Determining the Optimum Location for a Small to Medium Dog Kennel

Ian Fleury

Finding the Right Location to Move: Raising Dogs

- **Located in the United States**

- **Healthy State: environment should have hospitable conditions, favorable to animal's health, and NOT be full of hazards such as:**
  - **smog (polluted air)**
  - **rampant pesticide drops**
  - **dangerous ticks**
  - **other biting insects**

- **Favorable Breeding Laws: breeding laws differ from state to state and they range from quite lenient to very strict. Client is looking for areas that do NOT have:**
  - **stiff regulations**
  - **intrusive inspections**
  - **high annual fees**

Finding the Right Location to Move: Raising Dogs

- **State should allow her to attain her upper limit of:**
  - **20 dogs in residence**
  - **3 litters a year**
  - **15 puppy sales**

Finding the Right Location to Move: Recreation

- **Near to populated city**

- **City should have all or most of these features:**
  - **good walking parks**
  - **art exhibits**
  - **theatrical events**
  - **museums**
  - **animal preserve or zoo**
  - **pizza venues**
  - **taco vendors**
  - **domestic beer establishments**
  - **specialty coffee (fresh roasted)**
  - **espresso venues**

Finding the Right Location to Move: Points of Interest

- **Surrounding Cities: areas of real estate within 25 miles of the big city, but preferable not within it. Should be near locations where breeders are already established since this would indicate the proper amount of acreage and the right terrain for raising and breeding dogs.**

- **Vet Clinics: finding the top-rated veterinary clinics in the area is of the utmost importance for the dog's health and emergency care.**

- **Boarding Facilities: should be hospitable, well-run establishments where client can have dogs taken care of when she needs to travel or rotate breeding pairs.**

- **Pet Stores: locations to purchase high-quality dog food , grooming supplies, and other pet-related products. Client prefers the chain stores: PetSmart, Petco, and Petland since she already has membership with each of them and gets a breeder discount.**

Data Acquisition

- Used BeautifulSoup to scrape American Kennel Club website to retrieve table of the "*10 Best (and Worst) States for Your Pet's Health*."
- Used lxml scraper to retrieve table from www.animallaw.info (Michigan State University Animal Legal & Historical Center) detailing commercial pet breeders laws by state.
- Used the Foursquare API to search through a selection of cities looking for attractions that matched the keywords given by the client.
- Parsed tables from www.gomapper.com and www.withinhours.com, to obtain two lists of surrounding towns and cities within 25 miles of Wichita.

Data Acquisition

- Used BeautifulSoup to scrape the AKC website again to find a list of some of the breeders in the area and map out where they are situated.
- Used both the Foursquare and Yelp APIs to discover the locations of nearby vet clinics, pet stores, and boarding kennels.
- Used the Yelp API to display ratings, and customer reviews for the selected vets and boarding facilities.

Data Cleaning

- Added a 'Pet Health Score' column to the "*10 Best & Worst States*" dataset to give a positive weight of +1 to the best states and a negative weight of -1 to the worst states.

- Removed unnecessary and incomplete columns and rows from the table of "*Breeder Laws per State*." Converted long passages of text to numeric data separated into 3 categories.

- Removed redundant data from the list of attractions in the ten sample cities dataset.

- Removed duplicate entries and outliers from the two lists of cities and towns within 25 miles of Wichita. Corrected wrong latitude/longitude coordinates. Merged the two lists into one.

Data Cleaning

- Corrected a few details from AKC's list of local dog breeders, such as duplicate entries or wrong kennel names.
- Removed unnecessary columns of information from the vet clinic, boarding kennel, and pet store location data.
- Eliminated all pet stores except for "PetSmart", "Petco", and "Petland." Fixed entries with incorrect latitude/longitude coordinates.
- Created an additional column in the boarding kennel dataset to indicate which locations, retrieved from the Yelp API, also appeared on the Foursquare results.

Data Manipulation

After parsing the "Pet Breeder Laws per State" table from its website, I had to extract the numbers contained within the text.

State	Definition of commercial breeder and licensing requirements
Arizona	"Kennel" means an enclosed, controlled area, inaccessible to other animals, in which a person keeps, harbors or maintains five or more dogs under controlled conditions.A. R. S. § 11-1001A person operating a kennel shall obtain a permit issued by the board of supervisors of the county where the kennel is located except if each individual dog is licensed.Cost is $75 A. R. S. § 11-1009
California	As used in this article, "dog breeder," or "breeder" means a person, firm, partnership, corporation, or other association that has sold, transferred, or given away all or part of 3 or more litters or 20 or more dogs during the preceding 12 months that were bred and reared on the premises of the person, firm, partnership, corporation, or other association.

Converted to numeric data.

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year
Arizona	5	0	0
California	0	20	3

Data Manipulation

I then made some logical assumptions to fill in some of the blanks. Knowing that the average cocker spaniel litter is 5 puppies, I filled in empty "Maximum puppy sales" entries by multiplying 5 times the number of "Maximum litters ".

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year
Minnesota	10	0	6

6 x 5 = 30

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year
Minnesota	10	30	6

If the number of "Maximum litters" was empty then I would do the opposite operation and divide the "Maximum puppy sales" by 5 and use that number as the new litter amount.

Data Manipulation

Similarly, if there was no entry for "Maximum number of dogs", but there were numbers in either of the other categories, then I would assume a maximum dog allowance equal to at least one breeding pair, or 2 dogs.

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year
Pennsylvania	0	61	0

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year
Pennsylvania	2	61	12

Data Manipulation

I compared the new values to the client's desired goal of 20 dogs in residency, 3 litters per year, and 15 puppy sales per year. If a state's laws allowed the client to operate within her goal, then I would give that state a score of 1 for each feature that met or exceeded the client's upper limits, and 0 for each feature that did not meet it.

