Movielens_Project

October 21, 2020

[65]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
```

[66]:
```
df = pd.read_csv('movies.dat',sep='::',names=['Movie_id','Title','Genres'])
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:1: ParserWarning:
Falling back to the 'python' engine because the 'c' engine does not support
regex separators (separators > 1 char and different from '\s+' are interpreted
as regex); you can avoid this warning by specifying engine='python'.
  """Entry point for launching an IPython kernel.
```

[67]:
```
df.head()
```

[67]:

	Movie_id	Title	Genres
0	1	Toy Story (1995)	Animation|Children's|Comedy
1	2	Jumanji (1995)	Adventure|Children's|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama
4	5	Father of the Bride Part II (1995)	Comedy

[68]:
```
df1 = pd.read_csv('ratings.dat',sep='::
↪',names=['User_id','Movie_id','Rating','timestamp'])
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:1: ParserWarning:
Falling back to the 'python' engine because the 'c' engine does not support
regex separators (separators > 1 char and different from '\s+' are interpreted
as regex); you can avoid this warning by specifying engine='python'.
  """Entry point for launching an IPython kernel.
```

[69]:
```
df1.head()
```

[69]:

	User_id	Movie_id	Rating	timestamp
0	1	1193	5	978300760
1	1	661	3	978302109

```
2        1       914       3  978301968
3        1      3408       4  978300275
4        1      2355       5  978824291
```

[70]:
```
df2 = pd.read_csv('users.dat',sep='::
 ↪',names=['User_id','Gender','Age','Occupation','Zip_code'])
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:1: ParserWarning:
Falling back to the 'python' engine because the 'c' engine does not support
regex separators (separators > 1 char and different from '\s+' are interpreted
as regex); you can avoid this warning by specifying engine='python'.
  """Entry point for launching an IPython kernel.
```

[71]:
```
df2.head()
```

[71]:
```
   User_id Gender  Age  Occupation Zip_code
0        1      F    1          10    48067
1        2      M   56          16    70072
2        3      M   25          15    55117
3        4      M   45           7    02460
4        5      M   25          20    55455
```

[72]:
```
data = pd.merge(df,df1,on='Movie_id')
```

[73]:
```
data.head()
```

[73]:
```
   Movie_id             Title                        Genres  User_id  Rating  \
0         1  Toy Story (1995)  Animation|Children's|Comedy        1       5
1         1  Toy Story (1995)  Animation|Children's|Comedy        6       4
2         1  Toy Story (1995)  Animation|Children's|Comedy        8       4
3         1  Toy Story (1995)  Animation|Children's|Comedy        9       5
4         1  Toy Story (1995)  Animation|Children's|Comedy       10       5

   timestamp
0  978824268
1  978237008
2  978233496
3  978225952
4  978226474
```

[74]:
```
data = pd.merge(data,df2,on='User_id')
```

[75]:
```
data.head()
```

[75]:
```
   Movie_id                                Title  \
0         1                    Toy Story (1995)
1        48                   Pocahontas (1995)
```

```
2       150                            Apollo 13 (1995)
3       260  Star Wars: Episode IV - A New Hope (1977)
4       527                      Schindler's List (1993)

                                Genres  User_id  Rating  timestamp Gender  \
0           Animation|Children's|Comedy        1       5  978824268      F
1  Animation|Children's|Musical|Romance        1       5  978824351      F
2                                 Drama        1       5  978301777      F
3       Action|Adventure|Fantasy|Sci-Fi        1       4  978300760      F
4                             Drama|War        1       5  978824195      F

