Capstone Project

Machine learning for identification of commercial opportunities in Lincoln, Nebraska

Kevin Cheek

June 2019

Agenda

Introduction
Business Problem
Data
Methodology
Results
Conclusion/Discussion
References

Introduction

About Lincoln

- Capitol of Nebraska
- University of Nebraska at Lincoln (26k students)
- Population 287k
- Area of 96 square miles

Using FourSquare data, can we identify commercial development opportunities within the city of Lincoln?

Can we leverage this information to assist various stakeholders in directing development resources in an efficient way?

Business Problem

In this scenario, an entrepreneur has requested help with determining where to open a new Chinese restaurant in Lincoln.

We will be using machine learning principles to analyze FourSquare data. We will then apply the insights we gain to narrow the search and recommend specific areas of the city on which to focus.

Our target audience in this scenario is a specific entrepreneur with a specific need. However, the data insights could be used to inform a variety of other entities including commercial real estate brokers, developers, and various government agencies.

Data

Latitude and Longitude coordinates of a suitable city center for Lincoln, Nebraska

Coordinates of the city center will be obtained by converting an address to coordinates using the Nominatim geocoding tool

Latitude and Longitude of all points in a grid overlay of the city

The grid overlay and associated Latitude and Longitude coordinates will be calculated using Python.

Venue data from FourSquare for each region in the grid

Venue data will be retrieved from the FourSquare places api and will leverage the explore endpoint

Methodology

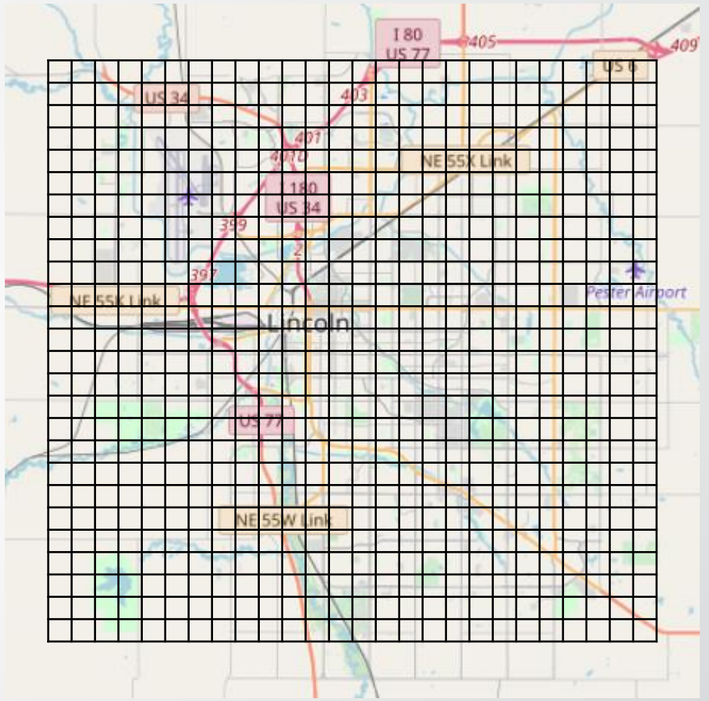
First, we find a point in the city which represents an approximate geographic center of the populated areas.

Next, we define a grid of ½ mile x ½ mile squares to overlay the city.

Using the grid overlay coordinate points, we retrieve FourSquare data, including up to 100 total venues, for each of these squares.

By total number of venues across the city, Chinese comes in as one of the top 15 types. It has the 6th highest total for food-related venue types.

We can be fairly confident from this data that Chinese food is popular in Lincoln and that we should dive deeper into the data.



Fast Food	169.0
Sandwiches	140.0
Pizza	128.0
Convenience Store	110.0
Mexican	87.0
Coffee Shop	87.0
Park	86.0
Bar	83.0
American	77.0
Hotel	70.0
Pharmacy	67.0
Gym / Fitness	64.0
Grocery Store	58.0
Chinese	57.0
Gas Station	54.0
Construction	48.0
Spa	44.0
Gym	42.0
Burgers	40.0
Ice Cream	39.0

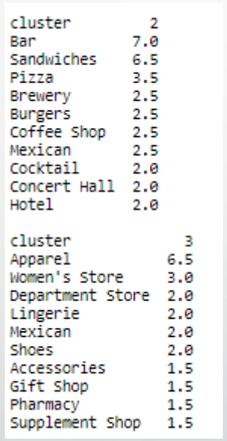
Methodology cont.

Using KMeans clustering, we group the grid squares into 6 different clusters based on which venues they contain.
The number of clusters allows us to optimize the differentiation between the clusters but still be able to easily identify different usage patterns within each cluster.

