▾ homework

1. Experienced Directors - Find out details of all the movies where the director has created atleast 15 movies in entire career.

```
import pandas as pd
```

Screenshot%202022-09-05%20at%2020.56.16.png

```
import pandas as pd
import numpy as np
!gdown 1s2TkjSpzNc4SyxqRrQleZyDIHlc7bxnd
!gdown 1Ws-_s1fHZ9nHfGLVUQurbHDvStePlEJm
movies = pd.read_csv('movies.csv', index_col=0)
directors = pd.read_csv('directors.csv',index_col=0)
data = movies.merge(directors, how='left', left_on='director_id',right_on='id')
data.drop(['director_id','id_y'],axis=1,inplace=True)
```

```
    Downloading...
    From: https://drive.google.com/uc?id=1s2TkjSpzNc4SyxqRrQleZyDIHlc7bxnd
    To: /Users/nikhilsanghi/Downloads/dsml-course-main-live/batches/May-Beg-Aug-Adv/movie
    100%|██████████████████████████████████████| 112k/112k [00:00<00:00, 537kB/s]
    Downloading...
    From: https://drive.google.com/uc?id=1Ws-_s1fHZ9nHfGLVUQurbHDvStePlEJm
    To: /Users/nikhilsanghi/Downloads/dsml-course-main-live/batches/May-Beg-Aug-Adv/direc
    100%|██████████████████████████████████████| 65.4k/65.4k [00:00<00:00, 2.90MB/s]
```

```
data
```

	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	6.9	4500

data.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1465 entries, 0 to 1464
Data columns (total 12 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   year           1465 non-null    int64
 8   month          1465 non-null    object
 9   day            1465 non-null    object
 10  director_name  1465 non-null    object
 11  gender         1341 non-null    object
dtypes: float64(1), int64(6), object(5)
memory usage: 148.8+ KB
```

data.describe(include="all")

	id_x	budget	popularity	revenue	title	vote_average	
count	1465.000000	1.465000e+03	1465.000000	1.465000e+03	1465	1465.000000	1
unique	NaN	NaN	NaN	NaN	1465	NaN	
top	NaN	NaN	NaN	NaN	Avatar	NaN	
freq	NaN	NaN	NaN	NaN	1	NaN	
mean	45225.191126	4.802295e+07	30.855973	1.432539e+08	NaN	6.368191	1
std	1189.096396	4.935541e+07	34.845214	2.064918e+08	NaN	0.818033	1
min	43597.000000	0.000000e+00	0.000000	0.000000e+00	NaN	3.000000	
25%	44236.000000	1.400000e+07	11.000000	1.738013e+07	NaN	5.900000	
50%	45022.000000	3.300000e+07	23.000000	7.578164e+07	NaN	6.400000	
75%	45990.000000	6.600000e+07	41.000000	1.792469e+08	NaN	6.900000	1
max	48395.000000	3.800000e+08	724.000000	2.787965e+09	NaN	8.300000	13

data

	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	Spider-Man 3	5.9	3576
...	...	...	...	...	...	...	...
1460	48363	0	3	321952	The Last Waltz	7.9	64
1461	48370	27000	19	3151130	Clerks	7.4	755

```
data["budget"]=(data["budget"]/1000000).round(2)
data
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	yea
0	43597	237.00	150	2787965087	Avatar	7.2	11800	200
1	43598	300.00	139	961000000	Pirates of the Caribbean: At World's End	6.9	4500	200
2	43599	245.00	107	880674609	Spectre	6.3	4466	20
3	43600	250.00	112	1084939099	The Dark Knight Rises	7.6	9106	20
4	43602	258.00	115	890871626	Spider-Man 3	5.9	3576	200
...	...	...	...	...	...	...	...	
1460	48363	0.00	3	321952	The Last Waltz	7.9	64	19
1461	48370	0.03	19	3151130	Clerks	7.4	755	199

```
data["revenue"]=(data["revenue"]/1000000).round(2)
data
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.00	150	2787.97	Avatar	7.2	11800	2009
1	43598	300.00	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	2007
2	43599	245.00	107	880.67	Spectre	6.3	4466	2015
3	43600	250.00	112	1084.94	The Dark Knight Rises	7.6	9106	2012
4	43602	258.00	115	890.87	Spider-Man 3	5.9	3576	2007
...	...	...	...	...	...	...	...	...
1460	48363	0.00	3	0.32	The Last Waltz	7.9	64	1978

