

Popularity Based Recommendation System

In [7]:

```
import numpy as np
import pandas as pd
```

In [9]:

```
ratings = pd.read_csv("D:/RISE - WPU/Internship/archive (2)/ratings.csv")
ratings.head()
```

Out[9]:

	userid	movied	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

In [10]:

```
movies = pd.read_csv("D:/RISE - WPU/Internship/archive (2)/movies.csv")
ratings.head()
```

Out[10]:

	userid	movied	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

In [14]:

```
# Merging both the datasets

ratings_movies = pd.merge(movies, ratings, on = 'movieId')
print(ratings_movies.shape)
ratings_movies

(100836, 6)
```

Out[14]:

	movied	title	genres	userid	rating	timestamp
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1	4.0	964982703
1	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	5	4.0	847434962
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	7	4.5	1106635946
3	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	15	2.5	1510577970
4	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	17	4.5	1305696483
5	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	18	3.5	1455209816
6	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	19	4.0	965705637
7	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	21	3.5	1407618878
8	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	27	3.0	962685262
9	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	31	5.0	850466616
10	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	32	3.0	856736119
11	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	33	3.0	939647444
12	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	40	5.0	832058959
13	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	43	5.0	848993983
14	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	44	3.0	869251860
15	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	45	4.0	951170182
16	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	46	5.0	834787906
17	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	50	3.0	1514238116
18	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	54	3.0	830247330
19	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	57	5.0	965796031
20	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	63	5.0	1443199669
21	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	64	4.0	1161520134
22	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	66	4.0	1104643957
23	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	68	2.5	1158531426
24	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	71	5.0	864737933
25	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	73	4.5	1464196374
26	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	76	0.5	1439165548
27	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	78	4.0	1252575124
28	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	82	2.5	1084467729
29	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	86	4.0	1344082549
...	...	...	...	...	...	...
100806	188301	Ant-Man and the Wasp (2018)	Action Adventure Comedy Fantasy Sci-Fi	596	4.0	1535709468
100807	188675	Dogman (2018)	Crime Drama	318	3.5	1535035470
100808	188751	Mamma Mia: Here We Go Again! (2018)	Comedy Romance	212	4.5	1532362151
100809	188797	Tag (2018)	Comedy	514	4.0	1535941279
100810	188833	The Man Who Killed Don Quixote (2018)	Adventure Comedy Fantasy	318	4.5	1535035971
100811	189043	Boundaries (2018)	Comedy Drama	338	2.5	1530148447
100812	189111	Spiral (2018)	Documentary	338	3.0	1530148343
100813	189333	Mission: Impossible - Fallout (2018)	Action Adventure Thriller	184	4.0	1537110051
100814	189333	Mission: Impossible - Fallout (2018)	Action Adventure Thriller	248	3.5	1534602181
100815	189381	SuperFly (2018)	Action Crime Thriller	318	2.5	1536097988
100816	189547	Iron Soldier (2010)	Action Sci-Fi	210	1.0	1528486011
100817	189713	BlackKkKlansman (2018)	Comedy Crime Drama	462	2.5	1536467299
100818	190183	The Darkest Minds (2018)	Sci-Fi Thriller	50	3.5	1533302021
100819	190207	Tilt (2011)	Drama Romance	338	1.5	1530148481
100820	190209	Jeff Ross Roasts the Border (2017)	Comedy	338	4.0	1530148487
100821	190213	John From (2015)	Drama	338	1.0	1530148478
100822	190215	Liquid Truth (2017)	Drama	338	1.5	1530148477
100823	190219	Bunny (1998)	Animation	338	1.0	1530148473
100824	190221	Hommage à Zgougou (et salut à Sabine Mamou) (2...	Documentary	338	1.0	1530148473
100825	191005	Gintama (2017)	Action Adventure Comedy Sci-Fi	184	4.5	1537109489
100826	193565	Gintama: The Movie (2010)	Action Animation Comedy Sci-Fi	184	3.5	1537098554
100827	193567	anohana: The Flower We Saw That Day - The Movi...	Animation Drama	184	3.0	1537099103
100828	193571	Silver Spoon (2014)	Comedy Drama	184	4.0	1537099392
100829	193573	Love Live! The School Idol Movie (2015)	Animation	184	4.0	1537099811
100830	193579	Jon Stewart Has Left the Building (2015)	Documentary	184	3.5	1537107259
100831	193581	Black Butler: Book of the Atlantic (2017)	Action Animation Comedy Fantasy	184	4.0	1537109082
100832	193583	No Game No Life: Zero (2017)	Animation Comedy Fantasy	184	3.5	1537109545
100833	193585	Flint (2017)	Drama	184	3.5	1537109805
100834	193587	Bungo Stray Dogs: Dead Apple (2018)	Action Animation	184	3.5	1537110021
100835	193609	Andrew Dice Clay: Dice Rules (1991)	Comedy	331	4.0	1537157606

