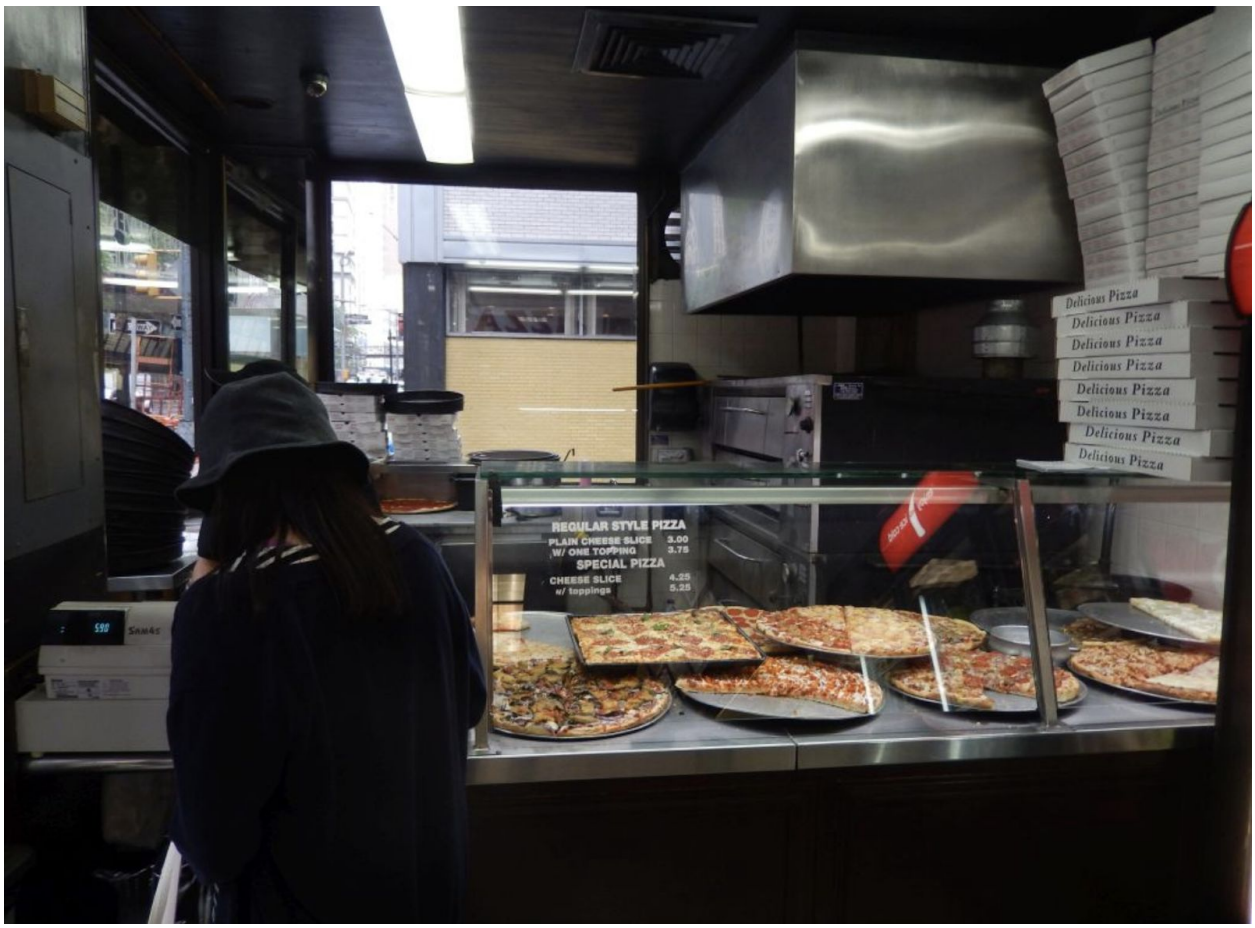
Applied Data Science Capstone by IBM/Coursera

The Battle of the Neighborhoods (Week 2)

YAPP (Yet Another Pizza Place)



Introduction

There are many aspects that can influence an investor's decision when it comes to selecting a location for a new restaurant, especially when the restaurant type is one of the most popular in a region, like for example Pizza Places in New England or the midwest of the US (Source). Parking space, Visibility, and other characteristics are very important but more obvious than other hidden features like population and the competitive landscape.

