | Eastland |
| 3177 | Enfield |
| 3220 | Everett |
| 3272 | Elmira Heights |
| 3388 | Emporia |
| 3391 | Eveleth |
| 3425 | El Paso |
| 3469 | Elmont |
| 3546 | Edgemoor |
| 3547 | Edgemoore |
| 3574 | Edgemoore |
| 3625 | Everton |
| 3629 | Everett |
| 3654 | East Haven |
| 3695 | Escawtawpa |
| 3696 | Eastport |
| 3748 | Elsberry |
| 3797 | Elmsford |
| 3807 | Edina |
| 3871 | Eureka |
| 3954 | Elmont |
| 3993 | Edison |
| 4040 | East Providence |

| | |
|------|--------------------|
| 4118 | East Orange |
| 4191 | East New York |
| 4214 | Ellwood City |
| 4221 | Emerson |
| 4245 | Exeter |
| 4287 | Eureka |
| 4304 | Evansville |
| 4306 | Ethell |
| 4340 | El Paso |
| 4388 | Exeter |
| 4389 | Exeter |
| 4413 | Edwards |
| 4432 | Elko |
| 4485 | El Cajon |
| 4486 | El Reno |
| 4557 | Elizaville |
| 4575 | Ellwood city |
| 4576 | Edwards AFB |
| 4594 | Estes Park |
| 4629 | Edina |
| 4662 | Edwards AFB |
| 4676 | Edinboro |
| 4690 | Excelsior Springs |
| 4730 | Elgin |
| 4804 | Eidson |
| 4829 | East Texas River |
| 4938 | Edison |
| 4975 | Elizabeth Township |
| 5003 | Eugene |
| 5017 | Elizabethtown |
| 5041 | Elmendorf AFB |
| 5063 | Ely |
| 5075 | El Paso |
| 5085 | Eatonton |
| 5113 | Eureka |
| 5136 | Elwood |
| 5160 | Edgartown |
| 5169 | Emporium |
| 5242 | El Paso |
| 5267 | Eustis |
| 5284 | Eatonton |
| 5317 | Eau Claire |
| 5371 | Edna |
| 5372 | Eatonton |
| 5428 | Euclid |
| 5479 | East Sparta |
| 5508 | Eldon |
| 5542 | Eglin AFB |
| 5570 | El Centro |
| 5573 | Elkader |
| 5575 | East Meadow |
| 5584 | Evansville |
| 5590 | East Helena |
| 5607 | El Paso |
| 5615 | Evansville |
| 5667 | Erie |
| 5749 | Eastchester |

| | |
|------|-----------------------|
| 5770 | El Paso |
| 5776 | Exeter |
| 5780 | Ellington |
| 5820 | Edgewood |
| 5846 | Eden Prairie |
| 5998 | El Dorado Hills |
| 6029 | Eastern border |
| 6085 | Enloe |
| 6111 | Ellenville |
| 6148 | Elk Park |
| 6175 | Englewood |
| 6218 | El Jebel |
| 6296 | East Windsor |
| 6308 | Eugene |
| 6324 | Elfin Forest |
| 6344 | El Paso |
| 6415 | El Dorado Hills |
| 6450 | Endwell |
| 6533 | Escondido |
| 6595 | El Cajon |
| 6604 | Elk River |
| 6676 | Easley |
| 6708 | Etowah |
| 6726 | East Winn |
| 6747 | Ely |
| 6769 | Elmer |
| 6845 | Enid |
| 6853 | East Corinth |
| 6873 | Elk Grove |
| 6905 | Echo Summit |
| 6915 | Eastern United States |
| 6952 | Exeter |
| 6955 | Elkland |
| 7043 | Everglades |
| 7065 | Economy Borough |
| 7111 | Efland |
| 7138 | Elkton |
| 7221 | Eden |
| 7226 | Evansville |
| 7229 | Encinal |
| 7232 | Elkton |
| 7251 | El Segundo |
| 7312 | Elkton |
| 7336 | Elkhart |
| 7337 | El Centro |
| 7354 | Eagle Butte |
| 7355 | Elkton |
| 7392 | Estill Springs |
| 7422 | Easthampton |
| 7425 | Erie |
| 7428 | Emporia |
| 7434 | Evansville |
| 7480 | Everett |
| 7487 | East Bernstadt |
| 7495 | Earlham |
| 7524 | Ewa Beach |
| 7537 | Elizabeth |

