

# **Coursera**

# **Capstone Project**

## **REPORT**

**Analyzing potential public  
working spaces in New York**

**Yaeji Myung**  
**June 2020**

# Introduction

As a consequence of the current pandemic, millions of people are facing the challenges of working from home. Many large companies like Google and Apple have extended their work from home policy to the end of the year, while others are considering extending their policy permanently. And while working from home has worked for certain employees, “to jump to the immediate conclusion that we should now reinvent workplace strategies overnight and leave a swath of our colleagues at home indefinitely” (Tim Oldman, Leesman) may not be the right approach. As cities begin to re-open, some employees find themselves permanently anchored to the home.

## **Business problem**

For this project I’m interested in exploring alternate public work spaces for employees outside of their home, mainly coffee shops. I have chosen coffee shops since they exist in abundance and can easily be used as a temporary ‘work’ or ‘meet’ space by any individual. This is of course, assuming things alleviate and individuals are able to go outside in a safe manner. The project will focus on neighborhoods in Manhattan, New York City.

## **Stakeholders**

The stakeholders for this project would be individuals or employees looking for alternate spaces, other than their homes, that can support their ability to work in a comfortable and optimal manner. The second would be developers looking to invest and open hybrid coffee/work space in new neighborhoods.

## Data Description:

- **Neighborhood Data**, including neighborhood names in Manhattan as well as latitude and longitude coordinates.
  - Source: “2014 New York City Neighborhood Names” ([https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)). Point file published by the NYC Department of City Planning in 2014.
- **Venue data**, including names, addresses, neighborhoods, venue description (in this case ‘coffee shop’), reviews and rating. We’ll be scrubbing the reviews\* to make sure the coffee shops are optimal to work in, looking for keywords such as ‘wifi’, ‘work’, ‘laptops’ and so forth.
  - Source: Foursquare API
  - *\* Unfortunately, with the personal account on Foursquare developer, I only have access to two photos and two tips (reviews). Since we don’t have access to the complete database, searching using keywords will not be completely accurate. For the purposes of this project, we’ll use the data that is available as a case study and hope that some initial findings will help us get the support of stakeholders to then purchase access to the rest of the database.*

# Methodology

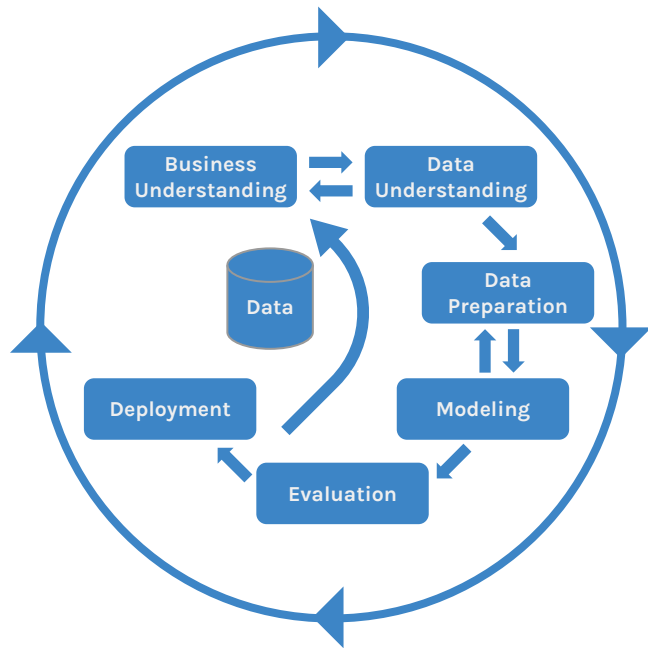
To structure this process, we will use the Cross Industry Process for Data Mining (CRISP-DM) methodology. The six steps are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

## Business Understanding

Typically, before beginning the process we would have gathered the major stakeholders together to understand the business problem. What are the main goals at hand? What are we hoping to accomplish and why? Since we don't have real stakeholders, we will move forward assuming we have a proper business understanding (outlined in page 2).

## Data Understanding

In this case we have identified our two data sources (outlined in page 3): the neighborhood data from the NYC Department of City Planning, and the Foursquare API. For this project, we were required to use the Foursquare API but in a typical project we would explore different data sets to make sure we're using the most optimal data source.



