MiniProject

March 30, 2018

1 Genres Trends Data Analysis

```
In [4]: import pandas as pd
```

2 First dataset is movies file

```
In [5]: #create a dataframe from movies file
        mdf = pd.read_csv('./movielens/movies.csv',sep = ',')

        #check the dataframe metadata
        mdf.describe()
        mdf.info()

        #check the data for last 5 records
        mdf.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27278 entries, 0 to 27277
Data columns (total 3 columns):
movieId    27278 non-null int64
title      27278 non-null object
genres     27278 non-null object
dtypes: int64(1), object(2)
memory usage: 639.4+ KB
```

```
Out[5]:    movieId                                title  \
        0        1                    Toy Story (1995)
        1        2                      Jumanji (1995)
        2        3             Grumpier Old Men (1995)
        3        4            Waiting to Exhale (1995)
        4        5  Father of the Bride Part II (1995)

                                              genres
        0  Adventure|Animation|Children|Comedy|Fantasy
        1                   Adventure|Children|Fantasy
        2                               Comedy|Romance
```

```
3                          Comedy|Drama|Romance
4                                       Comedy
```

Transformation *Extract the title and year in new columns*

```
In [6]: #create a new column with only the title
        mdf['NewTitle'] = mdf['title'].str.extract('(.*)\(.*', expand = True)

        #eliminate the spaces before and after the title
        mdf['NewTitle'] = mdf['NewTitle'].replace(mdf['NewTitle'].str.strip())

        mdf['Year'] = mdf['title'].str.extract('.*\((.*)\).*', expand =True)
```

