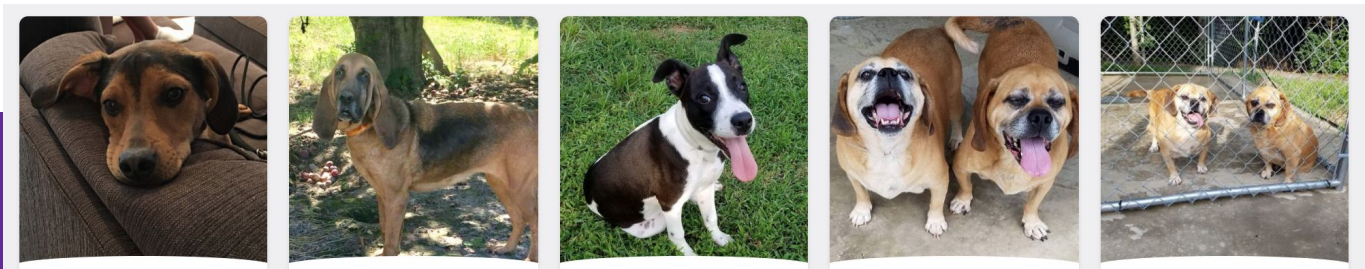


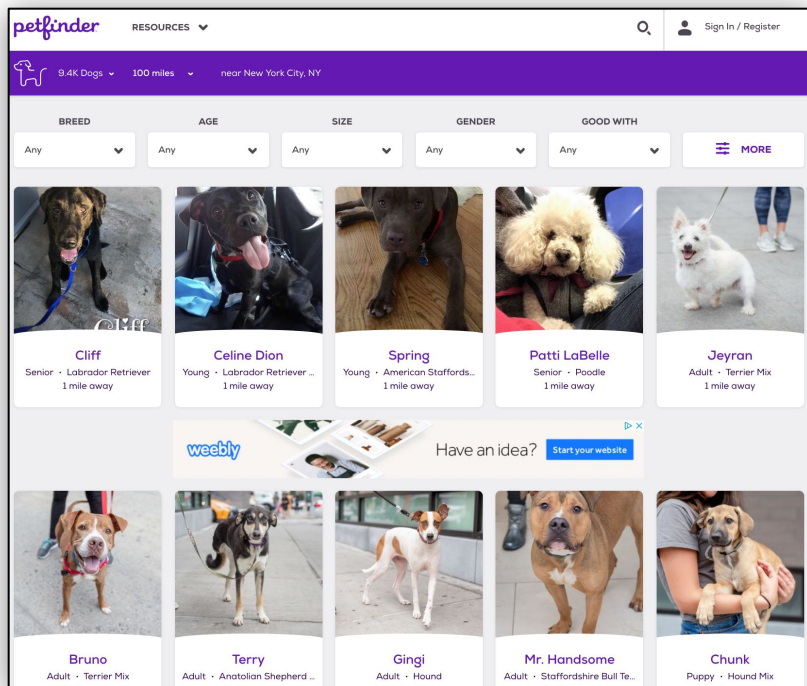
# WebScraping Project:

petfinder



Erin Dugan • 08.08.2018 • NYC-Data Science Academy

# overview



## petfinder.com

Largest online pet adoption website in North America

Almost 300,000 animals listed, including:

cats, dogs, fish, birds, horses, lizards, etc.