| | |
|------|-----------------|
| 7635 | El Toro |
| 7659 | El Granada |
| 7679 | Evart |
| 7688 | Edison |
| 7690 | Enterprise |
| 7739 | Elbridge |
| 7746 | Eaton Rapids |
| 7751 | Eastport |
| 7788 | Elizabethtown |
| 7790 | El Cerrito |
| 7845 | El Centro |
| 7857 | Electric City |
| 7859 | Eastgate |
| 7930 | Evington |
| 7942 | Everett |
| 7970 | Elma |
| 7975 | Eugene |
| 8015 | Earlville |
| 8092 | Edwards AFB |
| 8143 | East Aurora |
| 8159 | Eagle Creek |
| 8219 | El Paso |
| 8300 | Endicott |
| 8375 | East Tacoma |
| 8380 | Edmonds |
| 8497 | Enola |
| 8578 | Edinboro |
| 8609 | Elizabethtown |
| 8624 | Eagle Butte |
| 8643 | Effingham |
| 8790 | Emporia |
| 8791 | Evansville |
| 8826 | Ellisville |
| 8865 | Elizabeth |
| 8867 | Embarrass |
| 8868 | Eilmsport |
| 8871 | Eugene |
| 8888 | Elk Grove |
| 8923 | Elizabethtown |
| 8979 | Elkridge |
| 9005 | Elizabethtown |
| 9012 | E. Greenville |
| 9076 | Eureka |
| 9099 | Edwards AFB |
| 9102 | Edwards AFB |
| 9149 | Evansville |
| 9249 | Edmonds |
| 9299 | Everett |
| 9359 | El Paso |
| 9360 | Eugene |
| 9378 | Ely |
| 9382 | Eddington |
| 9398 | Edmonds |
| 9418 | Eau Claire |
| 9451 | East Providence |
| 9452 | East Providence |
| 9499 | Ellensburg |

| | |
|-------|---------------------|
| 9508 | Eau Claire |
| 9524 | East Setauket |
| 9535 | Ellsworth |
| 9557 | Eureka |
| 9606 | Englewood Cliffs |
| 9670 | Encino |
| 9800 | Elizabeth |
| 9813 | Everett |
| 9827 | Eugene |
| 9848 | Elkins |
| 9870 | Eden |
| 9886 | Ellensburg |
| 9897 | East Greenwich |
| 9902 | Elizabeth |
| 9954 | East Detroit |
| 9988 | East Mesa |
| 10142 | Edison |
| 10236 | Edinboro |
| 10291 | Elma |
| 10294 | Egg Harbor Township |
| 10298 | Eden Prairie |
| 10312 | El Centro |
| 10321 | El Paso |
| 10328 | Exeter |
| 10378 | Edmond |
| 10382 | Eden Prairie |
| 10393 | Eagle Butte |
| 10414 | Englewood Cliffs |
| 10428 | Edgemere |
| 10435 | Eunice |
| 10485 | Evansville |
| 10486 | Earth City |
| 10500 | Elmwood |
| 10533 | East Alton |
| 10570 | El Paso |
| 10595 | Eden Prairie |
| 10611 | Edmonds |
| 10679 | East Rochester |
| 10726 | Eustace |
| 10735 | Elk River |
| 10803 | Egan |
| 10880 | Evansville |
| 10909 | El Paso |
| 10911 | East Troy |
| 10937 | Enumclaw |
| 10958 | Edmonds |
| 10985 | Eureka |
| 11014 | El Dorado |
| 11039 | East Wenatchee |
| 11075 | East Woodstock |
| 11091 | Elizabethtown |
| 11166 | East Los Angeles |
| 11176 | Emporia |
| 11267 | Englewood |
| 11293 | Eagle County |
| 11380 | Eau Claire |
| 11390 | Edwards AFB |