   Age  Occupation Zip_code
0    1          10    48067
1    1          10    48067
2    1          10    48067
3    1          10    48067
4    1          10    48067
```

[76]:
```
Master_data = data.drop(['Genres','timestamp','Zip_code'],axis=1)
```

[77]:
```
Master_data.isnull().sum()
```

[77]:
```
Movie_id      0
Title         0
User_id       0
Rating        0
Gender        0
Age           0
Occupation    0
dtype: int64
```

[78]:
```
Master_data.duplicated().value_counts()
```

[78]:
```
False    1000209
dtype: int64
```

[79]:
```
Master_data.shape
```

[79]:
```
(1000209, 7)
```

[80]:
```
Master_data.dtypes
```

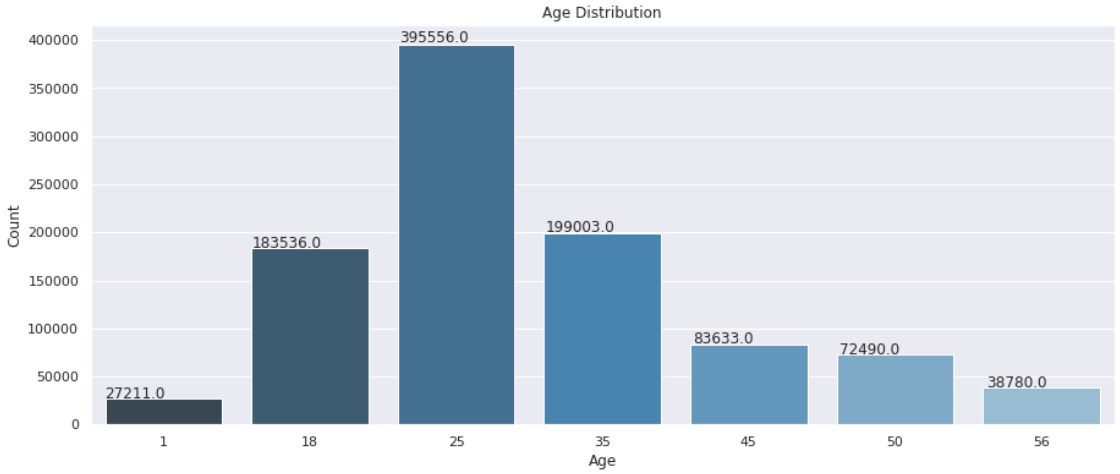
[80]:
```
Movie_id       int64
Title         object
User_id        int64
Rating         int64
Gender        object
```

```
Age            int64
Occupation     int64
dtype: object
```

[81]:
```
Age_distribution = Master_data.Age.value_counts()
Age_distribution
```

[81]:
```
25    395556
35    199003
18    183536
45     83633
50     72490
56     38780
1      27211
Name: Age, dtype: int64
```

[82]:
```
ax = sns.barplot(x=Age_distribution.index,y=Age_distribution.
 ↪values,palette='Blues_d')
fig = plt.gcf();
fig.set_size_inches(15,6)
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Age Distribution')

for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_height() * 1.
 ↪005));
```



Age Group of '25-34' contains the most user and less than 17 contains the least one.

[83]:
```
Toy_story_ratings = Master_data[Master_data['Movie_id']==1]
```

[84]:
```
Toy_story_ratings.head()
```

[84]:

	Movie_id	Title	User_id	Rating	Gender	Age	Occupation
0	1	Toy Story (1995)	1	5	F	1	10
53	1	Toy Story (1995)	6	4	F	50	9
124	1	Toy Story (1995)	8	4	M	25	12
263	1	Toy Story (1995)	9	5	M	25	17
369	1	Toy Story (1995)	10	5	F	35	1

[85]:
```
rating_count = Toy_story_ratings.Rating.value_counts()
rating_count
```

[85]:
```
4    835
5    820
3    345
2     61
1     16
Name: Rating, dtype: int64
```

[86]:
```
avg_rating = Toy_story_ratings.Rating.sum()/Toy_story_ratings.Rating.count()
avg_rating
```

[86]:
```
4.146846413095811
```

Average Rating of the Movie Toy Story is 4.14.

[87]:
```
Total_Ratings = Master_data.groupby(['Title'])['Rating'].count()
Total_Ratings[Total_Ratings>20]
```

[87]:
```
Title
$1,000,000 Duck (1971)                37
'Night Mother (1986)                  70
'Til There Was You (1997)             52
'burbs, The (1989)                   303
…And Justice for All (1979)        199
                                     …
Your Friends and Neighbors (1998)    109
Zed & Two Noughts, A (1985)           29
Zero Effect (1998)                   301
Zeus and Roxanne (1997)               23
eXistenZ (1999)                      410
Name: Rating, Length: 3011, dtype: int64
```

[88]:
```
Ratings_sum = Master_data.groupby(['Title'])['Rating'].sum()
Ratings_sum[Total_Ratings>20]
```

[88]:
```
Title
$1,000,000 Duck (1971)                112
'Night Mother (1986)                  236
'Til There Was You (1997)             140
'burbs, The (1989)                    882
…And Justice for All (1979)          739
                                      …
Your Friends and Neighbors (1998)     368
Zed & Two Noughts, A (1985)            99
Zero Effect (1998)                    1129
Zeus and Roxanne (1997)                58
eXistenZ (1999)                       1335
Name: Rating, Length: 3011, dtype: int64
```

[89]:
```
Movie_Rating = Ratings_sum[Total_Ratings>20]/Total_Ratings[Total_Ratings>20]
```

[90]:
```
Movie_Rating
```

[90]:
```
Title
$1,000,000 Duck (1971)                3.027027
'Night Mother (1986)                  3.371429
'Til There Was You (1997)             2.692308
'burbs, The (1989)                    2.910891
…And Justice for All (1979)          3.713568
                                        …
Your Friends and Neighbors (1998)     3.376147
Zed & Two Noughts, A (1985)           3.413793
Zero Effect (1998)                    3.750831
Zeus and Roxanne (1997)               2.521739
eXistenZ (1999)                       3.256098
Name: Rating, Length: 3011, dtype: float64
```

[91]:
```
Movie_Rating.sort_values(ascending=False).head(25)
```

[91]:
```
Title
Sanjuro (1962)
4.608696
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)
4.560510
Shawshank Redemption, The (1994)
4.554558
Godfather, The (1972)
4.524966
Close Shave, A (1995)
4.520548
Usual Suspects, The (1995)
4.517106
```

