BOSTON PET STORE BUSINESS

by Andres Macana

1. Introduction

In this time people are worried about pets, from spiders to horses, and now that people spend more time in social media, many of them decide to have a pet friend. The problem in many cities is access to the pet store, sometimes are far away from home. In this case, I'm not only want to analyze where the pet store is, but I also want to add two components, a coffee shop and a pizza place. Why? Because I think that the most popular venues can help to develop a new business helping with a big flux of people

2. Data

The data come from analyze Boston, an open source where different format like json and csv are available to the public. The hard part is cleaning the data, not all files are clear and easily can take hours.
I decide to go with csv file because in this the latitude and longitude were already separated.

[1] http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0a1d0a1a28_0.geojson

[2] http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0a1d0a1a28_0.csv

3. Methodology

- Get the libraries. First we need to load some libraries like a pandas, numpy, json, geopy, matplotlib and folium.

- Get the data. From the URLs above in the data section, we can see the information inside and start to make some preliminary data frames. We obtain a first data frame after cleaning:

	name	neighborhood_ID	latitude	longitude
0	Roslindale	15	42.2825152806	-71.126661984
1	Jamaica Plain	11	42.3077608417	-71.1150963368
2	Mission Hill	13	42.3317328217	-71.1026060494
3	Longwood	28	42.3386002118	-71.1051020737
4	Bay Village	33	42.3490945737	-71.0690314053
5	Leather District	27	42.3509742734	-71.0578275043
6	Chinatown	26	42.3490186627	-71.0616614624
7	North End	14	42.3654556539	-71.0538566471
8	Roxbury	16	42.3185099075	-71.0866426283
9	South End	32	42.3413417355	-71.0713687015
10	Back Bay	2	42.3498518529	-71.080534767

We can obtain another data frame by categories using Foursquare, helping us to understand the data.

	name	categories	lat	lng
0	Fornax Bread Company	Bakery	42.286170903	-71.1297603744
1	Roslindale Square	Plaza	42.2858151455	-71.1293042506
2	Roslindale Village Farmers Market	Farmers Market	42.2865337145	-71.1285093709
3	Delfino’s	Italian Restaurant	42.2871064436	-71.1294698416
4	Sophia's Grotto	Italian Restaurant	42.2866163419	-71.1298693998
5	Roslindale House Of Pizza	Pizza Place	42.2879894322	-71.1265487086
6	753 South	American Restaurant	42.2871082715	-71.1293410553
7	Effie's Kitchen	Greek Restaurant	42.2858985037	-71.1287414129
8	La Taqueria	Mexican Restaurant	42.27865	-71.1191
9	Blue Star Restaurant	Restaurant	42.2865955377	-71.1303937904
10	BK's Pub	Bar	42.2851666683	-71.129425084
11	Jimmies Ice Cream Cafe	Ice Cream Shop	42.285873481	-71.1291909767
12	Boston Cheese Cellar	Cheese Shop	42.2867859068	-71.1300612358
13	Napper Tandy's	Pub	42.2872416313	-71.1274545126
14	Target	Big Box Store	42.2882044205	-71.1266589544
15	SUBWAY	Sandwich Place	42.2859354187	-71.1286689297
16	BCYF- Flaherty Pool	Pool	42.2881334672	-71.1229132685
17	CVS pharmacy	Pharmacy	42.2847547128	-71.1309869567
18	Shanti Taste of India Roslindale	Indian Restaurant	42.2871526584	-71.1276700792
19	Birch Street Bistro	American Restaurant	42.2867761078	-71.1298665625
20	Romano's Pizzeria & Taqueria	Pizza Place	42.2857462516	-71.1290828608
21	Family Dollar	Discount Store	42.2863607385	-71.1298713185
22	Pet Cabaret	Pet Store	42.2814613492	-71.1334794052
23	J.J. Brannelly's	Bar	42.2802856295	-71.1346352984
24	Dunkin'	Donut Shop	42.2879118896	-71.1269911749
25	Village Market	Grocery Store	42.2861702613	-71.1294579771

