**Selecting best neighborhood in Seattle for a new daycare center**

Tze Yeoh

Oct 18, 2019

# Introduction

## Background:

As a parent, I've discovered that the is a huge shortage of daycare centers for working parents in the Seattle, WA market. Seattle has had a huge boom in population growth in the last 5 years, consistently posting top 5 in fastest growing cities in the USA. Consequently, this is creating strain on the local childcare industry. From local news reports, and through personal experience as well, there are huge waitlists in the city for childcare, with many parents driving in opposite direction from their commute, going out of the city for care. There is a huge pent up demand for additional daycare centers in the city. Therefore, it worth the time for any new childcare business to determine the best neighborhood to start a new center to optimize for matching supply with demand.

## Problem statement:

If I were to take the problem into my own hands, I would like to know the optimum neighborhood location to open my own daycare to optimize for the demand in the city. As with any new business, there is significant risk in starting the daycare in the incorrect neighborhood. Key factors in play are whether there is enough population in order to create the demand, and whether there are undesirable businesses nearby that would not be a good fit for the daycare.

# Data Acquisition & Feature Selection

## Data sources

### Foursquare Data

Venue data for each Seattle neighborhood is obtained from the Foursquare website. Data will contain venue name, venue type, longitude and latitude of the business. The dataset is also used to select neighborhoods that have lower concentration of venues that are undesirable like dive bars, night clubs and marijuana dispensaries, and have a higher concentration of venues that are residential, like apartment buildings. I utilized the Foursquare API to select for criteria venues and utilized it in my dataset.

### City of Seattle Data

Population data for each Seattle neighborhood is obtained from Seattle.gov website. Data contains detailed demographic data like population counts, households, average persons per household, etc. Dataset will provide an idea of demand for daycares in the neighborhood, where neighborhoods with higher population counts and densities are more ideal. I downloaded the latest data from the website and utilized it in my dataset.

## Data Cleaning

The City of Seattle dataset had approximate Latitude and Longitude coordinates for the respective city neighborhoods. As I was exploring the mapping, I noticed that the coordinates were sometimes to the edge of the neighborhoods versus in the middle. This could cause the data to be biased away from the center of the neighborhood, potentially having data spilled over to adjacent neighborhoods.

To properly center the neighborhood coordinates. I utilized Google Maps, searched for the specific neighborhoods, located and transposed the coordinates from the center of the neighborhood into the dataset I used for the project.

## Feature Selection

### Foursquare Data

For the purposes of the selecting daycare locations, it was important to have a few key venues as selection criteria within the API called for. Otherwise I would get venues that are not material to the analysis (example – bookstores, restaurants, parks, etc). Specifically, I selected for the following venues in my search:

|  |  |
| --- | --- |
| **Venue Categories** | **Venue Code** |
| Childcare service | #5744ccdfe4b0c0459246b4c7 |
| Bars | #4bf58dd8d48988d116941735 |
| Residences | #4e67e38e036454776db1fb3a |
| Nightlife Spot | #4d4b7105d754a06376d81259 |
| Marijuana Dispensary | #52c71aaf3cf9994f4e043d17 |

Table 1. Daycare venue search criteria

Full category list is found here: <https://developer.foursquare.com/docs/resources/categories>

* Childcare services were selected so that I could determine whether this is a concentration of other daycare services already in the neighborhood.
* Bars, Nightlife Spots, Dispensaries were selected so that I could determine the concentration of undesirable venues in the neighborhood.
* Residences were selected to assess the demand portion of the equation in the neighborhood. Higher concentration of residences should have higher demand for daycares.

### City of Seattle Data

The demographic information from the City of Seattle was important for further narrowing down of neighborhoods to ones with the highest concentration of population and households. I identified the features of “Population in Households” and “Family Households” as the important for selection of neighborhood.

There 2 features were selected as it reflects number of people that live in family unit, and the number of families counted, with the idea that the higher the counts in this feature, the higher change of the family having children.

## Other Considerations

Within the Foursquare API, is the variable “Radius”, which designates in meters the search range when initiating the Explore functionality. During the data exploration of the resulting JSON output, I discovered that I needed to experiment with different values of the Radius when considering the venues’ proximity to the coordinates of the neighborhoods. In order words, if a Radius value was too high or low, the venue would be distorted.

