

Coursera Capstone Project

IBM Applied Data Science

Opening a new Pizza Place in NYC

By: Julian Rieger

Januar 2021



Introduction

New York is one of the most diverse mega-cities of the world. People from different ethnics have settled down in New York since it founding in 1626. Since then it attracted people with any kind of backgrounds. This diversity is one of the main factors for the diversity of cultural offers within this city. Besides the diverse culture of the City New York has attracted major companies from all over the world. It is one of the centers for finance and banking, trade, real estate and many more.

In a city with an enourmous population density the market for all kinds of business such as resteraunts, bars, shops and many more is highly competitive. To be successful a business needs to invest lots of thoughts 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.

Problem

This report will aim to solve the question of where to locate an upper class pizza resteraunt. The Ferrari family is in the resteraunt business for over two decades in chicago but strives to relocate to New York. They are looking for the best three possibilities to open up a resteraunt. Thus, the target audience of this report is the Ferrari family who trusted the ABC Data science team with this success-critical task. The scope is to find an area, not an exact location or even venue. Due to its huge savings the Luigi family has no restrictions due to high rents in certain areas of NYC.

Measure for Success

The Luigi family is looking for a neighbourhood which is located in NYC. The area location should fulfill the requirements of as less as possible Pizza Places nearby. As a measure a k-means classification will be used. The Neighbourhoods will be split into category 0 (very few Pizza Places), category 1 (medium amount of Pizza Places) and category 2 (many Pizza Places). This data will be visualised to give a recommendation of where the Luigis could locate their resteraunt.

Data

To solve the problem the following data sources will be used. It will also briefly be explained how the data will be received and converted into a useable format.

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

1. Neighbourhoods of New York

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 This data is stored in a JSON format. It consists a list of features. Each feature equals one neighbourhood and contains the data in a dictionary format. We can loop through all of the items of the lists (all neighbourhoods) and can reframe our data into one pandas dataframe. The result can be seen below:

2. Foursquares API

Foursquare operates one of the richest and most used location database. It is used by over 125.000 developers worldwide. This data is accessible through an API. The foursquares API will deliver us the data of existing venues within the in step 1 retrieved Neighbourhoods of New York. This foursquares data will be used to find out how the competitor situation within each Neighbourhoods. Afterwards, we can use a k-means classification to give each Neighbourhood an indicator which resembles the

Methodology

Firstly, we will receive the data about the 306 Neighbourhoods of New York. This is done retrieving the data from the following link: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json .

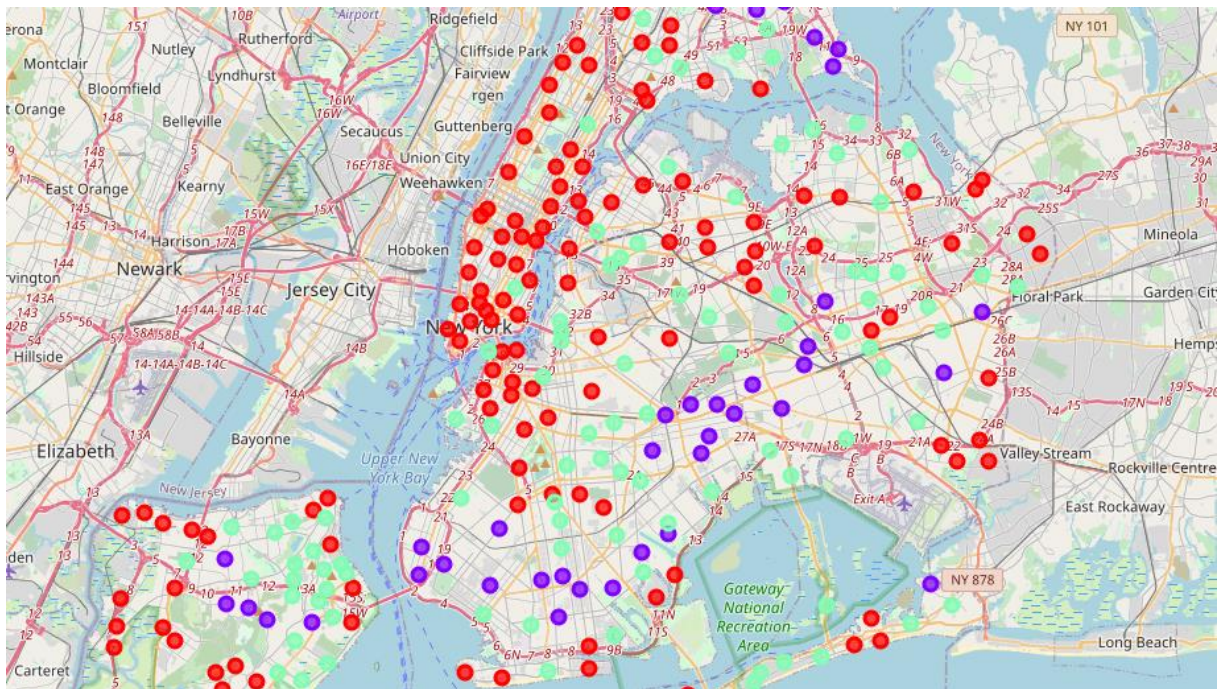
After creating a dataframe we can loop through the json file. The json file contains a list of features. For each feature the attributes will be stored within a new line of the dataframe. This allows us to work with the data within the pandas library. This data already contains latitude aswell as longitude. This is needed to use the Foursquares API to find competitor locations in the neighbourhood. After populating the Neighbourhood data into a Pandas DataFrame we will visualize the findings with a Folium map. This helps us to make a quick sensecheck of the downloaded data.

