Opening a restaurant in Portland, Oregon

Will Morgan Feb 2019

Will Morgan Feb, 2019
IBM Coursea Capstone

Table of Contents

Will Morgan Feb, 2019
IBM Coursea Capstone

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.

Will Morgan Feb, 2019
IBM Coursea Capstone

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 by extracting ‘New Units’ being built by neighborhood.

    method: Direct download

4 . Foursquare venue data

    source: www.foursquare.com

    purpose: To analyze restaurants in each neighborhood as well other types of businesses that may impact a new business.

    method: API calls

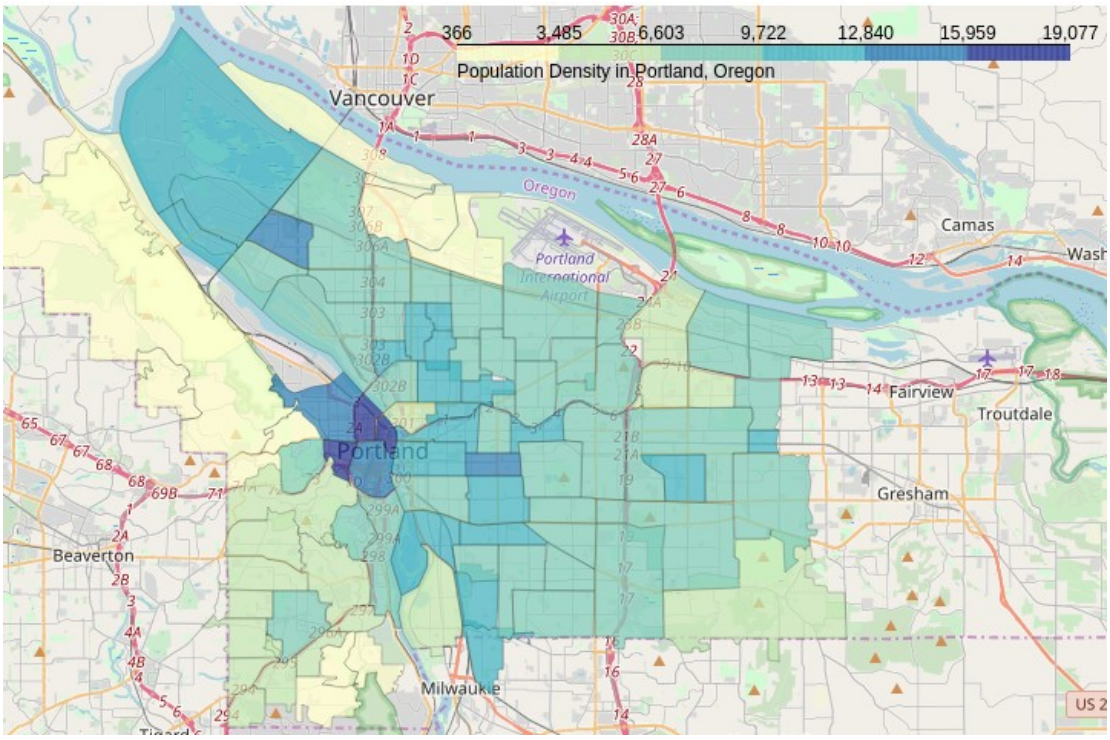
Will Morgan Feb, 2019
IBM Coursea Capstone

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:

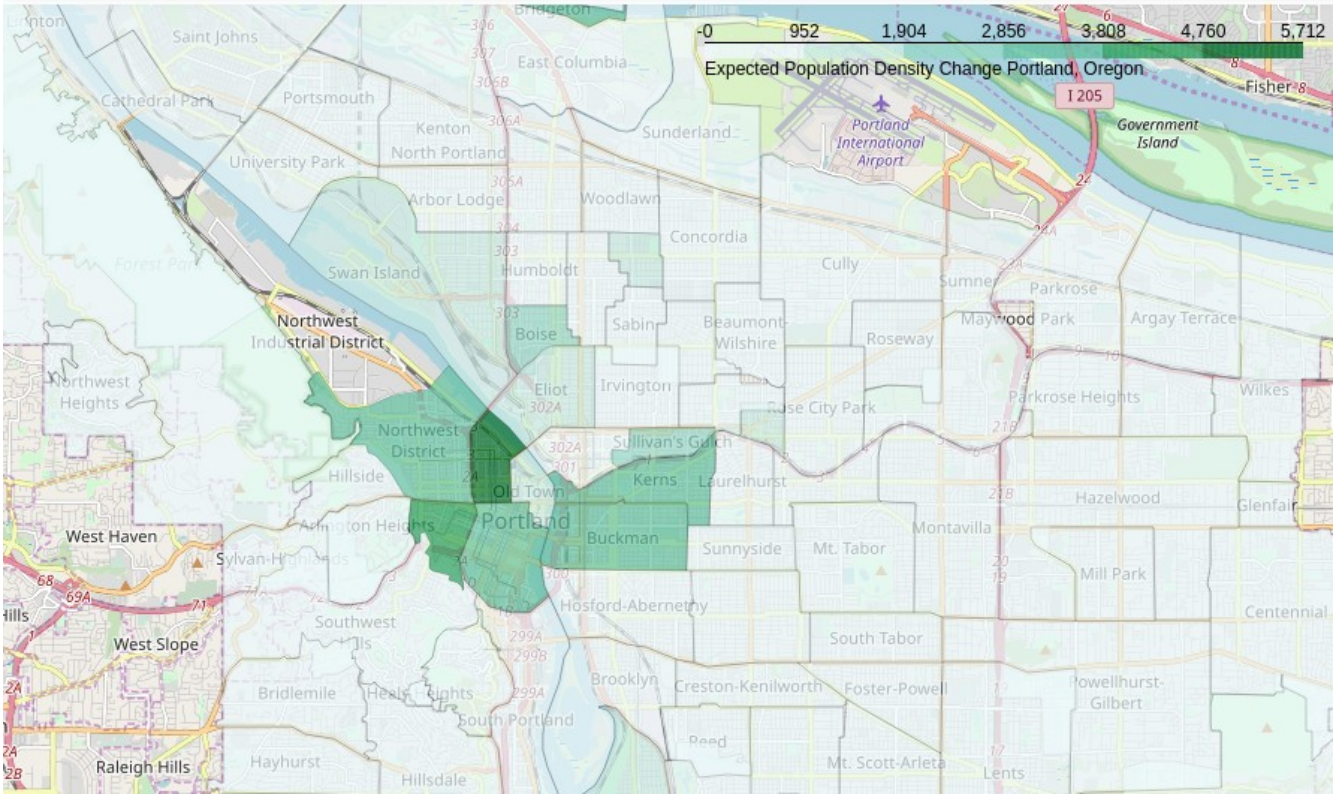
	Neighborhood	Adjusted population	Median age	Latitude	Longitude	Median household income	Walk score	New Units	5-year median price change	Adjusted population density
0	ALAMEDA	5641.0	41.0	45.548631	-122.636481	109616.0	65.0	30.0	26.0	9544.0
1	ARBOR LODGE	7245.0	36.0	45.571794	-122.690152	75223.0	72.0	443.0	33.0	8654.0
2	ARDENWALD-JOHNSON CREEK	2015.0	41.0	45.458516	-122.627539	95630.0	54.0	10.0	34.0	10273.0
3	ARGAY TERRACE	6035.0	40.0	45.552830	-122.523204	51990.0	45.0	58.0	52.0	6942.0
4	ARLINGTON HEIGHTS	1804.0	52.0	45.519496	-122.710667	113665.0	40.0	1.0	9.0	8934.0

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



Will Morgan Feb, 2019
IBM Coursea Capstone

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:

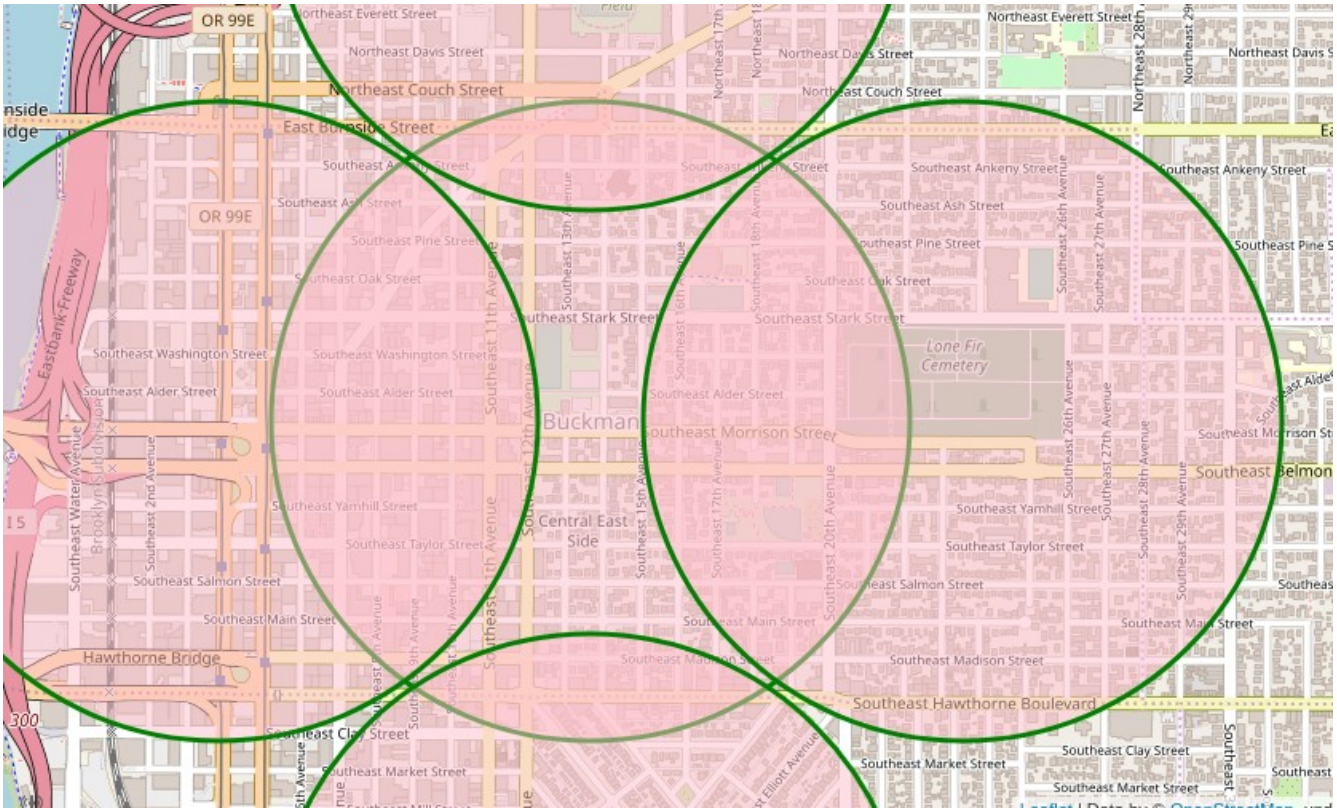


Units: people/sq.mile

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

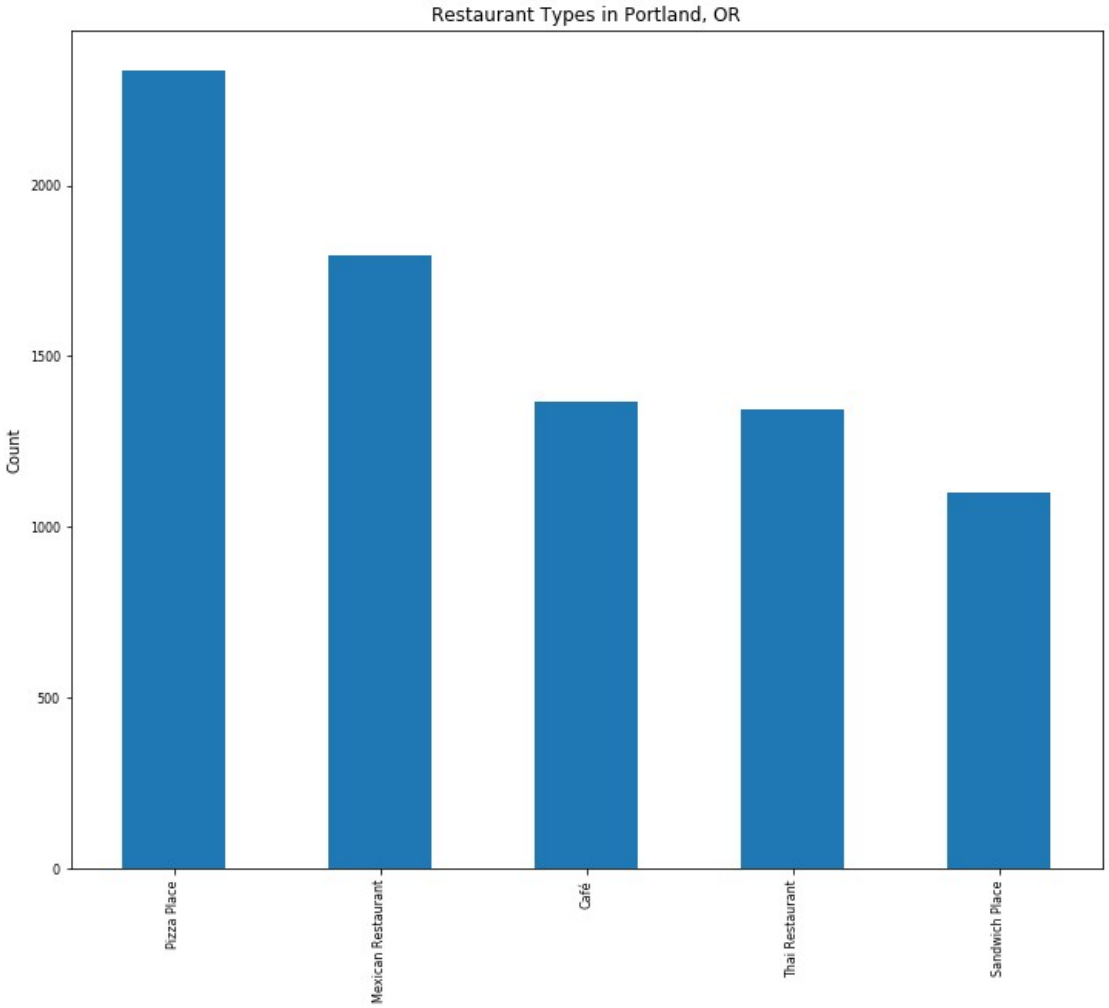
This was done for two reasons. First, it produces a more robust list of venues. Second, although some venues might technically be in other neighborhoods, they are still within walking distance, and should be taken into consideration when choosing a neighborhood to open a new restaurant.

