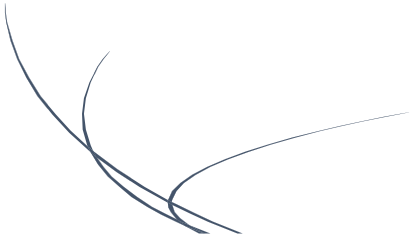


Exploring the Taste of New York

Utsav Sutariya

26 December, 2019



Introduction

Background

New York City is the most populous city in the United States, home to the headquarters of the United Nations and an important center for international diplomacy. It just might be the most diverse city on the planet, as it is home to over 8.6 million people and over 800 languages.

As quoted in an article - [What Food Tells Us About Culture](#)

“Traditional cuisine is passed down from one generation to the next. It also operates as an expression of cultural identity. Immigrants bring the food of their countries with them wherever they go and cooking traditional food is a way of preserving their culture when they move to new places.”

Problem

Undoubtedly, Food Diversity is an important part of an ethnically diverse metropolis. The idea of this project is to categorically segment the neighborhoods of New York City into major clusters and examine their cuisines. A desirable intention is to examine the neighborhood cluster's food habits and taste. Further examination might reveal if food has any relationship with the diversity of a neighborhood. This project will help to understand the diversity of a neighborhood by leveraging venue data from Foursquare's 'Places API' and 'k-means clustering' unsupervised machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighborhood.

Stakeholders

This quantifiable analysis can be used to understand the distribution of different cultures and cuisines over ‘the most diverse city on the planet – New York City’. Also, it can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.

Data

To examine the above said, following data sources will be used:

New York City Dataset

Link: https://geo.nyu.edu/catalog/nyu_2451_34572

This New York City Neighborhood Names point file was created as a guide to New York City’s neighborhoods that appear on the web resource, ‘New York: A City of Neighborhoods.’ Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale. This dataset will provide the addresses of neighborhood of NYC in json format. An extract of the json is as follows:

```
{'type': 'Feature',
'id': 'nyu_2451_34572.306',
'geometry': {'type': 'Point',
'coordinates': [-74.08173992211962, 40.61731079252983]}},
'geometry_name': 'geom',
'properties': {'name': 'Fox Hills',
'stacked': 2,
'annoline1': 'Fox',
'annoline2': 'Hills',
'annoline3': None,
'annoangle': 0.0,
'borough': 'Staten Island',
'bbox': [-74.08173992211962,
```

Foursquare API

Link: <https://developer.foursquare.com/docs>

Foursquare API, a location data provider, will be used to make RESTful API calls to retrieve data about venues in different neighborhoods. This is the link to [Foursquare Venue Category Hierarchy](#). Venues retrieved from all the neighborhoods are categorized broadly into ‘Arts & Entertainment’, ‘College & University’, ‘Event’, ‘Food’, ‘Nightlife Spot’, ‘Outdoors & Recreation’, etc. An extract of an API call is as follows:

```
'categories': [{ 'id': '4bf58dd8d48988d110941735',  
  'name': 'Italian Restaurant',  
  'pluralName': 'Italian Restaurants',  
  'shortName': 'Italian',  
  'icon': { 'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/italian_',  
    'suffix': '.png' },  
  'primary': True } ],  
'verified': False,  
'stats': { 'tipCount': 17 },  
'url': 'http://eccorestantny.com',  
'price': { 'tier': 4, 'message': 'Very Expensive', 'currency'
```