100836 rows × 6 columns

In [15]:

```
# Recommending Popular Movies
# Grouping movies together and finding mean ratings

ratings_movies.groupby('title')['rating'].mean().head()
```

Out[15]:

title	
'71 (2014)	4.0
'Hellboy': The Seeds of Creation (2004)	4.0
'Round Midnight (1986)	3.5
'Salem's Lot (2004)	5.0
'Til There Was You (1997)	4.0
Name: rating, dtype: float64	

In [16]: *# Sorting ratings from highest to Lowest*

```
ratings_movies.groupby('title')['rating'].mean().sort_values(ascending = False)
```

Out[16]:

title	
Karlson Returns (1970)	5.0
Winter in Prostokvashino (1984)	5.0
My Love (2006)	5.0
Sorority House Massacre II (1990)	5.0
Winnie the Pooh and the Day of Concern (1972)	5.0
Sorority House Massacre (1986)	5.0
Bill Hicks: Revelations (1993)	5.0
My Man Godfrey (1957)	5.0
Hellbenders (2012)	5.0
In the blue sea, in the white foam. (1984)	5.0
Won't You Be My Neighbor? (2018)	5.0
Red Sorghum (Hong gao liang) (1987)	5.0
Love Exposure (Ai No Mukidashi) (2008)	5.0
My Sassy Girl (Yeopgijeogin geunyeo) (2001)	5.0
The Love Bug (1997)	5.0
Ballad of Narayama, The (Narayama bushiko) (1983)	5.0
Heidi Fleiss: Hollywood Madam (1995)	5.0
Louis Theroux: Law & Disorder (2008)	5.0
Winnie the Pooh Goes Visiting (1971)	5.0
In the Realm of the Senses (Ai no corrida) (1976)	5.0
Winnie Pooh (1969)	5.0
Ex Drummer (2007)	5.0
Tom Segura: Mostly Stories (2016)	5.0
Tom and Jerry: A Nutcracker Tale (2007)	5.0
A Plasticine Crow (1981)	5.0
Tom and Jerry: Shiver Me Whiskers (2006)	5.0
Cosmic Scrat-tastrophe (2015)	5.0
Delirium (2014)	5.0
Lumberjack Man (2015)	5.0
Loving Vincent (2017)	5.0
	...
Cincinnati Kid, The (1965)	0.5
Wizards of the Lost Kingdom II (1989)	0.5
Begotten (1990)	0.5
Old Dogs (2009)	0.5
Spy Who Came in from the Cold, The (1965)	0.5
Oblivion 2: Backlash (1996)	0.5
Starcrash (a.k.a. Star Crash) (1978)	0.5
Baxter (1989)	0.5
Collector, The (1965)	0.5
Gypsy (1962)	0.5
3 Ninjas Knuckle Up (1995)	0.5
My Bloody Valentine (1981)	0.5
3 dev adam (Three Giant Men) (1973)	0.5
Superfast! (2015)	0.5
Crow, The: Wicked Prayer (2005)	0.5
Survivor (2015)	0.5
Mortal Kombat: The Journey Begins (1995)	0.5
Bad Santa 2 (2016)	0.5
Cyborg (1989)	0.5
Hard Ticket to Hawaii (1987)	0.5
Baby Boy (2001)	0.5
Midnight Chronicles (2009)	0.5
Haunted House 2, A (2014)	0.5
Maria Bamford: The Special Special Special! (2012)	0.5
Dead of Night (1945)	0.5
The Beast of Hollow Mountain (1956)	0.5
Follow Me, Boys! (1966)	0.5
The Butterfly Effect 3: Revelations (2009)	0.5
The Emoji Movie (2017)	0.5
Rust and Bone (De rouille et d'os) (2012)	0.5
Name: rating, Length: 9719, dtype: float64	