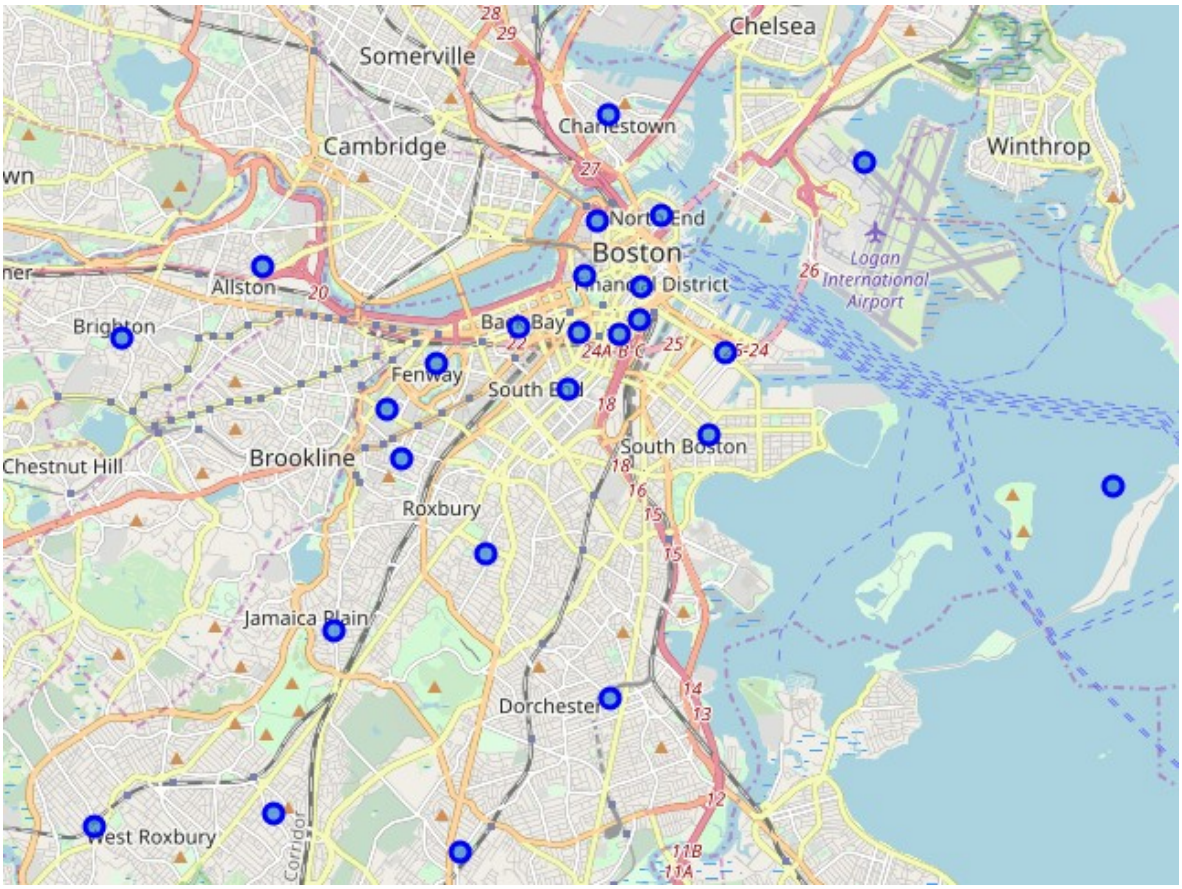
A more structured data frame with the coordinates of neighborhood and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roslindale	42.2825152806	-71.126661984	Fornax Bread Company	42.286170903	-71.1297603744	Bakery
1	Roslindale	42.2825152806	-71.126661984	Roslindale Square	42.2858151455	-71.1293042506	Plaza
2	Roslindale	42.2825152806	-71.126661984	Roslindale Village Farmers Market	42.2865337145	-71.1285093709	Farmers Market
3	Roslindale	42.2825152806	-71.126661984	Delfino’s	42.2871064436	-71.1294698416	Italian Restaurant
4	Roslindale	42.2825152806	-71.126661984	Sophia's Grotto	42.2866163419	-71.1298693998	Italian Restaurant
5	Roslindale	42.2825152806	-71.126661984	Roslindale House Of Pizza	42.2879894322	-71.1265487086	Pizza Place
6	Roslindale	42.2825152806	-71.126661984	753 South	42.2871082715	-71.1293410553	American Restaurant
7	Roslindale	42.2825152806	-71.126661984	Effie's Kitchen	42.2858985037	-71.1287414129	Greek Restaurant
8	Roslindale	42.2825152806	-71.126661984	La Taqueria	42.27865	-71.1191	Mexican Restaurant
9	Roslindale	42.2825152806	-71.126661984	Blue Star Restaurant	42.2865955377	-71.1303937904	Restaurant
10	Roslindale	42.2825152806	-71.126661984	BK's Pub	42.2851666683	-71.129425084	Bar

