2. Data

2.1. Data sources

The data for this project was collected from 3 main sources: Wikipedia, geopy python package and Foursquare. The list of New York neighborhoods was first obtained from Wikipedia, geographic co-ordinates for the neighborhoods were obtained using geopy and the details of restaurants located in each neighborhood were fetched using the Foursquare API.

2.2. Data Collection

The names of the neighborhoods were obtained from the wikipedia page, Neighborhoods in New York City [https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City]. The data was available in an HTML table with the New York community board names, area, population and neighborhoods in each board.

Community Board (CB)	Area km^2	Pop. Census 2010	Pop./ km^2	Neighborhoods
Bronx CB 1	7.17	91,497	12,761	Melrose, Mott Haven, Port Morris
Bronx CB 2	5.54	52,246	9,792	Hunts Point, Longwood
Bronx CB 3	4.07	79,762	19,598	Claremont, Concourse Village, Crotona Park, Morrisania
Bronx CB 4	5.28	146,441	27,735	Concourse, High Bridge
Bronx CB 5	3.55	128,200	36,145	Fordham, Morris Heights, Mount Hope, University Heights
Bronx CB 6	4.01	83,268	20,765	Bathgate, Belmont, East Tremont, West Farms
Bronx CB 7	4.84	139,286	28,778	Bedford Park, Norwood, University Heights
Bronx CB 8	8.83	101,731	11,521	Fieldston, Kingsbridge, Kingsbridge Heights, Marble Hill, Riverdale, Spuyten Duyvil, Van Cortlandt Village
Bronx CB 9	12.41	172,298	13,884	Bronx River, Bruckner, Castle Hill, Clason Point, Harding Park, Parkchester, Soundview, Unionport
Bronx CB 10	16.76	120,392	7,183	City Island, Co-op City, Locust Point, Pelham Bay, Silver Beach, Throgs Neck, Westchester Square
Bronx CB 11	9.32	113,232[2]	12,149	Allerton, Bronxdale, Indian Village, Laconia, Morris Park, Pelham Gardens, Pelham Parkway, Van Nest[3][4]

Figure 1: List of neighborhoods in New York (Wikipedia)

The table data from Figure 1 was scraped using the BeautifulSoup python library, cleaned and stored in a pandas dataframe.

Only Manhattan neighborhoods were chosen. The geographical coordinates for each of them were fetched using the geopy python package. The address input required by geopy was constructed using the neighborhood and borough names.

```
Battery Park City, Manhattan, New York 40.7110166 -74.0169369
Financial District, Manhattan, New York 40.7076124 -74.009378
TriBeCa, Manhattan, New York 40.7153802 -74.0093063
Chinatown, Manhattan, New York 40.7164913 -73.9962504
Greenwich Village, Manhattan, New York 40.7319802 -73.9965658
Little Italy, Manhattan, New York 40.7192728 -73.9982152
Lower East Side, Manhattan, New York 40.7159357 -73.9868057
NoHo, Manhattan, New York 40.7258746 -73.9939566
```

Figure 2: Neighborhood names, latitude and longitude

To get the list of restaurants and their details, the Foursquare API was used. The explore, venues and search API endpoints were used. The following figure shows the JSON response of the explore endpoint for a single neighborhood.

```
{'meta': {'code': 200, 'requestId': '5c6587e2351e3d4e704c2e54'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'},
    {'name': '$-$$$$', 'key': 'price'}]},
  'headerLocation': 'Battery Park City',
  'headerFullLocation': 'Battery Park City, New York',
  'headerLocationGranularity': 'neighborhood',
  'query': 'restaurants',
  'totalResults': 71,
  'suggestedBounds': {'ne': {'lat': 40.7155166045, 'lng': -74.0110113729468},
   'sw': {'lat': 40.7065165955, 'lng': -74.02286242705321}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '5362a2ae498e3b18c22334be',
       'name': 'Hudson Eats',
       'location': {'address': '225 Liberty St',
        'crossStreet': 'South End Avenue',
        'lat': 40.712802562960505,
        'lng': -74.01610221841605,
        'labeledLatLngs': [{'label': 'display',
          'lat': 40.712802562960505,
          'lng': -74.01610221841605}],
        'distance': 210,
        'postalCode': '10281',
        'cc': 'US',
        'city': 'New York',
        'state': 'NY',
        'country': 'United States',
        'formattedAddress': ['225 Liberty St (South End Avenue)',
         'New York, NY 10281',
         'United States']},
       'categories': [{'id': '4bf58dd8d48988d120951735',
         'name': 'Food Court',
         'pluralName': 'Food Courts',
         'shortName': 'Food Court',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/food_foodcourt_',
          'suffix': '.png'},
         'primary': True}],
```

Figure 3: API response for restaurants within a 500 metre radius of Battery Park City, Manhattan

The explore endpoint was used to fetch restaurants around a 500 metre radius of each Manhattan neighborhood. Only the Italian restaurants were chosen.

