



M1-04: DataFrames: Row Selection and Pandas “Cheat Sheet”

Part of the “Basics of Data Science with Python” Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-04/>

Course Catalog Dataset - Part 2

Continuing to use the same “Course Catalog” dataset, with the following structure:

Year	Term	YearTerm	Subject	Number	Title	...
2020	Fall	2021-fa	AAS	100	Intro Asian American Studies	...
...						

Dataset URL: <https://waf.cs.illinois.edu/discovery/course-catalog.csv>

Puzzle #3: Using the course catalog dataset that is stored in the DataFrame named `df`, complete the following puzzles:

- (a):
- | | |
|---------------------|---|
| Python: | |
| Description: | Select the first ten courses that are listed in the DataFrame. |
- (b):
- | | |
|---------------------|---|
| Python: | |
| Description: | Select seven courses with the largest course number. |
- (c):
- | | |
|---------------------|--|
| Python: | |
| Description: | Randomly select 1% (0.01) of all the courses offered at Illinois. |

Puzzle #4: Reversing the role and continuing to use the cheat sheet, describe the purpose of the following Python code when used with the course catalog dataset:

- (a):
- | | |
|---------------------|-------------------------------|
| Python: | <code>df.iloc[100:110]</code> |
| Description: | |
- (b):
- | | |
|---------------------|---------------------------------------|
| Python: | <code>df[df.Subject == "MATH"]</code> |
| Description: | |