**BringFido Webscraping Project High Level Design:**

All Dog Friendly Restaurants by State, scraping…

* Restaurant Name
* Restaurant Address
* Bones (rating denoted by whether bone 1-5 is filled or empty)
* Overall Rating? (only if I want to scrape another level down)
* Comments? (outside the scope of this project)

Analysis:

* Cities w/the most dog friendly restaurants
* Categorizing by words in restaurant name…
  + For example: dog, beagle, lazy, tavern, fish house, burger, Mexican, bistro, French, Italian, cafe, bakery, patio, park, brewery, brew house, inn, bar, waffle, chowder, garden, steak burger, shack, pub, tiki, hut, grill, restaurant, trattoria, beach, lake, woods, lounge, noodle, city, Chinese, Japanese, sushi, crepe, muffin, eggs, pancake, city, harbor, bay, break, cliff, seafood, etc
* Restaurants go out of business quickly if they are not successful. So, if we make the assumption that existing restaurants are successful restaurants, we can correlate pet friendly restaurant success to city and aspects of the restaurant name.
* Bone rating shows how much the restaurant is ‘liked’ by ONLY people who rated it on bringfido (only people who specifically want a dog friendly restaurant)
  + Bone rating will be used to measure ‘how’ successful the restaurant is, as perceived by dog people
  + It would be interesting to compare bone rating to rating given in general (not just from those wanting a dog friendly restaurant) by bringing in google or yelp ratings for each restaurant and comparing how the restaurant is perceived by dog people vs general people, but this is outside the scope of this project (should be mentioned in future work)
* Joining this data with census data would bring additional insight as to demographic characteristics of the cities w/the most dog-friendly restaurants, but this is out of scope for this project due to limited time (should be mentioned in future work)

Navigating web pages:

* Start w/URL that has state already selected (can also do for entire US, but start w/one state)
  + Hawaii – 495 – test population
  + Florida - 6,350 – target for this project
  + All US - 59,809 – future work
* 24 tiles are presented
* Show more brings up 24 more tiles, but URL does not change
* As ‘show more’ is selected, the html expands by 24 more tiles each time
* The navigation is to take the total number, divide by 24, and click the show more by that many times (ignore remainder because first set of 24 does not require a click.
* Then scrape data from all tiles

**--**🡺 Changing strategy due to issues encountered w/Selenium. Even though the link does not change when loading another page of results, if I go to a lower level (the individual restaurant), every restaurant has its own page in the format of [https://www.bringfido.com/restaurant/(#](https://www.bringfido.com/restaurant/() less than 75000) where the data I want can be scraped and simplifies to one page at a time w/no navigation. Therefore, if I’m willing to scrape all restaurants from bringfido, then I can visit each restaurant page through Scrapy instead of Selenium to get all the data and more quickly. Some numbers under 75000 do not bring up a restaurant page (they are assigned to cities or neighborhoods w/more than one restaurant appearing on the page). The scraping script will fail for these numbers because the html expected will not match what is presented. These records will simply be thrown out because they are not individual restaurants. Everything on these pages ultimately links to a different number so no restaurants will be missed by approaching the webscraping in this way.

I know I held on to my objective to power through w/Selenium for too long. But, I have a clear objective of what I want to do w/the resulting data and I can do it in chunks w/Scrapy in this way… thus, making it possible to start creating cleaning scripts, analyzing the facets of this data that I’m interested in, and scripts to create the graphs that go with on a subset of the data before the entire set is complete, and then re-run on the entire set.