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	ALAMEDA	Pizza Place	Café	Thai Restaurant	Mexican Restaurant	Bakery
1	ARBOR LODGE	Pizza Place	Mexican Restaurant	Thai Restaurant	Café	Chinese Restaurant
2	ARDENWALD-JOHNSON CREEK	Pizza Place	Café	American Restaurant	Thai Restaurant	Bakery
3	ARGAY TERRACE	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Food	Pizza Place
4	ARLINGTON HEIGHTS	Pizza Place	Café	Thai Restaurant	Sandwich Place	American Restaurant
5	ARNOLD CREEK	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Breakfast Spot
6	ASHCREEK	Pizza Place	Mexican Restaurant	Café	Mediterranean Restaurant	Fast Food Restaurant
7	BEAUMONT-WILSHIRE	Pizza Place	Café	Thai Restaurant	Mexican Restaurant	Restaurant
8	BOISE	Pizza Place	Café	Mexican Restaurant	Sandwich Place	Thai Restaurant
9	BRENTWOOD-DARLINGTON	Pizza Place	Mexican Restaurant	Chinese Restaurant	American Restaurant	Bakery
10	BRIDGETON	Fast Food Restaurant	American Restaurant	Sandwich Place	Seafood Restaurant	Breakfast Spot

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.

Will Morgan Feb, 2019
IBM Coursea Capstone



*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:

	Neighborhood	arts	hotels	university	highschool	nightlife	office
0	ALAMEDA	275	200	0	200	1800	900
1	ARBOR LODGE	100	500	0	100	2550	750
2	ARDENWALD-JOHNSON CREEK	225	0	0	0	675	700
3	ARGAY TERRACE	100	200	100	150	300	750
4	ARLINGTON HEIGHTS	575	200	0	0	1725	1200

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:

	Neighborhood	Adjusted population	Latitude	Longitude	Median household income	Walk score	add	subtract	others
0	ALAMEDA	5641.0	45.548631	-122.636481	109616.0	65.0	3375	-3000	-8650
1	ARBOR LODGE	7245.0	45.571794	-122.690152	75223.0	72.0	4000	-1800	-3650
2	ARDENWALD-JOHNSON CREEK	2015.0	45.458516	-122.627539	95630.0	54.0	1600	-1100	-3200
3	ARGAY TERRACE	6035.0	45.552830	-122.523204	51990.0	45.0	1600	-900	-2050
4	ARLINGTON HEIGHTS	1804.0	45.519496	-122.710667	113665.0	40.0	3700	-1800	-7250
5	ARNOLD CREEK	2963.0	45.446091	-122.689812	134282.0	17.0	500	-400	-850
6	ASHCREEK	4763.0	45.461163	-122.733365	92676.0	36.0	1500	-1100	-2250
7	BEAUMONT-WILSHIRE	6124.0	45.550391	-122.623694	112725.0	72.0	2925	-2500	-6450
8	BOISE	4430.0	45.550159	-122.671878	75505.0	92.0	8050	-2900	-7650
9	BRENTWOOD-DARLINGTON	12823.0	45.468707	-122.597633	59388.0	54.0	1000	-1700	-3300
10	BRIDGETON	71.0	45.602409	-122.668102	62263.0	30.0	2200	-400	-2150

