Applied Data Science

Capstone Project:

Moving to New York City

Tom Carminati

Introduction

A. Business Problem

You're planning a move to a new city (let's say NYC) and you'd like to find the neighborhoods in the new city that are most similar to your current neighborhood or one you've visited or lived in in the past

B. Stakeholders

- *2 sets of stakeholders- essentially looking at the same problem from different perspectives:*
- *The end user - This is someone thinking of moving to NYC and would like guidance on which neighborhoods to investigate*
- *The sponsor – Chamber of commerce, local realty board or some other agency looking to encourage people to move to NYC*

C. Background

Similar to the issue my daughter faced when she had to give up her apartment but wanted to find another in a similar neighborhood

Data, Tools and Methodology

Data

A. Data required

- *NYC neighborhood data - json file used in the NYC neighborhood lab*
- *foursquare data for existing zip code (or neighborhood) & target NYC neighborhoods*

B. Tools

- *Use the geopy library to convert addresses to latitude & longtitude*
- *folium.Map for generating maps*

Methodology

A. Exploratory data analysis

Primary data is:

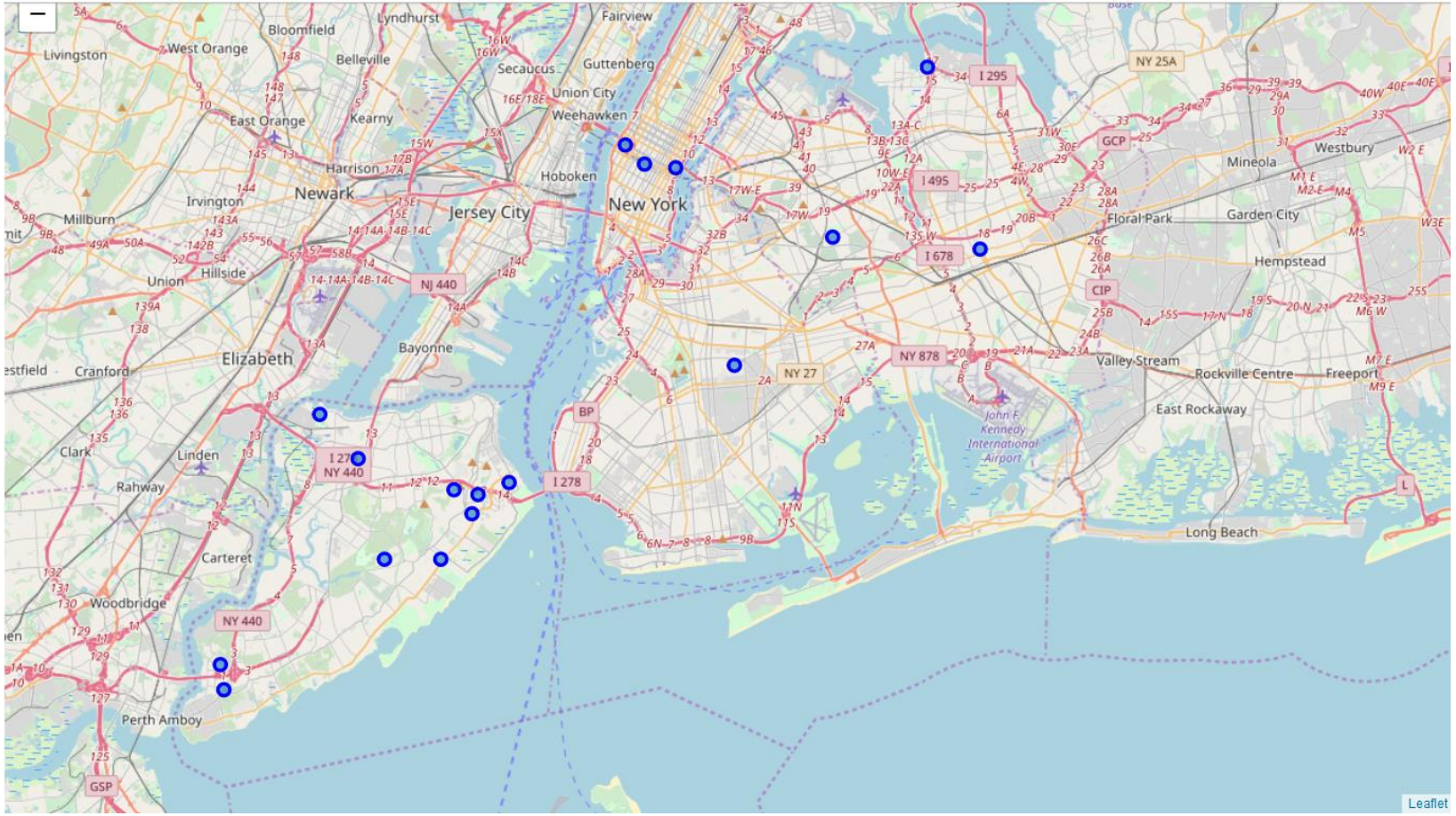
- NYC borough & neighborhood latitude, longitude & boundaries from .json file - for mapping & across to foursquare data
- latitude & longitude for the "Moving From" neighborhood
- foursquare venue data for each neighborhood

