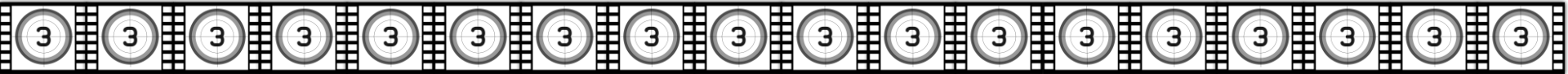


Technology, Culture & Movies

Katherine Potter



Dataset(s)

For this analysis, I will be using the IMDB Movie Dataset.



movieid	title	genres
1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
2	Jumanji (1995)	Adventure Children Fantasy
3	Grumpier Old Men (1995)	Comedy Romance
4	Waiting to Exhale (1995)	Comedy Drama Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action Crime Thriller
7	Sabrina (1995)	Comedy Romance
8	Tom and Huck (1995)	Adventure Children
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action Adventure Thriller
11	American President, The (1995)	Comedy Drama Romance
12	Dracula: Dead and Loving It (1995)	Comedy Horror
13	Balto (1995)	Adventure Animation Children
14	Nixon (1995)	Drama
15	Cutthroat Island (1995)	Action Adventure Romance
16	Casino (1995)	Crime Drama
17	Sense and Sensibility (1995)	Drama Romance
18	Four Rooms (1995)	Comedy
19	Ace Ventura: When Nature Calls (1995)	Comedy
20	Money Train (1995)	Action Comedy Crime Drama Thriller
21	Get Shorty (1995)	Comedy Crime Thriller
22	Copycat (1995)	Crime Drama Horror Mystery Thriller
23	Assassins (1995)	Action Crime Thriller
24	Powder (1995)	Drama Sci-Fi
25	Leaving Las Vegas (1995)	Drama Romance
26	Othello (1995)	Drama
27	Now and Then (1995)	Children Drama
28	Persuasion (1995)	Drama Romance
29	City of Lost Children, The (Cit�� des enfants perdus, La) (1995)	Adventure Drama Fantasy Mystery Sci-Fi
30	Shanghai Triad (Yao a yao dao waipo qiao) (1995)	Crime Drama
31	Dangerous Minds (1995)	Drama
32	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)	Mystery Sci-Fi Thriller
33	Wings of Courage (1995)	Adventure Romance IMAX
34	Babe (1995)	Children Drama
35	Carrington (1995)	Drama Romance
36	Dead Man Walking (1995)	Crime Drama

userid	movieid	rating	timestamp
	1	2	3.5
	1	29	3.5
	1	32	3.5
	1	47	3.5
	1	50	3.5
	1	112	3.5
	1	151	4
	1	223	4
	1	253	4
	1	260	4
	1	293	4
	1	296	4
	1	318	4
	1	337	3.5
	1	367	3.5
	1	541	4
	1	589	3.5
	1	593	3.5
	1	653	3
	1	919	3.5
	1	924	3.5
	1	1009	3.5
	1	1036	4
	1	1079	4
	1	1080	3.5
	1	1089	3.5
	1	1090	4
	1	1097	4
	1	1136	3.5
	1	1193	3.5
	1	1196	4.5
	1	1198	4.5
	1	1200	4
	1	1201	3
	1	1208	3.5
	1	1214	4

Motivation

Technology has changed every aspect of human life. I want to know how it has affected moviegoer preferences and which genres people enjoy the most.



Research Question(s)

Which genre has the most consistently high ratings?

- Are the genres that were popular in the early 1900s still popular today?
- Has the horror genre become more or less popular as people have come to love being scared?
- Let's find out!

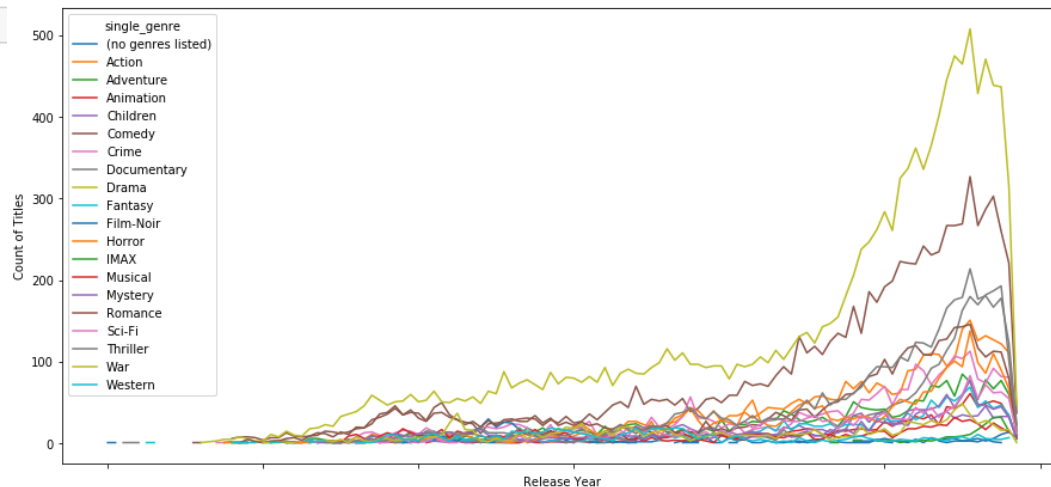


Findings : Digging Into the Data

- Over the given timeframe, 1891 to 2015, there were by far, more movies that fell into the drama genre than any of the others
- Overall, 49% of the movies in the dataset fell into the drama category
- Comedy came in second place followed by thrillers, romance, action, horror, & crime
 - Comedy - 30%
 - Romance - 15%
 - Crime - 11%
 - Thrillers - 15%
 - Action - 13%
 - Horror - 10%

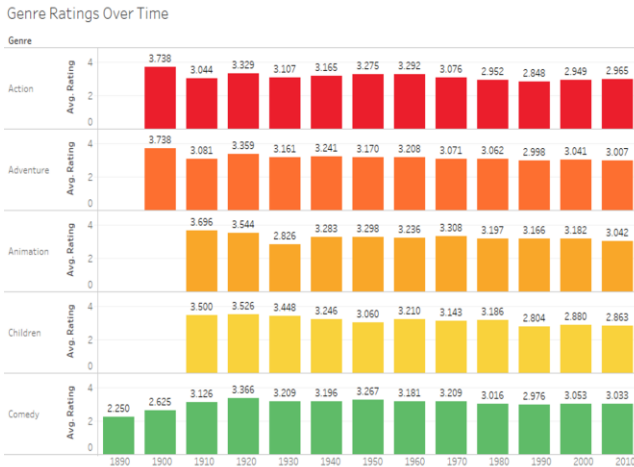
```
In [238]: exploded_genres['single_genre'].value_counts()
```

```
Out[238]: Drama      13340
Comedy      8371
Thriller    4177
Romance     4127
Action      3519
Crime       2938
Horror      2611
Documentary 2467
Adventure   2329
Sci-Fi      1741
Mystery     1513
Fantasy     1411
War         1194
Children    1139
Musical     1035
Animation   1027
Western     676
Film-Noir   330
(no genres listed) 237
IMAX        196
Name: single_genre, dtype: int64
```