In the case of high radius value (10,000 meters or more), the Foursquare API would return venues from other adjacent neighborhoods. In most cases, this caused the dataset to be harder to filter down, as it created too many false positives. Example – More neighborhoods will show up as having residences as their highest concentration of businesses, when the radius of Foursquare search was so large that it would return venues from other neighborhoods.

Conversely, for low radius value (100 meters or less), the API would return very limited venues in the response. Limited venues meant that many neighborhoods would be eliminated from consideration unnecessarily.

Therefore, after experimenting with the radius variable, across multiple iterations of the analysis. I selected 600 meters as the best Radius that generated the most consistent dataset with the highest level of reasonable accuracy.

# Methodology

## Overall Method

Overall, the method was to find and select all neighborhoods in the Seattle area, filter against venues that met selection criteria in Section 2.3, then finally filter against demographic data with higher populations to select best neighborhood. Step by step:

1. Identify all distinct neighborhoods in the Seattle area
2. Query Foursquare to find specific venues near neighborhoods, that match daycare criteria in Table 1 - Section 2.3.
3. Rank neighborhoods against venue criteria, showing most common to least common venue in Seattle.
4. Select for neighborhoods that have highest concentration of residences (1st most common)
5. Deselect for neighborhoods that have bars/nightclubs/dispensaries (2nd most common)
6. Rank remaining neighborhoods against demographic data, sorted with neighborhoods with highest number of population and families, with highest ranked neighborhood is selected as the neighborhood for daycare.

# Analysis

## Identify distinct neighborhoods in Seattle

Utilizing Seattle.gov data, the neighborhoods were plotted out in Folium to get a sense of where each neighborhood was. Figure 1 depicts all the neighborhoods in Seattle, with the center of the neighborhoods as blue circles.

