

# New York City: Where to live for food-loving tourists

Kevin Teo

## 1. Introduction

### 1.1 Problem statement

A common problem that I often face when travelling is having not enough time to try all the amazing food the city has to offer. For a city as big as New York, tourists might not have adequate time to visit all five boroughs to eat the amazing food. This paper aims to investigate 3 different food categories: pizza shops, cafés and American diners in the Manhattan borough and determine which neighborhood offers the best locations for each food category, thereby providing information to potential foodies on where to stay to be near to the food of their choice.

### 1.2 Background

The City of New York is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 784 km<sup>2</sup>, New York is the most densely populated major city in the United States (Maciag, 2013). New York City consists of five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island. Almost all ethnic cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods (Zelinsky, 1985). In New York City there are over 12 000 bodegas, delis and groceries and many among them are open 24 hours a day, 7 days a week.

## 2. Data

### 2.1 New York City borough and neighborhoods

As mentioned previously, New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, the datasheet retrieved from [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) was used, which contains a data set of all 306 neighborhoods. Further data processing and cleaning was done to isolate the neighborhoods in Manhattan, and the longitude and latitude of the neighborhoods of Manhattan was retrieved using the geopy library.

### 2.2 Restaurant details

By using foursquare and their API, restaurants are queried and their relevant information, such as name, rating and location, was used in conjunction with the Manhattan database to develop a comprehensive set of restaurants and their proximity to one another as well as their general neighborhood location.

With simple machine learning tools, the restaurants were segmented throughout the Manhattan neighborhood and will serve as information to tourists on where to stay in relation to the food options that they are interested in.

## 3. Methodology

### 3.1 Data retrieval and cleaning

The first task for this study was to gather the neighborhood details, such as name and coordinates. With the information read as a json data, the data frame columns were defined and instantiated as a pandas data frame named *dfNY*. The coordinates were retrieved by using the feature key in the *dfNY* data frame. Next, the data was sliced to obtain the neighborhood details for the borough of Manhattan. This data frame is named *mandf*.

### 3.2 Collection of foursquare data

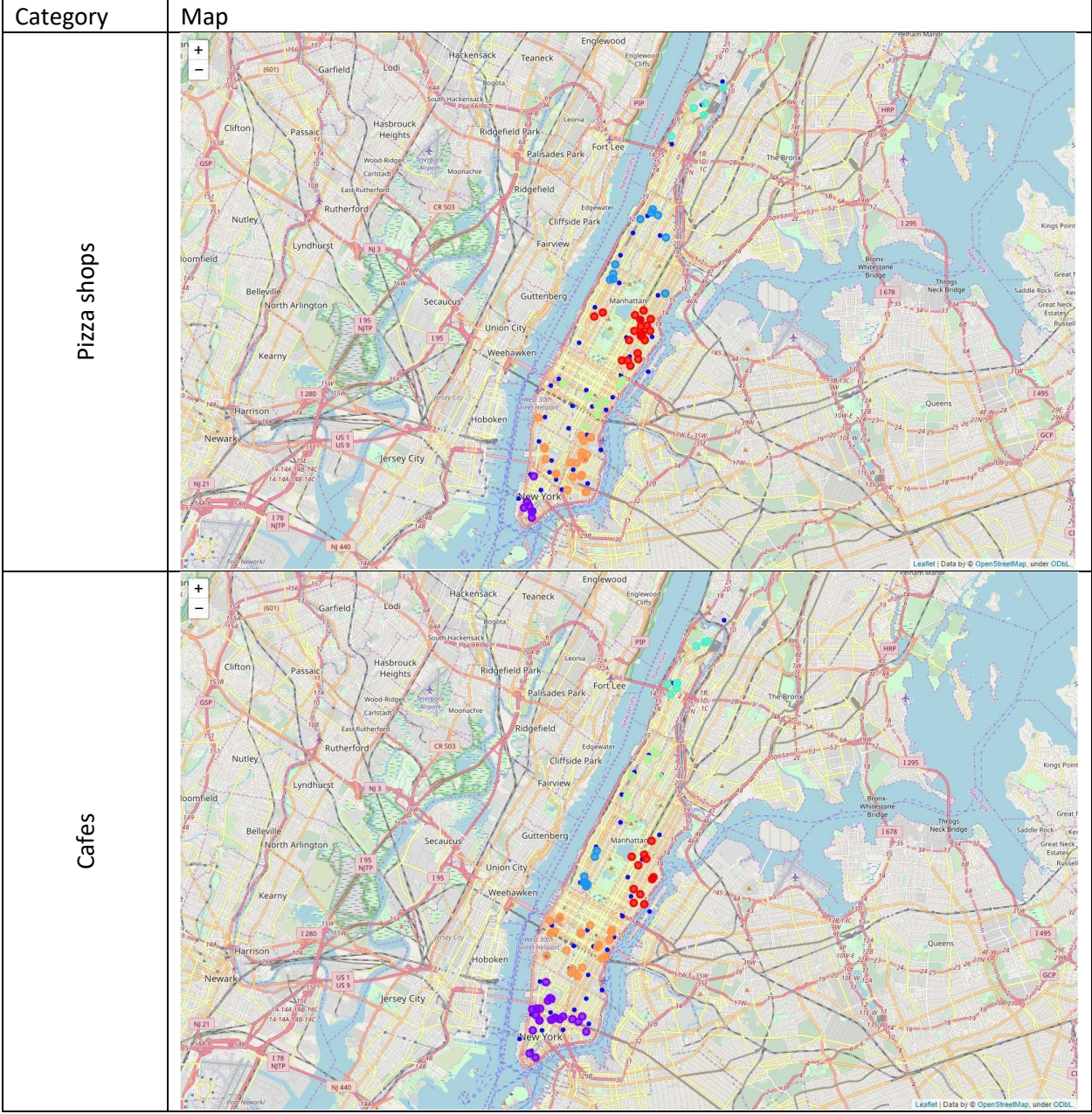
By using the foursquare API, a function – *getNearbyVenues*, was created in order to retrieve venues for each neighborhood location within a radius of 500 and a gathering limit of 200. *Manhattan\_venues*, was created after applying the function to the data frame *mandf*, and contains 338 unique categories, such as 128 Italian restaurants, 124 coffee shops, 78 American restaurants and 61 hotels.