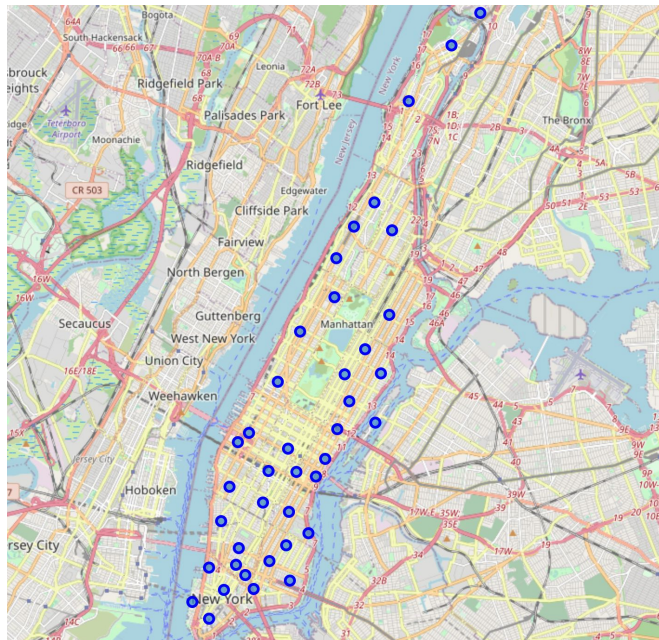
# Methodology

## Data Preparation

We will first download and prepare the first data set, neighborhoods in New York ([https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)). We can see how the data is structured, each 'neighborhood' has an associated 'id', 'borough', and 'coordinates' showing latitude and longitude. We will use jupyter notebook (in IBM Watson Studio) to import this data and transform it into a pandas dataframe.

We then create a table with four columns: 'Borough', 'Neighborhood', 'Latitude', and 'Longitude'. Each row is a different neighborhood. We can see that our current data set includes 5 boroughs and 306 neighborhoods. For our project, we will only focus on venues in Manhattan so we'll manipulate the table to remove the other 4 boroughs.

We can use *geocoder* to locate the coordinates of Manhattan and *folium* to make a quick visualization of the data. *Folium* allows us to easily overlay on a map each of the neighborhoods in Manhattan.



Manhattan

# Methodology

## Data Preparation

Next we want to access the Foursquare API, since we've set up a developer account we can use our 'CLIENT\_ID' and 'CLIENT\_SECRET' to access it. We create a GET request URL to access the venue data from each of the neighborhoods. For each of the venues, Foursquare API returns the venue 'id', 'name', 'address', 'categories', 'latitude', 'longitude', and other information.

We transform this data into another pandas dataframe. The result would be a table showing the venue name, venue id, category, latitude, and longitude. We can append this table to our first neighborhood dataframe so that for every venue we also have the corresponding neighborhood name, latitude, and longitude.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	4b4429abf964a52037f225e3	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	4baf59e8f964a520a6f93be3	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	4b79cc46f964a520c5122fe3	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	55f81cd2498ee903149fcc64	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	4b5357adf964a520319827e3	40.877136	-73.906666	Donut Shop

# Methodology

Data  
Preparation

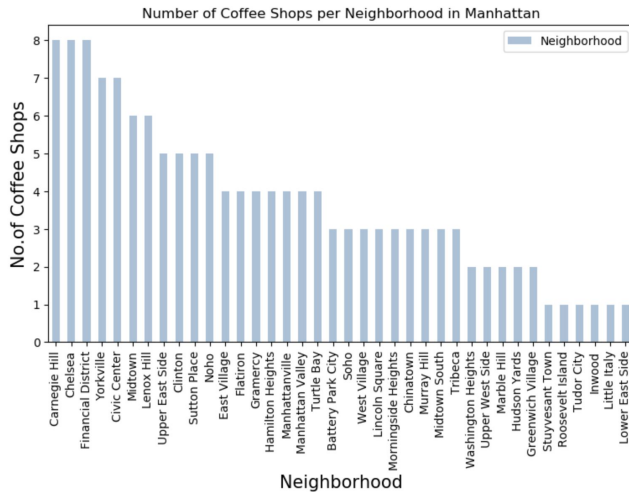
Modeling

## Data Preparation

We'll want to filter the data so that we are only seeing the venues with the category for 'coffee shop'. We now know that we have 140 coffee shops in Manhattan (within the Foursquare API).

