

Data Science

Peng Yang

Course Title

May 2, 2020

Which City is nice for a Nerdy Chinese Programmer after COVID -19?

Introduction

Background

Let's say you have never been to the US and you want to have only Chinese food while you are there. So you want to go to a place with a high density of Chinese restaurant around you.

Problem

The problem we aim to solve is to analyze the Chinese restaurants' locations in the major US cities. And also it is safer for Chinese are careful for COVID-19 because they all have face masks .

Intrest

We want to know which city have highest density of Chinese restaurants.

Data acquisition

We use Foursquare APIs to get LBS data.

Data API

The Places API offers real-time access to Foursquare's global database of rich venue data and user content to power your location-based experiences in your app or website.

Key Features

Feature	Description
---------	-------------

Access to Foursquare's Global Database	Get real-time access to over 105MM places available across 190 countries and 50 territories.
Power App Experiences	Use our custom API endpoints to power geo-tagging, venue search, venue recommendations, and more in your apps.
Descriptive Place Profiles	Leverage 70+ venue attributes and 900+ categories, sourced by the Foursquare consumer community.
Rich User Content	Create engaging location experiences with access to user-generated tips, tastes, photos & more.

Category

Since we only want to know the Chinese Restaurant. We need to limit the "category" param. If you set the category ID, the api will only return a hierarchical list of categories applied to venues. Categories that specify a list of "Supported Countries" can only appear on venues in those countries. All other categories are supported globally by default. The Category has 3 levels like a tree:

Food 4d4b7105d754a06374d81259		
	Afghan Restaurant 503288ae91d4c4b30a586d67	
	African Restaurant 4bf58dd8d48988d1c8941735	
	Ethiopian Restaurant 4bf58dd8d48988d10a941735	
	American Restaurant 4bf58dd8d48988d14e941735	
	New American Restaurant 4bf58dd8d48988d157941735	
	Asian Restaurant 4bf58dd8d48988d142941735	

	Burmese Restaurant 56aa371be4b08b9a8d573568	
	Cambodian Restaurant 52e81612bcb57f1066b7a03	
	Chinese Restaurant 4bf58dd8d48988d145941735	
		Anhui Restaurant 52af3a5e3cf9994f4e043bea Supported countries: CNHK-MOMYTWSG
		Beijing Restaurant 52af3a723cf9994f4e043bec Supported countries: CNHK-MOMYTWSGJP
		Chinese Breakfast Place 52af3a903cf9994f4e043bee Supported countries: CNHK-MOMYTWSG
		Taiwanese Restaurant 52af3b813cf9994f4e043c04 Supported countries: CNHK-MOMYTWSGUSCAJPAU
	

Exploratory Data Analysis

My main target here is to assess which city would have the highest Chinese food density. I used the Four Square API through the venues channel. I used the near query to get venues in the cities. Also, I use the CategoryID:4bf58dd8d48988d145941735 to set it to show only Chinese restaurant. An Example of my requests:

https://api.foursquare.com/v2/venues/explore&client_id=&client_secret=&v=20200401&New York,NY&limit=200&categoryId=4bf58dd8d48988d145941735