The details for each Italian restaurant were fetched using the venues endpoint of the API. The venue details response can be seen in Figure 4.

```
{'meta': {'code': 200, 'requestId': '5c6828039fb6b740fecfc76e'},
 'response': {'venue': {'id': '556cb8c6498e751e51fbcf41',
   'name': 'Parm',
   'contact': {'phone': '2127764927', 'formattedPhone': '(212) 776-4927'},
   'location': {'address': '250 Vesey St',
    'crossStreet': 'West St',
    'lat': 40.71451391801664,
    'lng': -74.01626377138824,
    'labeledLatLngs': [{'label': 'display',
      'lat': 40.71451391801664,
      'lng': -74.01626377138824}],
    'postalCode': '10080',
    'cc': 'US',
    'city': 'New York',
    'state': 'NY',
    'country': 'United States',
    'formattedAddress': ['250 Vesey St (West St)',
     'New York, NY 10080',
     'United States']},
   'canonicalUrl': 'https://foursquare.com/v/parm/556cb8c6498e751e51fbcf41',
   '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},
    {'id': '4bf58dd8d48988d1c5941735',
     'name': 'Sandwich Place',
     'pluralName': 'Sandwich Places',
     'shortName': 'Sandwiches',
     'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/deli_',
      'suffix': '.png'}}],
   'verified': True,
   'stats': {'tipCount': 67},
   'url': 'http://www.parmnyc.com',
   'price': {'tier': 2, 'message': 'Moderate', 'currency': '$'},
   'likes': {'count': 239,
    'groups': [{'type': 'others', 'count': 239, 'items': []}],
    'summary': '239 Likes'},
   'dislike': False,
   'ok': False,
   'rating': 8.5,
```

Figure 4: API response for venue details

The closest venues around each restaurant were also fetched to understand how they affect restaurant ratings and reviews. The search endpoint of the API was used to find the venues around a 250 metre radius of each restaurant. The raw response can be seen in Figure 5.

```
{'meta': {'code': 200, 'requestId': '5c63d3d01ed2192b5a2c1d53'},
 'response': {'venues': [{'name': 'Gild Hall',
    'hasPerk': False,
    'location': {'city': 'New York',
     'formattedAddress': ['15 Gold St (at Platt St)',
      'New York, NY 10038',
      'United States'],
     'lat': 40.7078871,
     'labeledLatLngs': [{'lat': 40.7078871,
       'label': 'display',
       'lng': -74.0071031}],
     'cc': 'US',
     'lng': -74.0071031,
     'address': '15 Gold St',
     'postalCode': '10038',
     'state': 'NY',
     'distance': 8,
     'country': 'United States',
     'crossStreet': 'at Platt St'},
    'id': '4a6e3e75f964a5204ed41fe3',
    'categories': [{'name': 'Hotel',
      'pluralName': 'Hotels',
      'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_',
       'suffix': '.png'},
      'primary': True,
      'shortName': 'Hotel',
      'id': '4bf58dd8d48988d1fa931735'}],
    'referralId': 'v-1550046160'},
   {'name': 'Felice 15 Gold Street',
    'delivery': {'provider': {'name': 'seamless',
      'icon': {'name': '/delivery_provider_seamless_20180129.png',
       'prefix': 'https://fastly.4sqi.net/img/general/cap/',
       'sizes': [40, 50]}},
     'id': '299433',
     'url': 'https://www.seamless.com/menu/felice-15-gold-st-new-york/299433?affiliate=1131&utm_source=foursquare-affili
ate-network&utm_medium=affiliate&utm_campaign=1131&utm_content=299433'},
    'hasPerk': False,
    'location': {'postalCode': '10038',
     'city': 'New York',
     'formattedAddress': ['15 Gold St', 'New York, NY 10038', 'United States'],
     'lat': 40.70783228127862,
     'labeledLatLngs': [{'lat': 40.70783228127862,
       'label': 'display',
       'lng': -74.00704045580463}],
     'cc': 'US',
```

Figure 5: API response for venues within a 250 metre radius of a restaurant