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

---

Jonathan Jettenberger-Burleson

February 10, 2021

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 neighborhoods in areas of the city that may interest 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 local customer base.

2. Data

2.1 Data sources

In this report, I will be mainly leveraging geospatial venue data gathered from Foursquare and rental data gathered from Zillow. The Foursquare API will 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 be acquired as a raw data file. 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 will also be used for modeling of neighborhoods. To assist in the cleaning of data, information about neighborhood names was gathered from PDX Listed. Missing rental data from Zillow was filled in with additional data from RENTCafé.

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, multiple requests were made in an attempt to increase the dataset size. A total of 130 requests were each centered on a neighborhood in Portland (Figure 1) and was searched in a radius of 1609 meters, or approximately 1 mile. Overlapping data points were removed. This method yielded a dataset of 3131 venues in the Portland area (Figure 2).

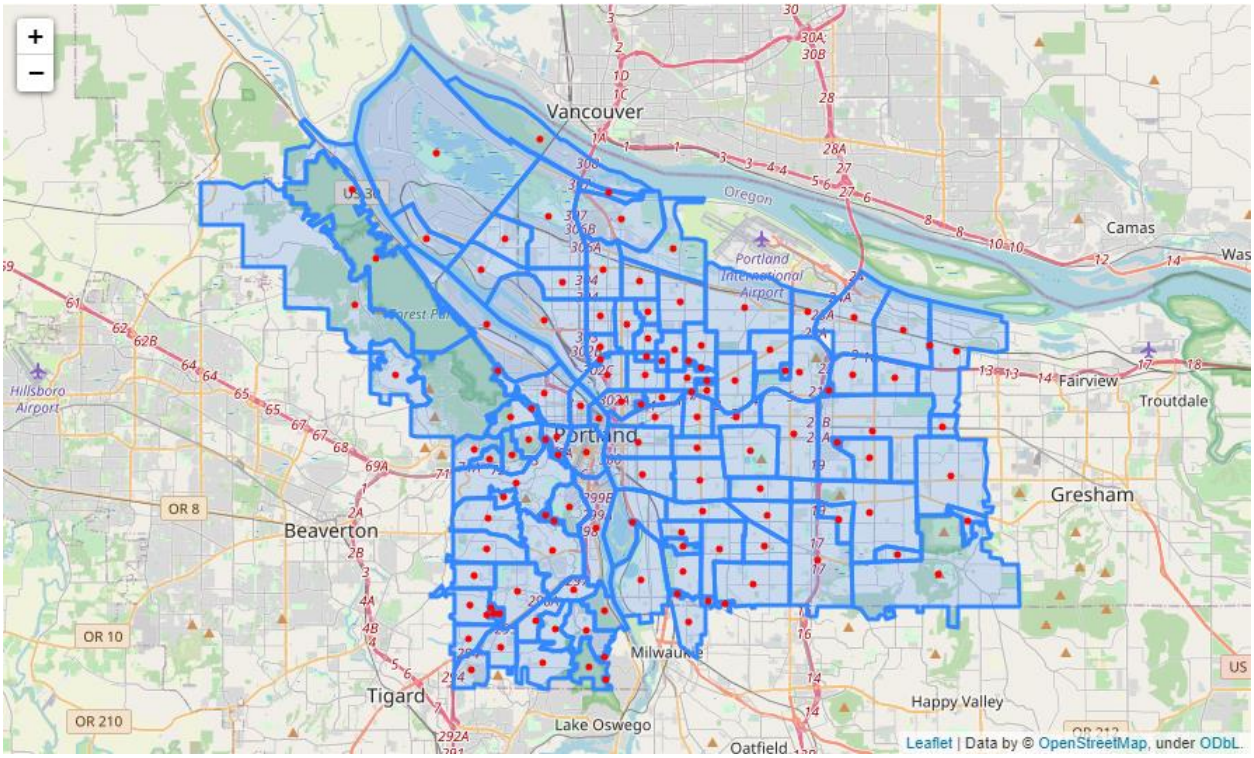

Figure 1. Boundary data for the neighborhoods of Portland, OR. The red points indicate the geospatial centers of each neighborhood.

