Movie Recommender System

Importimg Libraries

In [226]:

```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from ast import literal_eval
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
from nltk.stem.snowball import SnowballStemmer
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from surprise import Reader, Dataset, SVD
from surprise.model_selection import cross_validate

import warnings; warnings.simplefilter('ignore')
```

In [227]:

```
md = pd.read_csv('D:/Projects/movies_metadata.csv')
md.head()
```

Out[227]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	o
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	tt0114709	en	
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	tt0113497	en	
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	tt0113228	en	
3	False	NaN	16000000	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...	NaN	31357	tt0114885	en	
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[{'id': 35, 'name': 'Comedy'}]	NaN	11862	tt0113041	en	

5 rows × 24 columns

5 rows × 24 columns

In [228]:

```
# Preprocessing
```

In [229]:

```
md['genres'] = md['genres'].fillna('[]')
```

In [230]:

```
md.head(100)
```

Out[230]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	tt0114709
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	tt0113497
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	tt0113228
3	False	NaN	16000000	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...	NaN	31357	tt0114885
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[{'id': 35, 'name': 'Comedy'}]	NaN	11862	tt0113041
...	...	...	...	...	...	...	...
95	False	NaN	3000000	[{'id': 18, 'name': 'Drama'}]	NaN	406	tt0113247
96	False	NaN	0	[{'id': 28, 'name': 'Action'}, {'id': 12, 'nam...	NaN	45549	tt0111173
97	False	NaN	0	[{'id': 99, 'name': 'Documentary'}]	http://www.nickbroomfield.com/heidifleiss.html	63076	tt0113283
98	False	NaN	0	[{'id': 18, 'name': 'Drama'}, {'id': 53, 'name...	NaN	11062	tt0115907
				[{'id': 35, 'name':			

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id
99	False	NaN	7000000	'Comedy'}, {'id': 80, 'nam...	NaN	13685	tt0115734

100 rows × 24 columns

In [231]:

```
# Eval Example
```

In [232]:

```
list1 = '[1,2,3,4]'
```

In [233]:

```
list1
```

Out[233]:

```
'[1,2,3,4]'
```

In [234]:

```
list1[0]
```

Out[234]:

```
'['
```

In [235]:

```
list_eval = eval(list1)
```

In [236]:

```
list_eval
```

Out[236]:

```
[1, 2, 3, 4]
```

In [237]:

```
list_eval[0]
```

Out[237]:

```
1
```

In [238]:

```
literal_eval(list1)[0]
```

Out[238]:

```
1
```

In [239]:

```
md['genres'] = md['genres'].apply(literal_eval)
```

In [240]:

```
md.head()
```

Out[240]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	o

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	o
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	tt0114709	en	
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	tt0113497	en	
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	tt0113228	en	
3	False	NaN	16000000	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...	NaN	31357	tt0114885	en	
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[{'id': 35, 'name': 'Comedy'}]	NaN	11862	tt0113041	en	

5 rows × 24 columns

In [241]:

```
# Genres as List
```

In [242]:

```
md['genres'] = md['genres'].apply(lambda x: [i['name'] for i in x]if isinstance(x, list)
else [])
```

In [243]:

```
md.head()
```

Out[243]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	ori
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[Animation, Comedy, Family]	http://toystory.disney.com/toy-story	862	tt0114709	en	
1	False	NaN	65000000	[Adventure, Fantasy, Family]	NaN	8844	tt0113497	en	
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[Romance, Comedy]	NaN	15602	tt0113228	en	

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	ori
3	False	NaN	16000000	[Comedy, Drama, Romance]	NaN	31357	tt0114885	en	\
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[Comedy]	NaN	11862	tt0113041	en	