Being able to provide an idea of not only the number of similar businesses surrounding a targeted area but also know if these similar restaurants are part of the top 10 recommended restaurants would be very valuable for investors to know how strong the competitions could be and help them pick locations that would provide a higher potential for success.

We are selecting New Hampshire to focus this study because it is an area which I am familiar with and where I see interest by some small business investors that don't have a lot of access to technology when it comes to helping their decisions.

The Data

For this particular exercise we are planning to focus specifically on the state of New Hampshire and here are the different data sources we are planning to use for this project:

List of New Hampshire cities with their counties and zip codes

Source: https://www.zipcodestogo.com/New%20Hampshire

This data tells us all the different cities in New Hampshire their zip codes and counties. The city and zip code will be used to later get the location if the cities

New Hampshire population data by city:

Source: https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_New_Hampshire

This data would add the total population of each city that can be later used for determining the target customers for each potential location and determining its relevance

Latitude and longitude location data for cities in Nw Hampshire

Gathered by using the geopy which is a Python client for several popular geocoding web services, and the service we will be using is Nominatim. This data will allow us not only to map the cities for visualization but also to gather a list of popular businesses surrounding an area or location

FourSquare API: https://api.foursquare.com

We will use the FourSquare API to get data about the popular businesses surrounding a particular location. Part of the input for gathering this data will be the latitude and longitude gathered with geopy.

Methodology

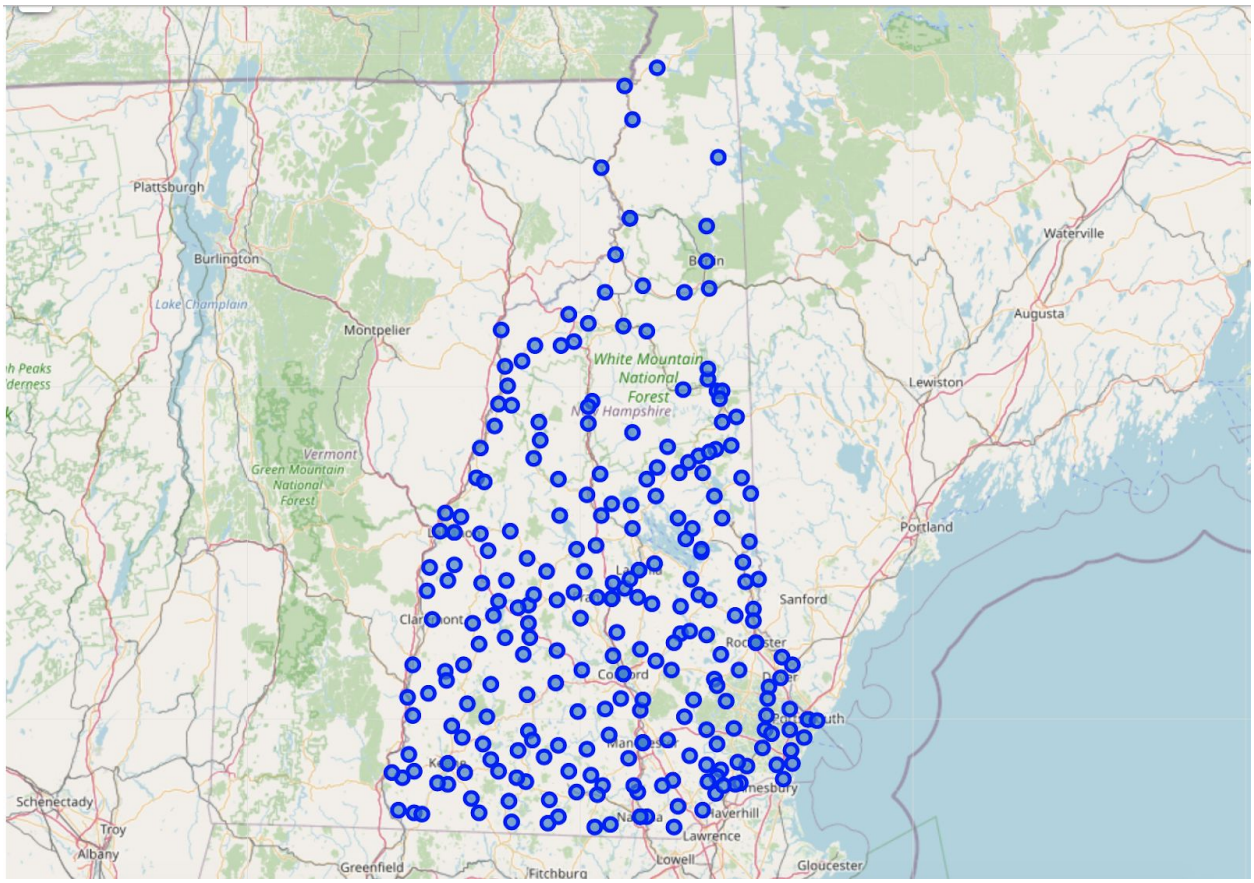
In this project, we will focus on detecting areas of the New Hampshire Sea Coast that have a variety of towns and cities with their population currently growing due to the increasing migration from the border state of Massachusetts of people mostly looking for lower property taxes and a more quiet way of life, and accompanying this growth, is a massive and increased on-demand for restaurants, especially fast-food restaurants, being the Pizza Places one of the most popular styles of food.

