Welcome to Week 14, Lecture 02!

Advanced Transformations with Pandas

05/25/22

Agenda

- Announcements/Assignments
- Transformations Overview
- CodeAlong: Mapping Yelp API Results - Part 2
    - Applying Advanced Transformations

- Quick Project 3 - Part 2 Clarifications

- Office Hours: Tackling the Project 3 - Part 2 assignment
    - Class Vote on Reviewing Live Today or Watching Recorded Video

Announcements

- 🏝️ **Break Week:**
  - Coding Dojo is closed for break next week (05/30 - 06/03)
  - No lecture, code reviews, or 1:1's.
  - Some TA support available (check break week section on stack schedule).

- **Previous Cohort's Data Enrichment** Lecture Recordings Playlist
  - For those who want to get ahead during the break.
  - Also added to Helpful Links tab on stack schedule.

Week 14 Assignments

This week's assignments:

- Efficient Yelp API Calls (Core)
- Project 3 - Part 2 (Core)
- Applying Advanced Transformations (Core)

Assignments turned in by Friday will get feedback Friday.
Assignments turned in by Sunday 06/05/02 at 11:59 PM will be consider on-time.

Advanced Transformations

Transformations to Perform

- Split 1 string column into multiple columns.
- Convert strings of dictionaries into actual dictionaries
- Use lambda functions with .apply
- Construct a function for extracting data from a column

CodeAlong

Mapping Yelp API Results - Part 2

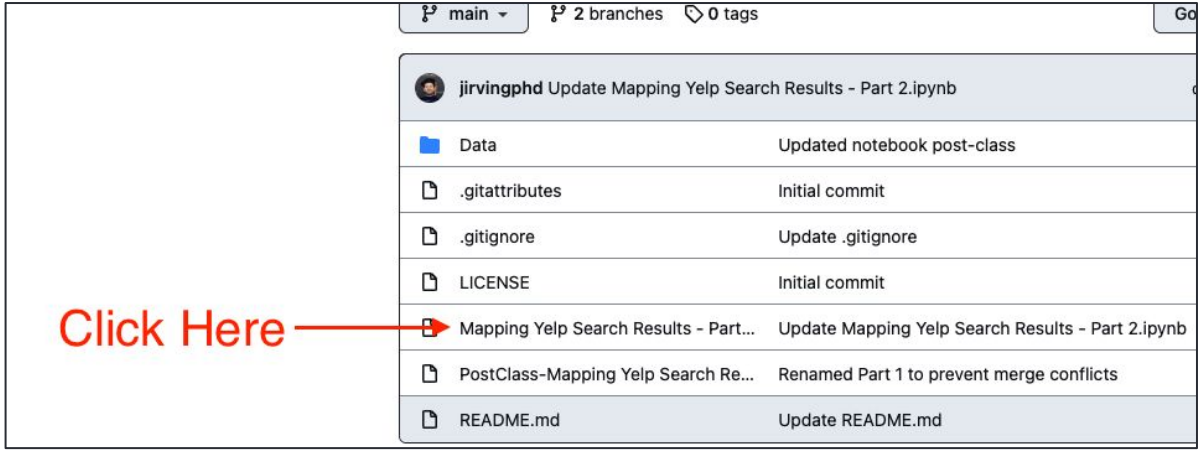
Activity Repository

- **We will be using the csv of our API results from last class and creating an interactive map.**
  - We will practice many of the advanced transformations steps.

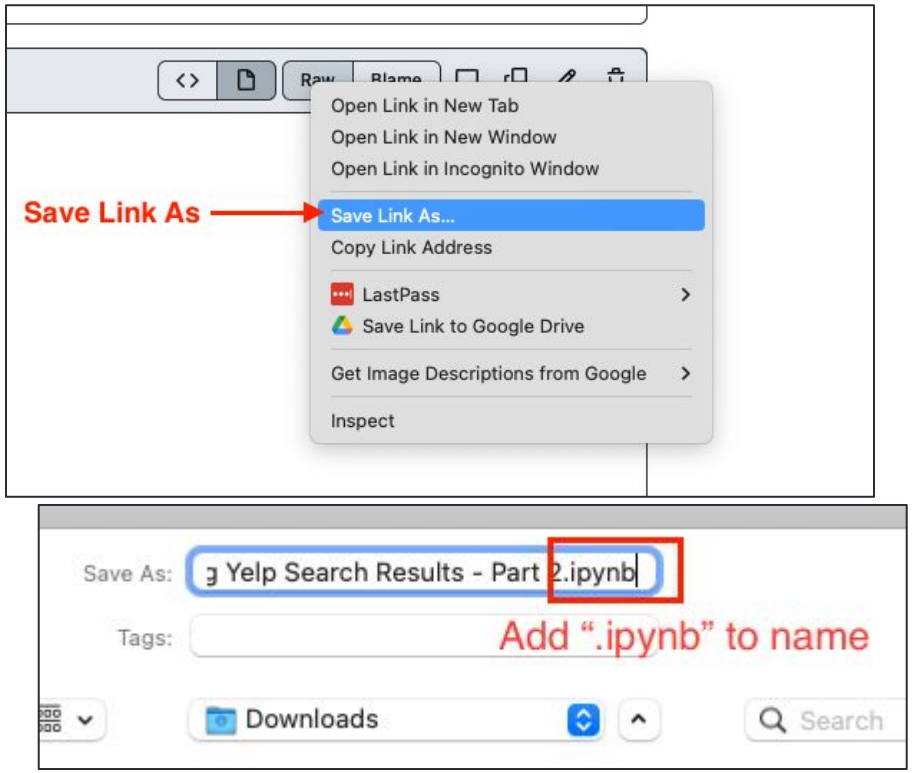
- 🗺️ CodeAlong: Mapping Yelp API Results - Part 2:
  - Same Repo with new Notebook: “Part2-Mapping Yelp Search Results.ipynb”
  - https://github.com/coding-dojo-data-science/data-enrichment-wk14-activity-mapping-yelp-api-results
    - Last’s Class Notebook & Data on Branch “05-23-22-class”

- If you are adding to your own copy to use your own data, we want to download JUST the new notebook file (see next slide).
  - If you did not do Part 1, you can use my Baltimore Burger data in the Data folder.

To Download Notebook File - #1



To Download Notebook File #2



Save the notebook link in the same folder as your part-1 notebook

Make sure to add ".ipynb" to file name!!!

Project 3 - Part 2

Quick Clarifications + Tips

PART 2 TIPS:

- **Make sure to import:**
  - from tqdm.notebook import tqdm_notebook
  - **import time**

- Functions should be defined BEFORE the loop and then just USED in the loop.

- Your EDA should be a **quick** peek at your results.
  - Do not worry about cleaning the data and dealing with null values

PART 1 TIPS (for those catching up)

- Replacing “\\N” with np.nan requires you import numpy as np (so no “ ” around np.nan)

- String methods throw an error if a cell has NaN
  - (like df[‘Genres’].str.contains(‘Documentary’))
  - Drop null values from genres before checking for documentary.

End of Lecture

Office Hours: Project 3 - Part 2 Demo

- ✅ I will record and share the demo.
- ❌ I will NOT share the notebook.

Recording: https://youtu.be/uD_HeaC_IF0