5 rows × 24 columns

In [244]:

```
md[md['vote_count'].notnull()]
```

Out[244]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_la
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[Animation, Comedy, Family]	http://toystory.disney.com/toy-story	862	tt0114709	
1	False	NaN	65000000	[Adventure, Fantasy, Family]	NaN	8844	tt0113497	
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[Romance, Comedy]	NaN	15602	tt0113228	
3	False	NaN	16000000	[Comedy, Drama, Romance]	NaN	31357	tt0114885	
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[Comedy]	NaN	11862	tt0113041	
...	...	...	...	...	...	...	...	
45461	False	NaN	0	[Drama, Family]	http://www.imdb.com/title/tt6209470/	439050	tt6209470	
45462	False	NaN	0	[Drama]	NaN	111109	tt2028550	
45463	False	NaN	0	[Action, Drama, Thriller]	NaN	67758	tt0303758	
45464	False	NaN	0	[]	NaN	227506	tt0008536	
45465	False	NaN	0	[]	NaN	461257	tt6980792	

45460 rows × 24 columns

In [245]:

```
vote_count = md[md['vote_count'].notnull()]['vote_count'].astype('int')
vote_count
```

Out[245]:

```
0        5415
1        2413
2          92
3          34
4         173
         ...
45461       1
45462       3
45463       6
45464       0
45465       0
Name: vote_count, Length: 45460, dtype: int32
```

In [246]:

```
vote_average = md[md['vote_average'].notnull()]['vote_average'].astype('int')
vote_average
```

Out[246]:

```
0        7
1        6
2        6
3        6
4        5
        ..
45461    4
45462    9
45463    3
45464    0
45465    0
Name: vote_average, Length: 45460, dtype: int32
```

In [247]:

```
top_movies = md.copy()
```

In [248]:

```
top_movies1 = top_movies.sort_values('vote_average', ascending = False).head(250)
```

In [249]:

```
# No Min otes Requirements
```

In [250]:

```
top_movies1
```

Out[250]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	origina
21642	False	NaN	0	[Documentary]	NaN	320849	tt0886500	
15710	False	NaN	0	[Documentary]	NaN	96451	tt1587373	

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	origina
22396	False	NaN	0	[Documentary]	NaN	72123	tt1341746	
22395	False	NaN	0	[Documentary]	http://www.marvinhamlischmovie.com/	230864	tt3011874	
35343	False	NaN	300000	[Comedy, Documentary, Music, TV Movie]	NaN	140595	tt0308213	
...	...	...	...	...	...	...	...	
35428	False	NaN	0	[Drama, Romance]	http://www.kuraitokorode.com/	206155	tt0872007	
24882	False	NaN	0	[]	NaN	273334	tt3575800	
4183	False	NaN	0	[Thriller, Drama]	NaN	88727	tt0157411	
36345	False	NaN	0	[History, Drama]	NaN	359154	tt4699592	
30212	False	NaN	0	[Documentary]	NaN	300179	tt3402078	

250 rows × 24 columns

In [251]:

```
# Min number of votes 1000
```

In [252]:

```
top_movies2 = top_movies[top_movies['vote_count']>1000]
```

In [253]:

```
top_movies2
```

Out[253]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[Animation, Comedy, Family]	http://toystory.disney.com/toy-story	862	tt01147
1	False	NaN	65000000	[Adventure, Fantasy,	NaN	8844	tt01134

Family]

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_
5	False	NaN	60000000	[Action, Crime, Drama, Thriller]	NaN	949	tt01132
9	False	{'id': 645, 'name': 'James Bond Collection', '...	58000000	[Adventure, Action, Thriller]	http://www.mgm.com/view/movie/757/Goldeneye/	710	tt01131
15	False	NaN	52000000	[Drama, Crime]	NaN	524	tt01126
...	...	...	...	...	...	...	
43644	False	NaN	34000000	[Action, Crime]	NaN	339403	tt38901
44009	False	{'id': 86066, 'name': 'Despicable Me Collectio...	80000000	[Action, Animation, Adventure, Family, Comedy]	http://www.despicable.me	324852	tt34690
44274	False	{'id': 173710, 'name': 'Planet of the Apes (Re...	152000000	[Drama, Science Fiction, War]	http://www.foxmovies.com/movies/war-for-the-pl...	281338	tt34509
44678	False	NaN	100000000	[Action, Drama, History, Thriller, War]	http://www.dunkirkmovie.com/	374720	tt50130
44842	False	{'id': 8650, 'name': 'Transformers Collection'...	260000000	[Action, Science Fiction, Thriller, Adventure]	http://www.transformersmovie.com/	335988	tt33713