The popularity of pizza places and the proximity between towns and cities make it very difficult for investors to find the right locations for starting a new business of such kind without having many competitors around with a loyal fan base. We would do the analysis taking Rockingham County as the area of study for this project due to the fact that this county groups most of the towns and cities of the Sea Coast receiving the majority of the migration from south of the state.

The first step was collecting the required data, starting with the New Hampshire towns and cities data with the corresponding zip codes. Secondly collecting the New Hampshire population by city and county. Once this data is on their corresponding data frames we used it to find the latitude and longitude coordinates for each town and city using the Nominatim provider via geopy.

	Zip Code	City	County	Latitude	Longitude
0	03031	Amherst	Hillsborough	42.864287	-71.625248
1	03032	Auburn	Rockingham	43.004529	-71.348398
2	03033	Brookline	Hillsborough	42.734810	-71.658127
3	03034	Candia	Rockingham	43.077916	-71.277237
4	03036	Chester	Rockingham	42.956753	-71.257285

Once all this data was loaded into the corresponding Pandas data frames we proceed to map the locations using the Folium python libraries in order to get familiar with the location of the towns and cities on the New Hampshire map.

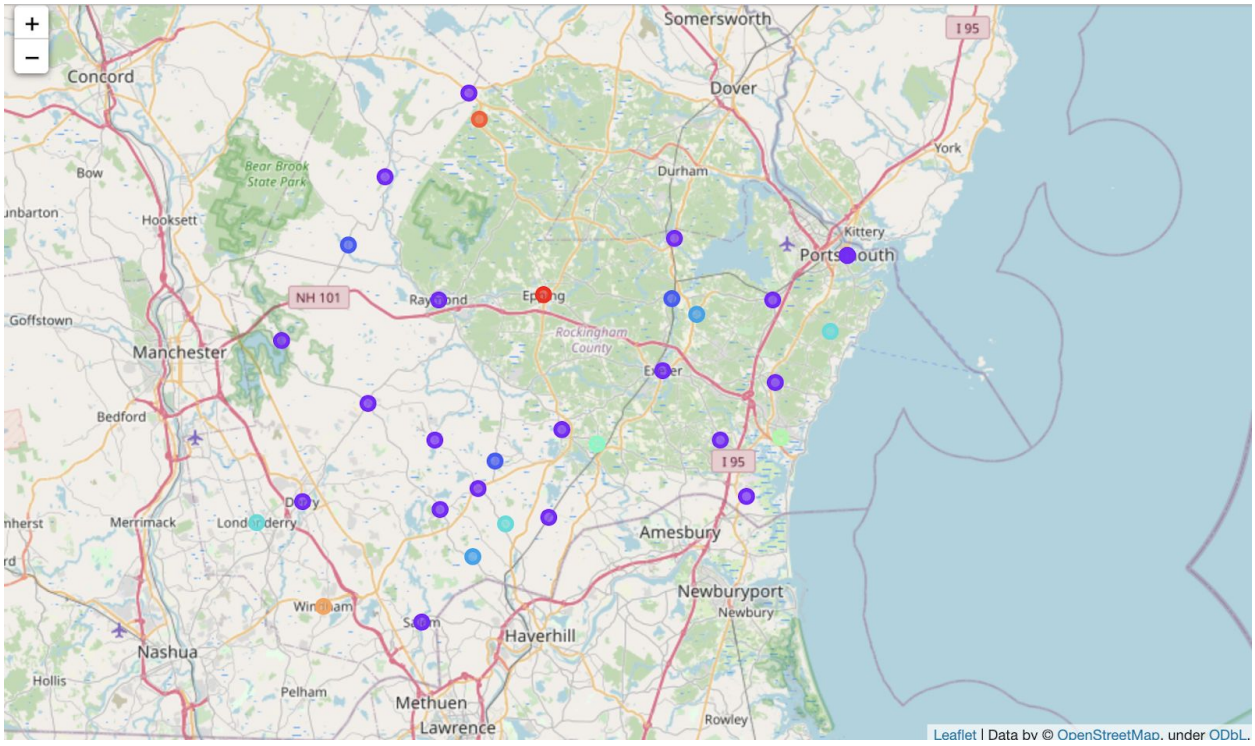