## Modeling

Let's go ahead and start modeling this data. We can create a quick bar graph to see the distribution of coffee shops per neighborhood. Carnegie Hill, Chelsea, and Financial District have the most number of coffee shops, eight for each neighborhood. We'll focus on these three neighborhoods for the rest of the analysis.



# Methodology

Data  
Preparation

Modeling

## Data Preparation

For each of the three neighborhoods, we'll use the 'venue\_id' to make additional GET request URL to access the tips for each of the venues. Tips are the reviews that Foursquare users have submitted per venue. We'll filter our request to retrieve the text of the review as well as some additional information about the user.

Remember, we can only retrieve two tips per venue. By the end of this process, for each of the neighborhoods we'll have a total of 16 reviews (two per coffee shop).

*\* Note: One of the venues for Carnegie did not return any tips. This may be because there have been 0 reviews submitted for this venue.*

	venue_id	text
0	599d7c7e61f0700d414c282b	Easily the best drip coffee on the UES. Shop is very small with minimal seating. Good source for reasonably priced coffee beans.
1	599d7c7e61f0700d414c282b	Nice coffee spot, great coffee and friendly barista! Try their raw bars Lenka - so yummy!
0	51ae935b7dd2de27a217609e	The Irving Farm coffee is excellent, and the rustic aesthetic is a perfect compliment to a robust brew. The service staff is personable and chill, and the cozy environment is the perfect city escape.
1	51ae935b7dd2de27a217609e	Just an all around great neighborhood spot for yummy pastries, good fresh roast coffee, good atmosphere/customer service, coziness, and wifi...within 60 sec walk from home! score.
0	5abb96b31f8ed641d150eafa	Finally! A new specialty coffee shop close to the UES 86th subway (on the same location as the recently closed Petite Shell). Everything tasted terrific so far, from espresso, cortado to batch brew.
1	5abb96b31f8ed641d150eafa	Very classy but welcoming interior. It was surprisingly empty and would be an excellent alternative to Cafe Jax if you like working at coffee shops.
0	57fd457a498e39cd34012586	Decent sized, friendly staff. Perfect stop for the morning commute.
1	57fd457a498e39cd34012586	Even though the official address is listed as 245 E. 93rd St., it's actually on second Avenue between 93rd and 94th Streets.
0	5330323b498e2836f173cf9e	The space is nice, and I easily found and used outlets for my devices. Enjoy a cold brew with a Magpies tart! Delicious.
1	5330323b498e2836f173cf9e	New to the locality - it's spacious, and has plenty of seating by the glass windows - for those who love their coffee with a view!
0	4da39a219935a0933d0bcd6f	The girls who work here are extremely friendly and are so pleasant when I stop in at 7:15 on my way to work. The best(less than \$5) bacon egg&cheese in the neighborhood! Great coffee and smoothies
1	4da39a219935a0933d0bcd6f	Fast service from very friendly staff. When I struggle to get going on a weekend morning, they save me. Prices are a little high, but also larger size cups than nearby places.
0	552a7fbb498e1e8662ac8146	Good local spot to pick up a quick breakfast scone and coffee or a smoothie for lunch. Very clean place with friendly help behind the counters.
1	552a7fbb498e1e8662ac8146	Great variety of juices, smoothies and açai bowls after my workouts at Synergy. They also have frozen yogurt for the kids younger and older. :)



We have now collected all of the tips for coffee shops in Carnegie Hill, Chelsea, and Financial District. We can create a visualization using word cloud in python. We will be using 'stopwords' in python to automatically remove any commonly used words such as 'the', 'a', 'in' so they are not shown in the word cloud.

We collate each of the rows from our dataframe into a text file (.txt), from there we can easily create a word cloud. The visualization shows the most frequent words used to review the coffee shops in each neighborhood. The larger the font, the more frequent it is used.

[illegible]

# Methodology

Modeling

## Modeling

The word cloud is helpful to see, but we want to make sure we're selecting coffee shops that have been reviewed by users as optimal to work in. Let's filter through our dataframe to find reviews using keywords such as 'work', 'wifi', 'laptops', and so forth.