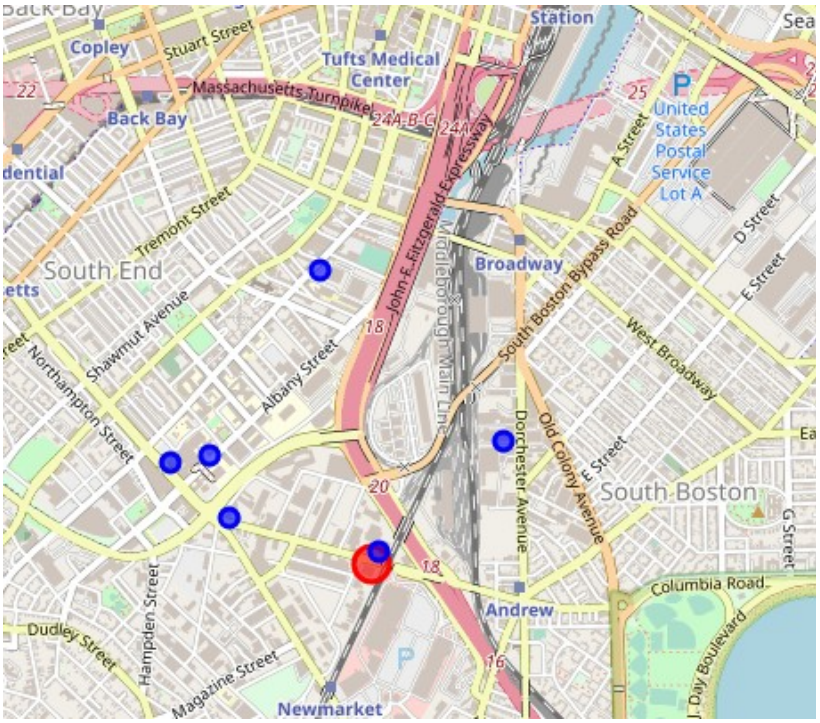
And a data frame with the most popular venues. In our case, this information is very important because we start to check the hot spots.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Allston	Korean Restaurant	Thai Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Pizza Place	Rental Car Location	Asian Restaurant	Seafood Restaurant
1 Back Bay	Coffee Shop	American Restaurant	Hotel	Italian Restaurant	Seafood Restaurant	Cosmetics Shop	Salon / Barbershop	Ice Cream Shop	Juice Bar	French Restaurant
2 Bay Village	Spa	Theater	American Restaurant	Italian Restaurant	Gym	Steakhouse	Coffee Shop	Hotel	Seafood Restaurant	Performing Arts Venue
3 Beacon Hill	Italian Restaurant	Hotel	Pizza Place	Sandwich Place	Coffee Shop	Gourmet Shop	Spa	French Restaurant	Park	Theater
4 Brighton	Pizza Place	Pub	Chinese Restaurant	Bakery	Café	Sushi Restaurant	Coffee Shop	Grocery Store	Greek Restaurant	Convenience Store
5 Charlestown	Pizza Place	National Park	Coffee Shop	Grocery Store	Gastropub	Pub	Gym	Convenience Store	Seafood Restaurant	Bus Station
6 Chinatown	Chinese Restaurant	Asian Restaurant	Bakery	Coffee Shop	Italian Restaurant	Sushi Restaurant	Theater	Vegetarian / Vegan Restaurant	Pizza Place	Sandwich Place
7 Dorchester	Vietnamese Restaurant	Café	Park	Sandwich Place	Food	Fast Food Restaurant	Southern / Soul Food Restaurar	Liquor Store	Market	Seafood Restaurant
8 Downtown	Coffee Shop	Historic Site	Italian Restaurant	Salad Place	Park	NewAmerican Restaurant	Sandwich Place	Seafood Restaurant	Hotel	American Restaurant
9 East Boston	Donut Shop	Coffee Shop	American Restaurant	Pizza Place	Airport Lounge	Brewery	Café	Airport	Airport Service	Border Crossing
10 Fenway	American Restaurant	Thai Restaurant	Restaurant	Burger Joint	Café	Hotel	Pizza Place	Lounge	Baseball Field	Sandwich Place