Also, as part of the first step, we extracted the data for Rockingham county into separate data frames.

The 2nd step consisted of using foursquare to obtain the data for exploring the venues in the cities and towns of Rockingham county. Out of the venues data, we found the recommended most popular venues and counted the top 10 most common types of venues for each town and city for Rockingham county.

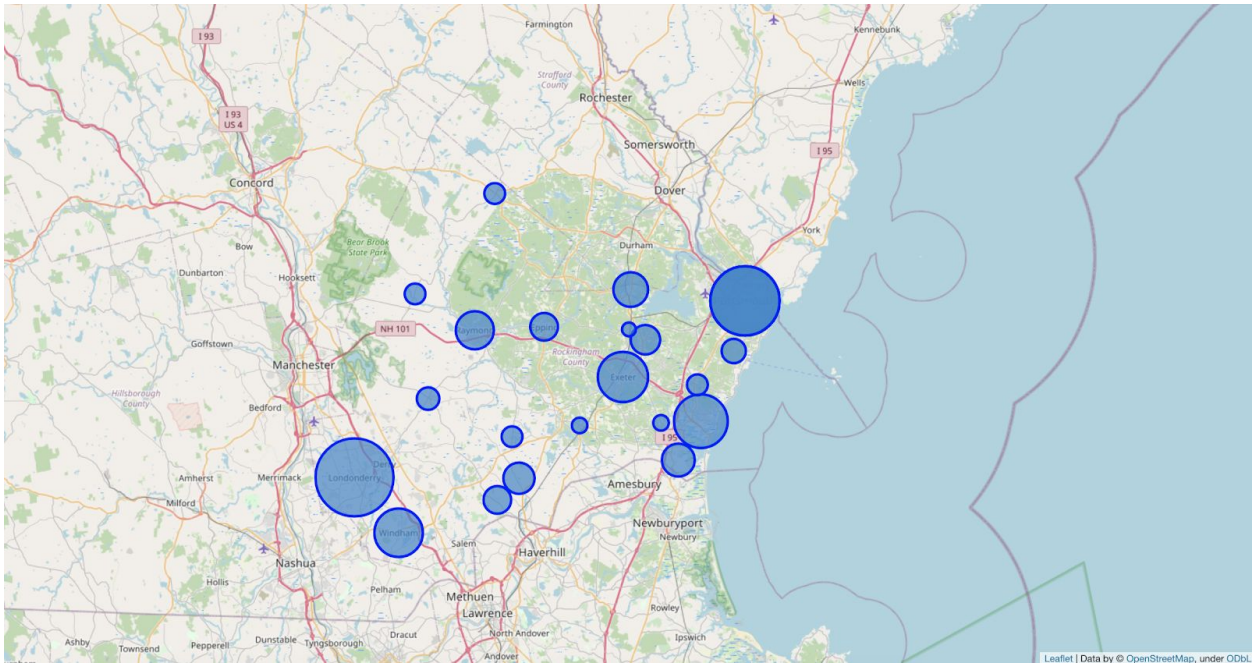
	Zip Code	City	County	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	03032	Auburn	Rockingham	43.004529	-71.348398	1.0	Pizza Place	Post Office	Construction & Landscaping	Supermarket	Deli / Bodega	Cheese Shop	Chinese Restaurant	Coffee Shop	Comic Shop	Convenience Store
1	03034	Candia	Rockingham	43.077916	-71.277237	1.0	Construction & Landscaping	Deli / Bodega	Wine Shop	Candy Store	Chinese Restaurant	Coffee Shop	Comic Shop	Convenience Store	Dance Studio	Diner
2	03036	Chester	Rockingham	42.956753	-71.257285	1.0	Diner	Food & Drink Shop	Ice Cream Shop	Wine Shop	Chinese Restaurant	Coffee Shop	Comic Shop	Construction & Landscaping	Convenience Store	Dance Studio
3	03037	Deerfield	Rockingham	43.130171	-71.238819	1.0	Pizza Place	Post Office	Restaurant	Bookstore	Dance Studio	Cheese Shop	Chinese Restaurant	Coffee Shop	Comic Shop	Construction & Landscaping
4	03038	Derry	Rockingham	42.880868	-71.326449	1.0	Pharmacy	Café	Italian Restaurant	Pizza Place	Video Store	Convenience Store	Salon / Barbershop	Diner	Bar	Park