** Analyzing the Genres if we have multiplication in the combintion of them for ex: ** *Romance|Drama and Drama|Romance*

```
In [7]: genres = mdf['genres'].unique().tolist()

        list_romance = []

        for i in genres:
            if 'Romance' in i:
                list_romance.append(i)

        len(genres)
        len(list_romance)

        def getKey(item):
                return item[0]

        sorted(list_romance,key=getKey)

Out[7]: ['Action|Adventure|Romance',
         'Adventure|Romance|IMAX',
         'Animation|Children|Drama|Musical|Romance',
         'Adventure|Comedy|Crime|Romance',
         'Action|Drama|Romance|War',
         'Action|Romance|Western',
         'Action|Drama|Romance',
         'Action|Comedy|Romance',
         'Animation|Children|Comedy|Romance',
         'Adventure|Children|Comedy|Fantasy|Romance',
         'Adventure|Children|Romance',
         'Action|Romance|Thriller',
         'Action|Adventure|Comedy|Romance|Thriller',
         'Action|Adventure|Crime|Drama|Romance|Thriller',
         'Action|Children|Romance',
         'Action|Adventure|Comedy|Romance',
         'Animation|Children|Fantasy|Musical|Romance|IMAX',
```

```
'Adventure|Animation|Children|Fantasy|Musical|Romance',
'Action|Adventure|Romance|Thriller',
'Adventure|Drama|Romance|War',
'Action|Adventure|Mystery|Romance|Thriller',
'Adventure|Romance|War',
'Adventure|Comedy|Romance|War',
'Adventure|Comedy|Crime|Drama|Romance',
'Animation|Children|Fantasy|Musical|Romance',
'Animation|Children|Comedy|Fantasy|Musical|Romance',
'Action|Romance',
'Action|Adventure|Comedy|Fantasy|Romance',
'Adventure|Drama|Romance',
'Action|Crime|Romance|Thriller',
'Action|Drama|Mystery|Romance|Thriller',
'Action|Romance|War|Western',
'Action|Romance|Sci-Fi|Thriller',
'Action|Crime|Romance',
'Adventure|Romance',
'Adventure|Comedy|Romance',
'Adventure|Animation|Children|Comedy|Drama|Musical|Romance',
'Adventure|Children|Drama|Romance',
'Animation|Children|Comedy|Musical|Romance',
'Adventure|Fantasy|Romance',
'Adventure|Drama|Fantasy|Romance',
'Adventure|Fantasy|Romance|Sci-Fi|Thriller',
'Action|Drama|Romance|Sci-Fi',
'Action|Comedy|Romance|Thriller',
'Action|Drama|Romance|Thriller',
'Adventure|Crime|Drama|Romance',
'Adventure|Musical|Romance',
'Adventure|Drama|Romance|Sci-Fi',
'Action|Comedy|Crime|Romance',
'Action|Crime|Mystery|Romance|Thriller',
'Adventure|Comedy|Mystery|Romance',
'Adventure|Animation|Children|Comedy|Fantasy|Romance',
'Action|Animation|Comedy|Crime|Drama|Romance|Thriller',
'Action|Adventure|Drama|Romance',
'Action|Adventure|Drama|Romance|Thriller',
'Action|Comedy|Romance|War',
'Action|Crime|Drama|Romance|Thriller',
'Action|Adventure|Comedy|Drama|Romance|Thriller',
'Action|Comedy|Fantasy|Romance',
'Action|Adventure|Fantasy|Romance',
'Action|Comedy|Drama|Romance',
'Adventure|Romance|War|Western',
'Action|Drama|Romance|Western',
'Adventure|Romance|Western',
'Animation|Drama|Romance',
```

```
'Adventure|Drama|Romance|Thriller|War',
'Adventure|Animation|Children|Comedy|Musical|Romance',
'Action|Adventure|Drama|Romance|War',
'Action|Adventure|Drama|Romance|Thriller|Western',
'Adventure|Horror|Romance|Sci-Fi',
'Action|Adventure|Romance|War',
'Adventure|Comedy|Musical|Romance',
'Action|Comedy|Romance|Western',
'Adventure|Animation|Children|Comedy|Drama|Fantasy|Musical|Romance',
'Action|Adventure|Comedy|Drama|Romance|War',
'Adventure|Comedy|Horror|Romance',
'Action|Fantasy|Horror|Romance',
'Adventure|Animation|Comedy|Fantasy|Romance',
'Action|Fantasy|Romance',
'Action|Crime|Drama|Romance',
'Adventure|Animation|Children|Musical|Romance',
'Action|Adventure|Drama|Fantasy|Romance',
'Action|Adventure|Animation|Comedy|Drama|Fantasy|Romance',
'Action|Animation|Comedy|Romance',
'Action|Adventure|Comedy|Crime|Romance|Thriller',
'Adventure|Animation|Fantasy|Romance',
'Adventure|Drama|Romance|Western',
'Action|Romance|Sci-Fi',
'Action|Adventure|Comedy|Musical|Romance|Thriller',
'Animation|Comedy|Fantasy|Musical|Romance',
'Action|Comedy|Horror|Romance|Sci-Fi',
'Action|Adventure|Comedy|Drama|Romance',
'Animation|Drama|Romance|Sci-Fi',
'Action|Drama|Fantasy|Romance',
'Adventure|Comedy|Drama|Fantasy|Romance',
'Adventure|Musical|Romance|Sci-Fi',
'Adventure|Animation|Comedy|Fantasy|Romance|Sci-Fi',
