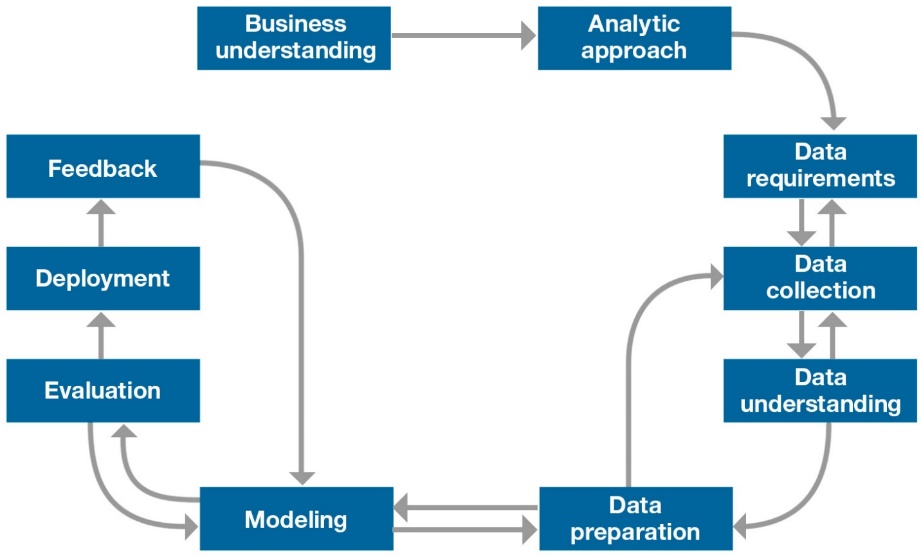
**IBM/COURSERA CAPSTONE PROJECT**

***Exploring the neighborhoods in New-York City that are the best equipped to fight the Covid-19 Pandemic***

*To answer this question, we will follow the recommended IBM Methodology for Data Science:*

******

**1. Business Understanding.**

**1.1. Background.**

The Coronavirus disease 2019 (COVID-19) is an [infectious disease](https://en.wikipedia.org/wiki/Infectious_disease) caused by [severe acute respiratory syndrome coronavirus 2](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2) (SARS-CoV-2). The disease was first identified in December 2019 in [Wuhan](https://en.wikipedia.org/wiki/Wuhan), the capital of China's [Hubei](https://en.wikipedia.org/wiki/Hubei) province, and has since spread globally, resulting in the ongoing [2019–20 coronavirus pandemic](https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic). As of 29 April 2020, [more than 3.13 million cases](https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic_cases/WHO_situation_reports) have been reported across 185 countries and territories, resulting in [more than 217,000 deaths](https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic_deaths/WHO_situation_reports).

Within this context, all countries have been hit, with some Regions of the World more or less severely impacted. New-York City in the US is one of the cities that have been strongly hit, with a death toll of approximately 12,509 as of today (out of a total of 58,864 deaths in the US) according to the WHO statistics.

**1.2. Business Problem.**

Hospitals across the state “have been urgently looking to expand capacity in advance of the continuing surge in the number of coronavirus patients and officials said they are planning to possibly shift confirmed virus patients from hospitals with dwindling numbers of available beds to hospitals elsewhere in the state” (ctmirror.org). At the time of writing this report (April 29th 2020) the situation has improved, but experts are dreading a second wave of infections. Within this context, it is important to have a closer overview of the neighborhoods in New-York City that are the best prepared to welcome infected patients and to fight against this pandemic, by looking at the hospital bed capacity at each neighborhood. To complete this analysis, I will also look at the cases of infected people by neighborhood in order to be able to compare with the hospitals bed capacity.

**1.3. Target Audience.**

The target audience would be medical and non-medical experts and analysts from the US Department of Health and Human Services (HHS) working for the City of New-York.

This study will give them a clear understanding of hospitals bed capacity at the different neighborhoods of NY City. The comparison with the number of infected cases at each neighborhood will give a better comprehension in the handling of the Covid-19 pandemic at the neighborhood level. This would also be helpful for making future predictions and for adopting a better crisis management approach in case of a second Covid-19 wave.

**2. Analytic Approach.**

We will adopt the following approach in an attempt to answer the problem:

* Collect the data about New York City.
* Collect the data about New-York City population for each neighborhood.
* Collect the data about the number of infected people for each neighborhood.
* Use the Foursquare API to get the list of hospitals at each neighborhood.
* Collect the hospital bed data.
* Perform Data Visualization statistical analysis.
* Analyze data by Clustering (using K-Means technique).
* Find the best value of K
* Visualize the neighborhood max density of hospital beds per 100 people.
* Visualize the neighborhood max density of hospital ICU beds per 100 people.
* Look at the number of infected people at each neighborhood and compare with the hospitals bed capacity at each neighborhood, along with some visualizations.
* Provide feedback, draw conclusions and open on future implications or questions for research.

**3. Data Requirements.**

As mentioned, we need to collect a variety of data for this study.

* **From public (online) data sources:** we will collect data about New-York City, its neighborhoods, its population. We will also collect data about the hospitals-bed capacity and about the number of infected cases in New-York City. We might have to scrap data from these sources.
* **From Fourthsquare API:** we will get the list of hospitals at each neighborhood by calling the Foursquare API, using the “venues” parameter.

**4. Data Collection.**

* New York city data: from Json file: <https://cocl.us/new_york_dataset>
* Population data for each neighborhood: Wikipedia: <https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City>
* List of hospitals at each neighborhood: through the Foursquare API.
* Hospital bed data: from the NYS Health Profile: <https://profiles.health.ny.gov/hospital/index#5.79/42.868/-76.809>
* List of confirmed Covid-19 cases by Borough: Wikipedia: <https://github.com/nychealth/coronavirus-data/blob/master/boro.csv>