

# **The Battle of Neighborhoods – Five Tech Cities**

**Alfredo Perez**

**June 15, 2020**

## **1. Introduction**

### **1.1 Background**

According to [The Washington Post](#) we're currently living through an unprecedented era of growth in tech jobs. The explosion of U.S. tech jobs is concentrated in five metro areas. These five cities are Seattle, Boston, San Francisco, San Jose, and San Diego. Many startups are increasingly calling these cities home and many companies are increasingly shifting to a tech-focused mindset to compete within their industries. This brings about an increase in jobs where a tech background is needed, making skills such as data science, coding, and agile project management more important than ever.

Only five metro areas – Boston, San Diego, San Francisco, San Jose, and Seattle – attained up to 90% of the 256,063 tech jobs created from 2005 to 2017, according to the Brookings Institution and the Information Technology and Innovation Foundation. The remaining 10% was divvied up in 377 other areas. The share of these jobs dropped significantly in would-be hubs such as Chicago, Durham, North Carolina, Philadelphia, Dallas, and Wichita.

### **1.2 Interest & Problem**

In this scenario, it is important to adopt machine learning tools to assist our Tech workforce (herein referred to as “techies”) interested in living and working in five popular and populated US cities: Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA to make wise and effective decisions in choosing the city with a high density of tech startups to apply for jobs. As a result, the business problem we are currently posing is how could we provide support to techies to apply to tech startups in Philadelphia, New York, Chicago, San Francisco, and Boston?

To solve this business problem, we will use the FourSquare API to collect data of each tech startup venue in the five US cities we chose for this analysis, Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA. We look forward to guiding our techies to new areas of opportunities, literally not figuratively and may the best city win.

## **2. Data acquisition and cleaning**

### **2.1 Data sources, cleaning and exploratory data analysis**

The main goal is to analyze the city with the highest density of tech startups by using the FourSquare API through the venues channel. Using the near query will provide a significant number of venues in each city. The Category ID 4bf58dd8d48988d125941735 will only show tech startups with a limit of 100 venues per query. This request was made for each city, Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA, and from the result the name and coordinates data was plotted for cluster map visualization. The center coordinates of all venues were calculated to attain the mean longitude and latitude. Then, the Euclidean distance from each venue to the mean coordinates was calculated as a point of reference for each potential area that would occupy most tech startups all around.

Exploring and recommending different cities based on the density of where these tech startups are located in Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA will be retrieved by accessing data through FourSquare API interface and arranged for visualization. Then, we will be able to recommend a final city for our techies to apply for jobs and potentially go to different interviews the same day or week.

## **3. Methodology Section**

Exploring and recommending different cities based on the density of where these tech startups are located in Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA will be retrieved by accessing data through FourSquare API interface and arranged for visualization. Then, we will be able to recommend a final city for our techies to apply for jobs and potentially go to different interviews the same day or week.

**The Methodology section comprises four stages:**

- 1. Collect Data**
- 2. Explore and Understand data visualization. Enjoy clicking on each circle to see the name of each tech startup**
- 3. Data Preparation before Euclidean distance**
- 4. Processing Data after Euclidean distance**

**Breakdown:**

- 1. Collect Data**

Using FourSquare we gather the following data.