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. Gaps in the data were filled from information collected from RENTCafé. The column titles for the rent information was standardized to Neighborhood and Rent. 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, data entries for neighborhood names had to be cleaned. There was one additional instance of a neighborhood being represented by its former name.

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 and RENTCafé, only 2 of the 130 areas on Figure 3 were missing rent data. One of these areas represented a smaller, unclaimed neighborhood that did not belong to any association. The other simply had no rental data. This could be due to a lack of reported rental information or even no rental properties to speak of.

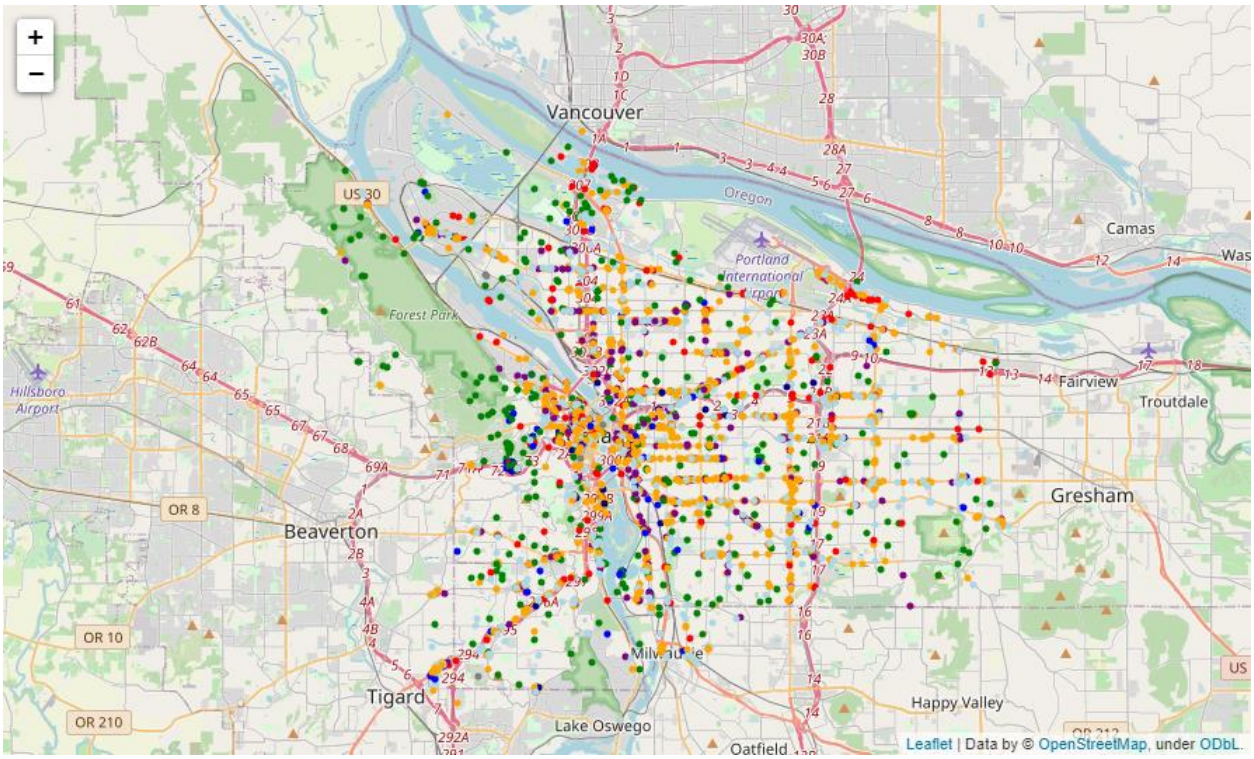

Figure 2. Map of Foursquare venue data points. Color-coded for each venue category.