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

$$((AdjustedPopulation + NewUnits) * .75) * MedianHouseHoldIncome)) + Contributions - (SameRestaurants + otherrestaurants) * WalkScore/1000000000$$

((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

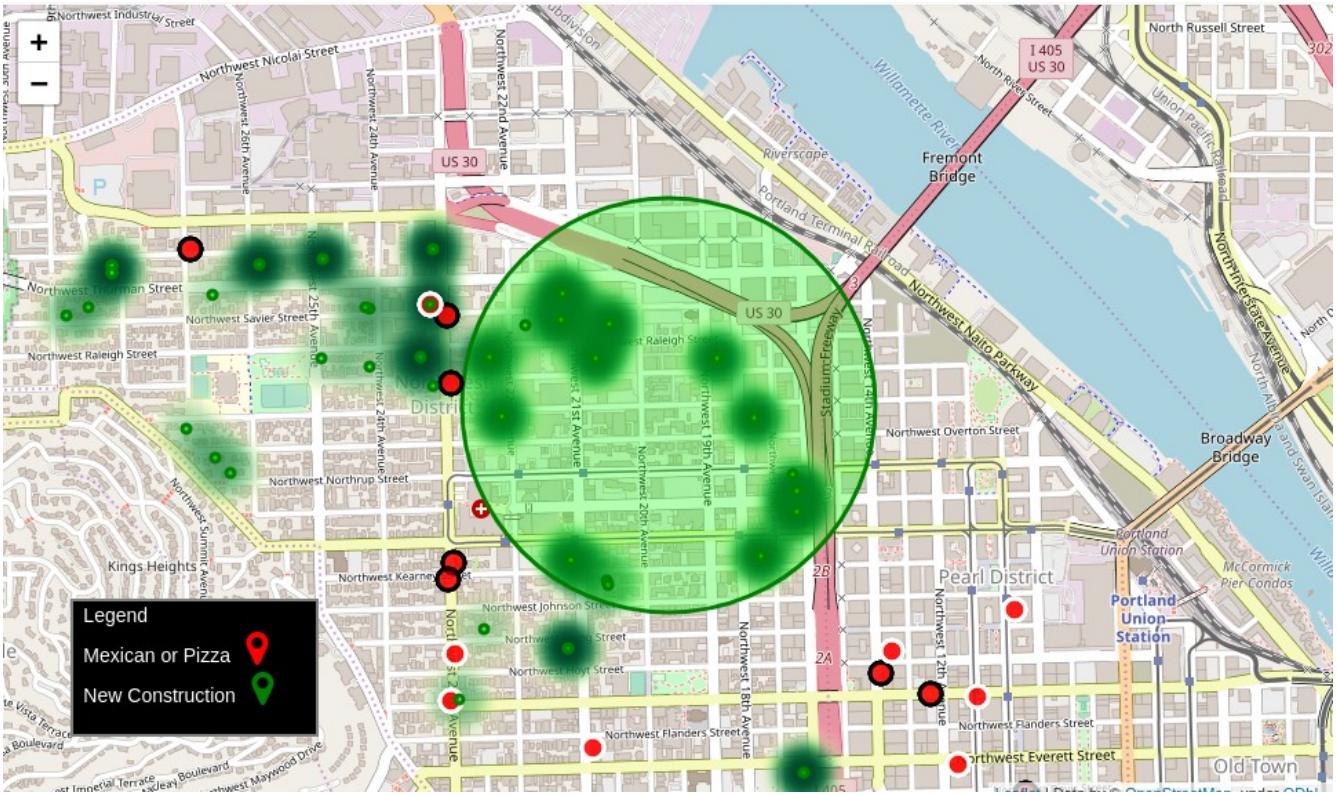
- An arbitrary to divisor to reduce the number into a digestible form.

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.”

	Neighborhood	total_score
57	NORTHWEST DISTRICT ASSOCIATION	95.551915
70	RICHMOND	68.000171
17	CONCORDIA	54.394694
31	HAZELWOOD	53.850306
52	MONTAVILLA	50.044402

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

    purpose: To analyze restaurants and other venues in each neighborhood

    method: API calls

Will Morgan Feb, 2019
IBM Coursea Capstone