```
Schindler's List (1993)
4.510417
Wrong Trousers, The (1993)
4.507937
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)
4.491489
Raiders of the Lost Ark (1981)
4.477725
Rear Window (1954)
4.476190
Paths of Glory (1957)
4.473913
Star Wars: Episode IV - A New Hope (1977)
4.453694
Third Man, The (1949)
4.452083
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)
4.449890
For All Mankind (1989)
4.444444
Wallace & Gromit: The Best of Aardman Animation (1996)
4.426941
To Kill a Mockingbird (1962)
4.425647
Double Indemnity (1944)
4.415608
Casablanca (1942)
4.412822
World of Apu, The (Apur Sansar) (1959)
4.410714
Sixth Sense, The (1999)
4.406263
Yojimbo (1961)
4.404651
Pather Panchali (1955)
4.404255
Lawrence of Arabia (1962)
4.401925
Name: Rating, dtype: float64
```

List of Top 25 Movies by Viewership Rating with Ratings per movie is greater than 20.

[92]:
```
User_2696_Rating = Master_data[Master_data['User_id']==2696]
User_2696_Rating
```

[92]:

	Movie_id	Title	User_id	\
991035	350	Client, The (1994)	2696	

991036	800	Lone Star (1996)	2696
991037	1092	Basic Instinct (1992)	2696
991038	1097	E.T. the Extra-Terrestrial (1982)	2696
991039	1258	Shining, The (1980)	2696
991040	1270	Back to the Future (1985)	2696
991041	1589	Cop Land (1997)	2696
991042	1617	L.A. Confidential (1997)	2696
991043	1625	Game, The (1997)	2696
991044	1644	I Know What You Did Last Summer (1997)	2696
991045	1645	Devil's Advocate, The (1997)	2696
991046	1711	Midnight in the Garden of Good and Evil (1997)	2696
991047	1783	Palmetto (1998)	2696
991048	1805	Wild Things (1998)	2696
991049	1892	Perfect Murder, A (1998)	2696
991050	2338	I Still Know What You Did Last Summer (1998)	2696
991051	2389	Psycho (1998)	2696
991052	2713	Lake Placid (1999)	2696
991053	3176	Talented Mr. Ripley, The (1999)	2696
991054	3386	JFK (1991)	2696

	Rating	Gender	Age	Occupation
991035	3	M	25	7
991036	5	M	25	7
991037	4	M	25	7
991038	3	M	25	7
991039	4	M	25	7
991040	2	M	25	7
991041	3	M	25	7
991042	4	M	25	7
991043	4	M	25	7
991044	2	M	25	7
991045	4	M	25	7
991046	4	M	25	7
991047	4	M	25	7
991048	4	M	25	7
991049	4	M	25	7
991050	2	M	25	7
991051	4	M	25	7
991052	1	M	25	7
991053	4	M	25	7
991054	1	M	25	7

[93]:
```
User_2696_Rating[['Movie_id','Title','Rating']]
```

[93]:

	Movie_id	Title	Rating
991035	350	Client, The (1994)	3
991036	800	Lone Star (1996)	5

