**Predicting a Successful Location for New York City’s First In-N-Out Burger**

Chris Melchione

April 28, 2020

1. **Introduction**
   1. **Background**

In-N-Out Burger is a major player in the hamburger industry. In fact, they have a 26% market share in the industry, which is second to only Five Guys. This is impressive considering that there is not a single In-N-Out Burger in my home state of New York. I often hear friends talk about how the first thing they would do upon arriving to California is eat at an In-N-Out Burger. This raised the question – how come there is not a location here? If there was to be one here, where should it be placed to be the most successful? Data science can be a key factor in answering this question. We can explore New York City demographics by borough, discover the most common types of venues in each borough, and determine where the biggest need is.

* 1. **Problem**

Normally, making an impactful expansion decision would be a labor-intensive process requiring hundreds of hours’ worth of planning, meetings, and research. With data science, we can create a template that is easy to understand which will tell us whether or not a specific location would be a good place to expand to or not. One key thing to remember is that this analysis is not an end point, rather it is a starting point that management can use to narrow in on results to save time.

* 1. **Interest**

Management of any company, not just In-N-Out Burger, would be interested in finding a way to simplify the decision-making process of expansion. This allows management to spend time and resources in other areas since they will be saving a drastic amount of time on the planning aspect for the perfect location.

1. **Data Acquisition and Cleaning**
   1. **Data Sources**

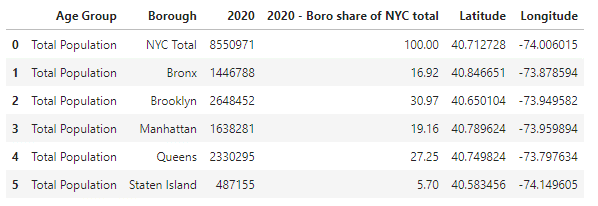
Data was sourced from the NYC Open Data website, link found [here.](https://data.cityofnewyork.us/City-Government/New-York-City-Population-by-Borough-1950-2040/xywu-7bv9/data) This is an excellent source for data regarding anything related to New York City. I chose a dataset that displayed the population from 1950 through 2040. For purposes of this project, we will only be focusing on the 2020 data.

* 1. **Data Cleaning**

Luckily, the data that was downloaded was already in a table format, so all that was left was cleaning the table to present it appropriately. Since the table displayed data each decade from 1950 through 2040, I ran a code that deleted all data except for that from the year 2020. In addition, I ran code to add the coordinates of each borough, then added a Latitude and Longitude column to the end of the table. This allowed me to create maps and run simultaneous analysis with Foursquare. I was very fortunate while cleansing the data that there were no significant outliers or missing values.

* 1. **Feature selection**

Prior to cleaning the data, the table consisted of 6 rows and 24 columns. After cleaning the data, the table consisted of 6 rows and 6 columns.



Above is the table after cleaning. You can see that it is easy to read and understand and has the added two columns at the end for latitude and longitude. This table provides an excellent breakout of the population of each New York City borough, as well as the percentage of the total population that respective borough takes up. At a quick glance, it is apparent that Brooklyn has the largest population while Staten Island has the smallest. Based on this, it would be easy to quickly say that Brooklyn would be the most ideal place to open an In-N-Out Burger, however, further analysis will show this may not necessarily be true.

1. **Methodology**
   1. **Data Analysis and Location Data**

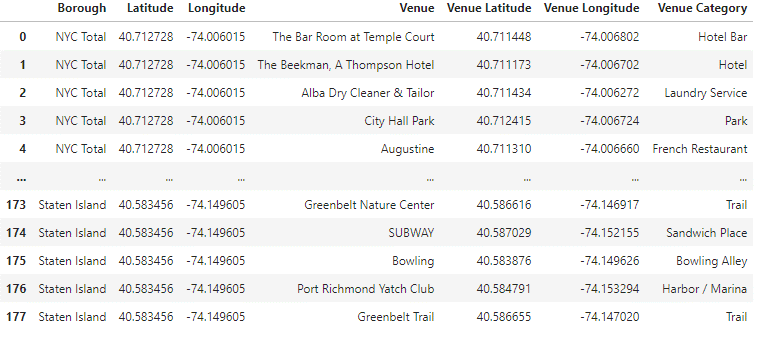
As mentioned earlier, Foursquare location data was used simultaneously with the dataset imported from the Open Data NYC website to find out more information regarding each borough of New York City. Foursquare was leveraged to find similarities between the boroughs, as well as identify the types of venues that were most common in each area. This was also an important step to help identify some of the higher traffic areas.



