Opening a restaurant in Portland, Oregon

Will Morgan Feb 2019

Table of Contents

I. Introduction…………………………………….……………..3

II. The Business Problem…………………….…………………3

III. Target Audience ……………………………….…………...3

IV. The Data……………………………….……….…………...4

V. Methodology………………………….……….…………….5

VI. Results……………………………….…………..………...10

VII. Discussion………………………………….……………..11

VIII. Conclusion……………………………………………….12

IX. References…………………………………………………12

I. Introduction

Portland, Oregon is known as a ‘foodie’ town. There many type of restaurants and the overall quality of the food rated very highly. There is a diverse range of options. Jeff represents an investment firm who is looking to break into the restaurant market in Portland, however Jeff’s firm does not know what kind of restaurant to open, nor do they know much about the city’s layout. Jeff’s firm wants to know what type of food will work and which neighborhood is the best neighborhood.

II. The Business Problem

The objective of this project is to collect data and analyze data on every neighborhood in Portland, Oregon. The data science methodology will be applied to answer two questions: What type of restaurant in Portland is very likely to succeed? And, Which neighborhood would be the best neighborhood to open a new restaurant?

III. Target Audience

The target audience of a report like this would be any investment firm, individual investor, or entrepreneur that is considering entering the restaurant business in Portland, Oregon. However, the same methodology could be used to analyze other cities to provide the same solution.

IV. The Data

In order to provide a solution for Jeff and his firm, various kinds of data will need to be collected and analyzed.

The data used in this project:

1. Geo spatial data for Portland, Oregon

source: <https://opendata.arcgis.com/datasets/9f50a605cf4945259b983fa35c993fe9_125.geojson>

purpose: For neighborhood boundaries and names

method: in-notebook download

2. Demographic Data for Portland, Oregon

source: <https://www.pdxmonthly.com/home-and-real-estate/2019/03/portland-neighborhoods-by-the-numbers-2019-the-city>

purpose: For population statistics, median house hold income, walk score, etc

method: Table scraped with Beautiful Soup

3. Residential Building Permit Data

source: <https://gis-pdx.opendata.arcgis.com/datasets/residential-building-permits/data>

purpose: Predicting future growth

method: Direct download

4 . Foursquare venue data

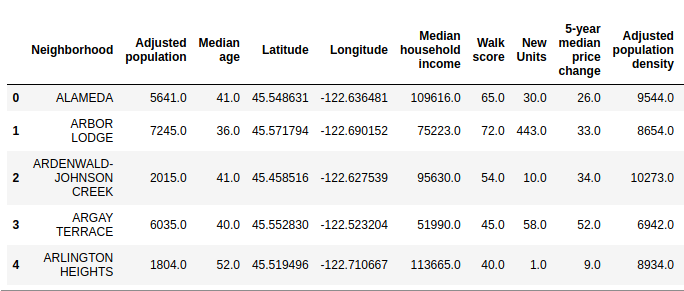
source: [www.foursquare.com](http://www.foursquare.com/)

