Where to start a business around Portland: Portland Area business and house price analysis

Tyler Stannard

Dec. 14, 2019

# Introduction

For residents of various localities who are interested in finding alternate income sources such as property or small business investments, a thorough understanding of the property prices, price trends, and local businesses is critical. In this case, Portland, Oregon, USA was chosen for the investment analysis because it is local to the author.

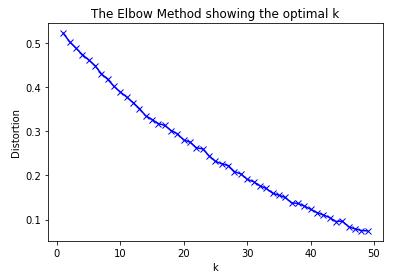
# Data

The US Census spatial data [[[1]](#endnote-1)], Zillow house value data [[[2]](#endnote-2)], and Foursquare API data [[[3]](#endnote-3)] were used to estimate the best Zip code in the Portland area to buy a house or start a business, and what type of business to buy. The Census spatial data was used to determine the zip code boundaries so an investor can easily visualize the locations being discussed. The Zillow house data was used to determine the house values of the zip codes, as well as their quarterly and annual price trends. The Foursquare API was used to determine the businesses available in each zip code so the investor can understand what type of business may be successful in the area or if the amenities they are interested in will be available in their area.

# Methodology

section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

First, the housing price data of Oregon based on zip code was downloaded from the Zillow research website. The zip codes from the Zillow data were chosen as the location anchors for the analysis. Next, Census data was found for Polygon map descriptions of each Zip code so that each Zip code could be plotted on the map for location visualization. Each Zip code was polled using the Python geolocator package to determine the GPS coordinates at the center of the zip code for loading data from the Foursquare API. Functions were written for turning the Foursquare data into a pandas data frame for easy analysis, and then a data frame containing each zip code, latitude, longitude, and venue was created. The number of venues per zip code was plotted to find the best zip codes for doing business based on convenience. The dataframe was then manipulated by one-hot encoding every venue type for every zip code for input into a machine learning algorithm, making the shape n\_venues x n\_categories. The mean one-hot of each zip code was taken to get the relative proportions of each type of venue per zip code, and the top 10 venues per zip code based on frequency were listed. A dataframe ranking the most common venues of each zip code was then created. A k-means analysis was attempted on the venue-location dataset. Based on the K-means “elbow” analysis shown in Figure 1, evaluating K-means w/1-50 K values, it appeared that the Kmeans algorithm could not group the venue types well based on zip code. The areas may have been too diverse with different types of venues for K-means to work well, since there were 184 venue types across only 748 venue locations. Next, a K-means elbow analysis of zip codes based on four Zillow price metrics was attempted and an elbow at K=8 was found as shown in Figure 2. The clusters were then plotted on the map using the folium API and the census polygon data. The venues of the areas with the fastest growing prices were then found and property purchase recommendations were given.

  
Figure 1: K-means elbow analysis on zip-code/venue data set does not show an elbow indicating the k-means algorithm doesn’t fit this dataset well.

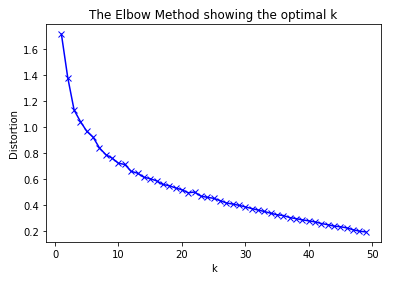
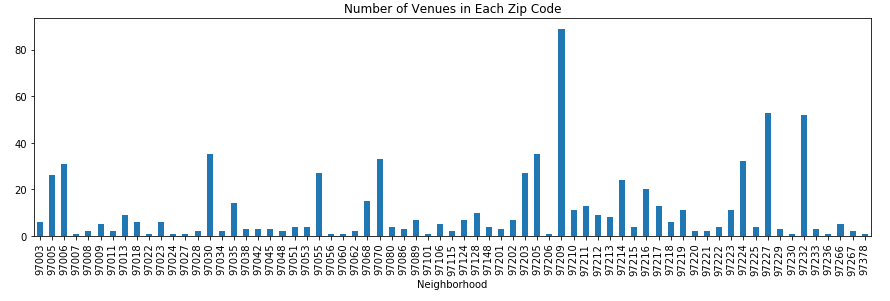


