



IBM Applied Data Science Capstone

Opening a pizza place in New York City, USA



Pizza: Popular all over the US

- Pizza one of the most favourite dishes in the US
- NYC already offers many places to buy pizza, that does not apply for the whole city
- This project aims to find the best neighborhood to open a pizza place in NYC

Business problem and target audience

- Objective of this project: Analyze and select the best location in NYC to open a new place to sell pizza using data science methodology
- Find most homogeneous and dense cluster of neighborhoods in NYC
- Target audience: Property developers and investors looking to invest new pizza places or Italian restaurants

Data acquisition and cleaning

- List of neighborhoods in New York City
- Latitude and longitude coordinates of those neighborhoods for plotting
- Venue data, particularly data related to pizza places and Italian restaurants from [foursquare.com](https://www.foursquare.com)

Methodology

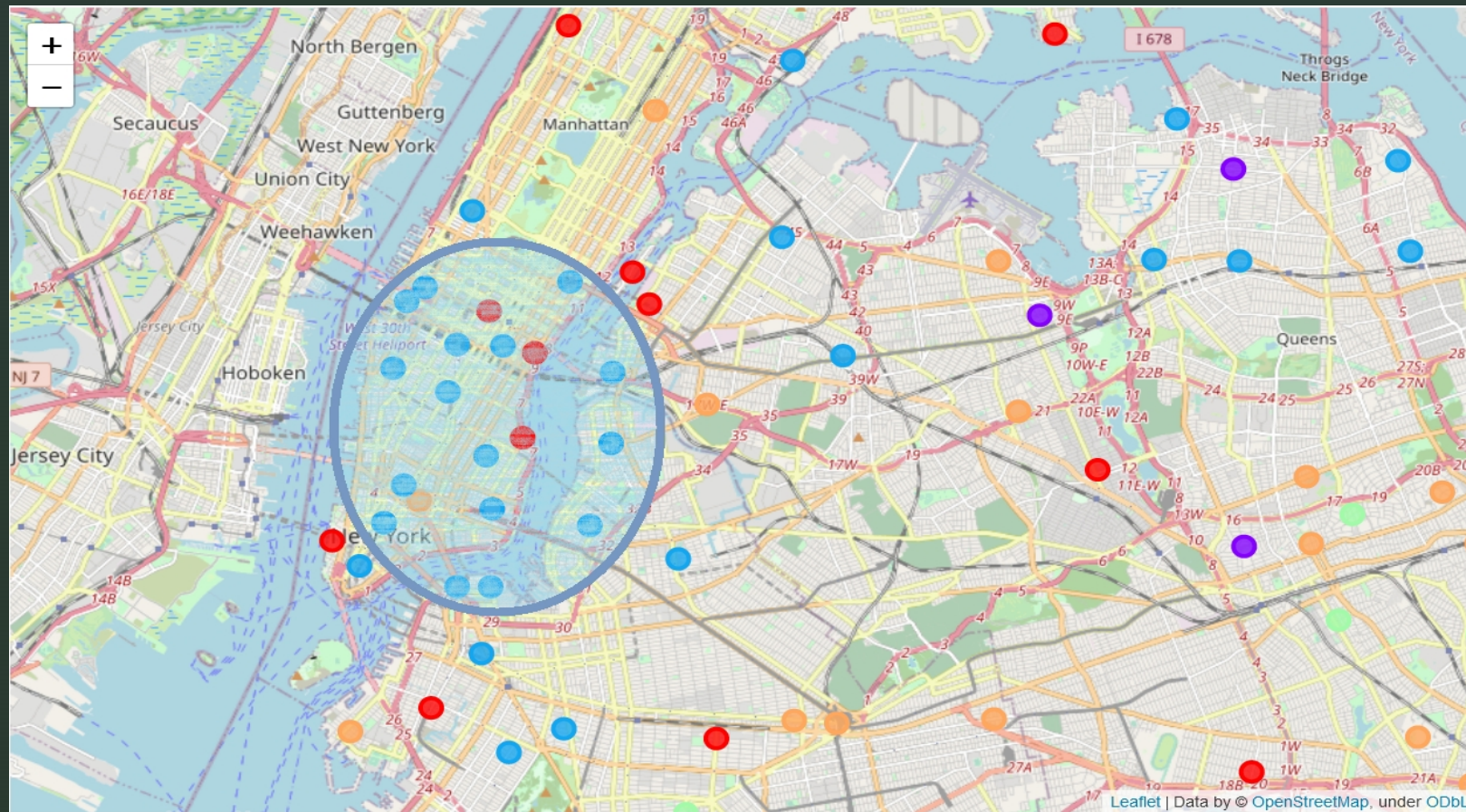
- Use Foursquare API to get the top 100 venues within a radius of 500 meters for each neighborhood in NYC
- Analyze each neighborhood, examine how many venues were returned and how many unique categories
- Take the mean of the frequency of occurrence of each venue category
- Use k-means clustering to find similar neighborhoods, of those with no pizza place in the top ten venues
- Finally, find most dense cluster of similar neighborhoods