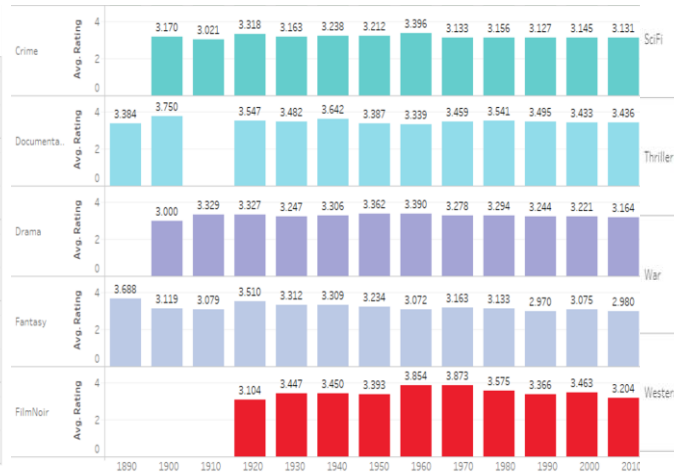


Findings : Genres by Decade

- The following visuals highlight the evolution of genre ratings by decade
- Most surprising to me were the ratings trends for horror and thriller genres
- Both horror & thriller movies received higher ratings in the early 1900's than in recent years
 - Perhaps this is because people have far more access to gore and scary movies and have become jaded or desensitized as a result



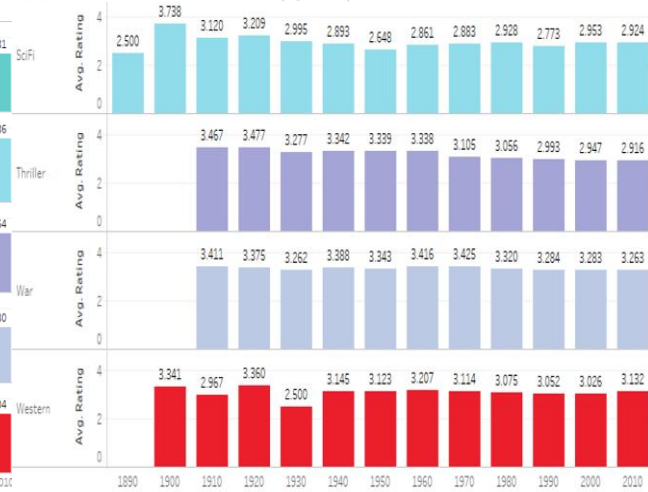
Average of Rating for each Calculation1 broken down by Genre. Color shows details about Genre. The data is filtered on Exclusions (Genre,Release Year), which keeps 1,811 members. The view is filtered on Genre, which excludes (no genres listed).



Average of Rating for each Calculation1 broken down by Genre. Color shows details about Genre. The data is filtered on Exclusions (Genre,Release Year), which keeps 1,811 members. The view is filtered on Genre, which excludes (no genres listed).



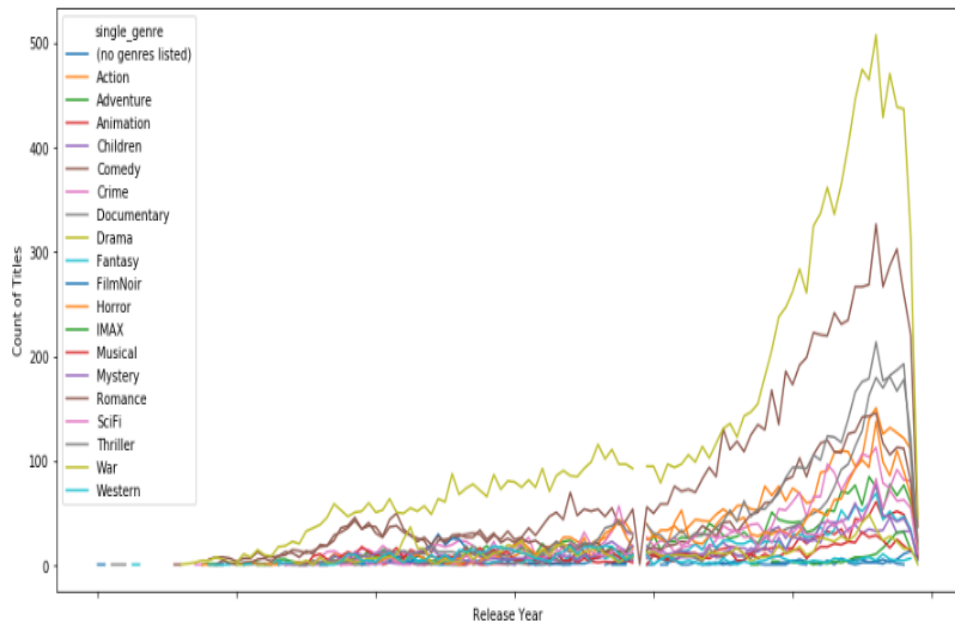
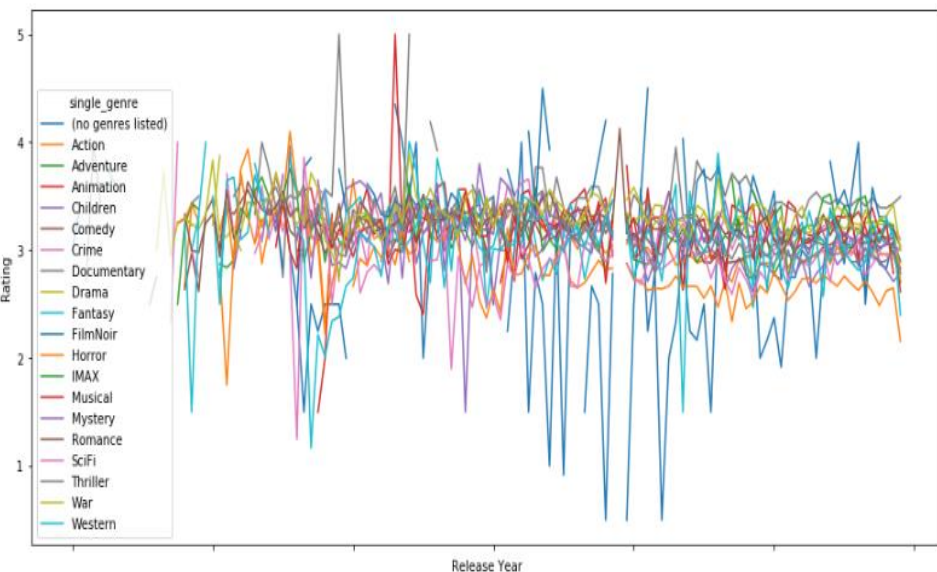
Average of Rating for each Calculation1 broken down by Genre. Color shows details about Genre. The data is filtered on Exclusions (Genre,Release Year), which keeps 1,811 members. The view is filtered on Genre, which excludes (no genres listed).