Data Science

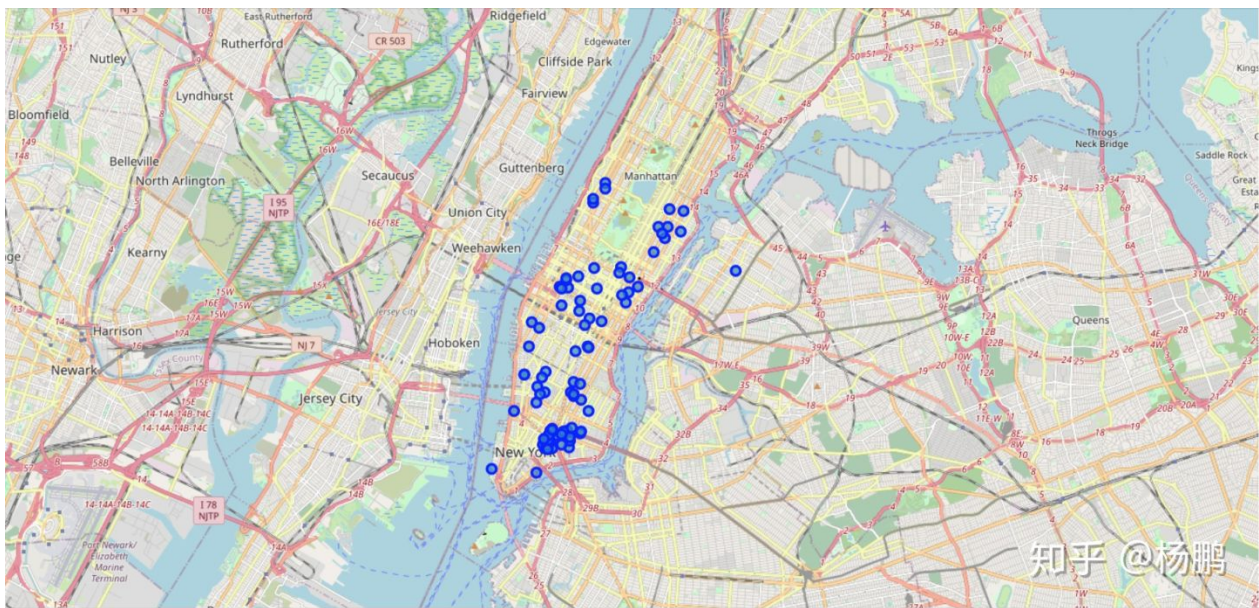
Moreover, I repeated this request for the 5 studied cities and got their top 200 venues. I saved the name and coordinate data only from the result and plotted them on the map for visual inspection.

Next, to get an indicator of the density of Chinese Restaurants, I calculated a center coordinate of the venues to get the mean longitude and latitude values. Then I calculated the mean of the Euclidean distance from each venue to the mean coordinates. That was my indicator; mean distance to the mean coordinate.

Result

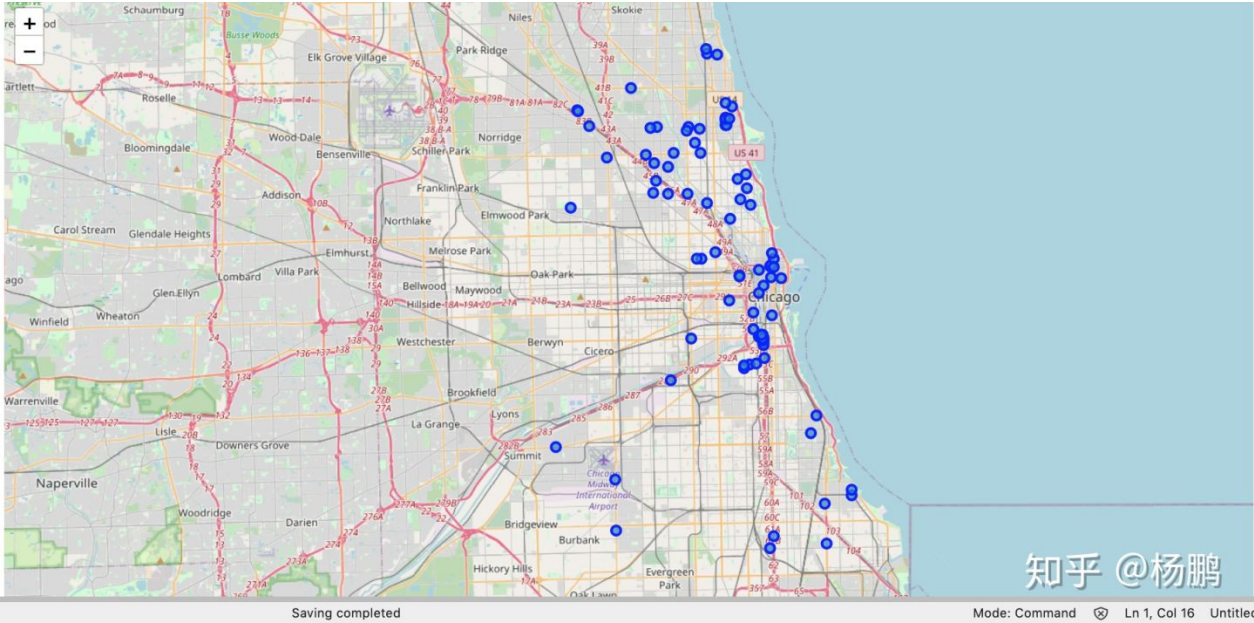
For our initial visual inspection we see that they all have multiple Chinese restaurants and often more than Foursquare would like to supply us. The following here are the pictures of the geoplot generated with folium:

New York:

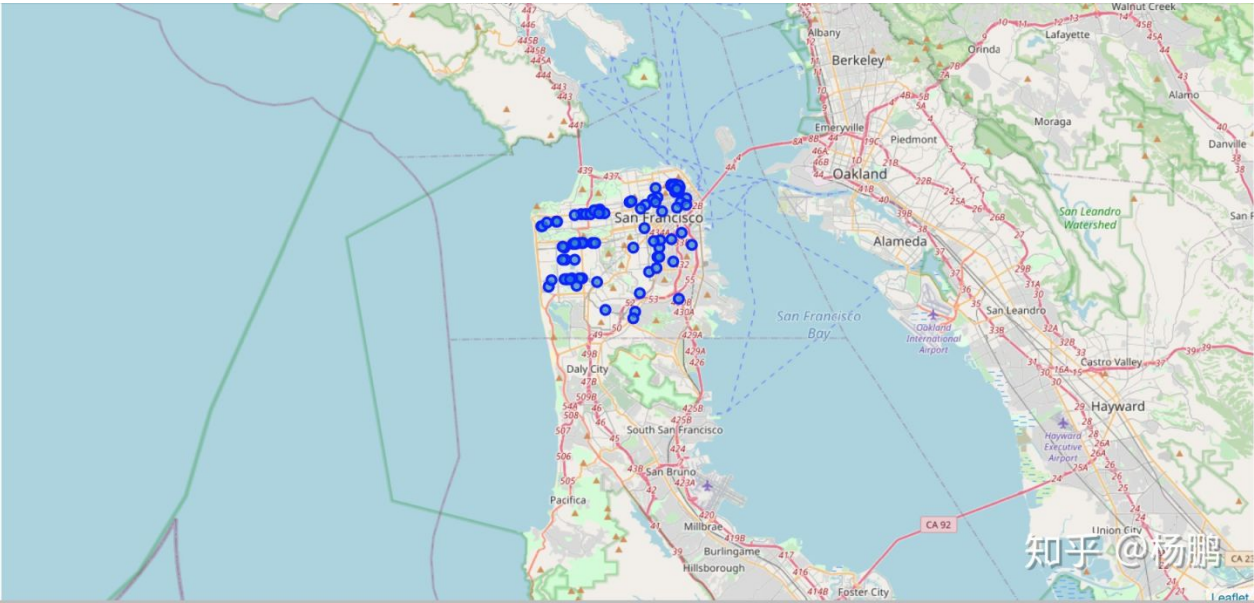


Chicago

Data Science

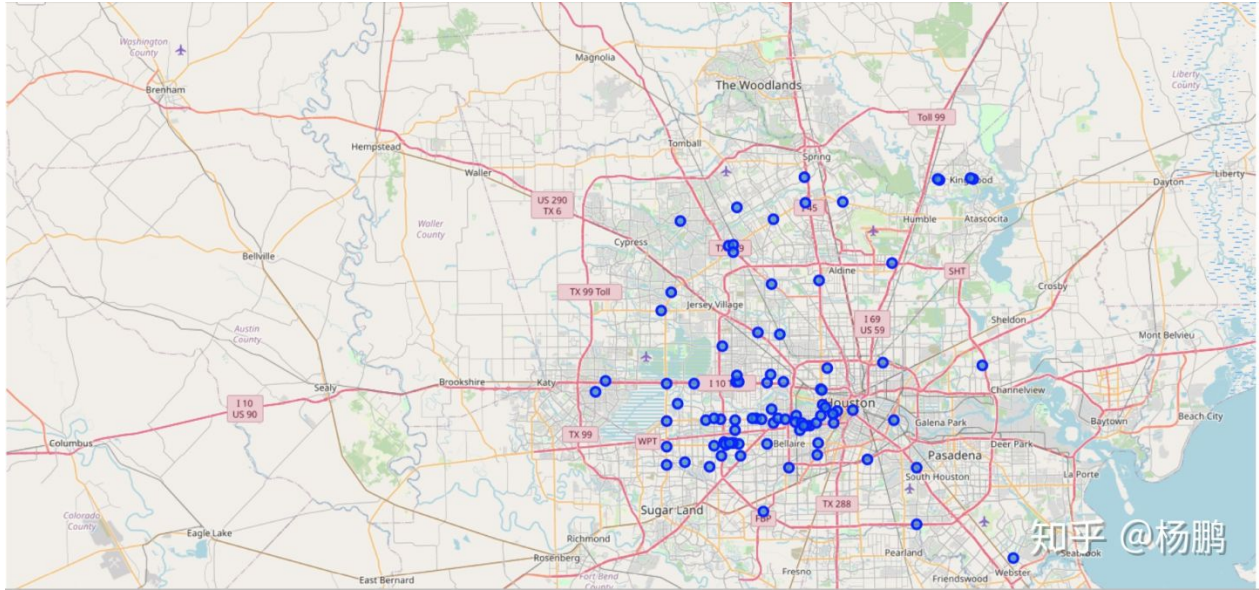


San Francisco

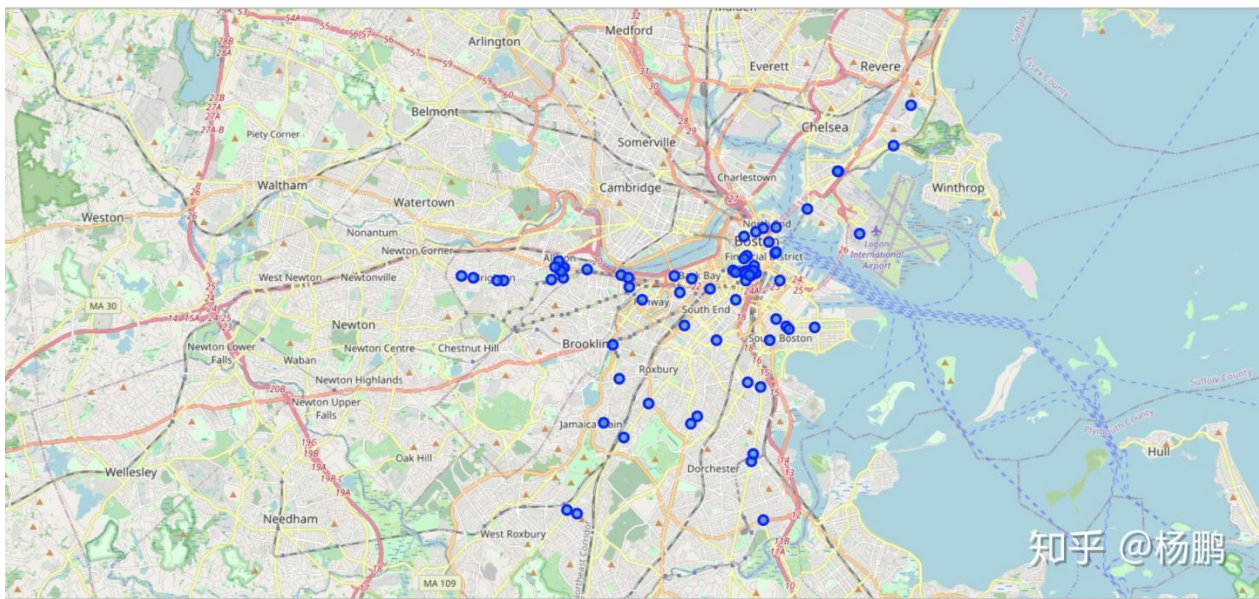


Houston

Data Science



Boston



Upon First inspection we see that New York, Houston and San Francisco are the most densely cities. In the next phase we Calculate the Mean coordinate and the mean distance to mean