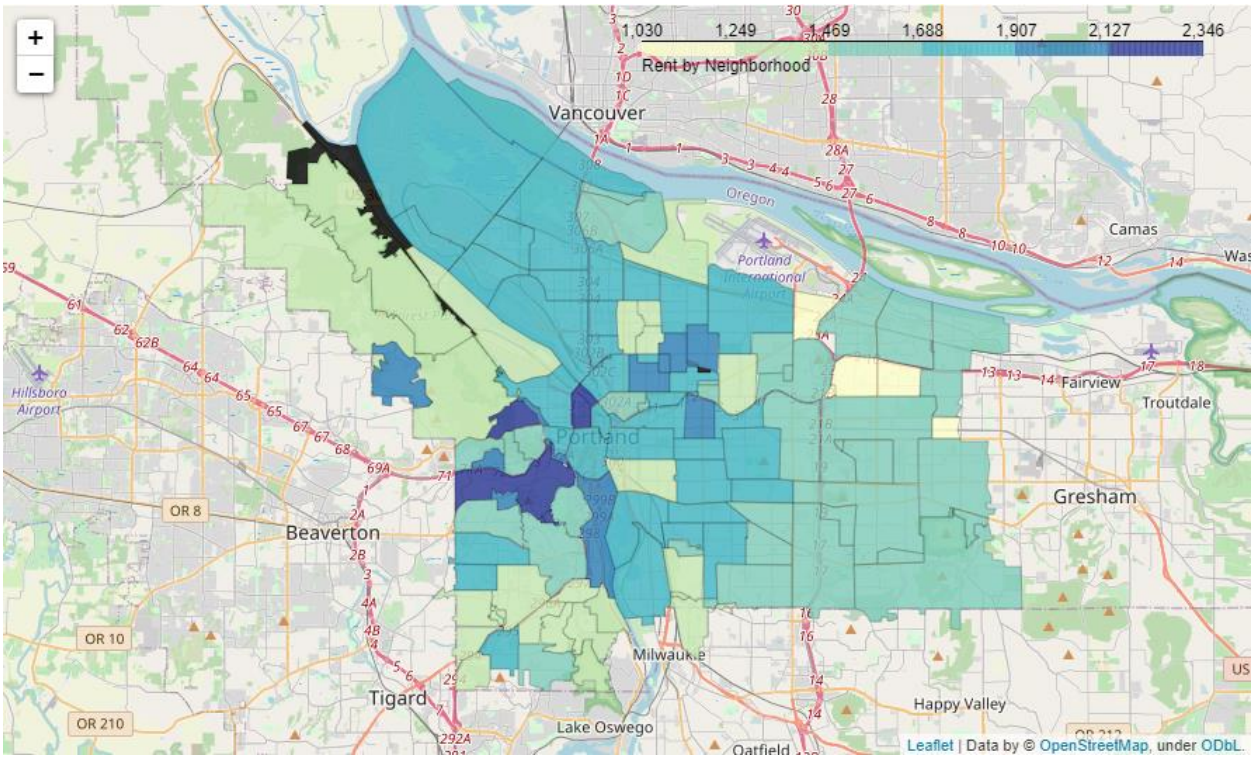

Figure 3. Heatmap representing average rent for neighborhoods in Portland, OR. A black zone represents no rental data for the area.

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. From a simple overlap of venue data and the rental data heat map, it should be apparent where downtown Portland is located. The dense congregation of venues and higher average rent areas is typical of a centrally located area. Each venue point is labeled with a subcategory and a main category for its type. Because of this, analyzing individual venues was deemed to not reveal interesting results in the scope of this analysis.

In order to yield more meaningful results, a larger priority was placed on the main categories that the venues are associated with. Of the 10 main categories for venues in the Foursquare data, 2 of them lacked data points and would be removed during data analysis. Before clustering, it is believed the two most represented categories, Food and Shop & Service, will be the most influential in recommending neighborhoods.