Average of Rating for each Calculation1 broken down by Genre. Color shows details about Genre. The data is filtered on Exclusions (Genre,Release Year), which keeps 1,811 members. The view is filtered on Genre, which excludes (no genres listed).

Findings : The Trend of Genres Over Time

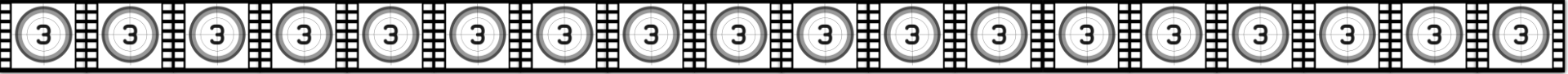
- Below are visuals for the trend of ratings by genre over time
- You can see the highs and lows of the genres that support the findings of this analysis
- Note: further analysis would be necessary in order to draw conclusions
 - Perhaps on a smaller, more comprehensive dataset
 - This analysis could be biased based on the dataset as the number of movies and ratings varied by year and by genre and could have skewed the end result



Findings : Which genre has the most consistently high ratings?

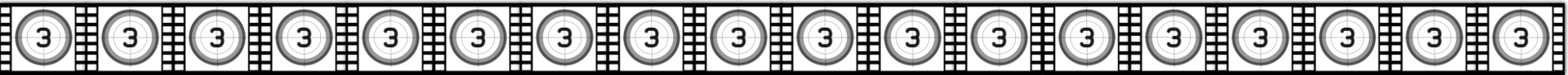
- Further analysis showed that, in fact, the most highly rated genres have been changing over the years
- Documentaries were the highest rated movies in the 1900's, 1920's, 1930's, 1940's, 1990's, & 2010's
- FilmNoir were the highest rated movies in the 1950's, 1960's, 1970's, 1980's, 2000's, and was the highest rated genre overall in this dataset
- **In conclusion, this analysis reveals that FilmNoir & Documentaries had the most consistently high ratings, as they were the highest rated genres throughout 6 decades each**

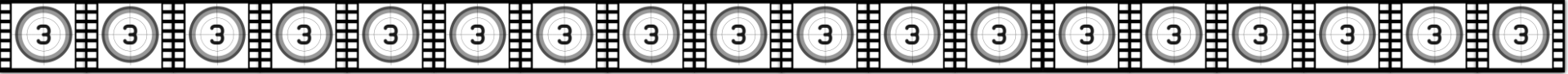
Genre	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	Grand Total
FilmNoir				3.104166667	3.446720621	3.450450045	3.392622459	3.853985616	3.872504913	3.575313893	3.366140128	3.462553314	3.203604825	3.49761864
Documentary	3.383928571	3.75		3.547138625	3.482013782	3.642274864	3.387439556	3.338617872	3.459452252	3.540907243	3.49526602	3.433193751	3.435702889	3.473106897
War			3.411124953	3.374599494	3.261531057	3.387537046	3.343441796	3.415500775	3.425449511	3.320023574	3.284372605	3.282907302	3.262582847	3.340366984
IMAX										3.409863946	3.30881179	3.307051763	3.283251156	3.311174326
Drama			3	3.328933187	3.327284706	3.247496562	3.306491285	3.362097634	3.38987726	3.277754946	3.294172478	3.221400772	3.163619508	3.289088327
Romance	2.928571429	3.166666667	3.150890975	3.402375389	3.216120397	3.317497241	3.297383179	3.374161667	3.233500511	3.204587879	3.158984813	3.176472793	3.145867253	3.24698679
Animation			3.696428571	3.544334627	2.825800742	3.283413224	3.297722504	3.236238423	3.307940595	3.196994012	3.165848849	3.182421238	3.041512805	3.208363655
Crime		3.170454545	3.020833333	3.317975679	3.162668464	3.237805936	3.211584043	3.396161405	3.133242311	3.156426563	3.127132445	3.144717909	3.131185002	3.195747765
Musical	3.428571429			3.154668261	3.185348829	3.164818708	3.241573531	3.184543825	3.229900182	3.088804327	3.186130512	3.208294991	3.172248053	3.187055792
Thriller			3.466666667	3.476593351	3.277192092	3.341630374	3.338797135	3.337812034	3.104936645	3.055534389	2.993307349	2.947332037	2.916340164	3.186662614
Fantasy		3.6875	3.119094488	3.079268293	3.509754704	3.312285099	3.309360104	3.233788898	3.07151026	3.162997137	3.133089284	2.969701639	3.075167325	3.180750775
Mystery				3	3.245600712	3.070007836	3.17436078	3.302917221	3.349101703	3.170872887	3.134864358	3.216682653	3.089158278	3.178327911
Comedy	2.25	2.625	3.125830172	3.365662403	3.208786959	3.196491772	3.26671029	3.180587802	3.292173204	3.015806302	2.976070121	3.053390503	3.032947308	3.144181886
Adventure		3.738188976	3.081389857	3.359484587	3.161462756	3.240742728	3.169799693	3.207996667	3.07075077	3.061934488	2.998026564	3.040713777	3.007443122	3.140185464
Children			3.5	3.52564158	3.44766178	3.245662043	3.060016317	3.209917256	3.142687837	3.185531231	2.803968866	2.879520513	2.862714027	3.111033126
Action		3.738188976	3.043554007	3.329098111	3.106945809	3.165151872	3.275112325	3.292313225	3.076395272	2.951902636	2.848012938	2.948856723	2.964673141	3.105330029
Western		3.340909091	2.966666667	3.360095594	2.499921226	3.144614407	3.123354613	3.207299777	3.113703776	3.074772942	3.052073122	3.026379792	3.131587551	3.05264432
SciFi	2.5	3.738188976	3.120238095	3.209456829	2.995495123	2.89344286	2.648012734	2.860512435	2.883210917	2.927636512	2.772646672	2.952826065	2.923630179	2.899314992
Horror			3.324642857	3.275839339	2.931172439	3.055265954	2.842854915	2.90529174	2.766852349	2.640825774	2.601935504	2.702504875	2.541197377	2.854663683
(no genres listed)	3.569444444			3.5	3.25	3.205555556	3.407142857	3.15	2.445833333	3.085482804	2.23	2.612847222	3.034642857	2.90734127
Grand Total	3.278521825	3.28937434	3.201497526	3.360675552	3.103720621	3.261743123	3.205729637	3.210519829	3.138789027	3.089868069	3.05663564	3.103083278	3.050963242	3.157802661



