# Finding an optimum location to open a tourist-based café in Manhattan

## Introduction: Business Problem & Background

In this project we will try to find an optimal location to open a new café in Manhattan, New York.

This café is a multinational chain and is known to be loved by customers. The owner of the cafe wants to open in Manhattan and wants to check what the ideal location to open the cafe is. This café is targeted for tourists and hence would it would be optimum to locate it where there are more tourist attractions nearby but at the same time are not overly crowded by cafes.

Hence, it's important to find locations that have enough tourist attractions as a first parameter. Once identified, these locations will have to be checked for existing cafes.

We will try and create location clusters which are most optimal for the required problem. High tourist points + low cafes is an optimum cluster of locations where the owner should open the cafe

## Data – Approach and Requirements

We are going to use the New York data (used in Week 3) + venues data from Foursquare to help us reach the desirable solution

Based on definition of our problem, factors that will influence our decision are:

- number of tourist attractions in the neighborhood – tourist attractions here are defined as one of these - 'Memorial Site', 'Museum', 'Garden', 'Art Gallery', 'Art Museum', 'Historic Site', 'History Museum', 'Monument / Landmark'
- existing number of cafés in these neighborhoods

Following data sources will be needed to extract/generate the required information:

- Manhattan locations will be extracted from existing data shared during the previous courses
- Various venues will be explored and tourist places + cafes will be explored in every neighborhood using Foursquare API
- coordinate of New York for the initial map will be hardcoded from data available online
- Initially top tourist areas will be identified, post which café data will be analyzed and ideal locations will be found

## Methodology

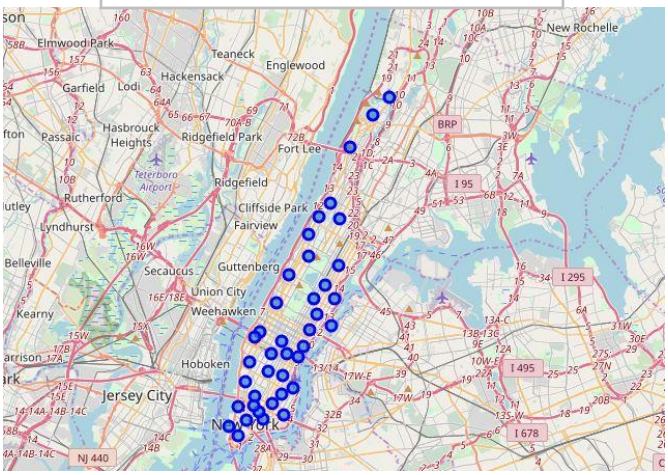
Initially started with Manhattan data and the various venue categories mentioned in the Foursquare API. After a qualitative study of the categories identified these as the tourism venues - 'Memorial Site', 'Museum', 'Garden', 'Art Gallery', 'Art Museum', 'Historic Site', 'History Museum', 'Monument / Landmark'

Once these were identified cleaned the data to find locations which had at least 1 tourism spot. Post this I used K means clustering to cluster these locations based on the number of cafes present and the number of tourist attractions present in each location. This helped identify the most optimal location

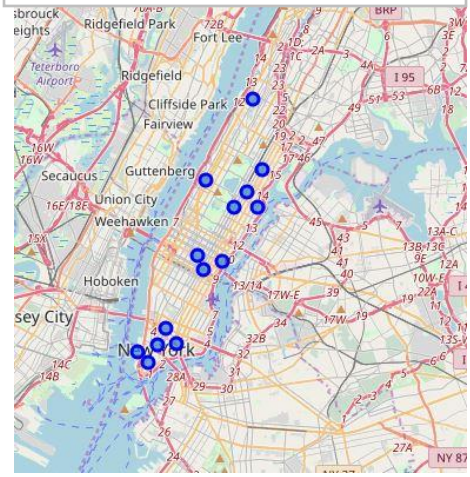
## Results

As a first step I identified locations in Manhattan which had tourist attractions nearby.