Figure 2: K-means elbow analysis on zip-code/venue data set does not show an elbow indicating the k-means algorithm fits this dataset reasonably well at K=8.

# Results

The number of venues were plotted for every zip code in Figure 3. The three zip codes with the most venues were 97209, 97227 and 97232. Many zip codes had fewer than 10 venues, which suggests the majority of the zip codes in the Portland-Vancouver area are fairly rural, with a few dense city areas.

In Figure 4, the mean of the mean Zillow home value index for each cluster of zip codes was plotted. Cluster 2 was the most expensive at over $400,000 while clusters 1 and 4 was the cheapest, with prices below $300,000. In Figure 4 right, the mean of the Zillow house price change metrics were plotted for each cluster. The metrics are “MoM,” month-over month, “QoQ,” quarter-over-quarter, and “YoY,” year-over-year. The most expensive cluster showed the most significant declines in prices recently based on all three price change metrics. The cheapest cluster showed the most significant increase in prices, suggesting that cluster 4 may represent a good buying opportunity for interested investors.

Figure 3: Number of venues per zip code in the Portland-Vancouver area.

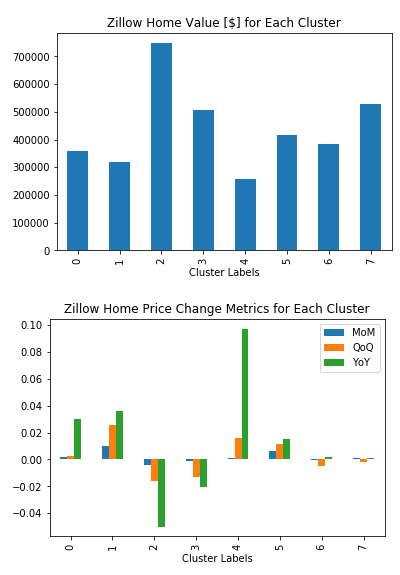
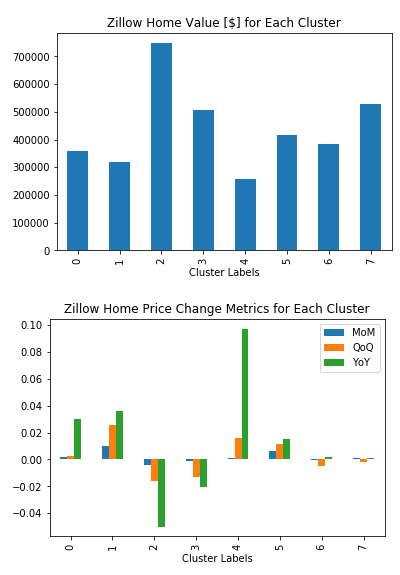


Figure 4: Left) Mean Zillow home value for each of the 8 k-means clusters. Right) Zillow price change metrics for each of the clusters.

The clusters were also visualized on the map using the folium API and the United States 2010 census polygon data. From this, it was clear that cluster 4 was a semi-rural area between Portland and Salem. It appeared that some clusters were grouped based on their proximity to the center of the city, which is expected since generally some of the highest priced locations are in or near the center of the city, and relatively cheaper locations surround. Cluster 4 contained two zip codes, 97101 and 97051. Figure 6 shows that 97101 only had a vineyard, but 97051 had some convenience stores and a fast food restaurant. This suggests that 97101 may be a good investment for a farm, and 97051 could be a potential real estate investment or a good place to add a small business such as a gas station.