Acknowledgements

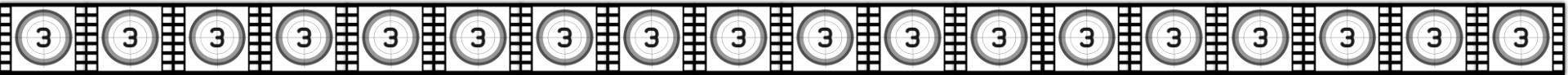
Special Thanks to the data scientists on my team who were gracious enough to listen to my project idea and hash it out with me as well as to lend a pair of fresh eyes and a bit of guidance whenever I asked.





References

I did not use any outside sources throughout this analysis. I relied on the movies and ratings datasets as well as a combination of the two in order to answer my questions.



```
In [194]: import pandas as pd
import numpy as np
import os.path
import re
import matplotlib.pyplot as plt
```

```
In [195]: #MOVIELENS set to path of movielens folder
MOVIELENS = "/Users/Katherine/Documents/PYTHON_FOR_DATA_SCIENCE_CLASS/Week 6 M
ini Project/ml-20m"

movies = pd.read_csv(os.path.join(MOVIELENS, 'movies.csv'), sep=',')
ratings = pd.read_csv(os.path.join(MOVIELENS, 'ratings.csv'), sep=',')
```

```
In [272]: # movies  
movies2 = movies[0:]  
movies2
```

Out[272]:

	movieid	title	genres
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
1	2	Jumanji (1995)	Adventure Children Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama Romance
4	5	Father of the Bride Part II (1995)	Comedy
5	6	Heat (1995)	Action Crime Thriller
6	7	Sabrina (1995)	Comedy Romance
7	8	Tom and Huck (1995)	Adventure Children
8	9	Sudden Death (1995)	Action
9	10	GoldenEye (1995)	Action Adventure Thriller
10	11	American President, The (1995)	Comedy Drama Romance
11	12	Dracula: Dead and Loving It (1995)	Comedy Horror
12	13	Balto (1995)	Adventure Animation Children
13	14	Nixon (1995)	Drama
14	15	Cutthroat Island (1995)	Action Adventure Romance
15	16	Casino (1995)	Crime Drama
16	17	Sense and Sensibility (1995)	Drama Romance
17	18	Four Rooms (1995)	Comedy
18	19	Ace Ventura: When Nature Calls (1995)	Comedy
19	20	Money Train (1995)	Action Comedy Crime Drama Thriller
20	21	Get Shorty (1995)	Comedy Crime Thriller
21	22	Copycat (1995)	Crime Drama Horror Mystery Thriller
22	23	Assassins (1995)	Action Crime Thriller
23	24	Powder (1995)	Drama SciFi
24	25	Leaving Las Vegas (1995)	Drama Romance
25	26	Othello (1995)	Drama
26	27	Now and Then (1995)	Children Drama
27	28	Persuasion (1995)	Drama Romance
28	29	City of Lost Children, The (Cité des enfants p...	Adventure Drama Fantasy Mystery SciFi
29	30	Shanghai Triad (Yao a yao yao dao waipo qiao) ...	Crime Drama
...
27248	131146	Werner Volles Rooäää (1999)	Animation Comedy
27249	131148	What A Man (2011)	Comedy Romance
27250	131150	7 Dwarves: The Forest Is Not Enough (2006)	Comedy

	movieid	title	genres
27251	131152	The Fat Spy (1966)	Comedy
27252	131154	Die Bademeister – Weiber, saufen, Leben retten...	Comedy
27253	131156	Ants in the Pants 2 (2002)	Comedy
27254	131158	Manta, Manta (1991)	Comedy
27255	131160	Oscar and the Lady in Pink (2009)	Drama
27256	131162	Por un puñado de besos (2014)	Drama Romance
27257	131164	Vietnam in HD (2011)	War
27258	131166	WWII IN HD (2009)	(no genres listed)
27259	131168	Phoenix (2014)	Drama
27260	131170	Parallels (2015)	SciFi
27261	131172	Closed Curtain (2013)	(no genres listed)
27262	131174	Gentlemen (2014)	Drama Romance Thriller
27263	131176	A Second Chance (2014)	Drama
27264	131180	Dead Rising: Watchtower (2015)	Action Horror Thriller
27265	131231	Standby (2014)	Comedy Romance
27266	131237	What Men Talk About (2010)	Comedy
27267	131239	Three Quarter Moon (2011)	Comedy Drama
27268	131241	Ants in the Pants (2000)	Comedy Romance
27269	131243	Werner Gekotzt wird später (2003)	Animation Comedy
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy
27271	131250	No More School (2000)	Comedy
27272	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror
27273	131254	Kein Bund für's Leben (2007)	Comedy
27274	131256	Feuer, Eis & Dosenbier (2002)	Comedy
27275	131258	The Pirates (2014)	Adventure
27276	131260	Rentun Ruusu (2001)	(no genres listed)
27277	131262	Innocence (2014)	Adventure Fantasy Horror

