In [1]:
```
import numpy as np
import pandas as pd
```

In [5]:
```
babynames = pd.read_csv('datasets/datasets/babynames/yob1880.txt', names=['name', 'sex
print(babynames)
```

```
           name sex  births
0          Mary   F    7065
1          Anna   F    2604
2          Emma   F    2003
3     Elizabeth   F    1939
4        Minnie   F    1746
...         ...  ..     ...
1995     Woodie   M       5
1996     Worthy   M       5
1997     Wright   M       5
1998       York   M       5
1999  Zachariah   M       5

[2000 rows x 3 columns]
```

In [12]:
```
years = range(1880, 2011)

piv = [pd.read_csv(f'datasets/datasets/babynames/yob{year}.txt', names=['name','sex',

names = pd.concat(piv, ignore_index=True)
print(names)
```

```
              name sex  births  year
0             Mary   F    7065  1880
1             Anna   F    2604  1880
2             Emma   F    2003  1880
3        Elizabeth   F    1939  1880
4           Minnie   F    1746  1880
...            ...  ..     ...   ...
1690779    Zymaire   M       5  2010
1690780     Zyonne   M       5  2010
1690781  Zyquarius   M       5  2010
1690782      Zyran   M       5  2010
1690783      Zzyzx   M       5  2010

[1690784 rows x 4 columns]
```

In [16]:
```
def sumbir(x):
    if(x.births < 10):
        return 1
    elif(x.births >= 10):
        return 0

names['sum'] = names.apply(lambda x: sumbir(x),axis='columns')
```

In [17]:
```
df = names.pivot_table(values='sum', index='year', columns='sex', aggfunc=sum)
print(df)
```

```
sex      F     M
year
1880   363   419
1881   357   403
1882   395   448
1883   403   387
1884   442   463
```

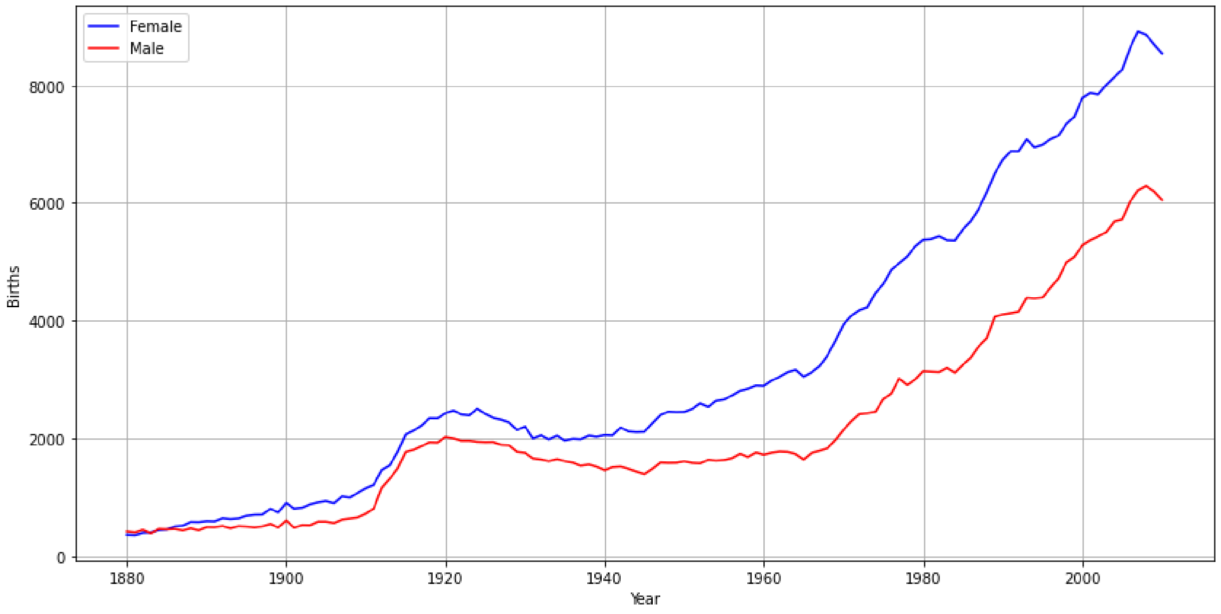
```
...    ...   ...
2006   8634  6024
2007   8916  6213
2008   8859  6287
2009   8695  6197
2010   8541  6052

[131 rows x 2 columns]
```

