

Lecture 6: pandas with pop music

LING 1340/2340: Data Science for Linguists

Na-Rae Han

Objectives

- ▶ Tools

- ◆ Using DataCamp courses, downloading csv files

- ▶ pandas library hands-on practice

DataCamp: download exercise .csv files

▶ <https://learn.datacamp.com/courses/intermediate-python-for-data-science>

- ▶ All .csv data files used in DataCamp exercise can be downloaded!
- ▶ Posted on course dashboard. Scroll down!
- ▶ You can download them and replicate exercises in your own Jupyter Notebook.

The screenshot shows the DataCamp course page for 'Intermediate Python'. At the top, it says 'INTERACTIVE COURSE' and 'Intermediate Python'. There are two buttons: 'Continue Course' (green) and 'Bookmark' (white with a bookmark icon). Below these, it lists course details: '4 hours', '18 Videos', '87 Exercises', '577,851 Participants', and '7,400 XP'. The 'Course Description' section states: 'Learning Python is crucial for any aspiring data science practitioner. Learn to visualize real data with Matplotlib's functions and get acquainted with data'. A yellow callout bubble points to the 'DATASETS' section, which lists 'Gapminder', 'Cars', and 'BRICS'. The 'PREREQUISITES' section lists 'Introduction to Python'. At the bottom, there is a progress bar for 'Case Study: Hacker Statistics' at 21%.

Download datasets and work locally!

pandas practice

Activity 3
40 mins



▶ 50 Years of Pop Music

- ◆ <https://www.kaylinpavlik.com/50-years-of-pop-music/>
- ◆ You tried out the CSV file 'billboard_lyrics_1964-2015.csv'



▶ In Class-Exercise-Repo, activity3/ folder:

- ◆ The CSV data file is already in the directory.
- ◆ You will find `pop_music_BLANK.ipynb`
- ← Copy it as `pop_music_YOURNAME.ipynb` and work on it.
- ← You are welcome to commit your version.

```
cp file1 file2  
makes a copy of file1 as file2.
```

Wrapping up

▶ To-do #5

- ◆ Continuation of pandas study notes
- ◆ Try a classmate's CSV

▶ Coming up: Project ideas

- ◆ Due Friday!
- ◆ Not sure? Come chat with Ashley and me

▶ PyLing this Wed! 6pm.

- ◆ Speaker: John Starr (Cornell Ling PhD)
- ◆ Hybrid: CL 2818 (will have pizza!) + Zoom