```
991037      1092                            Basic Instinct (1992)       4
991038      1097               E.T. the Extra-Terrestrial (1982)       3
991039      1258                              Shining, The (1980)       4
991040      1270                         Back to the Future (1985)      2
991041      1589                                  Cop Land (1997)       3
991042      1617                          L.A. Confidential (1997)      4
991043      1625                                Game, The (1997)        4
991044      1644           I Know What You Did Last Summer (1997)       2
991045      1645                      Devil's Advocate, The (1997)      4
991046      1711  Midnight in the Garden of Good and Evil (1997)       4
991047      1783                                Palmetto (1998)         4
991048      1805                              Wild Things (1998)        4
991049      1892                           Perfect Murder, A (1998)     4
991050      2338     I Still Know What You Did Last Summer (1998)       2
991051      2389                                   Psycho (1998)       4
991052      2713                               Lake Placid (1999)      1
991053      3176                  Talented Mr. Ripley, The (1999)      4
991054      3386                                    JFK (1991)         1
```

Ratings for all the movies rated by user with user_id=2696

[94]:
```
Genre_data = data.drop(['timestamp','Zip_code'],axis=1)
Genre_data
```

[94]:

	Movie_id	Title	Genres	User_id	Rating	Gender	Age
0	1	Toy Story (1995)	Animation|Children's|Comedy	1	5	F	1
1	48	Pocahontas (1995)	Animation|Children's|Musical|Romance	1	5	F	1
2	150	Apollo 13 (1995)	Drama	1	5	F	1
3	260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi	1	4	F	1
4	527	Schindler's List (1993)	Drama|War	1	5	F	1
…	…	…	…	…	…	…	…
1000204	3513	Rules of Engagement (2000)	Drama|Thriller	5727	4	M	25
1000205	3535	American Psycho (2000)	Comedy|Horror|Thriller	5727	2	M	25
1000206	3536	Keeping the Faith (2000)	Comedy|Romance	5727	5	M	25
1000207	3555	U-571 (2000)					
1000208	3578	Gladiator (2000)					

```
1000207                      Action|Thriller     5727       3      M   25
1000208                         Action|Drama     5727       5      M   25

         Occupation
0                10
1                10
2                10
3                10
4                10
…               …
1000204           4
1000205           4
1000206           4
1000207           4
1000208           4

[1000209 rows x 8 columns]
```

[95]:
```
genre_list = []
Genre_array = Genre_data['Genres'].unique()
Genre_array
```

[95]:
```
array(["Animation|Children's|Comedy",
       "Animation|Children's|Musical|Romance", 'Drama',
       'Action|Adventure|Fantasy|Sci-Fi', 'Drama|War', "Children's|Drama",
       "Animation|Children's|Comedy|Musical",
       "Animation|Children's|Musical", 'Crime|Drama|Thriller',
       'Animation', 'Animation|Comedy|Thriller', 'Musical|Romance',
       "Adventure|Children's|Drama|Musical", 'Musical',
       "Children's|Comedy|Musical", "Children's|Drama|Fantasy|Sci-Fi",
       'Action|Adventure|Comedy|Romance', 'Comedy|Sci-Fi',
       'Action|Adventure|Drama',
       "Adventure|Animation|Children's|Comedy|Musical", 'Drama|Romance',
       "Animation|Children's", 'Action|Drama|War', 'Comedy', 'Romance',
       'Action|Crime|Romance', 'Thriller', 'Comedy|Fantasy',
       'Comedy|Drama', "Children's|Comedy|Drama", 'Drama|Musical',
       'Drama|Romance|War|Western', 'Crime|Drama',
       'Action|Comedy|Western', 'Action|Romance|Thriller', 'Western',
       "Children's|Comedy", 'Adventure|Drama|Western', 'Comedy|Romance',
       'Comedy|Drama|Romance', 'Drama|Romance|War',
       "Children's|Comedy|Western",
       "Adventure|Animation|Children's|Musical", 'Action|Romance',
       'Action|Adventure|Romance|Sci-Fi|War', 'Comedy|Musical|Romance',
       'Drama|Romance|Thriller', "Adventure|Children's|Comedy",
       'Action|Adventure|Romance', "Children's|Fantasy|Musical",
       "Animation|Children's|Comedy|Musical|Romance",
       'Comedy|Fantasy|Romance', 'Action|Drama', 'Comedy|Musical',
```

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

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

