Recommending Portland Neighborhoods for Renters: A Geospatial Analysis of Rent and Venues

Jonathan Jettenberger-Burleson

April 27, 2020

1. Introduction

1.1 Background

The cost of a home is just one aspect that comes into play when calculating expenses. For many people the idea of buying a home can be either overwhelming or outside of their financial means. Renting is a good alternative that many decide is the optimal option for them. There are resources out there, such as Zillow that assist renters and home buyers alike in finding the right home. The monthly cost to rent can be affected by the unit size, location, number of bedrooms, and many other features. In addition, a unit's proximity to venues such as restaurants, parks, night clubs, and museums may play a role in pricing and, more importantly, a renter's interest in that location. Information relating monthly rent rates by neighborhood to nearby venues can be an important tool for renters.

1.2 Problem

A new home buyer that has decided that renting may be an option they wish to explore may weigh several criteria when looking for a place that is right for them. Other than the monthly rent, a location's proximity to certain venues can play a role in a renter's decision. This report will aim to recommend a neighborhood for a prospective renter based on venue type, venue proximity, and of course, monthly rent. In this report, we will focus on the Portland, Oregon area to minimize the scope.

1.3 Interest

This report is primarily of interest to both renters familiar with the Portland, Oregon area and those unfamiliar and new to the area. Home buyers and landlords looking to rent out their future property would also find this information of use in gauging the potential value of their investment based on proximity to venues. Those in real estate or city planning may also find this information of use when predicting rent, property value, or interest in an area. Business owners, especially small, local businesses, may find this report noteworthy in predicting their potential customer base.

2. Data

2.1 Data sources

In this report, I will be leveraging geospatial venue data gathered from Foursquare and rental data gathered from Zillow. The Foursquare API will mainly be utilized for venue types and their locations in the Portland area, where requests send back trending locations within the search radius. As the Zillow API does not allow for general searches for an area, information on rent from multiple properties and their locations will come from here. This data is periodically updated and is relevant at the time of this report. The Zillow service uses a proprietary algorithm to determine rent as an attribute they call Rent Zestimate. This rent value factors in a property's characteristics, unique features, on-market data, and off-market data. We will be selecting the

Zillow Rent Index (ZRI) data with neighborhood geography. Boundary data, found here, will also be used for modeling of neighborhoods. To assist in the cleaning of data, information about neighborhood names was gathered from PDX Listed.

2.2 Data cleaning and feature selection

Data requested from Foursquare provides much information about venues in an area. Information can be gathered based on location, other users, or even just by venue category. Foursquare also provides data on tips, hours, menus, photos, and events. Much of this information would be useful depending on the needs of the user, but are out of scope for this report.