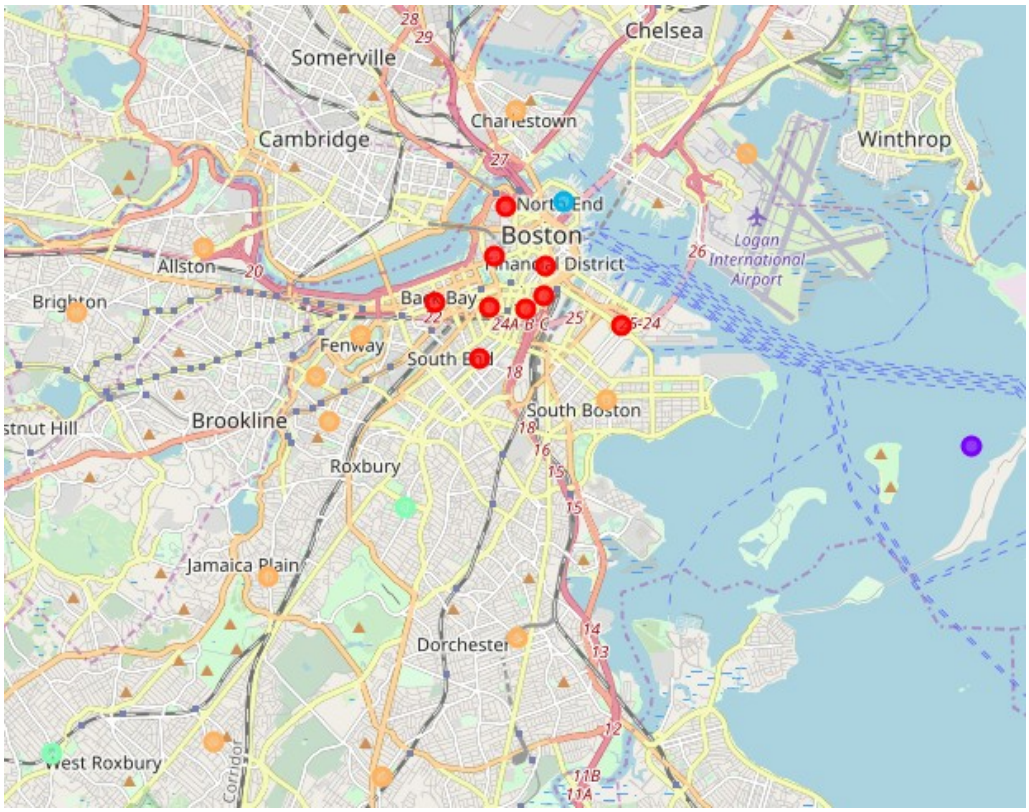
- Start to visualize. Now we can introduce a Boston city map related to business. Initially, I put a basic Boston map with the neighborhoods.