coordinate(MDMC). We represent the mean coordinate with a big green circle and distances with green lines

New York, NY

Mean Distance from Mean coordinates

Data Science

0.024481995330969415

Chicago, IL

Mean Distance from Mean coordinates

0.06863644136545263

San Francisco, CA

Mean Distance from Mean coordinates

0.038521524905265475

Houston, TX

Mean Distance from Mean coordinates

0.1340598491038303

Boston, MA

Mean Distance from Mean coordinates

0.03161295161570615

It seems Houston is worst for me an other programmer like me in US. Surprisingly Boston is almost same with San Francisco. We can predict New York is the best one.

New York:0.0244

Boston:0.0316

San Francisco:0.0385

Chicago:0.0686

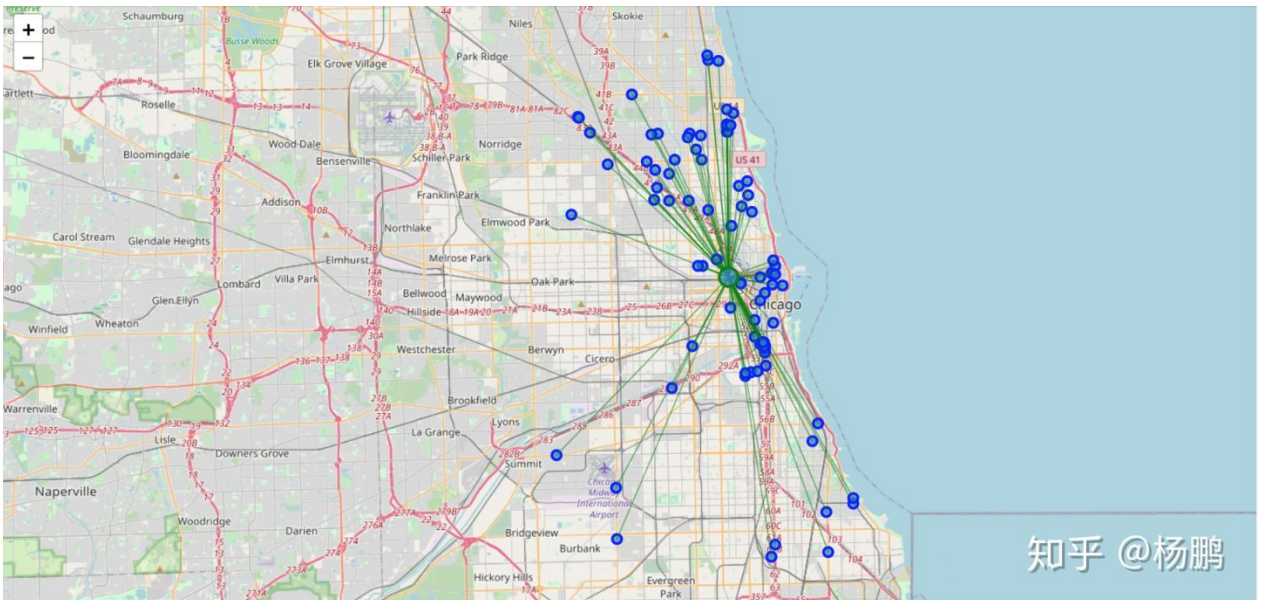
Houston: 0.1340

New York:

Data Science

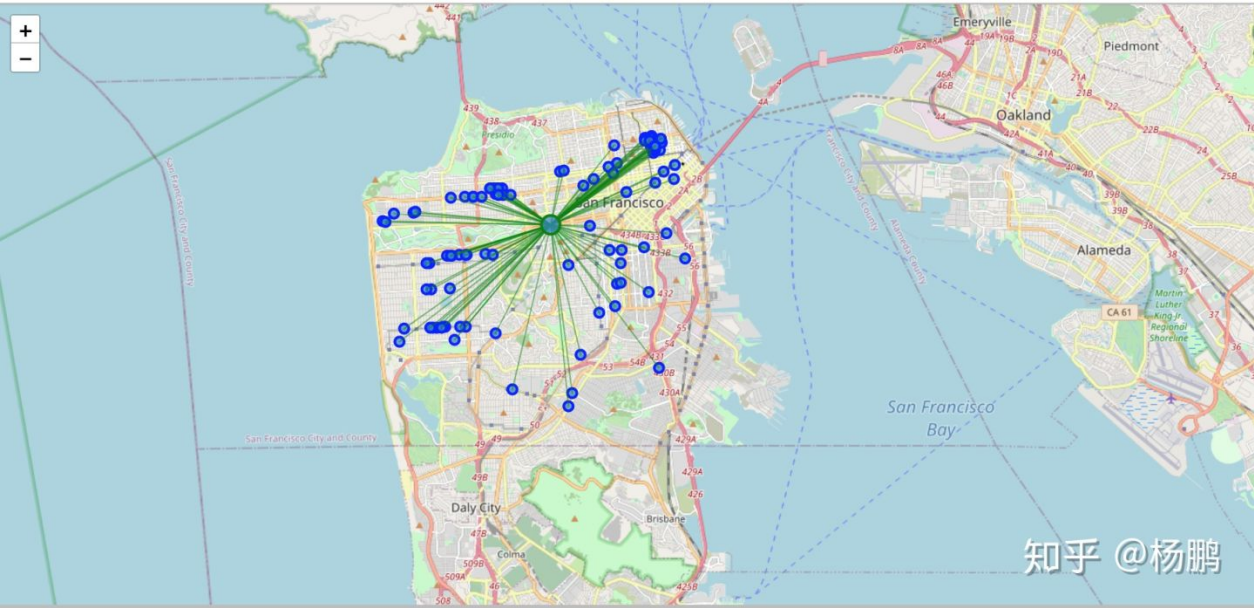


Chicago:

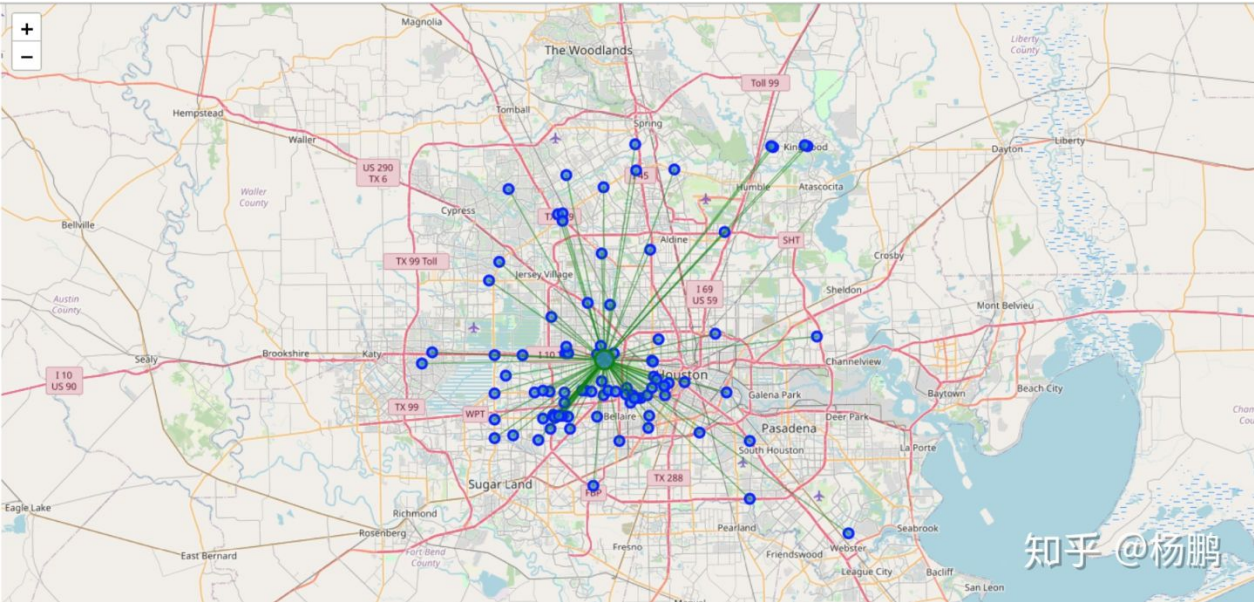


San Francisco:

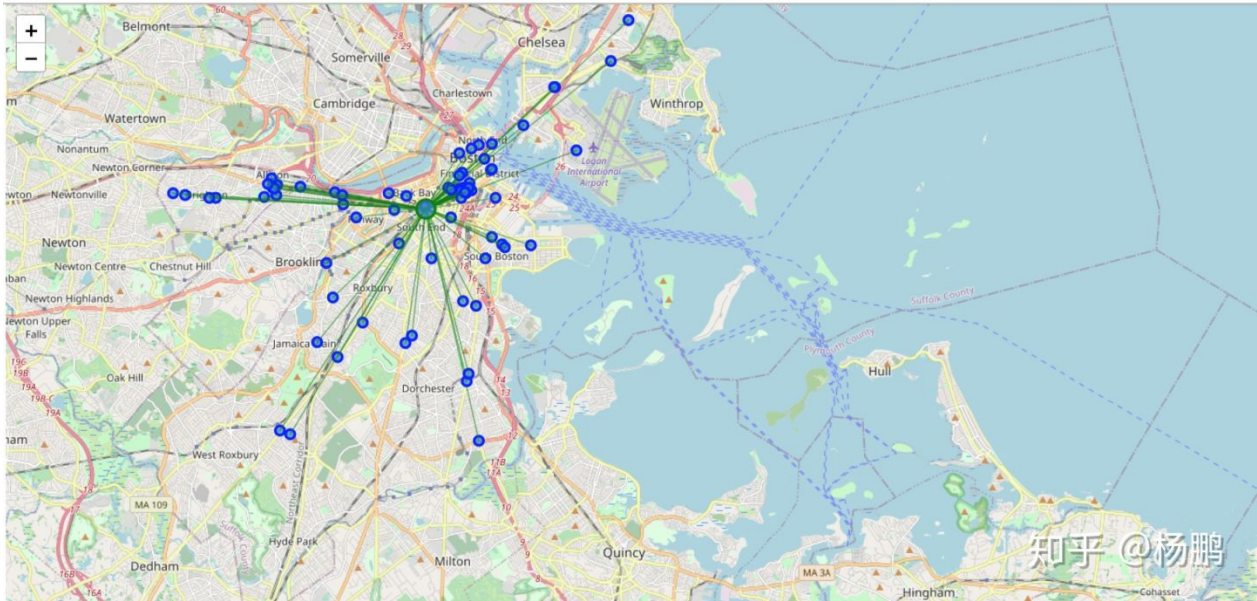
Data Science



Houston:



Boston:



Conclusion

Now there is no doubt that New York is the best place to Chinese foods in the US.