For coffee shops in Carnegie Hill, 4 of the reviews match the list of keywords we put together. Chelsea only returns 1 match and Financial District returns 0 matches. Taking a quick look at the reviews from Carnegie Hill we can see that these coffee shops can be used to work from.

```
In [241]: carnegie_key_list = carnegie_text_list.loc[(carnegie_text_list['Words'] == 'work') |
(carnegie_text_list['Words'] == 'wifi') |
(carnegie_text_list['Words'] == 'internet') |
(carnegie_text_list['Words'] == 'meet') |
(carnegie_text_list['Words'] == 'study') |
(carnegie_text_list['Words'] == 'laptop') |
(carnegie_text_list['Words'] == 'outlet') |
(carnegie_text_list['Words'] == 'outlets')]

carnegie_key_list
```

Out[241]:

	Words	Count
44	work	2
87	wifi	1
148	outlets	1

## Reviews

“...good atmosphere/customer service, coziness, and **wifi**...”

“...excellent alternative [...] if you **like working at coffee shops.**”

“...**easily found and used outlets for my devices.**”

text

Just an all around great neighborhood spot for yummy pastries, good fresh roast coffee, good atmosphere/customer service, coziness, and wifi...within 60 sec walk from home! score.

Very classy but welcoming interior. It was surprisingly empty and would be an excellent alternative to Cafe Jax if you like working at coffee shops.

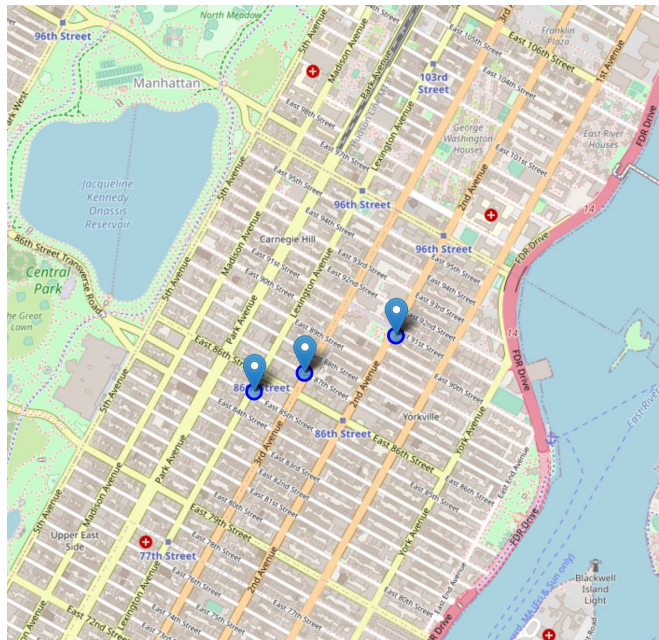
The space is nice, and I easily found and used outlets for my devices. Enjoy a cold brew with a Magpies tart! Delicious.

# Methodology

## Modeling

Since each of the reviews in our dataframe is matched to the 'Venue\_ID', we can see where these reviews are coming from. For Carnegie Hill, the coffee shops that are ideal to work from are 'Variety Coffee Roasters', 'Starbucks', and 'DTUT'.

Since we have the longitude and latitude of each of these venues, we can also map them out using *folium*.



Carnegie Hill

# Results

## Results

Out of all of the venues in Manhattan (within the Foursquare database), we found those that are labeled as 'coffee shop' and then managed to select the ones that were viable to work from, based on foursquare reviews. We filtered through the reviews using keywords to find whether users have previously worked from this venue. Based on the results, we were able to find three viable options for people to work from in Carnegie Hill.

Variety Coffee Roasters



Starbucks



DTUT



# Discussion

While the results of this project gave us some matches, the sample size was much too small to be certain of our findings. The biggest limitation would be the two tips per venue, as allowed by the Foursquare API. It is unclear whether Foursquare returns the highest liked reviews or simply randomizes the results. Either way, not every reviewer gives extensive description of the venue - so trying to determine which coffee shops are ideal to work from is quite limited.

As a case study, this project was successful since we were able to build the framework and the process to retrieve this information. If we could increase the number of reviews, the notebook is set up so that we can easily comb through the additional data and give more recommendations for optimal work spaces.

# Conclusion

Deployment

Business  
Understanding

In conclusion, for those employees who are looking for alternate work spaces in the office, we were able to find three optimal locations in Carnegie Hill. I don't doubt that there are plenty of other locations that we were not able to access through the Foursquare API.

Before deploying this solution, it is crucial to present this to our stakeholders to see if there is interest in this project. Using CRISP-DM, at this point we would cycle back to 'Business Understanding' and reassess our business goals and objectives.