

IBM Capstone Battle of the Neighborhoods

JUNE 1

Helena Lewczynska



Logo
Name

Table of Contents

Contents

Table of Contents	2
Introduction	3
Methodology	4
Data.....	5
Seattle Neighbourhood Data.....	5
Foursquare Location Data	6
Seattle pet data	7
Analysis	8
Exploratory Data Analysis	8
Further Data Analysis.....	11
Results and Discussion.....	13
Conclusion	14

Introduction

My partner and I have recently decided to move from Houston to Seattle. I know. expensive right! I have toyed with the idea of setting up a dog walking business as I love my furry friends and now have a pup of my own called Ellie.

So, the questions are:

- 1) Which neighborhoods suit my family and me?
- 2) Which neighborhoods in Seattle are most likely to require someone to walk their dog?
- 3) Do any of these neighborhoods provide the best of both worlds?

In this project, I will attempt to answer the above questions. First, I will explore the best neighborhoods that fit my own family life. Next, I will find which neighborhoods in Seattle would probably need a dog walker. Finally, I will try and find out which areas would provide the best of both!

To complete this task, I will use Foursquare location data to get the most common venue categories in each neighborhood and map which neighborhoods contain our favorite ones. Then I will use some data I found on Seattle pet populations to find which neighborhoods contain the most dogs and of what breed. Lastly, I will try to analyze which neighborhoods would best suit my lifestyle but also give me the ability to start up my dog walking business.

Methodology

My partner and I have recently decided to move from Houston to Seattle. I know, expensive right! I have toyed with the idea of setting up a dog walking business as I love my furry friends and now have a pup of my own called Ellie.

To complete this task, I will perform the following steps:

- Collecting data: Including Seattle Neighborhood data, venues nearby each neighborhood which will be generated by Foursquare API and pet data from the Seattle Open-Source Website
- Exploratory data analysis to better understand our working dataset.
- Creating Map of Seattle with Folium library overlaying neighborhood characteristics and pet data.
- Analyzing which neighborhoods, we should consider based on analysis of pets and venues.

Data

I did not find any dataset containing both Seattle neighborhoods and coordinates, so I found two datasets and merged them together:

1. Seattle neighborhood and zip code info:

<http://seattlearea.com/zip-codes/>

2. Seattle coordinates and zip code

<https://www.unitedstateszipcodes.org/wa/#zips-list>

In the next section I will describe preliminary data importing and preparation.

Seattle Neighborhood Data

Here's the top 5 rows of the neighborhood data:

	Zipcode	Neighborhood
0	98003	Federal Way
1	98005	Bellevue
2	98033	Kirkland
3	98037	Lynnwood
4	98040	Mercer Island

Here are the top 5 rows of the geospatial data:

	Zipcode	Latitude	Longitude
0	501	40.81	-73.04
1	544	40.81	-73.04
2	601	18.16	-66.72
3	602	18.38	-67.18
4	603	18.43	-67.15

Merging them together gives:

	Zipcode	Latitude	Longitude	Neighborhood
20	98110	47.64	-122.53	Bainbridge Island
17	98107	47.67	-122.38	Ballard
1	98005	47.62	-122.16	Bellevue
25	98121	47.62	-122.35	Belltown
9	98102	47.63	-122.32	Capital Hill

Foursquare Location Data

Now I will use the Foursquare API to identify venues in each neighborhood which I like the best. For generating the nearby venues, I have chosen a radius of 2km and limit to the top 80 venues. The venues dataset has 1557 rows and 7 columns.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bainbridge Island	47.64	-122.53	Storyville Coffee Roasting Studio	47.648323	-122.523960	Coffee Shop
1	Bainbridge Island	47.64	-122.53	Fletcher Bay Winery	47.648285	-122.525015	Wine Bar
2	Bainbridge Island	47.64	-122.53	Bainbridge Island Brewing	47.648299	-122.525042	Brewery
3	Bainbridge Island	47.64	-122.53	Jake's Pickup	47.636094	-122.516740	Café
4	Bainbridge Island	47.64	-122.53	Bainbridge Island Aquatic Center	47.638323	-122.522930	Gym Pool

Seattle pet data

For the Seattle data on pets, I have used the Open Data Program:

<https://data.seattle.gov/>

generated by the City of Seattle. This makes data openly available to the public for the purpose of increasing the quality of life for residents; increasing transparency, accountability, and comparability; promoting economic development and research; and improving internal performance management.