```
       'Adventure|Musical|Romance', "Children's|Comedy|Mystery",
       'Romance|War', 'Action|Comedy|Romance|Thriller',
       'Musical|Romance|War', "Animation|Children's|Comedy|Romance",
       'Comedy|Mystery|Romance', 'Action|Drama|Western',
       "Action|Animation|Children's|Sci-Fi|Thriller|War",
       'Comedy|Drama|Musical', 'Adventure|Comedy|Musical',
       'Action|Crime|Mystery|Thriller', 'Action|Adventure|Drama|Thriller',
       'Action|Adventure|Comedy|War', 'Mystery', 'Drama|Western',
       'Action|Adventure|Crime|Thriller',
       'Action|Mystery|Sci-Fi|Thriller',
       "Adventure|Children's|Comedy|Fantasy|Romance",
       "Adventure|Children's|Romance",
       "Action|Adventure|Animation|Children's|Fantasy",
       "Action|Adventure|Children's", "Adventure|Animation|Children's",
       'Musical|War', 'Action|Crime|Mystery',
       "Adventure|Animation|Children's|Fantasy", 'Comedy|Horror|Thriller',
       'Film-Noir', 'Crime|Film-Noir|Mystery', 'Drama|Film-Noir|Thriller',
       'Drama|Film-Noir', 'Action|Adventure|War', 'Crime|Drama|Romance',
       'Documentary|War', 'Sci-Fi|Thriller|War', 'Action|Comedy|Crime',
       'Crime|Horror', 'Drama|Romance|Sci-Fi', 'Crime|Mystery',
       'Comedy|Drama|Thriller', 'Crime|Horror|Thriller', 'Horror|Mystery',
       'Documentary|Drama', 'Drama|Horror|Thriller',
       'Comedy|Horror|Sci-Fi', "Action|Adventure|Children's|Fantasy",
       'Animation|Mystery', 'Comedy|Romance|Sci-Fi', 'Romance|Western',
       'Drama|Romance|Western', 'Comedy|Film-Noir|Thriller',
       'Film-Noir|Horror', 'Fantasy'], dtype=object)
```

[96]:
```
for genre in Genre_array:
    x = genre.split(sep='|')
    for i in range(len(x)):
        if x[i] not in genre_list:
            genre_list.append(x[i])
```

[97]:
```
genre_list
```

[97]:
```
['Animation',
 "Children's",
 'Comedy',
 'Musical',
 'Romance',
 'Drama',
 'Action',
 'Adventure',
 'Fantasy',
 'Sci-Fi',
 'War',
 'Crime',
```

```
 'Thriller',
 'Western',
 'Horror',
 'Mystery',
 'Documentary',
 'Film-Noir']
```

Different Genres present in the list.

[98]:
```
Genre = Genre_data['Genres']
Genre = Genre.str.get_dummies()
Genre
```

[98]:

	Action	Adventure	Animation	Children's	Comedy	Crime	Documentary	\
0	0	0	1	1	1	0	0	
1	0	0	1	1	0	0	0	
2	0	0	0	0	0	0	0	
3	1	1	0	0	0	0	0	
4	0	0	0	0	0	0	0	
…	…	…	…	…	…	…	…	
1000204	0	0	0	0	0	0	0	
1000205	0	0	0	0	1	0	0	
1000206	0	0	0	0	1	0	0	
1000207	1	0	0	0	0	0	0	
1000208	1	0	0	0	0	0	0	

	Drama	Fantasy	Film-Noir	Horror	Musical	Mystery	Romance	Sci-Fi	\
0	0	0	0	0	0	0	0	0	
1	0	0	0	0	1	0	1	0	
2	1	0	0	0	0	0	0	0	
3	0	1	0	0	0	0	0	1	
4	1	0	0	0	0	0	0	0	
…	…	…	…	…	…	…	…	…	
1000204	1	0	0	0	0	0	0	0	
1000205	0	0	0	1	0	0	0	0	
1000206	0	0	0	0	0	0	1	0	
1000207	0	0	0	0	0	0	0	0	
1000208	1	0	0	0	0	0	0	0	

	Thriller	War	Western
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	1	0
…	…	…	…
1000204	1	0	0

```
1000205         1    0        0
1000206         0    0        0
1000207         1    0        0
1000208         0    0        0