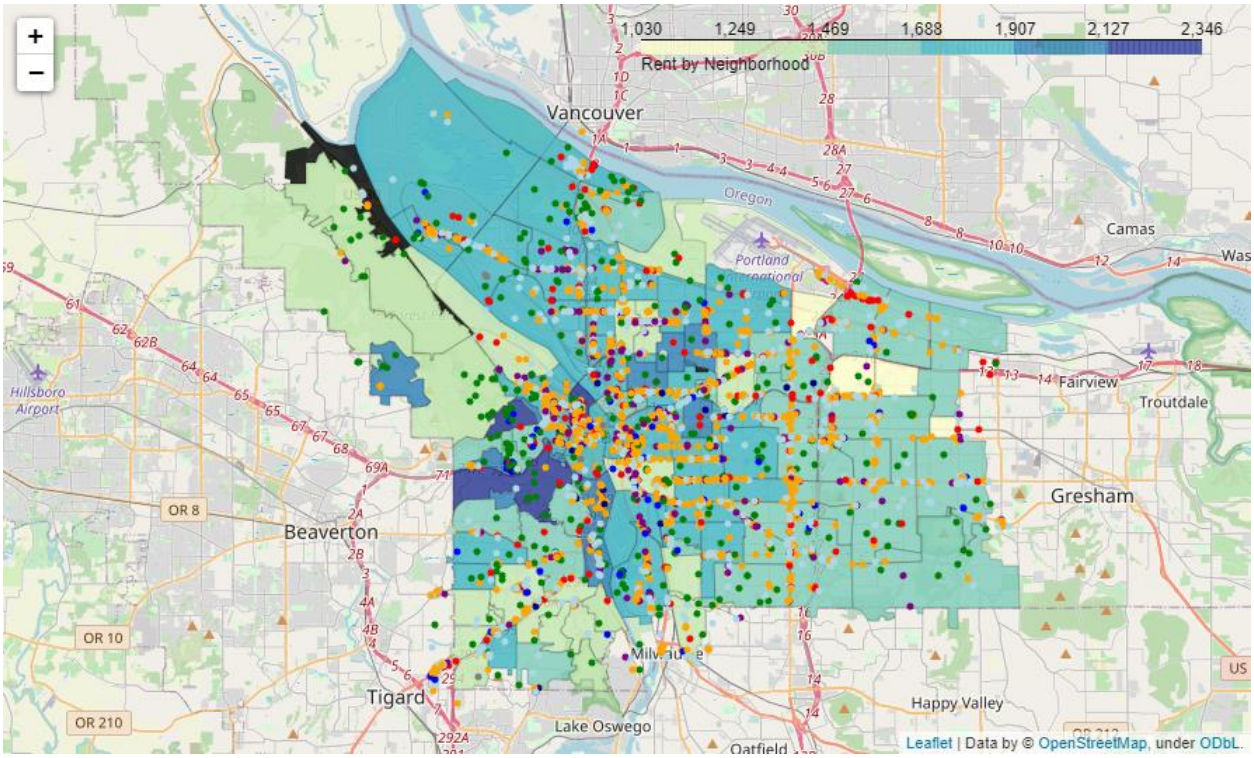

Figure 4. Venue points overlaid on average rent heat map.

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 grouped into 6 clusters. With the clustering labels generated, the data points in Figure 4 were given color and transformed into that which is displayed in Figure 6. A clear divide can be seen partitioning the venue points. The clustering has given us a reasonable spatial grouping by which to compare and recommend neighborhoods in different zones of the city. Should it be deemed necessary for a deeper analysis of the venue data, a more optimal number for the clusters can be calculated and used.