27278 rows × 3 columns


```
In [273]: ratings2 = ratings[0:]  
groupedratings = ratings2.groupby('movieId')['rating'].mean()  
groupedratings
```

```
Out[273]: movieId
1          3.921240
2          3.211977
3          3.151040
4          2.861393
5          3.064592
6          3.834930
7          3.366484
8          3.142049
9          3.004924
10         3.430029
11         3.667713
12         2.619766
13         3.272416
14         3.432082
15         2.721993
16         3.787455
17         3.968573
18         3.373631
19         2.607412
20         2.880754
21         3.581689
22         3.319400
23         3.148235
24         3.199849
25         3.689510
26         3.628857
27         3.413520
28         4.057546
29         3.952230
30         3.633880
...
131146     4.000000
131148     4.000000
131150     4.000000
131152     0.500000
131154     3.500000
131156     4.000000
131158     4.000000
131160     4.000000
131162     2.000000
131164     4.000000
131166     4.000000
131168     3.500000
131170     3.500000
131172     1.000000
131174     3.500000
131176     4.500000
131180     2.500000
131231     3.500000
131237     3.000000
131239     4.000000
131241     4.000000
131243     4.000000
131248     4.000000
131250     4.000000
131252     4.000000
```

131254	4.000000
131256	4.000000
131258	2.500000
131260	3.000000
131262	4.000000

Name: rating, Length: 26744, dtype: float64

```
In [274]: movies2 = movies[0:]
movies2['title'].str.split(r'\([^\\d]*(\\d+)[^\\d]*\\)')
title_arr = movies2['title'].str.split(r'\([^\\d]*(\\d+)[^\\d]*\\)')
has_year = movies2['title'].apply(lambda x: bool(re.search('\\([^\\d]*(\\d+)[^\\d]*\\)',x)))

filtered_years = movies2.loc[has_year]
filtered_years['movie_title'] = filtered_years.title.apply(lambda x: re.split(
'\\([^\\d]*(\\d+)[^\\d]*\\)',x)[0])
filtered_years['release_year'] = filtered_years.title.apply(lambda x: re.split(
'\\([^\\d]*(\\d+)[^\\d]*\\)',x)[-2])
filtered_years
# filtered_years['title'].nunique()
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:7: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
import sys
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

Out[274]:

	movieid	title	genres	movie_title	release_year
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
1	2	Jumanji (1995)	Adventure Children Fantasy	Jumanji	1995
2	3	Grumpier Old Men (1995)	Comedy Romance	Grumpier Old Men	1995
3	4	Waiting to Exhale (1995)	Comedy Drama Romance	Waiting to Exhale	1995
4	5	Father of the Bride Part II (1995)	Comedy	Father of the Bride Part II	1995
5	6	Heat (1995)	Action Crime Thriller	Heat	1995
6	7	Sabrina (1995)	Comedy Romance	Sabrina	1995
7	8	Tom and Huck (1995)	Adventure Children	Tom and Huck	1995
8	9	Sudden Death (1995)	Action	Sudden Death	1995
9	10	GoldenEye (1995)	Action Adventure Thriller	GoldenEye	1995
10	11	American President, The (1995)	Comedy Drama Romance	American President, The	1995
11	12	Dracula: Dead and Loving It (1995)	Comedy Horror	Dracula: Dead and Loving It	1995
12	13	Balto (1995)	Adventure Animation Children	Balto	1995
13	14	Nixon (1995)	Drama	Nixon	1995
14	15	Cutthroat Island (1995)	Action Adventure Romance	Cutthroat Island	1995
15	16	Casino (1995)	Crime Drama	Casino	1995
16	17	Sense and Sensibility (1995)	Drama Romance	Sense and Sensibility	1995
17	18	Four Rooms (1995)	Comedy	Four Rooms	1995
18	19	Ace Ventura: When Nature Calls (1995)	Comedy	Ace Ventura: When Nature Calls	1995

	movieid	title	genres	movie_title	release_year
19	20	Money Train (1995)	Action Comedy Crime Drama Thriller	Money Train	1995
20	21	Get Shorty (1995)	Comedy Crime Thriller	Get Shorty	1995
21	22	Copycat (1995)	Crime Drama Horror Mystery Thriller	Copycat	1995
22	23	Assassins (1995)	Action Crime Thriller	Assassins	1995
23	24	Powder (1995)	Drama SciFi	Powder	1995
24	25	Leaving Las Vegas (1995)	Drama Romance	Leaving Las Vegas	1995
25	26	Othello (1995)	Drama	Othello	1995
26	27	Now and Then (1995)	Children Drama	Now and Then	1995
27	28	Persuasion (1995)	Drama Romance	Persuasion	1995
28	29	City of Lost Children, The (Cité des enfants p...	Adventure Drama Fantasy Mystery SciFi	City of Lost Children, The	1995
29	30	Shanghai Triad (Yao a yao dao waipo qiao) ...	Crime Drama	Shanghai Triad	1995
...
27248	131146	Werner Volles Rooäää (1999)	Animation Comedy	Werner Volles Rooäää	1999
27249	131148	What A Man (2011)	Comedy Romance	What A Man	2011
27250	131150	7 Dwarves: The Forest Is Not Enough (2006)	Comedy	7 Dwarves: The Forest Is Not Enough	2006
27251	131152	The Fat Spy (1966)	Comedy	The Fat Spy	1966
27252	131154	Die Bademeister – Weiber, saufen, Leben retten...	Comedy	Die Bademeister – Weiber, saufen, Leben retten	1995
27253	131156	Ants in the Pants 2 (2002)	Comedy	Ants in the Pants 2	2002

	movieid	title	genres	movie_title	release_year
27254	131158	Manta, Manta (1991)	Comedy	Manta, Manta	1991
27255	131160	Oscar and the Lady in Pink (2009)	Drama	Oscar and the Lady in Pink	2009
27256	131162	Por un puñado de besos (2014)	Drama Romance	Por un puñado de besos	2014
27257	131164	Vietnam in HD (2011)	War	Vietnam in HD	2011
27258	131166	WWII IN HD (2009)	(no genres listed)	WWII IN HD	2009
27259	131168	Phoenix (2014)	Drama	Phoenix	2014
27260	131170	Parallels (2015)	SciFi	Parallels	2015
27261	131172	Closed Curtain (2013)	(no genres listed)	Closed Curtain	2013
27262	131174	Gentlemen (2014)	Drama Romance Thriller	Gentlemen	2014
27263	131176	A Second Chance (2014)	Drama	A Second Chance	2014
27264	131180	Dead Rising: Watchtower (2015)	Action Horror Thriller	Dead Rising: Watchtower	2015
27265	131231	Standby (2014)	Comedy Romance	Standby	2014
27266	131237	What Men Talk About (2010)	Comedy	What Men Talk About	2010
27267	131239	Three Quarter Moon (2011)	Comedy Drama	Three Quarter Moon	2011
27268	131241	Ants in the Pants (2000)	Comedy Romance	Ants in the Pants	2000
27269	131243	Werner Gekotzt wird später (2003)	Animation Comedy	Werner Gekotzt wird später	2003
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006
27271	131250	No More School (2000)	Comedy	No More School	2000

	movieid	title	genres	movie_title	release_year
27272	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	Forklift Driver Klaus: The First Day on the Job	2007
27273	131254	Kein Bund für's Leben (2007)	Comedy	Kein Bund für's Leben	2007
27274	131256	Feuer, Eis & Dosenbier (2002)	Comedy	Feuer, Eis & Dosenbier	2002
27275	131258	The Pirates (2014)	Adventure	The Pirates	2014
27276	131260	Rentun Ruusu (2001)	(no genres listed)	Rentun Ruusu	2001
27277	131262	Innocence (2014)	Adventure Fantasy Horror	Innocence	2014

27259 rows × 5 columns



