

My code requires urllib3, bs4, and psycopg to scrape and clean data from grad cafe. The initial scrape and clean only needs bs4 and urllib3 to get the data, but needs to check the database to see which records are new when a new scrape and clean is requested in the pipeline.

Scrape and clean also makes use of date utils, which is for formatting dates properly. I noticed some of the entries in the grad cafe included month and day, but not year, so I do a little check back and forth to figure out what year it's supposed to be and I add that to the record.

This was all part of scrape and clean, but after pylintng it made sense to break it off into its own utility for the simplicity and readability of the scrape and clean program.

Load data, and load new data also require psycopg. Load data does the initial push when the database is first being populated, and load new data uses it to put new records into the database.

Query data requires psycopg.sql to do database queries. The main app is dependent on all of these indirectly through the various utility apps. Only load data doesn't actually touch the main app. Flask and psycopg are the only direct imports to the main app aside from subprocess (which didn't show in the diagram for some reason).

