ETL: Adoptable Pets

**Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc)**

The three original data sources are: PetFinder API, “Most Popular Dog Breeds of 2018 (2019)” article from the *American Kennel Club* website, and the “Most Pet-Friendly Cities” article from the *Wallet Hub* website. The PetFinder API calls return nested dictionaries in JSON format. The “Most Popular Dog Breeds of 2018 (2019)” and “Most Pet-Friendly Cities” articles are formatted in HTML.

Wallet Hub —> using Chrome Inspect, copy and pasted HTML data to create an HTML file in the Visual Studio Code program. Used Beautiful Soup HTML Parser to ingest the HTML. The target data was captured within a HTML Table attribute/tag.

AKC —> used Beautiful Soup library in Python to extract the HTML data from the HTML Table attribute/tag in the article. Added a column for the year to dictate where the final

PetFinder API —> used Requests library in Python to extract the JSON data.

**Transform: what data cleaning or transformation was required**

* PetFinder API
  + Filtering data for only dogs
* Wallet Hub
* AKC

**Load: the final database, tables/collections, and why this was chosen.**

There is one final database named “Adoptable Dogs.” The database has three tables.

Appendix:

<https://www.petfinder.com/developers/>

AKC Staff. “Most Popular Dog Breeds of 2018 (2019).” *American Kennel Club*, 20th May, 2019, <https://www.akc.org/expert-advice/news/most-popular-dog-breeds-of-2018/>.

McCann, Adam. “Most Pet-Friendly Cities.” *Wallet Hub*, 7th August, 2018, <https://wallethub.com/edu/most-pet-friendly-cities/5562/>.