Where is Papa John Moving To?

By David Wedlake

July 1st, 2020

# Introduction

## Background

I love pizza! I like deep dish, New York style, thin crust and even Pizza rolls. COVID-19 has changed the way I work and live. I am one of the Americans that is now working from home full time and I don’t need to live near my job. With the new freedom to live anywhere in the US, I can now move to any city with a good internet connection and enjoy as much pizza as I want. My goal for this move is to find a medium sized city with the most Pizza places per capita. I would also like to find a starting point to find my new home with the most densely populated area so I can walk more.

## Problem

How do I find the best city to enjoy my love of Pizza? Is there a way to find the best city in the US with as many pizza joints as possible? I would love to eat pizza every day but what if everyone else wants to as well. I need to find a city where we can all enjoy pizza without crowding in on each other.

# Data Acquisition and Cleaning

## Data Sources

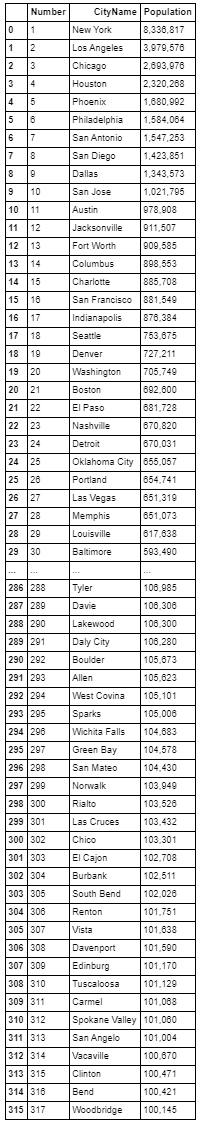
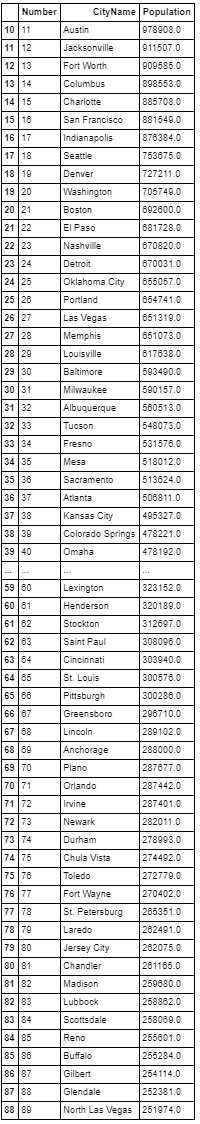
The Data for my project is in two sources. The first is Wikipedia for the cities in the US. It contains a list of the largest 300 cities. My second source will be FourSquare. It is one of the largest databases of location data and it will be a great source of Pizza shops.

## Data Cleaning

The data cleaning is in two phases. The first phase will import the US cities and then reduce the list to medium size cities. I have defined a medium city as one with a population between 250,000 and 1 million people. After I have a short list of cities, I can start the next phase of data cleaning. I will use the FourSquare data to find the number of pizza places within 5 km of city center. After I have a total count of pizza locations per city, I can calculate the per Capita number of Pizza locations.