| | |
|-------|----------------------|
| 11395 | El Cajon |
| 11459 | Epcot Center |
| 11470 | Eagan |
| 11567 | Elmira |
| 11584 | Ellensburg |
| 11622 | Elba |
| 11676 | Edmond |
| 11699 | Eustace |
| 11707 | Evansville |
| 11708 | Ellsworth |
| 11725 | East Amwell |
| 11726 | East Amwell Township |
| 11730 | Ellicott City |
| 11781 | Elkhart |
| 11890 | Erie |
| 11912 | Excelsior Springs |
| 11929 | El Paso |
| 11930 | El Paso |
| 11933 | East Machias |
| 11960 | Ettrick |
| 11974 | Erie |
| 11985 | Everett |
| 11988 | Everett |
| 12000 | El Dorado |
| 12032 | East Hartford |
| 12088 | Emory |
| 12119 | Edmond |
| 12133 | Escondido |
| 12170 | Eugene |
| 12190 | Everett |
| 12194 | Enumclaw |
| 12204 | Everett |
| 12258 | Everett |
| 12276 | Elgin |
| 12286 | El Monte |
| 12326 | Elgin |
| 12338 | Elk Grove |
| 12361 | Enterprise |
| 12549 | Edison |
| 12556 | Ellensburg |
| 12704 | Eureka |
| 12756 | Easton |
| 12775 | Escondido |
| 12799 | East Bethel |
| 12814 | Eagle River |
| 12858 | Estes Park |
| 12896 | Eureka |
| 12899 | Edwardsville |
| 13045 | Elk Grove |
| 13058 | Everett |
| 13082 | Edwards AFB |
| 13084 | Ely |
| 13085 | Elmira Heights |
| 13109 | East Rutherford |
| 13119 | Eagle Mountain |
| 13238 | Erie |
| 13245 | Enola |

| | |
|-------|---|
| 13264 | Encinitas: along I-5, from encinitas to oceans... |
| 13295 | E. Falmouth |
| 13302 | Edmond |
| 13441 | East Derry |
| 13451 | Everett |
| 13454 | East Liverpool |
| 13566 | El Paso |
| 13590 | East Stroudsburg |
| 13650 | Elkhart |
| 13658 | Everett |
| 13730 | Evansville |
| 13763 | El Cajon |
| 13865 | East Hoquiam |
| 13926 | Elko |
| 13978 | Escondido |
| 13979 | East Springfield |
| 14046 | Edgewood |
| 14091 | Eugene |
| 14121 | Eureka |
| 14151 | Ettersburg |
| 14164 | Elizabethtown |
| 14193 | East Lynne |
| 14204 | Eleanor |
| 14224 | Eureka |
| 14225 | East Northport |
| 14286 | Easton |
| 14311 | El Dorado Hills |
| 14339 | Elko |
| 14340 | Elko |
| 14384 | El Dorado Hills |
| 14385 | Eugene |
| 14408 | Everett |
| 14492 | Elk Grove |
| 14586 | Egg Harbor |
| 14600 | Estacada |
| 14606 | Elgin |
| 14628 | East Machias |
| 14657 | Everett |
| 14670 | Eules |
| 14673 | Everett |
| 14675 | Ewen |
| 14687 | Encino |
| 14711 | Elyria |
| 14717 | Endicott |
| 14766 | Elkton |
| 14789 | Elgin |
| 14805 | Eureka Springs |
| 14828 | El Paso |
| 14851 | Ellis |
| 14852 | Ellis |
| 14882 | Everson |
| 14921 | Erie |
| 15040 | Emmetsburg |
| 15062 | El Dorado |
| 15074 | El Centro |
| 15091 | El Dorado Hills |
| 15170 | Elgin |

