

Lecture 5: pandas with pop music

LING 1340/2340: Data Science for Linguists

Na-Rae Han

Objectives

- ▶ Housekeeping
 - ◆ GitHub: Resolving conflicts with pull requests
 - ◆ 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 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'. A yellow callout bubble points to the 'DATASETS' section, which lists 'Gapminder', 'Cars', and 'BRICS'. The 'Loops' chapter is highlighted with a purple bar, and its progress is shown as 6%. The 'Case Study: Hacker Statistics' chapter is shown with 21% progress.

Download datasets and work locally!

4 Loops 6%

There are several techniques you can use to repeatedly execute Python code. While loops are like repeated if statements, the for loop iterates over all kinds of data structures. Learn all about them in this chapter.

VIEW CHAPTER DETAILS **Continue Chapter**

5 Case Study: Hacker Statistics 21%

PREREQUISITES
Introduction to Python

DATASETS

- Gapminder
- Cars
- BRICS

pandas practice

Activity 3
50 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, activity2/ 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.
```

Homework 2: Process the ETS Corpus

- ▶ <https://naraehan.github.io/Data-Science-for-Linguists-2022/hw2>
- ▶ Corpus distributed via a private GitHub repo "Licensed-Data-Sets"
 - ◆ Clone directly to your laptop: no need to fork
- ▶ Submission via your own personal & private repo (shared with Na-Rae & the TAs)

Wrapping up

▶ To-do #4

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

▶ Next class:

- ◆ Pop music lyrics, round 2

▶ Coming up: Project ideas

- ◆ Due next Tuesday.
- ◆ Not sure? Come chat with me and the TAs

▶ Coming up: HW2 Process ETS Corpus

- ◆ Due next Thursday.
- ◆ This one is much bigger than HW1. Start soon!