Determination of Potential Thai Restaurant Locations in Portland, Oregon

An assessment using k-means clustering of neighborhoods

IBM Data Science Capstone Project

By Watts Dietrich

Introduction

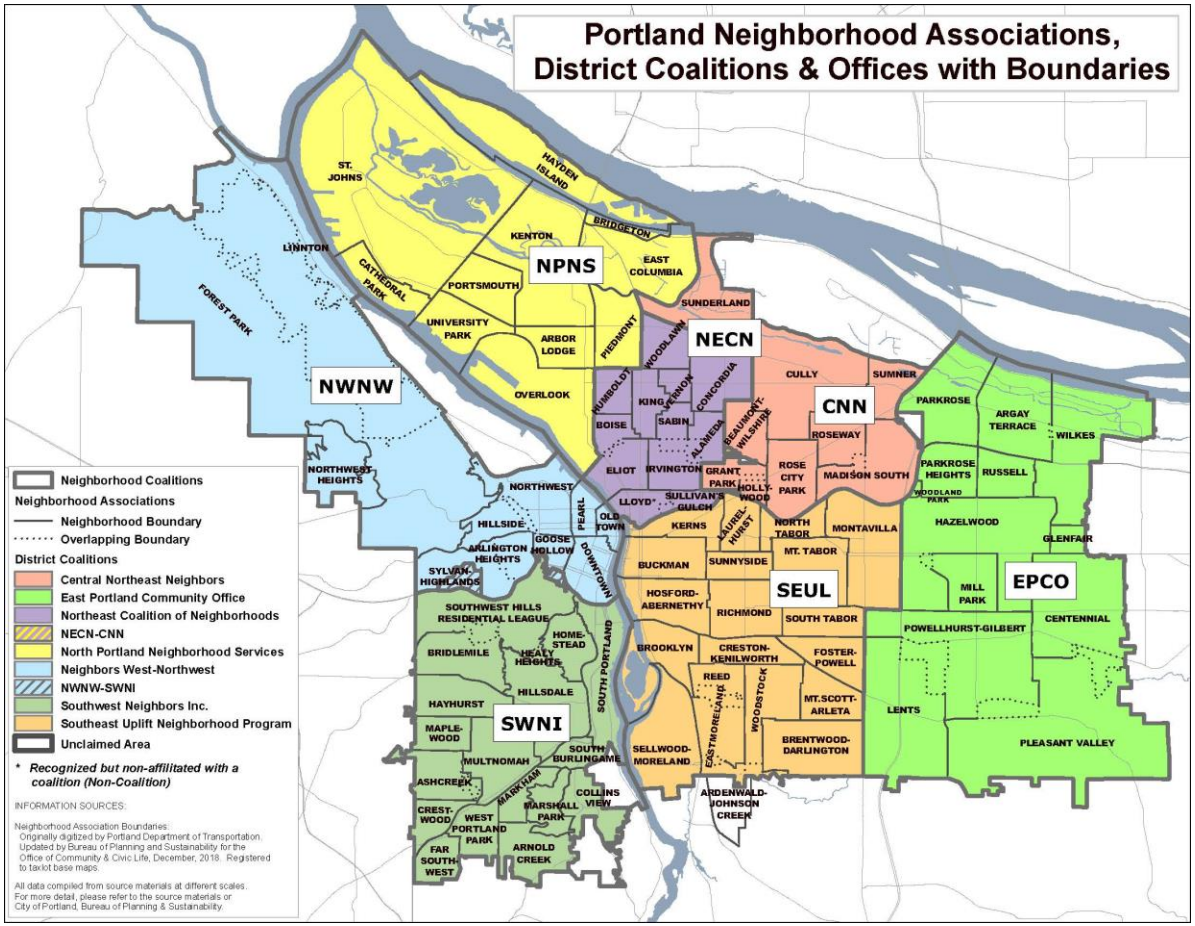
- Opening a new restaurant is often risky.
- Choosing a good location is a key factor for the success of the business.
- Location selection is not a straightforward problem to solve. Ideally, we want to find an area which maximizes access to potential customers while minimizing competition from similar restaurants.