In [17]: *# Recommending top 10 popular movies*

```
ratings_movies.groupby('title')['rating'].mean().sort_values(ascending = False).head(10)
```

Out[17]:

title	
Karlson Returns (1970)	5.0
Winter in Prostokvashino (1984)	5.0
My Love (2006)	5.0
Sorority House Massacre II (1990)	5.0
Winnie the Pooh and the Day of Concern (1972)	5.0
Sorority House Massacre (1986)	5.0
Bill Hicks: Revelations (1993)	5.0
My Man Godfrey (1957)	5.0
Hellbenders (2012)	5.0
In the blue sea, in the white foam. (1984)	5.0
Name: rating, dtype: float64	

In [18]: *# Calculating how many people have voted for each movie*

```
ratings_movies['title'].value_counts()
```

Out[18]:

Forrest Gump (1994)	329
Shawshank Redemption, The (1994)	317
Pulp Fiction (1994)	307
Silence of the Lambs, The (1991)	279
Matrix, The (1999)	278
Star Wars: Episode IV - A New Hope (1977)	251
Jurassic Park (1993)	238
Braveheart (1995)	237
Terminator 2: Judgment Day (1991)	224
Schindler's List (1993)	220
Fight Club (1999)	218
Toy Story (1995)	215
Star Wars: Episode V - The Empire Strikes Back (1980)	211
American Beauty (1999)	204
Usual Suspects, The (1995)	204
Seven (a.k.a. Se7en) (1995)	203
Independence Day (a.k.a. ID4) (1996)	202
Apollo 13 (1995)	201
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)	200
Lord of the Rings: The Fellowship of the Ring, The (2001)	198
Star Wars: Episode VI - Return of the Jedi (1983)	196
Godfather, The (1972)	192
Fugitive, The (1993)	190
Batman (1989)	189
Saving Private Ryan (1998)	188
Lord of the Rings: The Two Towers, The (2002)	188
Lord of the Rings: The Return of the King, The (2003)	185
Aladdin (1992)	183
Fargo (1996)	181
Sixth Sense, The (1999)	179
	...
Mambo Italiano (2003)	1
Ben X (2007)	1
Fur: An Imaginary Portrait of Diane Arbus (2006)	1
Shaolin Temple (Shao Lin si) (1976)	1
Company Men, The (2010)	1
UnHung Hero (2013)	1
Tim's Vermeer (2013)	1
Shining Through (1992)	1
Agony and the Ecstasy, The (1965)	1
Beyond the Valley of the Dolls (1970)	1
Children of the Night (1991)	1
Nothing Personal (1995)	1
Eve of Destruction (1991)	1
Wedding Banquet, The (Xi yan) (1993)	1
Love Me Tender (1956)	1
Charlie Countryman (2013)	1
Paisan (Paisà) (1946)	1
Mom's Night Out (2014)	1
Piano Teacher, The (La pianiste) (2001)	1
Skyline (2010)	1
The Good Boy (2016)	1
Unintentional Kidnapping of Mrs. Elfriede Ott, The (Die Unabsichtliche Entführung der Frau Elfriede Ott) (2010)	1
Who'll Stop the Rain (1978)	1
Late Shift, The (1996)	1
Jeff, Who Lives at Home (2012)	1
Cuckoo, The (Kukushka) (2002)	1
Crimson Pirate, The (1952)	1
I.O.U.S.A. (a.k.a. IOUSA) (2008)	1
Killer's Kiss (1955)	1
Johnny Eager (1942)	1
Name: title, Length: 9719, dtype: int64	

In [22]: *# Finding the movie rating and count of users*

```
rating_count['rating_counts'] = pd.DataFrame(ratings_movies.groupby('title')['rating'].count())
rating_count.head()
```