'Action|Crime|Drama|Romance|Western',
'Adventure|Animation|Drama|Romance|War',
'Adventure|Drama|Romance|Sci-Fi|Thriller',
'Adventure|Drama|Romance|Thriller',
'Action|Adventure|Drama|Fantasy|Horror|Romance|Sci-Fi',
'Adventure|Drama|Fantasy|Musical|Romance',
'Animation|Drama|Romance|Sci-Fi|War',
'Adventure|Comedy|Fantasy|Romance',
'Adventure|Animation|Children|Comedy|Fantasy|Musical|Romance',
'Action|Drama|Musical|Romance',
'Adventure|Comedy|Musical|Romance|Western',
'Animation|Comedy|Drama|Romance|Sci-Fi',
'Adventure|Romance|Thriller',
'Adventure|Animation|Children|Romance|Sci-Fi',
'Action|Animation|Comedy|Romance|Sci-Fi',
'Adventure|Comedy|Drama|Romance',
```

```
'Action|Comedy|Romance|Sci-Fi',
'Adventure|Romance|Sci-Fi',
'Action|Adventure|Fantasy|Horror|Romance',
'Adventure|Crime|Drama|Romance|War',
'Action|Documentary|Drama|Romance|War',
'Adventure|Fantasy|Musical|Romance',
'Action|Adventure|Romance|Western',
'Action|Adventure|Crime|Romance|Western',
'Action|Drama|Mystery|Romance|Sci-Fi',
'Action|Adventure|Animation|Children|Comedy|Romance',
'Adventure|Fantasy|Mystery|Romance|IMAX',
'Action|Adventure|Comedy|Romance|War',
'Animation|Comedy|Romance',
'Action|Adventure|Drama|Romance|Western',
'Animation|Comedy|Musical|Romance',
'Action|Adventure|Drama|Fantasy|Romance|Sci-Fi|Thriller',
'Action|Comedy|Horror|Romance',
'Action|Crime|Drama|Fantasy|Romance',
'Action|Children|Comedy|Romance',
'Action|Adventure|Drama|Musical|Romance|Thriller|War',
'Action|Animation|Comedy|Crime|Drama|Mystery|Romance|Thriller',
'Adventure|Fantasy|Romance|Sci-Fi',
'Action|Adventure|Fantasy|Romance|IMAX',
'Adventure|Fantasy|Horror|Romance|Sci-Fi|Thriller',
'Action|Comedy|Fantasy|Musical|Romance',
'Action|Comedy|Mystery|Romance',
'Animation|Children|Comedy|Fantasy|Musical|Romance|IMAX',
'Action|Musical|Romance',
'Action|Adventure|Comedy|Drama|Mystery|Romance|Thriller',
'Animation|Drama|Romance|War',
'Action|Drama|Horror|Romance|Sci-Fi',
'Adventure|Crime|Drama|Romance|Thriller',
'Action|Comedy|Drama|Romance|War',
'Adventure|Animation|Children|Comedy|Romance',
'Adventure|Mystery|Romance|Thriller',
'Animation|Musical|Romance',
'Action|Adventure|Mystery|Romance',
'Adventure|Comedy|Fantasy|Musical|Romance',
'Adventure|Drama|Fantasy|Romance|IMAX',
'Action|Adventure|Crime|Mystery|Romance',
'Action|Comedy|Drama|Musical|Romance',
'Adventure|Comedy|Fantasy|Romance|IMAX',
'Action|Romance|War',
'Action|Adventure|Fantasy|Musical|Romance',
'Action|Comedy|Crime|Drama|Romance',
'Adventure|Animation|Comedy|Fantasy|Musical|Romance',
'Action|Crime|Musical|Romance',
'Adventure|Animation|Romance',
```

```
'Adventure|Romance|Sci-Fi|IMAX',
'Adventure|Animation|Children|Comedy|Drama|Romance',
'Action|Comedy|Musical|Romance',
'Adventure|Comedy|Musical|Romance|Sci-Fi',
'Animation|Comedy|Drama|Romance',
'Animation|Romance',
'Action|Adventure|Crime|Drama|Romance',
'Action|Adventure|Fantasy|Romance|Sci-Fi',
'Animation|Children|Fantasy|Romance',
'Action|Comedy|Drama|Horror|Romance',
'Action|Crime|Mystery|Romance|Sci-Fi|Thriller',
'Animation|Romance|Sci-Fi',
'Adventure|Animation|Drama|Fantasy|Musical|Romance',
'Animation|Children|Musical|Romance',
'Animation|Children|Romance',
'Action|Comedy|Crime|Drama|Musical|Romance',
'Action|Comedy|Crime|Romance|Thriller',
'Action|Mystery|Romance|Western',
'Adventure|Crime|Mystery|Romance',
'Action|Mystery|Romance',
'Comedy|Romance',
'Comedy|Drama|Romance',
'Crime|Drama|Romance',
'Comedy|Horror|Romance',
'Comedy|Drama|Romance|War',
'Comedy|Drama|Romance|Thriller',
'Comedy|Mystery|Romance',
'Comedy|Romance|Thriller',
'Comedy|Drama|Fantasy|Romance|Thriller',
'Crime|Drama|Romance|Thriller',
'Comedy|Documentary|Romance',
'Comedy|Fantasy|Romance|Sci-Fi',
'Comedy|Musical|Romance',
'Comedy|Crime|Mystery|Romance|Thriller',
'Comedy|Drama|Musical|Romance',
'Crime|Mystery|Romance|Thriller',
'Children|Comedy|Romance',
'Comedy|Fantasy|Romance',
'Comedy|Drama|Fantasy|Romance',
'Crime|Drama|Fantasy|Film-Noir|Mystery|Romance',
'Comedy|Drama|Mystery|Romance',
'Comedy|Crime|Romance',
'Children|Comedy|Romance|Sci-Fi',
'Crime|Romance|Thriller',
'Crime|Drama|Mystery|Romance|Thriller',
'Crime|Drama|Romance|War',
'Comedy|Crime|Drama|Romance|Thriller',
'Comedy|Romance|Sci-Fi',