## objectives

Scrape subset of dog postings in NYC metro area

Explore distribution by age, breed, size, etc

Find adoption trends by region

# webscrapping

8000+ dogs from  
urban/suburban/rural areas in 5  
states

breed, age, size, health  
posting date, location, contact  
info

brief description

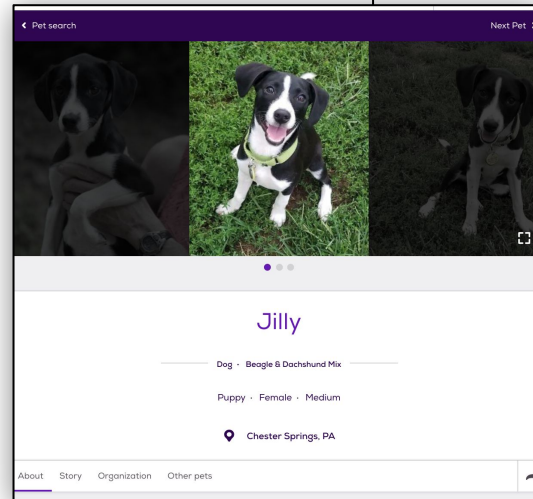
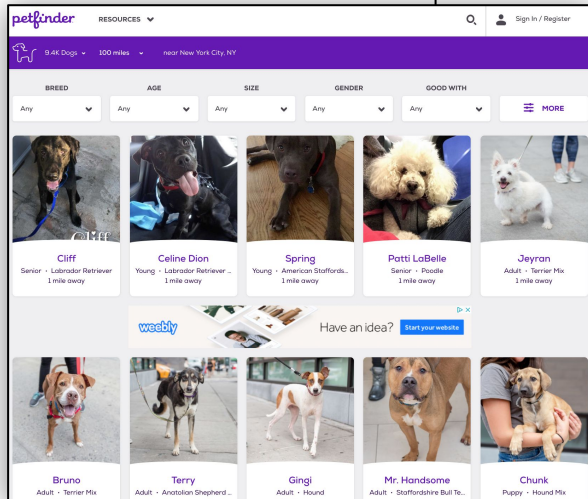
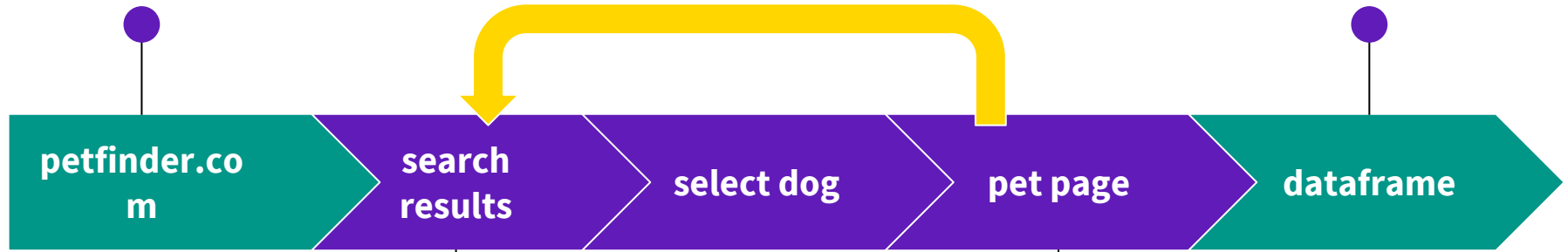
**All dogs available for adoption  
within 100-mile radius of White  
Plains NY**