1120 rows × 24 columns

In [254]:

```
top_movies2.sort_values('vote_average', ascending = False).head(250)
```

Out[254]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdl
314	False	NaN	25000000	[Drama, Crime]	NaN	278	tt0111
40251	False	NaN	0	[Romance, Animation, Drama]	https://www.funimationfilms.com/movie/yourname/	372058	tt5311

	adult	belongs_to_collection	budget	genres	homepage	id	imdl
834	False	{'id': 230, 'name': 'The Godfather Collection'...	6000000	[Drama, Crime]	http://www.thegodfather.com/	238	tt0068
1152	False	NaN	3000000	[Drama]	NaN	510	tt0073
1176	False	{'id': 119674, 'name': 'Psycho Collection', 'p...	806948	[Drama, Horror, Thriller]	NaN	539	tt0054
...	...	...	...	...	...	...	
11353	False	NaN	76000000	[Drama, Action, Thriller, Science Fiction]	http://www.universalstudiosentertainment.com/c...	9693	tt0206
24241	False	NaN	58800000	[War, Action]	http://www.americansnipermovie.com	190859	tt2179
13893	False	{'id': 1241, 'name': 'Harry Potter Collection'...	250000000	[Adventure, Fantasy, Family]	http://harrypotter.warnerbros.com/harrypottera...	767	tt0417
23561	False	NaN	18000000	[Drama, Comedy]	NaN	194662	tt2562
12368	False	NaN	15000000	[Comedy, Drama, Crime]	http://www.filminfocus.com/film/in_bruges	8321	tt0780

250 rows × 24 columns

In [255]:

```
vote_count = md[md['vote_count'].notnull()]['vote_count'].astype('int')
vote_averages = md[md['vote_average'].notnull()]['vote_average'].astype('int')
C = vote_averages.mean()
C
```

Out[255]:

```
5.244896612406511
```

In [256]:

```
m = vote_count.quantile(0.95)
m
```

Out[256]:

```
434.0
```

In [257]:

```
md['year'] = pd.to_datetime(md['release_date'], errors = 'coerce').apply(lambda x: str(x
).split('-')[0] if x != np.nan else np.nan)
```

In [258]:

```
top_movies
```

Out[258]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_la
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[Animation, Comedy, Family]	http://toystory.disney.com/toy-story	862	tt0114709	
1	False	NaN	65000000	[Adventure, Fantasy, Family]	NaN	8844	tt0113497	
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[Romance, Comedy]	NaN	15602	tt0113228	
3	False	NaN	16000000	[Comedy, Drama, Romance]	NaN	31357	tt0114885	
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[Comedy]	NaN	11862	tt0113041	
...	...	...	...	...	...	...	...	
45461	False	NaN	0	[Drama, Family]	http://www.imdb.com/title/tt6209470/	439050	tt6209470	
45462	False	NaN	0	[Drama]	NaN	111109	tt2028550	
45463	False	NaN	0	[Action, Drama, Thriller]	NaN	67758	tt0303758	
45464	False	NaN	0	[]	NaN	227506	tt0008536	
45465	False	NaN	0	[]	NaN	461257	tt6980792	

45466 rows × 24 columns

In [259]:

```
top_movies3 = top_movies[(top_movies['vote_count'] >= m) & (top_movies['vote_count'].not
null()) & (top_movies['vote_average'].notnull())][['title', 'vote_count', 'vote_average'
, 'popularity', 'genres']]
top_movies3['vote_count'] = top_movies3['vote_count'].astype('int')
top_movies3['vote_average'] = top_movies3['vote_average'].astype('int')
top_movies3.shape
```