A first look of the pet stores. We can observe that there are many places without a pet store. I use like a reference (red point) the Pet Store in 274 Southampton with a radius of 100m.



Now we can visualize the clusters made with a k = 5. And we can see that the most popular places are not close enough to the pet stores. I will use this characteristic to start planning where to open a new pet store.



- Get the information to start to clustering the neighborhoods. We can obtain a valuable information about the clustering and showing the results with the venues and the frequency for each one. Some examples are presented below.

```
----Allston----
                  venue  freq
0     Korean Restaurant  0.06
1               Bakery  0.04
2       Thai Restaurant  0.04
3    Chinese Restaurant  0.04
4           Coffee Shop  0.04
5      Sushi Restaurant  0.03
6           Pizza Place  0.03
7   Rental Car Location  0.03
8    Seafood Restaurant  0.02
9  Gym / Fitness Center  0.02


----Back Bay----
                  venue  freq
0          Coffee Shop  0.06
1  American Restaurant  0.06
2   Italian Restaurant  0.05
3               Hotel  0.04
4   Seafood Restaurant  0.04
5       Cosmetics Shop  0.03
6   Salon / Barbershop  0.03
7       Ice Cream Shop  0.03
8       Chocolate Shop  0.02
9            Juice Bar  0.02
```

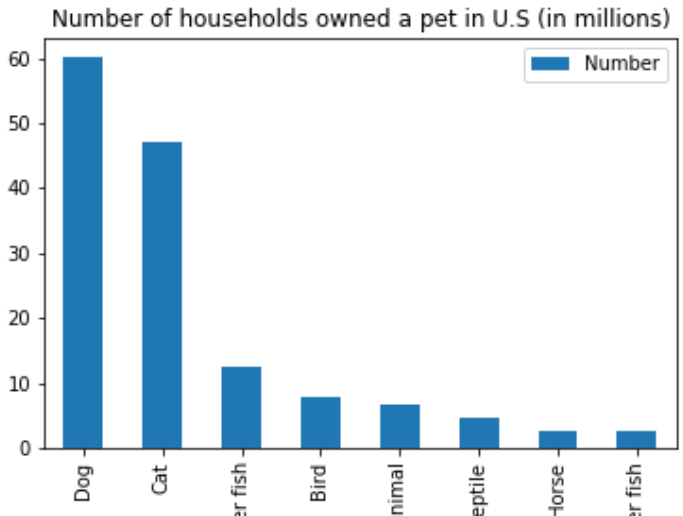
```
----Brighton----
                  venue  freq
0           Pizza Place  0.09
1                   Pub  0.07
2                 Café  0.05
3     Sushi Restaurant  0.05
4   Chinese Restaurant  0.05
5               Bakery  0.05
6          Coffee Shop  0.05
7       Sandwich Place  0.03
8  American Restaurant  0.03
9         Deli / Bodega  0.03


----Charlestown----
                venue  freq
0        Pizza Place  0.06
1          Gastropub  0.04
2                Pub  0.04
3                Gym  0.04
4      Grocery Store  0.04
5               Café  0.04
6        Coffee Shop  0.04
7      National Park  0.04
8  Convenience Store  0.04
```