-----Collect Data-----

Total number of tech startups in Philadelphia, PA = 107

Showing Top 100

-----Collect Data-----

Total number of tech startups in New York, NY = 180

Showing Top 100

-----Collect Data-----

Total number of tech startups in Chicago, IL = 141

Showing Top 100

-----Collect Data-----

Total number of tech startups in San Francisco, CA = 195

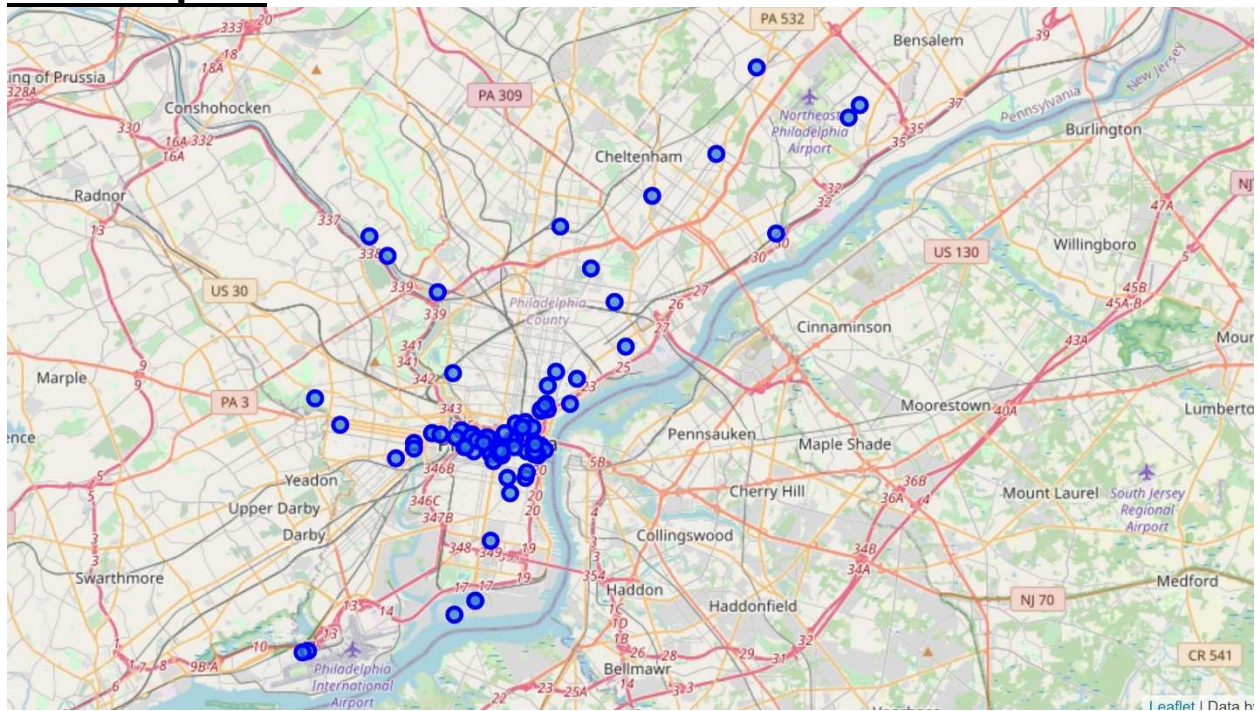
Showing Top 100

-----Collect Data-----