```
In [275]: # #filtered_movies = movies.loc[has_year]
filtered_genres = filtered_years[0:]
filtered_genres['genre'] = filtered_genres.genres.str.split('|')
filtered_genres['genre_len'] = filtered_genres['genre'].apply(len)

exploded_genres = filtered_genres.loc[filtered_genres.index.repeat(filtered_genres['genre_len'])]

# filtered_genres = filtered_genres.reset_index(drop=True)
# # print (filtered_genres)

exploded_genres['genre_index'] = exploded_genres.groupby(level=0)['genre'].cumcount().transform(lambda x:x)
exploded_genres['single_genre'] = exploded_genres.apply(lambda x: x.genre[x.genre_index],axis=1)
exploded_genres
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
This is separate from the ipykernel package so we can avoid doing imports until
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:4: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
after removing the cwd from sys.path.
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:12: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
if sys.path[0] == '':
```

```
C:\Users\Katherine\Anaconda3\lib\site-packages\ipykernel_launcher.py:13: SettingWithCopyWarning:
```

```
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: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
del sys.path[0]
```

Out[275]:

	movieid	title	genres	movie_title	release_year
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy Story	1995
1	2	Jumanji (1995)	Adventure Children Fantasy	Jumanji	1995
1	2	Jumanji (1995)	Adventure Children Fantasy	Jumanji	1995
1	2	Jumanji (1995)	Adventure Children Fantasy	Jumanji	1995
2	3	Grumpier Old Men (1995)	Comedy Romance	Grumpier Old Men	1995
2	3	Grumpier Old Men (1995)	Comedy Romance	Grumpier Old Men	1995
3	4	Waiting to Exhale (1995)	Comedy Drama Romance	Waiting to Exhale	1995
3	4	Waiting to Exhale (1995)	Comedy Drama Romance	Waiting to Exhale	1995
3	4	Waiting to Exhale (1995)	Comedy Drama Romance	Waiting to Exhale	1995
4	5	Father of the Bride Part II (1995)	Comedy	Father of the Bride Part II	1995

	movieid	title	genres	movie_title	release_year
5	6	Heat (1995)	Action Crime Thriller	Heat	1995
5	6	Heat (1995)	Action Crime Thriller	Heat	1995
5	6	Heat (1995)	Action Crime Thriller	Heat	1995
6	7	Sabrina (1995)	Comedy Romance	Sabrina	1995
6	7	Sabrina (1995)	Comedy Romance	Sabrina	1995
7	8	Tom and Huck (1995)	Adventure Children	Tom and Huck	1995
7	8	Tom and Huck (1995)	Adventure Children	Tom and Huck	1995
8	9	Sudden Death (1995)	Action	Sudden Death	1995
9	10	GoldenEye (1995)	Action Adventure Thriller	GoldenEye	1995
9	10	GoldenEye (1995)	Action Adventure Thriller	GoldenEye	1995
9	10	GoldenEye (1995)	Action Adventure Thriller	GoldenEye	1995
10	11	American President, The (1995)	Comedy Drama Romance	American President, The	1995
10	11	American President, The (1995)	Comedy Drama Romance	American President, The	1995
10	11	American President, The (1995)	Comedy Drama Romance	American President, The	1995
11	12	Dracula: Dead and Loving It (1995)	Comedy Horror	Dracula: Dead and Loving It	1995
11	12	Dracula: Dead and Loving It (1995)	Comedy Horror	Dracula: Dead and Loving It	1995
...
27262	131174	Gentlemen (2014)	Drama Romance Thriller	Gentlemen	2014

	movieid	title	genres	movie_title	release_year
27262	131174	Gentlemen (2014)	Drama Romance Thriller	Gentlemen	2014
27263	131176	A Second Chance (2014)	Drama	A Second Chance	2014
27264	131180	Dead Rising: Watchtower (2015)	Action Horror Thriller	Dead Rising: Watchtower	2015
27264	131180	Dead Rising: Watchtower (2015)	Action Horror Thriller	Dead Rising: Watchtower	2015
27264	131180	Dead Rising: Watchtower (2015)	Action Horror Thriller	Dead Rising: Watchtower	2015
27265	131231	Standby (2014)	Comedy Romance	Standby	2014
27265	131231	Standby (2014)	Comedy Romance	Standby	2014
27266	131237	What Men Talk About (2010)	Comedy	What Men Talk About	2010
27267	131239	Three Quarter Moon (2011)	Comedy Drama	Three Quarter Moon	2011
27267	131239	Three Quarter Moon (2011)	Comedy Drama	Three Quarter Moon	2011
27268	131241	Ants in the Pants (2000)	Comedy Romance	Ants in the Pants	2000
27268	131241	Ants in the Pants (2000)	Comedy Romance	Ants in the Pants	2000
27269	131243	Werner Gekotzt wird später (2003)	Animation Comedy	Werner Gekotzt wird später	2003
27269	131243	Werner Gekotzt wird später (2003)	Animation Comedy	Werner Gekotzt wird später	2003
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006

	movieid	title	genres	movie_title	release_year
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006
27270	131248	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Brother Bear 2	2006
27271	131250	No More School (2000)	Comedy	No More School	2000
27272	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	Forklift Driver Klaus: The First Day on the Job	2001
27272	131252	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	Forklift Driver Klaus: The First Day on the Job	2001
27273	131254	Kein Bund für's Leben (2007)	Comedy	Kein Bund für's Leben	2007
27274	131256	Feuer, Eis & Dosenbier (2002)	Comedy	Feuer, Eis & Dosenbier	2002
27275	131258	The Pirates (2014)	Adventure	The Pirates	2014
27276	131260	Rentun Ruusu (2001)	(no genres listed)	Rentun Ruusu	2001
27277	131262	Innocence (2014)	Adventure Fantasy Horror	Innocence	2014
27277	131262	Innocence (2014)	Adventure Fantasy Horror	Innocence	2014
27277	131262	Innocence (2014)	Adventure Fantasy Horror	Innocence	2014

54379 rows × 9 columns



