

# HỆ KHUYẾN NGHỊ

## BÀI TẬP THỰC HÀNH TUẦN 1

### LÀM VIỆC VỚI RATING DATA, THU THẬP DỮ LIỆU

#### 1. Quy định về việc nộp bài

- Thời gian: Được giảng viên thiết lập trên hệ thống Moodle.
- Hình thức nộp: Trên Moodle.
- Bài nộp bao gồm các file .ipynb trong một folder và nén lại thành một tập tin (.zip)
- Cách đặt tên: **BTTH1\_MSSV.zip**
- Công cụ thực hành: **Google Colab**
- Lưu ý: Sai qui định thì sẽ nhận 0 điểm.

#### 2. Nội dung thực hành

##### 2.1 Các thao tác cơ bản trên rating data

- Hướng dẫn sử dụng Google Colab
- Một số dạng bộ dữ liệu rating - Book rating dataset:
  - + Link download: <https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset?select=Ratings.csv>
  - + Rating scale: 1 → 10
  - + Books.csv

	A	B	C	D	E	F	G
1	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M
2	195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amaz	http://images.ar
3	2005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amaz	http://images.ar
4	60973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amaz	http://images.ar
5	374157065	Flu: The Story of the Great Influenza Pandemic	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amaz	http://images.ar
6	393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amaz	http://images.ar
7	399135782	The Kitchen God's Wife	Amy Tan	1991	Putnam Pub Group	http://images.amaz	http://images.ar
8	425176428	What If?: The World's Foremost Milita	Robert Cowley	2000	Berkley Publishing Group	http://images.amaz	http://images.ar
9	671870432	PLEADING GUILTY	Scott Turow	1993	Audioworks	http://images.amaz	http://images.ar
10	679425608	Under the Black Flag: The Romance and the Reality of Life among the Pirates	David Cordingly	1996	Random House	http://images.amaz	http://images.ar
11	074322678X	Where You'll Find Me: And Other Stories	Ann Beattie	2002	Scribner	http://images.amaz	http://images.ar
12	771074670	Nights Below Station Street	David Adams Richards	1988	Emblem Editions	http://images.amaz	http://images.ar
13	080652121X	Hitler's Secret Bankers: The Myth of Sv Adam Lebor		2000	Citadel Press	http://images.amaz	http://images.ar
14	887841740	The Middle Stories	Sheila Heti	2004	House of Anansi Press	http://images.amaz	http://images.ar
15	1552041778	Jane Doe	R. J. Kaiser	1999	Mira Books	http://images.amaz	http://images.ar
16	1558746218	A Second Chicken Soup for the Woman	Jack Canfield	1998	Health Communications	http://images.amaz	http://images.ar
17	1567407781	The Witchfinder (Amos Walker Mystery)	Loren D. Estleman	1998	Brilliance Audio - Trade	http://images.amaz	http://images.ar
18	1575663937	More Cunning Than Man: A Social History	Robert Hendrickson	1999	Kensington Publishing Corp	http://images.amaz	http://images.ar
19	1881320189	Goodbye to the Buttermilk Sky	Julia Oliver	1994	River City Pub	http://images.amaz	http://images.ar

## + Users.csv

	A	B	C
1	User-ID	Location	Age
2	1	nyc, new york, usa	
3	2	stockton, california, usa	18.0
4	3	moscow, yukon territory, russia	
5	4	porto, v.n.gaia, portugal	17.0
6	5	farnborough, hants, united kingdom	
7	6	santa monica, california, usa	61.0
8	7	washington, dc, usa	
9	8	timmins, ontario, canada	
10	9	germantown, tennessee, usa	
11	10	albacete, wisconsin, spain	26.0
12	11	melbourne, victoria, australia	14.0
13	12	fort bragg, california, usa	
14	13	barcelona, barcelona, spain	26.0
15	14	mediapolis, iowa, usa	
16	15	calgary, alberta, canada	

## +Ratings.csv

	A	B	C
1	User-ID	ISBN	Book-Rating
2	276725	034545104X	0
3	276726	155061224	5
4	276727	446520802	0
5	276729	052165615X	3
6	276729	521795028	6
7	276733	2080674722	0
8	276736	3257224281	8
9	276737	600570967	6
10	276744	038550120X	7
11	276745	342310538	10