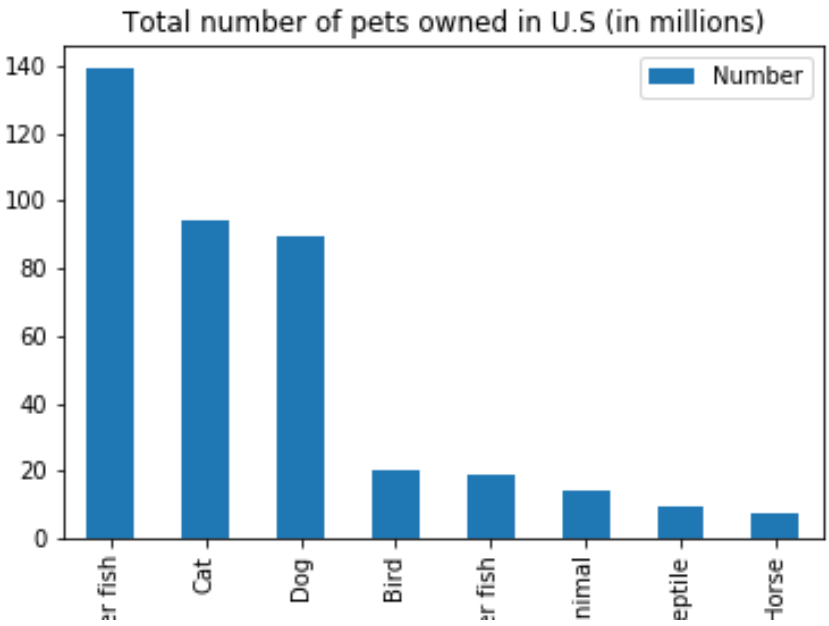
- Analize the clusters

Now we have an idea of the top business by neighborhood in Boston city. In this point, we can return to the idea presented in the introduction. Now we know the possibles neighborhoods where start a pet store without interfering with the other pet stores and using the popularity of the venue to attract the customers. According to Pets Statistics [3], sixty-eight percent own a pet (around 85 million) between 2017 and 2018.

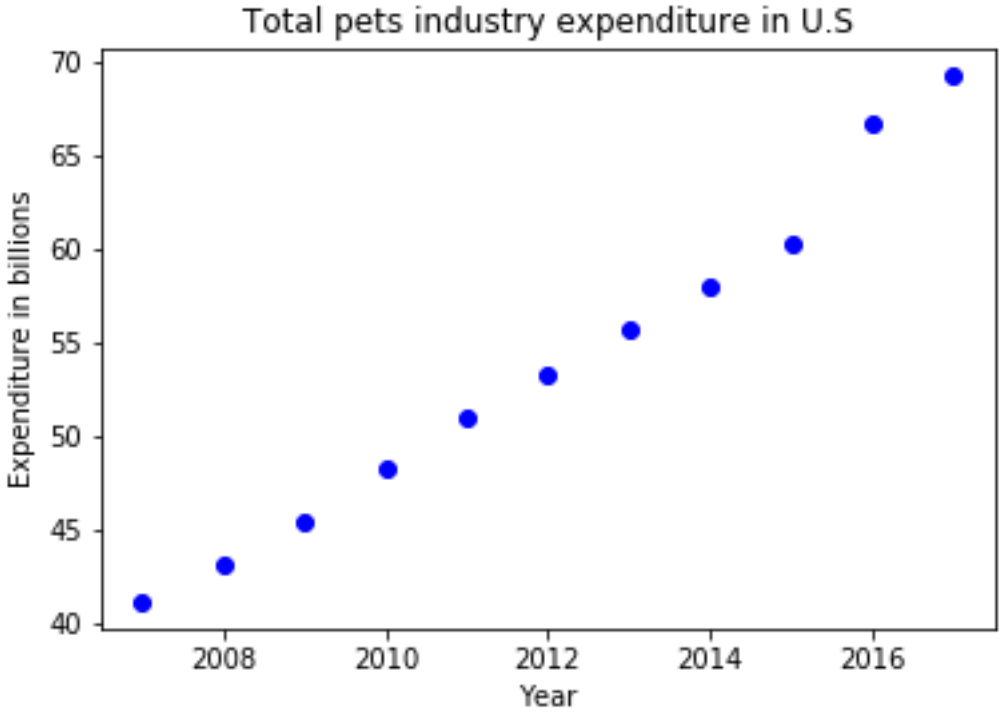
The number of U.S households that own a pet by type of animal [3] (millions)



Total number of pets owned in the U.S by type of animal [3] (millions)