| | |
|-------|------------------|
| 15179 | Elberfeld |
| 15196 | East Lansing |
| 15230 | Englewood Cliffs |
| 15304 | Erie |
| 15328 | Edison |
| 15408 | Ephrata |
| 15461 | Emory |
| 15585 | Elisabeth |
| 15620 | Evansville |
| 15621 | Erath |
| 15645 | El Centro |
| 15659 | Edwardsville |
| 15662 | Exton |
| 15672 | Edgewood |
| 15695 | Erie |
| 15746 | Ely |
| 15778 | Edinburg |
| 15793 | East Rutherford |
| 15803 | Escondido |
| 15838 | Eules |
| 16032 | Eureka Springs |
| 16042 | Elmira |
| 16048 | Ephrata |
| 16070 | Encino |
| 16075 | Escondido |
| 16112 | Everett |
| 16124 | Elko |
| 16130 | East Springfield |
| 16156 | Elk |
| 16185 | Evansville |
| 16187 | Elizabethtown |
| 16236 | Eugene |
| 16241 | Eugene |
| 16246 | East Killingly |
| 16307 | Eugene |
| 16315 | Eagle River |
| 16347 | Eagle Point |
| 16394 | Edinburg |
| 16395 | Evansville |
| 16421 | Eastchester |
| 16525 | Escondido |
| 16547 | El Cajon |
| 16553 | Ely |
| 16556 | Elkhart |
| 16557 | Ely |
| 16603 | Eugene |
| 16721 | Ewa Beach |
| 16771 | East Windsor |
| 16815 | Easton |
| 16871 | Evanston |
| 16936 | Enochville |
| 16965 | Elk City |
| 16966 | Elk City |
| 16970 | El Paso |
| 16999 | Everett |
| 17075 | Erie |
| 17149 | Elkton |

| | |
|-------|------------------|
| 17169 | El Cajon |
| 17199 | East Petersburg |
| 17269 | Eatontown |
| 17304 | Eugene |
| 17308 | Everett |
| 17338 | Erie |
| 17378 | Edmond |
| 17382 | Eldorado |
| 17399 | Everett |
| 17422 | Eugene |
| 17436 | East Springfield |
| 17480 | El Dorado Hills |
| 17514 | Everett |
| 17530 | Edmonds |
| 17652 | Elverta |
| 17678 | Elk City |
| 17679 | Eugene |
| 17684 | Escondido |
| 17716 | Eureka |
| 17732 | Easley |
| 17754 | Elmira |
| 17807 | Edgewood |
| 17838 | Elyria |
| 17900 | Eugene |
| 17933 | Ellinwood |
| 17936 | Ellensburg |
| 17938 | Ellensburg |
| 17943 | East Lansing |
| 17955 | Eagletown |
| 17957 | Everett |
| 17976 | East Brunswick |
| 18008 | El Paso |
| 18023 | Elizabeth |
| 18028 | Ellisville |
| 18037 | Eugene |
| 18042 | Edina |
| 18142 | Estes Park |
| 18151 | Enterprise |
| 18161 | Elizabeth City |
| 18182 | Evansville |
| 18215 | El Campo |
| 18224 | Eufaula |
| 18238 | Eagle River |
| 18239 | Eagle River |

Name: city, dtype: object

```
In [41]: # count number of reported cases for 'CIRCLE'
# and count the number of reported cases for 'LIGHT'

# print(temp_ufo)
circle_cities = temp_ufo['shape_reported'] == 'CIRCLE'
light_cities = temp_ufo['shape_reported'] == 'LIGHT'

# number of reported cases for 'CIRCLE'
print(len(temp_ufo[circle_cities]))

# number of reported cases for 'LIGHT'
print(len(temp_ufo[light_cities]))
```

1365
2803

```
In [42]: #determine what are the top three colors reported
temp_ufo['colors_reported'].value_counts().head(3)
```

```
Out[42]: colors_reported
RED      780
GREEN    531
ORANGE   528
Name: count, dtype: int64
```

```
In [45]: #count the number of shape reported by state and city

ufo.groupby(['State', 'City'])['Shape Reported'].size()
```

```
Out[45]: State  City
AK      Adak
1
      Alaska
2
      Anchorage
12
      Arctic
1
      Auke Bay
2
      Bering Sea
1
      Bethel
1
      Big Lake
1
      Cantwell
1
```

Question 2

IMDB ratings are from http://www.imdb.com/search/title?groups=top_1000&sort=user_rating&view=simple (http://www.imdb.com/search/title?groups=top_1000&sort=user_rating&view=simple)

Given the IMDB movies dataset and path, use the read_csv to read the data and assign to a variable 'movies'

```
In [46]: path = 'http://bit.ly/imdbratings'
```

```
In [47]: #read the dataset
movies = pd.read_csv(path)
# pd.set_option('display.max_rows', None)
```

```
In [48]: #show the top 10 rows
movies.head(10)
```

