Day_36_011223

January 23, 2024

[121]:
```
import pandas as pd
import numpy as np
```

[122]:
```
movies = pd.read_csv("movies.csv") # to choose index col throw an argument␣
↪index_col = 0
```

[123]:
```
directors = pd.read_csv("directors.csv")
```

[124]:
```
movies.shape
```

[124]:
```
(1465, 12)
```

[125]:
```
directors.shape
```

[125]:
```
(2349, 4)
```

[126]:
```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Unnamed: 0    1465 non-null   int64  
 1   id            1465 non-null   int64  
 2   budget        1465 non-null   int64  
 3   popularity    1465 non-null   int64  
 4   revenue       1465 non-null   int64  
 5   title         1465 non-null   object 
 6   vote_average  1465 non-null   float64
 7   vote_count    1465 non-null   int64  
 8   director_id   1465 non-null   int64  
 9   year          1465 non-null   int64  
 10  month         1465 non-null   object 
 11  day           1465 non-null   object 
dtypes: float64(1), int64(8), object(3)
memory usage: 137.5+ KB
```

[127]:
```
directors.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2349 entries, 0 to 2348
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Unnamed: 0     2349 non-null   int64
 1   director_name  2349 non-null   object
 2   id             2349 non-null   int64
 3   gender         1724 non-null   object
dtypes: int64(2), object(2)
memory usage: 73.5+ KB
```

[128]:
```
movies.drop('Unnamed: 0',axis=1,inplace=True)
```

[129]:
```
directors.drop('Unnamed: 0',axis=1,inplace=True)
```

[130]:
```
movies.sort_values('vote_count',ascending=False)
```

[130]:

	id	budget	popularity	revenue	title	vote_average	vote_count	director_id	year	month	day
59	43693	160000000	167	825532764	Inception	8.1	13752	4765	2010	Jul	Wednesday
45	43662	185000000	187	1004558444	The Dark Knight	8.2	12002	4765	2008	Jul	Wednesday
0	43597	237000000	150	2787965087	Avatar	7.2	11800	4762	2009	Dec	Thursday
58	43692	165000000	724	675120017	Interstellar	8.1	10867	4765			
178	43884	100000000	82	425368238	Django Unchained	7.8	10099	4927			
…	…	…	…	…	…	…	…	…	…	…	…
1431	47962	0	0	0	Walking and Talking	6.6	7	6204			
879	45373	0	0	0	The Magic Flute	6.9	6	4847			
1438	48145	500000	0	0	Everything Put Together	5.0	2	4773			
1440	48155	0	0	0	Alleluia! The Devil's Carnival	6.0	2	6056			
1378	47387	0	0	0	An Everlasting Piece	6.0	1	5037			

```
58    2014   Nov  Wednesday
178   2012   Dec    Tuesday
…     …    …          …
1431  1996   Jul  Wednesday
879   2006   Sep   Thursday
1438  2001   Nov     Friday
1440  2016   Mar    Tuesday
1378  2000   Dec     Friday

[1465 rows x 11 columns]
```

[131]:
```
movies.head()
```

[131]:
```
      id     budget  popularity     revenue  \
0  43597  237000000         150  2787965087
1  43598  300000000         139   961000000
2  43599  245000000         107   880674609
3  43600  250000000         112  1084939099
4  43602  258000000         115   890871626

                                    title  vote_average  vote_count  \
0                                  Avatar           7.2       11800
1  Pirates of the Caribbean: At World's End           6.9        4500
2                                 Spectre           6.3        4466
3                    The Dark Knight Rises           7.6        9106
4                             Spider-Man 3           5.9        3576

   director_id  year month       day
0         4762  2009   Dec  Thursday
1         4763  2007   May  Saturday
2         4764  2015   Oct    Monday
3         4765  2012   Jul    Monday
4         4767  2007   May   Tuesday
```

[132]:
```
directors.tail()
```

[132]:
```
         director_name    id gender
2344     Shane Carruth  7106   Male
2345  Neill Dela Llana  7107    NaN
2346       Scott Smith  7108    NaN
2347       Daniel Hsia  7109   Male
2348  Brian Herzlinger  7110   Male
```