[1000209 rows x 18 columns]
```

[99]:
```
movie_rating_genre = pd.concat([Genre_data,Genre],axis=1)
```

[100]:
```
movie_rating_genre.head()
```

[100]:
```
   Movie_id                                        Title  \
0         1                             Toy Story (1995)
1        48                            Pocahontas (1995)
2       150                             Apollo 13 (1995)
3       260  Star Wars: Episode IV - A New Hope (1977)
4       527                     Schindler's List (1993)

                                Genres  User_id  Rating Gender  Age  \
0           Animation|Children's|Comedy        1       5      F    1
1  Animation|Children's|Musical|Romance        1       5      F    1
2                                 Drama        1       5      F    1
3       Action|Adventure|Fantasy|Sci-Fi        1       4      F    1
4                           Drama|War        1       5      F    1

   Occupation  Action  Adventure  …  Fantasy  Film-Noir  Horror  Musical  \
0          10       0          0  …        0          0       0        0
1          10       0          0  …        0          0       0        1
2          10       0          0  …        0          0       0        0
3          10       1          1  …        1          0       0        0
4          10       0          0  …        0          0       0        0

   Mystery  Romance  Sci-Fi  Thriller  War  Western
0        0        0       0         0    0        0
1        0        1       0         0    0        0
2        0        0       0         0    0        0
3        0        0       1         0    0        0
4        0        0       0         0    1        0

[5 rows x 26 columns]
```

[101]:
```
movie_rating_genre.drop(['Genres','Title'],axis=1,inplace=True)
```

[102]:
```
movie_rating_genre
```

[102]:
```
         Movie_id  User_id  Rating Gender  Age  Occupation  Action  Adventure  \
0               1        1       5      F    1          10       0          0
```

```
1              48        1       5      F    1          10       0          0
2             150        1       5      F    1          10       0          0
3             260        1       4      F    1          10       1          1
4             527        1       5      F    1          10       0          0
…             …      …     …    …  …         …      …        …
1000204      3513     5727       4      M   25          4       0          0
1000205      3535     5727       2      M   25          4       0          0
1000206      3536     5727       5      M   25          4       0          0
1000207      3555     5727       3      M   25          4       1          0
1000208      3578     5727       5      M   25          4       1          0

         Animation  Children's  …  Fantasy  Film-Noir  Horror  Musical  \
0                1           1  …        0          0       0        0
1                1           1  …        0          0       0        1
2                0           0  …        0          0       0        0
3                0           0  …        1          0       0        0
4                0           0  …        0          0       0        0
…              …         …  …      …         …      …       …
1000204          0           0  …        0          0       0        0
1000205          0           0  …        0          0       1        0
1000206          0           0  …        0          0       0        0
1000207          0           0  …        0          0       0        0
1000208          0           0  …        0          0       0        0

         Mystery  Romance  Sci-Fi  Thriller  War  Western
0              0        0       0         0    0        0
1              0        1       0         0    0        0
2              0        0       0         0    0        0
3              0        0       1         0    0        0
4              0        0       0         0    1        0
…            …      …     …       …  …       …
1000204        0        0       0         1    0        0
1000205        0        0       0         1    0        0
1000206        0        1       0         0    0        0
1000207        0        0       0         1    0        0
1000208        0        0       0         0    0        0

[1000209 rows x 24 columns]
```

[103]:
```
movie_rating_genre.dtypes
```

[103]:
```
Movie_id        int64
User_id         int64
Rating          int64
Gender         object
Age             int64
Occupation      int64
```