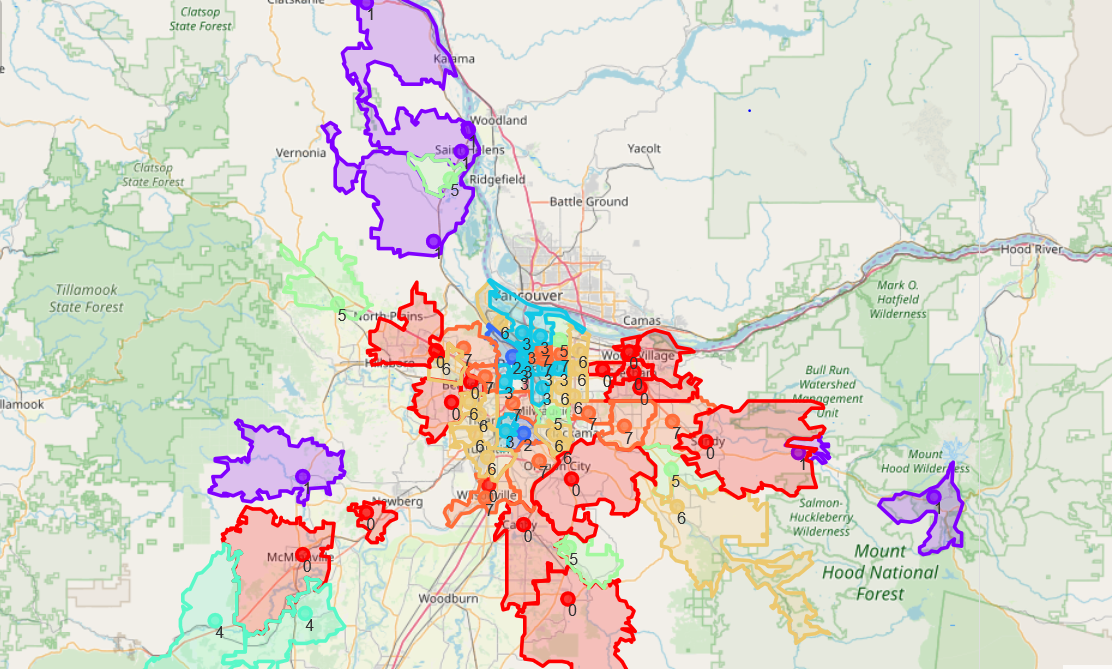


Figure 5: Visualization of house-price dataset clusters on map of Portland area. Cluster 4 is in lower left corner.

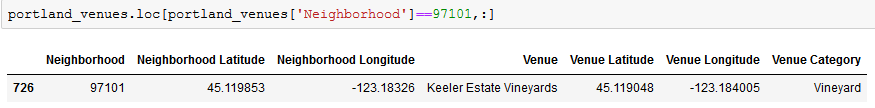


Figure 5: Venues of the two zip codes in the Portland area with the greatest price increases in recent history.

# Discussion

Zillow house price data was combined with Foursquare API business data and Census location data to visualize house prices and businesses in the Portland-Vancouver area of Oregon, United States. Two zip codes in a rural area between Portland and Salem were identified as good potential investments because of low prices and rapid price increases in the last year. The more expensive locations near the center of the city had flat-to-declining prices, suggesting they may not be good investments right now as the market may be cooling off there after many years of rapid and stable increases.

# Conclusion

Economic, geographical and business data of the Portland-Vancouver area were combined to determine the best potential investments in the area. 97101 and 97051 were determined to be the best property investments and potentially good places to start a small convenience store.

# References

1. . <http://www2.census.gov/geo/tiger/GENZ2010> [↑](#endnote-ref-1)
2. . <https://www.zillow.com/research/data/> [↑](#endnote-ref-2)
3. . <https://api.foursquare.com/v2/venues/explore> [↑](#endnote-ref-3)