purpose: To analyze restaurants and other venues in each neighborhood

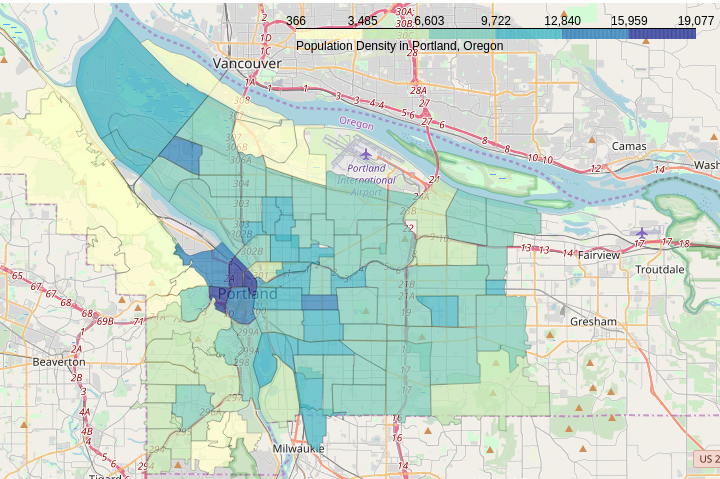
method: API calls

V. Methodology

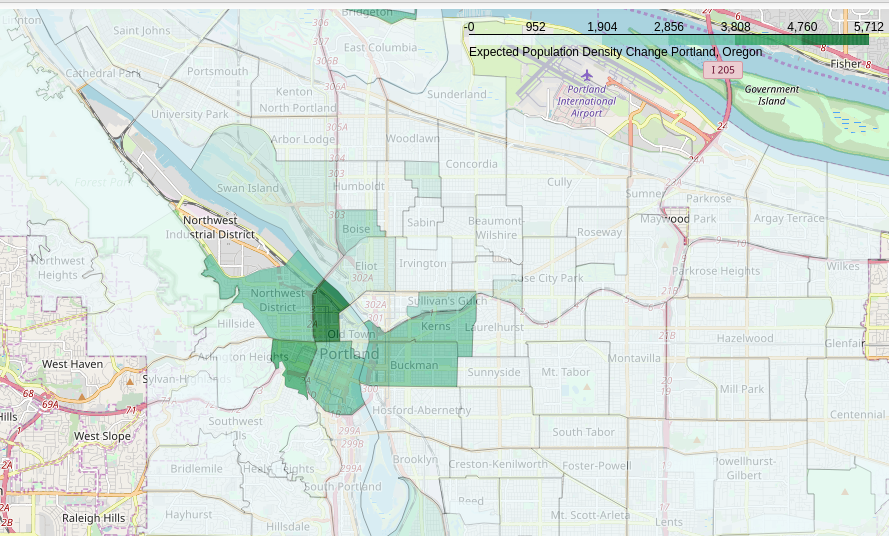
First a list of definitive neighborhoods was obtained from the demographic data then each of the neighborhoods corresponding coordinates were added using Open Street Map Api data. After filtering desirable columns from the demographic data, a base data set was produced:



Using this data, choropleth maps were produced to get an idea of how the population was distributed:



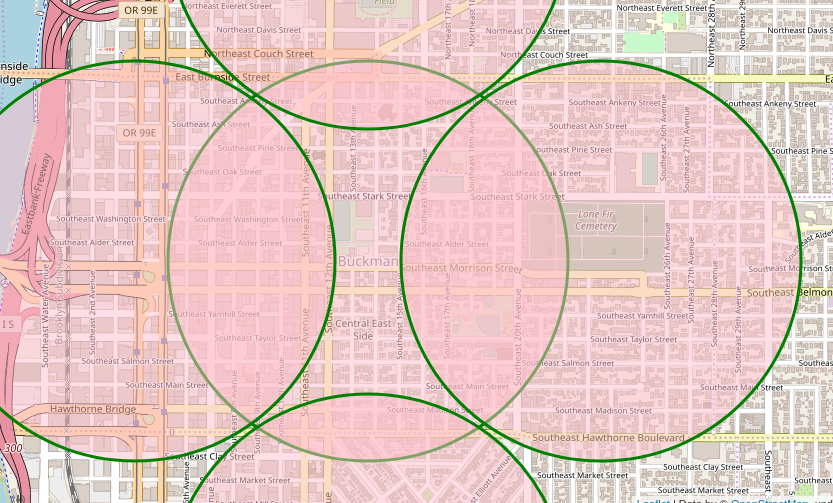
Then by grouping the neighborhoods together in the permit data set and summing the number of ‘NEW UNITS’, a new units column was added to the base data set. This number was multiplied by 1.5 to capture the fact that in Portland most people have a roommate, therefor, for every 1 new unit the population increases by 1.5 persons. With this information an expected population density was able to be calculated. Using these new columns, a choropleth map was produced to show areas with the expected change in density:

Units: people/sq.mile

At this point a general sense that opening a restaurant around the city center would be the correct decision. However, venue information still had not obtained to confirm this.

After some general exploratory analysis of the data, the focus switched to the objective answering the two questions in the business problem.

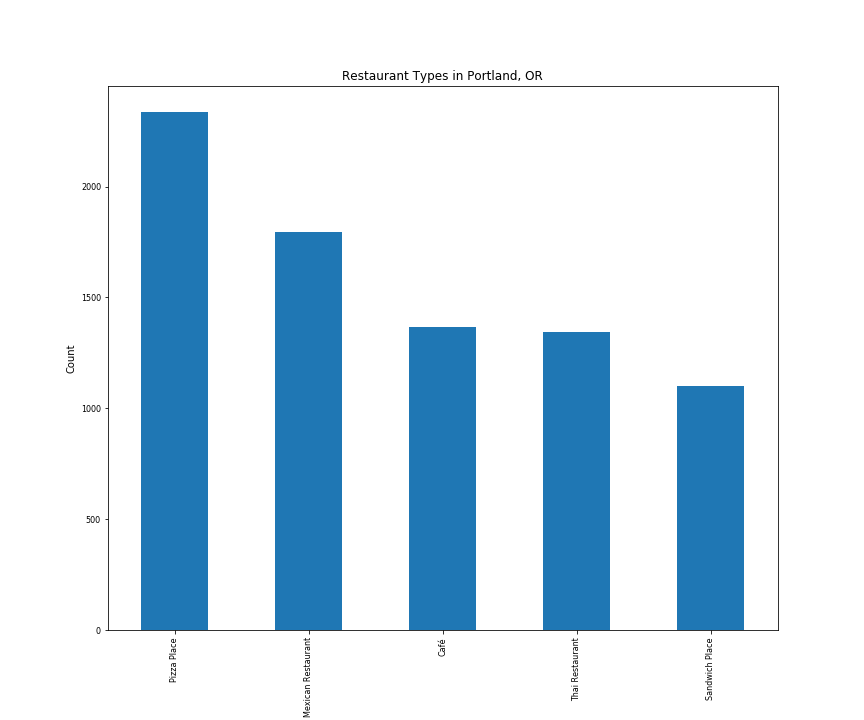
To answer the first question: “What kind of restaurant will have high likelihood of succeeding?’, data was collected from 5 points for each neighborhood. The neighborhood center, and +.01 points of its Geo-coordinates in either direction and duplicates from overlapping areas were dropped after their first occurrence. Visually it looks like this:

example: Buckman neighborhood

This was done for two reasons. First, it produces a more robust list of venues. Second, although some venues might technically be in other neighborhoods, they are still within walking distance, and should be taken into consideration when choosing a neighborhood to open a new restaurant.

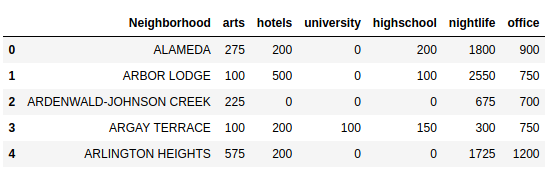
Using this data, analysis was performed and some visuals were produced:

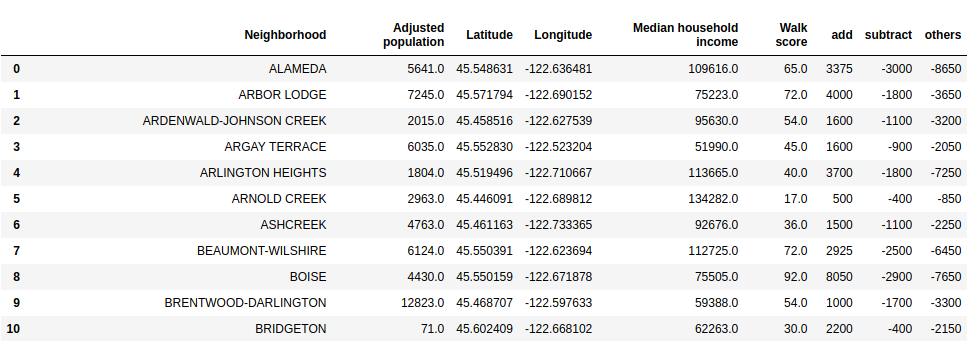
As shown above ‘Pizza Place’ and ‘Mexican Restaurant’ are the two most common types of restaurants of each of these neighborhoods. The assumption also generalizes to Portland as a whole.

 \*It is important to note that “Food Trucks” were omitted in this analysis for two reasons. First, the type of food was not available. Second, food trucks do not usually have normal hours so it is hard to asses their impact.

Based on this analysis Jeff and his firm would have two solid options when it comes to starting a new restaurant. Pizza or Mexican.

At this point, the kind of restaurant to open is known, however the best location is still unknown. To answer the second question, additional data was gathered with the foursquare API. For each neighborhood, a count of Arts and Entertainment venues, Hotels, Universities, High schools, nightlife spots, and office buildings was compiled. Then each of these venue types were multiplied by a weight to emphasize potential impact on business. Weights of 25,100,100,50,75,50 were applied respectively to each category to produce a ‘weights’ data set:

The weights are assumptions, however they are not arbitrary. Each venue has the potential to attract additional patrons to the area. Some more than others. For instance, a hotel will probably attract more people to the area than an art studio. However some venues are likely to have a negative impact on the business. Namely, other restaurants, especially restaurants of the same type. Therefore the foursquare data was used to count the number of pizza shops and Mexican restaurants in each neighborhood and a negative weight of 50 was applied. A negative weight of 25 was applied to all other restaurants in the neighborhood. These were the final components used in determining location:



A formula was devised in order to determine the best location :



((Adjusted Population +New Units) \* .75) \* Median House Hold Income)):

- This portion takes the current poplatuin level, adds the expected population growth (we assumed 1.5 people per new unit) and multiplies by .75. We multiply by .75 to try to capture the fact that it is not every person in the neighborhood that has the median income.

+Contributions

- This is the impact that was determined by venues that in theory would attract foot traffic.