```
data.loc[data["vote_average"]>=7]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.00	150	2787.97	Avatar	7.2	11800	2009
3	43600	250.00	112	1084.94	The Dark Knight Rises	7.6	9106	2012
8	43609	200.00	145	1065.66	Pirates of the Caribbean: Dead Man's Chest	7.0	5246	2006
14	43616	250.00	120	956.02	The Hobbit: The Battle of the Five Armies	7.1	4760	2014
16	43619	250.00	94	958.40	The Hobbit: The Desolation of Smaug	7.6	4524	2013
...	...	...	...	...	...	...	...	...

```
data[data["vote_average"]>=7][["title","director_name"]]
```

	title	director_name
0	Avatar	James Cameron
3	The Dark Knight Rises	Christopher Nolan
8	Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski
14	The Hobbit: The Battle of the Five Armies	Peter Jackson
16	The Hobbit: The Desolation of Smaug	Peter Jackson
...	...	...
1456	Eraserhead	David Lynch
1457	The Mighty	Peter Chelsom
1458	Pi	Darren Aronofsky
1460	The Last Waltz	Martin Scorsese
1461	Clerks	Kevin Smith

```
data.loc[data["vote_average"]>=7,["title","director_name"]]
```

	title	director_name
0	Avatar	James Cameron
3	The Dark Knight Rises	Christopher Nolan
8	Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski
14	The Hobbit: The Battle of the Five Armies	Peter Jackson
16	The Hobbit: The Desolation of Smaug	Peter Jackson
...	...	...
1456	Eraserhead	David Lynch
1457	The Mighty	Peter Chelsom
1458	Pi	Darren Aronofsky
1460	The Last Waltz	Martin Scorsese
1461	Clerks	Kevin Smith

363 rows × 2 columns

```
data.loc[(data["vote_average"]>=7)&(data["year"]>=2015)]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
30	43641	190.0	102	1506.25	Furious 7	7.3	4176	2015
78	43724	150.0	434	378.86	Mad Max: Fury Road	7.2	9427	2015
106	43773	135.0	100	532.95	The Revenant	7.3	6396	2015
162	43867	108.0	167	630.16	The Martian	7.6	7268	2015
312	44128	75.0	48	108.15	The Man from U.N.C.L.E.	7.1	2265	2015
394	44281	44.0	68	155.76	The Hateful Eight	7.6	4274	2015

13 Hours:

```
data.loc[(data["vote_average"]>=7)&(data["year"]>=2015),["title"]]
```

	title
30	Furious 7
78	Mad Max: Fury Road
106	The Revenant
162	The Martian
312	The Man from U.N.C.L.E.
394	The Hateful Eight
519	13 Hours: The Secret Soldiers of Benghazi
617	The Conjuring 2
625	The Intern
635	Bridge of Spies
808	Southpaw
833	Straight Outta Compton
839	The Big Short
1344	Race

```
data
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.00	150	2787.97	Avatar	7.2	11800	2009
1	43598	300.00	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	2007
2	43599	245.00	107	880.67	Spectre	6.3	4466	2015
3	43600	250.00	112	1084.94	The Dark Knight Rises	7.6	9106	2012
4	43602	258.00	115	890.87	Spider-Man 3	5.9	3576	2007
...	...	...	...	...	...	...	...	...

```
data.loc[(data["day"]=="Saturday")|(data["day"]=="Friday")]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	yea
1	43598	300.00	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	200
12	43614	380.00	135	1045.71	Pirates of the Caribbean: On Stranger Tides	6.4	4948	201
22	43627	200.00	35	783.77	Spider-Man 2	6.7	4321	200
25	43632	150.00	21	836.30	Transformers: Revenge of the Fallen	6.0	3138	200
40	43656	200.00	45	769.65	2012	5.6	4903	200
...	...	...	...	...	...	...	...	.
1457	48323	0.00	5	0.00	The Mighty	7.1	51	199
1458	48335	0.06	27	3.22	Pi	7.1	586	199