## Feature Selection

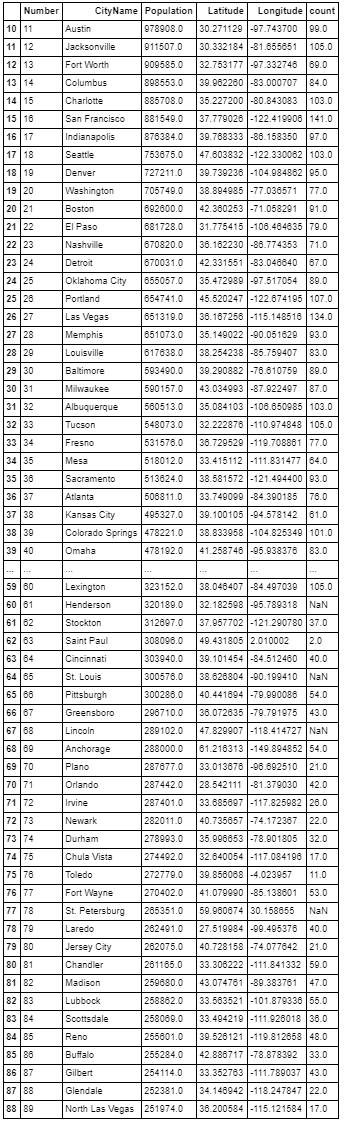
### Reduce the Number of City to Search



Remove cities less than 250 Thousand

Remove cities greater than 1 Million

### Add the Number of Locations to the Feature Set and Calculate per Capita



Divide the Count of Pizza Locations by the Population

# Exploratory Data Analysis

Now that I have a list of cities and the per capita Pizza shops in each city, I selected my new hometown. Based on the Population data and FourSquare data, my new hometown is Lexington KY. I used the location data from FourSquare to plot all the pizza shops in Lexington.

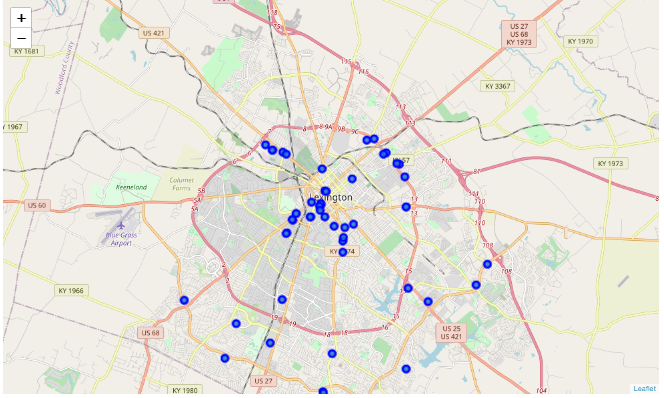


Figure -All Pizza Locations

# Regression models

To find the best location for my new house, I will use the K Mean regression method to find five groupings of restaurants. These groupings will allow me to find the area of two with the most pizza shops. After I have clustered the locations together, I can use the K Mean centroid function to find a location with the most eateries around it.

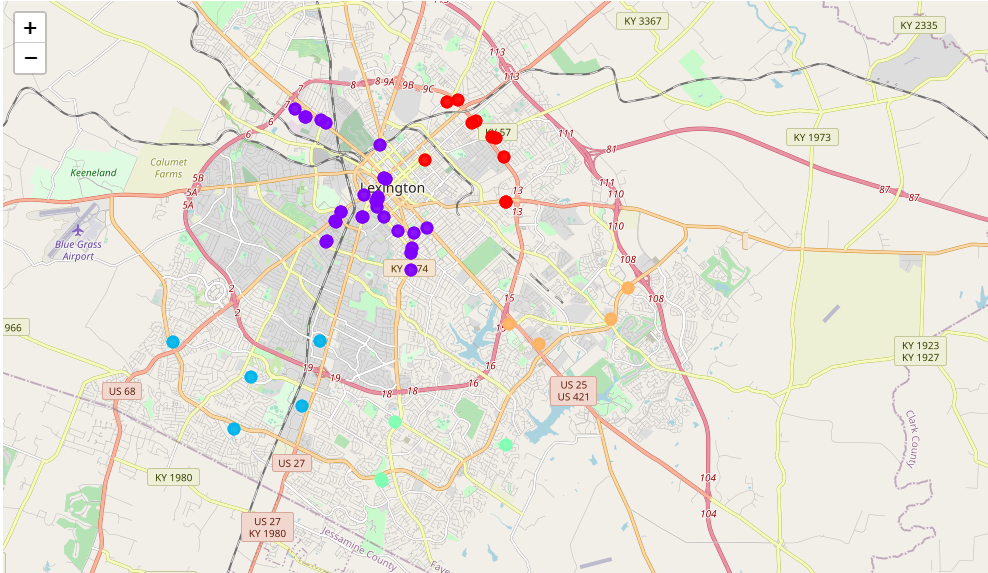
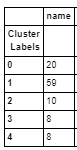


Figure - Cluster Location Count

Figure - Clustered Pizza Locations



# Conclusion

Based on the population data, FourSquare restaurant information, and the K Means regression model, I will be moving to Lexington KY near Center Court Condos (535 S. Upper St. Lexington KY 40508) and I will have 59 pizza locations around my new house.



# Appendix

## Github Code Link

<https://github.com/Maxizarius/datasciencecoursera/blob/master/The%20Battle%20of%20Neighborhoods.ipynb>