Total number of tech startups in Boston, MA = 130

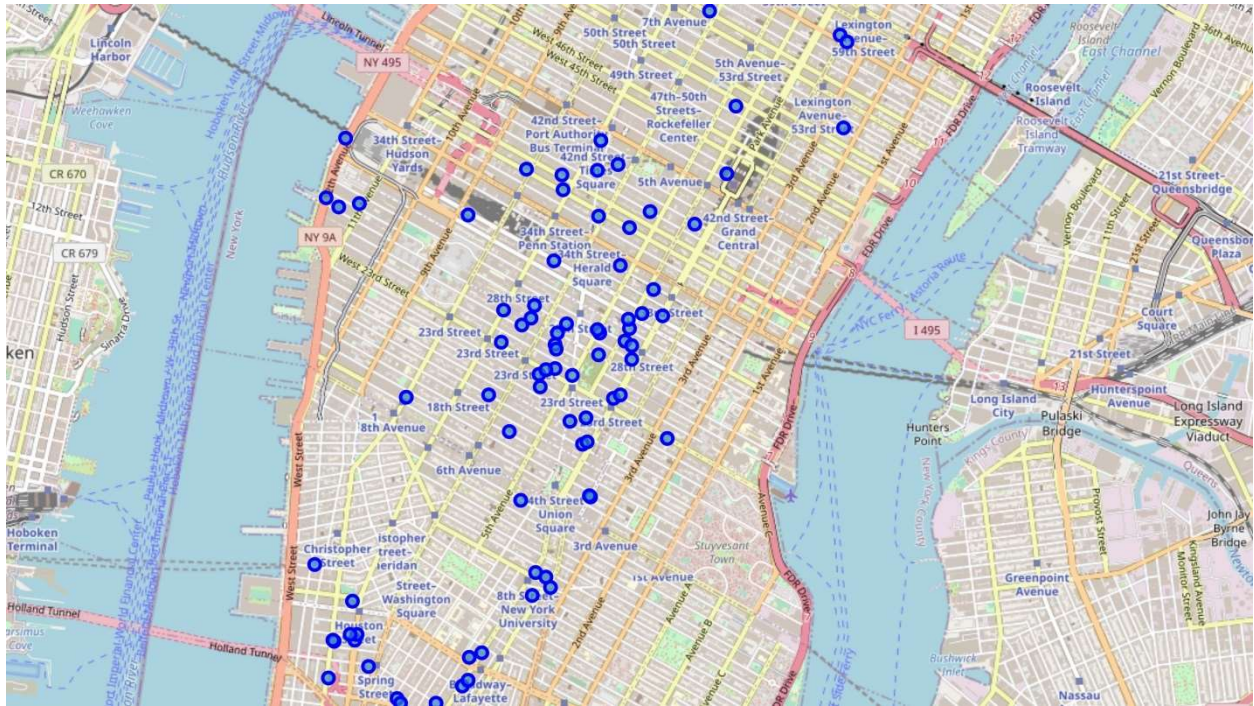
Showing Top 100

## Philadelphia

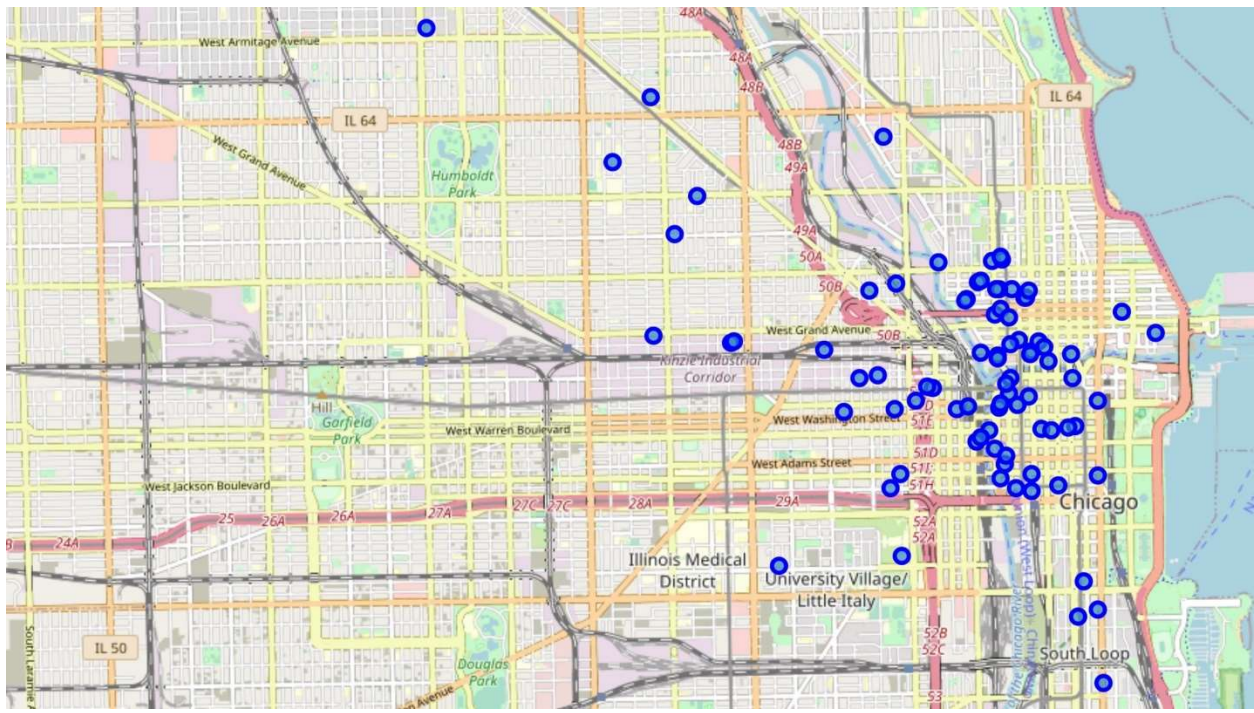




## New York, NY

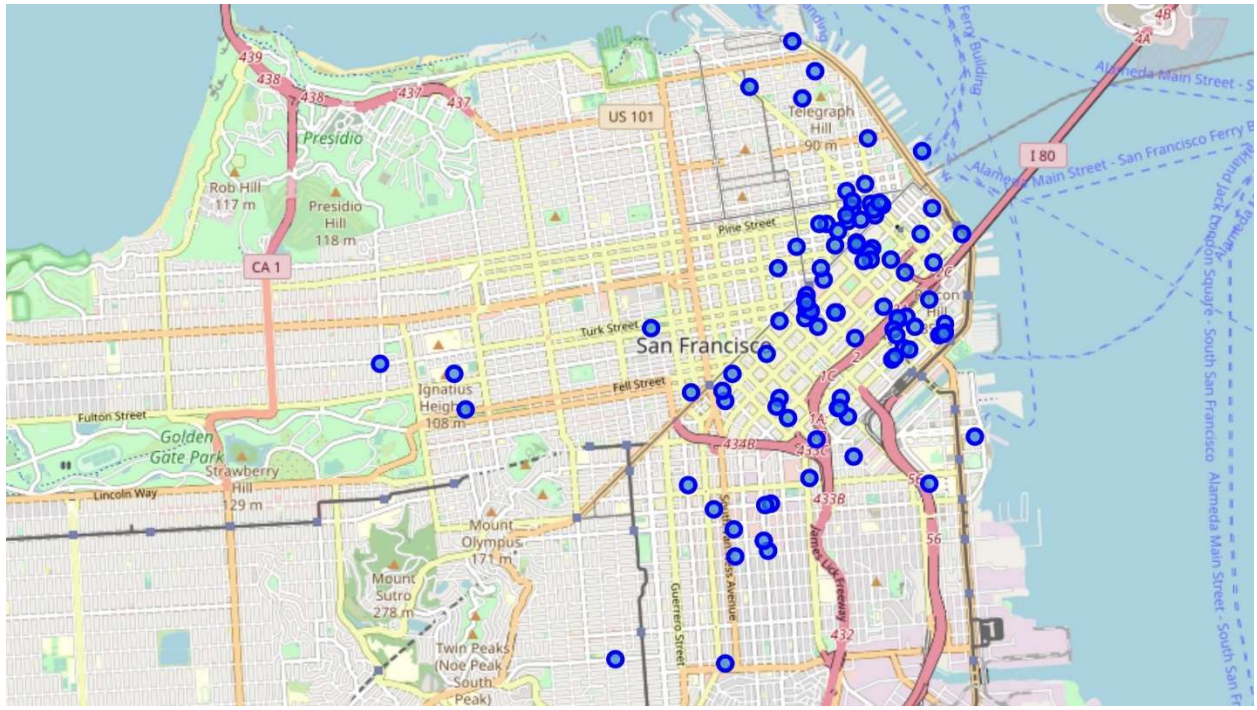


## Chicago, IL

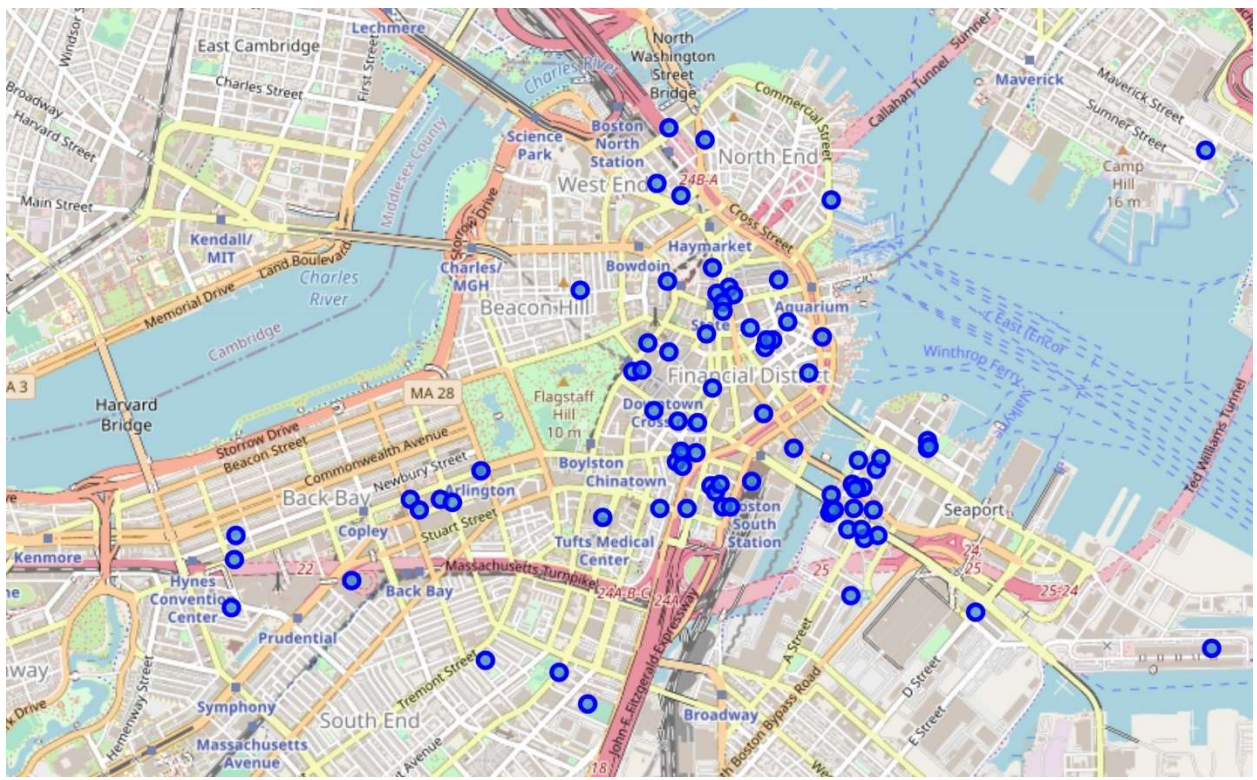




## San Francisco, CA



## Boston, MA



We can see that Boston, MA, New York, NY, and Philadelphia, PA are dense cities with tech startups. There is a way we can do better and have a concrete measure of this density by analyzing the size using a statistical approach. Attaining the mean location of all the tech startups which will be in high density when near and low density when far. Tracing the average distance of each venue to the mean coordinates would greatly determine the outliers and the best area to select when considering living and working in a location with the highest volume of tech startups.