1 Unique Count

[133]:
```
movies.title.nunique()  #Return the no of unique titles
```

[133]: 1465

[134]:
```
directors.id.nunique()
```

[134]: 2349

[135]:
```
movies.director_id.nunique()
```

[135]: 199

2 Whether all the directors in directors data is present in movies data

[136]:
```
np.all(movies.director_id.isin(directors.id))
```

[136]: True

3 Join both Movies and Directors table

[137]:
```
data = movies.merge(directors,left_on='director_id',right_on='id',how='left')
```

[138]:
```
data.head()
```

[138]:

	id_x	budget	popularity	revenue	\
0	43597	237000000	150	2787965087	
1	43598	300000000	139	961000000	
2	43599	245000000	107	880674609	
3	43600	250000000	112	1084939099	
4	43602	258000000	115	890871626	

	title	vote_average	vote_count	\
0	Avatar	7.2	11800	
1	Pirates of the Caribbean: At World's End	6.9	4500	
2	Spectre	6.3	4466	
3	The Dark Knight Rises	7.6	9106	
4	Spider-Man 3	5.9	3576	

	director_id	year	month	day	director_name	id_y	gender
0	4762	2009	Dec	Thursday	James Cameron	4762	Male
1	4763	2007	May	Saturday	Gore Verbinski	4763	Male
2	4764	2015	Oct	Monday	Sam Mendes	4764	Male
3	4765	2012	Jul	Monday	Christopher Nolan	4765	Male

```
4         4767  2007    May    Tuesday          Sam Raimi  4767    Male
```

[139]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 14 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   id_x           1465 non-null   int64
 1   budget         1465 non-null   int64
 2   popularity     1465 non-null   int64
 3   revenue        1465 non-null   int64
 4   title          1465 non-null   object
 5   vote_average   1465 non-null   float64
 6   vote_count     1465 non-null   int64
 7   director_id    1465 non-null   int64
 8   year           1465 non-null   int64
 9   month          1465 non-null   object
 10  day            1465 non-null   object
 11  director_name  1465 non-null   object
 12  id_y           1465 non-null   int64
 13  gender         1341 non-null   object
dtypes: float64(1), int64(8), object(5)
memory usage: 160.4+ KB
```

[140]:
```
data.drop(['id_y'],axis=1,inplace=True)
```

[141]:
```
data
```

[141]:

	id_x	budget	popularity	revenue	title	vote_average	vote_count
0	43597	237000000	150	2787965087	Avatar	7.2	11800
1	43598	300000000	139	961000000	Pirates of the Caribbean: At World's End	6.9	4500
2	43599	245000000	107	880674609	Spectre	6.3	4466
3	43600	250000000	112	1084939099	The Dark Knight Rises	7.6	9106
4	43602	258000000	115	890871626			
…	…	…	…	…			
1460	48363	0	3	321952			
1461	48370	27000	19	3151130			
1462	48375	0	7	0			
1463	48376	0	3	0			
1464	48395	220000	14	2040920			

```
4                                  Spider-Man 3           5.9        3576
…                                           …           …          …
1460                          The Last Waltz           7.9          64
1461                                  Clerks           7.4         755
1462                                Rampage            6.0         131
1463                                 Slacker           6.4          77
1464                             El Mariachi           6.6         238

      director_id  year month       day      director_name gender
0            4762  2009   Dec  Thursday      James Cameron   Male
1            4763  2007   May  Saturday     Gore Verbinski   Male
2            4764  2015   Oct    Monday         Sam Mendes   Male
3            4765  2012   Jul    Monday  Christopher Nolan   Male
4            4767  2007   May   Tuesday          Sam Raimi   Male
…             …    …     …       …                 …    …
1460         4809  1978   May    Monday    Martin Scorsese   Male
1461         5369  1994   Sep   Tuesday        Kevin Smith   Male
1462         5148  2009   Aug    Friday           Uwe Boll   Male
1463         5535  1990   Jul    Friday  Richard Linklater   Male
1464         5097  1992   Sep    Friday   Robert Rodriguez    NaN

[1465 rows x 13 columns]
```

[142]:
```
data.rename({'id_x':'movies_id'},axis=1,inplace=True)
```

[143]:
```
data
```

[143]:
```
      movies_id     budget  popularity     revenue  \
0         43597  237000000         150  2787965087
1         43598  300000000         139   961000000
2         43599  245000000         107   880674609
3         43600  250000000         112  1084939099
4         43602  258000000         115   890871626
…           …        …          …          …
1460      48363          0           3      321952
1461      48370      27000          19     3151130
1462      48375          0           7           0
1463      48376          0           3           0
1464      48395     220000          14     2040920