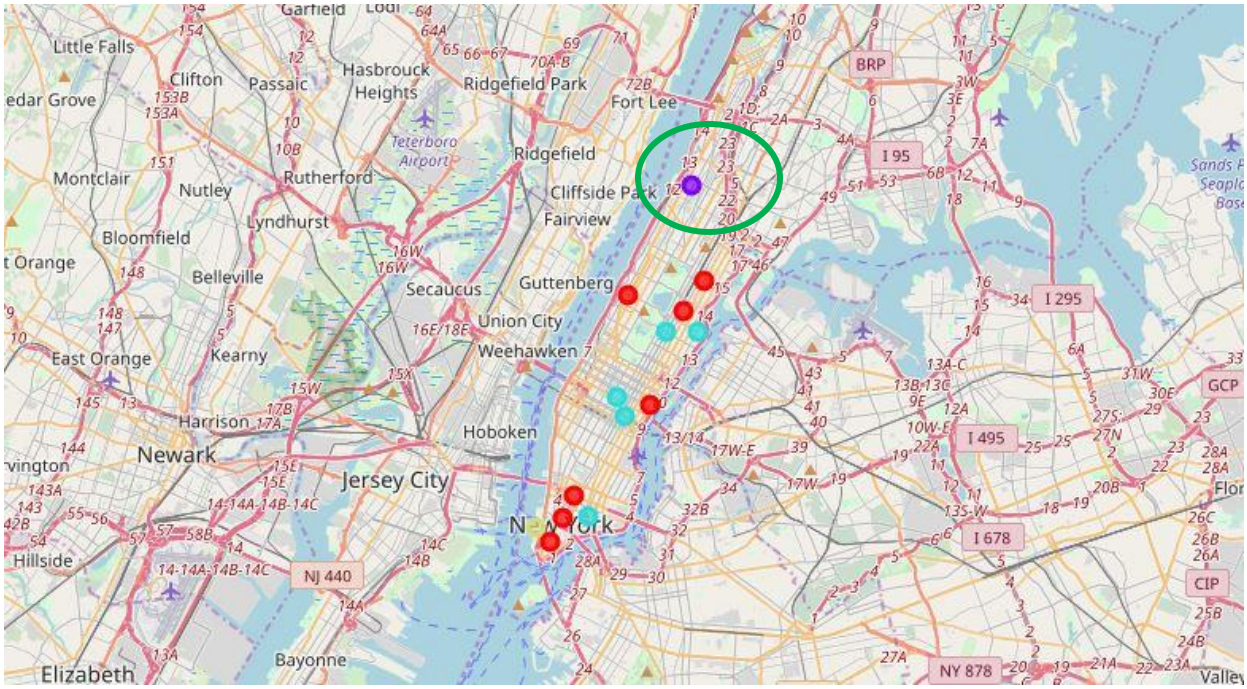
All Manhattan locations



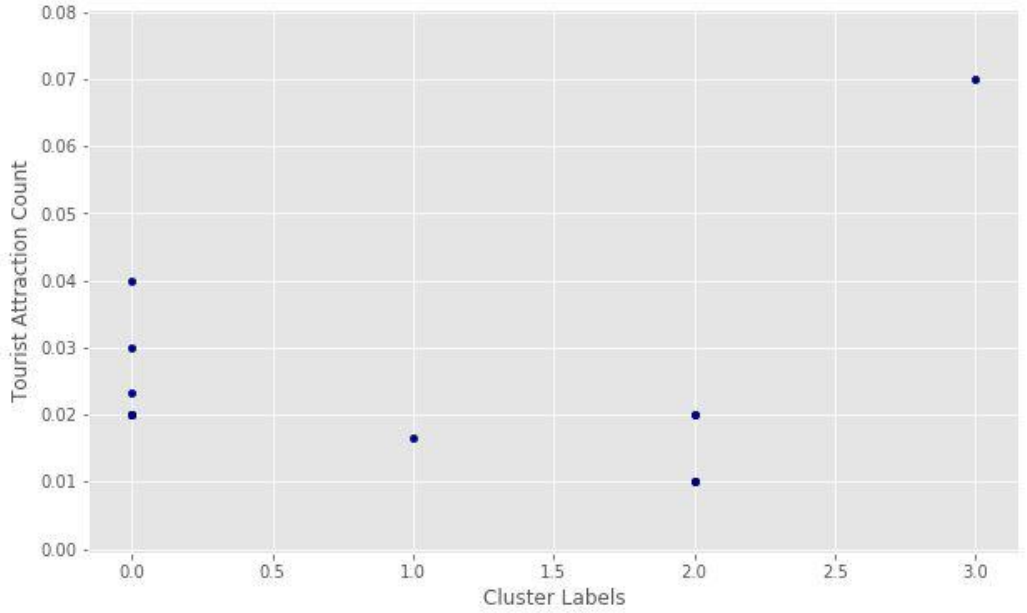

Locations with Tourist Attractions

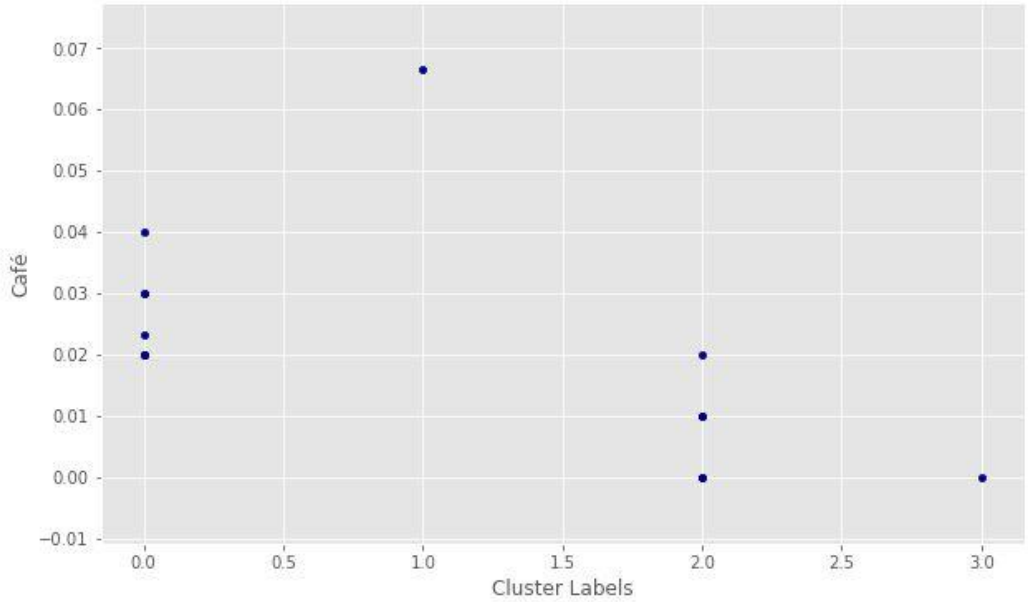



Finally, the clustering algorithm was used to create 4 clusters as below:




The location encircled in green resulted in the most optimal cluster – high no. of tourist spots + low number of cafes




Based on the above scatter plots we can thus define the clusters as -

Cluster 0 - Medium number of tourist attractions and medium number of cafes

Cluster 1 - Large Number of Cafes but low tourist attractions

Cluster 2 - Less tourist attraction and less cafes

Cluster 3 - Large number of tourist attraction and very few cafes

Thus, Battery Park resulted in the most optimal cluster

| | Cluster Labels | Neighborhood | Café | Tourist Attraction Count |
|---|---|---|---|---|
| 0 | 3 | Battery Park City | 0.000000 | 0.070000 |
| 1 | 0 | Carnegie Hill | 0.040000 | 0.020000 |
| 2 | 2 | Chinatown | 0.000000 | 0.020000 |
| 3 | 0 | Civic Center | 0.020000 | 0.020000 |
| 4 | 0 | East Harlem | 0.023256 | 0.023256 |
| 5 | 0 | Financial District | 0.030000 | 0.040000 |
| 6 | 1 | Hamilton Heights | 0.066667 | 0.016667 |
| 7 | 2 | Midtown | 0.020000 | 0.010000 |
| 8 | 2 | Murray Hill | 0.000000 | 0.010000 |
| 9 | 0 | Soho | 0.020000 | 0.020000 |
| 10 | 0 | Turtle Bay | 0.030000 | 0.020000 |
| 11 | 2 | Upper East Side | 0.010000 | 0.020000 |
| 12 | 0 | Upper West Side | 0.020000 | 0.030000 |
| 13 | 2 | Yorkville | 0.010000 | 0.010000 |

## Discussions & Recommendations

Thus, based on the above analysis we suggest the owner should open a new café at Battery Park. Within Battery Park, locations close to the top visited tourist spots are optimal

## Conclusion

Thus, location data along with clustering have been successfully used to identify the optimum location to open a new business. This proves the worth of data science techniques which can be applied to a large number of more deeper problems.

Further analysis in this could be –

- Identifying the exact address to open the café based on cost of property, etc
- Studying tourist trends, preferences to decide the menu of the café
- Creating café theme based on maximum kind of tourists likely to visit.

The applications of data science techniques are numerous and provide great value to solve a large number of problems in every day life.