```
Action          int64
Adventure       int64
Animation       int64
Children's      int64
Comedy          int64
Crime           int64
Documentary     int64
Drama           int64
Fantasy         int64
Film-Noir       int64
Horror          int64
Musical         int64
Mystery         int64
Romance         int64
Sci-Fi          int64
Thriller        int64
War             int64
Western         int64
dtype: object
```

[104]:
```
movie_rating_genre = pd.get_dummies(movie_rating_genre,columns=['Gender'])
```

[105]:
```
movie_rating_genre
```

[105]:

	Movie_id	User_id	Rating	Age	Occupation	Action	Adventure	\
0	1	1	5	1	10	0	0	
1	48	1	5	1	10	0	0	
2	150	1	5	1	10	0	0	
3	260	1	4	1	10	1	1	
4	527	1	5	1	10	0	0	
…	…	…	…	…	…	…	…	
1000204	3513	5727	4	25	4	0	0	
1000205	3535	5727	2	25	4	0	0	
1000206	3536	5727	5	25	4	0	0	
1000207	3555	5727	3	25	4	1	0	
1000208	3578	5727	5	25	4	1	0	

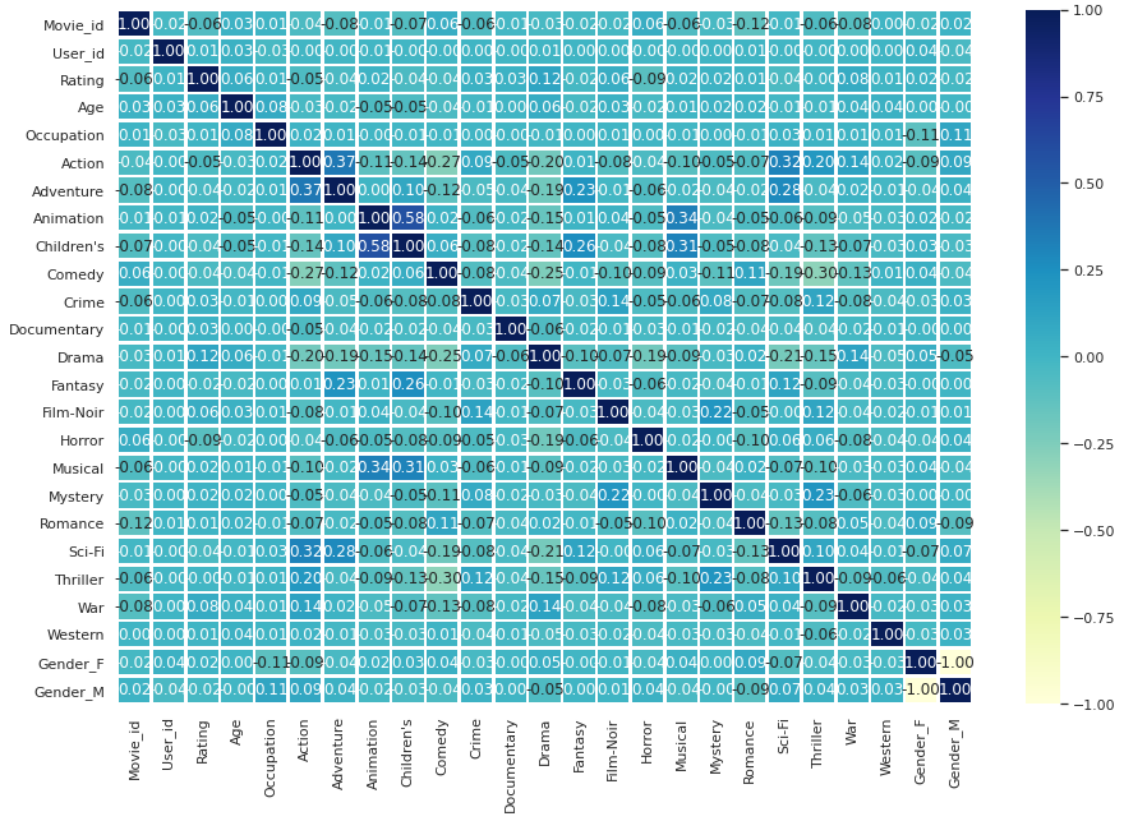
	Animation	Children's	Comedy	…	Horror	Musical	Mystery	\
0	1	1	1	…	0	0	0	
1	1	1	0	…	0	1	0	
2	0	0	0	…	0	0	0	
3	0	0	0	…	0	0	0	
4	0	0	0	…	0	0	0	
…	…	…	…	…	…	…	…	
1000204	0	0	0	…	0	0	0	
1000205	0	0	1	…	1	0	0	
1000206	0	0	1	…	0	0	0	

```
1000207          0           0       0  …       0        0        0
1000208          0           0       0  …       0        0        0

         Romance  Sci-Fi  Thriller  War  Western  Gender_F  Gender_M
0              0       0         0    0        0         1         0
1              1       0         0    0        0         1         0
2              0       0         0    0        0         1         0
3              0       1         0    0        0         1         0
4              0       0         0    1        0         1         0
…            …     …       …  …       …        …         …
1000204        0       0         1    0        0         0         1
1000205        0       0         1    0        0         0         1
1000206        1       0         0    0        0         0         1
1000207        0       0         1    0        0         0         1
1000208        0       0         0    0        0         0         1