I found the following dataset available:

<https://data.seattle.gov/Community/Seattle-Pet-Licenses/jguv-t9rb>

which is a list of active/current Seattle pet licenses, including animal type (species), pet's name, breed, and the owner's ZIP code.

	License Issue Date	License Number	Animal's Name	Species	Primary Breed	Secondary Breed	Zipcode
0	December 18 2015	S107948	Zen	Cat	Domestic Longhair	Mix	98117
13	May 16 2018	S126432	Chico	Cat	Domestic Shorthair	Mix	98122
15	July 10 2018	85322	Zen	Cat	Domestic Shorthair	Siamese	98105
16	August 13 2018	578368	Lucky	Cat	Oriental Shorthair	Mix	98106
17	August 30 2018	S121902	Buffy	Cat	Siamese	European Shorthair	98107

Analysis

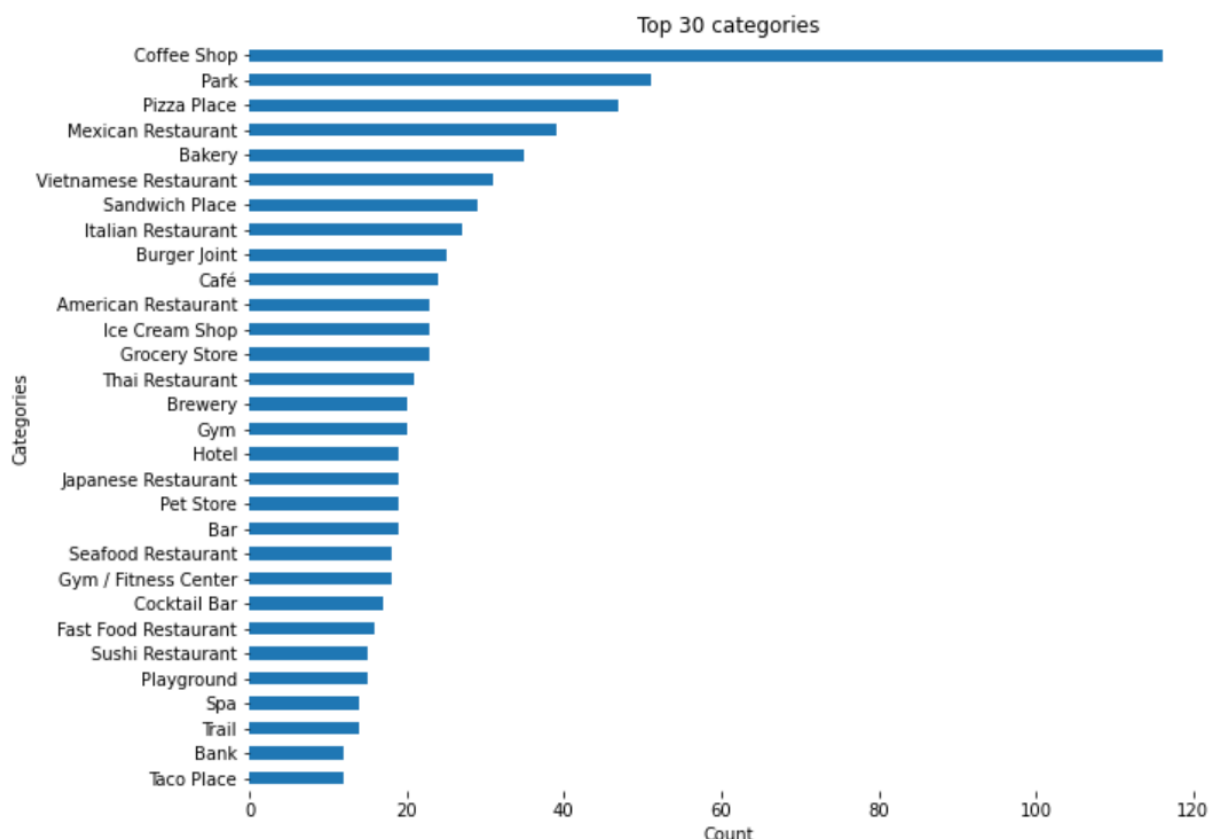
Exploratory Data Analysis

In the section, we will explore the venue and pet data and start with some exploratory data analysis.

Checking how many neighborhoods we successfully got venues from gives 21 which matches our neighborhood dataset.

Only 1 neighborhood: Bainbridge – had less than 12 venues and only 3 neighborhoods had less than 50 venues within a 2km radius.

The top 30 venues are:



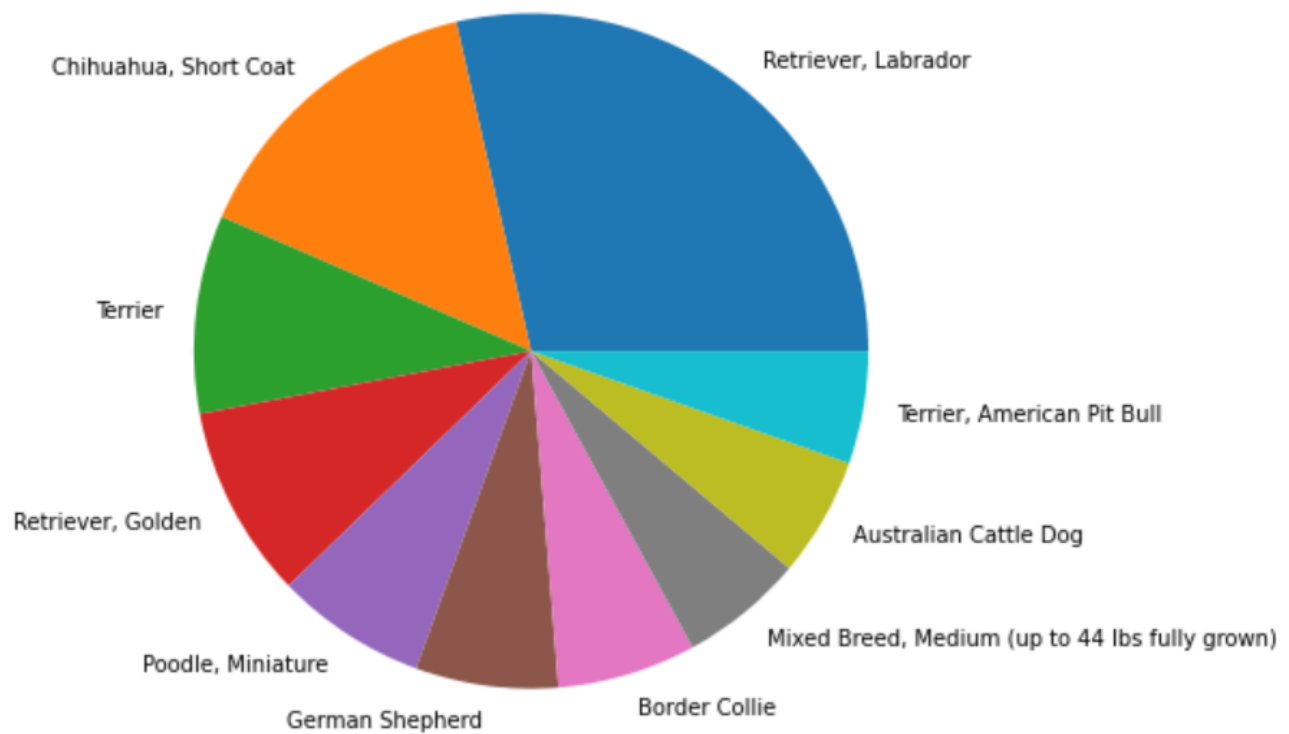
From the pet dataset, there are a total of 18110 dog licenses, 7873 cat licenses and 11 goat licenses in the City of Seattle.

