

```
In [1]: #importing the libraries
import numpy as np
import pandas as pd
```

```
In [2]: # Load in sub-dfs
#User data
df = pd.read_csv('udata.txt', sep='\t', header=None)
df.columns = ['user_id', 'item_id', 'rating', 'timestamp']

#Users
users = pd.read_csv('uuser.txt', sep='|', header=None)
users.columns = ['user_id', 'age', 'gender', 'occupation', 'zip_code']

#genre data
genre = pd.read_csv('ugenre.txt', sep='|', header=None)
genre.columns = ['genre', 'genre_id']

#occupation data
job = pd.read_csv('uoccupation.txt', sep='|', header=None)
job.columns = ['occupation']
job = job.reset_index().rename(columns={'index': 'job'})

#other items
items = pd.read_csv('uitem.txt', sep='|', header=None, encoding='latin-1')
items.columns = ['movie_id', 'movie_title', 'release_date', 'video_release_date',
                 'imdb_url', 'unknown', 'Action', 'Adventure', 'Animation', 'Children',
                 'Comedy', 'Crime', 'Documentary', 'Drama', 'Fantasy', 'Film_Noir',
                 'Horror', 'Musical', 'Mystery', 'Romance', 'Sci-Fi',
                 'Thriller', 'War', 'Western']

#Combine/clean data
data = df.merge(
    users, how='left', on='user_id').merge(
    job, how='left', on='occupation').merge(
    items, how='left', left_on='item_id', right_on='movie_id')
data[data.Western.isna()]
data = data[data.Western.notna()].astype({'Western': 'int'})
data.gender.unique()
data['gender_id'] = np.where(data.gender == 'M', 0, 1)

data.head()
```

```
Out[2]:
```

	user_id	item_id	rating	timestamp	age	gender	occupation	zip_code	job	movie_id	...	Film_Noir	Horror	Musical	Mystery	Ro
0	196	242	3	881250949	49	M	writer	55105	20	242	...	0	0	0	0	
1	186	302	3	891717742	39	F	executive	00000	6	302	...	1	0	0	1	
2	22	377	1	878887116	25	M	writer	40206	20	377	...	0	0	0	0	
3	244	51	2	880606923	28	M	technician	80525	19	51	...	0	0	0	0	
4	166	346	1	886397596	47	M	educator	55113	3	346	...	0	0	0	0	

```
In [3]: pivot_data = pd.pivot_table(data, values='rating', columns='movie_title', index='user_id')
pivot_data.drop(['unknown'], inplace=True, axis=1)
data2 = pivot_data.iloc[:, :500]
data2.shape, data.shape
```

```
Out[3]: ((943, 500), (100000, 34))
```

```
In [4]: print("-----")
print("List of 10 recommended movies to a user who has liked '2001: A Space Odyssey'")
print(data2.corr(method='pearson')['2001: A Space Odyssey (1968)'].sort_values(ascending=False).iloc[:10])
print()
print("-----")
print("List of 10 movies to NOT recommend a user who liked '2001: A Space Odyssey'")
print(data2.corr(method='pearson')['2001: A Space Odyssey (1968)'].sort_values().iloc[:10])
```

```
print()
print("-----")
```

List of 10 recommended movies to a user who has liked '2001: A Space Odyssey'

movie_title

Ciao, Professore! (1993)	1.000000
2001: A Space Odyssey (1968)	1.000000
Designated Mourner, The (1997)	1.000000
Ed's Next Move (1996)	1.000000
Faithful (1996)	1.000000
Dream With the Fishes (1997)	1.000000
Cement Garden, The (1993)	1.000000
Deep Rising (1998)	1.000000
Caro Diario (Dear Diary) (1994)	0.944911
Denise Calls Up (1995)	0.866025

Name: 2001: A Space Odyssey (1968), dtype: float64

List of 10 movies to NOT recommend a user who liked '2001: A Space Odyssey'

movie_title

American Dream (1990)	-1.000000
Collectionneuse, La (1967)	-1.000000
Duoluo tianshi (1995)	-1.000000
Curdled (1996)	-1.000000
Dream Man (1995)	-1.000000
1-900 (1994)	-0.981981
Clean Slate (Coup de Torchon) (1981)	-0.981981
Enfer, L' (1994)	-0.981981
Bitter Sugar (Azucar Amargo) (1996)	-0.970725
Brother Minister: The Assassination of Malcolm X (1994)	-0.962250

Name: 2001: A Space Odyssey (1968), dtype: float64