Out[48]:

| | 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....] |
| 5 | 8.9 | 12 Angry Men | NOT RATED | Drama | 96 | [u'Henry Fonda', u'Lee J. Cobb', u'Martin Bals...] |
| 6 | 8.9 | The Good, the Bad and the Ugly | NOT RATED | Western | 161 | [u'Clint Eastwood', u'Eli Wallach', u'Lee Van ...] |
| 7 | 8.9 | The Lord of the Rings: The Return of the King | PG-13 | Adventure | 201 | [u'Elijah Wood', u'Viggo Mortensen', u'Ian McK...] |
| 8 | 8.9 | Schindler's List | R | Biography | 195 | [u'Liam Neeson', u'Ralph Fiennes', u'Ben Kings...] |
| 9 | 8.9 | Fight Club | R | Drama | 139 | [u'Brad Pitt', u'Edward Norton', u'Helena Bonh...] |


```
In [49]: #show summary of the dataset  
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 979 entries, 0 to 978  
Data columns (total 6 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   star_rating     979 non-null   float64  
1   title           979 non-null   object  
2   content_rating  976 non-null   object  
3   genre           979 non-null   object  
4   duration        979 non-null   int64  
5   actors_list     979 non-null   object  
dtypes: float64(1), int64(1), object(4)  
memory usage: 46.0+ KB
```

```
In [50]: #determine the dimension of the dataset  
movies.shape
```

```
Out[50]: (979, 6)
```

```
In [51]: #show the data types for each coloumns  
movies.dtypes
```

```
Out[51]: star_rating    float64  
title                object  
content_rating       object  
genre                object  
duration             int64  
actors_list          object  
dtype: object
```

```
In [52]: #show all the column names  
movies.columns
```

```
Out[52]: Index(['star_rating', 'title', 'content_rating', 'genre', 'duration',  
               'actors_list'],  
              dtype='object')
```

In [54]: `# rename the following columns, 'star_rating' as 'stars_rating'`
`# 'content_rating' as 'content'`

```
temp_movies = movies.copy()
```

```
# print(temp_movies.columns)
```

```
for i in movies.columns:
    if i == "star_rating":
        temp_movies.rename(columns={
            i: "stars_rating"
        }, inplace=True)
```

```
    if i == "content_rating":
        temp_movies.rename(columns={
            i: "content"
        }, inplace=True)
```

```
print(temp_movies.columns)
```

```
Index(['stars_rating', 'title', 'content', 'genre', 'duration', 'actors_list'], dtype='object')
```

In [89]: `#given a list called 'col_names' change all the columns names to 'col_names'`
`# SKIP`

Out[89]:

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

In [55]: `#provide the statistical summary for 'star_rating' and 'duration'`
`movies[['star_rating', 'duration']].describe()`

Out[55]:

| | 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 [56]: `#show the data type for column 'star_rating'`
`movies['star_rating'].dtypes`

Out[56]: `dtype('float64')`

```
In [57]: ▶ #show the 5th row data for column 'content_rating'
movies['content_rating'][4]
```

Out[57]: 'R'

```
In [58]: ▶ #show all rows for movies duration more than 200 mins
movies[movies['duration'] > 200]
```

Out[58]:

| | star_rating | | title | content_rating | genre | duration | actors_list |
|----|-------------|--|---|----------------|-----------|----------|---|
| 7 | 8.9 | | The Lord of the Rings: The Return of the King | PG-13 | Adventure | 201 | [u'Elijah Wood', u'Viggo Mortensen', u'Ian McK... |
| 17 | 8.7 | | Seven Samurai | UNRATED | Drama | 207 | [u'Toshir\u00f4 Mifune', u'Takashi Shimura', u'K... |
| 78 | 8.4 | | Once Upon a Time in America | R | Crime | 229 | [u'Robert De Niro', u'James Woods', u'Elizabet... |
| 85 | 8.4 | | Lawrence of Arabia | PG | Adventure | 216 | [u"Peter O'Toole", u'Alec Guinness', u'Anthony... |
| | | | Lagaan: Once | | | | [u'Aamir Khan', |

```
In [59]: ▶ #what is the average movie duration
movies['duration'].mean()
```