Then for the 3rd step, we run k-means clustering with 5 clusters in order to group the cities based on their top 10 popular venues. With this data, we proceeded to present on a folium based map, and later explore each cluster looking for the ones with the least number of pizza places within their top 10 recommended places. At this point we found the population of the top locations to complement the previous data set and choose a set of optimal locations to be presented to the main stakeholders for further exploration and potential investment.



Results

After adding the population to the list of towns and cities with no pizza places within their top 10 recommended places wee proceed to map on a folium map increasing the size of the circle marker based on the population of each location we can see help visualize which cities present the best market opportunity for our the potential investors:



Discussion

Considering the previous data and visualization, we think that at list the top 10 towns and cities in our final list, which is sorted in descending order by population and excludes any city in which the top 10 recommended restaurants contain a pizza place, is a good reference and starting point for a potential investor interested on starting a pizza place restaurant in this New Hampshire Rockingham county. With this list, they can support their decision process and enhance their probability for success:

Zip Code	City	Latitude	Longitude	Population
03053	Londonderry	42.865087	-71.373953	24129
03801	Portsmouth	43.070222	-70.754862	20779
03802	Portsmouth	43.070222	-70.754862	20779
03804	Portsmouth	43.070222	-70.754862	20779
03842	Hampton	42.930438	-70.824896	15430
03833	Exeter	42.981477	-70.947834	14306
03087	Windham	42.800644	-71.304229	13592
03077	Raymond	43.036196	-71.183395	10138
03857	Newmarket	43.082931	-70.935974	8936
03874	Seabrook	42.884663	-70.860196	8693
03865	Plaistow	42.864094	-71.113350	7609
03885	Stratham	43.024413	-70.912281	7255
03811	Atkinson	42.838422	-71.147004	6751
03042	Epping	43.039446	-71.073025	6411
03870	Rye	43.012019	-70.771943	5298
03036	Chester	42.956753	-71.257285	4768
03819	Danville	42.912588	-71.124504	4387
03862	North Hampton	42.972589	-70.829776	4301
03261	Northwood	43.194664	-71.151581	4241
03034	Candia	43.077916	-71.277237	3909

Conclusion

The results of this project would add a lot of value for investors in the restaurant business by helping them focus on areas on which they would evaluate their opportunities, saving them time and money, having a closet idea of how big their competition could be.

The data on this study could be easily adapted to other regions and locations and other types of restaurants or even other types of businesses, by just performing minor changes on the code or the data sources.