# Applied Data Science Capstone

Assignment -

Predict hospital equipment needs in major cities in the United States

# Introduction

**Problem Statement**

Compare top cities in the United States with respect to business prospects for a hospital equipment seller

# Background

In the wake of new health emergencies, countries health system is going through a crisis. Some cities are worse than others with respect to availability of hospitals with critical care. As these cities ramp up to handle the current crisis as well to be prepared for the future, they'll need to build new hospitals and equip existing ones with the necessary equipment.

A seller for multiple brands of hospital critical care equipment would also have to scale up to be able to grab the available opportunities to sell. Since it's a time of crisis, he has to make the equipment instantly. He will find it very useful to be able to estimate the demand in advance. He can keep the inventory ready in the same city to shorten the delivery time and create a competitive advantage.

These more populous metropolitan areas in the country are definitely best places for this business. We'll try to find areas where we see a deficit and may need to build new hospitals or expand existing ones.

We will go through each step of this project and address them separately. I first outline the initial data preparation and describe future steps.

# Data

In order to compare the cities we need the following data

1. list of cities in the United States with population

2. For each city number of hospitals and geo location data for each hospital

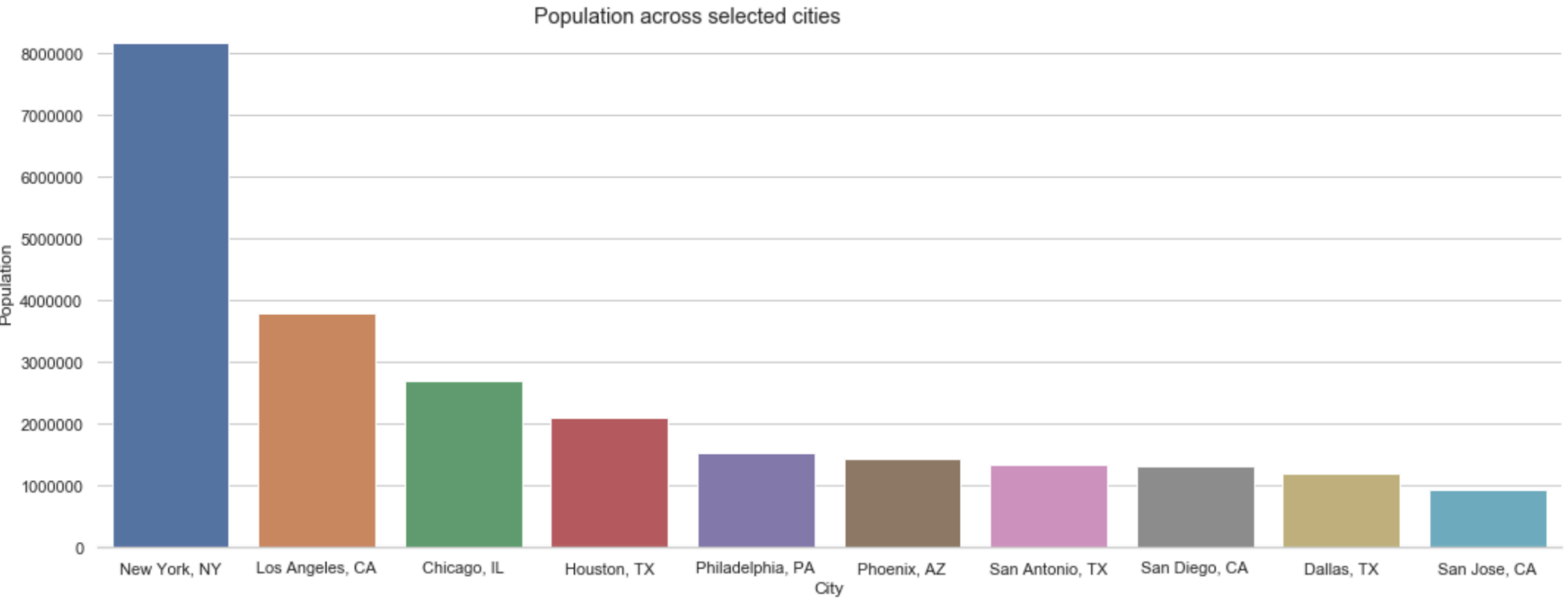
List of cities with population data will be obtained from the website - https://www.nlc.org/the-30-most-populous-cities

All data related to number of hospitals in each city and locations of each hospital will be obtained via the FourSquare API utilized via the Request library in Python.