Out[59]: 120.97957099080695

```
In [60]: ▶ #count the number of movies where actor 'Charlton Heston' acted in
len(movies[movies['actors_list'].str.contains("Charlton Heston")])
```

Out[60]: 4

```
In [61]: ▶ #what are the top 5 genre movies
movies['genre'].value_counts().head()
```

```
Out[61]: genre
Drama      278
Comedy     156
Action     136
Crime      124
Biography   77
Name: count, dtype: int64
```

```
In [62]: ▶ #which movie has the highest rating  
# movies[movies['star_rating'] == movies['star_rating'].max()][0]['title']  
  
ranked_best = movies.sort_values(by='star_rating', ascending=False)  
print(ranked_best['title'].head(1))
```

```
0    The Shawshank Redemption  
Name: title, dtype: object
```

```
In [63]: #which movie has the lowest rating  
movies[movies['star_rating'] == movies['star_rating'].min()]
```


Out[63]:

| | star_rating | title | content_rating | genre | duration | actors_list |
|-----|-------------|---|----------------|-----------|----------|---|
| 930 | 7.4 | Man on the Moon | R | Biography | 118 | [u'Jim Carrey', u'Danny DeVito', u'Gerry Becker'] |
| 931 | 7.4 | Mean Streets | R | Crime | 112 | [u'Robert De Niro', u'Harvey Keitel', u'David ... |
| 932 | 7.4 | Harry Potter and the Order of the Phoenix | PG-13 | Adventure | 138 | [u'Daniel Radcliffe', u'Emma Watson', u'Rupert... |
| 933 | 7.4 | Beetlejuice | PG | Comedy | 92 | [u'Alec Baldwin', u'Geena Davis', u'Michael Ke... |
| 934 | 7.4 | Crazy, Stupid, Love. | PG-13 | Comedy | 118 | [u'Steve Carell', u'Ryan Gosling', u'Julianne ... |
| 935 | 7.4 | Excalibur | R | Adventure | 140 | [u'Nigel Terry', u'Helen Mirren', u'Nicholas C... |
| 936 | 7.4 | True Grit | NaN | Adventure | 128 | [u'John Wayne', u'Kim Darby', u'Glen Campbell'] |
| 937 | 7.4 | Labyrinth | PG | Adventure | 101 | [u'David Bowie', u'Jennifer Connelly', u'Toby ... |
| 938 | 7.4 | Alice in Wonderland | G | Animation | 75 | [u'Kathryn Beaumont', u'Ed Wynn', u'Richard Ha... |
| 939 | 7.4 | Predestination | R | Drama | 97 | [u'Ethan Hawke', u'Sarah Snook', u'Noah Taylor'] |
| 940 | 7.4 | Much Ado About Nothing | PG-13 | Comedy | 111 | [u'Kenneth Branagh', u'Emma Thompson', u'Keanu... |
| 941 | 7.4 | A Bridge Too Far | PG | Drama | 175 | [u'Sean Connery', u'Ryan O'Neal', u'Michael Ca... |
| 942 | 7.4 | Raising Arizona | PG-13 | Comedy | 94 | [u'Nicolas Cage', u'Holly Hunter', u'Trey Wils... |
| 943 | 7.4 | The Bucket List | PG-13 | Adventure | 97 | [u'Jack Nicholson', u'Morgan Freeman', u'Sean ... |
| 944 | 7.4 | Terms of Endearment | R | Comedy | 132 | [u'Shirley MacLaine', u'Debra Winger', u'Jack ... |
| 945 | 7.4 | Take Shelter | R | Drama | 120 | [u'Michael Shannon', u'Jessica Chastain', u'Sh... |