We will be focusing on the latitude and longitude location of each venue, and the categories the venue falls under. The names of the venues will also be included for labeling. To restrict the data size for this report, we will search for venues with Portland as a city value. Due to the 100-entry limit on returned data per request for the Foursquare API, five separate requests were made in an attempt to increase the dataset size. Each request was centered on one of the five sections of Portland (Table 1) and was searched in a radius of 8046 meters, or approximately 5 miles. Overlapping data points were also removed. This method yielded a dataset of 369 venues in the Portland area (Figure 1).
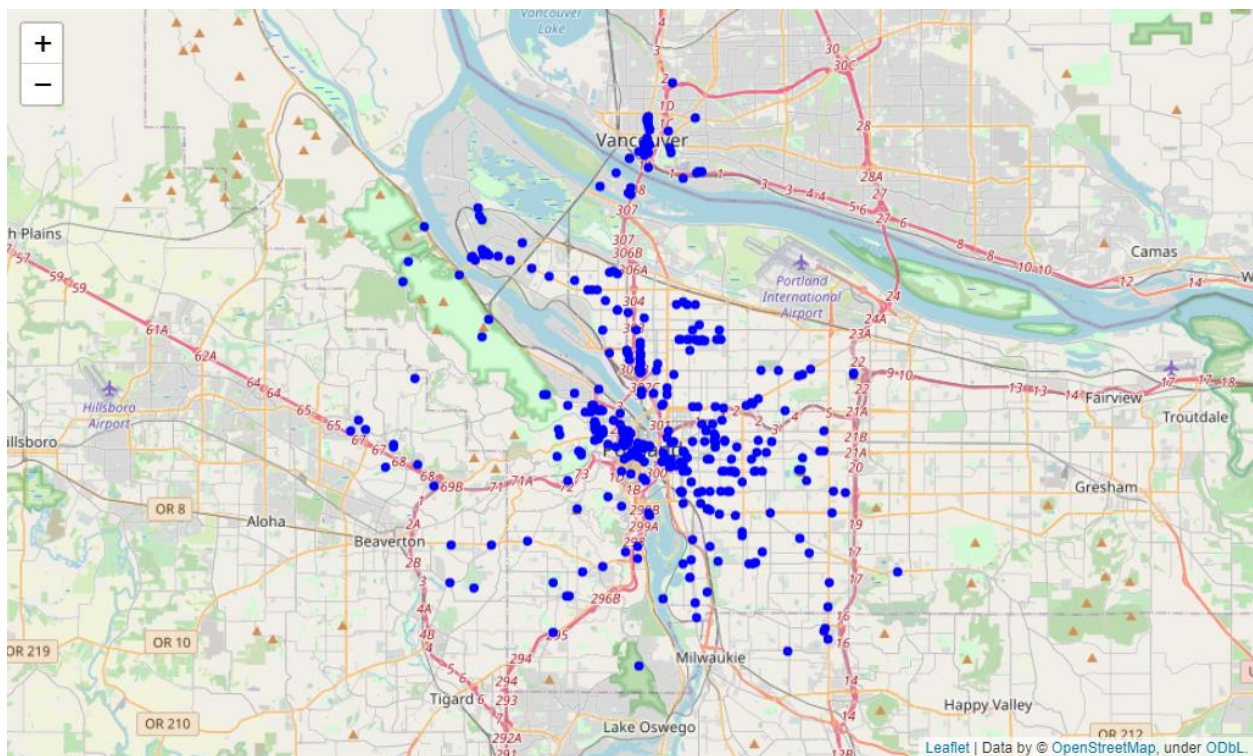

Figure 1. Map of Foursquare venue data points.

Section	Location (lat, lng)
Northwest	45.5586, -122.7609
Northeast	45.5676, -122.6179
North	45.6104, -122.7034
Southwest	45.4849, -122.7116
Southeast	45.4914, -122.5930

Zillow has detailed information on property characteristics and market data. The dataset collected has information from multiple locations. To refine the data, I began with selecting the subset focused on rental properties with the values City and State equaling Portland and OR, respectively. Due to the limited scope of this report, only the neighborhood names and the Rent Zestimate data will be examined. The column titles for this information was standardized to Neighborhood and Rent as well. In Portland, OR, many neighborhoods are represented by a home owners association or a league. This information was not represented between data sources. To match up the rent subset with the geospatial information gathered from the boundary data, 17 data entries for neighborhood names had to be cleaned because of this. There was one additional instance of a neighborhood being represented by its former name.
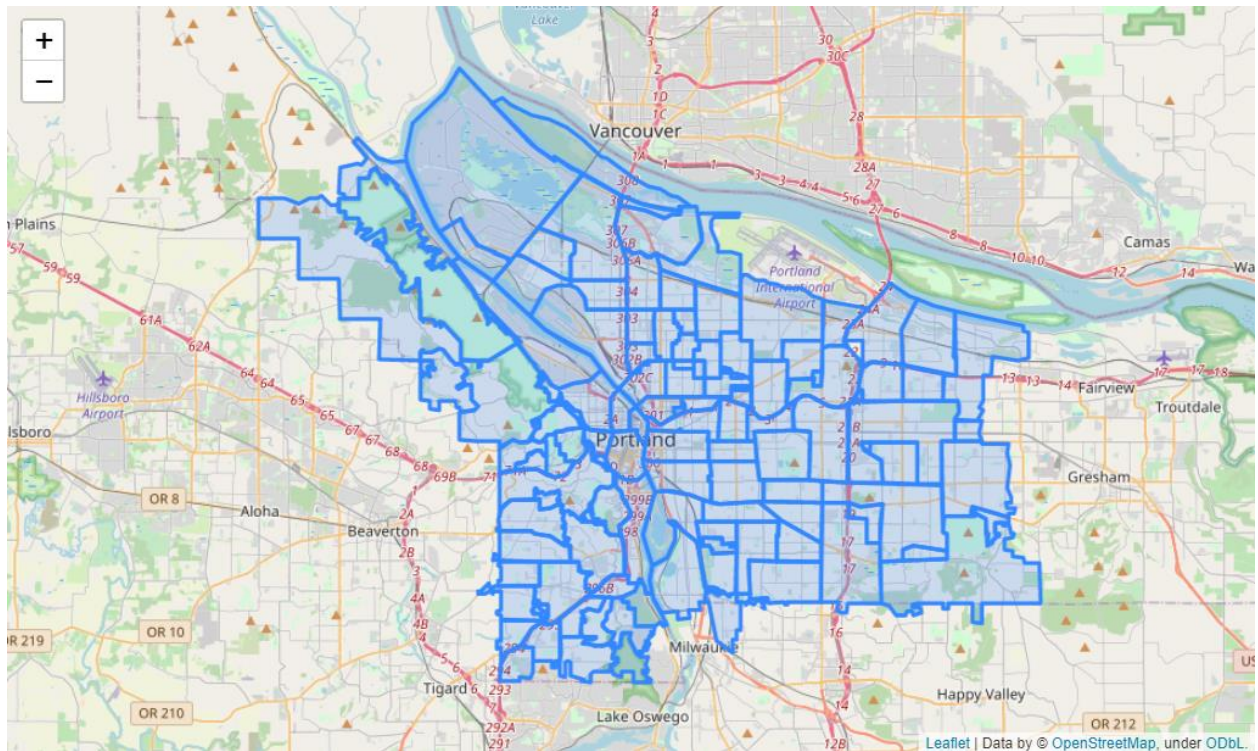

