Melvin Capital Investment into Manhattan, NYC

Nnamdi Osia

April 2021

1. Introduction

1.1 Background

Melvin Capital Management LP is an American investment management firm based in New York City. Melvin Capital invests primarily in tech and consumer stocks and is reported to have $8 billion in assets under management (AUM) as of January 2021. Melvin Capital came to popularity recently during the GameStop short squeeze of 2021, they sustained losses of 53%. To recoup their losses, Melvin Capital has been contemplating investing in the in the Manhattan Burroughs of New York City which it is a part of.

Manhattan, known regionally as the City and the urban core of the New York metropolitan area, is the most densely populated of the five boroughs of New York City, and coextensive with the County of New York, one of the original counties of the U.S. state of New York. Manhattan serves as the city's economic and administrative center, cultural identifier, and historical birthplace. The borough consists mostly of Manhattan Island, bounded by the Hudson, East, and Harlem rivers; as well as several small adjacent islands. Manhattan additionally contains Marble Hill, a small neighborhood now on the U.S. mainland, separated from the rest of Manhattan by the Harlem Ship Canal and later connected using landfill to the Bronx. Manhattan Island is divided into three informally bounded components, each aligned with the borough's long axis: Lower, Midtown, and Upper Manhattan.

Melvin Capital is looking to produce a list of venues by popularity so that they are able to determine which venue are needed and which venues are commonplace in New York City Burroughs.

1.2 Problem

Data that might contribute to determining which venues are lacking and which are commonplace include New York City data that contains borough, neighborhoods along with their latitudes and longitudes, population data from scraping Wikipedia, and venue information from foursquare API.

1.3 Interest

Obviously, hedge funds and venture capitalists would be interested in which venue are needed and which venues are commonplace. Also, individuals looking to startup business in New York City may also be interested in this project and data.

2. Data

2.1 Data Sources

We will be collecting data from following sources:

New York City data that contains borough, neighborhoods along with their latitudes and longitudes. Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This can be found online at https://cocl.us/new_york_dataset or https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json.

We can get the population data from scarping Wikipedia: https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City We are going to go through each of the links of neighborhood and find the population of each of them.

For additional insights, we can fetch venue information from foursquare API (https://api.foursquare.com/v2/venues).

2.2 Data Acquisition and Cleaning

The necessary librarys will need to be utilized and imported for data analysis. This includes 'Pandas', 'json','geopy', 'matplotlib', 'sklearn.cluster' and 'folium'. Numpy is used as a library to handle data in a vectorized manner. Pandas library is used for data analysis. Json library is used to handle JSON files. Nominatim is imported from geopy to help convert addresses into latitude and longitude values. The requests library is imported to handle requests used in FourSquare API. The json notmalize library is used to tranform JSON file into a pandas dataframe. Matplotlib is used for associated plotting modules and sklearn is used to import KMeans clustering. Finally, folium is used for map rendering.

We start of our data pre-processing by first downloading the data. We load and explore the data and define variable that includes key features. We then transform the data into a pandas frame. We loop through the data and fill the dataframe one loop at a time. We perform data validation to ensure all 5 boroughs and 306 neighborhoods contained. We finish by printing geographical coordinates of NYC.

3. Exploratory Data Analysis

3.1 Analyze Nearby Venues

We will be collecting data from following sources:

New York City data that contains borough, neighborhoods along with their latitudes and longitudes. Neighborhood

We begin by finding nearby venues for each neighborhood and create a new dataframe. Because we will be evaluating data for a bunch of neighborhoods, we create a function that will find this information within a predetermined radius using the name of the neighborhood and latitude and longitude of the neighborhood.

After we have found nearby venues for each neighborhood, we then check how many venues are returned for each neighborhood. Finally, we find out how many unique categories exist.

3.2 Find Top Common Venues

We setup the dataframe we will use for analyses similar to the previous dataframes we have setup earlier in this project. We add neighborhood column back to dataframe and move it to the first column. We then group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Finally, we print each neighborhood with top 5 common venues. We do the same for top 10 common venues.

3.3 Clustering

Kmeans clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of kmeans is to group data points into distinct non-overlapping subgroups.