Out[22]:

	rating	rating_counts
title		
'71 (2014)	4.0	1
'Hellboy: The Seeds of Creation (2004)	4.0	1
'Round Midnight (1986)	3.5	2
'Salem's Lot (2004)	5.0	1
'Til There Was You (1997)	4.0	2



```
In [23]: # Calculating cosine similarity

from math import *

def square_rooted(x):
    return round(sqrt(sum([a*a for a in x])),3)

def cosine_similarity(x,y):
    numerator = sum(a*b for a,b in zip(x,y))
    denominator = square_rooted(x) * square_rooted(y)
    return round(numerator/ float(denominator),3)

print(cosine_similarity([3,45,7,2],[2,54,13,15]))

0.972
```

Content Based Recommendation System

```
In [25]: %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

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

```
In [26]: movies_metadata = pd.read_csv('D:/RISE - WPU/Internship/archive (1)/movies_metadata.csv')
movies_metadata.head()
```

Out[26]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	original_title	overview	...	release_date	revenue	runtime	spoken_languages	status	tagline	title	video	vote_average	vote_count
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	{'id': 16, 'name': 'Animation'}, {'id': 35, 'name': 'Adventure'}, {'id': 14, 'name': 'Fantasy'}, {'id': 10, 'name': 'Family'}	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Story	Led by Woody, Andy's toys live happily in his world until...	...	1995-10-30	373554033.0	81.0	[{'iso_639_1': 'en', 'name': 'English'}]	Released		Toy Story	False	7.7	5415.0
1	False	NaN	65000000	{'id': 12, 'name': 'Adventure'}, {'id': 14, 'name': 'Fantasy'}, {'id': 10, 'name': 'Family'}	NaN	8844	tt0113497	en	Jumanji	When siblings Judy and Peter discover an enchanted board game that opens the door to a magical world of...	...	1995-12-15	262797249.0	104.0	[{'iso_639_1': 'en', 'name': 'English'}, {'iso_639_2': 'spa', 'name': 'Spanish'}]	Released	Roll the dice and unleash the excitement!	Jumanji	False	6.9	2413.0
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...', ...	0	{'id': 10749, 'name': 'Romance'}, {'id': 35, 'name': 'Comedy'}, {'id': 18, 'name': 'Drama'}	NaN	15602	tt0113228	en	Grumpier Old Men	A family wedding reignites the ancient feud between two brothers...	...	1995-12-22	0.0	101.0	[{'iso_639_1': 'en', 'name': 'English'}]	Released	Still Yelling. Still Fighting. Still Ready for...	Grumpier Old Men	False	6.5	92.0
3	False	NaN	16000000	{'id': 35, 'name': 'Comedy'}, {'id': 18, 'name': 'Drama'}	NaN	31357	tt0114885	en	Waiting to Exhale	Cheated on, mistreated and stepped on, the women of...	...	1995-12-22	81452156.0	127.0	[{'iso_639_1': 'en', 'name': 'English'}]	Released	Friends are the people who let you be yourself...	Waiting to Exhale	False	6.1	34.0
4	False	{'id': 96871, 'name': 'Father of the Bride Collection', ...	0	{'id': 35, 'name': 'Comedy'}	NaN	11862	tt0113041	en	Father of the Bride Part II	Just when George Banks has recovered from his previous wedding...	...	1995-02-10	76578911.0	106.0	[{'iso_639_1': 'en', 'name': 'English'}]	Released	Just When His World Is Back To Normal... He's Back To Normal...	Father of the Bride Part II	False	5.7	173.0

5 rows × 24 columns

```
In [27]: movies_metadata['genres'] = movies_metadata['genres'].fillna('').apply(literal_eval).apply(lambda x: [i['name'] for i in x] if isinstance(x, list) else [])
```

```
In [28]: vote_counts = movies_metadata[movies_metadata['vote_count'].notnull()]['vote_count'].astype('int')
vote_averages = movies_metadata[movies_metadata['vote_average'].notnull()]['vote_average'].astype('int')
vote_mean = vote_averages.mean()
vote_mean
```

Out[28]: 5.244896612406511

```
In [29]: vote_quantile = vote_counts.quantile(0.95)
vote_quantile
```

Out[29]: 434.0

```
In [30]: movies_metadata['year'] = pd.to_datetime(movies_metadata['release_date'], errors='coerce').apply(lambda x: str(x).split('-')[0] if x != np.nan else np.nan)
```

