Find Best Neighbourhood to Fight Covid-19 in New York City

Contents

[Find Best Neighbourhood to Fight Covid-19 in New York City 1](#_Toc40472999)

[**Step 1: Introduction to Business problem** 1](#_Toc40473000)

[**Step 2: Data** 1](#_Toc40473001)

[**Step 3: Approach** 1](#_Toc40473002)

[**Step 4: Data preparation** 1](#_Toc40473003)

**Step 1: Introduction to Business problem**

Right now, New York is one of the worst hit state by [COVID-19](https://en.wikipedia.org/wiki/Coronavirus_disease_2019) in USA. New York city is at the centre of the disaster. The hospitals are already stretched thin with patients overflowing. According to current report, the number of positive cases has already touched 350K and death toll at 27600.

I want to create something useful which would give some insight on this situation and can be used in case someone is looking to move to another neighbourhood within New York city. In this project we are going to determine which neighbourhood is best prepared for Covid-19, by finding out the best ratio of hospital beds per person for each neighbourhood in this city.

**Step 2: Data**

We will be collecting data from following sources:

1. New York City data that contains **borough**, **neighbourhoods** along with their **latitudes** and **longitudes**.
   * Data source: [NYC data set](https://cocl.us/new_york_dataset).
2. We are going to get **population** data from Scraping Wikipedia.
   * Data source: Wikipedia page of [NYC neighbourhood](https://en.wikipedia.org/wiki/neighborhoods_in_New_York_City).
   * We are going to go through each of the links of neighbourhood and find the population of each of them.
3. Hospital information is going to be fetched from foursquare API.
   * Data source: foursquare API
4. Hospital bed information is going to be fetched from **NYS Health Profile website** by Web Scraping.
   * Data source: [NYS Health Profile](https://profiles.health.ny.gov/).

**Step 3: Approach**

This is our approach to resolve issue:

* Collect the New York city data from [here](https://cocl.us/new_york_dataset).
* Collect population data for each neighbourhood by scraping Wikipedia.
* Using Foursquare API, we will get hospitals for each neighbourhood.
* Collect hospital bed data by scraping data from **NYS Health Profile**.
* Data Visualization and some statistical analysis.
* Analysing using Clustering (Specially K-Means).
* Find the best value of K
* Visualize the neighbourhood max density of hospital beds per 100 people.
* Visualize the neighbourhood max density of hospital ICU beds per 100 people.
* Inference from these results and related conclusions.

**Step 4: Data preparation**

Data used in the analysis are listed below:

* First, get the json data from [here](https://cocl.us/new_york_dataset), which will contain borough, neighbourhood, latitude and longitude information.
* neighbourhood data in New York City will be collected from scraping the [Wikipedia](https://en.wikipedia.org/wiki/neighborhoods_in_New_York_City) page. links given in the neighbourhood section of the table will be visited via scraper and find the population for each of them. Then data will be cleaned up and used to create a data frame containing borough, neighbourhood and population.
* Hospitals per neighbourhood information will be collected from foursquare API.
* We will collect bed and ICU capacity information from [NYS Health Profile website](https://profiles.health.ny.gov/). Will be using selenium-based scraping as this is a dynamic site.