Discussion and final recommendation

- The spot I will choose should be in a neighborhood where
 - the surrounding neighborhoods are as homogeneous as possible
 - the surrounding neighborhoods have as little pizza places as possible
 - the neighborhood itself and the surrounding neighborhoods are as densely populated as possible
 - None of the top ten revenues is a pizza place or Italian restaurant

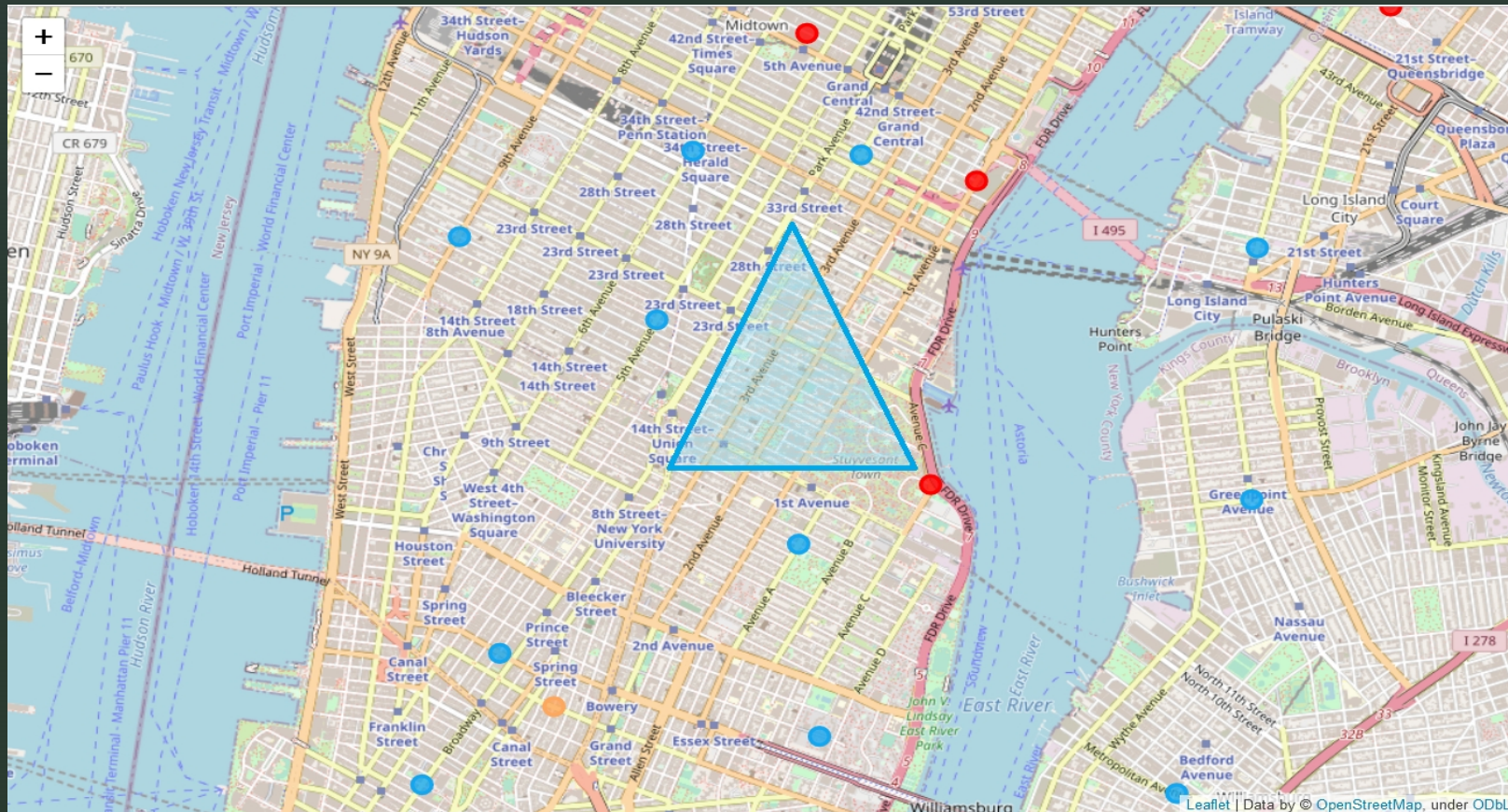
Results: Most homogeneous cluster

- Most homogeneous and dense cluster situated in the South of Manhattan:



Closer look: Most homogeneous cluster

- Nearby 23rd Street and 2nd Avenue in the south of Manhattan:



Limitations and future directions

- Just considered one major factor: Nearby a given neighborhood is no pizza place or Italian restaurant in the top ten venues
- May consider more factors: population density, income of inhabitants, housing prices