| | star_rating | title | content_rating | genre | duration | actors_list |
|-----|-------------|-----------------------------|----------------|-----------|----------|---|
| 946 | 7.4 | Far from Heaven | PG-13 | Drama | 107 | [u'Julianne Moore', u'Dennis Quaid', u'Dennis ... |
| 947 | 7.4 | Eraserhead | UNRATED | Drama | 89 | [u'Jack Nance', u'Charlotte Stewart', u'Allen ... |
| 948 | 7.4 | Frances Ha | R | Comedy | 86 | [u'Greta Gerwig', u'Mickey Sumner', u'Adam Dri... |
| 949 | 7.4 | Home Alone | PG | Comedy | 103 | [u'Macaulay Culkin', u'Joe Pesci', u'Daniel St... |
| 950 | 7.4 | Bound | R | Crime | 108 | [u'Jennifer Tilly', u'Gina Gershon', u'Joe Pan... |
| 951 | 7.4 | Sleepy Hollow | R | Drama | 105 | [u'Johnny Depp', u'Christina Ricci', u'Miranda... |
| 952 | 7.4 | Pirate Radio | R | Comedy | 117 | [u'Philip Seymour Hoffman', u'Bill Nighy', u'N... |
| 953 | 7.4 | The NeverEnding Story | PG | Adventure | 102 | [u'Noah Hathaway', u'Barret Oliver', u'Tami St... |
| 954 | 7.4 | X-Men | PG-13 | Action | 104 | [u'Patrick Stewart', u'Hugh Jackman', u'Ian Mc... |
| 955 | 7.4 | Zero Dark Thirty | R | Drama | 157 | [u'Jessica Chastain', u'Joel Edgerton', u'Chri... |
| 956 | 7.4 | Manhattan Murder Mystery | PG | Comedy | 104 | [u'Woody Allen', u'Diane Keaton', u'Jerry Adler'] |
| 957 | 7.4 | National Lampoon's Vacation | R | Comedy | 98 | [u'Chevy Chase', u'Beverly D'Angelo", u'Imogen... |
| 958 | 7.4 | My Sister's Keeper | PG-13 | Drama | 109 | [u'Cameron Diaz', u'Abigail Breslin', u'Alec B... |
| 959 | 7.4 | Deconstructing Harry | R | Comedy | 96 | [u'Woody Allen', u'Judy Davis', u'Julia Louis-... |
| 960 | 7.4 | The Way Way Back | PG-13 | Comedy | 103 | [u'Steve Carell', u'Toni Collette', u'Allison ... |
| 961 | 7.4 | Capote | R | Biography | 114 | [u'Philip Seymour Hoffman', u'Clifton Collins ... |

| | star_rating | title | content_rating | genre | duration | actors_list |
|-----|-------------|---|----------------|-----------|----------|---|
| 962 | 7.4 | Driving Miss Daisy | PG | Comedy | 99 | [u'Morgan Freeman', u'Jessica Tandy', u'Dan Ay... |
| 963 | 7.4 | La Femme Nikita | R | Action | 118 | [u'Anne Parillaud', u'Marc Duret', u'Patrick F... |
| 964 | 7.4 | Lincoln | PG-13 | Biography | 150 | [u'Daniel Day-Lewis', u'Sally Field', u'David ... |
| 965 | 7.4 | Limitless | PG-13 | Mystery | 105 | [u'Bradley Cooper', u'Anna Friel', u'Abbie Cor... |
| 966 | 7.4 | The Simpsons Movie | PG-13 | Animation | 87 | [u'Dan Castellana', u'Julie Kavner', u'Nancy... |
| 967 | 7.4 | The Rock | R | Action | 136 | [u'Sean Connery', u'Nicolas Cage', u'Ed Harris'] |
| 968 | 7.4 | The English Patient | R | Drama | 162 | [u'Ralph Fiennes', u'Juliette Binoche', u'Will... |
| 969 | 7.4 | Law Abiding Citizen | R | Crime | 109 | [u'Gerard Butler', u'Jamie Foxx', u'Leslie Bibb'] |
| 970 | 7.4 | Wonder Boys | R | Drama | 107 | [u'Michael Douglas', u'Tobey Maguire', u'Franc... |
| 971 | 7.4 | Death at a Funeral | R | Comedy | 90 | [u'Matthew Macfadyen', u'Peter Dinklage', u'Ew... |
| 972 | 7.4 | Blue Valentine | NC-17 | Drama | 112 | [u'Ryan Gosling', u'Michelle Williams', u'John... |
| 973 | 7.4 | The Cider House Rules | PG-13 | Drama | 126 | [u'Tobey Maguire', u'Charlize Theron', u'Micha... |
| 974 | 7.4 | Tootsie | PG | Comedy | 116 | [u'Dustin Hoffman', u'Jessica Lange', u'Teri G... |
| 975 | 7.4 | Back to the Future Part III | PG | Adventure | 118 | [u'Michael J. Fox', u'Christopher Lloyd', u'Ma... |
| 976 | 7.4 | Master and Commander: The Far Side of the World | PG-13 | Action | 138 | [u'Russell Crowe', u'Paul Bettany', u'Billy Bo... |
| 977 | 7.4 | Poltergeist | PG | Horror | 114 | [u'JoBeth Williams', u'Heather O'Rourke', u'Cr... |

