

Bài kiểm tra PySpark Dataframe

Chú ý:

- Ghi rõ họ tên ủa mình bên trong file notebook (nếu không sẽ được xem là không có bài)
- Với mỗi câu hỏi, show 100 dòng KHÔNG TRUNCATE
- Mỗi phim có thể có nhiều dòng phim, bạn cần tách riêng từng dòng phim của mỗi phim khi thực hiện tính toán
- Mỗi bạn submit 2 file: (1) file notebook, và (2) file pdf xuất ra từ notebook mà KHÔNG NÉN

Load dữ liệu từ 2 file **movies_small.csv** và **ratings_small.csv** và thực hiện các truy vấn sau:

1. Xuất số lượng phim được làm mỗi năm. Kết quả được sắp xếp theo năm và có định dạng sau đây. (1 điểm)

Year	Num_of_movies
------	---------------

```
+-----+-----+
|Year|Num_of_movies|
+-----+-----+
|    |13          |
|1902|1           |
|1903|1           |
|1908|1           |
|1915|1           |
|1916|4           |
|1917|1           |
|1919|1           |
|1920|2           |
|1921|1           |
|1922|1           |
|1923|4           |
|1924|5           |
|1925|4           |
|1926|5           |
|1927|7           |
|1928|4           |
|1929|4           |
|1930|5           |
|1931|14          |
|1932|9           |
|1933|12          |
|1934|11          |
|1935|13          |
|1936|18          |
|1937|16          |
|1938|15          |
```

2. Xuất số lượng phim thuộc về mỗi dòng phim được làm mỗi năm. Kết quả được sắp xếp theo năm. Số phim của mỗi dòng được trình bày trong từng cột như sau: (2 điểm)

Year	Action	Animation	Comedy	...
------	--------	-----------	--------	-----

year (no genres listed)	Action	Adventure	Animation	Children	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir	Horror	IMAX	Musical	Mystery	Romance	Sci-Fi	Thriller	War	Western
19	1	null	null	null	null	null	null	2	null	null	null	null	null	null	null	2	2	null	null
1902	null	1	1	null	null	null	null	null	1	1	null	null	null	null	null	1	null	null	null
1903	null	null	null	null	null	1	null	null	1	1	null	null	null	null	null	null	null	1	
1908	null	null	1	null	1	null	null	null	1	1	null	null	null	null	1	1	null	null	null
1915	null	null	null	null	null	null	null	1	1	1	null	null	null	null	1	1	1	1	null
1916	null	1	1	null	null	1	null	1	1	1	null	null	null	1	1	1	1	1	null
1917	null	null	null	null	1	1	null	null	1	1	1	null	null	null	1	1	1	1	null
1919	null	null	null	null	1	1	null	1	1	1	1	null	null	null	1	1	1	1	null
1920	null	null	null	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1921	null	null	null	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1922	null	null	null	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1923	null	1	null	null	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1924	null	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1925	null	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1926	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1927	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1928	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1929	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1930	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1931	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1932	null	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

3. Với mỗi userID, xuất điểm đánh giá trung bình của user đó cho mỗi dòng phim. Kết quả được sắp xếp theo userID. Kết quả của từng dòng phim được trình bày trong mỗi cột như sau: (2 điểm)

User_id	Action	Animation	Comedy	...
---------	--------	-----------	--------	-----

