

Which hotel to stay in New York?

(COVID-19 Outbreak)

Izidine Jaime

April 25, 2020

1. Introduction

1.1. Background

Due to the outbreak of Coronavirus in the world and the growing cases presented by World Health Organization (WHO), the United States has become the epicenter of this epidemic, with New York (NY) State being the state with the most confirmed positive cases. In order to help reduce the number of positive cases and the rapid spread of the virus (COVID-19), WHO has appointed specialist doctors from different countries to the United States, with the New York as its initial destination. However, the WHO delegation needs to find a strategic hotel to stay at.

1.2. Business Problem

The quick way of COVID-19 contagion has proved to be a major challenge for front line professionals, so the WHO delegation intends to take all precautions, staying in a Hotel located in the county with the highest number of confirmed cases of COVID-19 in NY State (to address the pandemic), but closest (less than 1.5 KM) to hospitals for first aid purposes.

1.3. Interested Audience

Clearly, the entire WHO team would be interested in this project and its accurate analysis and results. The same as anyone who is moving to New York State, would also be interested in this project in order to make a more preventive decision about which county or place to stay.

2. Data Understanding and Preparation

2.1. Data Source

In order to achieve the goal of the project we used the following data sources:

I. Geodata of Hotels and Hospitals in New York using Foursquare API

We will get all the hotels and hospitals located on the county with major confirmed cases of COVID-19 in NY State, using Foursquare API.

II. Dataset of New York State Statewide COVID-19 Testing

This dataset includes information on the number of tests of individuals for COVID-19 infection performed in New York State beginning March 1, 2020, when the first case of COVID-19 was identified in the state. The primary goal of publishing this dataset is to provide users timely information about local disease spread and reporting of positive cases.

The Dataset can be found on: <https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing/xdss-u53e>

2.2. Data Collection and Preparation

Once downloaded Data from Dataset of New York State Statewide COVID-19 Testing and read it into a *pandas* dataframe, we get the result as show in Table 1.

Table 1. New York State COVID-19 Testing Data

Test Date	County	New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
04/18/2020	Albany	34	682	305	7032
04/18/2020	Allegany	0	30	7	355
04/18/2020	Bronx	804	28823	2162	58605
04/18/2020	Broome	6	192	52	1383
04/18/2020	Cattaraugus	0	34	17	471

When we look at the size of our dataframe we see that our dataset is composed by 2976 rows and 6 columns, where the columns are distributed in:

- **Test Date** - The date the test result was processed by the NYS Electronic Clinical Laboratory Reporting System (ECLRS);

- **County** - The county of residence for the person tested;
- **New Positives** - The number of new persons tested positive for COVID-19 infection on the test date in each county;
- **Cumulative Number of Positives** - Running total for the number of persons tested positive for COVID-19 infection in each county as of the test date;
- **Total Number of Tests Performed** - The number of tests of individuals performed on the test date in each county. This total includes positives, negatives, and inconclusive results;
- **Cumulative Number of Tests Performed** - Running total for the number of tests of individuals performed in each county as of the last update to the dataset. This total includes positives, negatives, and inconclusive results.

To simplify our analysis, we rename the columns for shortnames and create a new dataframe with only the needed columns.

Table 2. First 5 rows of new table prepared for Analysis

Test Date	County	Total Positives	Tests Performed
04/18/2020	Albany	682	7032
04/18/2020	Allegany	30	355
04/18/2020	Bronx	28823	58605
04/18/2020	Broome	192	1383
04/18/2020	Cattaraugus	34	471

3. Exploratory Data Analysis

3.1. Top Counties of Tested and Confirmed Cases

If we look at Table 2, we cannot have a clear understanding of the Counties with the highest number of Tests performed and Confirmed cases, since the table only show the first 5 rows.

To make things clear, we sort the table in descending order by **Total Positives** and present the 10 first rows (see Table 3).

Table 3. Top 10 Counties by Total Positives

Test Date	County	Total Positives	Tests Performed
04/18/2020	Queens	41237	80404
04/18/2020	Kings	36482	74825
04/18/2020	Nassau	30013	69293
04/18/2020	Bronx	28823	58605
04/18/2020	Suffolk	26888	65320
04/18/2020	Westchester	23803	71741
04/18/2020	New York	18220	46204
04/18/2020	Richmond	9674	23535
04/18/2020	Rockland	9364	21949
04/18/2020	Orange	6379	18499

Now we can easily see the counties with the highest number of confirmed cases. The results suggest that the greater the number of Tests Performed, the greater the number of positive cases. Is that a direct relationship? Let's analyze the Top 15 counties by Total Positives and the Top 15 counties by Tests Performed.