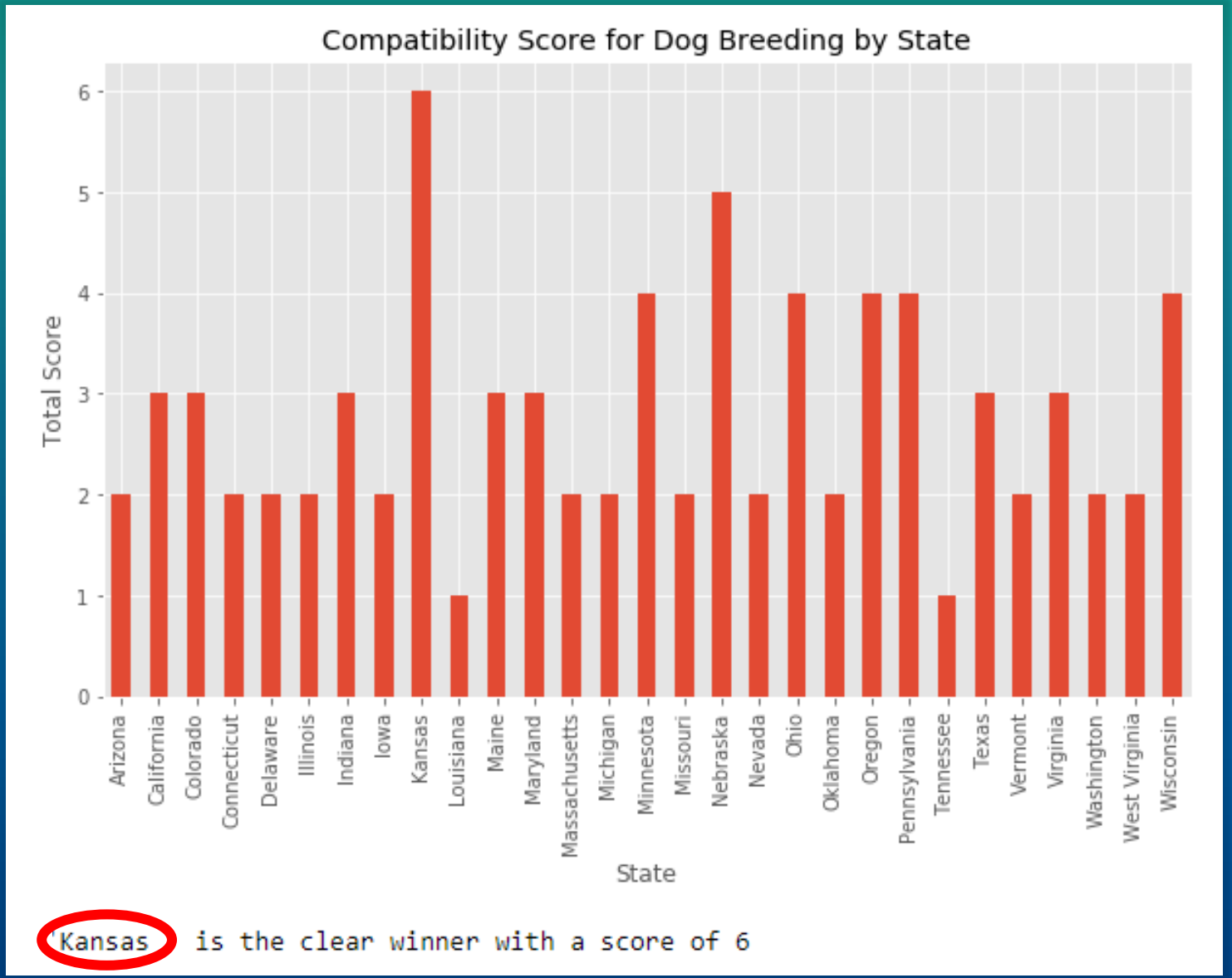
State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year	Definition of commercial breeder and licensing requirements	Pet Health Score
Nebraska	4	31	4	Commercial breeder means any one of the follow...	1.0

(>=20) (>=15) (>=3)

State	Maximum number of dogs	Maximum puppy sales per year	Maximum litters per year	Definition of commercial breeder and licensing requirements	Pet Health Score	Total Score
Nebraska	0	1	1	Commercial breeder means any one of the follow...	1.0	3.0

The scores that each state received were then added, along with their "Pet Health Score", to give a "Total Score" result.

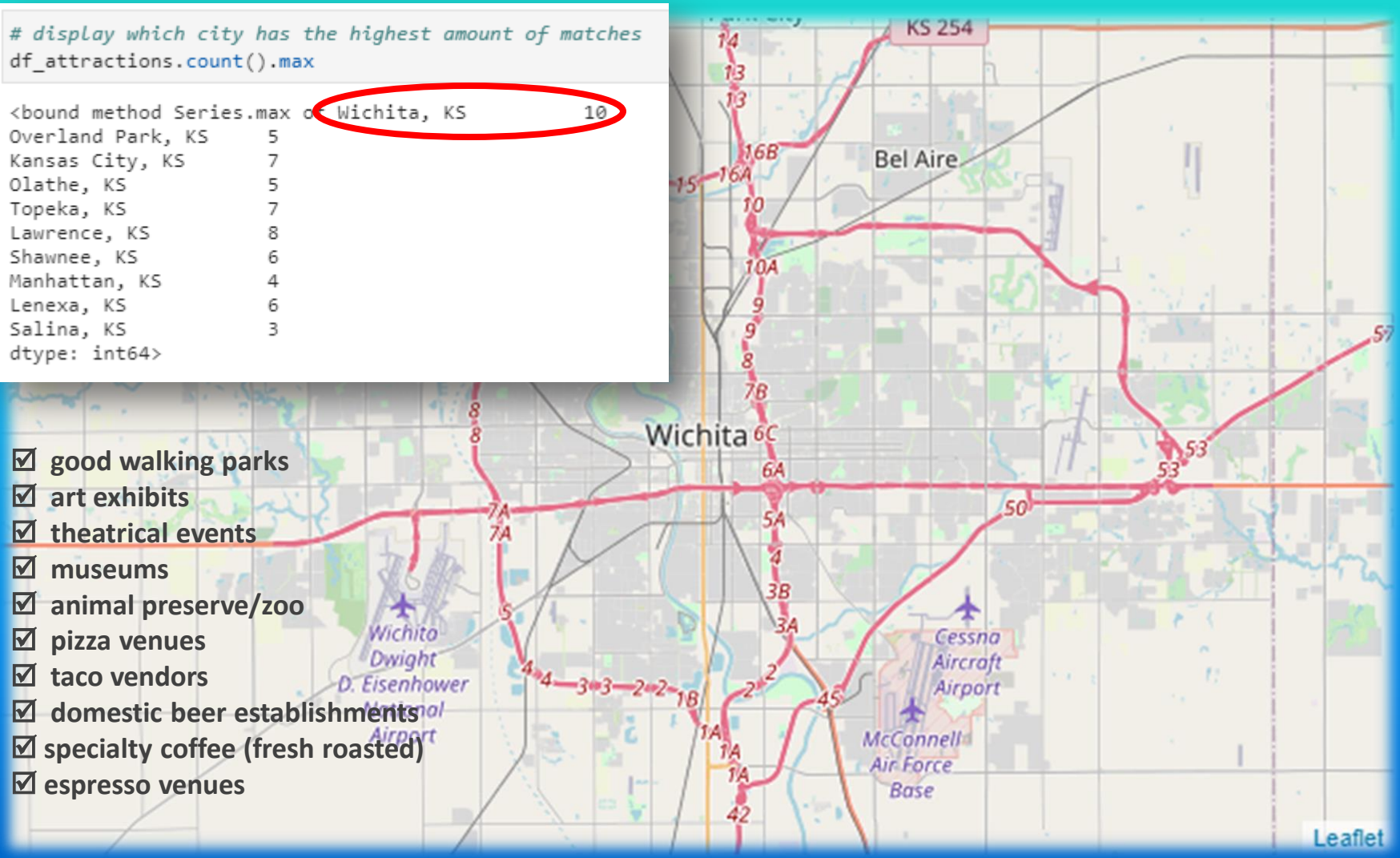
Visualizing the State Scores



Determining City with Best Attractions

```
# display which city has the highest amount of matches
df_attractions.count().max
```