In [18]:

```
import matplotlib.pyplot as plt

plt.figure(figsize=(14, 7))
plt.plot(df.index, df['F'], label='Female', color='blue')
plt.plot(df.index, df['M'], label='Male', color='red')
plt.xlabel('Year')
plt.ylabel('Births')
plt.legend()
plt.grid(True)
plt.show()
```



In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [3]:

```
pd.options.display.float_format = '{:.4f}'.format

unames = ['user_id', 'gender', 'age', 'occupation', 'zip']
users = pd.read_table('datasets/datasets/movielens/users.dat', sep='::', header=None,

rnames = ['user_id', 'movie_id', 'rating', 'timestamp']
ratings = pd.read_table('datasets/datasets/movielens/ratings.dat', sep='::', header=N

mnames = ['movie_id', 'title', 'genres']
movies = pd.read_table('datasets/datasets/movielens/movies.dat', sep='::', header=Non
```

```
<ipython-input-3-0382b8b996bd>:4: ParserWarning: Falling back to the 'python' engine b
ecause the 'c' engine does not support regex separators (separators > 1 char and diffe
rent from '₩s+' are interpreted as regex); you can avoid this warning by specifying en
gine='python'.
  users = pd.read_table('datasets/datasets/movielens/users.dat', sep='::', header=Non
e, names=unames)
<ipython-input-3-0382b8b996bd>:7: ParserWarning: Falling back to the 'python' engine b
```

```
ecause the 'c' engine does not support regex separators (separators > 1 char and diffe
rent from '₩s+' are interpreted as regex); you can avoid this warning by specifying en
gine='python'.
  ratings = pd.read_table('datasets/datasets/movielens/ratings.dat', sep='::', header=
None, names=rnames)
<ipython-input-3-0382b8b996bd>:10: ParserWarning: Falling back to the 'python' engine
because the 'c' engine does not support regex separators (separators > 1 char and diff
erent from '₩s+' are interpreted as regex); you can avoid this warning by specifying e
ngine='python'.
  movies = pd.read_table('datasets/datasets/movielens/movies.dat', sep='::', header=No
ne, names=mnames)
```

In [4]:
```
data = pd.merge(pd.merge(ratings, users), movies)
print(data)
```

```
         user_id  movie_id  rating  timestamp gender  age  occupation    zip  ₩
0              1      1193       5  978300760      F    1          10  48067
1              2      1193       5  978298413      M   56          16  70072
2             12      1193       4  978220179      M   25          12  32793
3             15      1193       4  978199279      M   25           7  22903
4             17      1193       5  978158471      M   50           1  95350
...          ...       ...     ...        ...    ...  ...         ...    ...
1000204     5949      2198       5  958846401      M   18          17  47901
1000205     5675      2703       3  976029116      M   35          14  30030
1000206     5780      2845       1  958153068      M   18          17  92886
1000207     5851      3607       5  957756608      F   18          20  55410
1000208     5938      2909       4  957273353      M   25           1  35401

                                          title                genres
0             One Flew Over the Cuckoo's Nest (1975)                 Drama
1             One Flew Over the Cuckoo's Nest (1975)                 Drama
2             One Flew Over the Cuckoo's Nest (1975)                 Drama
3             One Flew Over the Cuckoo's Nest (1975)                 Drama
4             One Flew Over the Cuckoo's Nest (1975)                 Drama
...                                          ...                   ...
1000204                        Modulations (1998)           Documentary
1000205                     Broken Vessels (1998)                 Drama
1000206                         White Boys (1999)                 Drama
1000207                   One Little Indian (1973)  Comedy|Drama|Western
1000208  Five Wives, Three Secretaries and Me (1998)           Documentary

[1000209 rows x 10 columns]
```