Figure 2. Boundary data for the neighborhoods of Portland, OR.

The neighborhood boundary data includes 25 entries that represent overlapping of the neighborhoods. In these instances, available rental data was averaged for the overlapping areas. From the rental data from Zillow, only 93 of the 130 areas on Figure 3 were filled. This still left some areas of the map with no rental data. A few of these areas represent unclaimed neighborhoods that are not claimed by any association. The other areas simply have no rental data on Zillow. This could be due to a lack of rental information posted on Zillow, or even no rental properties to speak of.
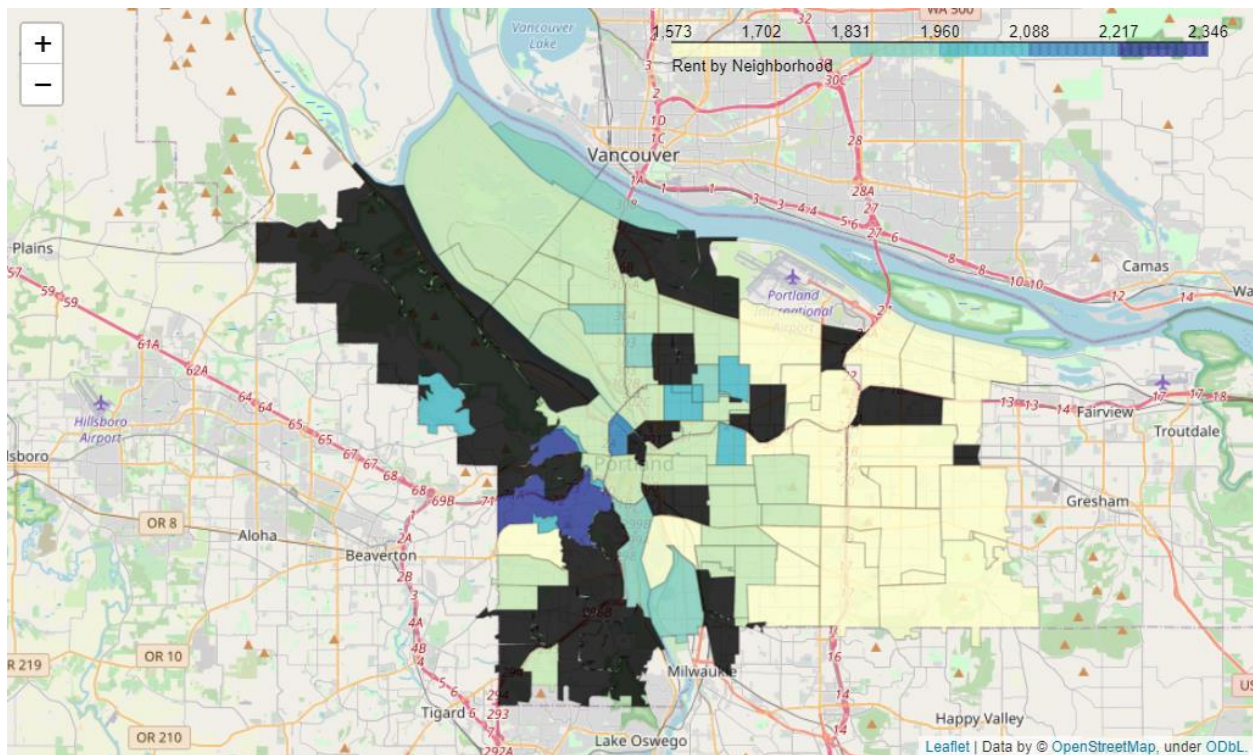