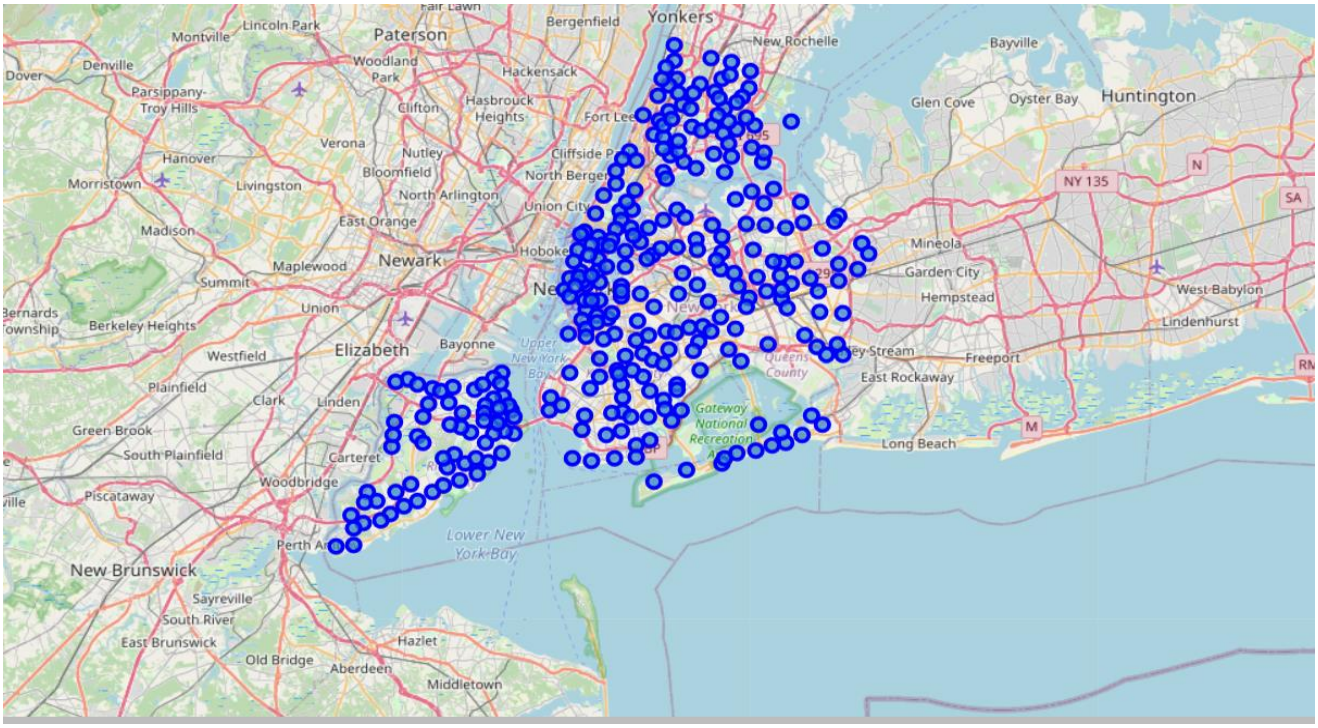
In the step of data analyses in machine learning, we often need to prepare our data in specific ways before feeding into a machine learning model. One of the examples is to perform a One-Hot encoding on categorical data. One Hot Encoding is a process in the data processing that is applied to categorical data, to convert it into a binary vector representation for use in machine learning algorithms.

4. Results

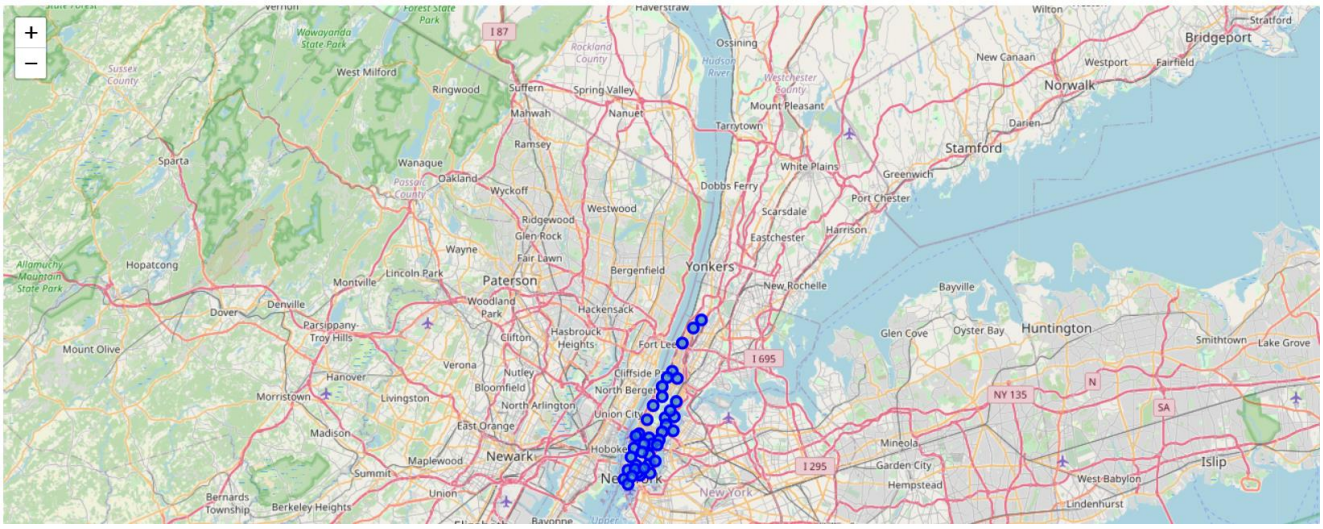
During our data analysis, we verified that the dataframe we used showed 5 boroughs and 306 neighborhoods. The geograpical coordinate of New York City are 40.7127281, -74.0060152. A quick Google Maps search will verify this location as City Hall Park, arguably the epicenter of NYC.