In [5]:
```
mean = data.pivot_table('rating', index='title', columns='gender', aggfunc='mean')
print(mean)
```

```
gender                                          F       M
title
$1,000,000 Duck (1971)                     3.3750  2.7619
'Night Mother (1986)                       3.3889  3.3529
'Til There Was You (1997)                  2.6757  2.7333
'burbs, The (1989)                         2.7935  2.9621
...And Justice for All (1979)              3.8286  3.6890
...                                           ...     ...
Zed & Two Noughts, A (1985)                3.5000  3.3810
Zero Effect (1998)                         3.8644  3.7231
Zero Kelvin (Kj�rlighetens kj�tere) (1995)    NaN  3.5000
Zeus and Roxanne (1997)                    2.7778  2.3571
eXistenZ (1999)                            3.0986  3.2891

[3706 rows x 2 columns]
```

In [6]:
```
mean['diff'] = abs(mean['M'] - mean['F'])
print(mean)
```

```
gender                                           F      M   diff
title
$1,000,000 Duck (1971)                      3.3750 2.7619 0.6131
'Night Mother (1986)                        3.3889 3.3529 0.0359
'Til There Was You (1997)                   2.6757 2.7333 0.0577
'burbs, The (1989)                          2.7935 2.9621 0.1686
...And Justice for All (1979)               3.8286 3.6890 0.1395
...                                            ...    ...    ...
Zed & Two Noughts, A (1985)                 3.5000 3.3810 0.1190
Zero Effect (1998)                          3.8644 3.7231 0.1413
Zero Kelvin (Kj�rlighetens kj�tere) (1995)    NaN 3.5000    NaN
Zeus and Roxanne (1997)                     2.7778 2.3571 0.4206
eXistenZ (1999)                             3.0986 3.2891 0.1905

[3706 rows x 3 columns]
```

In [7]:

```
mean2 = mean.sort_values(by='diff', ascending=False)
print(mean2)
```

```
gender                                                F      M   diff
title
Tigrero: A Film That Was Never Made (1994)       1.0000 4.3333 3.3333
Spiders, The (Die Spinnen, 1. Teil: Der Goldene... 4.0000 1.0000 3.0000
Neon Bible, The (1995)                           1.0000 4.0000 3.0000
James Dean Story, The (1957)                     4.0000 1.0000 3.0000
Country Life (1994)                              5.0000 2.0000 3.0000
...                                                 ...    ...    ...
With Friends Like These... (1998)                   NaN 4.0000    NaN
Wooden Man's Bride, The (Wu Kui) (1994)             NaN 3.0000    NaN
Year of the Horse (1997)                            NaN 3.2500    NaN
Zachariah (1971)                                    NaN 3.5000    NaN
Zero Kelvin (Kj�rlighetens kj�tere) (1995)          NaN 3.5000    NaN

[3706 rows x 3 columns]
```

In [10]:

```
print(mean2[:10])
# mean2.head(10)
```

```
gender                                                F      M   diff
title
Tigrero: A Film That Was Never Made (1994)       1.0000 4.3333 3.3333
Spiders, The (Die Spinnen, 1. Teil: Der Goldene... 4.0000 1.0000 3.0000
Neon Bible, The (1995)                           1.0000 4.0000 3.0000
James Dean Story, The (1957)                     4.0000 1.0000 3.0000
Country Life (1994)                              5.0000 2.0000 3.0000
Enfer, L' (1994)                                 1.0000 3.7500 2.7500
Babyfever (1994)                                 3.6667 1.0000 2.6667
Stalingrad (1993)                                1.0000 3.5938 2.5938
Woman of Paris, A (1923)                         5.0000 2.4286 2.5714
Cobra (1925)                                     4.0000 1.5000 2.5000
```

In [11]:

```
rating = data.groupby('title').size()
print(rating)
```

```
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
                                             ...
Zed & Two Noughts, A (1985)                    29
Zero Effect (1998)                            301
Zero Kelvin (Kj�rlighetens kj�tere) (1995)      2
Zeus and Roxanne (1997)                        23
```

```
eXistenZ (1999)                                   410
Length: 3706, dtype: int64
```

In [12]:

```
rating2 = rating.index[rating >= 100]
print(rating2)
```

```
Index([''burbs, The (1989)', '...And Justice for All (1979)',
       '10 Things I Hate About You (1999)', '101 Dalmatians (1961)',
       '101 Dalmatians (1996)', '12 Angry Men (1957)',
       '13th Warrior, The (1999)', '2 Days in the Valley (1996)',
       '20 Dates (1998)', '20,000 Leagues Under the Sea (1954)',
       ...
       'Yellow Submarine (1968)', 'Yojimbo (1961)', 'You've Got Mail (1998)',
       'Young Frankenstein (1974)', 'Young Guns (1988)',
       'Young Guns II (1990)', 'Young Sherlock Holmes (1985)',
       'Your Friends and Neighbors (1998)', 'Zero Effect (1998)',
       'eXistenZ (1999)'],
      dtype='object', name='title', length=2019)
```

