

# EjercicioS5

May 2, 2020

## 1 Proyecto Final: Sesión 5

### 1.0.1 Cargamos el dataset y lo unimos en ún unico dataframe

```
[1]: import pandas as pd

r_cols = ['user_id', 'movie_id', 'rating']
ratings = pd.read_csv("dataset_unido/u.data", sep='\t', names=r_cols,
    ↳ usecols=range(3), encoding="ISO-8859-1")

m_cols = ['movie_id', 'title']
movies = pd.read_csv('dataset_unido/u.item', sep='|', names=m_cols,
    ↳ usecols=range(2), encoding="ISO-8859-1")

# combinamos ambos datasets para tener el
ratings = pd.merge(movies, ratings)

# Pivotamos la tabla para que la matriz tenga : fila por usuario y columna por
    ↳ película
movieRatings = ratings.
    ↳ pivot_table(index=['user_id'], columns=['title'], values='rating')
movieRatings.head()
```

```
[1]: title      'Til There Was You (1997)  1-900 (1994)  101 Dalmatians (1996)  \
user_id
1              NaN              NaN              2.0
2              NaN              NaN              NaN
3              NaN              NaN              NaN
4              NaN              NaN              NaN
5              NaN              NaN              2.0

title      12 Angry Men (1957)  187 (1997)  2 Days in the Valley (1996)  \
user_id
1              5.0              NaN              NaN
2              NaN              NaN              NaN
3              NaN              2.0              NaN
```

4	NaN	NaN	NaN
5	NaN	NaN	NaN

title 20,000 Leagues Under the Sea (1954) 2001: A Space Odyssey (1968) \

user\_id

1	3.0	4.0
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	4.0

title 3 Ninjas: High Noon At Mega Mountain (1998) 39 Steps, The (1935) \

user\_id

1	NaN	NaN
2	1.0	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN

title ... Yankee Zulu (1994) Year of the Horse (1997) \

user\_id ...

1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN

title You So Crazy (1994) Young Frankenstein (1974) Young Guns (1988) \

user\_id

1	NaN	5.0	3.0
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
5	NaN	4.0	NaN

title Young Guns II (1990) Young Poisoner's Handbook, The (1995) \

user\_id

1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN

title Zeus and Roxanne (1997) unknown Á köldum klaka (Cold Fever) (1994)

user\_id

1	NaN	4.0	NaN
2	NaN	NaN	NaN

3	NaN	NaN	NaN
4	NaN	NaN	NaN
5	NaN	4.0	NaN

[5 rows x 1664 columns]

### 1.0.2 Exploramos el dataframe ya combinando y damos una breve descripción de la composición de este (número de películas y usuarios y rango de puntuaciones)

```
[2]: print("Número de usuarios: {}; Número de películas: {}".format(movieRatings.  
    ↳shape[0], movieRatings.shape[1]))
```

Número de usuarios: 943; Número de películas: 1664

Vemos que el dataset consta de 943 filas (usuarios) y 1664 columnas (películas)

```
[3]: print("Puntuación mínima de cualquier película: {}".format(movieRatings.min().  
    ↳min()))
```

Puntuación mínima de cualquier película: 1.0

```
[4]: print("Puntuación mínima de cualquier película: {}".format(movieRatings.max().  
    ↳max()))
```

Puntuación mínima de cualquier película: 5.0

Vemos que el rango de puntuaciones varía entre [1, 5]

### 1.0.3 Guardamos ese dataframe en un CSV

```
[5]: movieRatings.to_csv('dataset_unido.csv')
```