Analysis of the cluster data makes it clear which are residential and which are commercial.
We can use this data to eliminate the residential areas from consideration.
We can define the remaining clusters by their most prevalent venue types and overall venue density.

cluster	0
Fast Food	1.524590
Sandwiches	0.819672
Pizza	0.721311
American	0.606557
Pharmacy	0.573770
Hotel	0.557377
Convenience Store	0.540984
Chinese	0.491803
Mexican	0.491803
Grocery Store	0.442623

cluster	1
Park	0.068
Bar	0.066
Fast Food	0.056
Baseball Field	0.054
Pizza	0.054
Gym / Fitness	0.048
Convenience Store	0.046
Golf Course	0.046
Lake	0.046
Sandwiches	0.038



cluster	2
Bar	7.0
Sandwiches	6.5
Pizza	3.5
Brewery	2.5
Burgers	2.5
Coffee Shop	2.5
Mexican	2.5
Cocktail	2.0
Concert Hall	2.0
Hotel	2.0

cluster	3
Apparel	6.5
Women's Store	3.0
Department Store	2.0
Lingerie	2.0
Mexican	2.0
Shoes	2.0
Accessories	1.5
Gift Shop	1.5
Pharmacy	1.5
Supplement Shop	1.5

cluster	4
Sandwiches	2.500000
Coffee Shop	2.250000
Fast Food	1.666667
Mexican	1.500000
Pizza	1.083333
American	1.000000
Bar	1.000000
Hotel	1.000000
Mobile Phones	1.000000
Grocery Store	0.666667

cluster	5
Convenience Store	0.484848
Pizza	0.373737
Sandwiches	0.262626
Park	0.252525
Coffee Shop	0.242424
Fast Food	0.242424
Mexican	0.202020
American	0.151515
Spa	0.151515
Gas Station	0.141414

Cluster 0:
Tier 3 commercial
Low-Medium venue density
Primarily Food and Service venues

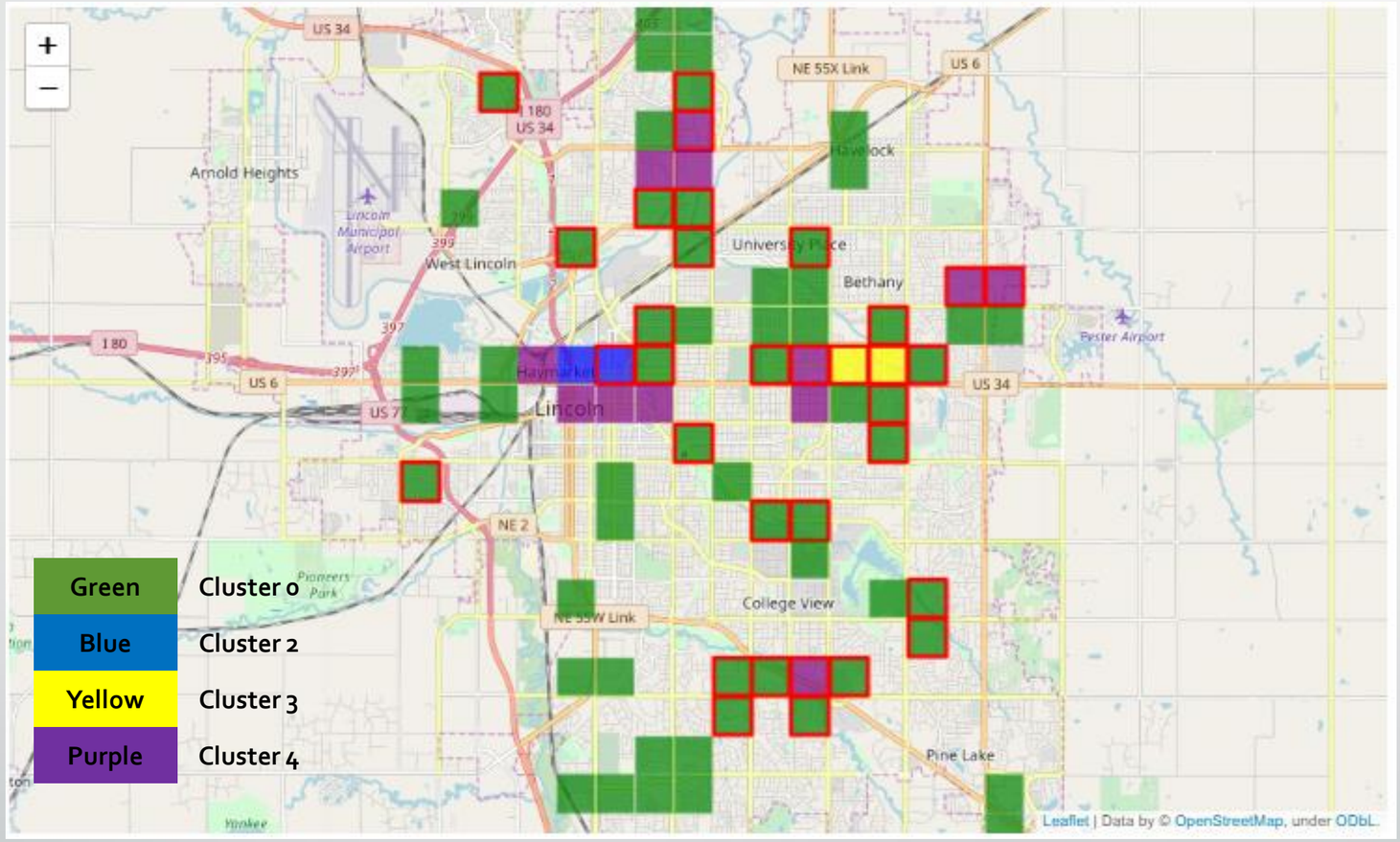
Cluster 2:
Tier 1 commercial
High venue density
Primarily Bars/Food/Entertainment

Cluster 3:
Tier 1 commercial
High venue density
Primarily Shopping/Food

Cluster 4:
Tier 2 commercial
Medium venue density
Primarily Food

Methodology cont.