```

```
'Comedy|Horror|Romance|Thriller',
'Comedy|Drama|Romance|Western',
'Comedy|Crime|Romance|Thriller',
'Children|Comedy|Fantasy|Romance',
'Comedy|Crime|Drama|Romance',
'Comedy|Crime|Mystery|Romance',
'Comedy|Romance|Sci-Fi|Thriller',
'Comedy|Musical|Romance|Western',
'Comedy|Crime|Drama|Mystery|Romance',
'Comedy|Romance|War',
'Crime|Drama|Film-Noir|Romance|Thriller',
'Comedy|Fantasy|Musical|Romance',
'Comedy|Crime|Drama|Musical|Mystery|Romance',
'Crime|Drama|Film-Noir|Romance',
'Crime|Drama|Mystery|Romance',
'Children|Musical|Romance',
'Crime|Film-Noir|Romance',
'Crime|Drama|Musical|Romance',
'Children|Comedy|Musical|Romance',
'Comedy|Romance|Western',
'Comedy|Drama|Romance|Sci-Fi',
'Comedy|Drama|Mystery|Romance|Thriller',
'Comedy|Drama|Horror|Romance|Thriller',
'Comedy|Crime|Drama|Horror|Mystery|Romance|Thriller',
'Comedy|Drama|Fantasy|Musical|Romance',
'Comedy|Fantasy|Mystery|Romance',
'Crime|Film-Noir|Mystery|Romance',
'Comedy|Horror|Romance|Sci-Fi',
'Comedy|Drama|Musical|Romance|War',
'Comedy|Crime|Film-Noir|Mystery|Romance|Thriller',
'Crime|Drama|Film-Noir|Mystery|Romance',
'Children|Comedy|Drama|Musical|Romance',
'Comedy|Romance|Thriller|Western',
'Crime|Film-Noir|Mystery|Romance|Thriller',
'Children|Fantasy|Musical|Romance',
'Crime|Film-Noir|Horror|Romance|Thriller',
'Comedy|Drama|Fantasy|Mystery|Romance',
'Comedy|Documentary|Drama|Romance',
'Crime|Horror|Romance',
'Children|Drama|Romance',
'Comedy|Musical|Romance|War',
'Children|Drama|Fantasy|Musical|Romance',
'Comedy|Crime|Drama|Musical|Romance',
'Comedy|Horror|Musical|Mystery|Romance',
'Crime|Romance',
'Crime|Mystery|Romance',
'Children|Romance',
'Comedy|Drama|Film-Noir|Romance',
```

```
'Children|Fantasy|Romance',
'Crime|Horror|Mystery|Romance|Thriller',
'Comedy|Mystery|Romance|Western',
'Children|Fantasy|Horror|Romance',
'Comedy|Fantasy|Horror|Romance',
'Comedy|Mystery|Romance|Thriller',
'Comedy|Drama|Horror|Romance',
'Crime|Drama|Fantasy|Romance|Thriller',
'Comedy|Fantasy|Romance|Thriller',
'Drama|Romance',
'Drama|Mystery|Romance',
'Drama|Musical|Romance',
'Drama|Fantasy|Romance',
'Drama|Romance|War|Western',
'Drama|Film-Noir|Romance',
'Drama|Horror|Romance|Thriller',
'Drama|Romance|War',
'Drama|Mystery|Romance|Thriller',
'Drama|Mystery|Romance|Sci-Fi|Thriller',
'Drama|Romance|Western',
'Drama|Romance|Thriller',
'Drama|Fantasy|Mystery|Romance',
'Drama|Romance|Sci-Fi|Thriller',
'Drama|Film-Noir|Mystery|Romance',
'Drama|Romance|Sci-Fi',
'Drama|Fantasy|Mystery|Romance|Thriller',
'Drama|Fantasy|Musical|Romance',
'Documentary|Drama|Romance',
'Drama|Horror|Romance',
'Drama|Fantasy|Romance|War',
'Drama|Romance|Thriller|War',
'Drama|Fantasy|Romance|Sci-Fi',
'Drama|Horror|Mystery|Romance|Thriller',
'Drama|Mystery|Romance|War',
'Drama|Fantasy|Horror|Romance',
'Drama|Fantasy|Mystery|Romance|Sci-Fi',
'Drama|Musical|Romance|War',
'Drama|Film-Noir|Romance|Thriller',
'Documentary|Romance',
'Drama|Fantasy|Romance|Thriller',
'Drama|Fantasy|Horror|Romance|Thriller',
'Drama|Fantasy|Horror|Mystery|Romance',
'Drama|Musical|Romance|Sci-Fi',
'Drama|Musical|Romance|IMAX',
'Drama|Film-Noir|Mystery|Romance|Sci-Fi',
'Drama|Fantasy|Film-Noir|Mystery|Romance|War',
'Documentary|Romance|War',
'Film-Noir|Romance|Thriller',
```

```
          'Fantasy|Horror|Romance|Thriller',
          'Fantasy|Horror|Mystery|Romance',
          'Fantasy|Romance',
          'Fantasy|Musical|Romance',
          'Fantasy|Mystery|Romance|Thriller',
          'Fantasy|Romance|Sci-Fi',
          'Fantasy|Romance|Thriller',
          'Fantasy|Romance|Thriller|IMAX',
          'Film-Noir|Mystery|Romance|Thriller',
          'Horror|Romance',
          'Horror|Romance|Thriller',
          'Horror|Mystery|Romance',
          'Horror|Romance|Sci-Fi',
          'Horror|Mystery|Romance|Sci-Fi',
          'Musical|Romance',
          'Mystery|Romance|Thriller',
          'Musical|Romance|War',
          'Mystery|Romance|Sci-Fi|Thriller',
          'Musical|Romance|Western',
          'Mystery|Romance|Western',
          'Mystery|Romance',
          'Mystery|Romance|Sci-Fi',
          'Romance',
          'Romance|Western',
          'Romance|War',
          'Romance|Thriller',
          'Romance|Sci-Fi',
          'Romance|Sci-Fi|Thriller']
```