The Problem

- In this presentation, I seek to propose some potential solutions to a hypothetical question: Which neighborhoods in Portland, Oregon would be best for a new Thai restaurant?

Background

- Portland is a city of about 650,000 people.
- The city is divided into 95 neighborhoods



Background

- There are many factors we could examine to help determine the best neighborhoods for a new Thai restaurant. Examples include:
  - Proximity to competition – How many Thai restaurants already exist in the neighborhood?
  - Proximity to complementary businesses – While being too near other Thai restaurants may be bad, being near other businesses (e.g. bars, shops, entertainment) is often a good thing.
  - Median household income – Low income neighborhoods will generally house fewer and less-frequent customers.
- This analysis will concentrate primarily on minimizing direct competition by finding neighborhoods which lack Thai restaurants, with some consideration to median income.

Data Description

- A list of neighborhoods and their GPS coordinates was obtained from https://en.wikipedia.org/wiki/Neighborhoods_of_Portland,_Oregon
- Median household income data were obtained from a 2018 report on Portland Real Estate
- Data on existing Thai restaurant locations were obtained via the Foursquare API.

Methodology - Approach

- To answer the question of where to locate a new Thai restaurant, I decided to focus on determining neighborhoods where there are few or no existing Thai restaurants and to cross-reference these result with data on median household income to find neighborhoods where competition is low and income is high.

Methodology - Exploration

- First, a list of all 95 Portland neighborhoods and their geographic coordinates was compiled into a Pandas dataframe.
- Venue data for the city was imported via the Foursquare API and grouped by neighborhood.
- Exploratory data analysis found that there are 25 Thai restaurants listed in Portland and that they are spread across many neighborhoods.
- I had originally intended to analyze Indian restaurants, rather than Thai. However, Foursquare lists only four venues in the “Indian Restaurant” category. This lack of data led me to change my project slightly. It is worth noting that a real-world analysis would need to find a way to deal with this problem.

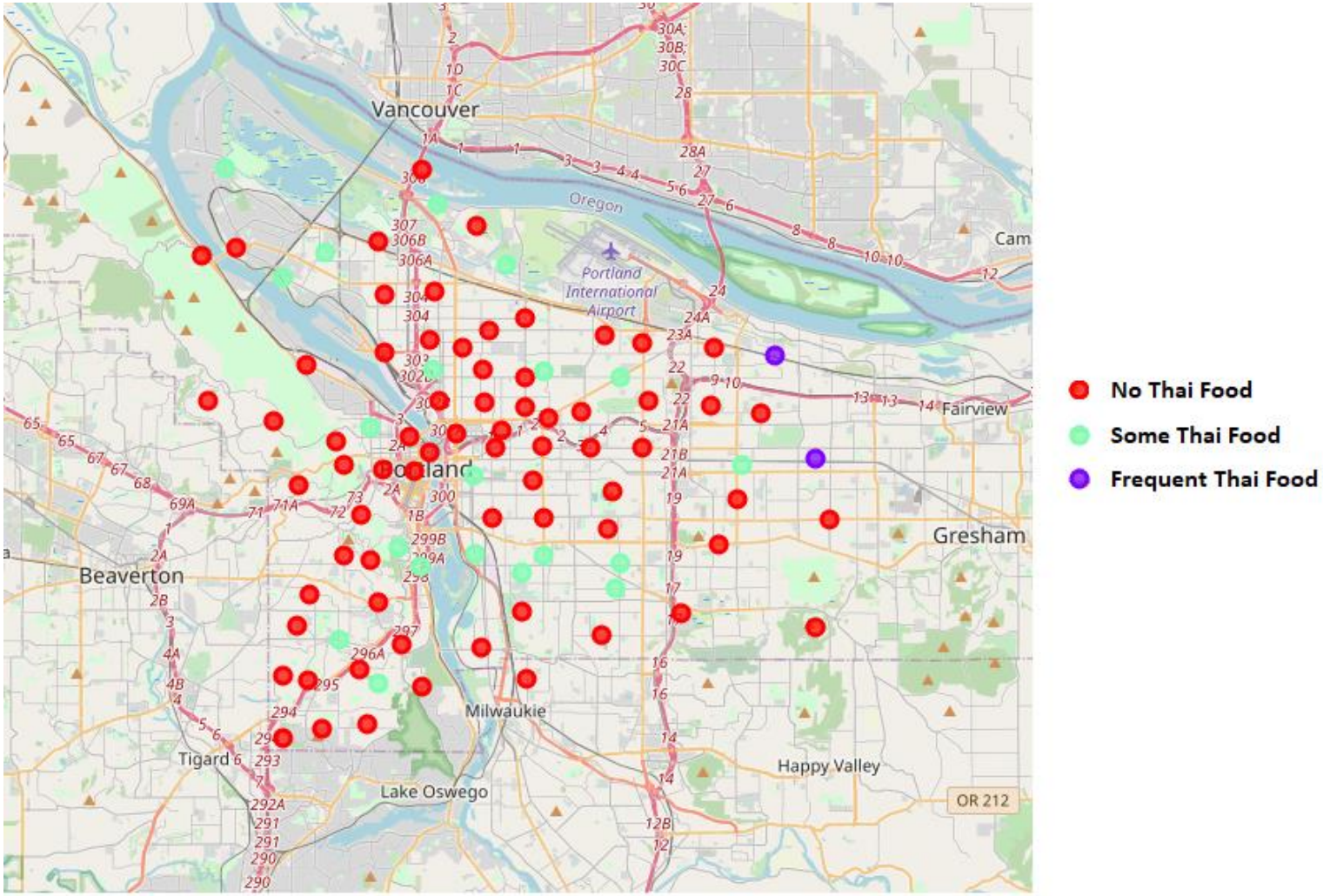
Methodology – Clustering of Neighborhoods

