

Recommending a neighbourhood to open Italian Restaurant in New York City

IBM Data Science - Capstone

CHRISTODOULOS CHRISTODOULOU

Introduction

- Italian cuisine is very popular
- Selecting a location for the opening of a new restaurant is very important
- Find a neighbourhood with not many similar options can increase the customers and the income
- Help aspiring restaurant owners to select a place for their new investment

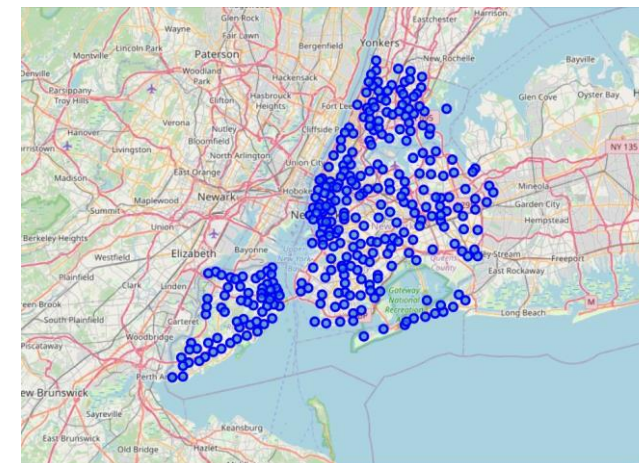


Data

- New York City neighbourhoods from NYU Spatial Data Repository
- Coordinates from Geocoder package
- Venues for each neighbourhood from Foursquare API

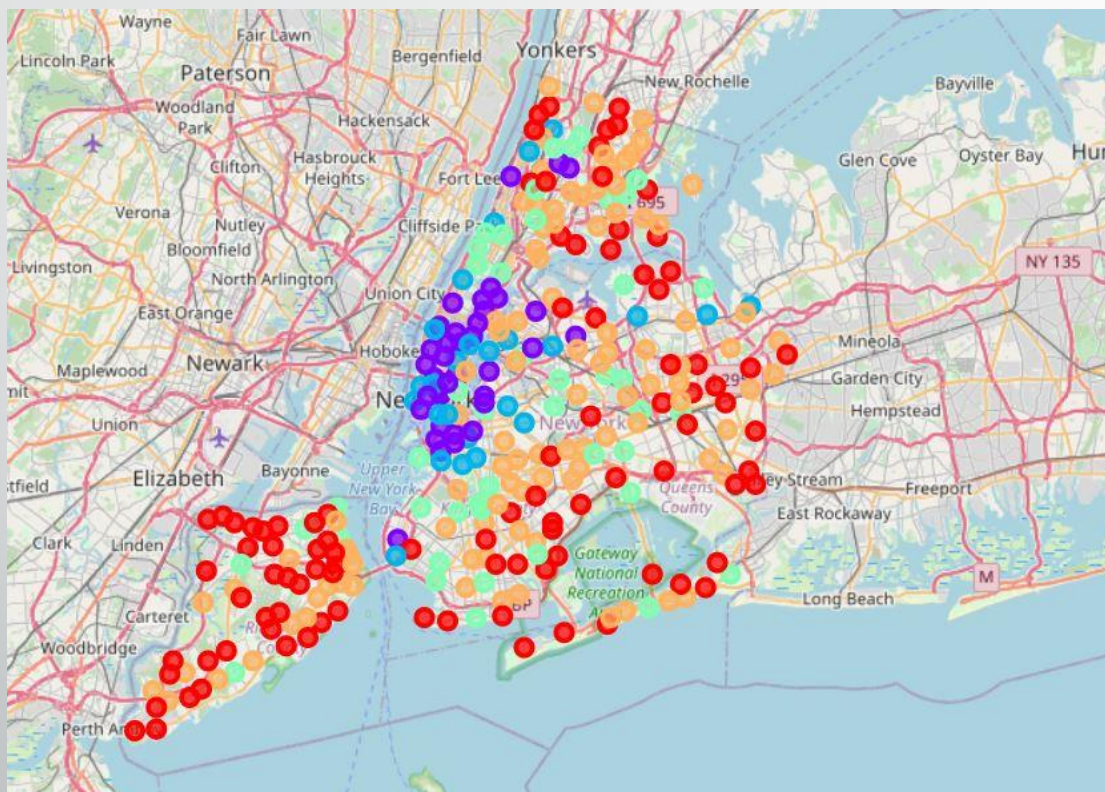
Methodology

- Collect names and locations of the neighborhoods
- Explore venues for each neighbourhood
- Perform data cleaning to create a data frame including the neighbourhoods and the number of existed Italian restaurnts
- Apply K-Means algorithm to cluster neighborhoods with respect to the presence of Italian restaurants



	Neighborhood	Italian Restaurant
0	Allerton	30
1	Annadale	15
2	Arden Heights	5
3	Arlington	6
4	Arrochar	20

Results



	Cluster	Italian Restaurants
0	Red	0-14
1	Orange	15-30
2	Light Blue	31-52
3	Light Green	53-80
4	Purple	80-130

Discussion

- Neighbourhoods away from centre (Manhattan) have less existed restaurants
- Choice should be done in red cluster but to be close enough to the centre where other venues are present too
- Astoria Height neighbourhood is a good recommendation based the above criteria

Conclusion

- A methodology to select location for opening a new Italian restaurant was presented using machine learning
- Recommendation based on number of other similar options to the neighbourhood and distance from the centre
- Further improvement is to make the clustering of neighbourhoods by taking into account the average rating of Italian restaurants on each neighbourhood too