The code shown above was one of the most important elements of location data. This code found the top 100 venues within 500 meters of each borough. Once this data was found, it became easier to start manipulating the data to find the answers that were being searched for,

* 1. **Data Manipulation/ Exploratory Data Analysis**

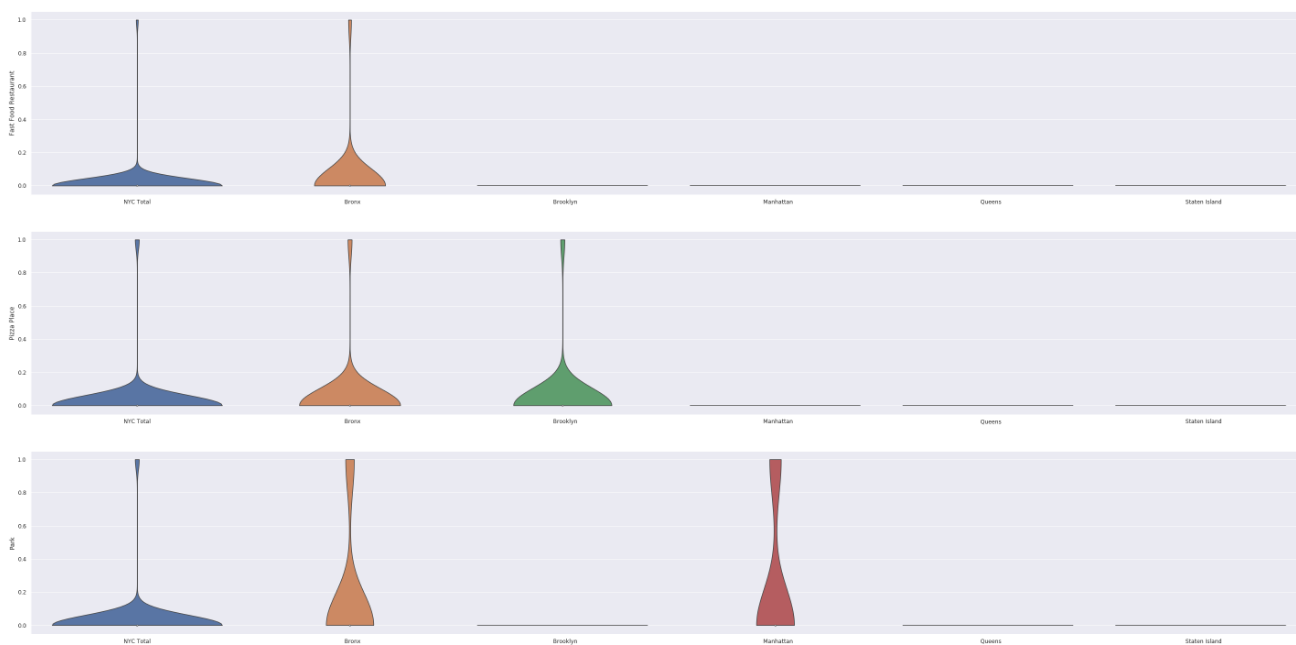
After finding the top 100 venues within 500 meters of each borough, I created a new dataframe for all of the venues. This allowed me to find the latitude and longitude of each venue, the name of the venue, and the category of the venue, as shown below.