### 3.3 Slicing and segmenting the foursquare data

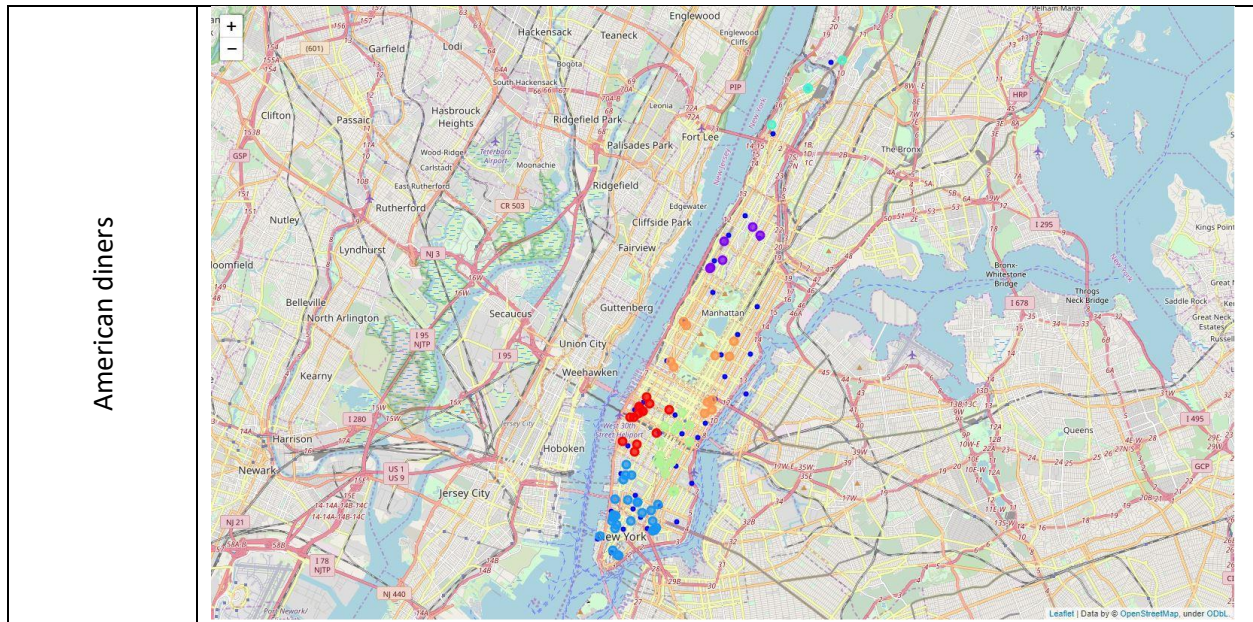
Since the intention of this study was to determine the best locations to stay in order to try pizza shops, cafés and American diners, the data frame was sliced, and three separate data frames were created: one for each category. They were named *manhattan\_pizza*, *manhattan\_coffee*, *manhattan\_american* respectively.

For each individual data frame, a machine learning tool called k-means clustering was used in order to partition the venues into 6 segments. A map was created using folium and the results was overlaid onto the map. This resulted in a cluster of venues within a few neighborhoods, which will allow tourist to decide between a few neighborhoods to stay in to try specific stores or to stay in a neighborhood with the largest volume of venues.

4. Results







## 5. Discussion

The original objective of this study is achieved by finally analyzing how the venues are segmented. In this section, certain suggestions are made for each category on where to stay to get the best “bang for your buck”, and fully utilizing the time an individual might have in New York City.

It was found that for people who intend to try pizza options, the best locations to stay are near the red and orange venues. Some of the neighborhoods in that location include Lower East Side and Carnegie Hill. Between these two clusters, they contain 55.4% of the sourced pizza locations.

It was found that for people who intend to try coffee options, the best locations to stay are near the purple and orange venues. Some of the neighborhoods in that location include Tudor City and Tribeca. Between these two clusters, they contain 57.3% of the sourced café locations.

It was found that for people who intend to try American diner options, the best locations to stay are near the blue venues. Some of the neighborhoods in that location include Chinatown and Tribeca. For this cluster, it contains 33.3% of the sourced diner locations.

## 6. Conclusion

The study concludes and shows that it is possible to provide relevant data to tourist for them to make an informed decision on where to stay. The intersection of neighborhood for all three food option seem to be in the south end of Manhattan. This could be a possible location to stay in order to be able to try as many different cuisines as possible. Doing so will reduce the headache of deciding on where to stay when on

vacation. A possible development of the project can incorporate other factors like Airbnb prices and crime frequency of each neighborhood to give a clearer picture to the end-user.

This list is by no means an exhaustive list. Further expansion of this project to include more food options, or a wider radius, or a larger segmentation number are all viable options. Due to the nature of the code, these variables are all easily interchangeable and suited for the intention of the end-user.

## 7. References

- Mike Maciag (October 2, 2013). "Mapping the Nation's Most Densely Populated Cities". *Governing – The States and Localities*. Archived from the original on August 27, 2016. Retrieved September 18, 2019.
- Zelinsky, W. (1985). "The roving palate: North America's ethnic restaurant cuisines". *Geoforum*. 16: 51–72. doi:10.1016/0016-7185(85)90006-5.