Opening a Food Cart in Portland, OR after the Pandemic

12.14.2020

Grant Langlois
Coursera Capstone Project

Introduction

In the wake of the 2020 COVID-19 pandemic, many brick and mortar restaurants have been forced to close down due to revenue loss from the restrictions put in place to curb the spread of COVID-19. As such, independent restaurant ownership has a higher barrier to entry than ever before. A path forward for the independent restaurateur lies in the food cart scene, which has a large footprint in Portland, OR. Outside of reduced overall foot traffic, managing a food cart does not share many of the financial burdens of opening a restaurant that are so critically hampered by the uncertainties in revenue from reductions in customers and employees.

The stakes are always elevated when it comes to independent restaurant ownership, and the large food truck footprint in Portland, OR, means competition can be fierce. If you wish to open your first food-service establishment, or you wish to pivot from your current brick and mortar location, the question becomes: where should you open a food cart to optimize customer engagement while minimizing competition? The goal of this report is to provide a brief overlook of what types of neighborhoods exist in Portland, OR and which neighborhoods provide your best chance for success.

Data

Starting with a list of neighborhoods in Portland, OR (derived from scraping these data with the BeautifulSoup4 python package) the intent was to gather a list of nearby venue data within each neighborhood via calls to the RESTful Foursquare API in order to segment the city into clusters of similar neighborhoods using a K-means algorithm. The choice of best number of cluster centers was determined iteratively by comparing results from both an elbow plot and a silhouette score.

Based on nearby attractions (specifically food cart options), the goal is to recommend the top five neighborhoods to open a food cart based on descriptive statistics and the resultant clustering data. The recommendations are also based, in part, on a few naive assumptions to be defined later in the Discussion section of the report.

Methodology

Raw Data Aggregation from Wikipedia using BeautifulSoup4

We first use BeautifulSoup4 to parse the raw Wikipedia HTML into a list of dictionaries containing each neighborhood's name, latitude and longitude. This specific data structure makes it easy to convert the data into a pandas DataFrame. Helpfully, each neighborhood has its own subpage on Wikipedia as well, listing the location in latitude and longitude, as well as its population density (among other metrics). Thus, for each neighborhood, we also scrape its Wikipedia subpage for these data instead of utilizing a geocoder or another dataset.

The Wikipedia scraping fails to find population densities for 10/99 neighborhoods due to this value not existing on the subpages for these neighborhoods. For simplicity, we replace all missing values with the mean population density of the remaining 89 neighborhoods:

	Neighborhood	**Population Density (per sq. mi.)**	**Latitude**	**Longitude**
0	Alameda	8200	45.5482	-122.6307
1	Alberta Arts District	5818 (*mean*)	45.55905	-122.64286
2	Arbor Lodge	7000	45.57354	-122.69247
3	Ardenwald-Johnson Creek	5818 (*mean*)	45.4554	-122.6298
4	Argay	3100	45.55475	-122.52114
...	...	...	...	...

Nearby Venue Data via Foursquare API

We make a call to the Foursquare API's "Venues" backend to gather categorical data of the top 100 nearby venues within 500m of the neighborhood center. "Neighborhood center" is defined as the neighborhood's latitude and longitude in the above dataset.

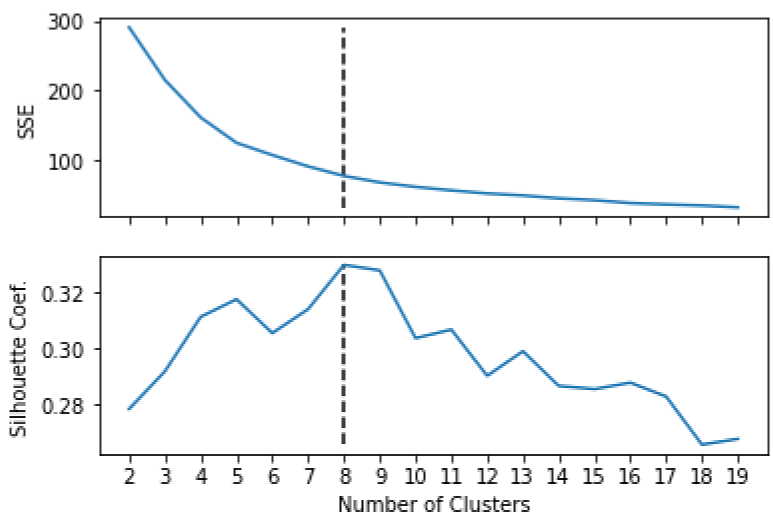
	Neighborhood	Venue	Venue Latitude	Venue Longitude	Category
0	Alameda	Marianne's	45.547526	-122.630483	Sports Bar
1	Alameda	TriMet Bus Line 24	45.548481	-122.630690	Bus Line
2	Alameda	Joes Back Yard BBQ!	45.550229	-122.629402	BBQ Joint
3	Alameda	Tacovore	45.548444	-122.624638	Taco Place
4	Alameda	Beaux Berry	45.548367	-122.624389	Ice Cream Shop
...	...	...	...	...	...

