



**IBM Developer  
SKILLS NETWORK**

[\(img\)](#)

# **Coursera Capstone Project - Battle of Neighborhoods**

**IBM Data Science Professional Certificate**

**TITLE: Finding oportunities for opening a delivery-based supermarket in New York City, considering a post-COVID scenario**



[\(img\)](#)

**Author: Tiago de Moraes**

**June 2021**

# Introduction

This project aims to answer a very concerning question that most countries around the world are questioning themselves: "how will the world be like after the Covid-19 crisis is behind us?". In other words, how will the economy be shaped after the social distancing, the obligatory use of masks and constant fear of going outside is over? Many specialists believe that humanity will eventually move on and resume their pre-Covid routines, such as going in restaurants and enjoying events in closed spaces.

## But what new habits acquired during the pandemic will persist?

That is exactly the question that many entrepreneurs are asking themselves right now. They want to find in first-hand opportunities to invest in business that caters to new acquired habits that will carry on after the pandemic is over. One of such habits that most people agree that will persist is the delivery of groceries and general goods, specially in big cities.

Two key aspects may explain why that might be the case:

- During 2020, many business had to quickly adapt their supply and delivery chains in order to cater to new demands of a world hit by storm by a new disease that forced people to stay inside their homes for many months. That established a new sales and after-sales channels that had to mature in a record time. By 2021, that investment will probably not be lost and will continue to be used by many businesses.
- Customers in general got used to order many things online during 2020, with online shopping taking an exponential growth in a matter of weeks. This type of convenience is very appealing and even people that aren't normally very keen to shop online (middle-aged people) are already familiarized with this system. In 2021 this type of convenience might as well persist as a major delivery sales channel for many businesses and will no longer be a differential, but a mandatory revenue stream for any new business to come.

Considering that scenario, this final project explores the best locations for a delivery-focused market throughout New York City. As with any business, opening a new market requires thorough market study considering many key aspects in particular. The location of the market is one of the most important factors that will affect the outcome of the business. Hence, this project will attempt to find optimal locations to install a new delivery-focused market in New York, considering its neighborhoods profile and local venues locations.

## Objective

Using Data Science methodology and instruments such as Data Analysis and Visualization, the main goal of this Capstone project is to analyze and select the best locations in the city of New York for the deployment of a delivery-focused market. This project aims to provide solutions to answer a core business question: *\*\*Where in New York City should the investor consider to install a delivery-based market? \*\**

## Data

To make such analysis, the following data will be queried:

- Geospatial data of the city of New York, containing the neighborhoods and boroughs information;
- Latitude and longitude coordinates of those neighborhoods for plot purposes and to get local venue data.

- Venue data, particularly data related to nearby markets, in order to to perform further analysis of the neighborhoods.

This project will require using of many data science skills, from web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium), along with extensive exploratory data analysis and statistical testing for model assessment using machine learning techniques.

## Methodology

- Data will be collected from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) ([https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)), wrangled and then processed into a Pandas dataframe.
- FourSquare API will be used to locate all venues and then filtered by markets in New York. Ratings, tips, and likes by users will be counted and added to the dataframe, and then data will be sorted based on rankings.
- Data be will be visually assessed using Python libraries.
- Maps will be also generated for a more pleaseant visual analysis
- k-means algorithms will be used for cluesterization of spatial data