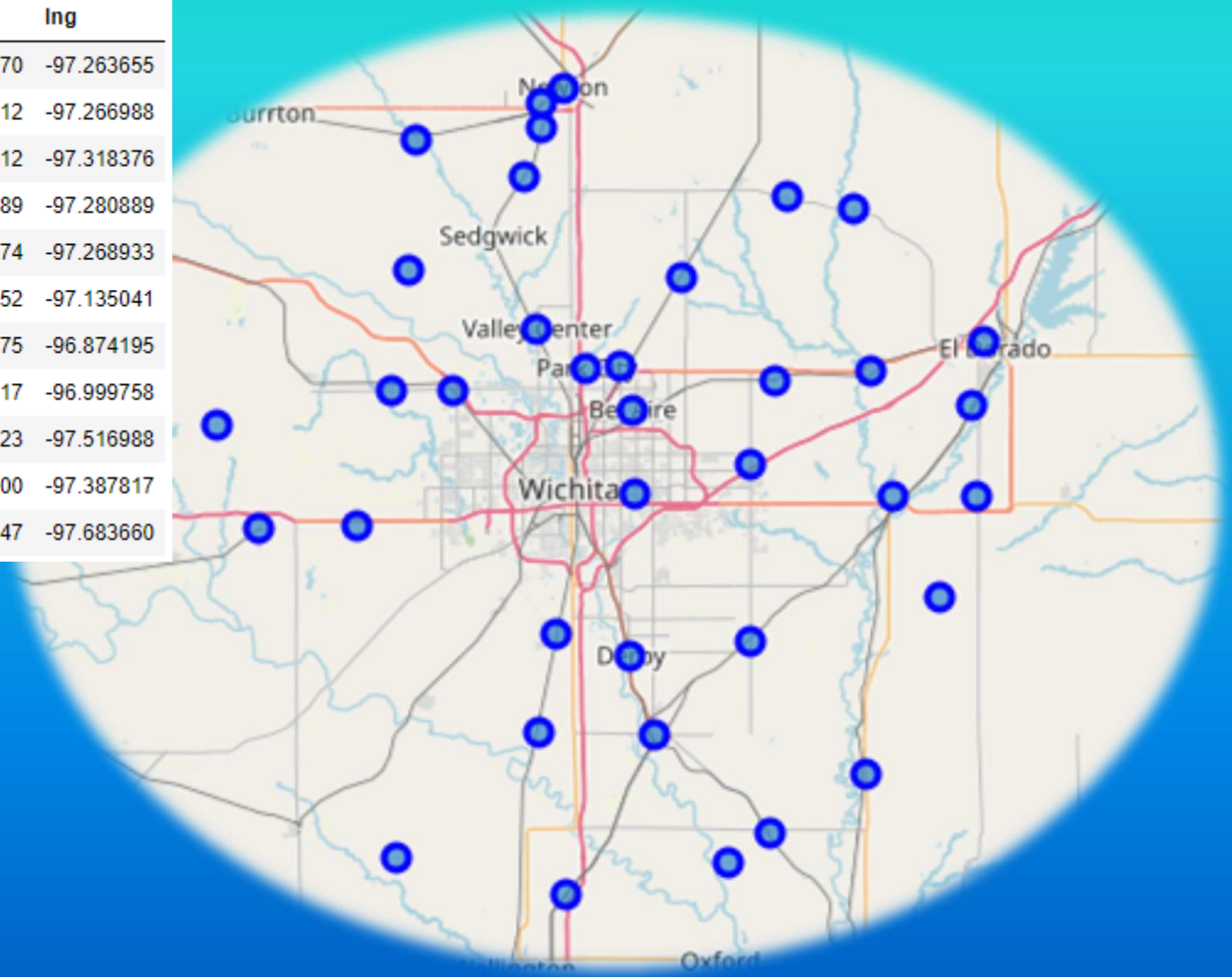
```
<bound method Series.max of Wichita, KS          10
Overland Park, KS     5
Kansas City, KS       7
Olathe, KS            5
Topeka, KS            7
Lawrence, KS          8
Shawnee, KS           6
Manhattan, KS         4
Lenexa, KS            6
Salina, KS            3
dtype: int64>
```



- ☑ good walking parks
- ☑ art exhibits
- ☑ theatrical events
- ☑ museums
- ☑ animal preserve/zoo
- ☑ pizza venues
- ☑ taco vendors
- ☑ domestic beer establishments
- ☑ specialty coffee (fresh roasted)
- ☑ espresso venues

Locating Surrounding Cities

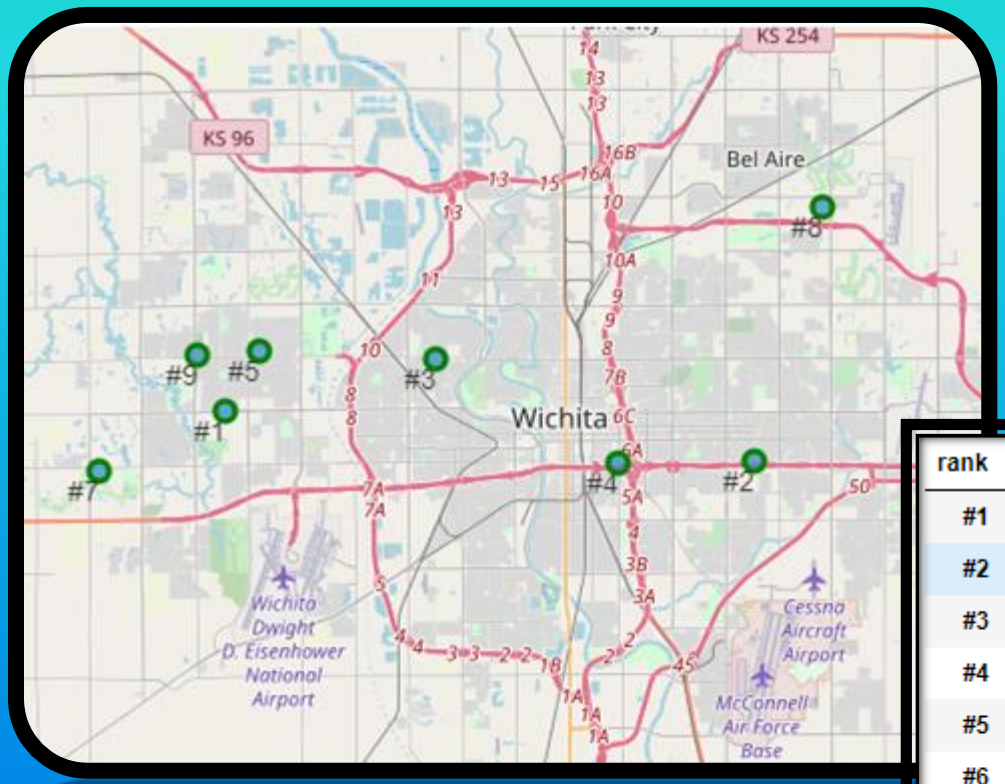
	name	Wichita (mi)	lat	lng
0	Eastborough	4	37.688070	-97.263655
1	Bel Aire	6.2	37.762512	-97.266988
2	Park City	7.9	37.800012	-97.318376
3	Kechi	8.2	37.800889	-97.280889
4	Derby	10.4	37.545574	-97.268933
5	Rose Hill	13.8	37.558352	-97.135041
6	El Dorado	26.9	37.823475	-96.874195
7	Towanda	19.9	37.797517	-96.999758
8	Bentley	16.6	37.886123	-97.516988
9	Putnam	19.3	37.968900	-97.387817
10	Garden Plain	19.1	37.658347	-97.683660