The API call fails to return venues for three neighborhoods: Bridlemile, Forest Park, and St. Johns. Thus, also in the interest of simplicity, we will exclude these neighborhoods from the analysis. One-hot encoding of the venue categories is done in anticipation of using part of these data to cluster the neighborhoods. We encode each category as its fraction relative to the total number of venues in that neighborhood, and then join these data to the original neighborhood dataset.

Feature Set and Unsupervised Clustering

For the feature set, we select population density, latitude, longitude, and fraction of venues that are food trucks. Latitude and longitude are included here because Portland, OR is segmented into six subsections (reclassification of the sixth sextant was finalized in May 2020).

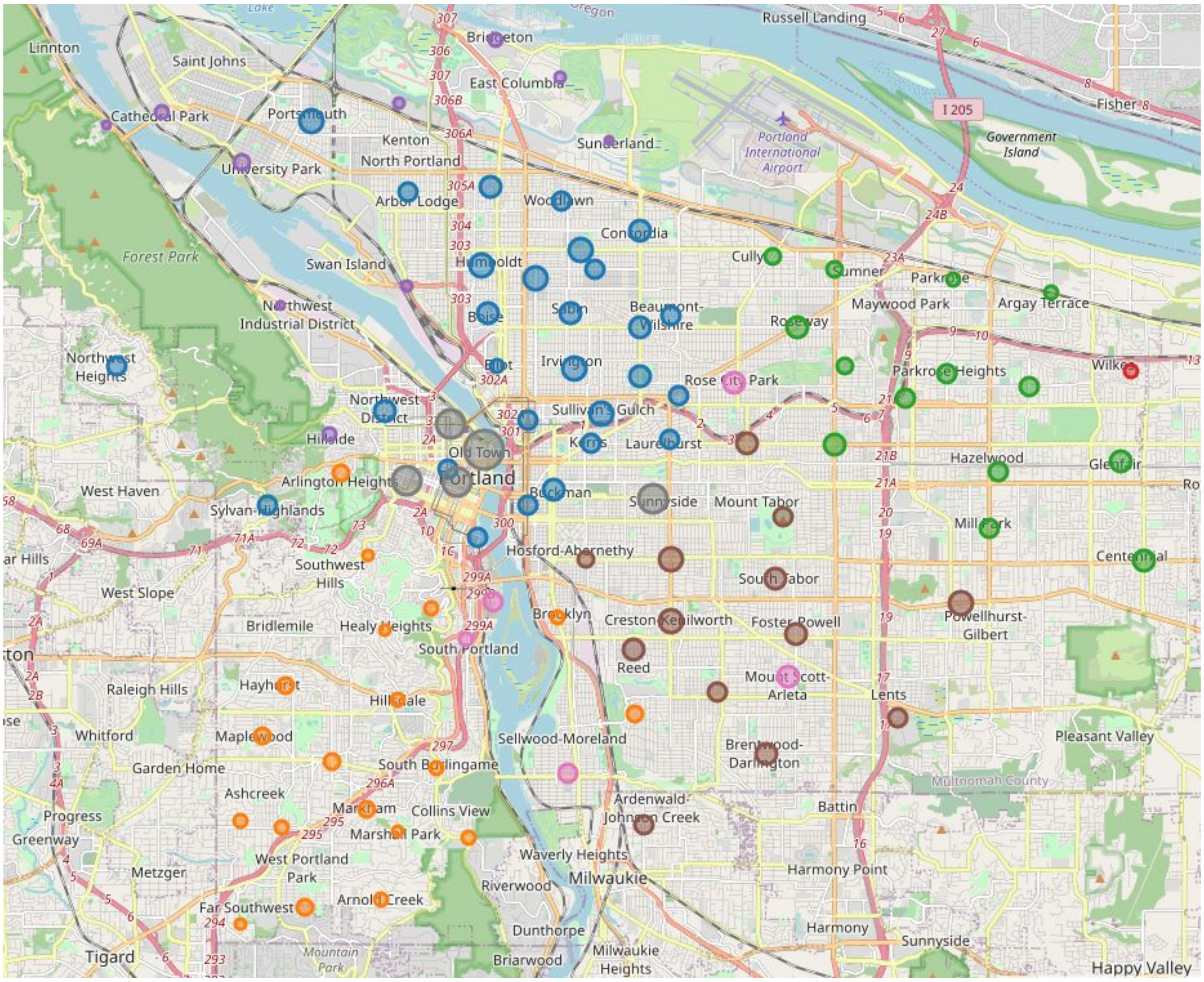
The machine learning algorithm chosen for this analysis was K-Means clustering, an unsupervised technique. Considering both the sum of squared errors (SSE) and the silhouette score, the optimum number of clusters to be used was determined to be eight:



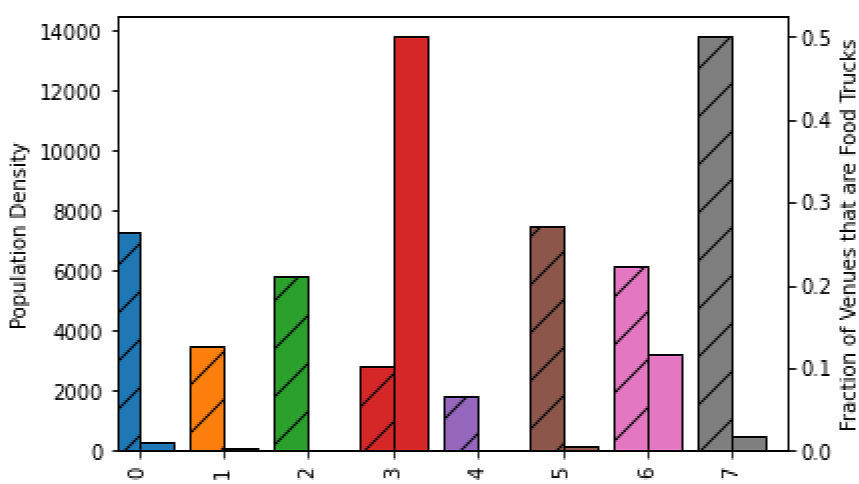
Discussion

Clustering Results

Unsurprisingly, the clustering does somewhat reflect the defined sectioning of the city, but it does separate out areas like downtown (grey), industrial neighborhoods located near the Willamette/Columbia river junction (purple), and retail areas near the Willamette River (pink). In the plot below, the marker size is defined by the relative population density.



With this summarized dataset, we can look at the mean population density and fraction of venues that are food trucks within each cluster to select the most promising clusters:



Recommendations

We can quickly rule out cluster 3 since it has one of the lowest population densities and the highest relative number of food trucks. Cluster 6 also looks less appealing as it has the second-highest fraction of food trucks despite having a reasonable population density. We then look to the three clusters with the highest population densities: 7, 5, and 0, respectively below.

	neighborhood	pop_density	Food Truck
65	Old Town Chinatown	19000.0	0.027778
24	Downtown Portland	13000.0	0.030000
69	Pearl District	13000.0	0.000000
32	Goose Hollow	12000.0	0.000000
90	Sunnyside	12000.0	0.035714

	neighborhood	pop_density	Food Truck
44	Irvington	10000.0	0.000000
93	Vernon	10000.0	0.022727
87	Sullivan's Gulch	9500.0	0.000000
47	King	9300.0	0.030303
43	Humboldt	9200.0	0.024390

	neighborhood	pop_density	Food Truck
21	Creston-Kenilworth	10000.0	0.041667
72	Powellhurst-Gilbert	8900.0	0.000000
74	Richmond	8900.0	0.000000
30	Foster-Powell	8300.0	0.000000
61	North Tabor	8100.0	0.000000

Despite hosting the highest population densities, cluster 7 is dominated by neighborhoods near Downtown Portland, arguably the area hit hardest by the pandemic due to its mixture of high-density office districts and tourism-related ventures.

Discounting cluster 7, the five neighborhoods with the highest population densities and lowest number of food trucks are:

1. Irvington
2. Sullivan's Gulch
3. Powellhurst-Gilbert
4. Richmond
5. Foster-Powell

The above list serves as the most attractive neighborhoods for opening a food truck in terms of areas with high population density and no competition.

Additional Considerations

There are clearly other considerations when opening a business, mainly associated with available budget and local ordinances. However, the clustering gives a reasonable first-order approximation for reducing the nearly 100 possible neighborhoods into a short list of reasonable selections.

One way to potentially improve your odds of success using this analysis would be to also consider the density of restaurants in general instead of just food trucks, since many have pivoted to offering delivery options in the wake of the pandemic. Another potential improvement could be to consider the densities of retail venues to estimate nearby "foot traffic". These data are available in the original data set derived from the Foursquare API, but would require cleaning the dataset to create supersection of venue categories.

Conclusion

K-means clustering offers a reasonable way to quickly drill down through the nearly 100 possible neighborhoods in Portland, OR to select a few neighborhoods that optimize the chance of your food truck's success. Utilizing the Foursquare API to categorize nearby popular venues in each neighborhood, we can conclude based on their high population densities and low level of competition, that the five most reasonable neighborhoods for further consideration are:

1. Irvington
2. Sullivan's Gulch
3. Powellhurst-Gilbert
4. Richmond
5. Foster-Powell