** Conclusion of the above step: ** *the genres are not dublicated, as the list that represent the genres is ordered alphabetically*

** Verify for the missing values and delete them **

```
In [8]: print (mdf.info())
        print (mdf.isnull().any())
        filter_isnull = mdf['NewTitle'].isnull()
        mdf[filter_isnull]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27278 entries, 0 to 27277
Data columns (total 5 columns):
movieId     27278 non-null int64
title       27278 non-null object
genres      27278 non-null object
NewTitle    27261 non-null object
Year        27261 non-null object
dtypes: int64(1), object(4)
memory usage: 1.0+ MB
```

```
None
movieId     False
title       False
genres      False
NewTitle     True
Year         True
dtype: bool
```

Out[8]:

	movieId	title	genres	NewTitle	Year
10593	40697	Babylon 5	Sci-Fi	NaN	NaN
23617	112406	Brazil: In the Shadow of the Stadiums	Documentary	NaN	NaN
23824	113190	Slaying the Badger	Documentary	NaN	NaN
24286	115133	Tatort: Im Schmerz geboren	Crime	NaN	NaN
24412	115685	National Theatre Live: Frankenstein	Drama|Fantasy	NaN	NaN
26115	125571	The Court-Martial of Jackie Robinson	(no genres listed)	NaN	NaN
26127	125632	In Our Garden	(no genres listed)	NaN	NaN
26180	125958	Stephen Fry In America - New World	(no genres listed)	NaN	NaN
26335	126438	Two: The Story of Roman & Nyro	Documentary|Drama	NaN	NaN
26395	126929	Li'l Quinquin	(no genres listed)	NaN	NaN
26432	127005	A Year Along the Abandoned Road	(no genres listed)	NaN	NaN
26749	128612	Body/Cialo	Comedy|Drama|Mystery	NaN	NaN
26784	128734	Polskie gówno	Comedy|Musical	NaN	NaN
26963	129651	The Third Reich: The Rise & Fall	(no genres listed)	NaN	NaN
26974	129705	My Own Man	(no genres listed)	NaN	NaN
27027	129887	Moving Alan	(no genres listed)	NaN	NaN
27114	130454	Michael Laudrup - en Fodboldspiller	(no genres listed)	NaN	NaN

In [9]: *#clean up the dataframe from NaN values*

```
mdf = mdf.dropna()
mdf.isnull().any()
```

```
Out[9]: movieId     False
        title       False
        genres      False
        NewTitle    False
        Year        False
        dtype: bool
```

3 Second dataset file to be used is Rating

```
In [10]: rdf = pd.read_csv('./movielens/ratings.csv',sep = ',', parse_dates = ['timestamp'])
         rdf.head()
```

```
Out[10]:    userId  movieId  rating   timestamp
         0       1        2     3.5  1112486027
         1       1       29     3.5  1112484676
         2       1       32     3.5  1112484819
         3       1       47     3.5  1112484727
         4       1       50     3.5  1112484580
```

```
In [11]: rdf.info()
         rdf.isnull().any()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000263 entries, 0 to 20000262
Data columns (total 4 columns):
userId       int64
movieId      int64
rating       float64
timestamp    object
dtypes: float64(1), int64(2), object(1)
memory usage: 610.4+ MB
```

```
Out[11]: userId       False
         movieId      False
         rating       False
         timestamp    False
         dtype: bool
```