Figure 3. Heatmap representing average rent for neighborhoods in Portland, OR.

3. Methodology

3.1 Exploratory data analysis and merging data

In analyzing the raw data, it is clear we have a reasonable spread to our data points. As seen in Figure 4, there are some clear outliers. These venue points with locations outside of the neighborhood boundaries is due to the way in which venues are requested. By searching in a radius from a given set of coordinates, we have introduced venues in other neighborhoods.
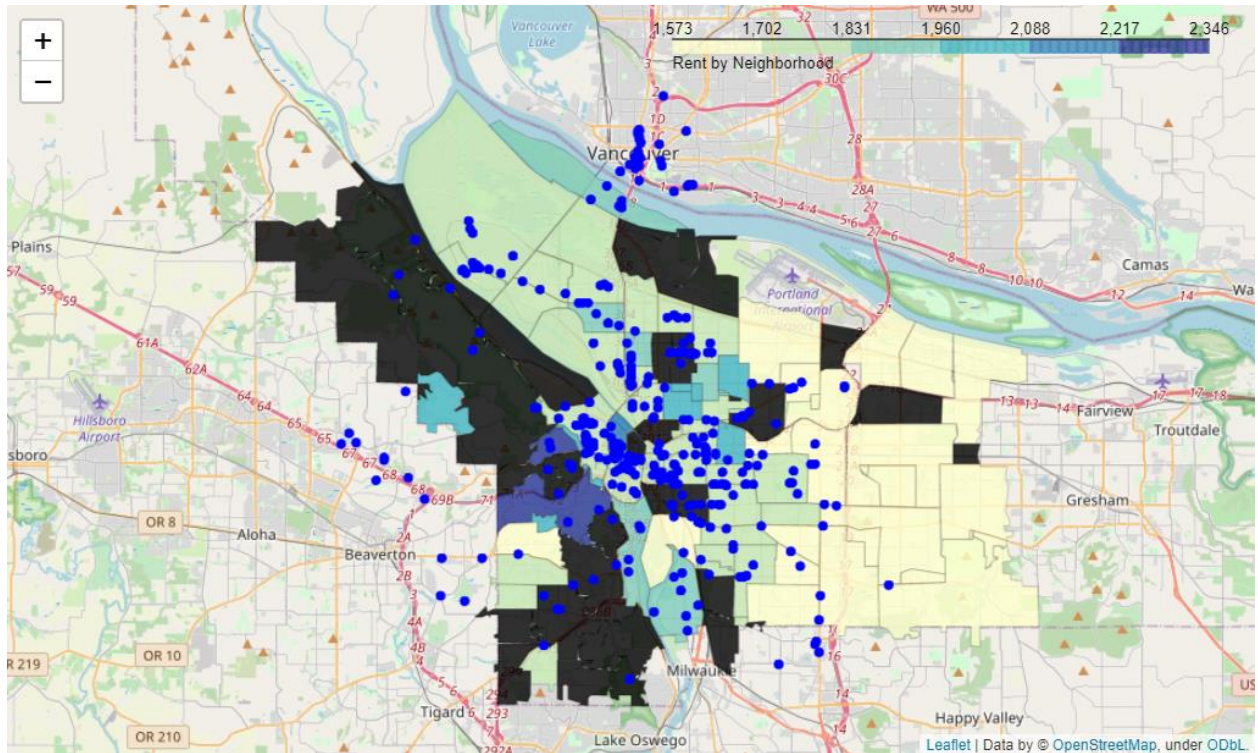
Figure 4. Venue points overlaid on average rent heatmap.