userID (no genres listed)	Action	Adventure	Animation	Children	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir
1	4.322222222222222	4.3882352941176475	4.689655172413793	4.5476190476190474	4.27710843373494	4.355555555555555	null	4.520411764705882	4.297872340425532	5.0
2	3.9545454545454546	4.166666666666667	null	null	4.0	3.8	4.333333333333333	3.8823529411764706	null	null
3	3.5714285714285716	2.727272727272727	0.5	0.5	1.0	0.5	null	0.75	3.375	null
4	3.32	3.6551724137931036	4.0	3.8	3.5096153846153846	3.814814814814815	4.0	3.4833333333333334	3.6842105263157894	4.0
5	3.111111111111111	3.25	4.333333333333333	4.111111111111111	3.466666666666667	3.833333333333333	null	3.8	4.142857142857143	null
6	3.609375	3.893617021276596	4.071428571428571	3.617021276595745	3.370878740151748	3.2857142857142856	null	3.6142857142857143	3.5384615384615383	2.5
7	3.2578125	3.314814814814815	3.392857142857143	3.2	3.163265306122449	3.3076923076923075	null	3.1315789473684212	3.0652173913043477	3.25
8	3.3333333333333335	3.5454545454545454	5.0	4.25	3.2083333333333335	3.888888888888889	null	3.789473684210526	3.25	null
9	3.125	3.8	4.0	4.0	3.6666666666666665	3.142857142857143	null	3.4285714285714284	5.0	4.0
10	3.5	3.5806451612903225	3.866666666666667	3.607142857142857	3.2658227848101267	3.1153846153846154	null	3.1527777777777777	3.4411764705882355	null
11	3.5434782608695654	3.5555555555555554	null	null	3.4166666666666665	3.6923076923076925	null	4.166666666666667	2.0	null
12	4.0	3.0	3.0	null	4.3055555555555555	3.0	null	4.5625	4.25	null
13	4.2	4.166666666666667	null	null	3.272727272727273	3.6666666666666665	null	3.5	3.0	null
14	3.3333333333333335	3.4444444444444446	null	3.0	2.7058823529411766	2.7142857142857144	null	3.7083333333333335	1.0	null
15	3.203389830804745	3.3421052631578947	2.9545454545454546	2.6904761904761907	3.357142857142857	3.5789473684210527	null	3.740740740740741	2.90625	null
16	3.52	3.775	3.9583333333333335	4.0	3.74	3.75	4.0	3.7291666666666665	3.9375	4.0
17	4.230769230769231	4.285714285714286	4.4	4.214285714285714	4.157894736842105	4.25	3.5	4.186274509893922	4.323529411764706	3.75
18	3.588516746411483	3.6338028169014805	3.7500645161290325	3.5428571428571427	3.5087567567567566	3.9347826086956523	3.9444444444444446	3.894485412848068	3.644230769230769	3.875
19	2.7266666666666666	2.815286624203215	2.9375	2.690878431372549	2.64021164021164	2.9027777777777777	null	2.609271523178086	2.8333333333333335	3.333333333333
20	2.75531914893617	3.441806465116279	4.184615384615385	3.9090909090909091	3.6714285714285713	3.24	3.75	3.8080591549295775	3.8645833333333335	1.5
21	4.0	3.4557522123893807	3.663841807906047	3.2413793103448274	3.2049180327868854	3.201492537313433	3.1700821917808217	2.952127659574468	3.0873015873015874	3.5
22	2.7833333333333333	2.3461538461538463	2.7083333333333335	2.75	2.4871794871794872	2.533333333333333	2.0	2.6095890410958904	1.875	4.5
23	3.537037037037037	3.388888888888889	3.8	3.5	3.4166666666666665	3.6538461538461537	3.25	3.642857142857143	3.8214285714285716	3.9545454545454546
24	3.642857142857143	3.6515151515151514	4.25	4.25	3.59	3.9130434782608696	null	3.6875	3.7333333333333334	null

4. Với mỗi phim, xuất ra tên, năm, số lượng ratings, và điểm rating trung bình (từ mọi users) cho bộ phim đó. Kết quả được sắp xếp theo năm, sau đó theo tên phim như sau: (2 điểm)

Year	Movie_name	Num_rating	Average_rating
------	------------	------------	----------------