Locating the Local Breeders

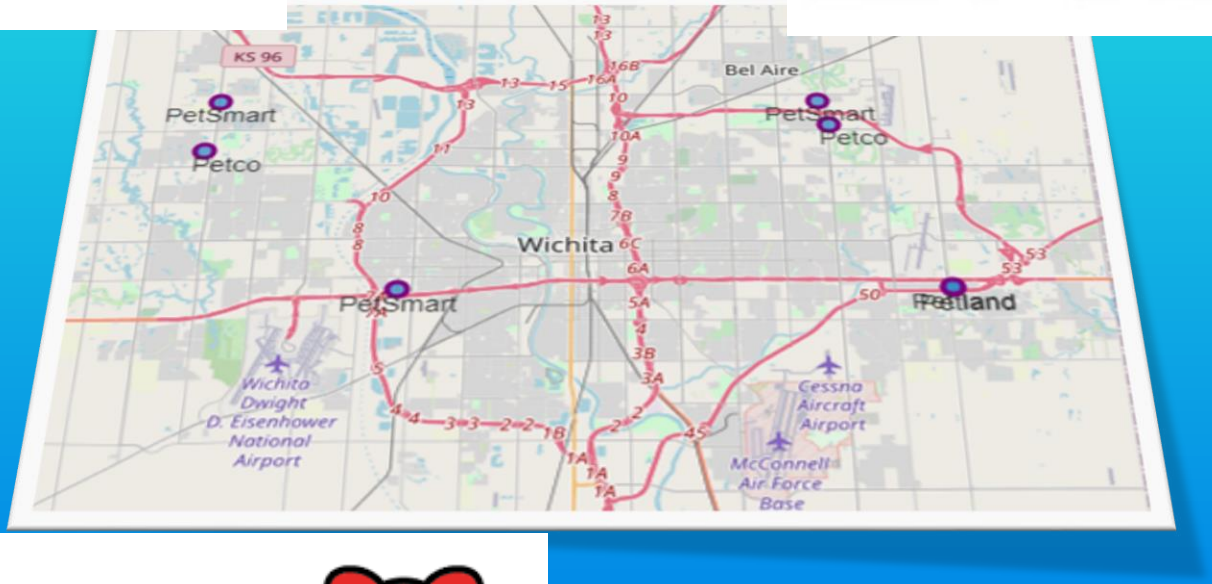
	kennel name
0	GladeMist Chihuahuas
1	Waltman's Bulldogs (English) & Waltman's Chin...
2	Xyomar Standard Schnauzers
3	Komondor Puppies
4	Nightwatch Black Russian Terriers
5	Creekwell Shepherds
6	Brenglora Bulldogs
7	Miniature American Shepherds at Coyote Creek
8	Cavalier King Charles Spaniel Male Puppies
9	Rottweiler Puppies
10	Labrador Retriever Puppies

Locating the Top Rated Veterinary Clinics



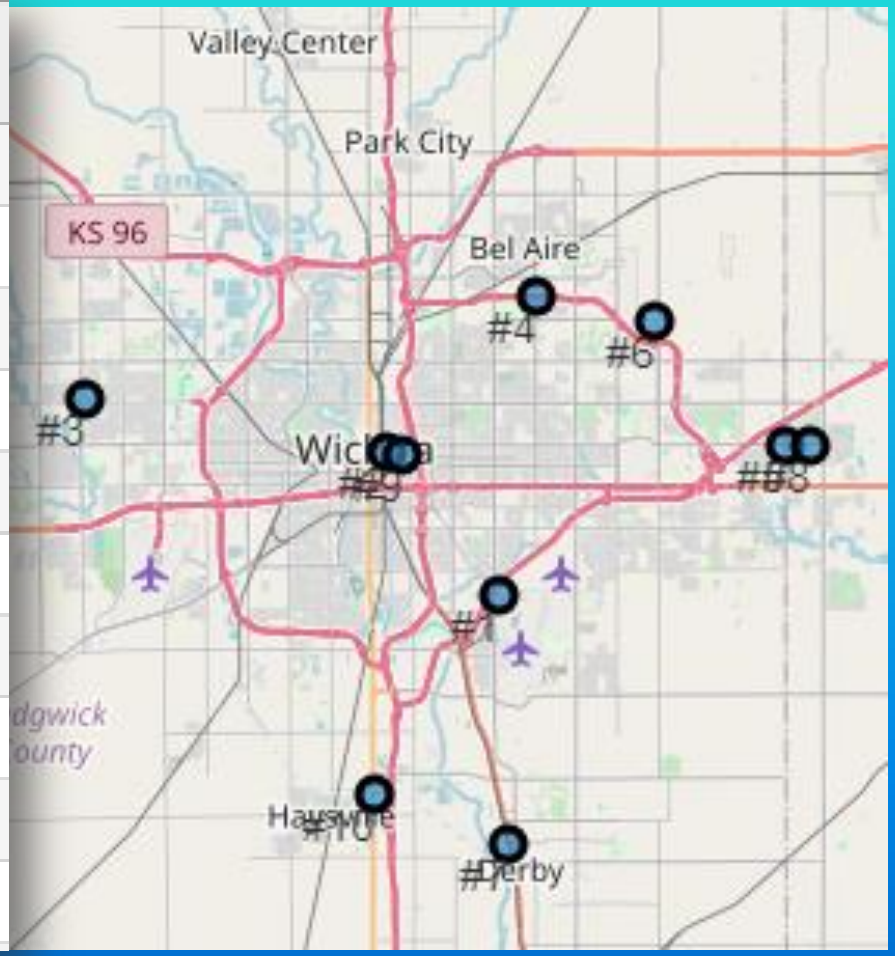
rank	
#1	Best Friend's Pet Clinic
#2	Skaer Veterinary Clinic
#3	Indian Hills Animal Clinics (loc. 1)
#4	Air Capital Veterinary Clinic
#5	All Creatures Veterinary Hospital
#6	Mulvane Animal Clinic
#7	Animal Hospital at Auburn Hills
#8	Banfield Pet Hospital
#9	Indian Hills Animal Clinic
#10	El Paso Animal Hospital (loc. 2)

