

Applied Data Science Capstone Project

(week 4)

Introduction & Business Problem

Starting a new business is a big risk. However, there are a lot of things that we can do to mitigate that risk. One of these things is using data science to find out the prime location to set up shop. If you pick a location with high foot traffic, affordable rent and that's quite safe to boot, you can limit the risk involved.

In this report, we'll be analysing New York City to find the optimal location for a Belgian restaurant. In order to find the best possible neighbourhood, we'd need to consult multiple data sources. As New York City is quite large, it's more efficient to start looking at the city on a borough-level, rather than a neighbourhood-level. As such, this report will include a first analysis of on borough-level to determine the best borough, before we drill down further and examine each neighbourhood.

Data

Borough-level data

When setting up a business, it's crucial that the surrounding area is lively and willing to spend. As such, we'll need to compare each borough based on **GDP** and **total population**. This data is available on Wikipedia (https://en.wikipedia.org/wiki/Boroughs_of_New_York_City) in a table format. We can scrape this data into a dataframe to use it for further analysis.

However, these aren't the only important metrics to compare. A new business benefits from a safe environment, which means we also have to take crime rates into account. We can get a fragmented overview of the **number of crimes per borough** through the NYPD-website

(<https://www1.nyc.gov/site/nypd/stats/crime-statistics/borough-and-precinct-crime-stats.page>). The NYPD only lists these per borough, so we'll have to compile the Excel-files ourselves before we are able to conduct a full analysis.

Ideally, we will find a borough that has a high GDP & population, while having a relatively low crime rate.

Neighbourhood-level data

Once the ideal borough has been found, we can zoom in and have a look at the best neighbourhood in that borough. There are a few important factors that will influence the decision.

First of all, we need to have a look at the **other businesses in the neighbourhood**. We could either look for a neighbourhood with a high density of restaurants of multiple cuisines, as this would be a neighbourhood that's frequented by foodies who are looking to try something new.

On the other hand, we could also look for a neighbourhood that has a very small number of restaurants so that the competition is rather limited. However, we have to keep in mind that having an existing Belgian restaurant in a neighbourhood would make business harder than it has to be. All of this data will be obtained through the Foursquare API

(<https://api.foursquare.com>)

Another thing to keep in mind is the average rent in the neighbourhood. Rentcafe gives an overview of the average rent per neighbourhood. You can consult the Manhattan rent per neighbourhood here: <https://www.rentcafe.com/average-rent-market-trends/us/ny/manhattan/>.

It is worth noting that these are not the rental prices for retail spaces, which are generally more expensive. However, as there was no trustworthy retail space rental data readily available, this data will give us an indication of how expensive it will be to set up our business in that neighbourhood.