- Movie rating dataset: Movielens <https://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

+ Link download: <https://grouplens.org/datasets/movielens/>

+ Rating scale: 1 → 5

## + Movies.csv

	A	B	C
1	movieid	title	genres
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
3	2	Jumanji (1995)	Adventure Children Fantasy
4	3	Grumpier Old Men (1995)	Comedy Romance
5	4	Waiting to Exhale (1995)	Comedy Drama Romance
6	5	Father of the Bride Part II (1995)	Comedy
7	6	Heat (1995)	Action Crime Thriller
8	7	Sabrina (1995)	Comedy Romance
9	8	Tom and Huck (1995)	Adventure Children
10	9	Sudden Death (1995)	Action
11	10	GoldenEye (1995)	Action Adventure Thriller
12	11	American President, The (1995)	Comedy Drama Romance
13	12	Dracula: Dead and Loving It (1995)	Comedy Horror
14	13	Balto (1995)	Adventure Animation Children
15	14	Nixon (1995)	Drama
16	15	Cutthroat Island (1995)	Action Adventure Romance
17	16	Casino (1995)	Crime Drama
18	17	Sense and Sensibility (1995)	Drama Romance
19	18	Four Rooms (1995)	Comedy
20	19	Ace Ventura: When Nature Calls (1995)	Comedy

## + Users.csv

	A	B	C	D
	Name Box	movieId	tag	timestamp
2	2	60756	funny	1445714994
3	2	60756	Highly quotable	1445714996
4	2	60756	will ferrell	1445714992
5	2	89774	Boxing story	1445715207
6	2	89774	MMA	1445715200
7	2	89774	Tom Hardy	1445715205
8	2	106782	drugs	1445715054
9	2	106782	Leonardo DiCaprio	1445715051
10	2	106782	Martin Scorsese	1445715056
11	7	48516	way too long	1169687325
12	18	431	Al Pacino	1462138765
13	18	431	gangster	1462138749
14	18	431	mafia	1462138755
15	18	1221	Al Pacino	1461699306
16	18	1221	Mafia	1461699303
17	18	5995	holocaust	1455735472

## +Ratings.csv

1	userId	movieId	rating	timestamp
2	1	1	4.0	964982703
3	1	3	4.0	964981247
4	1	6	4.0	964982224
5	1	47	5.0	964983815
6	1	50	5.0	964982931
7	1	70	3.0	964982400
8	1	101	5.0	964980868
9	1	110	4.0	964982176
10	1	151	5.0	964984041
11	1	157	5.0	964984100
12	1	163	5.0	964983650
13	1	216	5.0	964981208

- Các thao tác trên rating data

- Dataframe: Các thao tác như thống kê, sắp xếp, visualization, merge, join, concat, thêm cột, thêm hàng, xoá cột, xoá hàng, ...

Đọc thêm: <https://pandas.pydata.org/docs/reference/frame.html>

- Thao tác cơ bản trên ma trận: Các phép toán ma trận với số, các phép toán giữa hai ma trận như nhân, chia, tích vô hướng,... - SciPy Sparse Data:

Đọc thêm: [https://www.w3schools.com/python/scipy/scipy\\_sparse\\_data.php](https://www.w3schools.com/python/scipy/scipy_sparse_data.php)

- Yêu cầu: Thực hiện các yêu cầu trong file **lab1.jpynb** ([Tải file trên website môn học](#))

## 2.1 Sử dụng thư viện để thu thập các tin tức

- Quan sát hướng dẫn sử dụng **BeautifulSoup** để thu thập dữ liệu
- Thực hiện yêu cầu:

+ Thu thập 1 tập dữ liệu khoảng 5.000 tin tức từ báo DANTRI (<https://dantri.com.vn>)

+ Bao gồm các thuộc tính như sau:

STT	Tiêu đề tài báo	Nội dung bài báo	Tác giả	Thể loại	Thời gian đăng bài	Tổng số lượt reaction
1						
...						