Transformation for Rating dataframe

```
In [12]: rdf['rdate'] = pd.to_datetime(rdf['timestamp'],unit = 's')
         rdf.head()
```

```
Out[12]:    userId  movieId  rating   timestamp               rdate
         0       1        2     3.5  1112486027 2005-04-02 23:53:47
```

```
1       1       29      3.5  1112484676 2005-04-02 23:31:16
2       1       32      3.5  1112484819 2005-04-02 23:33:39
3       1       47      3.5  1112484727 2005-04-02 23:32:07
4       1       50      3.5  1112484580 2005-04-02 23:29:40
```

4 Merging the dataframes to get the rating value in the movie dataframe

```
In [13]: mdf = mdf.merge(rdf, on = 'movieId', how = 'inner')
         mdf.head()
```

```
Out[13]:    movieId            title                                        genres  \
         0        1  Toy Story (1995)  Adventure|Animation|Children|Comedy|Fantasy
         1        1  Toy Story (1995)  Adventure|Animation|Children|Comedy|Fantasy
         2        1  Toy Story (1995)  Adventure|Animation|Children|Comedy|Fantasy
         3        1  Toy Story (1995)  Adventure|Animation|Children|Comedy|Fantasy
         4        1  Toy Story (1995)  Adventure|Animation|Children|Comedy|Fantasy

              NewTitle  Year  userId  rating   timestamp               rdate
         0  Toy Story   1995       3     4.0   944919407 1999-12-11 13:36:47
         1  Toy Story   1995       6     5.0   858275452 1997-03-13 17:50:52
         2  Toy Story   1995       8     4.0   833981871 1996-06-05 13:37:51
         3  Toy Story   1995      10     4.0   943497887 1999-11-25 02:44:47
         4  Toy Story   1995      11     4.5  1230858821 2009-01-02 01:13:41
```

CleanUp and delete the uneccesary fields from the dataframe

```
In [14]: mdf.pop('title')
         mdf.pop('timestamp')
         mdf.head()
```

```
Out[14]:    movieId                                        genres    NewTitle  Year  \
         0        1  Adventure|Animation|Children|Comedy|Fantasy  Toy Story   1995
         1        1  Adventure|Animation|Children|Comedy|Fantasy  Toy Story   1995
         2        1  Adventure|Animation|Children|Comedy|Fantasy  Toy Story   1995
         3        1  Adventure|Animation|Children|Comedy|Fantasy  Toy Story   1995
         4        1  Adventure|Animation|Children|Comedy|Fantasy  Toy Story   1995

            userId  rating               rdate
         0       3     4.0 1999-12-11 13:36:47
         1       6     5.0 1997-03-13 17:50:52
         2       8     4.0 1996-06-05 13:37:51
         3      10     4.0 1999-11-25 02:44:47
         4      11     4.5 2009-01-02 01:13:41
```

Group the data for analysis by genres and year

```
In [15]: r1_df = mdf[['genres','Year','rating']].groupby(['genres','Year'], as_index = False).a
         r1_df.head()
```

Out[15]:

	genres	Year	rating	
			AvgRating	CountRatings
0	(no genres listed)	1891	3.000000	1
1	(no genres listed)	1893	3.375000	4
2	(no genres listed)	1898	4.333333	3
3	(no genres listed)	1912	3.500000	1
4	(no genres listed)	1917	3.500000	1

Check the data grouped by Genres only, then build the result data set with two measures