Having defined the clusters, we can generate a map showing the plot squares to be considered.

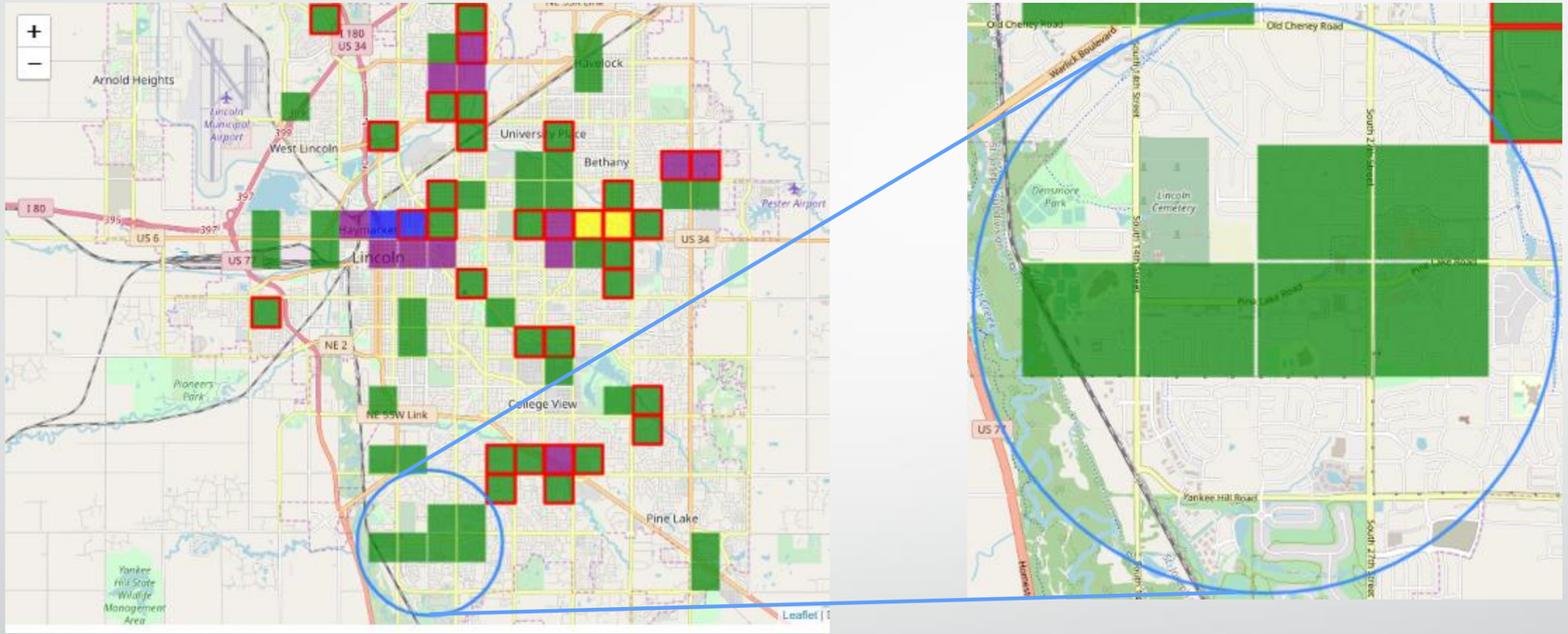


The squares are color coded by cluster.

Squares with an existing Chinese venue are given a red border.

Results

Several areas of Lincoln should be able to support a new Chinese restaurant. The majority of the viable regions either already have a Chinese venue or are located in close proximity to another region which already has one. From our map, it appears the southwest corner of the city has the lowest saturation of Chinese venues.



The highlighted area has a promising cluster of Tier 3 commercial areas without any existing Chinese venues. The area is surrounded by dense residential areas and includes a mid-sized shopping mall, grocery, and several fast food venues. It is also located at a crossroads of 2 major north-south roads and 1 major east-west road and has no existing Chinese venues.

Conclusion/Discussion

We were able to use FourSquare data to classify the commercial regions of the city and compare their venue makeup to identify possible commercial development opportunities.

We were also able to apply our insights to a specific scenario and provide our entrepreneur with a specific area on which to focus time and resources.

This project was successful but we could improve the quality of our recommendation by including additional information in our model. This could include income data, rent prices, and zoning data among others.

References

Lincoln, NE statistics
https://en.wikipedia.org/wiki/Lincoln,_Nebraska

FourSquare api
https://developer.foursquare.com/docs/api

Nominatim geocoder
https://wiki.openstreetmap.org/wiki/Nominatim

Thank You