**Search**: dogs within  
100-mi of White Plains NY

**Selenium**: step through 40  
dogs per page, 217 pages

**Python**: clean & parse  
data for analysis



**Selenium & Xpath**:  
scrape visible &  
hidden data within  
HTML code

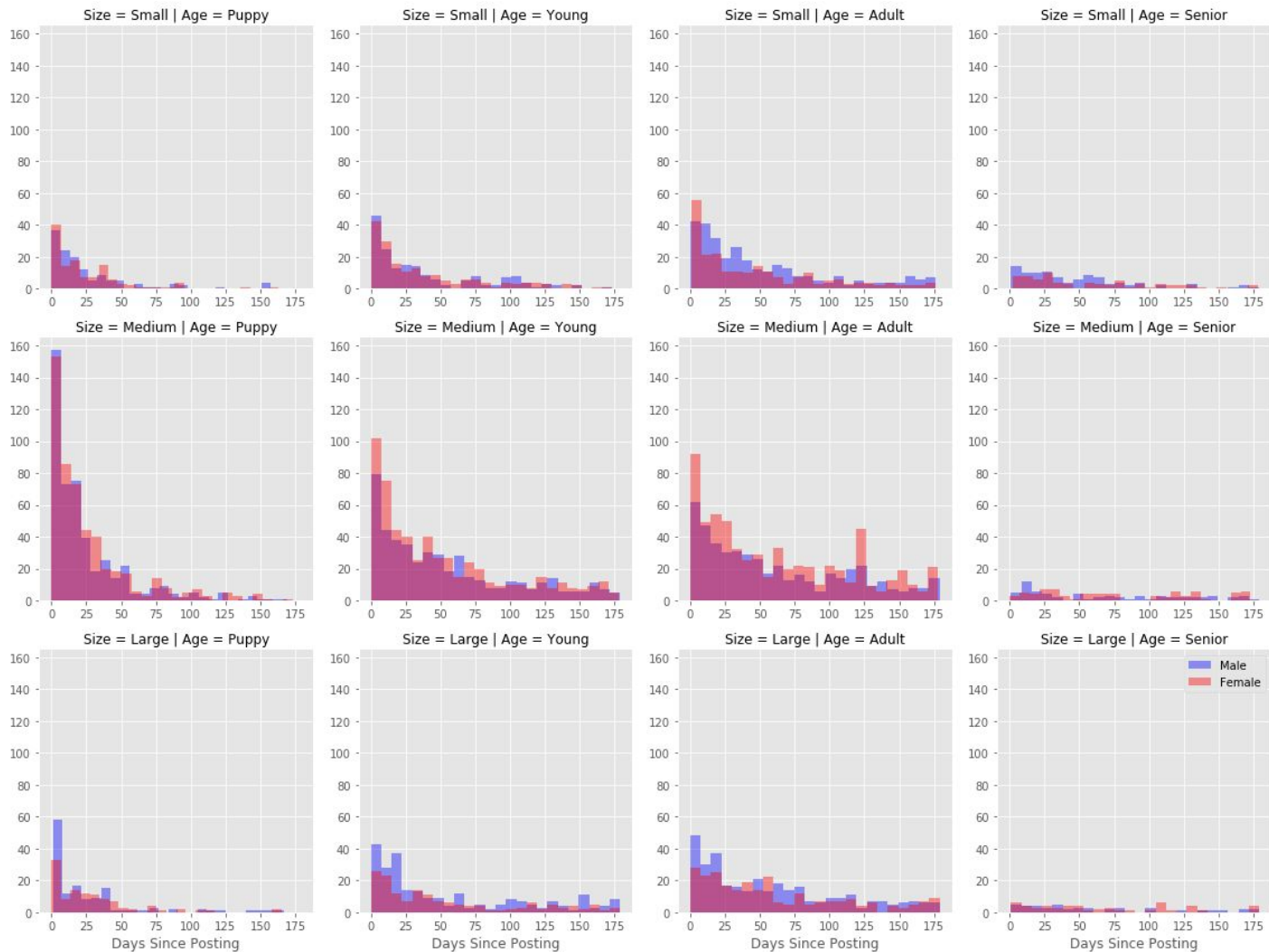




# adoption trends

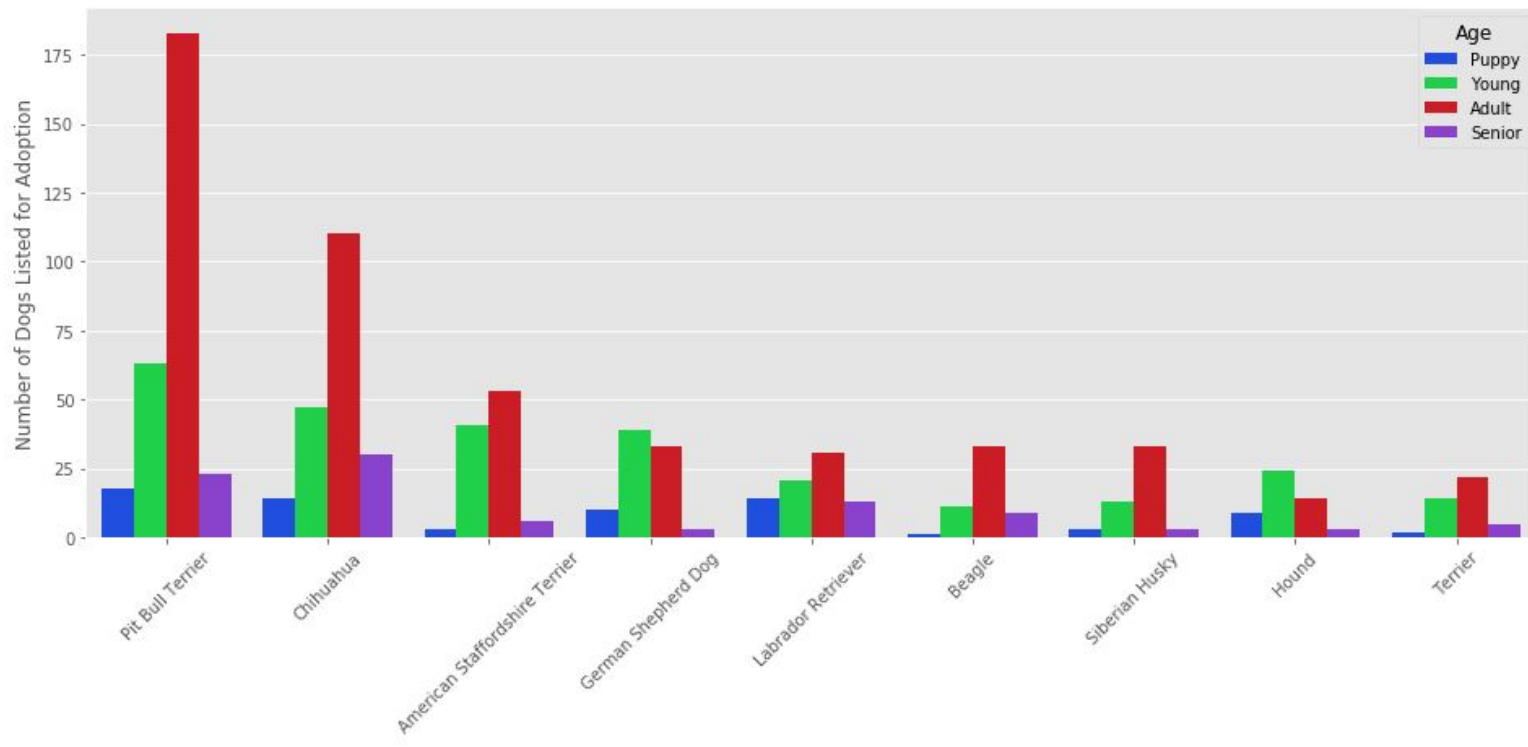
Number of days dogs listed on petfinder.com (before adoption)

How quickly does each category get adopted?



# demographics

Most popular purebred dogs in NYC region



## A close-up photograph of a grey and white pit bull dog. The dog has a white blaze on its face and is wearing a red collar. It is sitting on green grass.





# Continued Work

---

## Future Applications

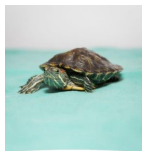
- Analyze other pet types & regional areas
- Evaluate most popular mixed dog breeds
- Help to find homes for pets evacuated from devastated areas due to fires, hurricanes, etc.
- Use similar analysis for other search sites: Dating apps, Employment search, etc.



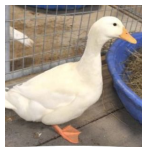
**Yuri**  
Lionhead & Lop Eared Mix  
44 miles away



**Waffles**  
Domestic Short Hair  
26 miles away



**Leah**  
Turtle  
225 miles away



**PUCK**  
Duck  
932 miles away



**Bingo**  
Quarterhorse  
840 miles away



**Atlas**  
Vietnamese Pot Bellied  
440 miles away