

Coursera Capstone

IBM Applied Data Science Capstone

Opening a Veterinary Clinic in the Bay Area

Shahbaz Masood

Feb 2019

Introduction

An uncle of mine is a veterinarian and wants to open up a clinic in the Bay Area (California) where the his clinic is most likely to succeed. This analysis aims to help all veterinarians to decide where to open up a pet clinic in the bay area.

Business Problem

The location of any consumer-oriented retail business is a key factor in its success. Same is true for a pet clinic. An ideal location should consist of many factors like pet friendliness of area, number of dog parks, number of pet food and service shops etc. The following analysis helps solve veterinarians to pick a spot where the clinic would have the most customers while considering the above mentioned factors.

Data

The analysis uses the data set **Bay_Area_cities.json** which was downloaded from the [github](#) repository of *FrankBowers43*. This json is used to get the names of all the neighborhoods in the bay area. The names are located in the features.properties.NAME10 node.

Methodology

I extracted the names of the Bay Area neighborhoods in the JSON file in a DataFrame.

These names are then fed to the Google Geocoder API to fetch the longitude and latitude values of each neighborhood. The locations are then fed into the foursquare API to get the 5 venue categories from the FourSquare API (i.e Dog Run ,Pet Café ,Pet Service ,Pet Store and 'Veterinarian'). A function is written to perform this task. It

returns a DataFrame which houses all the neighborhoods and all the venues that were found for that neighbourhood.

The FourSquare API returns some venues where the city attribute in the result is not found. The function that fetches the venues caters for this.

Some data cleaning is performed on the data.

- Changed the string Café to Cafe
- Filtered to the relevant categories only
- Filtered to keep the rows based either of the following conditions
 - The venue city is equal to the neighborhood
 - The venue city is 'Not found'
 - The neighborhood contains Richmond if venue city is Richmond (this is because Richmond in the bay area data is populated as Richmond, North Richmond and East Richmond)
- Duplicates are removed based on the Neighborhood,Venue_Name,Venue_City and Venue Address

This leaves us with a data set of 850 venues. Here is a snapshot of the data.

```
[43]:
```

	Neighborhood	Venue	Venue_City	Venue_Address	Venue_Type
3	Green Valley	Ruff Dog Hotel & Daycare	Not found	California United States	Pet Store
14	Green Valley	Grooming Angels	Not found	California United States	Pet Store
25	Boulder Creek	Boulder Creek Veterinary Clinic	Boulder Creek	12870 Highway 9 Boulder Creek, CA 95006 United...	Veterinarian
27	Boulder Creek	Boulder Creek Feed & Supply	Boulder Creek	13101 Highway 9 (Forest Ave) Boulder Creek, CA...	Pet Store
54	Sunol	Happiness Country Kennels	Sunol	5815 Mission Rd Sunol, CA 94586 United States	Pet Service

Dummy variables are created for the Venue_Type column which results in this:

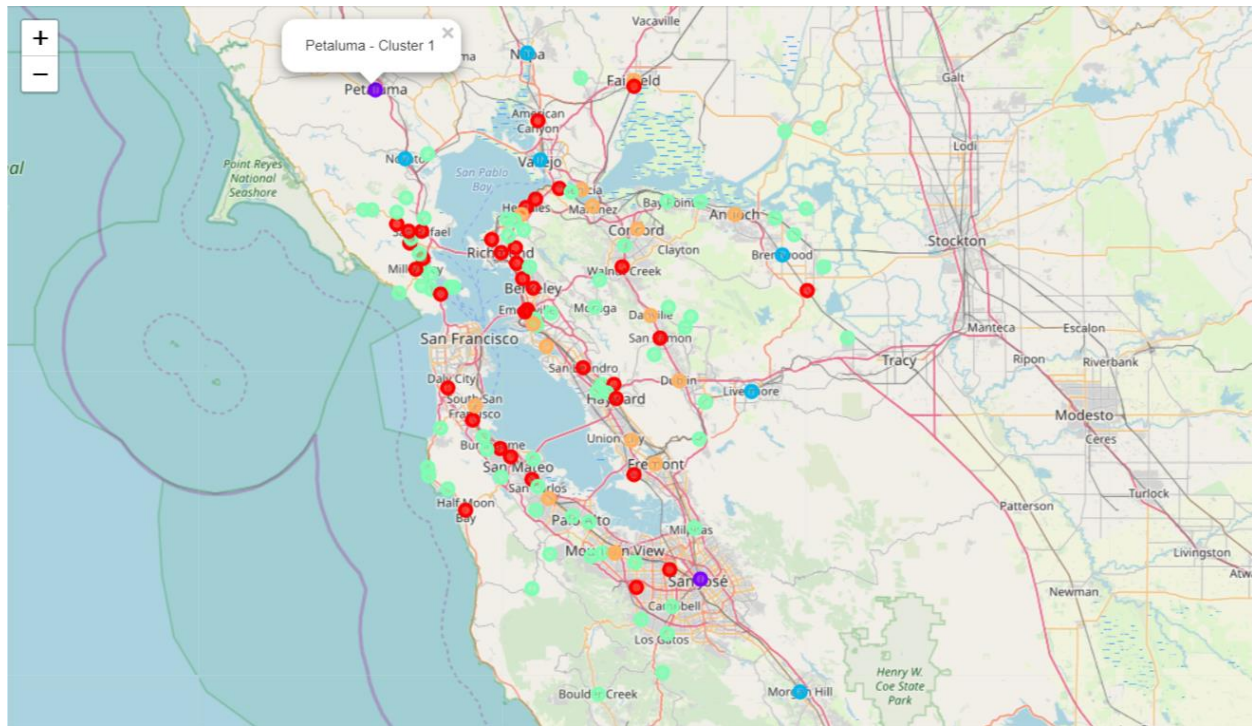
```
[46]:
```

	Neighborhood	Dog Run	Pet Cafe	Pet Service	Pet Store	Veterinarian
3	Green Valley	0	0	0	1	0
14	Green Valley	0	0	0	1	0
25	Boulder Creek	0	0	0	0	1
27	Boulder Creek	0	0	0	1	0
54	Sunol	0	0	1	0	0

Now the idea here is that the place which has more of Dog Run, Pet Café, Pet Service and Pet Stores and less of Veterinians, is the place where there is a shortage of vets and thus should prove to be an ideal place for a vets clinic.

As a result the columns Dog Run, Pet Café, Pet Service and Pet Stores is summed and the result is subtracted from the Veterinarian column. The max of the result is where the ideal location should be.

I did that and performed kMeans on the analysis on the difference of the values to find the neighborhoods with a similar difference.



Results

The clusters showed that San Jose and Petaluma are the best places to open up a Vet Clinic.

Discussion

Even though the results show San Jose and Petaluma but a physical survey of the area is strongly advised. The population of the two districts should also be considered in making the decision along with commercial rents etc.

Having said that, the analysis does a good job in defining the starting point for conducting the research. It also highlights the areas to avoid.

Conclusion

The analysis provides a starting point for doing a physical survey of a good choice of starting and establishing vet clinic.