The top 3 most popular pet names are:

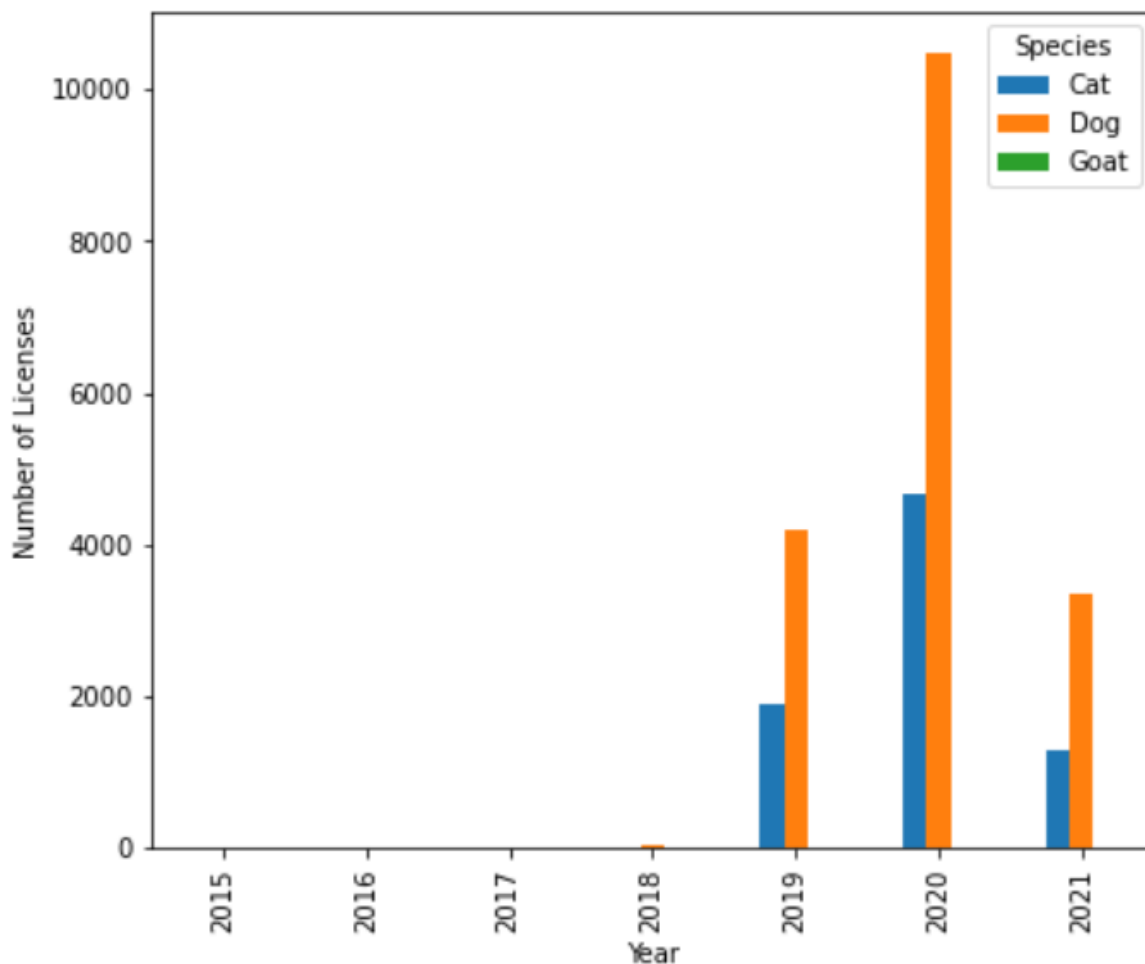
- Dogs: Luna, Charlie, Lucy

-
- Cats: Luna, Lily, Charlie
 - Goats: Estelle, Coco, Charlotte

The following pie chart shows the most common species of dog:



This bar chart helps us show pet licenses from the year 2015-2021:



The final plot does not tell us not much data was collected in the years before 2019 so the License Issue data probably won't be useful

That's my initial exploratory analysis done - lets delve deeper into the dataset.