                                      title  vote_average  vote_count  \
0                                    Avatar           7.2       11800
1     Pirates of the Caribbean: At World's End           6.9        4500
2                                   Spectre           6.3        4466
3                       The Dark Knight Rises           7.6        9106
4                                  Spider-Man 3           5.9        3576
…                                           …           …          …
```

```
1460                          The Last Waltz          7.9          64
1461                                  Clerks          7.4         755
1462                                 Rampage          6.0         131
1463                                 Slacker          6.4          77
1464                              El Mariachi          6.6         238

      director_id  year month       day      director_name gender
0            4762  2009   Dec  Thursday      James Cameron   Male
1            4763  2007   May  Saturday     Gore Verbinski   Male
2            4764  2015   Oct    Monday         Sam Mendes   Male
3            4765  2012   Jul    Monday  Christopher Nolan   Male
4            4767  2007   May   Tuesday          Sam Raimi   Male
…             …    …     …        …                …     …
1460         4809  1978   May    Monday    Martin Scorsese   Male
1461         5369  1994   Sep   Tuesday        Kevin Smith   Male
1462         5148  2009   Aug    Friday           Uwe Boll   Male
1463         5535  1990   Jul    Friday  Richard Linklater   Male
1464         5097  1992   Sep    Friday   Robert Rodriguez    NaN

[1465 rows x 13 columns]
```

4 Describe function will give the description of Numeric data

[144]:
```
data.describe()
```

[144]:
```
          movies_id        budget   popularity       revenue  vote_average  \
count   1465.000000  1.465000e+03  1465.000000  1.465000e+03   1465.000000
mean   45225.191126  4.802295e+07    30.855973  1.432539e+08      6.368191
std     1189.096396  4.935541e+07    34.845214  2.064918e+08      0.818033
min    43597.000000  0.000000e+00     0.000000  0.000000e+00      3.000000
25%    44236.000000  1.400000e+07    11.000000  1.738013e+07      5.900000
50%    45022.000000  3.300000e+07    23.000000  7.578164e+07      6.400000
75%    45990.000000  6.600000e+07    41.000000  1.792469e+08      6.900000
max    48395.000000  3.800000e+08   724.000000  2.787965e+09      8.300000

         vote_count  director_id         year
count   1465.000000  1465.000000  1465.000000
mean    1146.396587  5040.192491  2002.615017
std     1578.077438   258.059631     8.680141
min        1.000000  4762.000000  1976.000000
25%      216.000000  4845.000000  1998.000000
50%      571.000000  4964.000000  2004.000000
75%     1387.000000  5179.000000  2009.000000
max    13752.000000  6204.000000  2016.000000
```

5 Describe for non-numeric data

[145]:
```
data.describe(include=object)
```

[145]:

	title	month	day	director_name	gender
count	1465	1465	1465	1465	1341
unique	1465	12	7	199	2
top	Avatar	Dec	Friday	Steven Spielberg	Male
freq	1	193	654	26	1309

6 Changing the number into Millions, Lakhs, Thousands (Short form)

Currency Meter	
10 Lakhs	1 Million
1 Crore	10 Million
100 Crore	1 Billion

[146]:
```
data['budget']=data['budget']/1000000
```

[147]:
```
data
```

[147]:

	movies_id	budget	popularity	revenue	\
0	43597	237.000	150	2787965087	
1	43598	300.000	139	961000000	
2	43599	245.000	107	880674609	
3	43600	250.000	112	1084939099	
4	43602	258.000	115	890871626	
…	…	…	…	…	
1460	48363	0.000	3	321952	
1461	48370	0.027	19	3151130	
1462	48375	0.000	7	0	
1463	48376	0.000	3	0	
1464	48395	0.220	14	2040920	

	title	vote_average	vote_count	\
0	Avatar	7.2	11800	
1	Pirates of the Caribbean: At World's End	6.9	4500	
2	Spectre	6.3	4466	
3	The Dark Knight Rises	7.6	9106	
4	Spider-Man 3	5.9	3576	
…	…	…	…	
1460	The Last Waltz	7.9	64	
1461	Clerks	7.4	755	
1462	Rampage	6.0	131	