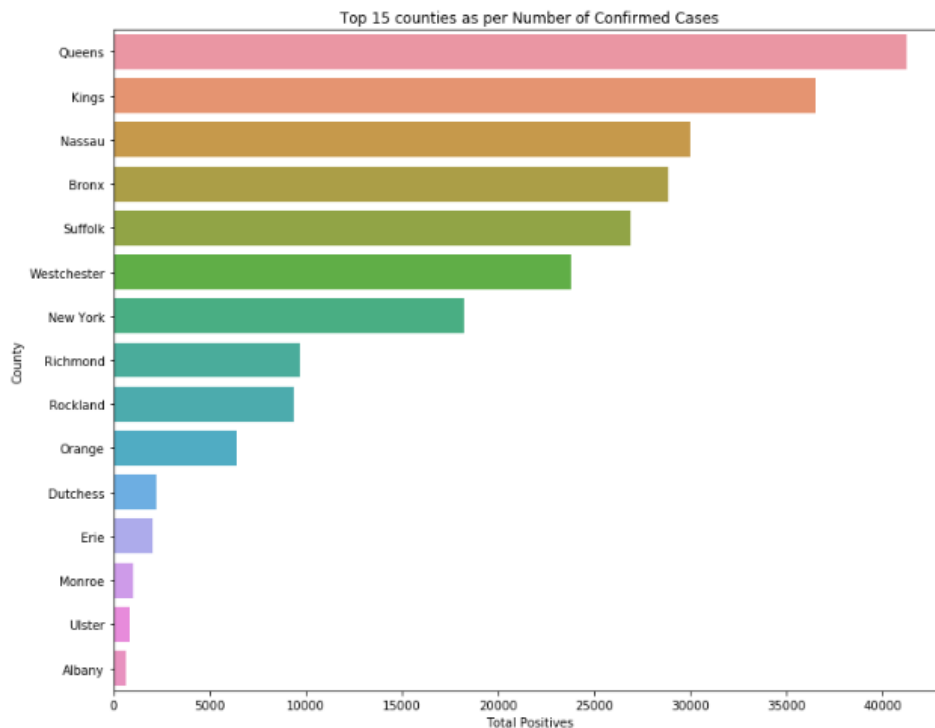


Figure 1. Top 15 counties by Total Positives

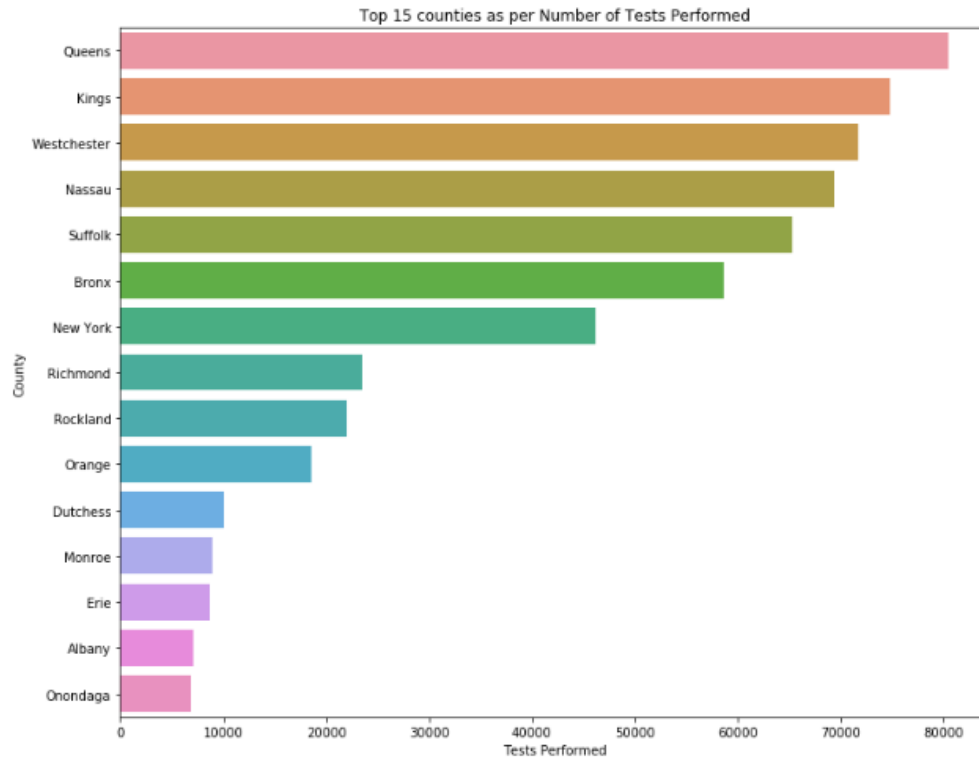


Figure 2. Top 15 counties by Tests Performed

If we evaluate figures 1 and 2, we can easily conclude that there is a direct relationship between the tests performed and the confirmed cases. **Queens** is the County with highest numbers of *Tests Performed* and *Confirmed cases*. Let's plot and analyse the Daily Increase in Number of Confirmed Cases and the Tests Performed in Queens (Figure 3).

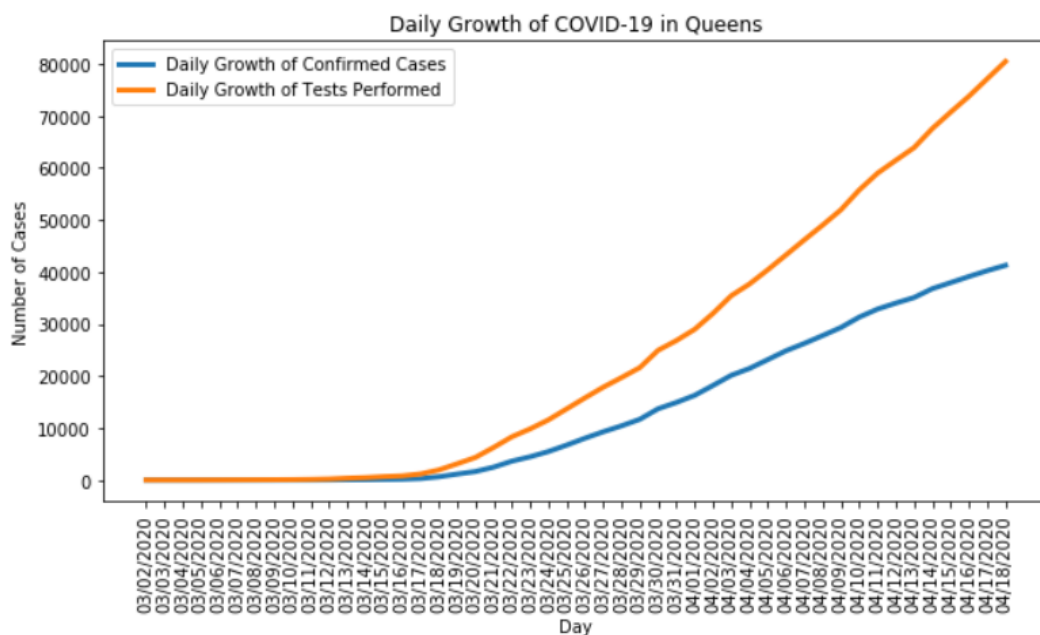


Figure 3. Daily Growth of COVID-19 in Queens

Queens is the fourth most densely populated county among **New York City's** boroughs, as well as in the United States. As you probably realize, Queens is also a borough of New York City. Looking back to our Business Problem, **Queens** match with the requirements, so let's search and explore all the hotels of Queens as well as the latitude and longitude coordinates of each Hotels using **FOURSQUARE**.

4. Load and Explore Hotels with Foursquare API

Firstly, we define our Foursquare Credentials and Version, I mean CLIENT_ID, CLIENT_SECRET and VERSION and a URL to send a request to the API to search for Hotels that is within 2.5 km from Queens coordinates. We get the results in JSON format and transform it into a *pandas* dataframe. Then we define information of interest such as *name*, *categories* and *location* to filter dataframe.

On the new dataframe we can now visualize the Hotels that are nearby Queens filtering only the relevant columns sorted by *distance*.