Year	Movie_name	Num_rating	Average_rating
	Babylon 5	2	2.25
	Black Mirror	1	5.0
	Cosmos	2	4.5
	Death Note: Desu nôto (2006-2007)	1	5.0
	Generation Iron 2	1	3.5
	Hyena Road	1	2.0
	Maria Bamford: Old Baby	1	1.0
	Moonlight	1	5.0
	Nocturnal Animals	2	3.0
	Paterson	1	4.5
	Ready Player One	12	3.5
	The Adventures of Sherlock Holmes and Doctor Watson	1	5.0
	The OA	1	4.0
1902	Trip to the Moon, A (Voyage dans la lune, Le) (1902)	20	3.5
1903	The Great Train Robbery (1903)	4	2.5
1908	The Electric Hotel (1908)	3	4.0
1915	Birth of a Nation, The (1915)	2	2.0
1916	20,000 Leagues Under the Sea (1916)	6	3.75
1916	Intolerance: Love's Struggle Throughout the Ages (1916)	1	3.5
1916	Rink, The (1916)	1	3.5
1916	Snow White (1916)	2	3.5
1917	Immigrant, The (1917)	1	4.5
1919	Daddy Long Legs (1919)	2	2.0
1920	Cabinet of Dr. Caligari, The (Cabinet des Dr. Caligari., Das) (1920)	21	3.857142857142857
1920	One Week (1920)	1	3.5
1921	Kid, The (1921)	10	4.1
1922	Nosferatu (Nosferatu, eine Symphonie des Grauens) (1922)	16	3.53125
1923	Hunchback of Notre Dame, The (1923)	2	3.0
1923	Our Hospitality (1923)	2	3.25
1923	Safety Last! (1923)	6	4.0
1923	Woman of Paris, A (1923)	2	4.0
1924	Aelita: The Queen of Mars (Aelita) (1924)	7	4.5
1924	Greed (1924)	1	4.0

5. Với mỗi userID, xuất ra (a) tên dòng phim nhận được đánh giá (rating) trung bình cao nhất từ user đó, (b) điểm rating trung bình của user đó cho dòng phim này, (c) danh sách top 5 bộ phim thuộc dòng phim này nhận được rating trung bình cao nhất từ mọi user mà user này chưa từng xem (chưa có đánh giá), và (c) điểm đánh giá trung bình của mọi user cho 5 bộ phim này. Ví dụ Film-Noir là dòng phim nhận được rating cao nhất từ user #1, bạn cần xuất ra danh sách 5 bộ phim thuộc về thể loại Film-Noir nhận rating trung bình cao nhất từ mọi user mà user #1 chưa xem (chưa đánh giá). Kết quả được sắp xếp theo userID như sau: (3 điểm)

User_id	Highest_rated_genre_name	Top_5_unrated_movies_with_highest_rating	Avg_rating_from_all_user
---------	--------------------------	--	--------------------------