```
1463                                Slacker           6.4          77
1464                           El Mariachi           6.6         238

      director_id  year month       day      director_name gender
0            4762  2009   Dec  Thursday      James Cameron   Male
1            4763  2007   May  Saturday     Gore Verbinski   Male
2            4764  2015   Oct    Monday         Sam Mendes   Male
3            4765  2012   Jul    Monday  Christopher Nolan   Male
4            4767  2007   May   Tuesday          Sam Raimi   Male
…             …    …     …       …                …    …
1460         4809  1978   May    Monday    Martin Scorsese   Male
1461         5369  1994   Sep   Tuesday        Kevin Smith   Male
1462         5148  2009   Aug    Friday           Uwe Boll   Male
1463         5535  1990   Jul    Friday  Richard Linklater   Male
1464         5097  1992   Sep    Friday   Robert Rodriguez    NaN

[1465 rows x 13 columns]
```

7 Find out highly rated movies and there director details

[148]:
```
a = data.loc[data.vote_average > 7]
a[['title','vote_average','vote_count','year']]
```

[148]:
```
                                        title  vote_average  vote_count  \
0                                      Avatar           7.2       11800
3                       The Dark Knight Rises           7.6        9106
14   The Hobbit: The Battle of the Five Armies           7.1        4760
16         The Hobbit: The Desolation of Smaug           7.6        4524
19                                     Titanic           7.5        7562
…                                          …             …           …
1456                             Eraserhead           7.5         485
1457                             The Mighty           7.1          51
1458                                     Pi           7.1         586
1460                          The Last Waltz           7.9          64
1461                                  Clerks           7.4         755

      year
0     2009
3     2012
14    2014
16    2013
19    1997
…     …
1456  1977
1457  1998
1458  1998
```

```
1460  1978
1461  1994

[301 rows x 4 columns]
```

8 Highly rated movies release after 2014

[149]:
```
b = data.loc[(data.vote_average > 7) & (data.year > 2014)]
```

[150]:
```
b.reset_index()
```

[150]:

	index	movies_id	budget	popularity	revenue	title	vote_average	vote_count	director_id	year	month	day	director_name	gender
0	30	43641	190.0	102	1506249360	Furious 7	7.3	4176	4794	2015	Apr	Wednesday	James Wan	Male
1	78	43724	150.0	434	378858340	Mad Max: Fury Road	7.2	9427	4845	2015	May	Wednesday	George Miller	Male
2	106	43773	135.0	100	532950503	The Revenant	7.3	6396	4874	2015	Dec	Friday	Alejandro González Iñárritu	Male
3	162	43867	108.0	167	630161890	The Martian	7.6	7268	4779	2015	Sep	Wednesday	Ridley Scott	Male
4	312	44128	75.0	48	108145109	The Man from U.N.C.L.E.	7.1	2265	4888	2015	Aug	Thursday	Guy Ritchie	Male
5	394	44281	44.0	68	155760117	The Hateful Eight	7.6	4274	4927	2015	Dec	Friday	Quentin Tarantino	Male
6	625	44770	35.0	53	194564672	The Intern	7.1	1881	4978	2015	Sep	Thursday		
7	635	44784	40.0	48	165478348	Bridge of Spies	7.2	2583	4799	2015	Oct	Thursday		
8	808	45194	30.0	65	91709827	Southpaw	7.3	2067	5034	2015	Jun	Monday		
9	833	45293	28.0	61	201634991	Straight Outta Compton	7.7	1355	5033	2015	Aug	Thursday		
10	839	45301	28.0	57	133346506	The Big Short	7.3	2607	4925	2015	Dec	Friday		
11	1344	47181	5.0	22	24804129	Race	7.1	478	5008	2016	Feb	Friday		