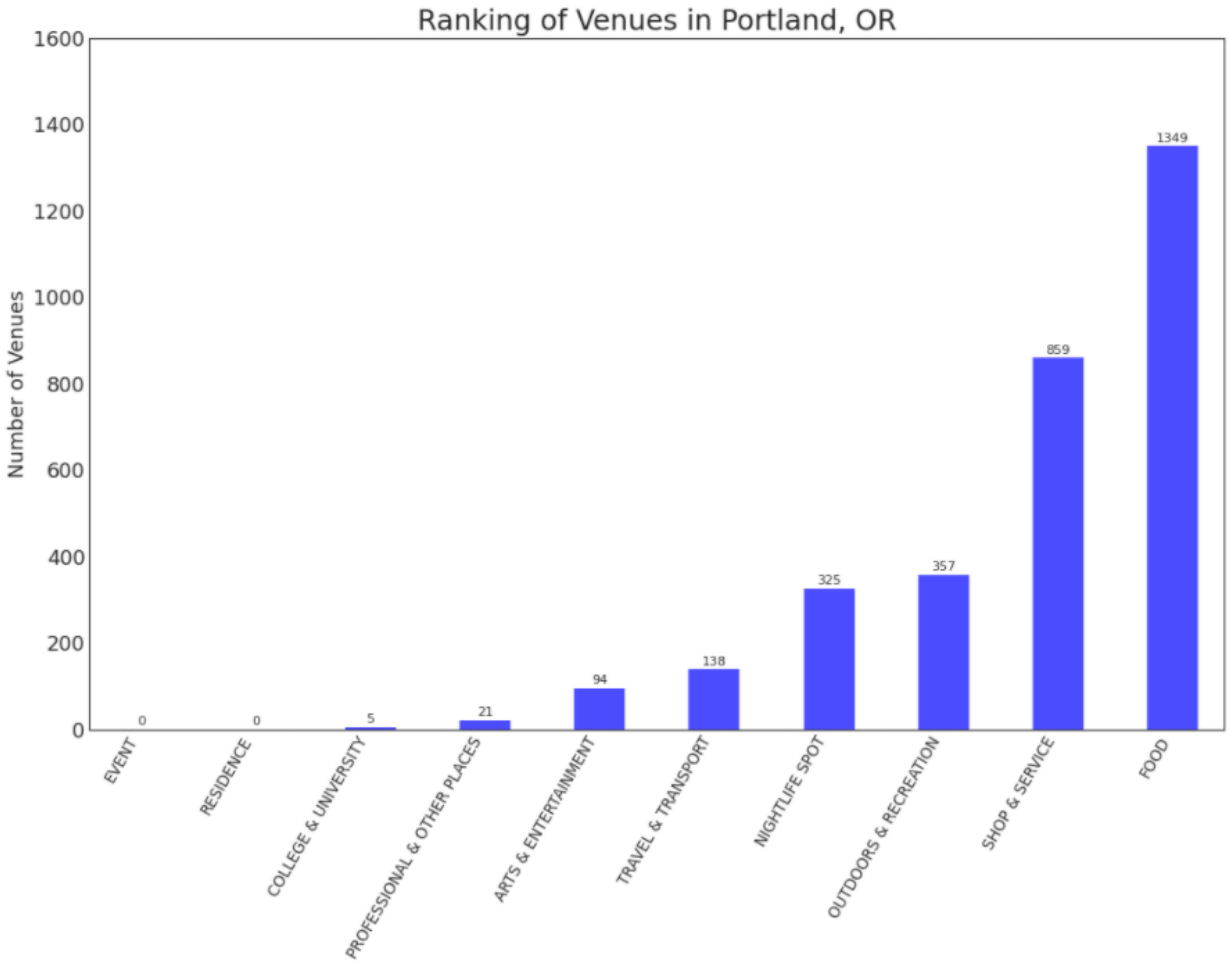

Figure 5. Table of venue distribution across categories.