- K-means clustering was performed on the list of neighborhoods. The feature used to build the clusters was the number of Thai restaurants in the neighborhood.
- I chose to set k=3 for this analysis. This naturally tends to yield different clusters for low, medium, and high numbers of Thai restaurants.
- Folium was then used to visualize the map of clustered neighborhoods.

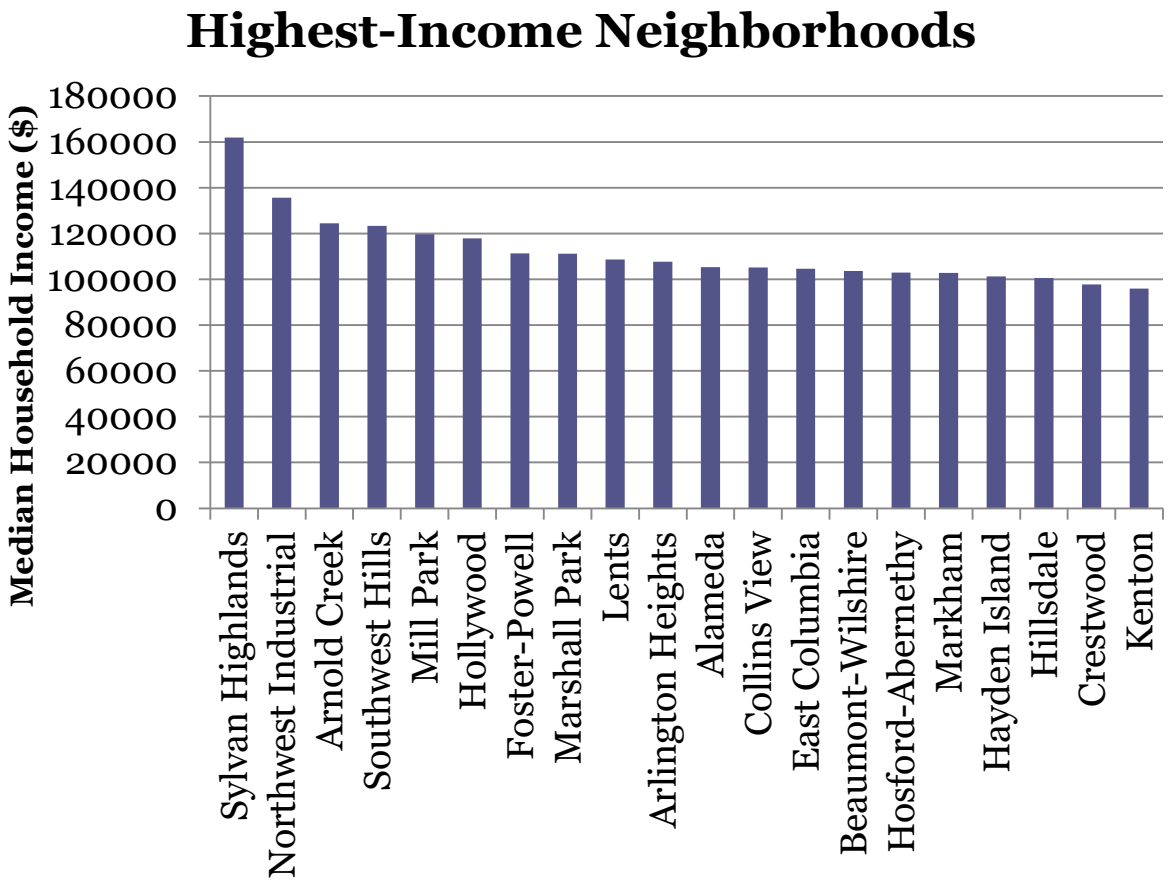
Results

- K-means clustering yielded the following clusters:
  - Cluster 0: No existing Thai restaurants
  - Cluster 1: High frequency of Thai restaurants
  - Cluster 2: Lower frequency of Thai restaurants
- The number of neighborhoods in each cluster:
  - Cluster 0: 69 neighborhoods
  - Cluster 1: 2 neighborhoods
  - Cluster 2: 20 neighborhoods

Results – Map of Clustered Neighborhoods



Results – Neighborhoods Ranked by Median Household Income



Discussion

- To minimize competition, we should focus on neighborhoods in cluster 0 (those neighborhoods with no existing Thai restaurants).
- There are only 25 listed Thai restaurants in the 95 neighborhoods, so a large number of neighborhoods fall into cluster 0 (69 out of 95)
- More information is needed to further narrow our options.

Discussion

- One way to further narrow our location search is to cross-reference the cluster-0 neighborhoods with the median household income of the neighborhood.
  - Generally, higher-income families are more likely to frequently dine out at nice restaurants.
  - While income is likely far from a perfect metric for predicting customer engagement, but it is a good place to start.