- (Same Restaurant + other restaurants)

- This is negative impact of having similar restaurants in the area and having restaurants in general in the area.

\* Walk Score

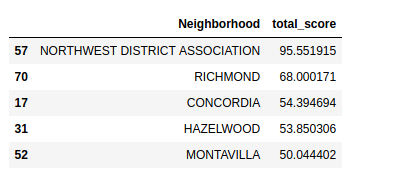
- This captures the ‘walk-ability’ of the neighborhood.

/ 1000000000

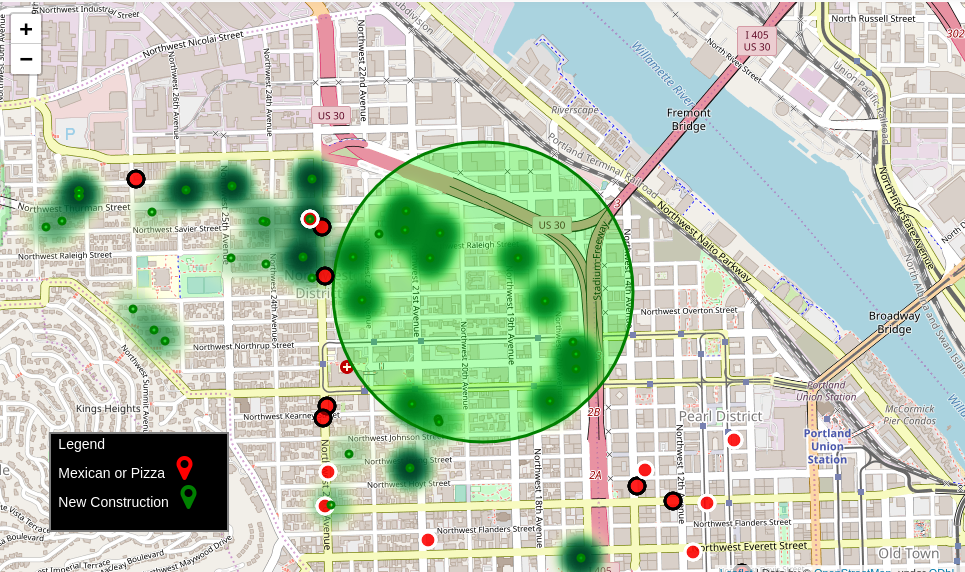
- An arbitrary to divisor to reduce the number into a digestible form.

VI. Results

After extensive analysis, it was determined that opening either a Mexican Restaurant or Pizza Place was the right option. Based on the formula devised, it was determined that the best neighborhood to open in would be “NORTHWEST DISTRICT ASSOCIATION.”



The map of also supports the findings:



This final map strongly supports the conclusion that either a pizza place or Mexican restaurant would fair well in the North West neighborhood. Particularly in the light green circle. The green dots denote new construction, the darker the green, the higher the number of new units. The red dots show existing pizza and Mexican restaurants.

VII. Discussion

Based on this analysis it appears that opening either a pizza place or a Mexican restaurant in the North West neighborhood will likely by a successful venture. Multiple factors were taken into consideration: population, expected growth, income, competition , attractions etc. in order to be a precise as possible.

**For future research:**

While this projected included many factors, there are still addition factors that could be taken into consideration. For example, the rental/lease rates could be taken into consideration for each neighborhood, this could impact profitability. Also, a deeper dive into the contributing and detracting venues could prove to be valuable. For example, for each hotel, the capacity and fill rate could be taken into consideration when assigning weights, or the size of each office building could also be taken into consideration.

VIII. Conclusion

At the beginning of this project, a business problem was identified. The appropriate data requirements were determined and the data was collected. After wrangling the data and performing extensive analysis, a solution was produced. This solution would be presented to Jeff and his firm and they would proceed with their decisions for opening a new restaurant.

IX. References

1. Geo spatial data for Portland, Oregon

source: <https://opendata.arcgis.com/datasets/9f50a605cf4945259b983fa35c993fe9_125.geojson>

purpose: For neighborhood boundaries and names

method: in-notebook download

2. Demographic Data for Portland, Oregon

source: <https://www.pdxmonthly.com/home-and-real-estate/2019/03/portland-neighborhoods-by-the-numbers-2019-the-city>

purpose: For population statistics, median house hold income, walk score, etc

method: Table scraped with Beautiful Soup

3. Residential Building Permit Data

source: <https://gis-pdx.opendata.arcgis.com/datasets/residential-building-permits/data>

purpose: Predicting future growth

method: Direct download

4 . Foursquare venue data

source: [www.foursquare.com](http://www.foursquare.com/)

purpose: To analyze restaurants and other venues in each neighborhood

method: API calls