| | star_rating | title | content_rating | genre | duration | actors_list |
|-----|-------------|-------------|----------------|-------|----------|---|
| 978 | 7.4 | Wall Street | R | Crime | 126 | [u'Charlie Sheen', u'Michael Douglas', u'Tamar... |

```
In [64]:  # group by genre and content rating and calculate the mean for duration  
movies.groupby(['genre', 'content_rating'])['duration'].mean()
```

```

Out[64]: genre      content_rating
Action      APPROVED      143.333333
            G            178.000000
            GP           144.000000
            NOT RATED    129.500000
            PASSED       98.000000
            PG           119.727273
            PG-13        130.204545
            R            123.850746
            UNRATED      110.666667
Adventure   APPROVED      158.333333
            G            162.000000
            NOT RATED    113.200000
            PASSED       102.000000
            PG           133.952381
            PG-13        143.913043
            R            124.882353
            UNRATED      136.000000
Animation   APPROVED      84.666667
            G            93.150000
            NOT RATED    91.000000
            PG           99.360000
            PG-13        104.200000
            R            101.000000
            UNRATED      89.000000
Biography   APPROVED      111.000000
            G            143.000000
            GP           172.000000
            NOT RATED    96.000000
            PG           126.000000
            PG-13        133.241379
            R            132.138889
Comedy       APPROVED      108.333333
            G            86.000000
            GP           91.000000
            NC-17        95.000000
            NOT RATED    129.875000
            PASSED       83.666667
            PG           100.956522
            PG-13        106.565217
            R            107.561644
            UNRATED      103.750000
            X            84.000000
Crime        APPROVED      102.833333
            NC-17        106.000000
            NOT RATED    109.428571
            PASSED       107.000000
            PG           108.166667
            PG-13        120.750000
            R            127.114943
            UNRATED      112.909091
            X            136.000000
Drama        APPROVED      119.583333
            G            170.666667
            NC-17        138.250000
            NOT RATED    128.166667
            PASSED       172.000000

```

| | | |
|-----------|-----------|------------|
| | PG | 125.360000 |
| | PG-13 | 127.181818 |
| | R | 125.790210 |
| | TV-MA | 131.000000 |
| | UNRATED | 118.777778 |
| | X | 113.000000 |
| Family | G | 100.000000 |
| | PG | 115.000000 |
| Fantasy | R | 112.000000 |
| Film-Noir | APPROVED | 88.000000 |
| | NOT RATED | 93.000000 |
| | UNRATED | 111.000000 |
| History | UNRATED | 66.000000 |
| Horror | APPROVED | 82.500000 |
| | NC-17 | 85.000000 |
| | NOT RATED | 75.000000 |
| | PG | 114.000000 |
| | PG-13 | 111.500000 |
| | R | 108.062500 |
| | UNRATED | 98.000000 |
| | X | 92.000000 |
| Mystery | APPROVED | 118.000000 |
| | G | 160.000000 |
| | NOT RATED | 130.000000 |
| | PG | 138.000000 |
| | PG-13 | 99.000000 |
| | R | 111.000000 |
| | UNRATED | 86.000000 |
| Sci-Fi | APPROVED | 92.000000 |
| | PG-13 | 91.000000 |
| | R | 120.666667 |
| Thriller | APPROVED | 108.000000 |
| | PG | 120.000000 |
| | R | 114.333333 |
| Western | APPROVED | 132.000000 |
| | NOT RATED | 151.000000 |
| | PG | 110.000000 |
| | PG-13 | 175.000000 |
| | R | 133.666667 |

Name: duration, dtype: float64