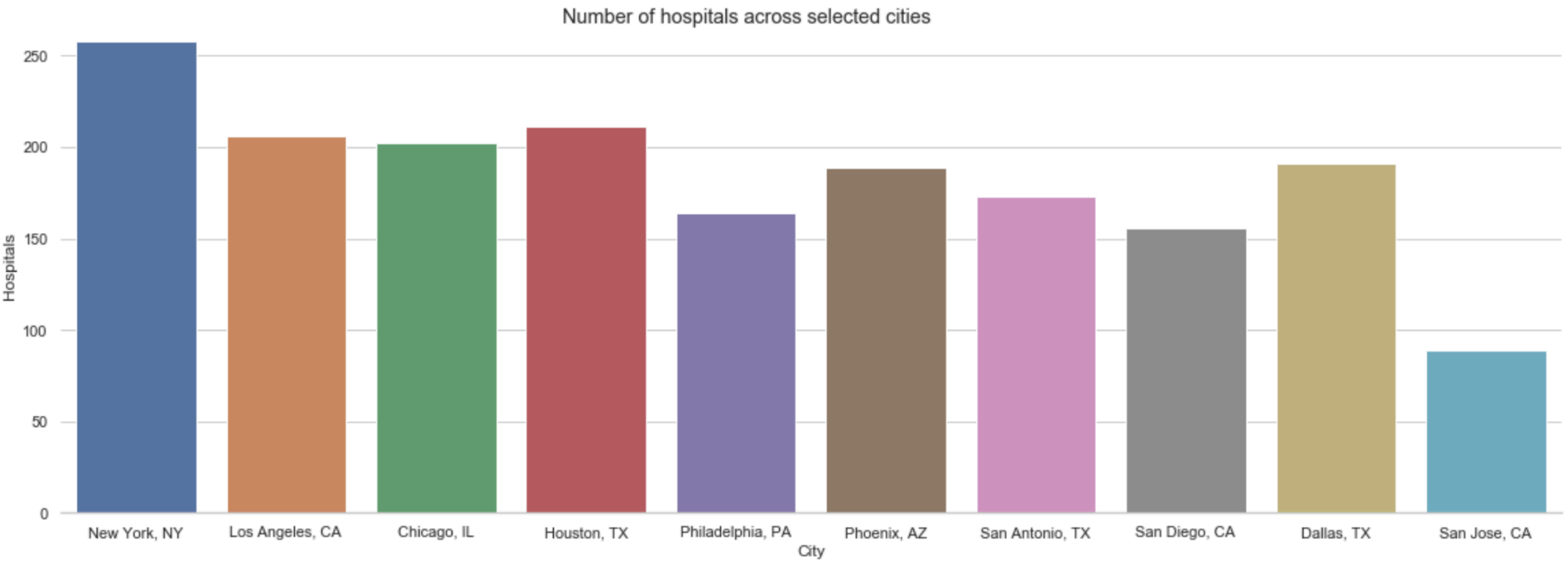
# Methodology

- Cities and population D=data will be collected from https://www.nlc.org/the-30-most-populous-cities and cleaned and processed into a dataframe.
- FourSquare be used to locate all hospitals in each of the cities. Hospital details with geolocation data is added to the dataframe.
- Data is processed to calculate additional values like hospital density for each city
- Finally, the data be will be visually assessed using graphing from various Python libraries.