User_id	Highest_rated_genre_name	User_avg_rating	Top_5_unrated_movies_with_highest_rating	Avg_ratings_from_all_user
1	Film-Noir	5.0	[Rififi (Du rififi chez les hommes) (1955), Long Goodbye, The (1973), 13 Tzmetti (2005), You Only Live Once (1937), Sunset Blvd. (a.k.a. Sunset Bou...	[4.75, 4.67, 4.5, 4.5, 4.33]
2	Romance	4.5	[Spellbound (2011), Jump In! (2007), Duel in the Sun (1946), Laggies (2014), Love and Pigeons (1985)]	[5.0, 5.0, 5.0, 5.0, 5.0]
3	Mystery	5.0	[Scooby-Doo! Abrecadabra-Doo (2010), PK (2014), Adventures Of Sherlock Holmes And Dr. Watson: The Twentieth Century Approches (1986), Get Low (200...	[5.0, 5.0, 5.0, 5.0, 5.0]
4	Horror	4.25	[Sorority House Massacre II (1990), Bloodsucking Bastards (2015), Hollywood Chainsaw Hookers (1988), Slumber Party Massacre II (1987), Breed, The (...]	[5.0, 5.0, 5.0, 5.0, 5.0]
5	Musical	4.4	[Babes in Toyland (1934), Very Potter Sequel, A (2010), True Stories (1996), Woman Is a Woman, A (femme est une femme, Une) (1961), Into the Woods ...]	[5.0, 5.0, 5.0, 5.0, 5.0]
6	IMAX	4.67	[More (1998), Happy Feet Two (2011), Journey to the West: Conquering the Demons (Daai wa sai you chi Chui mo chun kei) (2013), Final Destination 5 ...]	[5.0, 5.0, 4.75, 4.5, 4.25]
7	Horror	4.0	[Bloodsucking Bastards (2015), Vampire in Venice (Nosferatu a Venezia) (Nosferatu in Venice) (1986), Sorority House Massacre II (1990), 'Salem's Lo...	[5.0, 5.0, 5.0, 5.0, 5.0]
8	Animation	5.0	[Nasu: Summer in Andalusia (2003), Garden of Words, The (Koto no ha no nawa) (2013), Wonder Woman (2009), My Life as McDull (Hak dau goo si) (2001)...	[5.0, 5.0, 5.0, 5.0, 5.0]
9	Fantasy	5.0	[The Bremen Town Musicians (1969), Stand, The (1994), Nine Lives of Tomas Katz, The (2000), 7 Faces of Dr. Lao (1964), L.A. Slasher (2015)]	[5.0, 5.0, 5.0, 5.0, 5.0]
10	Animation	3.67	[Runaway Brain (1995), Investigation Held by Kolobki (1986), Return to Treasure Island (1968), Lesson Faust (1994), Into the Forest of Fireflies' ...]	[5.0, 5.0, 5.0, 5.0, 5.0]
11	Horror	5.0	[Stand, The (1994), Slumber Party Massacre II (1987), Only Lovers Left Alive (2013), Vaniac Cop 2 (1990), 'Salem's Lot (2004)]	[5.0, 5.0, 5.0, 5.0, 5.0]
12	Romance	4.57	[Battle For Sevastopol (2015), My Sassy Girl (Yeogijeogin geunyeo) (2001), Continental Divide (1981), Woman Is a Woman, A (femme est une femme, Un...	[5.0, 5.0, 5.0, 5.0, 5.0]
13	Mystery	4.5	[The Adventures of Sherlock Holmes and Doctor Watson: King of Blackmalters (1980), 'Salem's Lot (2004), The Editor (2015), Galaxy of Terror (Quest)...	[5.0, 5.0, 5.0, 5.0, 5.0]
14	Mystery	4.5	[Colourful (Karafulu) (2010), Get Low (2009), Mother (Maeo) (2009), Scooby-Doo! Curse of the Lake Monster (2010), 'Salem's Lot (2004)]	[5.0, 5.0, 5.0, 5.0, 5.0]
15	War	4.1	[Che: Part One (2008), Eichmann (2007), Guy X (2005), Battle Royale 2: Requiem (Batoru rowaiaru II: Chinkonka) (2003), Human Condition III, The (Ni...	[5.0, 5.0, 5.0, 5.0, 5.0]
16	IMAX	4.12	[Happy Feet Two (2011), More (1998), Journey to the West: Conquering the Demons (Daai wa sai you chi Chui mo chun kei) (2013), Final Destination 5 ...]	[5.0, 5.0, 4.75, 4.5, 4.25]
17	War	4.44	[Max Manus (2008), Che: Part Two (2008), Eichmann (2007), Red Sorghum (Hong gao liang) (1987), Che: Part One (2008)]	[5.0, 5.0, 5.0, 5.0, 5.0]