In [18]:

```
total_movies = len(movies.title)
rating3 = (len(rating2) / total_movies) * 100
print(rating3)
```

```
51.995879474633014
```

In [22]:

```
final = data[['timestamp', 'rating']]
final
```

Out[22]:

	timestamp	rating
0	978300760	5
1	978298413	5
2	978220179	4
3	978199279	4
4	978158471	5
...	...	...
1000204	958846401	5
1000205	976029116	3
1000206	958153068	1
1000207	957756608	5
1000208	957273353	4

1000209 rows × 2 columns

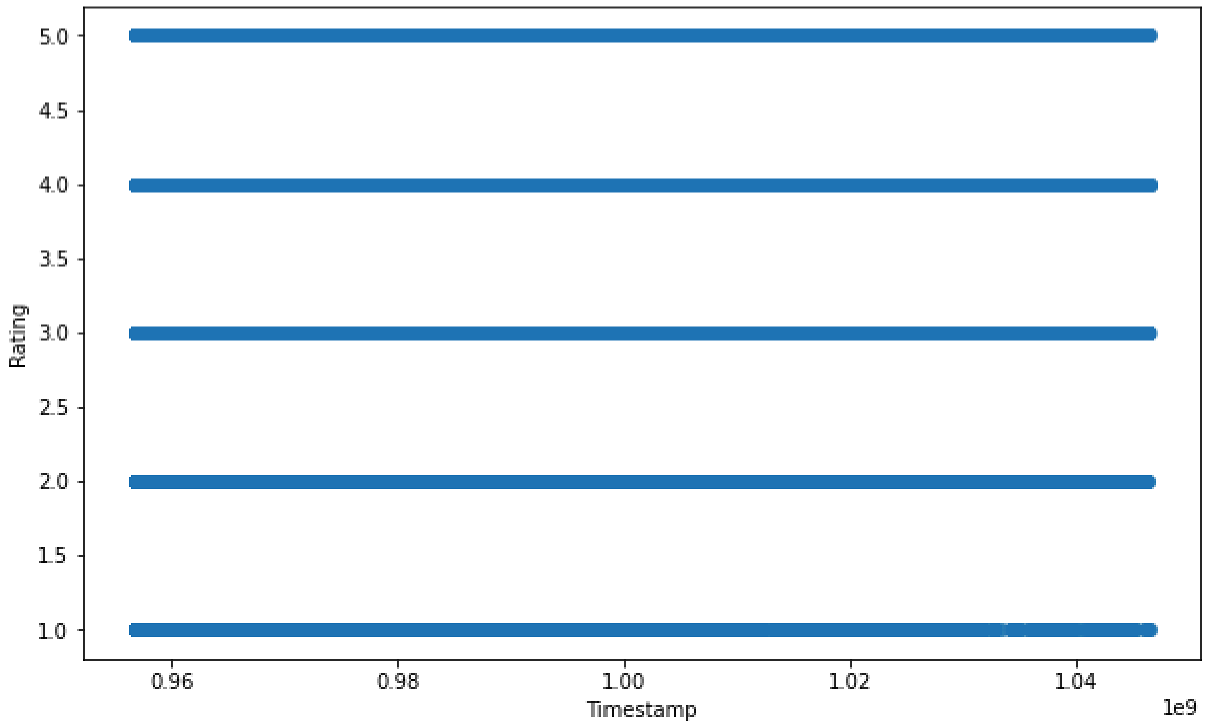
In [26]:

```
corr = data['rating'].corr(data['timestamp'])
print('상관관계: %.4f' %(corr))
```

```
상관관계: -0.0268
```

In [30]:

```
plt.figure(figsize=(10, 6))
plt.scatter(data['timestamp'], data['rating'], alpha=0.5)
plt.xlabel('Timestamp')
plt.ylabel('Rating')
plt.show()
```



```
In [ ]:
```