Next we will use the Foursquare API. This will allow us to get the top 100 venues within a 2000m radius of the neighbourhood. To run through the data download we will need a Foursquare Developer Account. After creating the account and an app on Foursquare we receive an ID key aswell as a secret key. Those are essential for the API calls we need to obtain the data from the Foursquares API. Afterwards, we loop through all the Neighbourhoods and make on API call for each Neighbourhood adding the longitude and latitude parameters aswell as the radius of 2000m and maximum amount of venues (100). All of those venues will be added to a list. Then we can filter on „Pizza Place“. This allows us to find all the pizza places within the Neighbourhoods.

Finally, we will perform a clustering on the data. We will use the k-means clustering. It identifies k number of centroids and allocates the specific Neighbourhoods to one of the centroids. Due to its simplicity it is one of the most popular unsupervised machine learning algorithms. The frequency of occurrence of Pizza Places within the Neighbourhood will be the main factor for the k-means clustering. This clustering will help us to define suitable and less suitable Neighbourhoods to help us answer the question of where to best locate a new Pizza Place. Finally, the results will be visualized showing a marker for each Neighbourhood.

Results

The results have been visualized in the following graphic. Green points show neighbourhoods with a high density of pizza places. Violet places have a medium density of Pizza Places whereas the red markers show the neighbourhoods with very few Pizza Places.



Discussion of results

We can see a lot of red markers within the Borough of Manhattan. This means that the density of Pizza Places directly in Manhattan seems to be low. Therefore, Manhattan could be one of the more competitor-less areas of New York for Pizza Places.

It seems that more outside of the city there is a higher density of Pizza Places. Therefore, the direct center of the city could be a more attractive choice to open a Pizza Place since there are less direct competitors.

Recommendation and observations

Interesting locations seem to be various areas of Manhattan, although other areas are placed with red markers as well. Depending on the pricing strategy of the restaurant there could be interesting places more outside of the city centre such as in the Bronx or Queens.

One could assume that in the areas where most of New York's population lives (almost everywhere but in the center of Manhattan) the density of restaurants looks higher than in the direct city center. Maybe the demand for food is higher in those areas.

Conclusion

After giving an introduction into the topic in the first chapter of this report we discussed the problem which this result aims to solve. The aim was to find an area which could be suitable to open a new Pizza Place. Stakeholder of this report was the Luigi Family although this kind of report would be useful for everyone who would aim to open an Italian restaurant within New York City. After addressing the results of this report we came to the following recommendation. A pizza place should be located in either Manhattan or other areas where the red markers are located. Besides looking at this visualization of the results another location analysis should be done. There are more factors such as distance to food markets for supply of ingredients as well as crime rates or even more specific factors such as rent prices or median income of the neighbourhood to develop a fitting pricing strategy for the restaurant. This report should be used as the first filter of finding the right location for a Pizza Place.