Out[259]:

```
(2274, 5)
```

In [260]:

```
def weighted_rating(x):
    v = x['vote_count']
    R = x['vote_average']
    return (v/(v+m) * R) + (m/(m+v) * C)
```

In [261]:

```
top_movies3['weight_rate'] = top_movies3.apply(weighted_rating, axis = 1)
```

In [262]:

```
top_movies3.head()
```

Out[262]:

	title	vote_count	vote_average	popularity	genres	weight_rate
0	Toy Story	5415	7	21.9469	[Animation, Comedy, Family]	6.869770
1	Jumanji	2413	6	17.0155	[Adventure, Fantasy, Family]	5.884891
5	Heat	1886	7	17.9249	[Action, Crime, Drama, Thriller]	6.671675
9	GoldenEye	1194	6	14.686	[Adventure, Action, Thriller]	5.798701
15	Casino	1343	7	10.1374	[Drama, Crime]	6.571348

In [263]:

```
top_movies3 = top_movies3.sort_values('weight_rate', ascending = False).head(10)
```

In [264]:

```
top_movies3.head(10)
```

Out[264]:

	title	vote_count	vote_average	popularity	genres	weight_rate
15480	Inception	14075	8	29.1081	[Action, Thriller, Science Fiction, Mystery, A...	7.917588
12481	The Dark Knight	12269	8	123.167	[Drama, Action, Crime, Thriller]	7.905871
22879	Interstellar	11187	8	32.2135	[Adventure, Drama, Science Fiction]	7.897107
2843	Fight Club	9678	8	63.8696	[Drama]	7.881753
4863	The Lord of the Rings: The Fellowship of the Ring	8892	8	32.0707	[Adventure, Fantasy, Action]	7.871787
292	Pulp Fiction	8670	8	140.95	[Thriller, Crime]	7.868660
314	The Shawshank Redemption	8358	8	51.6454	[Drama, Crime]	7.864000
7000	The Lord of the Rings: The Return of the King	8226	8	29.3244	[Adventure, Fantasy, Action]	7.861927
351	Forrest Gump	8147	8	48.3072	[Comedy, Drama, Romance]	7.860656
5814	The Lord of the Rings: The Two Towers	7641	8	29.4235	[Adventure, Fantasy, Action]	7.851924

Top Movies

In [265]:

```
# Genre = Romance
```

In [266]:

```
genre_TM = top_movies.apply(lambda x: pd.Series(x['genres']), axis=1).stack().reset_inde
x(level=1, drop=True)
genre_TM.name = 'genre'
genre_top_movies = top_movies.drop('genres', axis=1).join(genre_TM)
```

In [267]:

```
genre_top_movies
```

Out[267]:

	adult	belongs_to_collection	budget	homepage	id	imdb_id	original_language	original_ti
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Sto
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Sto
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Sto
1	False	NaN	65000000	NaN	8844	tt0113497	en	Juma
1	False	NaN	65000000	NaN	8844	tt0113497	en	Juma
...	...	...	...	...	...	...	...	
45463	False	NaN	0	NaN	67758	tt0303758	en	Betray
45463	False	NaN	0	NaN	67758	tt0303758	en	Betray
45463	False	NaN	0	NaN	67758	tt0303758	en	Betray
45464	False	NaN	0	NaN	227506	tt0008536	en	Sata likuyushcl
45465	False	NaN	0	NaN	461257	tt6980792	en	Queeran