The thing my family enjoys the most is the outdoors. Therefor I am most interested in Neighborhoods with **Trails, Parks** and **Scenic Lookouts**.

We love a good brewery - especially having a beer at the end of a long bike ride. So, let's look for **Breweries**.

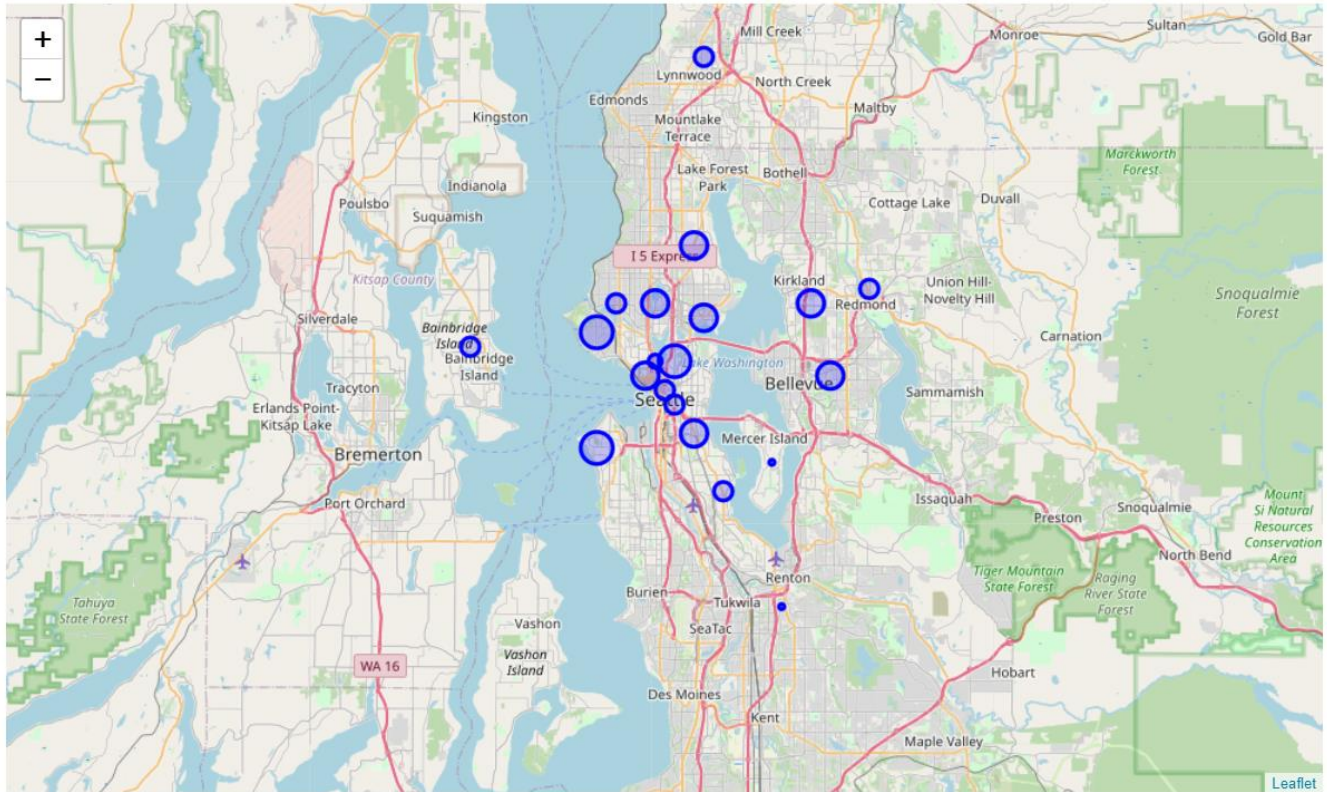
One last thing - our favorite food is **burgers** or **sushi**