```
data.loc[(data["year"]>=2015) & (data["vote_average"]>=7) & ((data["month"]=="Oct")|(data[
```

```
data.loc[data["month"].isin(["Dec","Nov","Oct"])]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.00	150	2787.97	Avatar	7.2	11800	2009
2	43599	245.00	107	880.67	Spectre	6.3	4466	2015
7	43608	200.00	107	586.09	Quantum of Solace	6.1	2965	2008
14	43616	250.00	120	956.02	The Hobbit: The Battle of the Five Armies	7.1	4760	2014
16	43619	250.00	94	958.40	The Hobbit: The Desolation of Smaug	7.6	4524	2013
...	...	...	...	...	...	...	...	...
1451	48268	0.20	13	4.51	Swingers	6.8	253	1996
1452	48274	0.00	5	2.61	Three	6.3	31	2010
1454	48300	0.15	8	2.89	Tadpole	5.5	19	2002

```
data.loc[(data["year"]>=2015) & (data["vote_average"]>=7) & (data["month"].isin(["Dec","No
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
106	43773	135.0	100	532.95	The Revenant	7.3	6396	2015
394	44281	44.0	68	155.76	The Hateful Eight	7.6	4274	2015

```
data.loc[data["title"]<"Avengers"]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.0	150	2787.97	Avatar	7.2	11800	2009
23	43629	200.0	78	1025.49	Alice in Wonderland	6.4	4645	2010
40	43656	200.0	45	769.65	2012	5.6	4903	2009

```
data.loc[data["title"].str.contains("The",case=False)]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
1	43598	300.00	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	2007
3	43600	250.00	112	1084.94	The Dark Knight Rises	7.6	9106	2012
8	43609	200.00	145	1065.66	Pirates of the Caribbean: Dead Man's Chest	7.0	5246	2006
9	43610	255.00	49	89.29	The Lone Ranger	5.9	2311	2013
11	43612	225.00	53	419.65	The Chronicles of Narnia: Prince Caspian	6.3	1630	2008
...								

```
data.loc[data["title"].str.startswith("Batman")]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year	m
5	43606	250.0	155	873.26	Batman v Superman: Dawn of Justice	5.7	7004	2016	
74	43716	150.0	115	374.22	Batman Begins	7.5	7359	2005	
128	43807	125.0	50	238.21	Batman & Robin	4.2	1418	1997	
					Batman				

```
top_10_popular_movies=data.sort_values(["popularity"],ascending=False)
top_10_popular_movies.loc[:,["title"]]
```

	title
58	Interstellar
78	Mad Max: Fury Road
119	Pirates of the Caribbean: The Curse of the Bla...
120	The Hunger Games: Mockingjay - Part 1
45	The Dark Knight
...	...
1437	Finishing The Game
1448	Chuck & Buck
1438	Everything Put Together
1378	An Everlasting Piece
1440	Alleluia! The Devil's Carnival

1465 rows × 1 columns

```
data.sort_values(["popularity"],ascending=False)[["title"]]
```

	title
58	Interstellar
78	Mad Max: Fury Road
119	Pirates of the Caribbean: The Curse of the Bla...
120	The Hunger Games: Mockingjay - Part 1
45	The Dark Knight
...	...
1437	Finishing The Game
1448	Chuck & Buck
1438	Everything Put Together
1378	An Everlasting Piece
1440	Alleluia! The Devil's Carnival

1465 rows × 1 columns

```
data.loc[data["director_name"]=="Christopher Nolan",["title"]]
```

	title
3	The Dark Knight Rises
45	The Dark Knight
58	Interstellar
59	Inception
74	Batman Begins
565	Insomnia
641	The Prestige

```
data["director_name"].value_counts()
```

```
Steven Spielberg        26
Martin Scorsese         19
Clint Eastwood          19
Woody Allen             18
Ridley Scott            16
                        ..
Tim Hill                 5
Jonathan Liebesman       5
Roman Polanski           5
Larry Charles            5
Nicole Holofcener        5
Name: director_name, Length: 199, dtype: int64
```