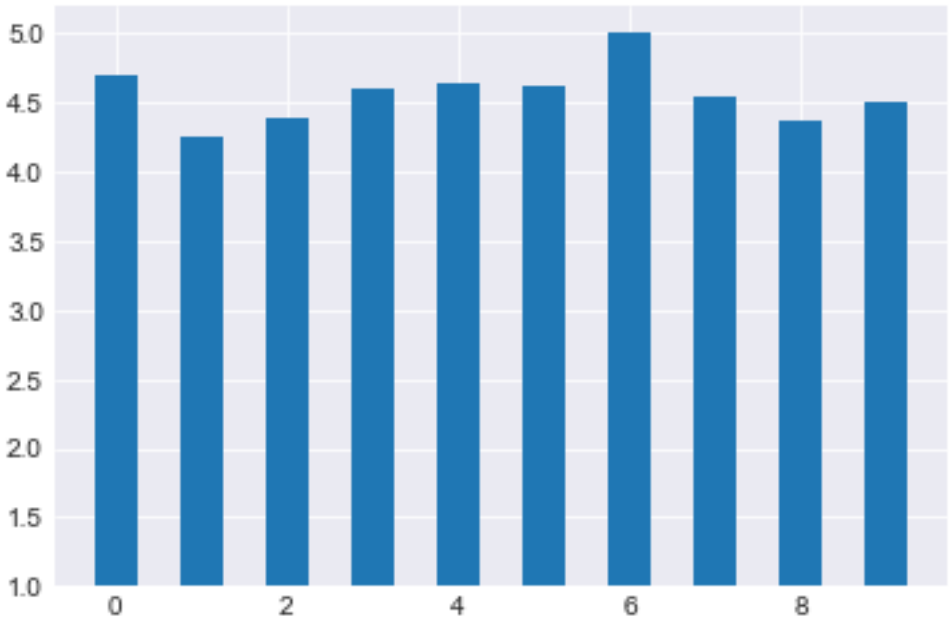
In [16]:
```
#r1_df[r1_df['genres'] == 'Drama|Romance']

r2_df = mdf[['genres','rating']].groupby(['genres'], as_index = False).agg({'rating':
    columns = ({'mean': 'AvgRating','count':'CountRatings'}))
#r2_df.head()
r3_df  = r2_df[r2_df[('rating','CountRatings')] > 3].sort_values(by = [('rating', 'Cou
r3_df = r3_df.reset_index()
r3_df['counter'] = range(len(r3_df))
#r2_df.info()
#sorted(r2_df)
r3_df

r_df = mdf[['genres','rating','Year']][mdf['genres'].isin( r3_df['genres'].values)].gr
    columns = ({'mean': 'AvgRating','count':'CountRatings'}))
r_df.head()
#r_df.genres.unique()
```

Out[16]:

	genres	Year	rating	
			AvgRating	CountRatings
0	Action|Adventure|Sci-Fi	1916	3.257143	70
1	Action|Adventure|Sci-Fi	1936	3.040000	25
2	Action|Adventure|Sci-Fi	1960	3.699373	2869
3	Action|Adventure|Sci-Fi	1965	2.857143	14
4	Action|Adventure|Sci-Fi	1967	1.900000	5

5 Best Mean rating values for top 10 genres

In [134]:
```
#r2_df[('rating','AvgRating')].corr(r2_df[('rating','CountRatings')])
import matplotlib.pyplot as plt

plt.bar(r3_df['counter'].values,r3_df[('rating','AvgRating')].values, 0.5,1)
plt.show()
print (r3_df[['counter','genres']])
```



	counter	genres
0	0	Drama
1	1	Comedy
2	2	Comedy|Romance
3	3	Comedy|Drama
4	4	Drama|Romance
5	5	Comedy|Drama|Romance
6	6	Crime|Drama
7	7	Action|Adventure|Sci-Fi
8	8	Action|Adventure|Thriller
9	9	Action|Crime|Thriller

Checking if there is any correlation between genres

```
In [18]: mdf['rating'][mdf['genres'] == 'Drama'].corr(mdf['rating'][mdf['genres'] == 'Crime|Dra
         #mdf['rating'][mdf['genres'] == 'Comedy']
```