Mean distance from mean coordinates grounds itself in euclidean distance (represents the shortest distance between two points) though in this case for accuracy, intent, and purposes it is the mean which will be utilized. Potential processing of this data can lend itself to interpretations based on Pattern classification, k-Nearest neighbor, and Distance function which are outside of the scope of this analysis. As you will see below, we represent the mean coordinates with a large yellow circle and distances with black lines.

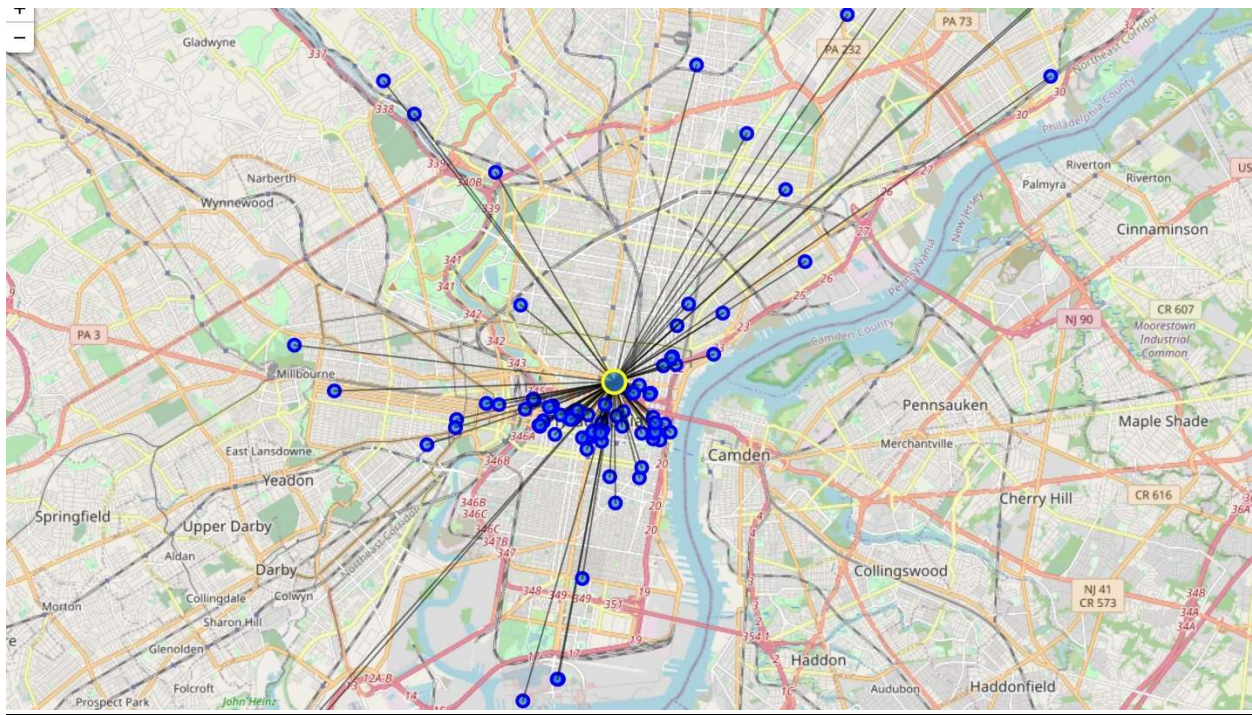
## **Analysis**

```
Philadelphia, PA
Mean Distance from Mean coordinates(MD)
0.03704758874316186
New York, NY
Mean Distance from Mean coordinates(MD)
0.01588455816583397
Chicago, IL
Mean Distance from Mean coordinates(MD)
0.029379499925633813
San Francisco, CA
Mean Distance from Mean coordinates(MD)
0.013971154430047825
Boston, MA
Mean Distance from Mean coordinates(MD)
0.01866261368241032
```

