

	movieId		title		genres	userId	movieId	rating	timestamp
0	1		Toy Story (1995)	Adventure Animation Children Comedy Fantasy		0	1	1	4.0 964982703
1	2		Jumanji (1995)	Adventure Children Fantasy		1	1	3	4.0 964981247
2	3		Grumpier Old Men (1995)	Comedy Romance		2	1	6	4.0 964982224
3	4		Waiting to Exhale (1995)	Comedy Drama Romance		3	1	47	5.0 964983815
4	5		Father of the Bride Part II (1995)	Comedy		4	1	50	5.0 964982931
5	6		Heat (1995)	Action Crime Thriller		5	1	70	3.0 964982400
6	7		Sabrina (1995)	Comedy Romance		6	1	101	5.0 964980868
7	8		Tom and Huck (1995)	Adventure Children		7	1	110	4.0 964982176
8	9		Sudden Death (1995)	Action		8	1	151	5.0 964984041
9	10		GoldenEye (1995)	Action Adventure Thriller		9	1	157	5.0 964984100

Merge on MovieID

	userId	movieId	rating	timestamp		title		genres
0	1	1	4.0	964982703		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
1	5	1	4.0	847434962		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
2	7	1	4.5	1106635946		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
3	15	1	2.5	1510577970		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
4	17	1	4.5	1305696483		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
5	18	1	3.5	1455209816		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
6	19	1	4.0	965705637		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
7	21	1	3.5	1407618878		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
8	27	1	3.0	962685262		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
9	31	1	5.0	850466616		Toy Story (1995)		Adventure Animation Children Comedy Fantasy

Drop userID, movieID, rating, timestamp

		title		genres
0		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
1		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
2		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
3		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
4		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
5		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
6		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
7		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
8		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
9		Toy Story (1995)		Adventure Animation Children Comedy Fantasy

Keep unique

		title		genres
0		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
215		Grumpier Old Men (1995)		Comedy Romance
267		Heat (1995)		Action Crime Thriller
369		Seven (a.k.a. Se7en) (1995)		Mystery Thriller
572		Usual Suspects, The (1995)		Crime Mystery Thriller
776		From Dusk Till Dawn (1996)		Action Comedy Horror Thriller
831		Bottle Rocket (1996)		Adventure Comedy Crime Romance
854		Braveheart (1995)		Action Drama War
1091		Rob Roy (1995)		Action Drama Romance War
1135		Canadian Bacon (1995)		Comedy War

Brute force reindex

		title		genres
		Toy Story (1995)		Adventure Animation Children Comedy Fantasy
		Grumpier Old Men (1995)		Comedy Romance
		Heat (1995)		Action Crime Thriller
		Seven (a.k.a. Se7en) (1995)		Mystery Thriller
		Usual Suspects, The (1995)		Crime Mystery Thriller
		From Dusk Till Dawn (1996)		Action Comedy Horror Thriller
		Bottle Rocket (1996)		Adventure Comedy Crime Romance
		Braveheart (1995)		Action Drama War
		Rob Roy (1995)		Action Drama Romance War
		Canadian Bacon (1995)		Comedy War

.mean(rating)(title), count(title), output with title index

	title	rating	rating_counts
	"11'09""01 - September 11 (2002)"	4.000000	2
	'71 (2014)	4.000000	1
	'Hellboy': The Seeds of Creation (2004)	4.000000	1
	'Round Midnight (1986)	3.500000	2
	'Salem's Lot (2004)	5.000000	1
	'Til There Was You (1997)	4.000000	2
	'Tis the Season for Love (2015)	1.500000	1
	'burbs, The (1989)	3.176471	17
	'night Mother (1986)	3.000000	1
	(500) Days of Summer (2009)	3.666667	42

title	"11'09""01 - September 11 (2002)"	'71 (2014)	'Hellboy': The Seeds of Creation (2004)	'Round Midnight (1986)	\
userId					
1	NaN	NaN		NaN	NaN
2	NaN	NaN		NaN	NaN
3	NaN	NaN		NaN	NaN
4	NaN	NaN		NaN	NaN
5	NaN	NaN		NaN	NaN
6	NaN	NaN		NaN	NaN
7	NaN	NaN		NaN	NaN
8	NaN	NaN		NaN	NaN
9	NaN	NaN		NaN	NaN
10	NaN	NaN		NaN	NaN

corelation matrix

title	"11'09""01 - September 11 (2002)"	'71 (2014)
title		
"11'09""01 - September 11 (2002)"	NaN	NaN
'71 (2014)	NaN	NaN
'Hellboy': The Seeds of Creation (2004)	NaN	NaN
'Round Midnight (1986)	NaN	NaN
'Salem's Lot (2004)	NaN	NaN
'Til There Was You (1997)	NaN	NaN

query on user movie

User input -> Query

	correlation
title	
(500) Days of Summer (2009)	-0.944911
10 Things I Hate About You (1999)	-0.960769
101 Dalmatians (One Hundred and One Dalmatians)...	0.277350
12 Years a Slave (2013)	1.000000
13 Going on 30 (2004)	-0.866025
15 Minutes (2001)	-1.000000
16 Blocks (2006)	1.000000
2 Fast 2 Furious (Fast and the Furious 2, The) ...	0.960769
2001: A Space Odyssey (1968)	-0.174078
2012 (2009)	0.419314

join

	correlation	rating	rating_counts
title			
(500) Days of Summer (2009)	-0.944911	3.666667	42
10 Things I Hate About You (1999)	-0.960769	3.527778	54
101 Dalmatians (One Hundred and One Dalmatians)...	0.277350	3.431818	44
12 Years a Slave (2013)	1.000000	3.625000	16
13 Going on 30 (2004)	-0.866025	3.190476	21
15 Minutes (2001)	-1.000000	3.000000	10
16 Blocks (2006)	1.000000	3.000000	9
2 Fast 2 Furious (Fast and the Furious 2, The) ...	0.960769	2.605263	19
2001: A Space Odyssey (1968)	-0.174078	3.894495	109
2012 (2009)	0.419314	2.619048	21

join, sort so the correlation desc, rating desc, rating count desc

	correlation	rating	rating_counts		genres
title					
Stand by Me (1986)	1.0	4.005495	91		Adventure Drama
American President, The (1995)	1.0	3.671429	70		Comedy Drama Romance
Citizen Kane (1941)	1.0	4.043478	69		Drama Mystery
Legends of the Fall (1994)	1.0	3.397859	68		Drama Romance War Western
District 9 (2009)	1.0	3.776923	65		Mystery Sci-Fi Thriller
Vertigo (1958)	1.0	4.025000	60		Drama Mystery Romance Thriller