```
6                 Nancy Meyers  Female
7              Steven Spielberg    Male
8                 Antoine Fuqua    Male
9                  F. Gary Gray    Male
10                   Adam McKay    Male
11               Stephen Hopkins    Male
```

9 Find the movies release on either friday's or Sunday's

[151]:
```
c = data.loc[(data.day == 'Friday') | (data.day == 'Sunday')]
```

[152]:
```
c
```

[152]:
```
      movies_id  budget  popularity    revenue  \
22        43627  200.00          35  783766341
25        43632  150.00          21  836297228
53        43672  175.00          44  264218220
61        43696   38.00           6  207283925
65        43701  160.00          21  181674817
…           …       …           …          …
1458      48335    0.06          27    3221152
1459      48359    0.00           2          0
1462      48375    0.00           7          0
1463      48376    0.00           3          0
1464      48395    0.22          14    2040920

                                  title  vote_average  vote_count  \
22                          Spider-Man 2           6.7        4321
25    Transformers: Revenge of the Fallen           6.0        3138
53                             Waterworld           5.9         992
61                The Fast and the Furious           6.6        3428
65                               Poseidon           5.5         583
…                                     …             …           …
1458                                Pi           7.1         586
1459                    George Washington           6.4          36
1462                              Rampage           6.0         131
1463                              Slacker           6.4          77
1464                           El Mariachi           6.6         238

      director_id  year month     day       director_name gender
22           4767  2004   Jun  Friday           Sam Raimi   Male
25           4788  2009   Jun  Friday         Michael Bay   Male
53           4814  1995   Jul  Friday      Kevin Reynolds    NaN
61           4810  2001   Jun  Friday           Rob Cohen   Male
65           4833  2006   May  Friday   Wolfgang Petersen   Male
…             …     …     …       …                   …      …
```

```
1458         4881  1998   Jul  Friday    Darren Aronofsky   Male
1459         5231  2000   Oct  Sunday  David Gordon Green   Male
1462         5148  2009   Aug  Friday           Uwe Boll   Male
1463         5535  1990   Jul  Friday   Richard Linklater   Male
1464         5097  1992   Sep  Friday    Robert Rodriguez    NaN

[700 rows x 13 columns]
```

10 Display top 5 Popular movies

[153]:
```
data.sort_values('popularity',ascending=False).head()
```

[153]:
```
     movies_id  budget  popularity     revenue  \
58       43692   165.0         724   675120017
78       43724   150.0         434   378858340
119      43796   140.0         271   655011224
120      43797   125.0         206   752100229
45       43662   185.0         187  1004558444

                                                title  vote_average  \
58                                       Interstellar           8.1
78                                   Mad Max: Fury Road           7.2
119  Pirates of the Caribbean: The Curse of the Bla…           7.5
120              The Hunger Games: Mockingjay - Part 1           6.6
45                                     The Dark Knight           8.2

     vote_count  director_id  year month        day      director_name gender
58        10867         4765  2014   Nov  Wednesday  Christopher Nolan   Male
78         9427         4845  2015   May  Wednesday      George Miller   Male
119        6985         4763  2003   Jul  Wednesday     Gore Verbinski   Male
120        5584         4831  2014   Nov    Tuesday   Francis Lawrence   Male
45        12002         4765  2008   Jul  Wednesday  Christopher Nolan   Male
```

11 Convert all Males directors into 0 and Female Directors to 1 in your dataframe

[154]:
```
def change_to_num(gender):
        if gender == 'Male':
            return 0
        else:
            return 1

data['gender'] = data['gender'].apply(change_to_num)
```

[155]:
```
data.head()
```

[155]:

	movies_id	budget	popularity	revenue	title	vote_average	vote_count	director_id	year	month	day	director_name	gender
0	43597	237.0	150	2787965087	Avatar	7.2	11800	4762	2009	Dec	Thursday	James Cameron	0
1	43598	300.0	139	961000000	Pirates of the Caribbean: At World's End	6.9	4500	4763	2007	May	Saturday	Gore Verbinski	0
2	43599	245.0	107	880674609	Spectre	6.3	4466	4764	2015	Oct	Monday	Sam Mendes	0
3	43600	250.0	112	1084939099	The Dark Knight Rises	7.6	9106	4765	2012	Jul	Monday	Christopher Nolan	0
4	43602	258.0	115	890871626	Spider-Man 3	5.9	3576	4767	2007	May	Tuesday	Sam Raimi	0

12 Find the sum of Revenue and Budget

[166]:
```
data['revenue'].sum()/1000000
```

[166]: 209866.997305

[ ]:
```

```