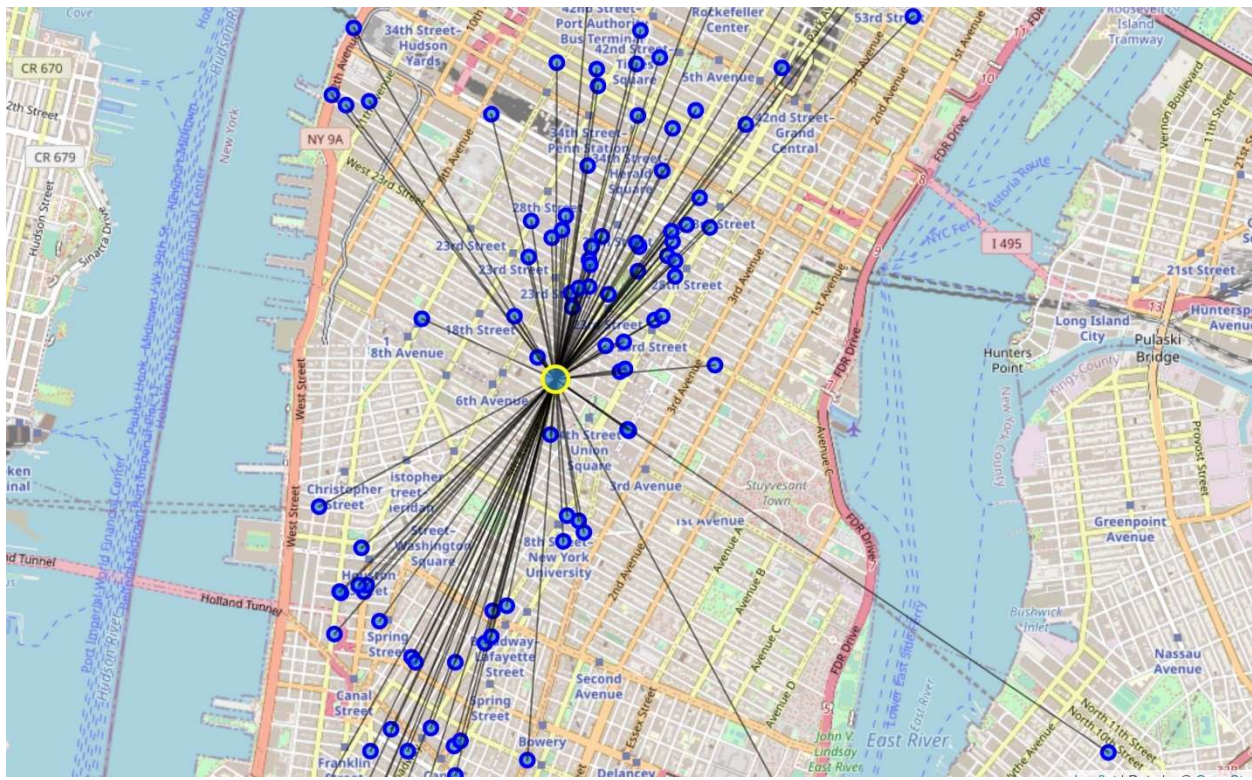
---



## Philadelphia

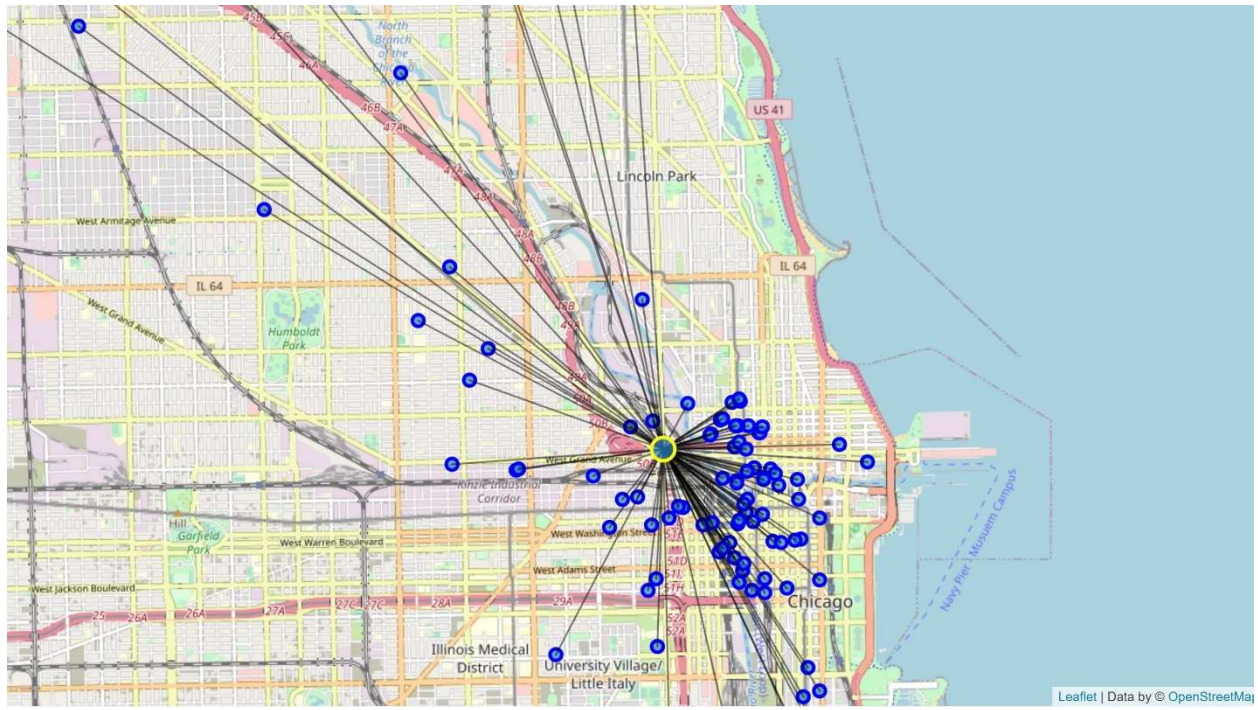


## New York, NY

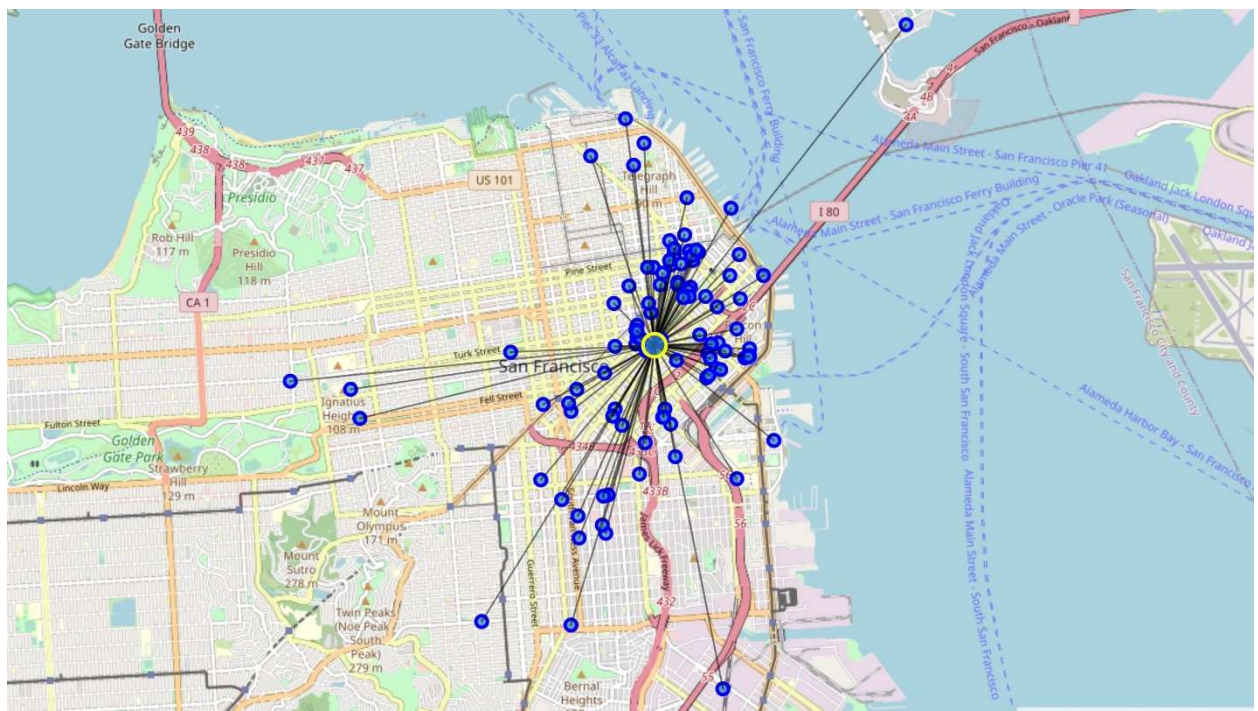




## Chicago, IL

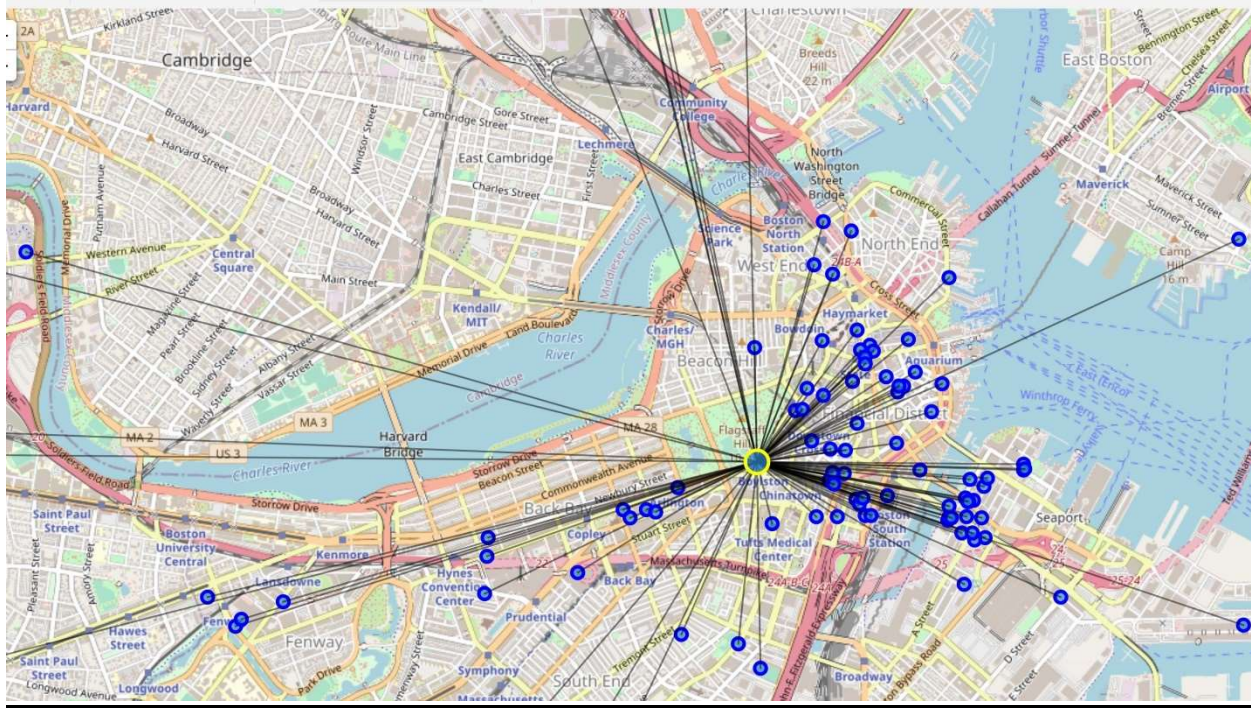


## San Francisco, CA





## **Boston**



We can see that San Francisco, CA is the best city for any techie! It's amazing how by proximity it provides the most tech startups in a single location which will help all techies to apply to jobs around the area. The maps of the other four remaining cities help pinpoint the exact location where techies can find the most opportunities if they are still interested in applying to jobs in those cities. You can also see clearly see the outliers in each map which are too far from the center of the mean.

## **Results and Discussion section**

Tech startups have grown exponentially and from the five cities we analyzed San Francisco, CA still holds it's grounds as one of the most sought out locations for tech startups and for potential