Discussion

- Narrowing our search to high-income neighborhoods with no existing Thai places still yields a fairly long list.
  - There are 42 neighborhoods that have no Thai places and a median household income above $60,000 (the median income for the U.S.)
- Looking at the top of this list, I would recommend starting our location search in the following five neighborhoods:
  - 1. Wilkes
  - 2. Parkrose
  - 3. Arnold Creek
  - 4. Sunderland
  - 5. Mt. Scott Arleta

Discussion – Questions for Continued Research

- The analysis and recommendations presented here are a good start, but are by no means a complete picture. For the sake of simplicity, I have neglected some factors that could play a role in location selection.
- Some other factors that could be considered:
  - Spatial distance to nearest competition
  - Proximity to complementary businesses
  - Property values (how much will it cost to rent a location?)
  - Ethnic demographics (where is the local Thai population concentrated?)
- Finally, the large number of neighborhoods compared to the number of existing Thai places (95 neighborhoods, 25 Thai restaurants) may suggest our competition analysis is too granular. It may be better to look at larger areas of the city rather than individual neighborhoods.

Conclusion

- If our goal is to find a location for a new Thai restaurant in Portland with minimal direct competition and proximity to high-income customers, we should start our search in the neighborhoods of Wilkes, Parkrose, Arnold Creek, Sunderland, and Mt. Scott Arleta.