Map for 40.7127281, -74.0060152

Using Folium, we can create map and superimpose cities in order to visualize data as well. We create the map of New York using latitude and longitude values. We add markers to the map and cluster neighborhoods in NYC and the borough of Manhattan. We display the geographical coordinates for Manhattan as well.



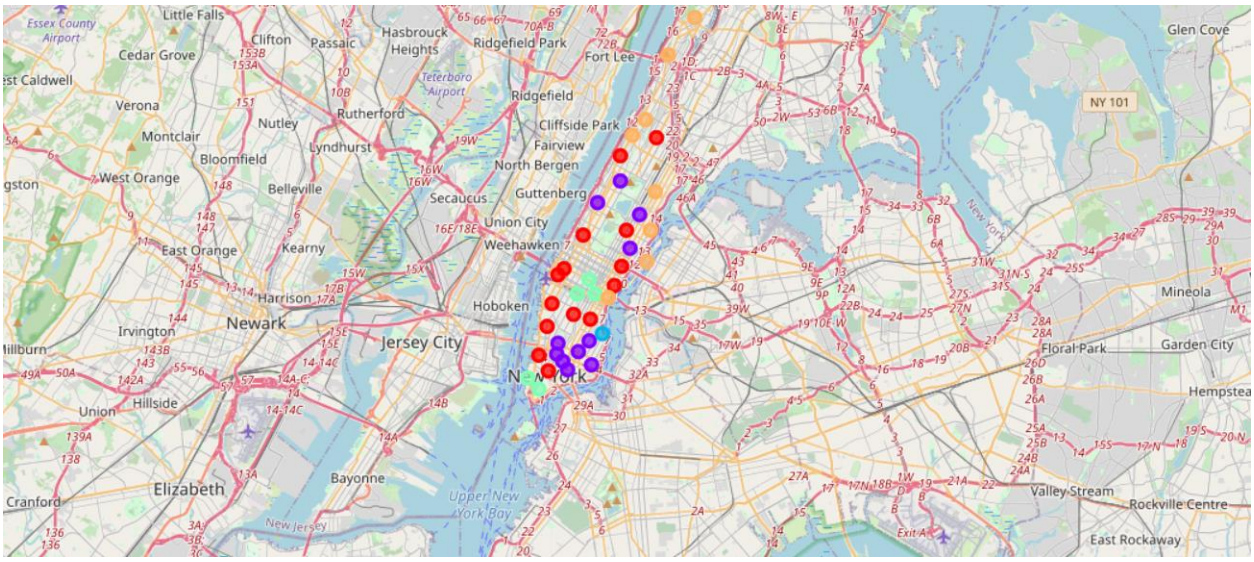
The geograpical coordinate of Manhattan are 40.7896239, -73.9598939.



Using FourSquare API, we are able to visualize some additional data. After we make the request to have a better picture of what venues are nearby, we are able to break down into the categories of venues found nearby. Using FourSquare to check the API, we are able to further explore these neighborhoods of Manhattan.

Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea
Greenwich Village
East Village
Lower East Side
Tribeca
Little Italy
Soho
West Village
Manhattan Valley
Morningside Heights
Gramercy
Battery Park City
Financial District
Carnegie Hill
Noho
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards

Further analysis shows us a total number of about 3235 venues in Manhattan alone. Out of the 3235 venues, 325 were found to be unique. For each neighborhood, we print the top 5 most co mmon venues. Our analysis shows us that Hotel, Coffee Shops, Cafes, Italian and American Re staurants were the most popular venues. Doing a cluster analysis will give us some additional i nformation and takeaways.



Out of all the cluster examined, there is one cluster in Stuyvesant which is unique, due to the v enues located here that are not associated with all of New York. These venues include a Bar, P ark, Coffee Shop, Baseball Field, Pet Service, Gas Station, Farmers Market, Bistro, Gym / Fitn ess Center and Cocktail Bar.

5. Conclusion

Based on our analysis, I think that we can conclude that a great investment idea for Melvin Capital would be one of the following: a Bar, Park, Coffee Shop, Baseball Field, Pet Service, Gas Station, Farmers Market, Bistro, Gym / Fitness Center and Cocktail Bar. This will offer some exclusivity as these venues are rare in other parts of NYC.