3.2 Clustering

In order to analyze the relationship between the locations of venues and their proximity to neighborhoods, the venue points were put through a k-means clustering algorithm. The venue data was clustered into 5 clusters, matching the number of sections of the city in which venue data points was gathered. With the clustering labels generated, the data points in Figure 4 were given color and transformed into that which is displayed in Figure 5. A clear divide can be seen partitioning the venue points.
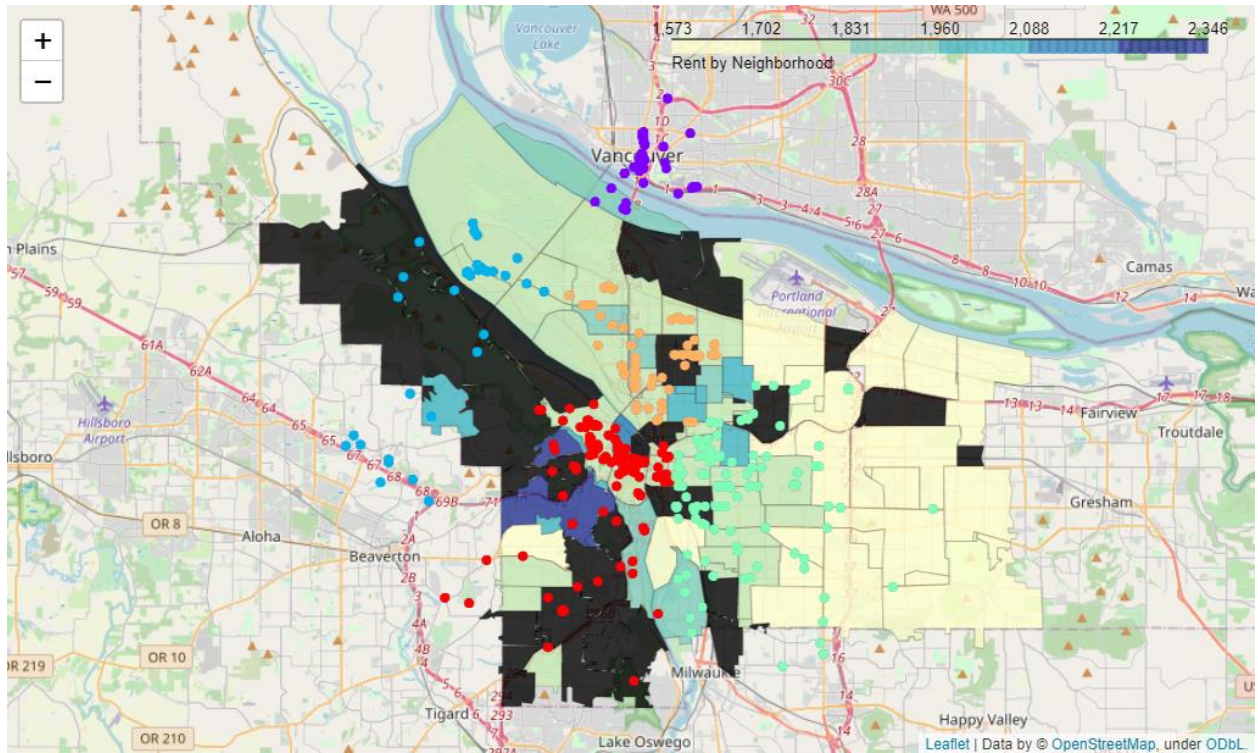
Figure 5. Venue data points clustered over the average rent heatmap.

4. Results

4.1 Cluster 1

The first cluster is representative of the Southwest section of Portland. There are many venues with a reasonable spread of category type, with multiple options for each type. Most of the venues are clustered in the downtown area, but many of them are spaced further out. This area of the city does contain the highest priced rental properties though, with some neighborhoods with rent topping the chart at an average of $2346/month. While there appears to be enough rental data for this area, there are several patches with no rental data, especially in the southern half of this cluster.