B. Inferential statistical testing

- *K-means analysis to determine clusters of similar neighborhoods based on the venue data from foursquare This is used to find the neighborhoods most similar to the "Moving From" neighborhood*

Map of NYC neighborhoods in same cluster as “Moving from” neighborhood

- i.e., similar neighborhoods based on foursquare venue data



Extract of similar neighborhoods

- *from K-means clustering with foursquare venue data*

(not all neighborhoods shown)

Borough	Neighborhood	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
Staten Island	Port Ivory	Bar	Wings Joint	Frozen Yogurt Shop	French Restaurant	Food Truck	Food & Drink Shop	Food	Fish Market	Fish & Chips Shop	Field
Brooklyn	Wingate	Fast Food Restaurant	BBQ Joint	Gym / Fitness Center	Discount Store	Juice Bar	Donut Shop	Fish & Chips Shop	Field	Pharmacy	Food Truck
Staten Island	Graniteville	Food Truck	Bus Stop	Grocery Store	Wings Joint	Fried Chicken Joint	Food & Drink Shop	Food	Fish Market	Fish & Chips Shop	Field
Staten Island	Old Town	Italian Restaurant	Pharmacy	Bakery	American Restaurant	Gas Station	Optical Shop	Pizza Place	Food & Drink Shop	Food	Fish Market
Staten Island	Grant City	Wings Joint	Pizza Place	Food & Drink Shop	Tanning Salon	Fast Food Restaurant	Grocery Store	Health & Beauty Service	Arts & Crafts Store	Event Space	Dessert Shop
Staten Island	Charleston	Big Box Store	Furniture / Home Store	Grocery Store	Cosmetics Shop	Restaurant	Arts & Crafts Store	Pizza Place	Gift Shop	Bakery	Food & Drink Shop
Manhattan	Midtown South	Lingerie Store	Street Food Gathering	Korean Restaurant	Food Truck	Grocery Store	Clothing Store	Cosmetics Shop	Dessert Shop	Italian Restaurant	Discount Store
Staten Island	Lighthouse Hill	Spa	Moving Target	Italian Restaurant	Café	Art Museum	Trail	Massage Studio	Wings Joint	Falafel Restaurant	Food
Staten Island	Richmond Valley	Bank	Deli / Bodega	Coffee Shop	Mexican Restaurant	Smoothie Shop	Fast Food Restaurant	Train Station	Sandwich Place	Convenience Store	Dog Run
Queens	Malba	Scenic Lookout	Rest Area	Rock Club	Tennis Court	Wings Joint	Event Space	Food	Fish Market	Fish & Chips Shop	Field
Manhattan	Hudson Yards	American Restaurant	Food Truck	Theater	Hotel	Residential Building (Apartment / Condo)	Supermarket	Music School	Gym / Fitness Center	Park	Fish Market
Moving From	60640	Breakfast Spot	Jazz Club	Ethiopian Restaurant	Gay Bar	Hot Dog Joint	Massage Studio	Hotel	Arcade	American Restaurant	Food

What I found out

Results

- *I was having trouble getting foursquare to return the venue data for all 307 neighborhoods. Since this is just a demo program, I tailed the list of neighborhoods to 100.*
- *Using 100 neighborhoods and creating 10 K-means clusters - the cluster containing the "Moving From" neighborhood had 17 target neighborhoods*
- *Since the "Moving From" neighborhood I used in the test has a suburban feel, I wasn't surprised that most of the target neighborhoods are in outlying boroughs, especially Staten Island. But 3 of the target neighborhoods are in Manhattan.*

Discussion

A. Observations

- *K-means analysis on foursquare venue data does not provide what are to me intuitive groupings of the data points.*
- *It's not like a K-means analysis on a 2-dimension graph of data points, where the groupings make sense visually.*

B. Recommendations

- *If this were a real commercial tool there would need to be some changes.*
- *1. The NYC data could be pre-loaded. Maybe a .CSV file that's read in at start-up.*
- *2. The "Moving from" neighborhood would be entered as field on a website.*
- *3. The foursquare venue data would be needed for all 307 neighborhoods. I was using a Sandbox version of foursquare, and this may have been a factor.*

Conclusion

- *For me, this was a meaningful and useful extension of the labs I completed for this course.*
- *It allowed me to answer a question I posed (admittedly, in demo form) and gave me a feel for what I could potentially do.*
- *It also showed me some of the limitations of my skills and the demo/sandbox implementation.*