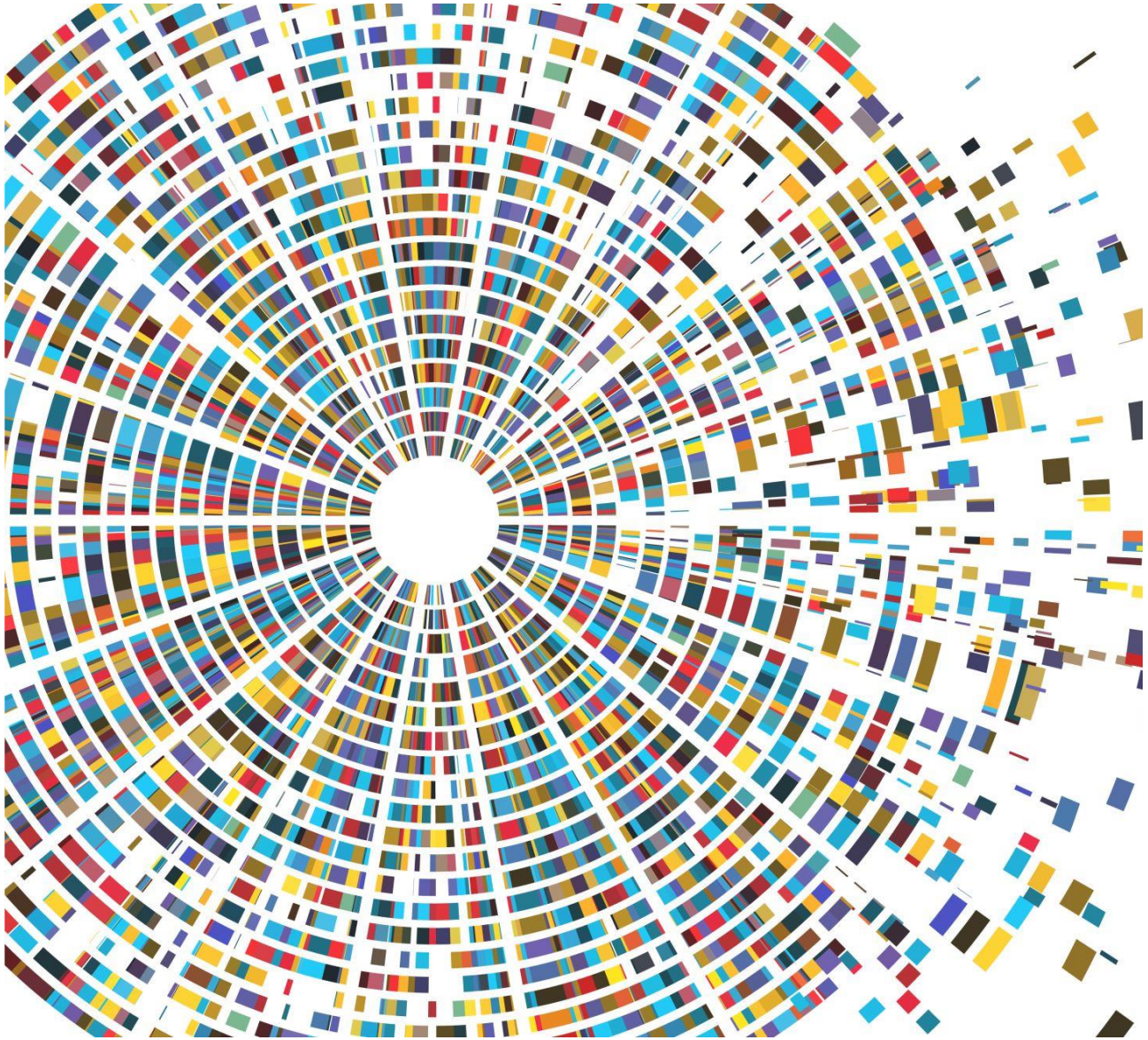


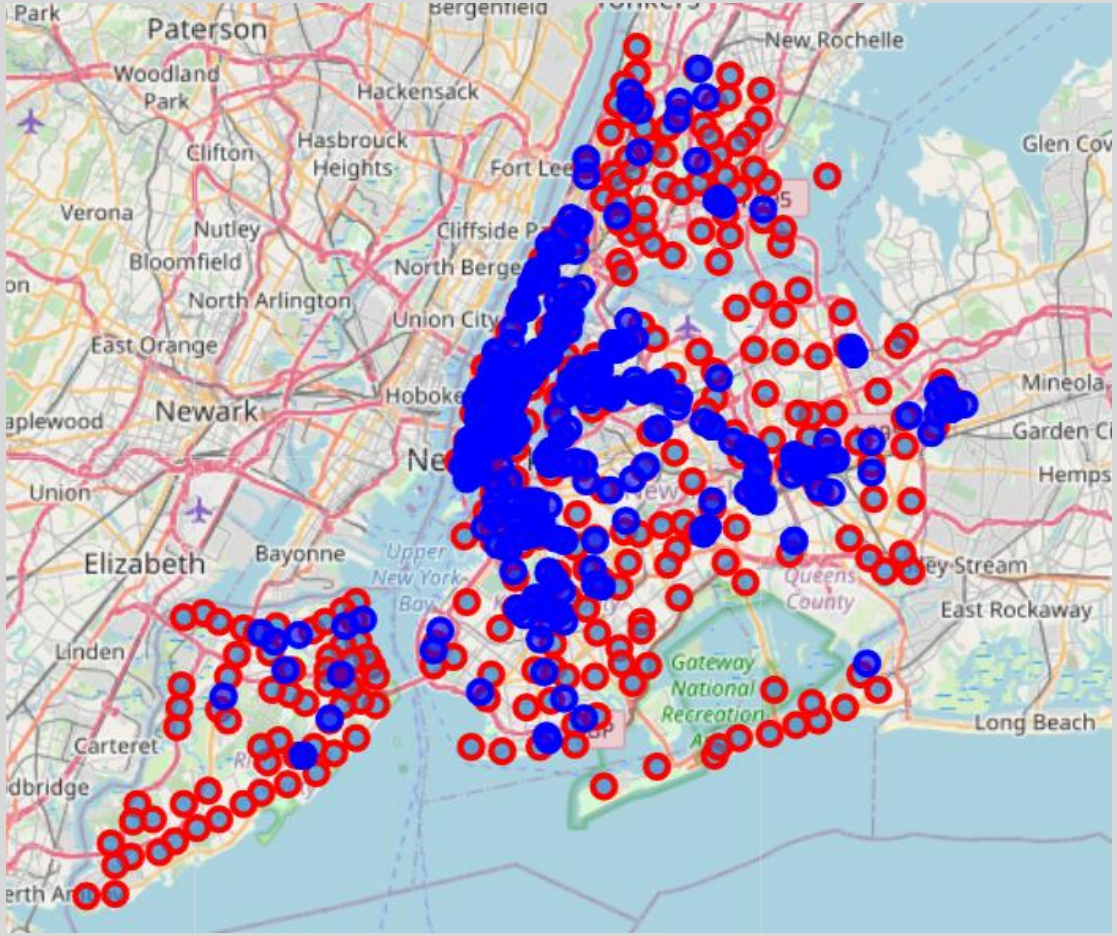
DISTRIBUTION OF INDIAN RESTAURANTS IN NYC

Business Understanding and Data

- Business owner looking to open new restaurant want to establish the density of Indian restaurants in NYC
- Frequency will be looked at based on different neighborhood regions
- Data obtained from foursquare api and maps from https://cocl.us/new_york_dataset and https://maps.princeton.edu/catalog/nyu-2451-34561
- K-means algorithms used to analyse the data

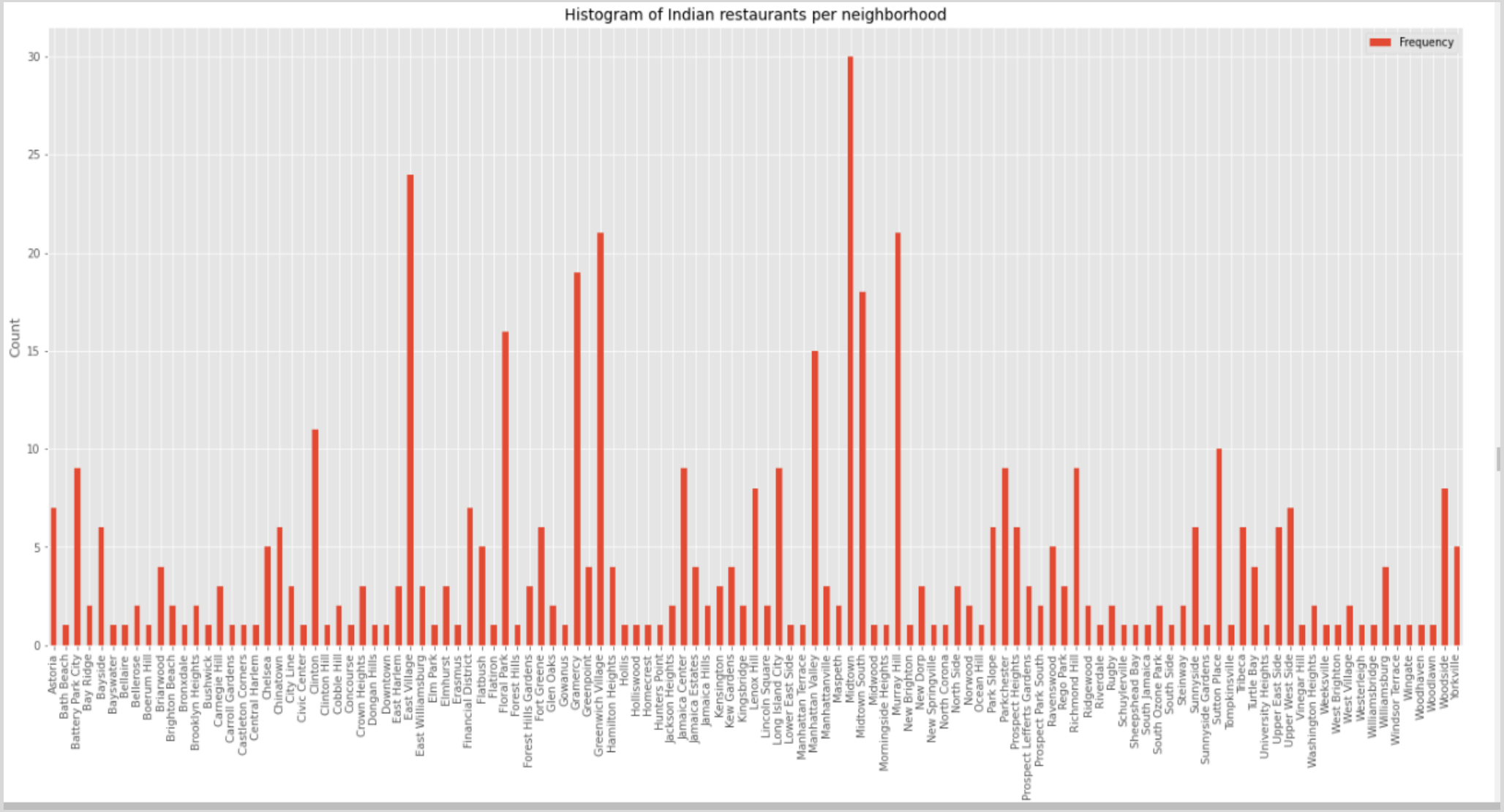
Restaurants by location across the city

- Restaurant data was obtained from foursquare api with the condition that the categoryId was for Indian restaurants within a search radius of 750m

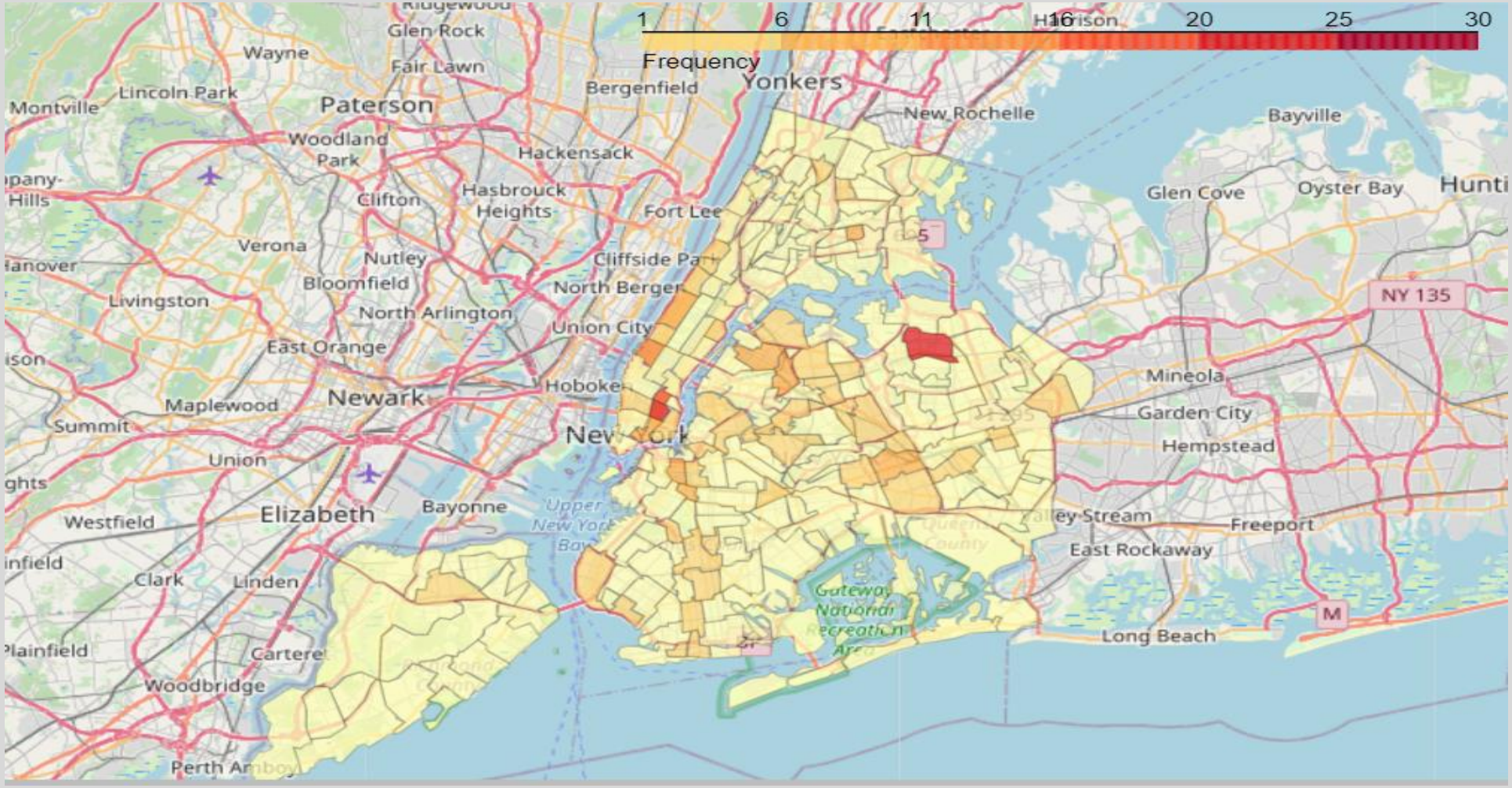


Red indicates neighborhood locations and the blue dots the restaurants

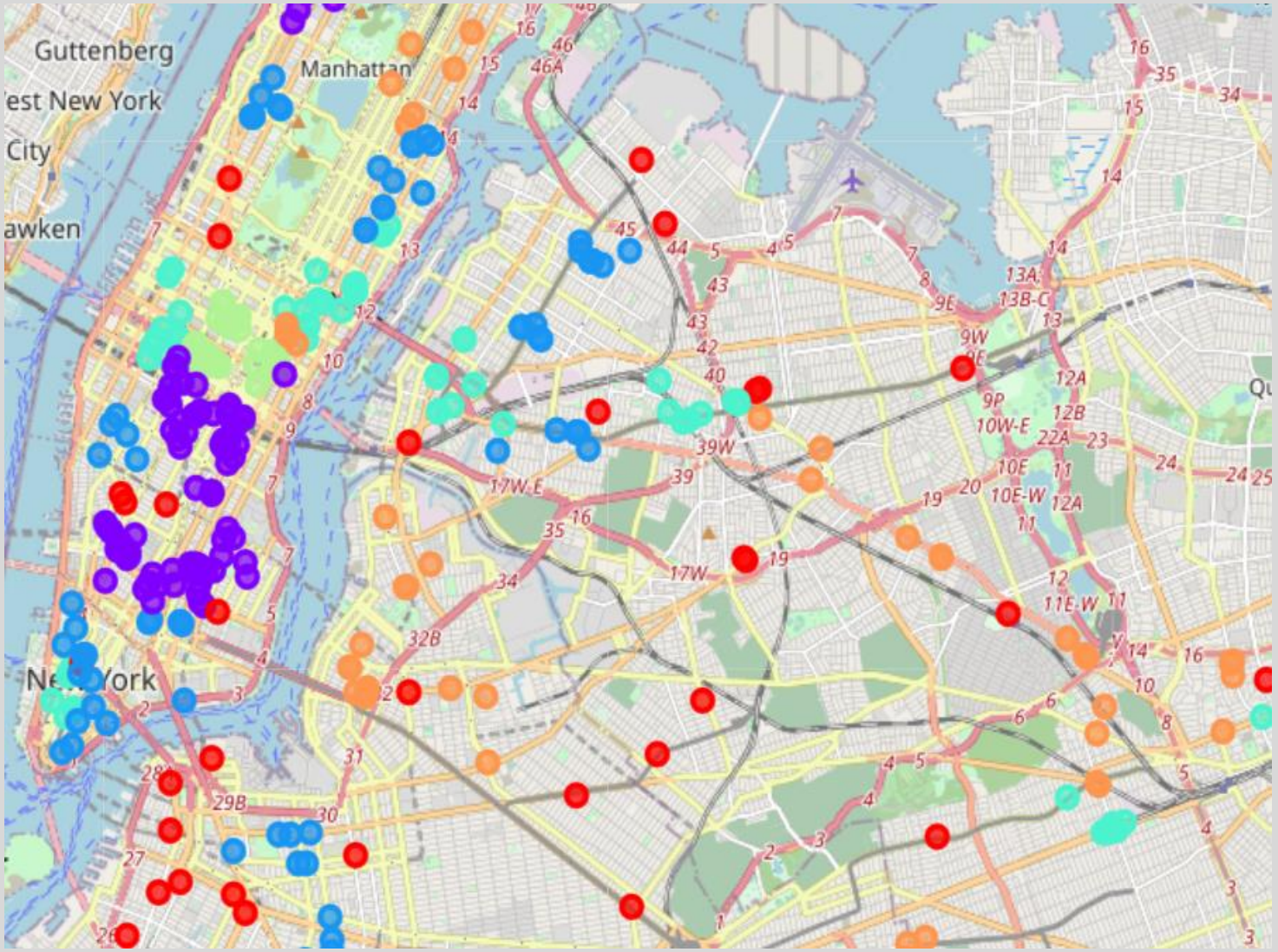
Frequency distributions: Bar chart



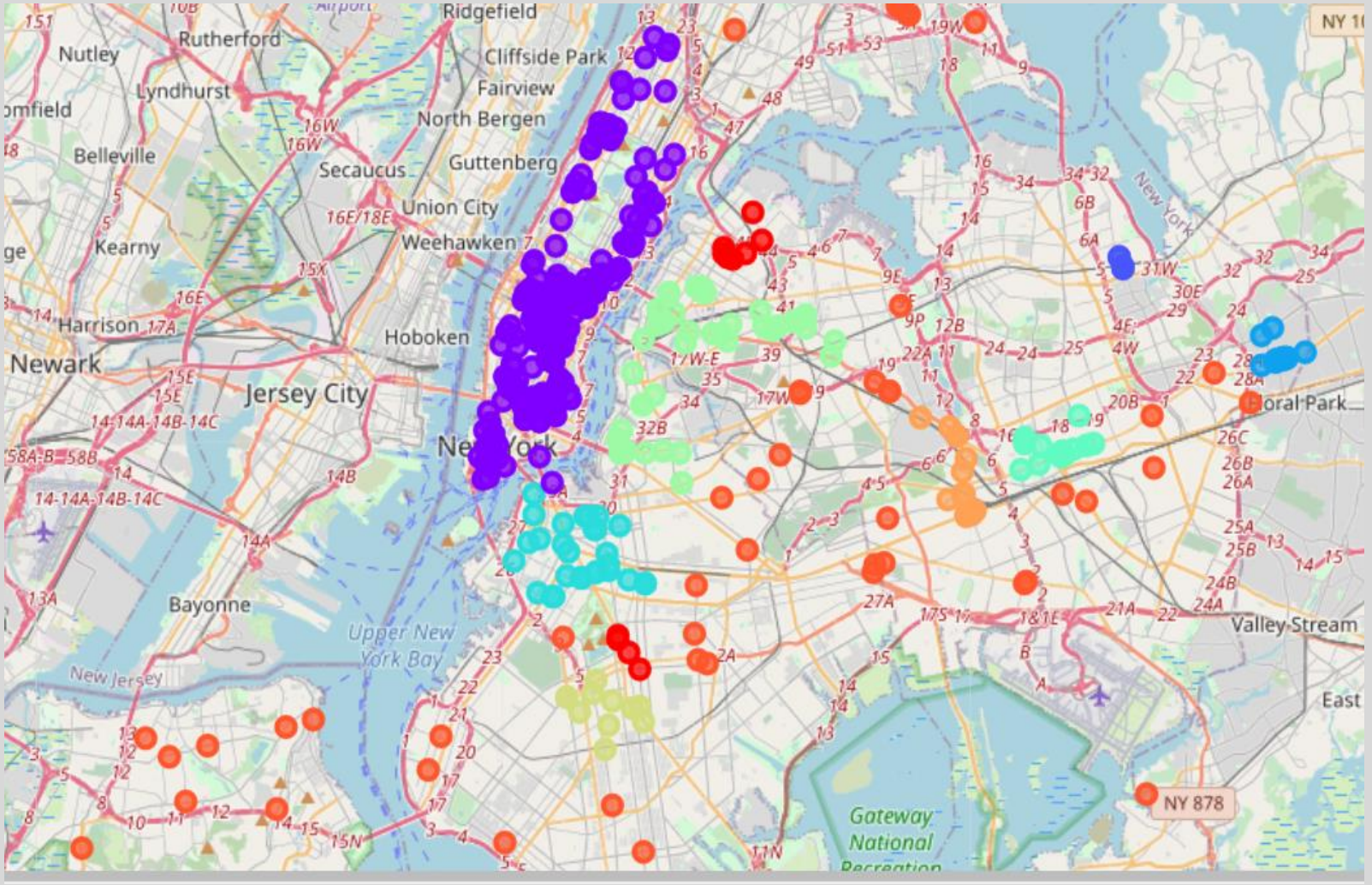
Frequency distributions: Heat map



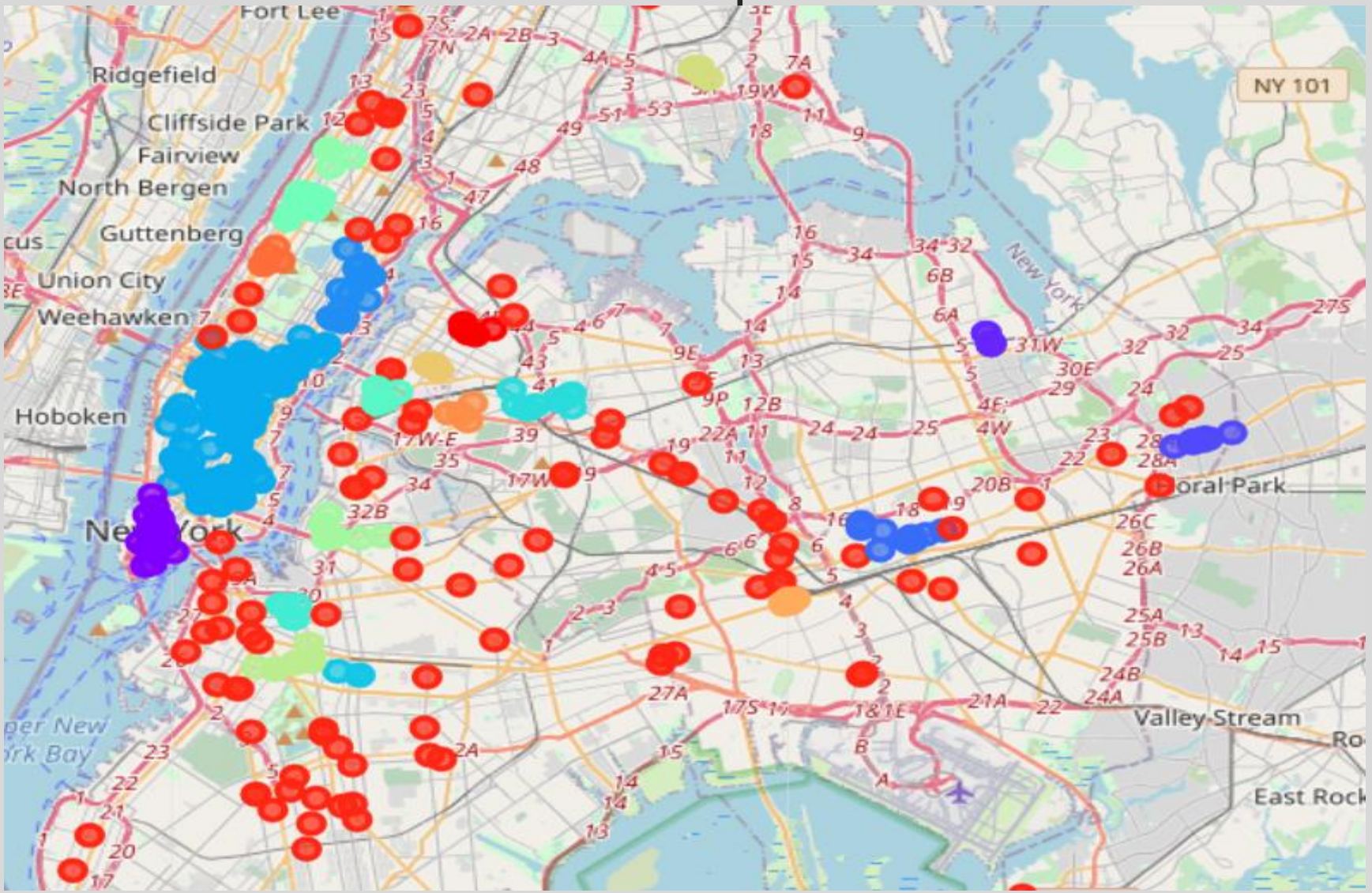
K-means clustering



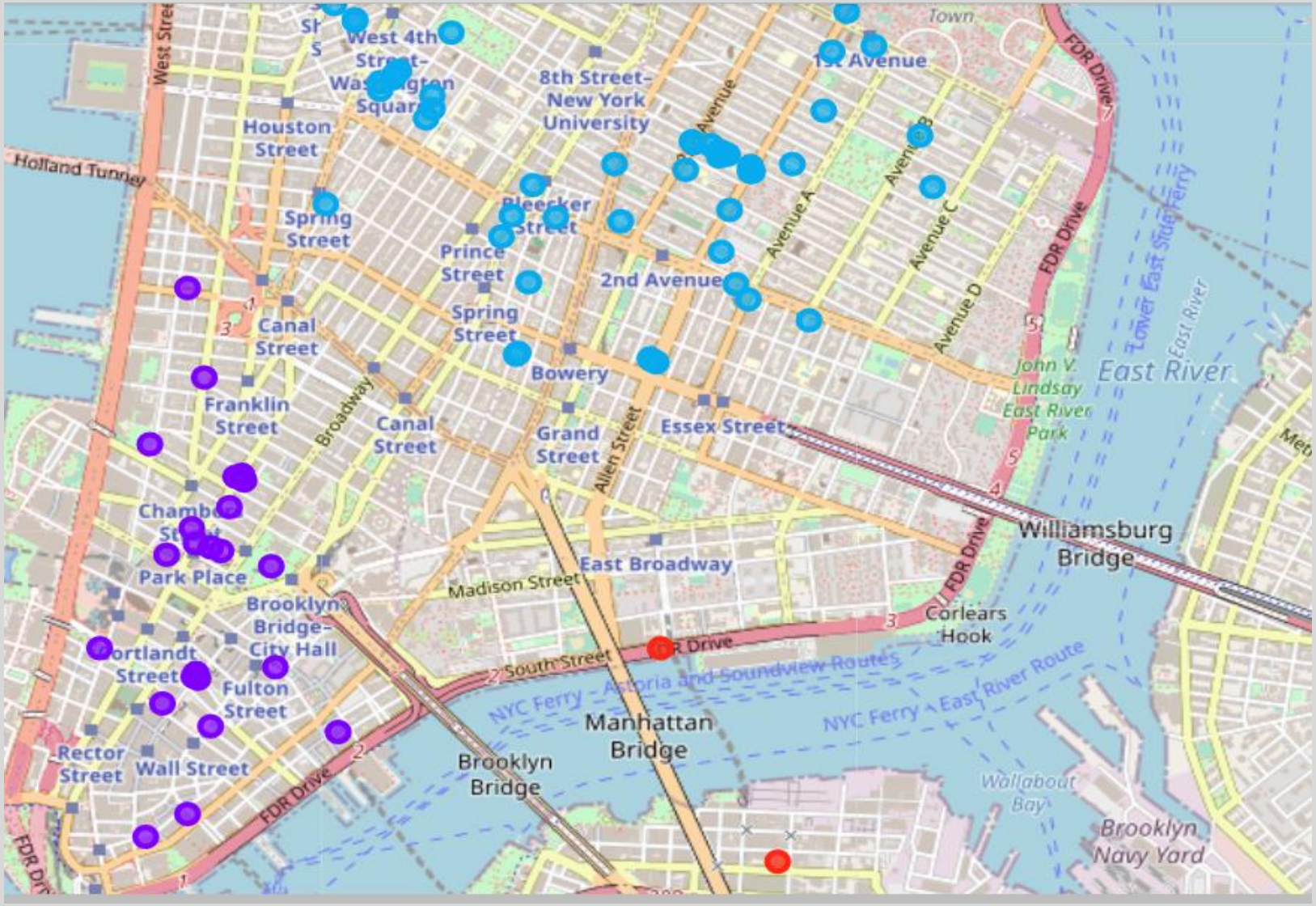
K-means with dBScan: ellipse =0.2



K-means with dBScan: ellipse =0.1



K-means with dBScan: ellipse =0.1 expanded lower manhatten



Discussion

- Manhattan has the highest density of restaurants: Midtown, East village, Greenwich Village and Murray Hill >20 Indian restaurants
- The choropleth map gives an indication on a broader scale on the distribution of restaurants on a visual scale.
- k-means clustering shows patterns on a local scale, and larger scale.
- While some regions do have a high density of Indian restaurants in the vicinity, there are also sparse intercluster areas where perhaps it is not favourable to open a business.
- However on an intracluster bases games can be identified

Conclusions

- Exercise is useful in determining the distribution of restaurants and identifying the best and worst places to set up based on nearest neighbours
- However, it is clearly just an initial step and more rigorous analysis need on points:
    - Restaurant quality
    - Population density in a given area
    - Prices