```
In [278]: df1 = pd.DataFrame(groupedratings)
df1['index'] = [i for i in range(len(df1))]
df2 = pd.DataFrame(exploded_genres)
df2['index'] = [i for i in range(len(df2))]

movieratings = df1.merge(df2, on = 'movieId')
movieratings
```

Out[278]:

	movieid	rating	index_x	title	genres	movie
0	1	3.921240	0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy
1	1	3.921240	0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy
2	1	3.921240	0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy
3	1	3.921240	0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy
4	1	3.921240	0	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	Toy
5	2	3.211977	1	Jumanji (1995)	Adventure Children Fantasy	Ju
6	2	3.211977	1	Jumanji (1995)	Adventure Children Fantasy	Ju
7	2	3.211977	1	Jumanji (1995)	Adventure Children Fantasy	Ju
8	3	3.151040	2	Grumpier Old Men (1995)	Comedy Romance	Grui Old
9	3	3.151040	2	Grumpier Old Men (1995)	Comedy Romance	Grui Old
10	4	2.861393	3	Waiting to Exhale (1995)	Comedy Drama Romance	Wait E
11	4	2.861393	3	Waiting to Exhale (1995)	Comedy Drama Romance	Wait E
12	4	2.861393	3	Waiting to Exhale (1995)	Comedy Drama Romance	Wait E
13	5	3.064592	4	Father of the Bride Part II (1995)	Comedy	Fatl the F

	movieid	rating	index_x	title	genres	movie
14	6	3.834930	5	Heat (1995)	Action Crime Thriller	
15	6	3.834930	5	Heat (1995)	Action Crime Thriller	
16	6	3.834930	5	Heat (1995)	Action Crime Thriller	
17	7	3.366484	6	Sabrina (1995)	Comedy Romance	Sa
18	7	3.366484	6	Sabrina (1995)	Comedy Romance	Sa
19	8	3.142049	7	Tom and Huck (1995)	Adventure Children	Tom
20	8	3.142049	7	Tom and Huck (1995)	Adventure Children	Tom
21	9	3.004924	8	Sudden Death (1995)	Action	Su [
22	10	3.430029	9	GoldenEye (1995)	Action Adventure Thriller	Golde
23	10	3.430029	9	GoldenEye (1995)	Action Adventure Thriller	Golde
24	10	3.430029	9	GoldenEye (1995)	Action Adventure Thriller	Golde
25	11	3.667713	10	American President, The (1995)	Comedy Drama Romance	Ame Pres
26	11	3.667713	10	American President, The (1995)	Comedy Drama Romance	Ame Pres
27	11	3.667713	10	American President, The (1995)	Comedy Drama Romance	Ame Pres
28	12	2.619766	11	Dracula: Dead and Loving It (1995)	Comedy Horror	Dræ Dear Lov
29	12	2.619766	11	Dracula: Dead and Loving It (1995)	Comedy Horror	Dræ Dear Lov
...	
53362	131174	3.500000	26728	Gentlemen (2014)	Drama Romance Thriller	Gentle

	movieid	rating	index_x	title	genres	movie
53363	131174	3.500000	26728	Gentlemen (2014)	Drama Romance Thriller	Gentl
53364	131176	4.500000	26729	A Second Chance (2014)	Drama	A Se Cr
53365	131180	2.500000	26730	Dead Rising: Watchtower (2015)	Action Horror Thriller	R Watchi
53366	131180	2.500000	26730	Dead Rising: Watchtower (2015)	Action Horror Thriller	R Watchi
53367	131180	2.500000	26730	Dead Rising: Watchtower (2015)	Action Horror Thriller	R Watchi
53368	131231	3.500000	26731	Standby (2014)	Comedy Romance	Sta
53369	131231	3.500000	26731	Standby (2014)	Comedy Romance	Sta
53370	131237	3.000000	26732	What Men Talk About (2010)	Comedy	What Talk /
53371	131239	4.000000	26733	Three Quarter Moon (2011)	Comedy Drama	Qu I
53372	131239	4.000000	26733	Three Quarter Moon (2011)	Comedy Drama	Qu I
53373	131241	4.000000	26734	Ants in the Pants (2000)	Comedy Romance	Ants I
53374	131241	4.000000	26734	Ants in the Pants (2000)	Comedy Romance	Ants I
53375	131243	4.000000	26735	Werner Gekotzt wird später (2003)	Animation Comedy	W Ge wird s
53376	131243	4.000000	26735	Werner Gekotzt wird später (2003)	Animation Comedy	W Ge wird s
53377	131248	4.000000	26736	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Br B

	movieid	rating	index_x	title	genres	movie
53378	131248	4.000000	26736	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Br B
53379	131248	4.000000	26736	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Br B
53380	131248	4.000000	26736	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Br B
53381	131248	4.000000	26736	Brother Bear 2 (2006)	Adventure Animation Children Comedy Fantasy	Br B
53382	131250	4.000000	26737	No More School (2000)	Comedy	No S
53383	131252	4.000000	26738	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	F [Klaus Firs on th
53384	131252	4.000000	26738	Forklift Driver Klaus: The First Day on the Jo...	Comedy Horror	F [Klaus Firs on th
53385	131254	4.000000	26739	Kein Bund für's Leben (2007)	Comedy	Kein für's L
53386	131256	4.000000	26740	Feuer, Eis & Dosenbier (2002)	Comedy	Feue Dose
53387	131258	2.500000	26741	The Pirates (2014)	Adventure	The P
53388	131260	3.000000	26742	Rentun Ruusu (2001)	(no genres listed)	Ri F
53389	131262	4.000000	26743	Innocence (2014)	Adventure Fantasy Horror	Innoc
53390	131262	4.000000	26743	Innocence (2014)	Adventure Fantasy Horror	Innoc
53391	131262	4.000000	26743	Innocence (2014)	Adventure Fantasy Horror	Innoc

53392 rows × 12 columns