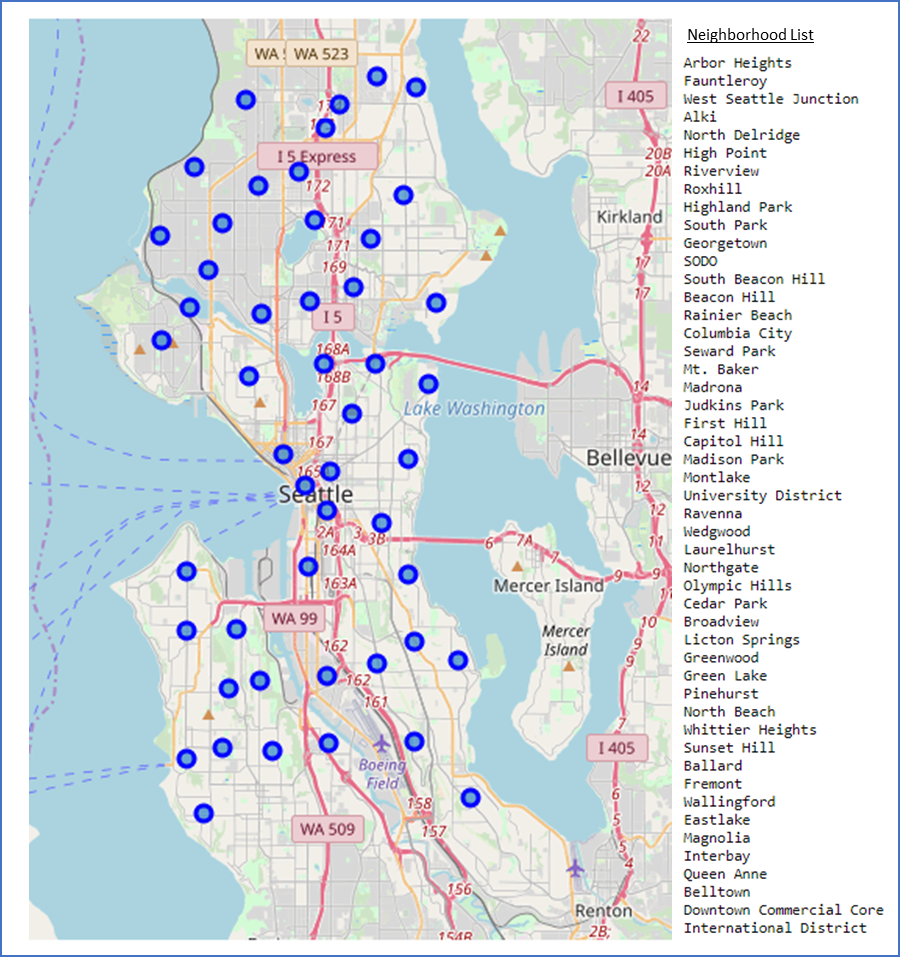


Figure 1. Neighborhoods of Seattle plotted

## Foursquare Query

Using the Foursquare explore API, the coordinates of every neighborhood are ran through the an algorithm, searching for venues meeting selection criteria defined in Table 1- Section 2.3 (Childcare services, Bars, Residences, etc.). Furthermore, the search radius of was limited to 600 meters, as explained in Section 2.4.

The output of the query is a listing of all venues meeting the search criteria for selecting a daycare, with the following sample header in Table 2.

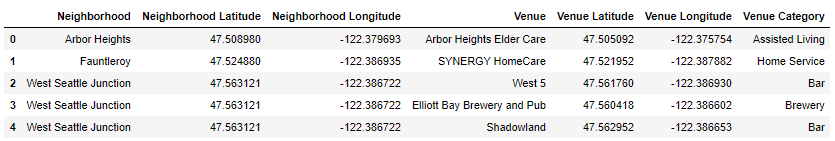


Table 2. Header of all venues near neighborhoods that meet daycare search criteria

## Ranking venues into each neighborhood

Using the one-hot encoding method to the dataset, I grouped and arranged the venues from most common venue in the neighborhood, to the least common venue in the neighborhood. The objective of this ranking is seeing the respective neighborhoods differs in composition of their daycare criteria.

The output of this ranking is displayed in the sample header in Table 3.



Table 3. Header of neighborhoods and their respective most common to least common venues.

## Filtering for highest residences venues

Filtering for neighborhoods with “residences” as the 1st most common venues, I came up with a shortlist of 12 neighborhoods. The reasoning is, if we prioritize our selection criteria for neighborhoods with the most amount of residences (homes, apartments, & condos), we will have a higher chance of families that need daycare.

The output of this filtering is as displayed in Table 4.

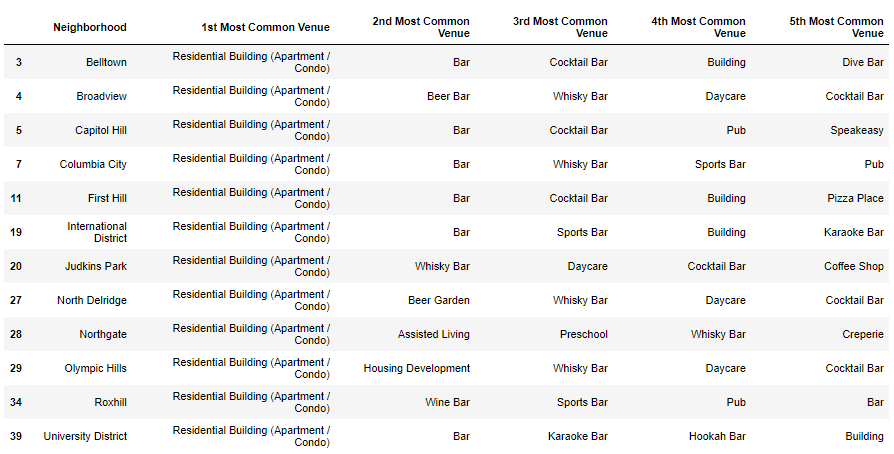


Table 4. Shortlist of neighborhoods with the 1st common venue being residences

## Deselecting for remaining neighborhoods

To further narrow down the listing of potential neighborhoods, I filtered away neighborhoods with alcoholic, dance club and/or marijuana related venues in their 2nd most common venue field.

The output of this further filtering is as displayed in Table 5.