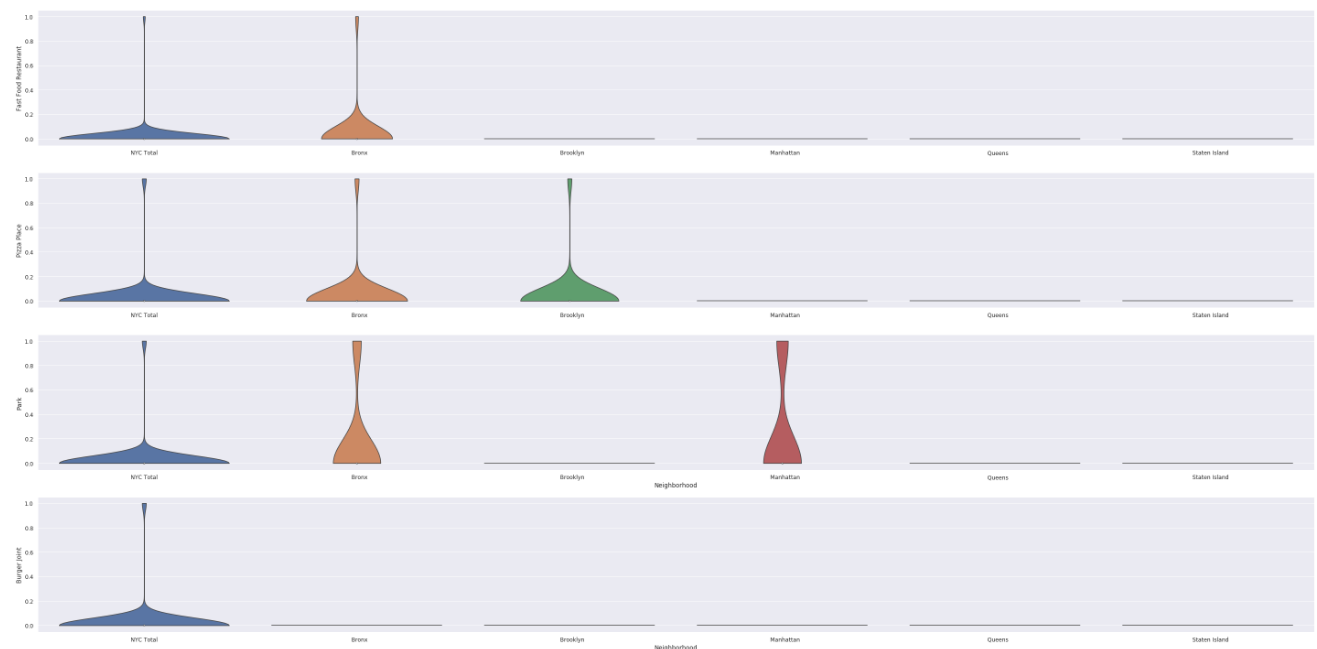
We can now see some of the most popular venues that are in each borough of New York City. After this, we can explore the data even more by grouping the total number of unique venue category types, then breaking them out as columns for each borough. With each new line of code, the data got clearer and clearer. More specific screenshots will be shown in the Results section of this report.

* 1. **Statistical Testing**

I used a lot of violin plots as visuals to display frequency distributions for different types of venues in each borough. For example, see the screenshot below which breaks out the frequency of Fast Food Restaurants, Pizza Places, and Parks in each borough



This plot tells us that the Bronx is a very high traffic area with regards to parks, pizza places, and fast food restaurants. This was an important piece of my analysis. I decided to take this one step further and display a fourth criteria which was “Burger Joints” to display how frequent burger places were with respect to any other kind of venue.



As you can see from the plot above, burger places are not very common in any borough with respect to other types of venues.

1. **Results**
   1. **The Chosen Borough**

Based on the analysis shown above, I decided that the Bronx would make the most sense for a new In-N-Out Burger. It was important to run this extra analysis rather than just basing the decision off of which borough had the highest population. This analysis shows that the Bronx has very high traffic for other fast food restaurants but does not have a lot of burger places. As a result, it can be inferred that In-N-Out would be successful because Bronx natives love fast food, and there are not many other burger places in the area.

1. **Discussion**

It's not a surprise that another fast food joint would find success in New York City. What is clear from this analysis is that the boroughs of New York City are very large, and manually researching and visiting locations would have been an incredibly time-consuming task. With this data, we now have a starting point and we know where to look first. We have made inferences along the way in this project by determining that the Bronx would be the best fit for an In-N-Out Burger. However, there is no right or wrong answer here. The role of data analysis in this specific project is to steer a course of action for a new In-N-Out location, to meet the criteria of being in boroughs that have other abundant venues, and to narrow the search down to just a few boroughs rather than all five. We have successfully accomplished all of these tasks.

1. **Conclusion**

There are countless numbers of ways that this analysis could have been performed, which is the beauty of data science. Anyone can recreate this project and tweak it to highlight other facts and features. I chose to approach the project in this way because in my opinion it was concise, kept things as simple as possible, and reached an answer that could be supported by the analysis. Keep in mind that these results are not necessarily an end point, but instead a starting point that will allow management to have a head start in finding the perfect location.

Without leveraging data to provide insight into decisions like these, the process would have taken significantly longer. Another benefit of this is that we now use this notebook as a template if we wanted to expand into other new areas.