techies who are looking for the most opportunities all around in a single location it's definitely worth your time and efforts to send job applications and set up interviews all around until you get hired!

Five cities were analyzed. Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA, which were all within our scope to determine which of these will have the tightest cluster of tech startups and increase your hiring possibilities. The data provided also gives you pinpoint accuracy of each tech startup which you can research and decide if that company is one

which aligns with your values and goals. If you were to look into the map in San Francisco, CA you can see that Pivotal HQ, PlayHaven South HQ, and Bridgefy SF HQ are relatively very close to each other and if you were to live around that location you would be strategically around many other potential companies.

Let's look into Philadelphia, PA where you can find a small cluster of tech startups like Appadoo, NewQuest USA, TTR Data Recovery Services - Philadelphia which are relatively clustered together and if it's a city you were considering to live and work then you can definitely explore your opportunities. You can take this same concept and look into other clusters in other cities and find your own opportunities!

## **Conclusion**

The increase in tech jobs and tech startups shows us how entrepreneurial our society has become. Continuous improvements and innovations in technology will help gear us towards more connectivity and easier access to information with which we will become a better society. Change takes time. It is certain that more and more tech startups will emerge as time goes on. I look forward towards your own self growth and to use this data to look into the different tech start ups in the five cities which were analyzed (Philadelphia, PA, New York, NY, Chicago, IL, San Francisco, CA, and Boston, MA,) and motivate you to research a few companies and take a look into their different products and services!