Locating the Pet Stores

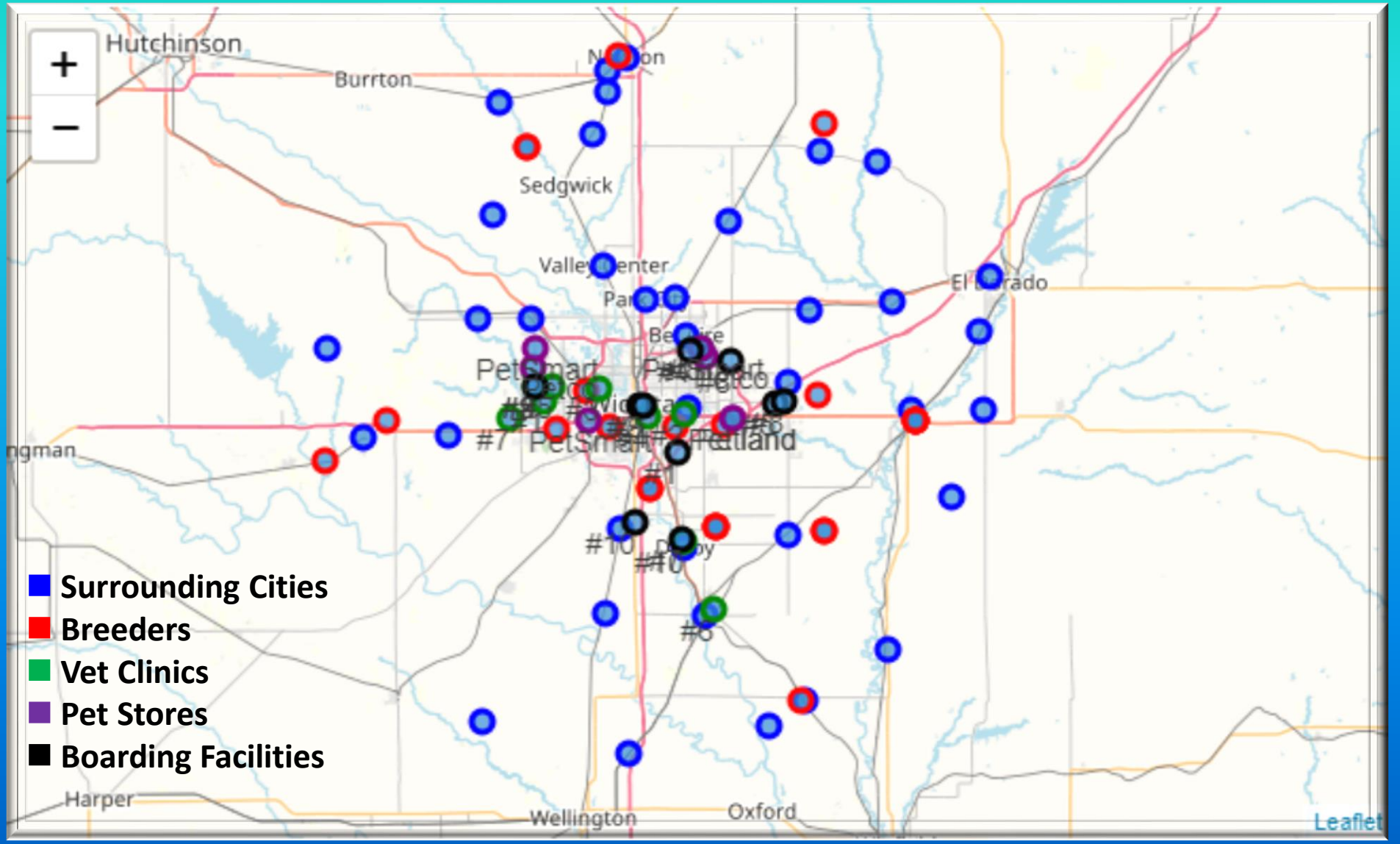


Locating the Top Rated Boarding Facilities

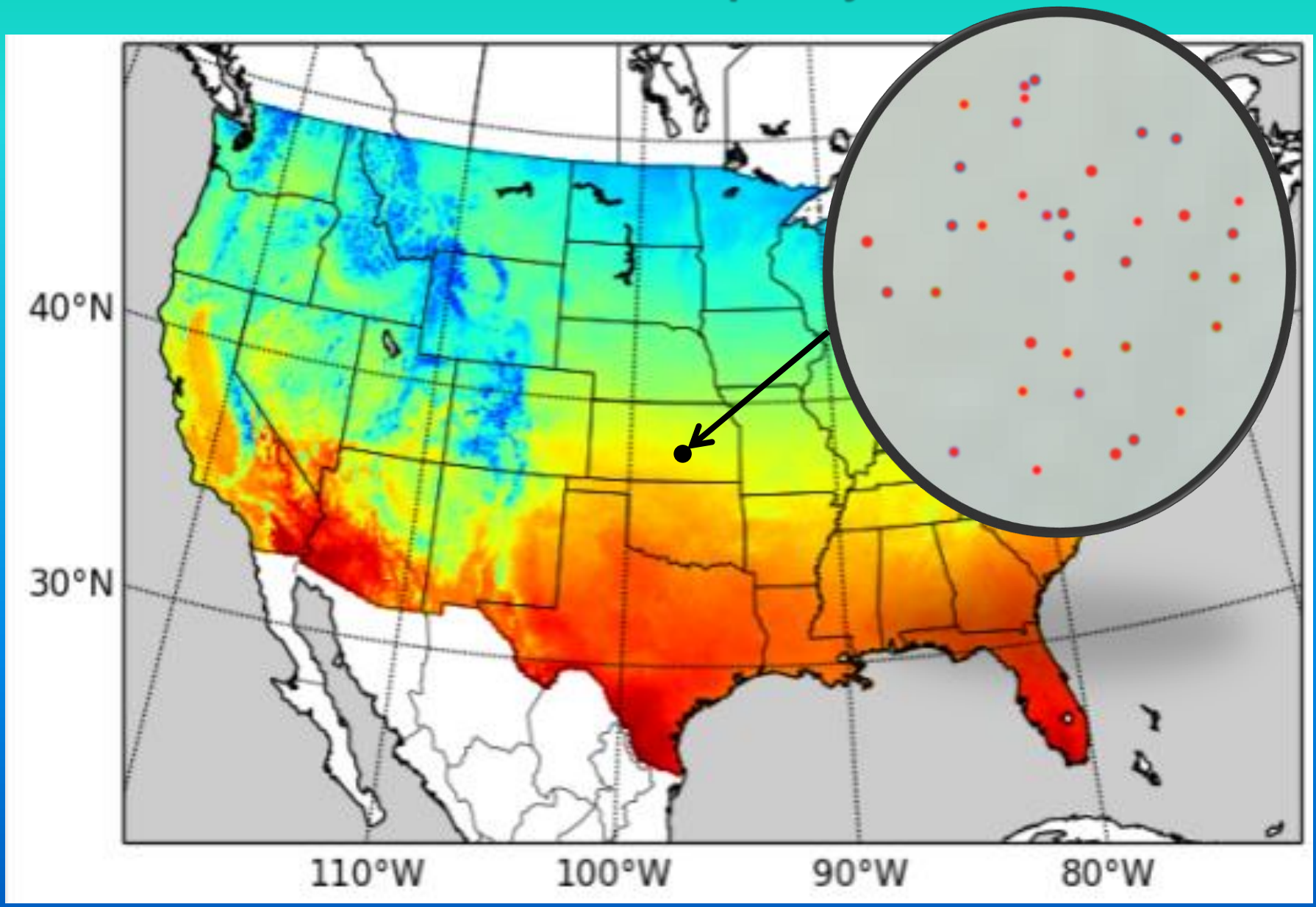
	name
1	Pet Services and Supplies
2	Dog Days Of Sommer
3	Indian Hills Animal Clinic
4	Heartland Animal Hospital
5	Kutter Pet Care Center
6	Sydney's Pet Resort
7	Diggin' Inn Pet Resort
8	ıal Hospital and Pet Resort
9	Doggy Day Care
10	Hound Dog Hotel