Further Data Analysis

In this section, I put my favorite categories into a separate column and calculate the frequency of occurrence of each category:

	Neighborhood	Park	Trail	Scenic Lookout	Brewery	Sushi Restaurant	Burger Joint	Total	Zipcode	Latitude	Longitude
0	Bainbridge Island	1	0	1	1	0	0	3	98110	47.64	-122.53
1	Ballard	0	0	0	1	1	1	3	98107	47.67	-122.38
2	Bellevue	1	0	0	1	1	1	4	98005	47.62	-122.16
3	Belltown	1	0	1	1	1	0	4	98121	47.62	-122.35
4	Capital Hill	1	1	1	0	1	1	5	98102	47.63	-122.32
5	Columbia City	1	1	0	0	0	1	3	98118	47.54	-122.27
6	Federal Way	0	0	0	0	0	1	1	98003	47.30	-122.31
7	Greenwood	1	1	0	0	1	1	4	98103	47.67	-122.34
8	Kirkland	1	1	0	1	0	1	4	98033	47.67	-122.18
9	Lynnwood	1	0	0	0	1	1	3	98037	47.84	-122.29
10	Magnolia	1	1	1	1	0	1	5	98199	47.65	-122.40
11	Mercer Island	1	0	0	0	0	0	1	98040	47.56	-122.22
12	Mount Baker	1	1	1	1	0	0	4	98144	47.58	-122.30
13	Northgate	0	1	0	1	1	1	4	98125	47.71	-122.30
14	Pioneer Square	1	0	0	0	1	1	3	98104	47.60	-122.32
15	Redmond	1	1	0	0	0	1	3	98052	47.68	-122.12
16	Renton	0	0	0	0	1	0	1	98055	47.46	-122.21
17	Seattle	0	0	0	1	1	1	3	98101	47.61	-122.33
18	South Lake Union	1	0	1	0	0	0	2	98109	47.63	-122.34
19	University District	1	1	0	1	0	1	4	98105	47.66	-122.29
20	West Seattle	1	0	1	1	1	1	5	98116	47.57	-122.40

Now we can plot the data using folium to highlight which neighborhoods top all my favorite venues.