93548 rows × 24 columns

In [268]:

```
def build_chart(genre, percentile = 0.85):
    df = genre_top_movies[genre_top_movies['genre'] == genre]
    vote_counts = df[df['vote_count'].notnull()]['vote_count'].astype('int')
    vote_averages = df[df['vote_average'].notnull()]['vote_average'].astype('int')
    C = vote_averages.mean()
    m = vote_counts.quantile(percentile)

    #top_movies3 = top_movies[(top_movies['vote_count'] >= m) & (top_movies['vote_count'].notnull()) & (top_movies['vote_average'].notnull())][['title', 'vote_count', 'vote_average', 'popularity', 'genres']]

    qualified = df[(df['vote_count'] >= m) & (df['vote_count'].notnull()) & (df['vote_average'].notnull())][['title', 'vote_count', 'vote_average', 'popularity']]
    qualified['vote_count'] = qualified['vote_count'].astype('int')
    qualified['vote_average'] = qualified['vote_average'].astype('int')

    qualified['wr'] = qualified.apply(lambda x: (x['vote_count']/(x['vote_count'] + m) * x['vote_average']) + (m/(m+x['vote_count'])*C), axis=1)
    qualified = qualified.sort_values('wr', ascending=False).head(250)

    return qualified
```

Top Genre Movies

In [269]:

```
build_chart('Animation').head(10)
```

Out[269]:

	title	vote_count	vote_average	popularity	wr
359	The Lion King	5520	8	21.6058	7.909339
5481	Spirited Away	3968	8	41.0489	7.875933
9698	Howl's Moving Castle	2049	8	16.136	7.772103
2884	Princess Mononoke	2041	8	17.1667	7.771305
5833	My Neighbor Totoro	1730	8	13.5073	7.735274
40251	Your Name.	1030	8	34.461252	7.589820
5553	Grave of the Fireflies	974	8	0.010902	7.570962
19901	Paperman	734	8	7.19863	7.465676
39386	Piper	487	8	11.243161	7.285132
20779	Wolf Children	483	8	10.2495	7.281198

In [270]:

```
build_chart('Family').head(10)
```

Out[270]:

	title	vote_count	vote_average	popularity	wr
1225	Back to the Future	6239	8	25.7785	7.893053
359	The Lion King	5520	8	21.6058	7.879754
5481	Spirited Away	3968	8	41.0489	7.835635
5833	My Neighbor Totoro	1730	8	13.5073	7.650968
926	It's a Wonderful Life	1103	8	15.0316	7.490637
19901	Paperman	734	8	7.19863	7.301918
39386	Piper	487	8	11.243161	7.071694
20779	Wolf Children	483	8	10.2495	7.066709
31658	Feast	420	8	7.36566	6.980490
25044	Song of the Sea	420	8	6.96736	6.980490

In [271]:

```
build_chart('Action').head(10)
```

Out[271]:

	title	vote_count	vote_average	popularity	wr
15480	Inception	14075	8	29.1081	7.955099
12481	The Dark Knight	12269	8	123.167	7.948610
4863	The Lord of the Rings: The Fellowship of the Ring	8892	8	32.0707	7.929579
7000	The Lord of the Rings: The Return of the King	8226	8	29.3244	7.924031
5814	The Lord of the Rings: The Two Towers	7641	8	29.4235	7.918382
256	Star Wars	6778	8	42.1497	7.908327
1154	The Empire Strikes Back	5998	8	19.471	7.896841
4135	Scarface	3017	8	11.2997	7.802046
9430	Oldboy	2000	8	10.6169	7.711649
1910	Seven Samurai	892	8	15.0178	7.426145

Content Based Recommender

In [272]:

```
links_small = pd.read_csv('D:/Projects/links_small.csv')
links_small = links_small[links_small['tmdbId'].notnull()]['tmdbId'].astype('int')
```

In [273]:

```
top_movies = top_movies.drop([19730., 29503, 35587])
```

In [274]:

```
#Check EDA Notebook for how and why I got these indices
top_movies['id'] = top_movies['id'].astype('int')
```

In [275]:

```
top_movies4 = top_movies[top_movies['id'].isin(links_small)]
top_movies4.shape
```