[1000209 rows x 25 columns]
```

[106]:
```
movie_rating_genre.dtypes
```

[106]:
```
Movie_id       int64
User_id        int64
Rating         int64
Age            int64
Occupation     int64
Action         int64
Adventure      int64
Animation      int64
Children's     int64
Comedy         int64
Crime          int64
Documentary    int64
Drama          int64
Fantasy        int64
Film-Noir      int64
Horror         int64
Musical        int64
Mystery        int64
Romance        int64
Sci-Fi         int64
Thriller       int64
War            int64
Western        int64
Gender_F       uint8
Gender_M       uint8
dtype: object
```

[107]:
```
f,ax = plt.subplots(figsize=(15,10))
sns.heatmap(movie_rating_genre.corr(), annot = True, linewidths=1.5 , fmt = '.
↪2f',ax=ax,cmap='YlGnBu')
plt.show()
```



From Heatmap we can see the correlation of Ratings with other columns

[108]:
```
movie_rating_genre.describe()
```

[108]:

	Movie_id	User_id	Rating	Age	Occupation	\
count	1.000209e+06	1.000209e+06	1.000209e+06	1.000209e+06	1.000209e+06	
mean	1.865540e+03	3.024512e+03	3.581564e+00	2.973831e+01	8.036138e+00	
std	1.096041e+03	1.728413e+03	1.117102e+00	1.175198e+01	6.531336e+00	
min	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	0.000000e+00	
25%	1.030000e+03	1.506000e+03	3.000000e+00	2.500000e+01	2.000000e+00	
50%	1.835000e+03	3.070000e+03	4.000000e+00	2.500000e+01	7.000000e+00	
75%	2.770000e+03	4.476000e+03	4.000000e+00	3.500000e+01	1.400000e+01	
max	3.952000e+03	6.040000e+03	5.000000e+00	5.600000e+01	2.000000e+01	

	Action	Adventure	Animation	Children's	Comedy	\
count	1.000209e+06	1.000209e+06	1.000209e+06	1.000209e+06	1.000209e+06	

```
mean   2.574032e-01  1.339250e-01  4.328395e-02  7.217092e-02  3.565055e-01
std    4.372036e-01  3.405719e-01  2.034957e-01  2.587708e-01  4.789672e-01
min    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
25%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
50%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
75%    1.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  1.000000e+00
max    1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00

       …        Horror       Musical       Mystery       Romance  \
count  …  1.000209e+06  1.000209e+06  1.000209e+06  1.000209e+06
mean   …  7.637004e-02  4.152432e-02  4.016960e-02  1.474922e-01
std    …  2.655894e-01  1.994996e-01  1.963569e-01  3.545960e-01
min    …  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
25%    …  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
50%    …  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
75%    …  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
max    …  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00

             Sci-Fi      Thriller           War       Western      Gender_F  \
count  1.000209e+06  1.000209e+06  1.000209e+06  1.000209e+06  1.000209e+06
mean   1.572611e-01  1.896404e-01  6.851268e-02  2.067868e-02  2.463885e-01
std    3.640470e-01  3.920166e-01  2.526237e-01  1.423063e-01  4.309076e-01
min    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
25%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
50%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
75%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
max    1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00

          Gender_M
count  1.000209e+06
mean   7.536115e-01
std    4.309076e-01
min    0.000000e+00
25%    1.000000e+00
50%    1.000000e+00
75%    1.000000e+00
max    1.000000e+00

[8 rows x 25 columns]
```

[109]:
```
Y = movie_rating_genre['Rating']
```

[110]:
```
X = movie_rating_genre.drop('Rating',axis=1)
```

[111]:
```
from sklearn.model_selection import train_test_split
```

[112]:
```
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.
↪2,random_state=0)
```

[113]:
```
from sklearn.linear_model import LinearRegression
```

[114]:
```
regressor = LinearRegression()
```

[115]:
```
regressor.fit(X_train,Y_train)
```

[115]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

[116]:
```
Y_pred = regressor.predict(X_test)
```

[117]:
```
regressor.score(X_test,Y_test)
```

[117]:
```
0.040333162013190704
```

[118]:
```
Prediction_df = pd.DataFrame({'TEST':Y_test,'Predicted':Y_pred})
```

[119]:
```
Prediction_df.head()
```

[119]:

	TEST	Predicted
324271	4	3.576581
818637	3	3.703874
148677	5	3.830291
778790	4	3.701040
525489	5	3.681163

[ ]:
```

```