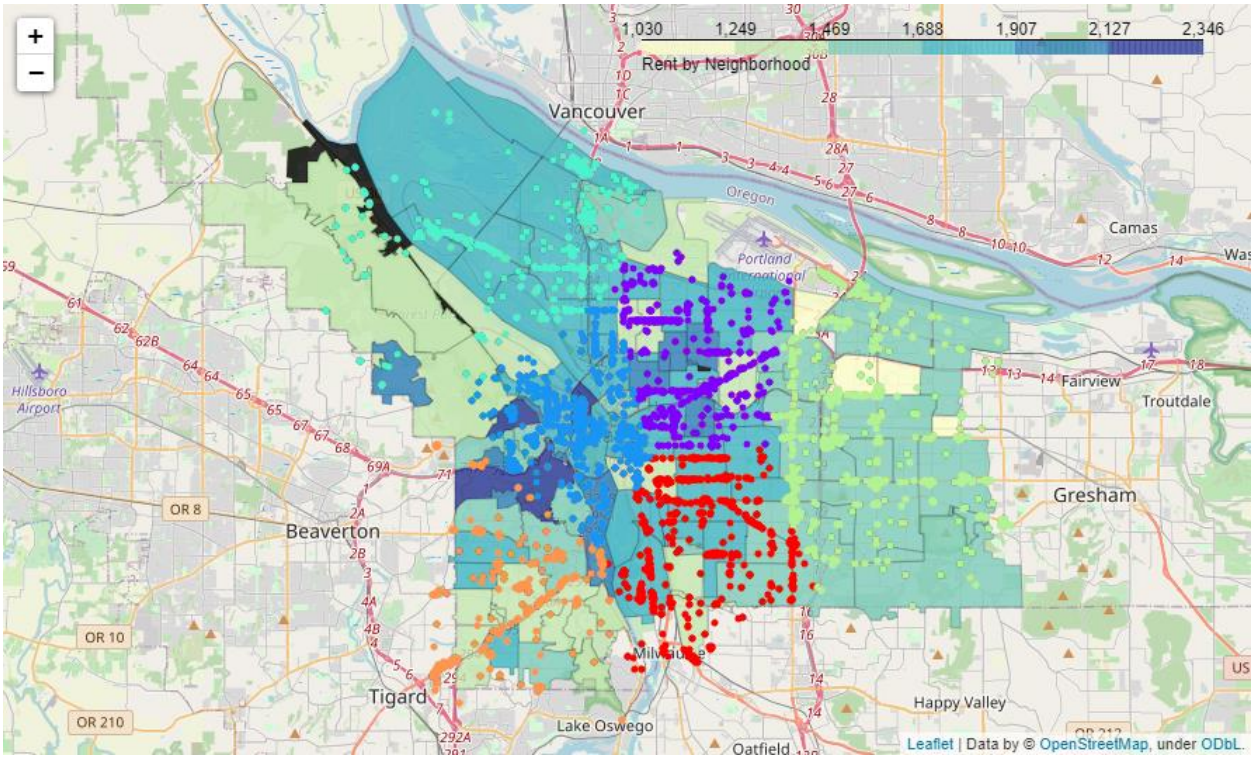

Figure 6. Venue data points clustered over the average rent heatmap.