Screenshot%202022-09-06%20at%2000.28.18.png

```
data.groupby("director_name")["budget"].max()
```

```
director_name
Adam McKay                    100.0
Adam Shankman                  80.0
Alejandro González Iñárritu   135.0
Alex Proyas                   140.0
Alexander Payne                30.0
                              ...
Wes Craven                     40.0
Wolfgang Petersen             175.0
Woody Allen                    30.0
Zack Snyder                   250.0
Zhang Yimou                    94.0
Name: budget, Length: 199, dtype: float64
```

```
data.groupby("director_name")["title"].count()
```

```
director_name
Adam McKay                    6
Adam Shankman                 8
Alejandro González Iñárritu   6
Alex Proyas                   5
Alexander Payne               5
                              ..
```

```
        Wes Craven                       10
        Wolfgang Petersen                 7
        Woody Allen                      18
        Zack Snyder                       7
        Zhang Yimou                       6
        Name: title, Length: 199, dtype: int64
```

```
data.groupby("director_name")["year"].min()
```

```
        director_name
        Adam McKay                       2004
        Adam Shankman                    2001
        Alejandro González Iñárritu      2000
        Alex Proyas                      1994
        Alexander Payne                  1999
                                          ...
        Wes Craven                       1984
        Wolfgang Petersen                1981
        Woody Allen                      1977
        Zack Snyder                      2004
        Zhang Yimou                      2002
        Name: year, Length: 199, dtype: int64
```

```
data.groupby("director_name")["year"].max()
```

```
        director_name
        Adam McKay                       2015
        Adam Shankman                    2012
        Alejandro González Iñárritu      2015
        Alex Proyas                      2016
        Alexander Payne                  2013
                                          ...
        Wes Craven                       2011
        Wolfgang Petersen                2006
        Woody Allen                      2013
        Zack Snyder                      2016
        Zhang Yimou                      2014
        Name: year, Length: 199, dtype: int64
```

```
df_agg=data.groupby("director_name")[["title","year"]].aggregate({"title":"count","year":[
df_agg
```

	title	year	
	count	max	min
director_name			
Adam McKay	6	2015	2004
Adam Shankman	8	2012	2001
Alejandro González Iñárritu	6	2015	2000
Alex Proyas	5	2016	1994
Alexander Payne	5	2013	1999
...	...	...	...

```
df_agg.columns
```

```
MultiIndex([('title', 'count'),
            ( 'year',   'max'),
            ( 'year',   'min')],
           )
```

```
df_agg.columns=["_".join(col) for col in df_agg.columns]
```
199 rows × 3 columns

```
df_agg
```

	title_count	year_max	year_min
director_name			
Adam McKay	6	2015	2004
Adam Shankman	8	2012	2001
Alejandro González Iñárritu	6	2015	2000
Alex Proyas	5	2016	1994
Alexander Payne	5	2013	1999
...	...	...	...
Wes Craven	10	2011	1984
Wolfgang Petersen	7	2006	1981
Woody Allen	18	2013	1977
Zack Snyder	7	2016	2004
Zhang Yimou	6	2014	2002

199 rows × 3 columns

```
df_agg.reset_index(inplace=True)
```

df_agg

	director_name	title_count	year_max	year_min
0	Adam McKay	6	2015	2004
1	Adam Shankman	8	2012	2001
2	Alejandro González Iñárritu	6	2015	2000
3	Alex Proyas	5	2016	1994
4	Alexander Payne	5	2013	1999
...	...	...	...	...
194	Wes Craven	10	2011	1984
195	Wolfgang Petersen	7	2006	1981
196	Woody Allen	18	2013	1977
197	Zack Snyder	7	2016	2004
198	Zhang Yimou	6	2014	2002

199 rows × 4 columns

```
df_agg["years_active"]=df_agg["year_max"]-df_agg["year_min"]
```

df_agg

	director_name	title_count	year_max	year_min	years_active
0	Adam McKay	6	2015	2004	11
1	Adam Shankman	8	2012	2001	11
2	Alejandro González Iñárritu	6	2015	2000	15
3	Alex Proyas	5	2016	1994	22
4	Alexander Payne	5	2013	1999	14
...	...	...	...	...	...
194	Wes Craven	10	2011	1984	27
195	Wolfgang Petersen	7	2006	1981	25
196	Woody Allen	18	2013	1977	36
197	Zack Snyder	7	2016	2004	12
198	Zhang Yimou	6	2014	2002	12