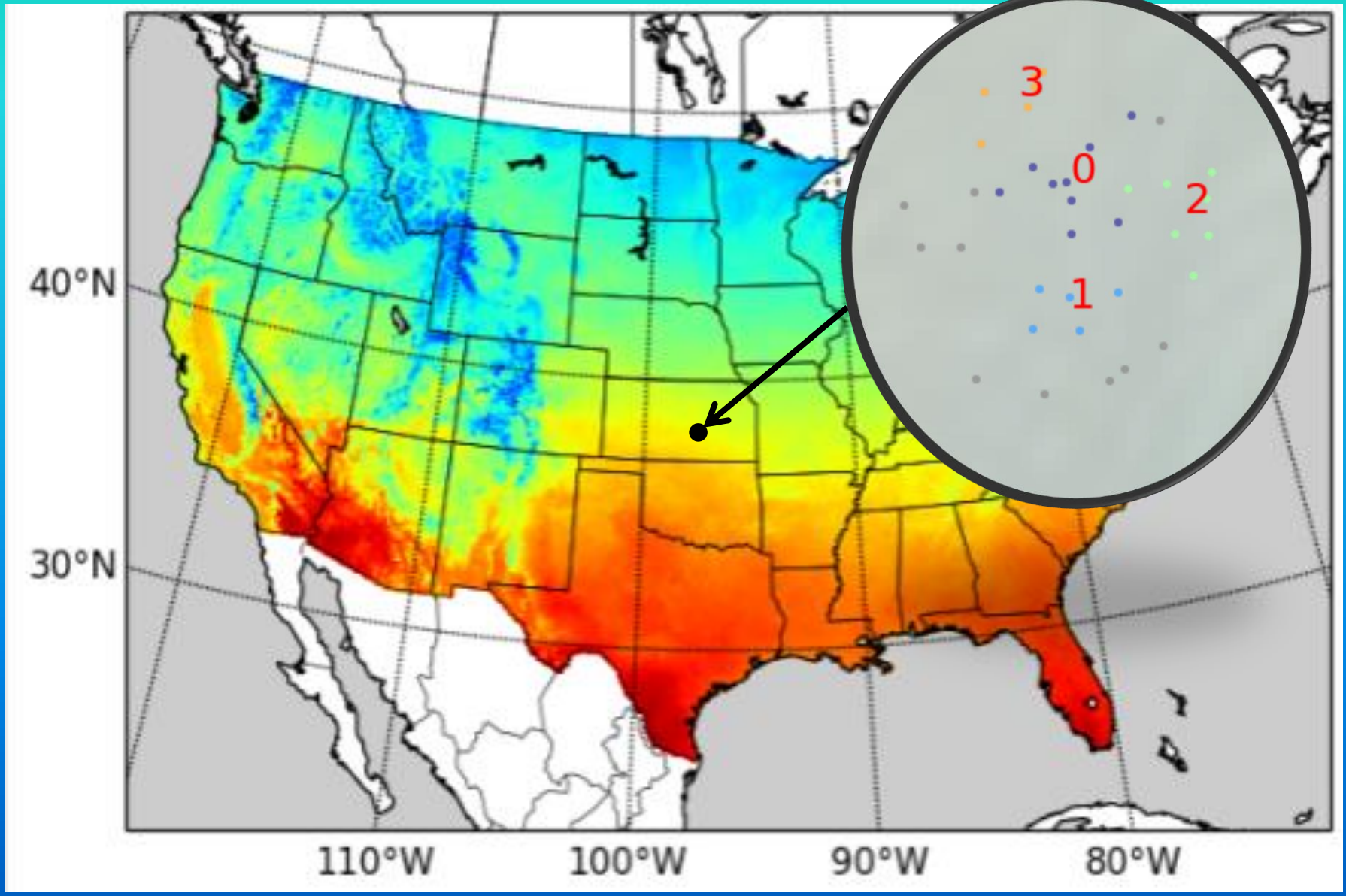
Visualizing All of the Points of Interest Together