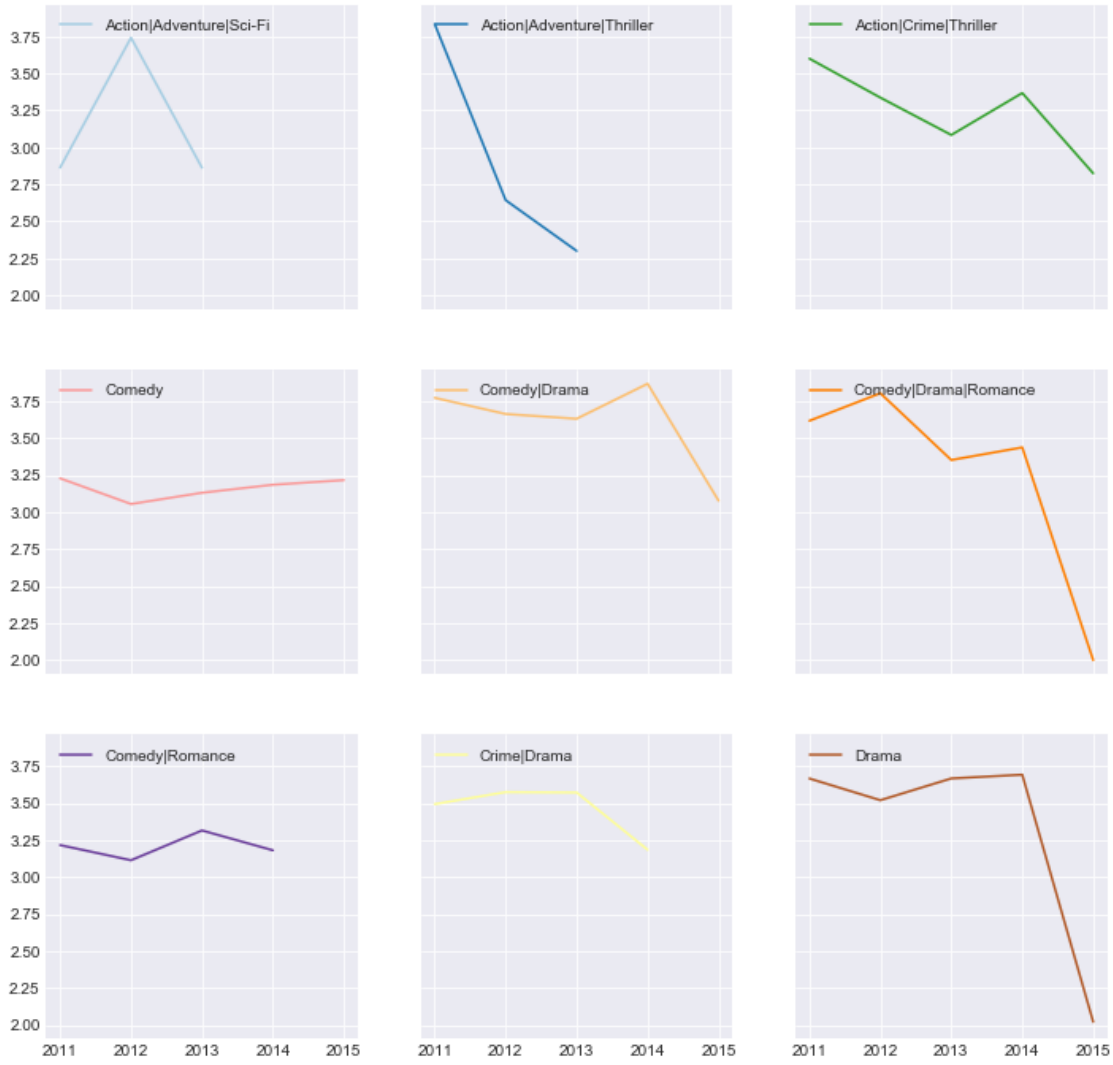
```
Out[18]: nan
```

Checking the trend of a genre for last 5 years of data

```
In [ ]: import matplotlib.pyplot as plt

        plt.plot(r_df['Year'][(r_df['genres'] == 'Comedy') & ( r_df['Year'] > '2010')].head(),r
        #plt.plot(r1_df['Year'].head(),r1_df[('rating','CountRatings')].head())
        plt.show()
```

6 Showing the top 9 genres and their trends line for last 5 years

In [126]:
```
import numpy as np


        r_df = r_df[r_df['Year'] > '2010'].sort_values(by = ['Year','genres'])

        g_list = np.array(r_df['genres'].unique())
        #type(g_list)
        #print (len(g_list))



        # Initialize the figure
        plt.style.use('seaborn-darkgrid')

        colors = plt.cm.Paired(np.linspace(0,1,9))    #third option

        fig, axes = plt.subplots(3, 3, sharex=True, sharey=True, figsize = (12,12))


        for c,ax in enumerate(fig.axes):
            #for r in enumerate(fig.axes):
            p_df = r_df[(r_df['genres'] == g_list[c]) ]
            ax.plot(p_df['Year'],p_df[('rating','AvgRating')],label = g_list[c], color = colo
            ax.legend(loc="upper left")



        #print (item)
        #print (g_list[i])

        plt.show()


        #r_df[(r_df['Year'] > '2010')]
```



7 Conclusion

The filmaker should focus on creating movies whith genres that get a big interest in the last 5 years. As we see from above graphical respresentation we can identify 3 most stable and with high rating genres through the period of 5 years: Comedy | Drama, Comedy, Comedy | Romance.