Table 2. Results collected on Cluster 1.

Number of venues	109
Number of unique categories	58
Rent average	High

4.2 Cluster 2

The second cluster represents the most northern area of Portland. This area consists of several larger neighborhoods and has relatively average rent. This cluster has not that many venues, but very little overlap in their types. Most of the venues in this cluster actually reside outside of the Portland area, but are near each other.

Table 3. Results collected on Cluster 2.

Number of venues	45
Number of unique categories	43
Rent average	Average

4.3 Cluster 3

The third cluster is representative of the Northwest section of Portland. This cluster contains the least amount of venues and category types. There is also a good deal of distance between most of the venues. Most of the area resides in neighborhoods that have no rental data. Of what little data there is, the average rent appears to be about average.

Table 4. Results collected on Cluster 3.

Number of venues	37
Number of unique categories	25
Rent average	Average (not much data)

4.4 Cluster 4

The fourth cluster is representative of the Southeast section of Portland. This cluster has the lowest average rent, with rent as low as $1573/month. There are many venues with a similar uniqueness in their categories as cluster 1. Venues are spread out a good deal, but there is some clustering within proximity to the downtown area. As you go east in this cluster, rent averages go down.

Table 5. Results collected on Cluster 4.

Number of venues	110
Number of unique categories	64
Rent average	Low

4.5 Cluster 5

The fifth cluster is representative of the Northeast and some of the Northern section of Portland. This area has relatively average rent. Venues are spread out with two areas of clustering along major streets, North Mississippi Avenue and Northeast Alberta Street. There is a fair amount of venues in this area, with a decent amount of spread in their categories.

Table 6. Results collected on Cluster 5.

Number of venues	68
Number of unique categories	42
Rent average	Average

4. Discussion

4.1 Observations

With the clustering of the rent data from Zillow, it is clear which neighborhoods of the city of Portland, OR are more expensive. It follows that these areas have a larger amount of trending venues. What is interesting to note, the areas with low rent share this trend. Cluster 1 with its high rental areas contains downtown Portland. Right across the Willamette River is Cluster 4 with its low rent. The other cluster, representing the areas with average relative rent, span the northern half of Portland. The lack of data in these northern regions could have played a role in their standing in this comparison. Cluster 3 may be an exception though, as it contains Forest Park. This park covers much of the area, actually dividing the region. This could be the reason for the lack of data, but also plays a role in the distance between venues. On the up side, this area contains the majority of the park and trail venues in the entire city.

4.2 Recommendations

For someone looking for a property to rent, it is important to take their budget in to account. If a renter is looking to budget, the Southeast section of Portland is the obvious choice. This area boasts the lowest rent between $1500 and $2000/month, while still being close to downtown. Venues are spread out, but neighborhoods in the northwestern part of this area show more clustering with closer venues. If budget is not a concern, the Southwestern section of Portland is the right choice. With the downtown area, these neighborhoods contain numerous venues, especially if including its proximity to other sections of the city. Renters should be prepared to see over $2000/month in some areas, if this is the area they decide on.

5. Conclusions

We set out to recommend neighborhoods for potential renters in the Portland, OR area. Data from Foursquare and Zillow was leveraged to find a relationship between average rent by neighborhood and location of trending venues. Foursquare provides much information on venues, but needs some cleaning and data analysis to yield meaningful results from a data science point of view. Data gathered from Zillow needed cleaning in order to match up with boundary information. Once paired up, the heatmaps generated show a clear trend in the location of venues and the number of venue categories represented. The majority of venues are situated in areas of high and low rental data. This may be indicative of their quality, but that is something that could be explored in the future. It is clear that someone looking to rent should be recommended either a high rent or low rent neighborhood in the southern half of Portland, depending on their budget. Depending on how the data is interpreted, other interested parties may see an opportunity in the northern half of the city. Business owners, for example, may see less competing local venues. On the other hand, these areas do not stand out as much as those on the high and low end of the rent spectrum. Moving forward, more data must be collected to fill in the missing areas of the heatmap.