DBSCAN: Basemap Projection



DBSCAN: Basemap Projection with Clusters



Running the K-Means Algorithm

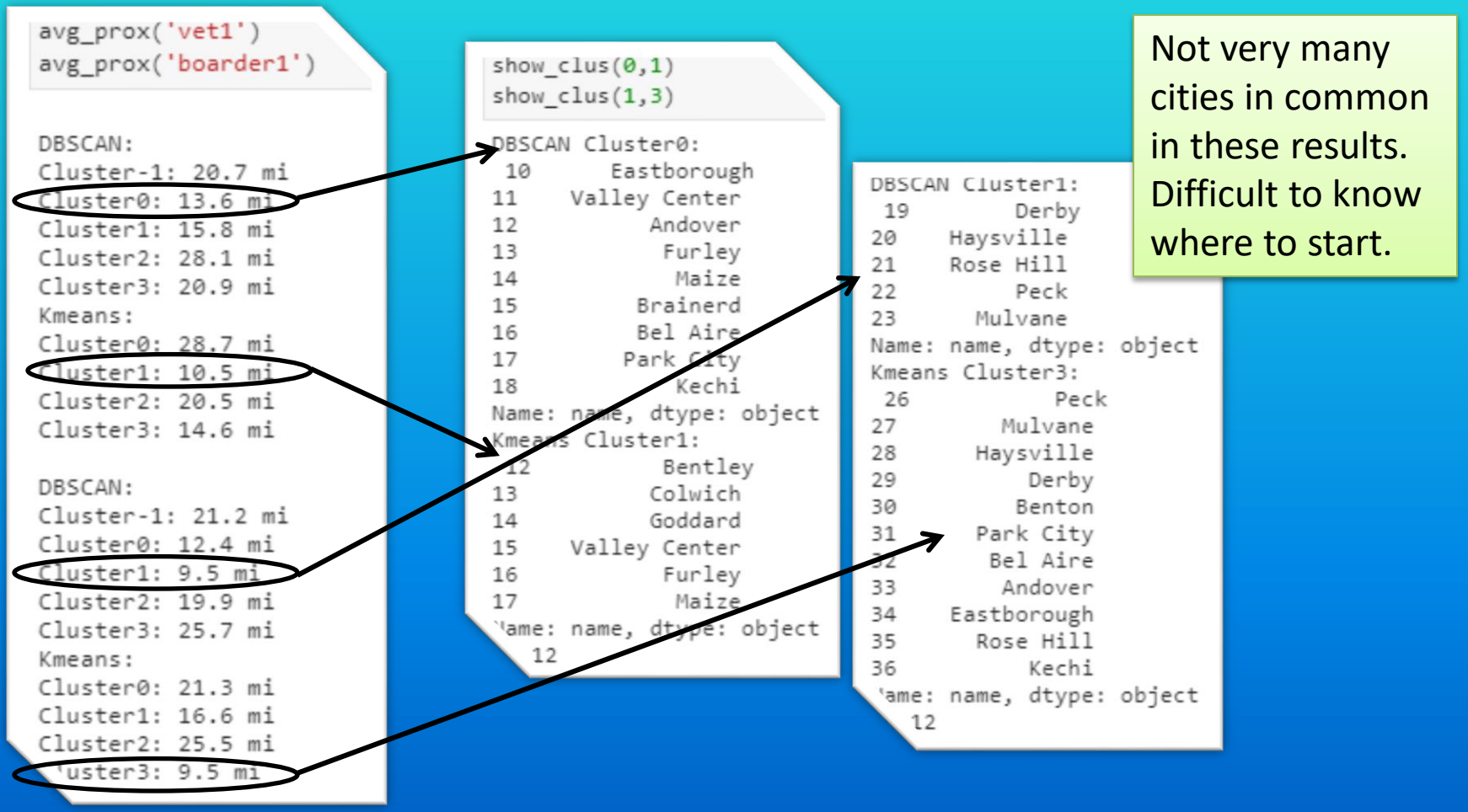
```
array([[ 0.99625018,  0.80403635,  0.59166208, -0.03325073,  0.41715964,
         0.36118272, -0.96687464, -0.0788286 , -0.0788286 ,  1.17716996,
         0.86933713, -0.03325073, -0.69638426, -0.69638426, -0.69638426,
         0.86364674, -0.96687464,  1.58969526,  1.58969526, -0.0788286 ,
        -0.0788286 , -0.0788286 ,  0.69254566,  1.08229488,  0.70230689,
         1.18628027,  0.83019794,  1.1915786 , -1.28076461,  0.95519003,
                                                            .96992638,
                                                            .63495083,
                                                            .44157394,
                                                            .57340814,
        -0.2153085 ,  0.52661936, -0.35131478, -0.35131478,  1.0626933 ,
         1.3287679 ,  1.28335376,  1.03634848,  1.03634848,  1.03634848,
         1.19899301,  0.52661936,  0.04630147,  0.04630147, -0.35131478,
        -0.35131478, -0.35131478,  1.2659048 ,  1.20014544,  0.77989547,
         1.01632583,  1.00006603,  1.10282777,  1.08730516,  1.27752102,
         0.14298684,  1.12385266,  1.18062147,  1.04630244,  0.96477512,
         1.12385266,  0.25101548,  0.07109735,  0.03125554,  1.18759234,
         0.01570363,  0.94456599,  1.34640569,  1.01908584,  0.47457467,
         0.47055173,  0.9051257 ,  0.19769969],
```

```
# retrieve labels for each point in the model and save to a variable
k_means_labels = k_means.labels_

k_means_labels
```

```
array([0, 0, 0, 0, 0, 2, 1, 2, 2, 1, 3, 1, 3, 1, 1, 0, 3, 3, 3, 3, 3, 3,
       3, 3, 0, 0, 3, 0, 0, 0, 0, 2, 1, 2, 2, 2, 2], dtype=int32)
```

Running the avg_prox() and show_clus() functions to determine average distance per cluster from the top-rated vet clinic and top-rated boarding facility



Running the avg_prox() and show_clus() functions again to determine average distance per cluster from the top two rated vet clinic and top two rated boarding facility

```
avg_prox('vet2')
avg_prox('boarder2')


DBSCAN:
Cluster-1: 22.1 mi
Cluster0: 10.1 mi
Cluster1: 11.8 mi
Cluster2: 18.8 mi
Cluster3: 23.3 mi
Kmeans:
Cluster0: 21.2 mi
Cluster1: 15.2 mi
Cluster2: 24.3 mi
Cluster3: 9.2 mi

DBSCAN:
Cluster-1: 21.6 mi
Cluster0: 10.2 mi
Cluster1: 12.8 mi
Cluster2: 21.5 mi
Cluster3: 21.7 mi
Kmeans:
Cluster0: 23.4 mi
Cluster1: 13.0 mi
Cluster2: 22.6 mi
Cluster3: 10.3 mi
```

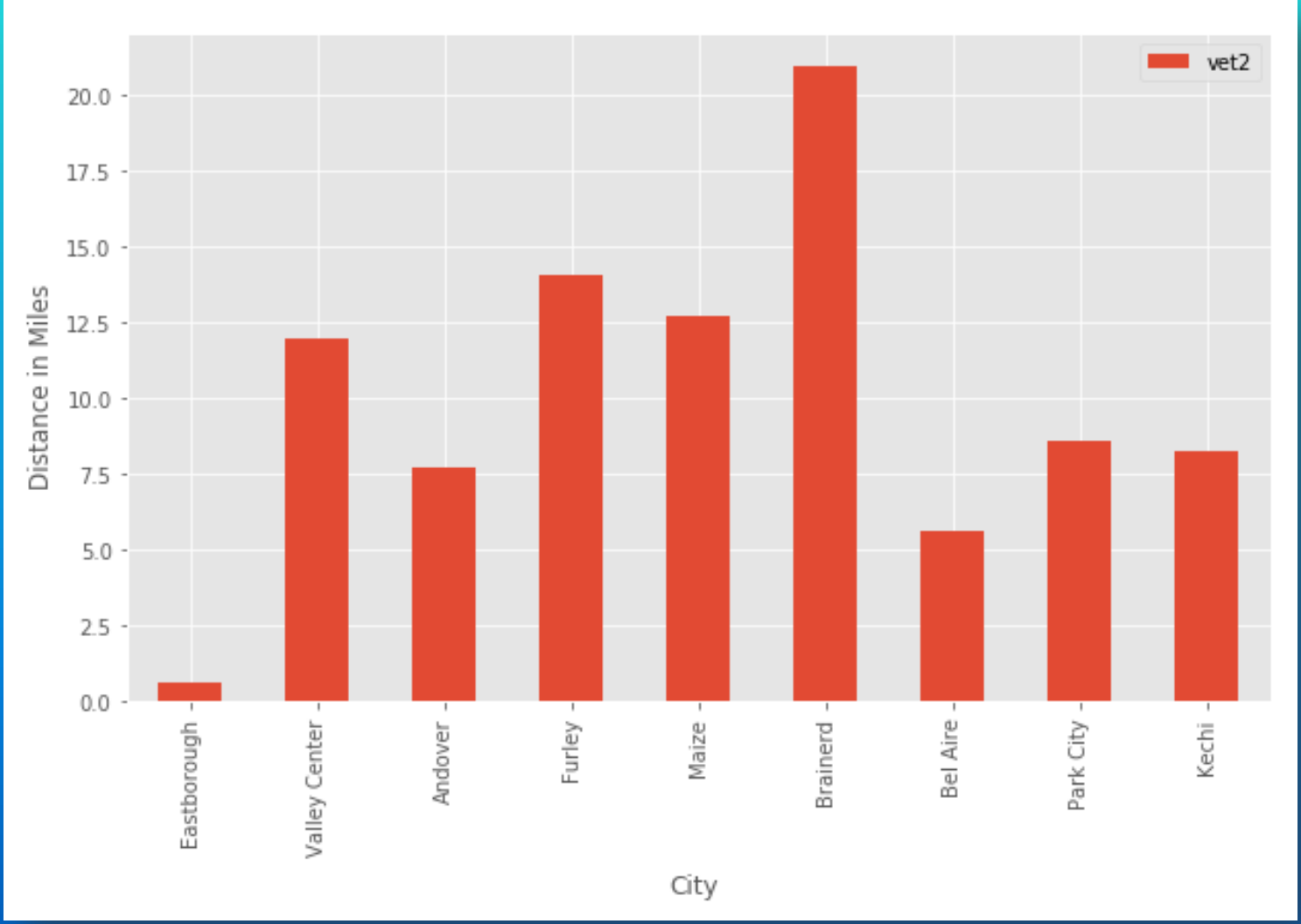
```
show_clus(0,3)

DBSCAN Cluster0:
10      Eastborough
11     Valley Center
12          Andover
13           Furley
14            Maize
15         Brainerd
16          Bel Aire
17         Park City
18            Kechi
Name: name, dtype: object
Kmeans Cluster3:
26             Peck
27         Mulvane
28       Haysville
29           Derby
30          Benton
31       Park City
32        Bel Aire
33         Andover
34     Eastborough
35       Rose Hill
36           Kechi
Name: name, dtype: object
12
```