Table 4. Hotels that are nearby Queens

name	categories	lat	lng	distance
Ski Und Golfresort Hotel Riml Obergurgl	Apres Ski Bar	40.749071	-73.812082	1221
Bajwa Hotels	Hotel	40.758244	-73.807939	1278
Hotel Aramie	None	40.757008	-73.810927	1377
Jinfeng Hotel	Hotel	40.757871	-73.820400	2118
Sheraton Laguardia East Hotel	Music Venue	40.765380	-73.817360	2401
Navaratna Hotel In CBD	Indian Restaurant	40.756033	-73.825229	2427
Navaratna Hotel	Indian Restaurant	40.756033	-73.825229	2427
Double tree hotel	Hotel	40.753359	-73.828031	2593
Michelle Hotel	Hotel Bar	40.766609	-73.773380	2770
Sheraton Hotel Flushing	Building	40.757730	-73.829285	2810

Let's use **Folium**, a great visualization library, to create a map of Queens with the Hotels nearby, so we can better visualize the *geolocation* and *distance* of the Hotels to Queens.

4.1. Explore and Analyze Each Hotels

We already know that Queens is the county with the most confirmed cases of COVID-19, so our main goal is to find out which hotel in Queens is closest to a hospital. As we can see in figure 4, we created a Folium map, using the dataframe with data provided by Foursquare. We set the geolocation of Hotels with blue circle marker, and Queens with red circle marker. The **Ski Und Golfresort Hotel Riml Obergurgl** is the closest Hotel of **Queens** coordinates. But how close is this Hotel to Hospitals? What about other Hotels?

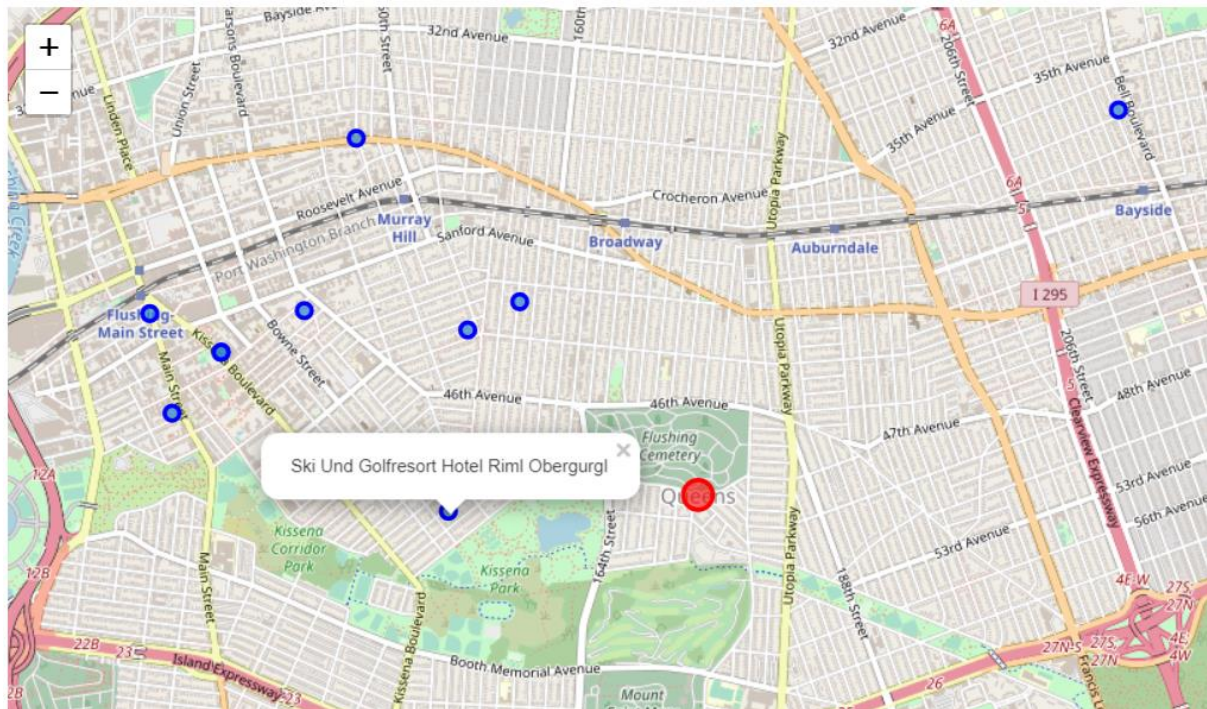


Figure 4. Hotels Geolocation nearby Queens

The approach used to answer these questions was:

1. Create a function that use Foursquare API to explore all hospitals with a proximity less than 1.5km (round = 1500) to a specific Hotel;
2. Run the function created on each Hotel and create a new dataframe
3. Group all hotel information and hospitals nearby in a single dataframe, including geolocation (Latitude, Longitude) and distance (see Figure 6);

Table 5. First 5 rows of Hotel and Hospitals Dataframe

Hotel	Hotel Latitude	Hotel Longitude	Hotel distance	Hospital	Hospital Latitude	Hospital Longitude	Hospital distance	Hospital Category
Ski Und Golfresort Hotel Riml Obergurgl	40.749071	-73.812082	1221	NewYork-Presbyterian Queens	40.747248	-73.825336	1136	Hospital
Ski Und Golfresort Hotel Riml Obergurgl	40.749071	-73.812082	1221	Flushing Hospital	40.755268	-73.816618	788	Hospital
Ski Und Golfresort Hotel Riml Obergurgl	40.749071	-73.812082	1221	Flushing Hospital Human Resources	40.755171	-73.816555	776	Hospital
Ski Und Golfresort Hotel Riml Obergurgl	40.749071	-73.812082	1221	FHMC 3W	40.755448	-73.816804	813	Hospital
Ski Und Golfresort Hotel Riml Obergurgl	40.749071	-73.812082	1221	FHMC 4N1	40.755623	-73.817022	839	Hospital

- Use Descriptive Statistical Analysis functions (*value_counts()* and *describe()*) to check how many Hospitals were returned for each Hotel and analyze the variables on the dataframe and get value of the minimum distance on *Hotel distance* column;

Table 6. Number of Hospitals 1.5 km nearby Hotels

	Number of Hospitals
Sheraton Hotel Flushing	18
Jinfeng Hotel	17
Navaratna Hotel In CBD	17
Double tree hotel	17
Navaratna Hotel	17
Ski Und Golfresort Hotel Riml Obergurgl	15
Hotel Aramie	7
Michelle Hotel	7
Sheraton Laguardia East Hotel	7
Bajwa Hotels	6

Table 7. Summary of Dataframe using description method

	Hotel Latitude	Hotel Longitude	Hotel distance	Hospital Latitude	Hospital Longitude	Hospital distance
count	128.000000	128.000000	128.000000	128.000000	128.000000	128.000000
mean	40.756591	-73.819131	2226.593750	40.753384	-73.819029	950.062500
std	0.004300	0.012901	546.100931	0.007166	0.012792	270.979871
min	40.749071	-73.829285	1221.000000	40.746428	-73.845345	324.000000
25%	40.755365	-73.828031	2118.000000	40.746760	-73.825212	768.750000
50%	40.756033	-73.825229	2427.000000	40.755171	-73.824523	1004.000000
75%	40.757871	-73.812082	2593.000000	40.755623	-73.816804	1135.250000
max	40.766609	-73.773380	2810.000000	40.779088	-73.768463	1481.000000

5. Get the Hotel and Hospital details, including names and distance.

Hotel	Hotel Latitude	Hotel Longitude	Hotel distance	Hospital	Hospital Latitude	Hospital Longitude	Hospital distance	Hospital Category
Double tree hotel	40.753359	-73.828031	2593	Flushing Hospital Pediatric Clinic	40.751156	-73.825506	324	Hospital

5. Results and Recommendations

Our results show that the **Double tree hotel** is best hotel that the WHO Delegation should go based on the requirements of the *Business Problem*. Including *Flushing Hospital Pediatric Clinic*, which is the closest Hospital to **Double tree hotel** with 324 metres, **Double tree hotel** has 17 Hospitals nearby.

You can feel free to consider another hypothesis, like to take closest Hotel of Queens coordinates: **Ski Und Golfresort Hotel Riml Obergurgl**, and evaluate the minimum distance to each Hospital nearby. Ski Und Golfresort Hotel Riml Obergurgl has 15 Hotels nearby.

6. Conclusion

The correct analytical approach depends on the business requirements for the problem. If we read objectively our business problem, we'll realize that the solution we are looking for, is based on the current status of COVID-19 in New York State. Descriptive analysis is the correct approach to solve this kind of problem, so we used it and the result was great. Another interesting point is the power of Foursquare API to deliver incredible results, including geolocation data. Foursquare API made it easy to find the best Hotel that the WHO Delegation could stay.