```
In [58]: exploded_genres.groupby('release_year')['single_genre'].value_counts()  
# exploded_genres[2] = 'count'
```

```

Out[58]: release_year  single_genre
1891      (no genres listed)      1
1893      (no genres listed)      1
1894      Documentary            1
          Musical                1
1895      Comedy                1
          Documentary            1
1896      Documentary            1
          Romance                1
1898      (no genres listed)      3
          Fantasy                1
          SciFi                 1
1899      Fantasy                1
1900      Romance                1
1901      Documentary            1
1902      Action                1
          Adventure             1
          Fantasy                1
          SciFi                 1
1903      Crime                 1
          Western                1
1905      Comedy                1
          Fantasy                1
1909      Comedy                1
          Crime                 1
          Documentary            1
          Drama                 1
1910      Drama                 2
          SciFi                 2
          Horror                 1
1912      Drama                 3
          ..
2014      Crime                 81
          Romance                80
          Horror                 67
          Adventure             60
          SciFi                 55
          Fantasy                35
          Animation             33
          Children               33
          Mystery                32
          War                    17
          IMAX                   15
          (no genres listed)     14
          Musical                14
          Western                7
2015      Drama                 46
          Comedy                37
          Thriller               23
          Action                 21
          Adventure             16
          Documentary            15
          SciFi                 12
          Crime                 10
          Children               8
          Horror                 8
          Romance                8

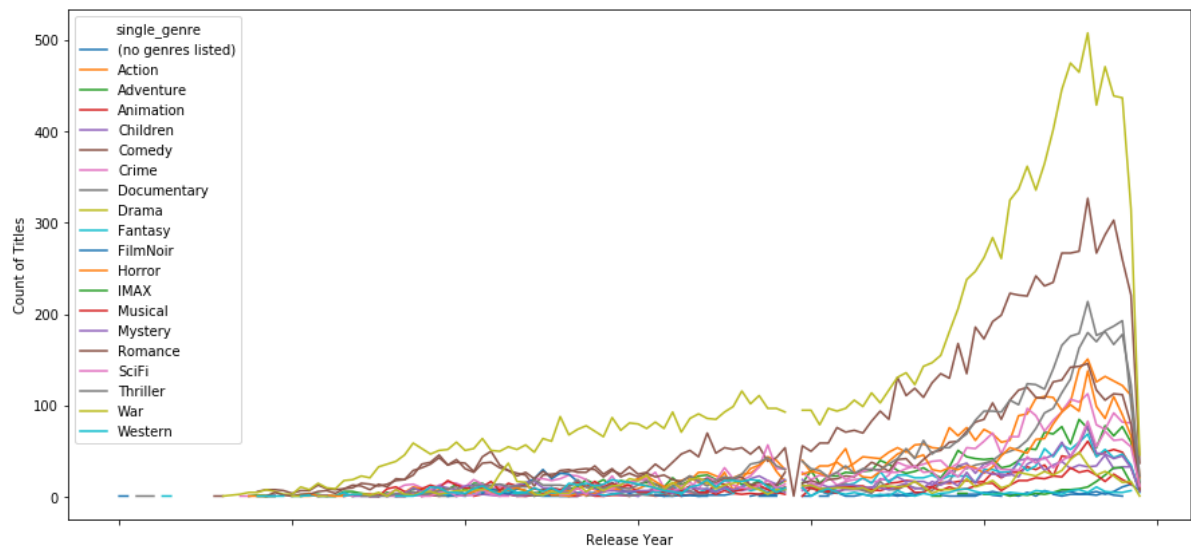
```

(no genres listed)	7
Animation	6
Mystery	6
Fantasy	5
War	1

Name: single_genre, Length: 1817, dtype: int64

```
In [281]: fig, ax = plt.subplots(figsize=(15,7))
exploded_genres.groupby(['release_year', 'single_genre']).count()['title'].unstack().plot(ax=ax)
ax.set_xlabel('Release Year')
ax.set_ylabel('Count of Titles')
```

Out[281]: Text(0,0.5,'Count of Titles')



```
In [279]: movieratings2 = movieratings.groupby(['release_year', 'single_genre'])['rating']  
          .mean()  
          movieratings2
```



```

Out[279]: release_year  single_genre
1891      (no genres listed)  3.000000
1893      (no genres listed)  3.375000
1894      Documentary        2.714286
          Musical            3.428571
1895      Comedy             2.250000
          Documentary        4.000000
1896      Documentary        3.437500
          Romance            2.928571
1898      (no genres listed)  4.333333
          Fantasy            3.750000
          SciFi              2.500000
1899      Fantasy            3.625000
1900      Romance            3.166667
1901      Documentary        5.000000
1902      Action             3.738189
          Adventure          3.738189
          Fantasy            3.738189
          SciFi              3.738189
1903      Crime              3.340909
          Western            3.340909
1905      Comedy             2.500000
          Fantasy            2.500000
1909      Comedy             2.750000
          Crime              3.000000
          Documentary        2.500000
          Drama              3.000000
1910      Drama              3.733333
          Horror             3.466667
          SciFi              2.983333
1912      (no genres listed)  3.500000
          ...
2014      Comedy             3.063935
          Crime              3.196525
          Documentary        3.433191
          Drama              3.236819
          Fantasy            3.197120
          Horror             2.649285
          IMAX               3.220480
          Musical            2.785698
          Mystery            3.030124
          Romance            3.192565
          SciFi              2.927532
          Thriller           3.002945
          War                3.428255
          Western            3.117457
2015      (no genres listed)  2.896429
          Action             2.891214
          Adventure          2.703389
          Animation          2.616987
          Children           2.789966
          Comedy             2.827056
          Crime              3.106252
          Documentary        3.495833
          Drama              2.731007
          Fantasy            2.398232
          Horror             2.156796

```

Mystery	3.002564
Romance	3.033714
SciFi	3.039751
Thriller	2.782928
War	3.000000

Name: rating, Length: 1812, dtype: float64

```
In [282]: fig, ax = plt.subplots(figsize=(15,7))
movieratings.groupby(['release_year', 'single_genre']).mean()['rating'].unstack(
).plot(ax=ax)
ax.set_xlabel('Release Year')
ax.set_ylabel('Rating')

ax.xaxis.set_tick_params(labelbottom=True)
ax.yaxis.set_tick_params(labelbottom=True)
```