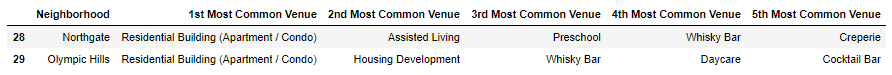


Table 5. Final shortlist of neighborhoods with non-children friendly venues filtered away.

## Conclusion: Ranking and selection of target neighborhood

Finally, I dived back into the Seattle.gov data, and extracted key demographic information from the neighborhoods shortlist in Table 5 – Section 4.5. The neighborhoods were sorted by highest number of population and families, with highest ranked neighborhood is selected as the neighborhood for daycare.

As a result, the “Olympic Hills” neighborhood was selected as the target neighborhood, as described by the data in Table 6.

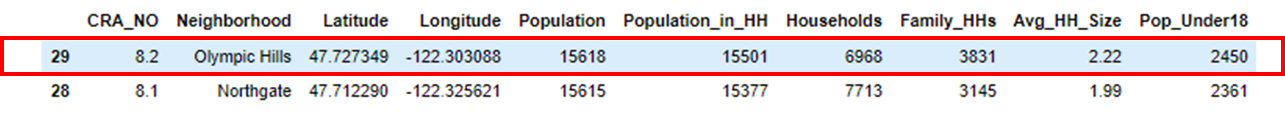


Table 6. Selected neighborhood for daycare, after ranking by demographic data

# Conclusion & Final Notes

In this study, I combined datasets from the City of Seattle, and the Foursquare API to analyze and determine the neighborhood best suited for a new daycare. Starting with coordinates of each neighborhood in Seattle, I mapped the various venues that were positive towards daycare clients (residences) against venues that we negative to clients (alcoholic, dance club and/or marijuana related). Once I had a shortlist of neighborhoods, I selected the final target neighborhood based on demographic data from the City of Seattle, which was “Olympic Hills”.

As an additional exercise in curiosity, I was interested to see which specific area inside the target neighborhood would be good candidates for the specific placement of the daycare. I isolated the neighborhood in folium and mapped out the venues that were positive (green) vs the negative (red) to daycares in the neighborhood. Then I drew a rough specific search area that tried to maximize the positives and avoid the negatives (green line) in Figure 2. Any further analysis would probably require a physical visit to the area and more detailed neighborhood studies to select the final storefront location. However, this project has allowed us to use publicly available data to narrow down to the candidate neighborhood that gives the best chance of success in starting a daycare.

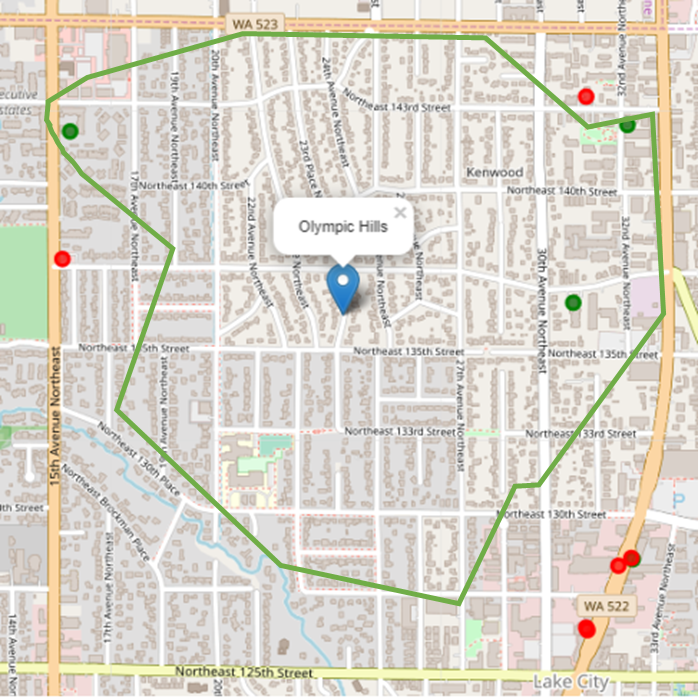


Figure 2. Selected neighborhood for daycare, with possible search zone for specific location