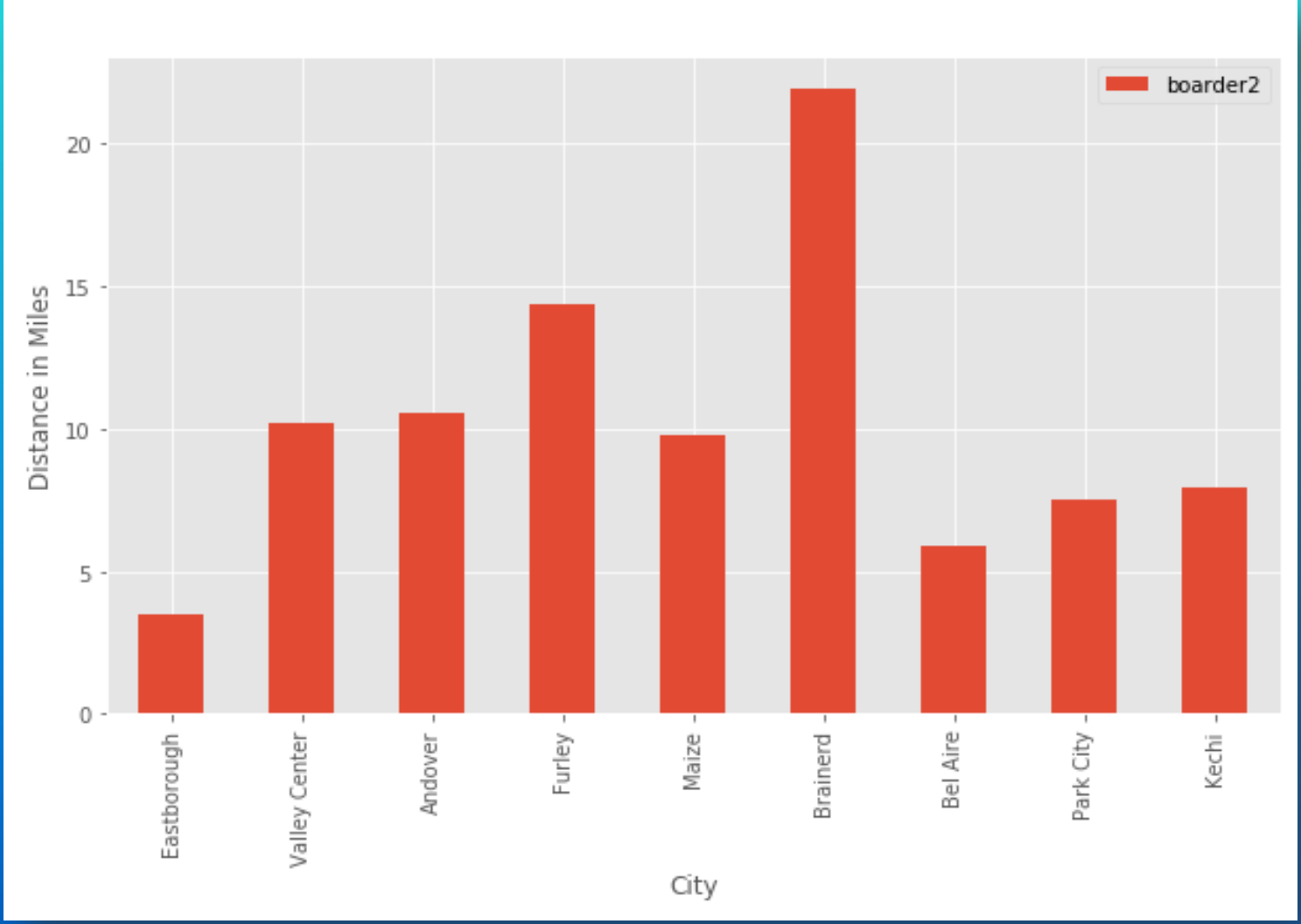
There are quite a bit more commonalities among these results.

Eastborough,
Andover,
Bel Aire,
Park City,
and **Kechi**
are all common to both models.

Bar Graph Showing Distances per City within DBSCAN Cluster 0 in relation to ‘vet2’



Bar Graph Showing Distances per City within DBSCAN Cluster 0 in relation to ‘boarder2’



Conclusion and Recommendation

- Using various procedures from the Data Science Methodology, I was able to take the client from having no particular destination in which to move her kennel to having a finely-tuned set of results to choose from.

- Once a city was chosen as a central hub, I was able to:
  - search for surrounding cities and towns in which to live
  - map out how the other breeders in the area were clustered
  - find the specific locations of veterinarians, pet boarders, and pet stores.

- After retrieving and storing all of this data, I was able to fit the dataset into two unsupervised, machine-learning models: DBSCAN and K-Means.

Conclusion and Recommendation

- I then ran two functions that I created to display the clusters from each model according to their average distance from whichever point of interest the client prefers. In this case, it was the top-rated vet clinic and top-rated boarding facility.

- Step by step, in this case study, it was discovered that:
  - Kansas is the best choice of states from the available data
  - Wichita is the most desirable city to be in proximity to in order to fit the client's specifications.
  - When measuring distance in proximity to the number two rated vet and boarding facility, there were more results that were in common, namely the cities of:
    - Eastborough
    - Andover
    - Bel Aire
    - Park City
    - Kechi

Conclusion and Recommendation

- I would eliminate Eastborough as an option because it is an enclave located within the city of Wichita. Bel Aire is also a bit too close to the city, where acreage would be more difficult to attain.

- My recommendation is that the client begins her search for real estate in one of the following areas first: Andover, Park City, or Kechi. If suitable land is not available in any of those locations, then I would suggest moving the search out to the second closest clusters produced by both the models and working her way out from there.