18	Western	4.18	[Trinity and Sartana Are Coming (1972), Duel in the Sun (1946), 7 Faces of Dr. Lao (1964), Bone Tomahawk (2015), Ox-Bow Incident, The (1943)]	[5.0, 5.0, 5.0, 4.5, 4.5]
19	Film-Noir	3.33	[Rififi (Du rififi chez les hommes) (1955), Long Goodbye, The (1973), You Only Live Once (1937), 13 Tzmetti (2005), Sunset Blvd. (a.k.a. Sunset Bou...	[4.75, 4.67, 4.5, 4.5, 4.33]
20	IMAX	4.6	[Happy Feet Two (2011), More (1998), Journey to the West: Conquering the Demons (Daai wa sai you chi Chui mo chun kei) (2013), Final Destination 5 ...]	[5.0, 5.0, 4.75, 4.5, 4.25]
21	Western	4.0	[7 Faces of Dr. Lao (1964), Trinity and Sartana Are Coming (1972), Duel in the Sun (1946), Scalphunters, The (1968), Ox-Bow Incident, The (1943)]	[5.0, 5.0, 5.0, 4.5, 4.5]
22	Film-Noir	4.5	[Rififi (Du rififi chez les hommes) (1955), Long Goodbye, The (1973), 13 Tzmetti (2005), You Only Live Once (1937), Sunset Blvd. (a.k.a. Sunset Bou...	[4.75, 4.67, 4.5, 4.5, 4.33]
23	Mystery	4.67	[Get Low (2009), 7 Faces of Dr. Lao (1964), Scooby-Doo! Curse of the Lake Monster (2010), Adventures Of Sherlock Holmes And Dr. Watson: The Twentie...	[5.0, 5.0, 5.0, 5.0, 5.0]
24	Animation	4.25	[Scooby-Doo Goes Hollywood (1979), A Detective Story (2003), Karlson Returns (1970), Kung Fu Panda: Secrets of the Masters (2011), Winnie Pooh (1969)]	[5.0, 5.0, 5.0, 5.0, 5.0]
25	Romance	5.0	[Love and Pigeons (1985), Rain (2001), Open Hearts (Elsker dig for evigt) (2002), PK (2014), Saving Face (2004)]	[5.0, 5.0, 5.0, 5.0, 5.0]
26	Horror	4.0	['Salem's Lot (2004), The Girl with All the Gifts (2016), Buzzard (2015), Breed, The (2008), Only Lovers Left Alive (2013)]	[5.0, 5.0, 5.0, 5.0, 5.0]
27	Mystery	4.33	[The Editor (2015), 'Salem's Lot (2004), PK (2014), Colourful (Karafulu) (2010), Scooby-Doo! Abrecadabra-Doo (2010)]	[5.0, 5.0, 5.0, 5.0, 5.0]
28	IMAX	3.77	[More (1998), Happy Feet Two (2011), Journey to the West: Conquering the Demons (Daai wa sai you chi Chui mo chun kei) (2013), Final Destination 5 ...]	[5.0, 5.0, 4.75, 4.5, 4.25]
29	Film-Noir	5.0	[Rififi (Du rififi chez les hommes) (1955), Long Goodbye, The (1973), 13 Tzmetti (2005), You Only Live Once (1937), Sunset Blvd. (a.k.a. Sunset Bou...	[4.75, 4.67, 4.5, 4.5, 4.33]
30	War	5.0	[King of Hearts (1966), Battle For Sevastopol (2015), Red Sorghum (Hong gao liang) (1987), Guy X (2005), Eichmann (2007)]	[5.0, 5.0, 5.0, 5.0, 5.0]
31	War	5.0	[Che: Part One (2008), Come and See (Idi i smotri) (1985), Guy X (2005), Che: Part Two (2008), Red Sorghum (Hong gao liang) (1987)]	[5.0, 5.0, 5.0, 5.0, 5.0]
32	Crime	4.05	[Nother (Maeo) (2009), Ex Drummer (2007), Sonatine (Sonachine) (1993), Sherlock Holmes and Dr. Watson: Acquaintance (1979), Little Murders (1971)]	[5.0, 5.0, 5.0, 5.0, 5.0]
33	Film-Noir	5.0	[Rififi (Du rififi chez les hommes) (1955), Long Goodbye, The (1973), 13 Tzmetti (2005), You Only Live Once (1937), Sunset Blvd. (a.k.a. Sunset Bou...	[4.75, 4.67, 4.5, 4.5, 4.33]
34	Adventure	3.81	[The Love Bug (1997), Monster Squad, The (1987), SnowLake, the White Gorilla (2011), Delirium (2014), Prkiyucheniya Kapitana Vrungelya (1979)]	[5.0, 5.0, 5.0, 5.0, 5.0]