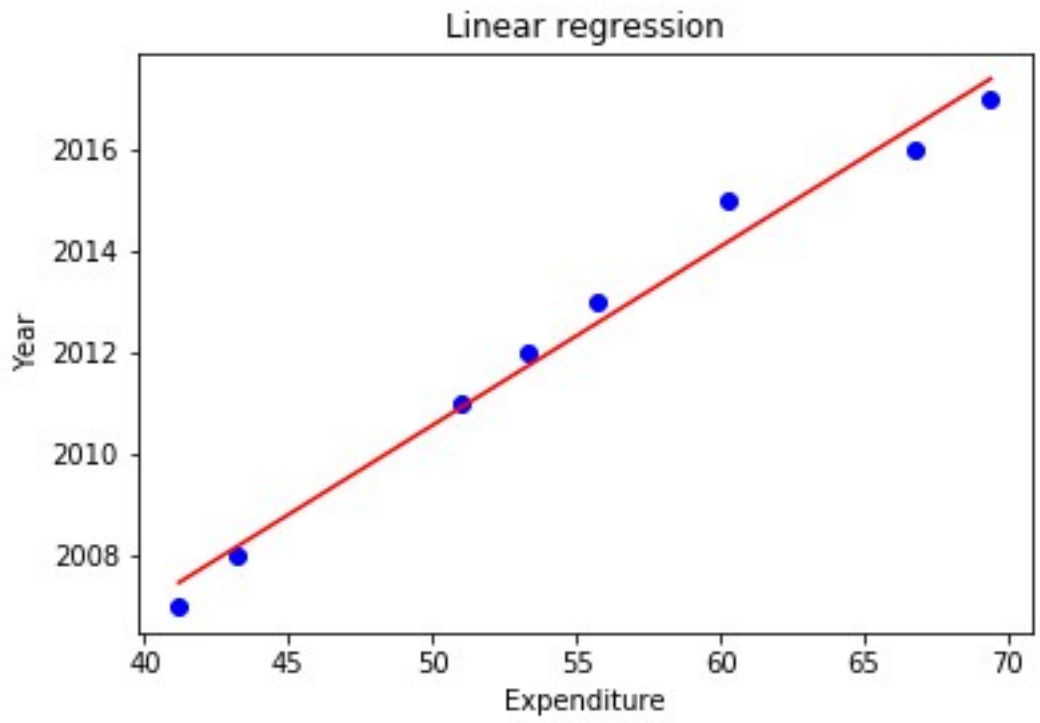


Total of U.S pet industry expenditures, 2007-2017 [3] ($billions)



4. Results

Using linear log regression we can sow that the number of pets in the U.S is increasing every year and we can make a prediction of the amount of money expended. These results show us that open a pet store will be a good idea. The clustering algorithm helps us too to make a good choice of places where the pet store will be open and attaching close to the most popular venues in the city of Boston. This attaches will be helpful because we can assure a flux of possibles customers.



We obtain:
MSE = 0.12
R2-Score = 0.97

5. Discussion

The mix between Foursquare and K-means clustering is really helpful obtaining information about the most popular places and where we can find them. In the cities, clustering shows us how society is moving or how society start to make clusters naturally, but the algorithm can clarify some issues about these clusters.

6. Conclusion

The pets industry graphic shows us that this business grows up, the linear regression shows us that with great confidence that this kind of business is not a bad idea because people will have more pets and they will be worried about their pets. With the help of clustering analysis we can target some places in the city of Boston where open a pet store will be more successful, that's meaning that we can use this targets far away from other stores and we can select a place close to another business where the flux of people is bigger, taking the information about the most common venues.

7. References

[1] http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0a1d0a1a28_0.geojson

[2] http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0a1d0a1a28_0.csv

[3] https://www.iii.org/fact-statistic/facts-statistics-pet-statistics#NUMBER%20OF%20U.S.%20HOUSEHOLDS%20THAT%20OWN%20A%20PET,%20BY%20TYPE%20OF%20ANIMAL