4. Results

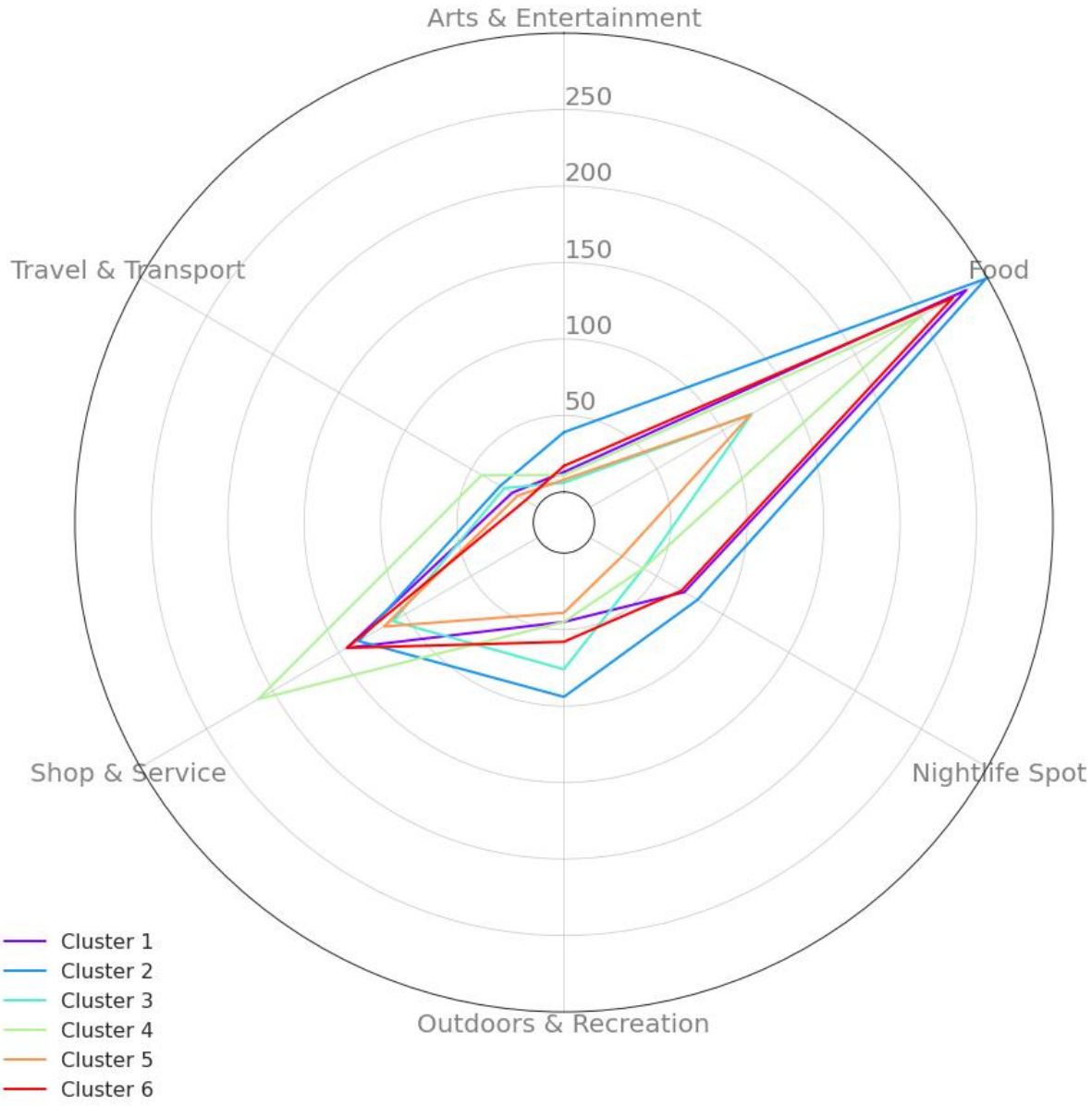

Figure 7. Comparison of venue type distribution in each cluster.

4.1 Cluster 1

The first cluster is representative of the Northeast section of Portland. There are many venues with a reasonable spread of category type, with multiple options for each type. The largest grouping of venues are clustered on and around a major street: Northeast Sandy Boulevard. This area of the city does contain an average spread of rental prices, with neighborhoods spanning nearly the entire range of rent values. This cluster has the second most Food venues and the third most Nightlife Spot venues, but has a below average number of Outdoors & Recreation venues.

4.2 Cluster 2

The second cluster represents the downtown area of Portland. This area includes some of the most expensive neighborhoods in the rental data, with rent ranging from average to high and as much as $2346/month. The venues are centered on the downtown area and includes the most amount of Food, Nightlife Spot, Arts & Entertainment, and Outdoors & Recreation venues in the city. While having the least spread compared to the other clusters, cluster 2 actually contains the most venues.

4.3 Cluster 3

The third cluster is representative of the Northwest section of Portland. This cluster contains the least amount of venues. There is also a good deal of distance between most of the venues, with most residing along North Lombard Street. One neighborhood in this area lacks rental data, which may indicate that the area may not be as well represented as the others. A large part of the cluster also spans Forest Park, a large urban park just northwest of downtown Portland. This area has about average rent.

4.4 Cluster 4

The fourth cluster is representative of the Eastern section of Portland. This cluster has average to low rent, with rent as low as $1030/month. This area has the most Shop & Service venues while having some of the lowest number of venues in Outdoors & Recreation and Nightlife Spot. Venues are spread out a good deal, but there is some clustering within proximity to the major intersections. Neighborhoods with the lowest rent in the city are located toward the north end of this cluster.

4.5 Cluster 5

The fifth cluster is representative of the Southwest area of Portland. This area ranges nearly the entire spectrum of rent in the city, with rent gradually increasing as you move north, closer to the downtown area. Venues are spread out and centered on major intersections and the I5. This cluster has the least amount of venues and is not well represented in any category.

4.6 Cluster 6

The sixth and final cluster is representative of the Southeast area of Portland. This area has average rent, with rent decreasing as you move away from the downtown area. The area also has a high amount of Food and Shop & Service venues. Venues appear to be centered on every major avenue of the cluster.

5. Discussion

5.1 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, though the lower rent areas also have a similar clustering of venues. The clusters with the highest rent also has the most amount of Food venues, possibly indicating a high correlation between the two. The same can be said about Nightlife venues, but to a lesser degree. Arts & Entertainment and Travel & Transport venues appear to have the least amount of correlation to average rent, though this could also be due to the presence of fewer points.

Cluster 2 with its high rental areas contains downtown Portland. Right across the Willamette River is Cluster 1 and Cluster 6 with its relatively lower average rent. The other clusters represent the areas with average to low relative rent, spanning to the north, southwest, and east 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.

5.2 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.

6. Conclusions

We set out to recommend neighborhoods for potential renters in the Portland, OR area. Data mainly 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. Additional rental data also had to be obtained to cover the entire city. 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 southwestern quadrant 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, incorporating other metrics such as crime rates, property value trends, and schooling could provide more accurate recommendations.