The simple recommender

Juan David Serna Valderrama

Read .CSV file

In [13]:

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

df = pd.read_csv("C:/Users/juand/OneDrive/Escritorio/Recommendation Systems with Python/Data/movies_metadata.csv")
df.head()
```

```
c:\users\juand\appdata\local\programs\python\python38\lib\site-packages\IPython\core\interactiveshell.py:3145: DtypeWarning: Columns (10) have mixed types.Spe
cify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[13]:

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	original_title	overview	...	release_date	revenue	runtime	spoken_languag
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Story	Led by Woody, Andy's toys live happily in his ...	...	1995-10-30	373554033.0	81.0	[{'iso_639_1': 'e 'name': 'Englis
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	tt0113497	en	Jumanji	When siblings Judy and Peter discover an encha...	...	1995-12-15	262797249.0	104.0	[{'iso_639_1': 'e 'name': 'Englis {'is
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	tt0113228	en	Grumpier Old Men	A family wedding reignites the ancient feud be...	...	1995-12-22	0.0	101.0	[{'iso_639_1': 'e 'name': 'Englis
3	False	NaN	16000000	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...	NaN	31357	tt0114885	en	Waiting to Exhale	Cheated on, mistreated and stepped on, the wom...	...	1995-12-22	81452156.0	127.0	[{'iso_639_1': 'e 'name': 'Englis
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	[{'id': 35, 'name': 'Comedy'}]	NaN	11862	tt0113041	en	Father of the Bride Part II	Just when George Banks has recovered from his ...	...	1995-02-10	76578911.0	106.0	[{'iso_639_1': 'e 'name': 'Englis

5 rows × 24 columns

The steps for building the Simple Recommender System

1. Choose a metric (or score) to rate the movies on

2. Decide on the prerequisites for the movie to be featured on the chart

3. Calculate the score for every movie that satisfies the conditions

4. Output the list of movies in decreasing order of their scores

In [14]:

```python
#Calculate the number of votes garnered by the 80th percentile movie
m = df['vote_count'].quantile(0.80)
m
```

Out[14]:

```
50.0
```

We will only consider movies that are greater than 45 minutes and less than 300 minutes in length.

In [15]:

```python
#Only consider movies longer than 45 minutes and shorter than 300 minutes
q_movies = df[(df['runtime'] >= 45) & (df['runtime']) <= 300]

#Only consider movies that have garnered more than m votes
q_movies = q_movies[q_movies['vote_count'] > m]

q_movies.shape
```

Out[15]:

```
(9048, 24)
```

Calculating the score

In [16]:

```python
# Calculate C
C = df['vote_average'].mean()
C
```

Out[16]:

```
5.618207215134185
```

Define a function that computes the rating for a movie

In [17]:

```python
# Function to compute the IMDB weighted rating for each movie
def weighted_rating(x, m=m, C=C):
    v = x['vote_count']
    R = x['vote_average']

    return (v/(v+m) * R) + (m/(m+v) * C)
```

Compute the score using the weighted_rating function defined above

In [18]:

```python
q_movies['score'] = q_movies.apply(weighted_rating, axis=1)
```

Sort movies in descending order of their scores

In [19]:

```python
q_movies = q_movies.sort_values('score', ascending=False)

#Print the top 25 movies
q_movies[['title', 'vote_count', 'vote_average', 'score', 'runtime']].head(25)
```

Out[19]:

	title	vote_count	vote_average	score	runtime
10309	Dilwale Dulhania Le Jayenge	661.0	9.1	8.855148	190.0
314	The Shawshank Redemption	8358.0	8.5	8.482863	142.0
834	The Godfather	6024.0	8.5	8.476278	175.0
40251	Your Name.	1030.0	8.5	8.366584	106.0
12481	The Dark Knight	12269.0	8.3	8.289115	152.0
2843	Fight Club	9678.0	8.3	8.286216	139.0
292	Pulp Fiction	8670.0	8.3	8.284623	154.0
522	Schindler's List	4436.0	8.3	8.270109	195.0
23673	Whiplash	4376.0	8.3	8.269704	105.0
5481	Spirited Away	3968.0	8.3	8.266628	125.0
2211	Life Is Beautiful	3643.0	8.3	8.263691	116.0
1178	The Godfather: Part II	3418.0	8.3	8.261335	200.0
1152	One Flew Over the Cuckoo's Nest	3001.0	8.3	8.256051	133.0
1176	Psycho	2405.0	8.3	8.245381	109.0
351	Forrest Gump	8147.0	8.2	8.184252	142.0
1184	Once Upon a Time in America	1104.0	8.3	8.183804	229.0
1154	The Empire Strikes Back	5998.0	8.2	8.178656	124.0
18465	The Intouchables	5410.0	8.2	8.176357	112.0
289	Leon: The Professional	4293.0	8.2	8.170276	110.0
3030	The Green Mile	4166.0	8.2	8.169381	189.0
1170	GoodFellas	3211.0	8.2	8.160414	145.0
2216	American History X	3120.0	8.2	8.159278	119.0
1161	12 Angry Men	2130.0	8.2	8.140785	96.0
9698	Howl's Moving Castle	2049.0	8.2	8.138499	119.0
2884	Princess Mononoke	2041.0	8.2	8.138264	134.0

In [ ]: