Data Analytics Assignment 7:

Part 1: Collaborative Filtering

Initial dataset of movies and ratings by users for movies:

```
movies.head()
```

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

```
Ratings.head()
```

	userId	movieId	rating	timestamp
0	12882	1	4.0	1147195252
1	12882	32	3.5	1147195307
2	12882	47	5.0	1147195343
3	12882	50	5.0	1147185499
4	12882	110	4.5	1147195239

Code snippet for recommending top 5 movies for a user:

```
def User_item_score1(user):
    Movie_seen_by_user = check.columns[check[check.index==user].notna().any()].tolist()
    a = sim_user_30_m[sim_user_30_m.index==user].values
    b = a.squeeze().tolist()
    d = Movie_user[Movie_user.index.isin(b)]
    l = ','.join(d.values)
    Movie_seen_by_similar_users = l.split(',')
    Movies_under_consideration = list(set(Movie_seen_by_similar_users)-set(list(map(str, Movie_seen_by_user))))
    Movies_under_consideration = list(map(int, Movies_under_consideration))
    score = []
    for item in Movies_under_consideration:
        c = final_movie.loc[:,item]
        d = c[c.index.isin(b)]
        f = d[d.notnull()]
        avg_user = Mean.loc[Mean['userId'] == user,'rating'].values[0]
        index = f.index.values.squeeze().tolist()
        corr = similarity_with_movie.loc[user,index]
        fin = pd.concat([f, corr], axis=1)
        fin.columns = ['adg_score','correlation']
        fin['score']=fin.apply(lambda x:x['adg_score'] * x['correlation'],axis=1)
        nume = fin['score'].sum()
        deno = fin['correlation'].sum()
        final_score = avg_user + (nume/deno)
        score.append(final_score)
    data = pd.DataFrame({'movieId':Movies_under_consideration,'score':score})
    top_5_recommendation = data.sort_values(by='score',ascending=False).head(5)
    Movie_Name = top_5_recommendation.merge(movies, how='inner', on='movieId')
    Movie_Names = Movie_Name.title.values.tolist()
    return Movie_Names
```

Result:

```
Enter the user id to whom you want to recommend : 370

The Recommendations for User Id :  370

Band of Brothers (2001)
Godfather: Part II, The (1974)
Wallace & Gromit: The Wrong Trousers (1993)
Bicycle Thieves (a.k.a. The Bicycle Thief) (a.k.a. The Bicycle Thieves) (Ladri di biciclette) (1948)
Spirited Away (Sen to Chihiro no kamikakushi) (2001)
```

Part 2: Content Based System

Initial dataset:

```
df = pd.read_csv('u.data', sep='\t', names=['user_id','item_id','rating','titmestamp'])
df.head()
```

	user_id	item_id	rating	titmestamp
0	0	50	5	881250949
1	0	172	5	881250949
2	0	133	1	881250949
3	196	242	3	881250949
4	186	302	3	891717742

```
movie_titles = pd.read_csv('Movie_Titles')
movie_titles.head()
```

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

Calculation of correlation between movies:

```
similar_to_air_force_one=movie_matrix.corrwith(AFO_user_rating)
similar_to_air_force_one.head()
```

```
title
'Til There Was You (1997)    0.867722
1-900 (1994)                      NaN
101 Dalmatians (1996)        0.221943
12 Angry Men (1957)          0.228031
187 (1997)                   0.294232
dtype: float64
```

```
similar_to_contact = movie_matrix.corrwith(contact_user_rating)
similar_to_contact.head()
```

```
title
'Til There Was You (1997)    0.904534
1-900 (1994)                      NaN
101 Dalmatians (1996)       -0.108441
12 Angry Men (1957)          0.022265
187 (1997)                   0.135512
dtype: float64
```

Result:

```
corr_AFO[corr_AFO['number_of_ratings'] > 100].sort_values(by='correlation', ascending=False).head(5)
```

title	correlation	number_of_ratings
Air Force One (1997)	1.000000	431
Hunt for Red October, The (1990)	0.554383	227
Firm, The (1993)	0.526743	151
Murder at 1600 (1997)	0.514906	218
Eraser (1996)	0.500606	206

```
corr_contact[corr_contact['number_of_ratings'] > 100].sort_values(by='Correlation', ascending=False).head(5)
```

title	Correlation	number_of_ratings
Contact (1997)	1.000000	509
Philadelphia (1993)	0.446509	137
Mask, The (1994)	0.418328	129
Young Guns (1988)	0.388839	101
Sling Blade (1996)	0.384840	136