For this model, I chose to focus the data related to the schools and the high schools of New York City.  
I chose this city since its area has the largest amount of known coronavirus cases in the USA and probably in the entire world. I arranged the schools' data in the model's database based on the following features:

1. Name of School, Neighborhood and Borough – In order to get the list of schools in NYC, I used the same database which was introduced in Assignment 3,derived from the link <https://geo.nyu.edu/catalog/nyu_2451_34572> , which has the information about the lat-lon locations of each neighborhood in the city. By that data, I used the Foursquare API explore method to scroll over the neighborhoods and scrap the basic information of each school in NYC while filtering the venues for the relevant categories ("School", "High School", "Religious School" and etc.)
2. Latitude and Longitude of the school – By the process that was mentioned in section 1, I could also retrieve the lat-lon locations of each school in the database.
3. The Zip Code Area of the school and its amount of Coronavirus Cases Per 1000 people – Based on the map and the CSV data which are introduced in the article: "Coronavirus in New York City, Tracking the spread of the pandemic", link: <https://projects.thecity.nyc/2020_03_covid-19-tracker/>.   
   CSV files for zip code lat-lon locations and the cases for zip code – at my Github repository:

<https://github.com/Avichai1125/Coursera_Capstone/blob/master/uszips.csv>  
<https://github.com/Avichai1125/Coursera_Capstone/blob/master/cases_per_zip.csv>

1. The composition of the neighboring venues near the school – The model assumes that schools are likely to be opened at the last stages of the exit strategy, and therefore examining the types of the closest venues to the school's area is a good key for examining the risk of the virus' relapse after the school reactivation. The venues will be retrieved by the Foursquare API explore method and they will be counted and divided into main groups based on their categories, each represented by a column in the database: Restaurants, Shops and Stores, Outdoor Venues, Religious Sites, Other Indoor Venues.
2. The Risk Factor – A weighted sum of the features mentioned in the sections of 3 and 4 which represents the level of risk at re-opening the school. The weights will be determined by the importance of each factor to the risk itself when a high weight represents a higher threat.

For Example, for the data in row 20 of the database:

|  |  |
| --- | --- |
| **Column Name** | **Value** |
| School | P.S 207 |
| Borough | Manhattan |
| Neighborhood | Marble Hill |
| Latitude | 40.87831 |
| Longitude | -73.90597 |
| Zip Code | 10463 |
| Cases\_Per\_1000 | 20 |
| Restaurants | 12 |
| Shops and Stores | 11 |
| Outdoor Venues | 0 |
| Religious Sites | 0 |
| Other Indoor Venues | 3 |
| Risk | 75 |

The division of the venues based on their Foursquare's categories:

|  |  |  |
| --- | --- | --- |
| **Group Name** | **Foursquare Categories** | **Total** |
| Restaurants | Pizza Place (x2), Caribbean Restaurant, Mexican Restaurant, Spanish Restaurant, Sandwich Place (x3), Seafood Restaurant, American Restaurant, Steakhouse, Café | 12 |
| Shops and Stores | Coffee Shop, Supermarket (x2), Dount Shop (x2), Department Store, Discount Store, Candy Store, Supplement Shop, Ice Cream Shop, Miscellaneous Shop | 11 |
| Outdoor Venues | - | 0 |
| Religious Sites | - | 0 |
| Other Indoor Venues | Yoga Studio, Gym (x2) | 3 |

Determining the risk value:

|  |  |  |
| --- | --- | --- |
| **Feature** | **Weight for feature** | **Feature\*Weight** |
| Cases\_Per\_1000 | 1 | 20 |
| Restaurants | 2.5 | 12\*2.5 = 30 |
| Shops and Stores | 2 | 22 |
| Outdoor Venues | 0.8 | 0 |
| Religious Sites | 2.5 | 0 |
| Other Indoor Venues | 1 | 3 |
|  | Total: | 75 |