Out[275]:

```
(9099, 24)
```

In [276]:

```
top_movies4.head()
```

Out[276]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	ori
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[Animation, Comedy, Family]	http://toystory.disney.com/toy-story	862	tt0114709	en	
1	False	NaN	65000000	[Adventure, Fantasy, Family]	NaN	8844	tt0113497	en	

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	ori
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[Romance, Comedy]	NaN	15602	tt0113228	en	
3	False	NaN	16000000	[Comedy, Drama, Romance]	NaN	31357	tt0114885	en	\
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[Comedy]	NaN	11862	tt0113041	en	

5 rows × 24 columns

Movie Description Based Recommender

In [277]:

```
top_movies4['tagline'] = top_movies4['tagline'].fillna('')
top_movies4['description'] = top_movies4['overview'] + top_movies4['tagline']
top_movies4['description'] = top_movies4['description'].fillna('')
```

In [278]:

```
tf = TfidfVectorizer(analyzer = 'word',ngram_range = (1, 2), min_df = 0, stop_words = 'english')
tfidf_matrix = tf.fit_transform(top_movies4['description'])
```

In [279]:

```
tfidf_matrix
```

Out[279]:

```
<9099x268124 sparse matrix of type '<class 'numpy.float64'>'
 with 540591 stored elements in Compressed Sparse Row format>
```

In [280]:

```
tfidf_matrix.shape
```

Out[280]:

```
(9099, 268124)
```

Cosine Similarity

In [281]:

```
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)
```

In [282]:

```
cosine_sim
```

Out[282]:

```
array([[1.        , 0.00680476, 0.        , ..., 0.        , 0.00344913,
        0.        ],
       [0.00680476, 1.        , 0.01531062, ..., 0.00357057, 0.00762326,
        0.        ],
       [0.        , 0.01531062, 1.        , ..., 0.        , 0.00286535,
        0.00472155],
       ...,
       [0.        , 0.00357057, 0.        , ..., 1.        , 0.07811616,
        0.        ],
       [0.00344913, 0.00762326, 0.00286535, ..., 0.07811616, 1.        ,
        0.        ],
       [0.        , 0.        , 0.00472155, ..., 0.        , 0.        ,
        1.        ]])
```

In [283]:

```
cosine_sim[0]
```

Out[283]:

```
array([1.        , 0.00680476, 0.        , ..., 0.        , 0.00344913,
       0.        ])
```

In [284]:

```
top_movies = top_movies4.reset_index()
titles = top_movies4['title']
indices = pd.Series(top_movies4.index, index = top_movies4['title'])
```

In [285]:

```
def get_recommendations(title):
    idx = indices[title]
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key = lambda x: x[1], reverse = True)
    sim_scores = sim_scores[1:31]
    movie_indices = [i[0] for i in sim_scores]
    return titles.iloc[movie_indices]
```

In [286]:

```
get_recommendations('GoldenEye').head(10)
```

Out[286]:

```
7330                          Octopussy
2875                   Live and Let Die
5658                      Casino Royale
7333              Never Say Never Again
2873                  For Your Eyes Only
2263                    A View to a Kill
3517        The Man with the Golden Gun
7329                 You Only Live Twice
2874                     Licence to Kill
3511    On Her Majesty's Secret Service
Name: title, dtype: object
```

In [287]:

```
get_recommendations('The Apartment').head()
```

Out[287]:

```
2241                     The Rugrats Movie
7497                      Duck, You Sucker
11238    The Wind That Shakes the Barley
5422                           FearDotCom
5464                          The Yearling
Name: title, dtype: object
```

In [295]:

```
get_recommendations('The Godfather').head(10)
```

Out[295]:

```
2697                         In Too Deep
2147                     Blame It on Rio
7991                       The Snake Pit
4305                 Written on the Wind
338                 The Baby-Sitters Club
3040                       Stuart Little
6875                           Silverado
16621                No Strings Attached
2731     The Adventures of Milo and Otis
24243                               Tusk
Name: title, dtype: object
```