199 rows × 5 columns

```
df_agg["movies_per_year"]=df_agg["title_count"]/df_agg["years_active"]
```

```python
df_agg.sort_values("movies_per_year",ascending=False).head(10)
```

	director_name	title_count	year_max	year_min	years_active	movies_per_year
190	Tyler Perry	9	2013	2006	7	1.285714
73	Jason Friedberg	5	2010	2006	4	1.250000
169	Shawn Levy	11	2014	2002	12	0.916667
158	Robert Rodriguez	16	2014	1992	22	0.727273
1	Adam Shankman	8	2012	2001	11	0.727273
156	Robert Luketic	6	2010	2001	9	0.666667
179	Steven Spielberg	26	2016	1977	39	0.666667
8	Andy Fickman	6	2015	2006	9	0.666667
112	Marc Forster	8	2013	2001	12	0.666667
81	Joel Schumacher	13	2007	1987	20	0.650000

```python
data.loc[data["budget"]>=100]
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.0	150	2787.97	Avatar	7.2	11800	2009
1	43598	300.0	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	2007
2	43599	245.0	107	880.67	Spectre	6.3	4466	2015
3	43600	250.0	112	1084.94	The Dark Knight Rises	7.6	9106	2012
4	43602	258.0	115	890.87	Spider-Man 3	5.9	3576	2007
...	...	...	...	...	...	...	...	...
219	43960	100.0	34	235.93	A.I. Artificial Intelligence	6.8	1974	2001

```python
df_high_budget_dir=data.groupby("director_name").filter(lambda x:x["budget"].max() >=100)
```

df_high_budget_dir

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
0	43597	237.00	150	2787.97	Avatar	7.2	11800	2009
1	43598	300.00	139	961.00	Pirates of the Caribbean: At World's End	6.9	4500	2007
2	43599	245.00	107	880.67	Spectre	6.3	4466	2015
3	43600	250.00	112	1084.94	The Dark Knight Rises	7.6	9106	2012
4	43602	258.00	115	890.87	Spider-Man 3	5.9	3576	2007
...	...	...	...	...	...	...	...	...
1450	48267	0.40	33	100.00	Mad Max	6.6	1213	1979
1451	48268	0.20	13	4.51	Swingers	6.8	253	1996

[[]]

df_high_budget_dir[df_high_budget_dir["director_name"]=="Doug Liman"]

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
52	43671	178.0	79	370.54	Edge of Tomorrow	7.6	4858	2014
148	43845	110.0	44	478.21	Mr. & Mrs. Smith	6.5	2965	2005
251	44016	85.0	21	222.23	Jumper	5.9	1799	2008
399	44291	60.0	86	214.03	The Bourne Identity	7.3	3583	2002
544	44588	22.0	12	24.19	Fair Game	6.5	235	2010

df_high_budget_dir[df_high_budget_dir["director_name"]=="Darren Aronofsky"]

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year	m
113	43786	125.00	46	362.64	Noah	5.6	2350	2014	
751	45065	35.00	24	15.30	The Fountain	6.8	827	2006	
					Swan				
					Wrestler				

```
data[data["budget"]>100].iloc[:, [2,3]]=1
```

```
/Users/nikhilsanghi/opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexing.py
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/u
  self._setitem_single_column(loc, value, pi)
```

```
data.loc[data["budget"]>100, [2,3]]
```

	2	3
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
...	...	...

```
data[data["budget"]>100].iloc[10:20, [2,3]]
```

	popularity	revenue
10	99	662.85
11	53	419.65
12	135	1045.71
13	52	624.03
14	120	956.02
15	37	310.67
16	94	958.40
17	42	372.23
18	61	550.00
19	100	1845.03

```
data
```

	id_x	budget	popularity	revenue	title	vote_average	vote_count	year
					Pirates of At World's			
2	43599	245.00	107	880.67	Spectre	6.3	4466	2015
3	43600	250.00	112	1084.94	Knight	7.6	9106	2012
4	43602	258.00	115	890.87	Spider- Man 3	5.9	3576	2007
...	...	...	...	...	...	...	...	...
1460	48363	0.00	3	0.32	The Last Waltz	7.9	64	1978

Colab paid products - Cancel contracts here