```
In [31]: qualified = movies_metadata[(movies_metadata['vote_count'] >= vote_quantile) & (movies_metadata['vote_count'].notnull()) & (movies_metadata['vote_average'].notnull())][['title', 'year', 'vote_count', 'vote_average', 'popularity', 'genres']]
qualified['vote_count'] = qualified['vote_count'].astype('int')
qualified['vote_average'] = qualified['vote_average'].astype('int')
qualified.shape
```

Out[31]: (2274, 6)

```
In [ ]: # To qualify to be considered for the chart, a movie has to have at Least 434 votes. We also see that the average rating for a movie on is 5.244 on a scale of 10. 2274 Movies qualify to be on our chart.
```

```
In [33]: def weighted_rating(x):
    cnt = x['vote_count']
    avg = x['vote_average']
    return (cnt/(cnt+vote_quantile) * avg) + (vote_quantile/(vote_quantile+cnt) * vote_mean)
```

```
In [34]: qualified['wr'] = qualified.apply(weighted_rating, axis=1)
```

```
In [35]: qualified = qualified.sort_values('wr', ascending=False).head(250)
```

```
In [37]: qualified.head(10)
```

Out[37]:

	title	year	vote_count	vote_average	popularity	genres	wr
15480	Inception	2010	14075	8	29.1081	[Action, Thriller, Science Fiction, Mystery, Adventure, Drama]	7.917588
12481	The Dark Knight	2008	12269	8	123.167	[Drama, Action, Crime, Thriller]	7.905871
22879	Interstellar	2014	11187	8	32.2135	[Adventure, Drama, Science Fiction]	7.897107
2843	Fight Club	1999	9678	8	63.8696	[Drama]	7.881753
4863	The Lord of the Rings: The Fellowship of the Ring	2001	8892	8	32.0707	[Adventure, Fantasy, Action]	7.871787
292	Pulp Fiction	1994	8670	8	140.95	[Thriller, Crime]	7.868660
314	The Shawshank Redemption	1994	8358	8	51.6454	[Drama, Crime]	7.864000
7000	The Lord of the Rings: The Return of the King	2003	8226	8	29.3244	[Adventure, Fantasy, Action]	7.861927
351	Forrest Gump	1994	8147	8	48.3072	[Comedy, Drama, Romance]	7.860656
5814	The Lord of the Rings: The Two Towers	2002	7641	8	29.4235	[Adventure, Fantasy, Action]	7.851924

```
In [38]: # For particular genres

s = movies_metadata.apply(lambda x: pd.Series(x['genres']),axis=1).stack().reset_index(level=1, drop=True)
s.name = 'genre'
gen_md = movies_metadata.drop('genres', axis=1).join(s)
```

```
In [39]: def build_chart(genre, percentile=0.85):
    df = gen_md[gen_md['genre'] == genre]
    vote_counts = df[df['vote_count'].notnull()]['vote_count'].astype('int')
    vote_averages = df[df['vote_average'].notnull()]['vote_average'].astype('int')
    vote_mean = vote_averages.mean()
    vote_quantile = vote_counts.quantile(percentile)

    qualified = df[(df['vote_count'] >= vote_quantile) & (df['vote_count'].notnull()) & (df['vote_average'].notnull())][['title', 'year', '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']+vote_quantile) * x['vote_average']) + (vote_quantile/(vote_quantile+x['vote_count']) * vote_mean), axis=1)
    qualified = qualified.sort_values('wr', ascending=False).head(250)

    return qualified
```

```
In [41]: build_chart('Adventure').head(10)
```

Out[41]:

	title	year	vote_count	vote_average	popularity	wr
15480	Inception	2010	14075	8	29.1081	7.906526
22879	Interstellar	2014	11187	8	32.2135	7.883426
4863	The Lord of the Rings: The Fellowship of the Ring	2001	8892	8	32.0707	7.854939
7000	The Lord of the Rings: The Return of the King	2003	8226	8	29.3244	7.843867
5814	The Lord of the Rings: The Two Towers	2002	7641	8	29.4235	7.832647
256	Star Wars	1977	6778	8	42.1497	7.812801
1225	Back to the Future	1985	6239	8	25.7785	7.797828
1154	The Empire Strikes Back	1980	5998	8	19.471	7.790329
5481	Spirited Away	2001	3968	8	41.0489	7.695056
9698	Howl's Moving Castle	2004	2049	8	16.136	7.465435

The top radventure movie according to our metrics is Inception.