


An aerial photograph of the New York City skyline, featuring a dense cluster of skyscrapers. The Empire State Building is the most prominent structure in the center, with its iconic Art Deco spire reaching towards the sky. Other notable buildings include the United Nations Secretariat Building to the left and the 111 West 57th Street skyscraper to the right. The city extends to the horizon under a clear blue sky with light clouds.

COURSERA CAPSTONE PROJECT

APPLIED DATA SCIENCE

INTRODUCTION

- NYC has an enormous population density
- Restaurant market is highly competitive.
- To be successful a business needs to invest in its strategy, pricing aswell as **location**. Analyzing the market is one of the key methods to set up a business for long-term success.



- Question: where to locate an upper class pizza restaurant in NYC?

- Business case: Ferrari Family wants to open up a Pizza Place in NYC but is unsure where to locate it.

DATA

To solve the problem the following data sources will be used

1. Data about Neighbourhoods of New York provided in .json format
2. Data about venues retrieved from Foursquares API

DATA

To solve the problem the following data sources will be used

1. Data about Neighbourhoods of New York provided in .json format
2. Data about venues retrieved from Foursquares API

DATA

306 Neighbourhoods of New York from this link: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Essentiell step:

Data population from JSON format into Pandas DataFrame

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643

DATA

Foursquares API: <https://www.foursquares.com>

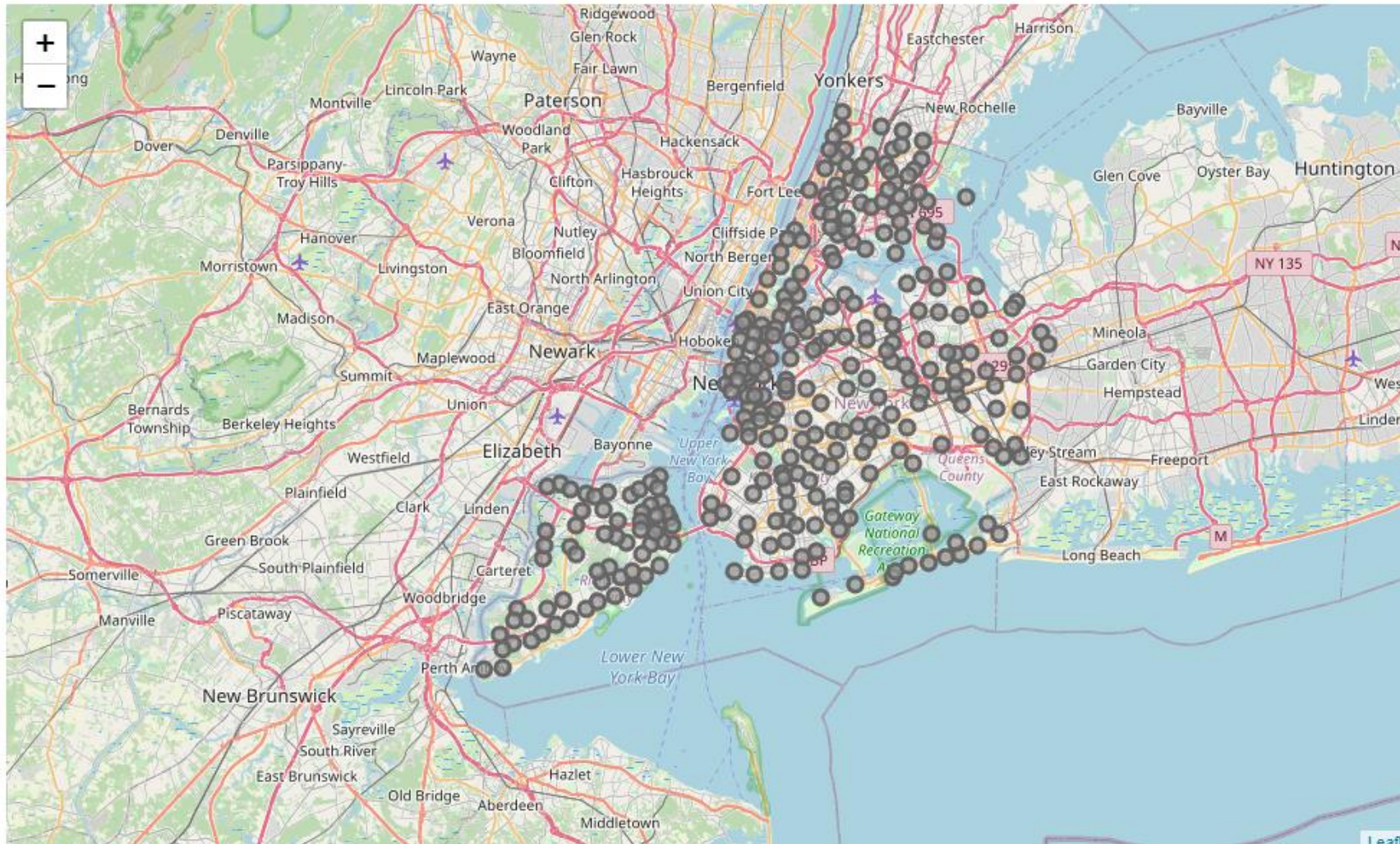
Essentiell step:

API calls from Foursquares API. Data population into pandas DataFrame

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643

DATA

Visualization as sanity check...



METHODOLOGY

1. Data crawling
2. Data preperation
3. k-means clustering
4. Result visualization

METHODOLOGY k-means clustering

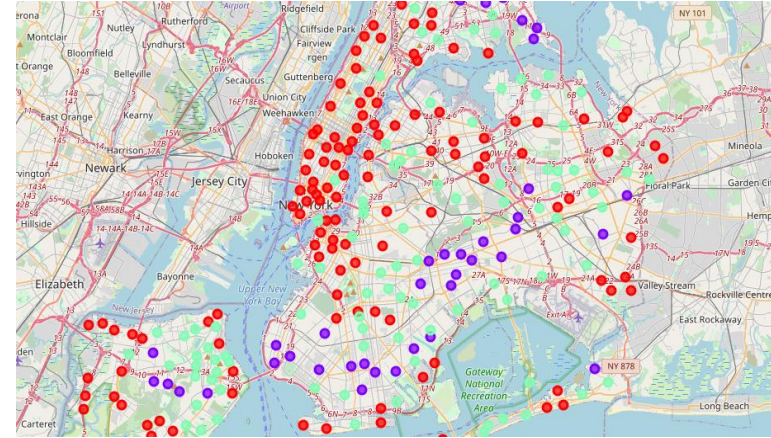
1. k-means clustering identifies k number of centroids and allocates the specific Neighbourhoods to one of the clusters
2. Due to its simplicity it is one of the most popular unsupervised machine learning algorithms.
3. The frequency of occurrence of Pizza Places within the Neighbourhood will be the main factor for the k-means clustering.

RESULT of k-means clustering



RESULT DISCUSSION

1. Lot of Neighbourhoods with no big competition in NYC
Manhattan
2. Many Pizza places in suburban areas where most of population lives -> not good choices
3. This analysis is only the first filter. Another criteries such as demand within the Neighbourhoods or other information should be obtained. But red Neighbourhoods seem suitable!



CONCLUSION

- Various neighbourhoods with competitor-less markers
- Interestingly high amount of Pizza places in suburban areas, not in city center
- More factors to be taken into consideration before giving a bulletproof recommendation for a location but first filter applicable with this report