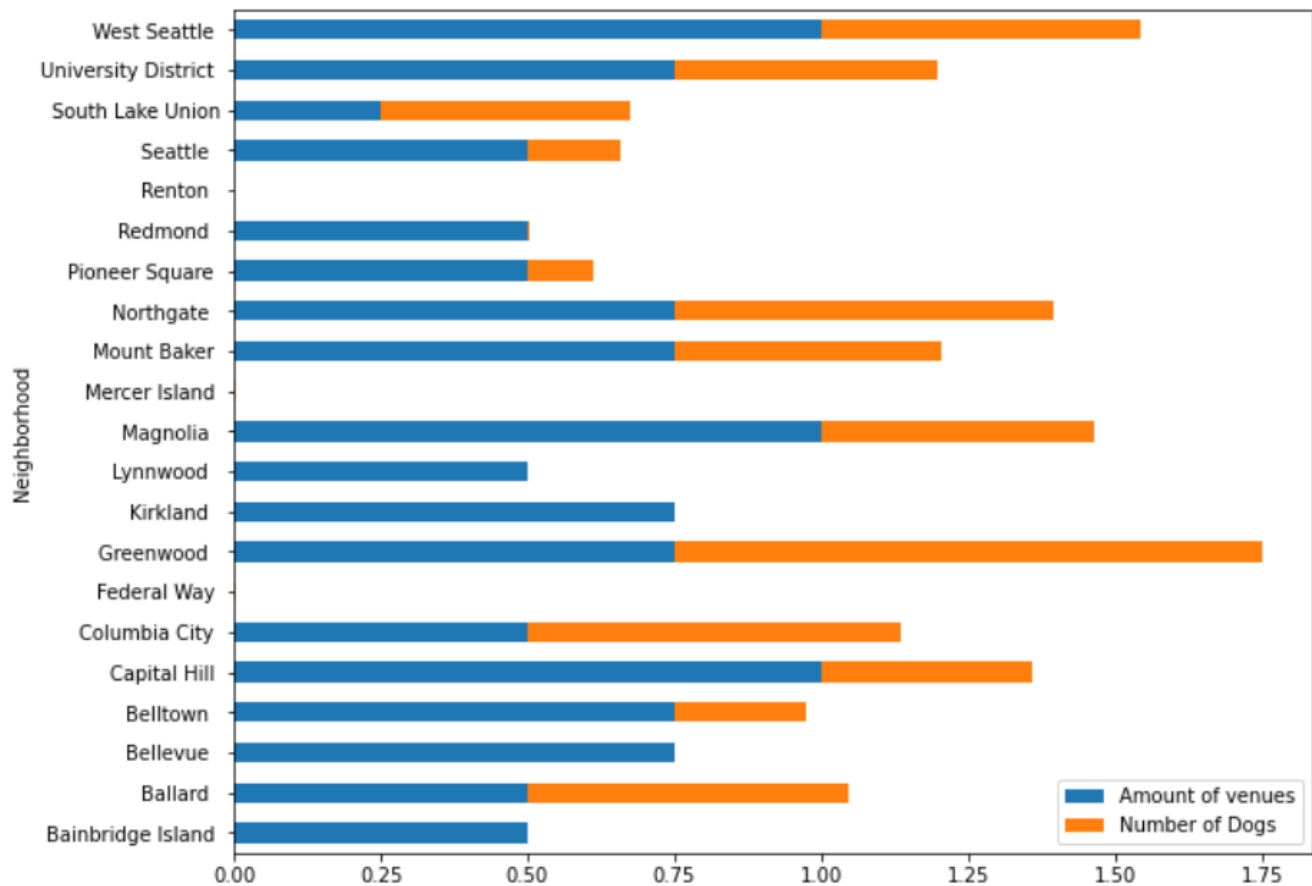
I will then explore these Neighborhoods compared to the amount of dog licenses and see any are a good match with top venues and number of dogs - which could prove a good location for my family to move to.

Grouping the data by neighborhood and dog licenses:

	Neighborhood	Species	dog counts
0	Greenwood	Dog	1510
1	Northgate	Dog	975
2	Columbia City	Dog	961
3	Ballard	Dog	825
4	West Seattle	Dog	822

Greenwood, Northgate, and Columbia city have the highest counts of dog licenses.

None of these top 3 match with the best neighborhoods according to our favorite venues. So, I normalized the data and plot it as a bar graph to see if it gave better results:



By normalizing and stacking we can see that **Greenwood** gives a good amount of our favorite venues and number of dogs. We also might want to consider **West Seattle** and **Magnolia** as Neighbourhoods to further analyses too!

Results and Discussion

From the exploration above, we find that among the Seattle neighborhoods **Magnolia, West Seattle and Capitol Hill** seem to be the best Neighborhoods in terms of parks, views, trails, breweries, sushi and burgers!

Since there are 21 neighborhoods in Seattle, we did not look deep into each one. We then found that **Greenwood, Northgate and Columbia** areas have the highest counts of dog licenses.

As there was no neighbourhood which was the same from both results, both datasets we concatenated and normalized to graphically show the neighbourhoods that resulted in the highest counts of dog licenses with the largest venue counts. *Greenwood* was our winner, with *West Seattle* and *Magnolia* also of interest.

Always note that, although this analysis gives insight into the issue concerned, since our data are collected online from multiple sources, we should always be critical about the confidence of the result and only take them as a reference.

Conclusion

If this project, we are trying to find the neighbourhoods in Seattle which would best suit our family life and which have the potential for a dog walking business. We did find some neighborhoods in Seattle with both our favorite destinations and a good puppy population.

It should be acknowledged that the result of this analysis is according to the categories of venues nearby the center of the neighborhood, and there are other factors I would also consider before making up my mind, things like neighborhood wealth, crime rates, parking availability and so on.

I hope you had a good time through this journey of exploring the neighborhoods of Seattle with me!