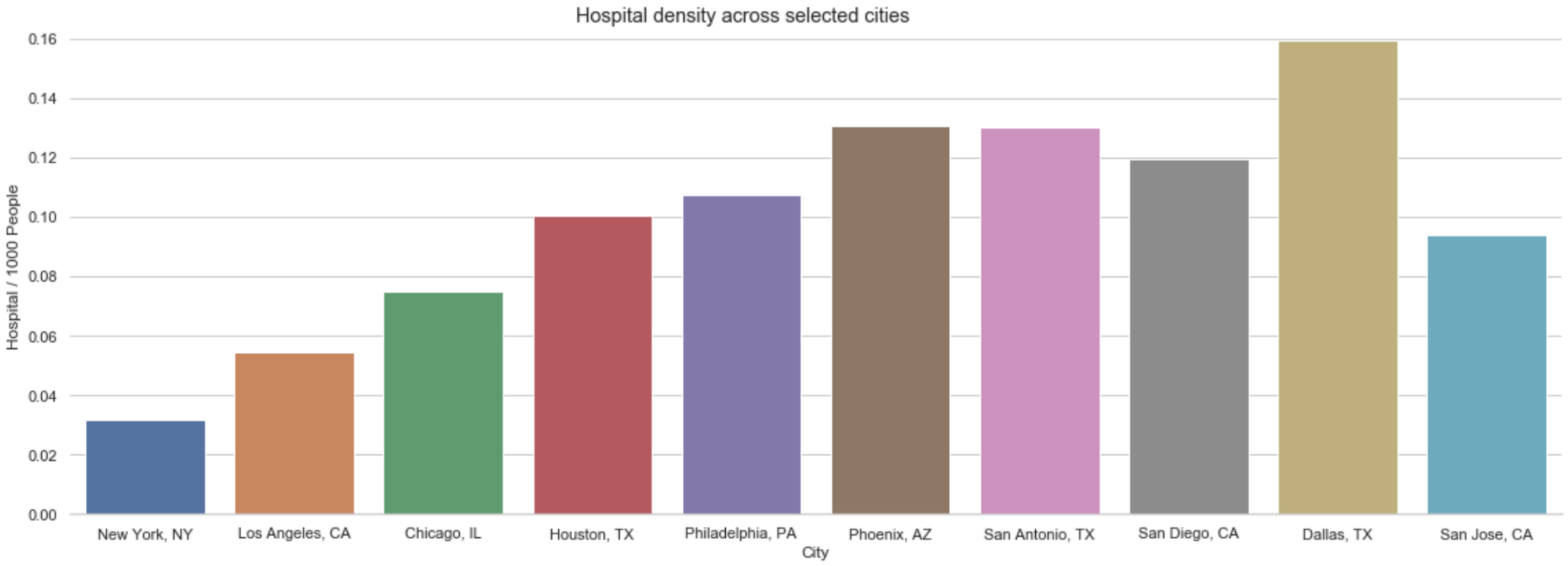
# Population across selected cities



# Number of hospitals across selected cities

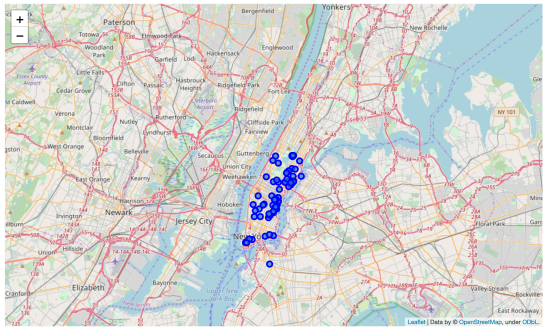


# Hospital density across selected cities

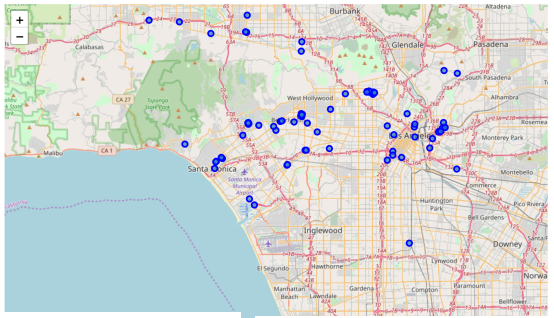


# Distribution of hospitals across cities

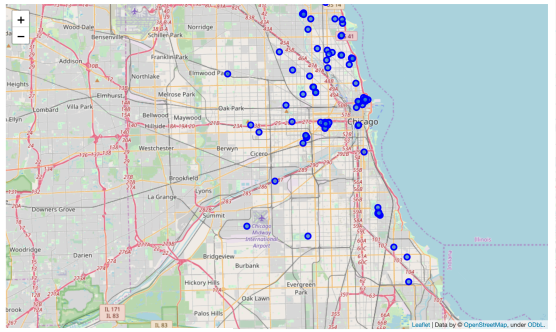
Hospitals in New York, NY



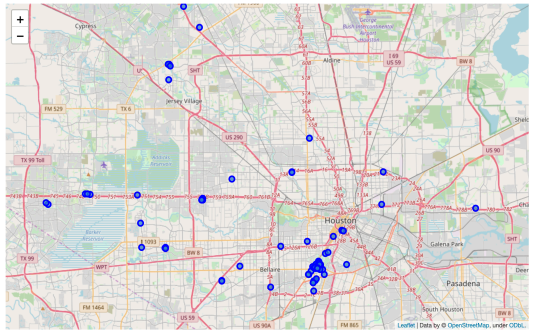

Hospitals in Los Angeles, CA



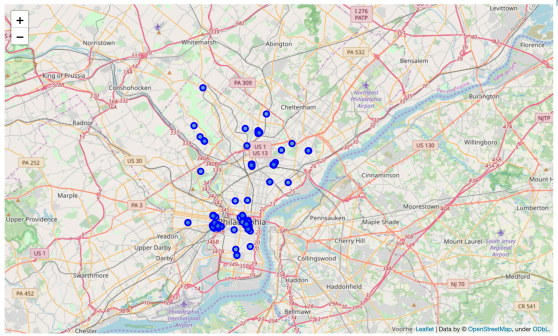

Hospitals in Chicago, IL



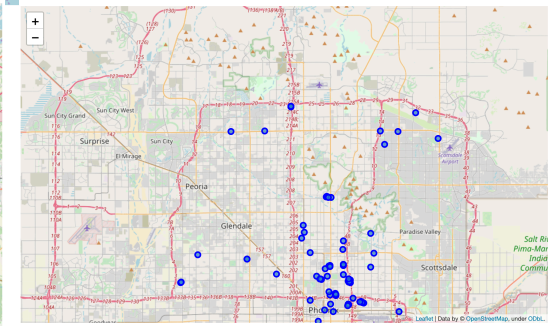

Hospitals in Houston, TX



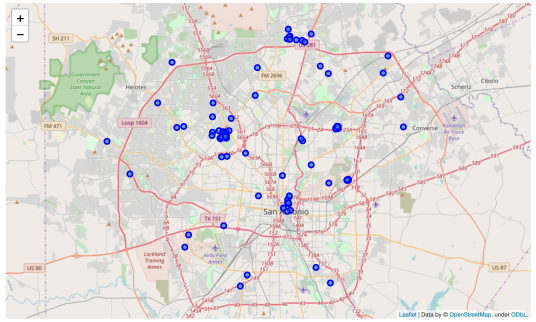

Hospitals in Philadelphia, PA



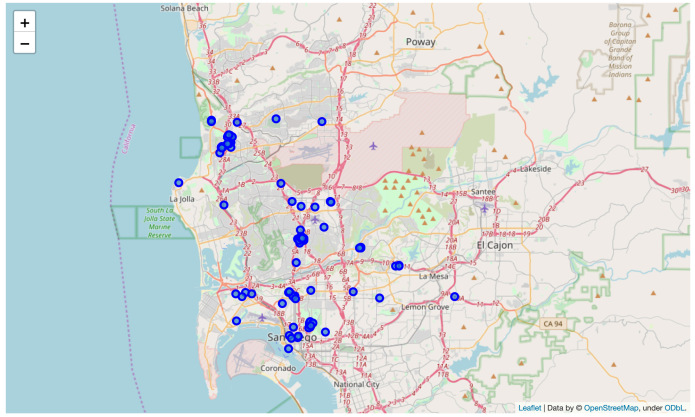

Hospitals in Phoenix, AZ



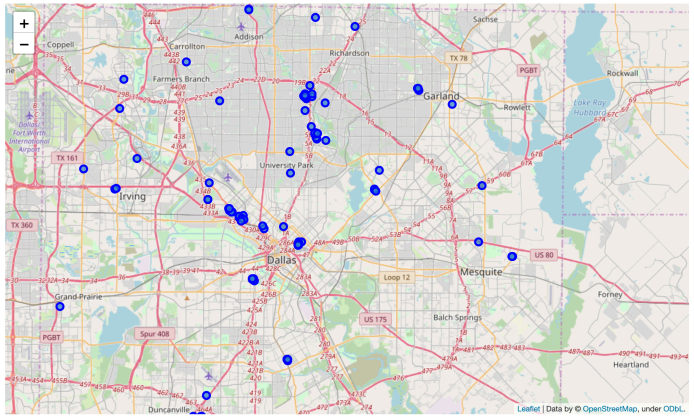

Hospitals in San Antonio, TX



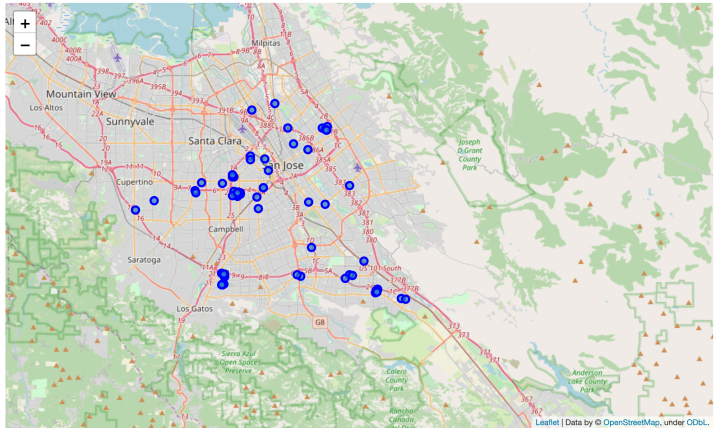

Hospitals in San Diego, CA




Hospitals in Dallas, TX




Hospitals in San Jose, CA



# Results and Discussion

Based on our analysis above, we can draw a number of conclusions that will be useful to aid any business person.

We scraped the web for a list of most populous cities in the United States. For this assignment we picked the top 10 cities. For each of these cities we used the Foursquare API to get the number of hospitals and the details of each hospital such as latitude and longitude values After collecting data from the Foursquare and web, we got a list of 120 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their as well as names to combine them and remove all the outliers. This collected a number of records for each of the 10 cities

We processed the data to calculate hospital density in each city and compared them visually along with population data and number of hospitals For each city hospital locations were plotted on the map. This provides clue as to how the hospitals are scattered in each of the cities. Further analysis can be done at a more granular neighbourhood level to suggest ideal locations for new hospitals

Hospital equipment seller who's looking for cities which may need to set up more hospitals will find New York as the right place

# Conclusion

The purpose of this project was to compare the cities in the United States with respect to the business prospects for a hospital equipment seller. The cities have been identified using Foursquare and data from web scraping and have been plotted on bar charts and on the maps. These plots reveals that New York has the least number of hospitals per 1000 people. To prepare for current or future health crisis, New York may have to invest more in their health systems. This may have to be taken up sooner and provides the right information to the hospital equipment seller to take the right business actions and to optimize his inventory