

# The perceived quality of neighborhood amenities

A case study set in The Hague, Netherlands

for:

IBM Applied Data Science Capstone

April 2020

By: Henry Veenbergen

# Introduction

- The Hague is a city located on the west coast of The Netherlands. With around 550,000 inhabitants, it is the country's 3rd largest city.
- The city houses the country's government, houses of parliament and many of the country's civil servants in the many ministries. These are mostly located in the city centre.
- The heart of the city is complemented by large numbers of shops, bars and restaurants.
- On the western end of the city are the Scheveningen & Kijkduin beach areas, which are very popular with tourists and locals alike
- In and around all this are all the neighborhoods, which together form a highly diverse city.



# The problem

- As in most cities, the neighborhoods in The Hague vary greatly, for example in terms of:
  - Wealth
  - Ethnic composition
  - Age distribution
  - Degree of urbanisation
- At the same time, each of these neighborhoods would have many of the same types of amenities, such as:
  - Restaurants
  - Café's & Bars
  - Groceries
  - Bakeries

While one would intuitively expect some of the neighborhood characteristics, most notably wealth, to affect the quality of amenities, this may not necessarily be the case. The lack of a readily available, objective way of measuring quality complicates the issue.

The question therefore is twofold:

- Does the perceived quality of amenities vary between neighborhoods?
- Can key neighborhood characteristics explain this difference?

These questions may be relevant for a range of people, including entrepreneurs looking to open a new business and (prospective) residents.

# The data

The data that will be used for this project comprises:

1. Information from the Dutch National Statistics office (CBS) with respect to Dutch cities and neighborhoods. This dataset contains a wealth of key demographic and economic data for each neighborhood in The Hague.
2. Geodata for each of the neighborhoods in The Hague
3. Foursquare venue data for each of the neighborhoods in The Hague
4. Foursquare ratings data for each venue (where available)

Foursquare allows for a deeper look into each of the venues we have retrieved. We are looking for data that says something about the quality of the venue. The best available metric is customer rating. However, there are several problems with this, including:

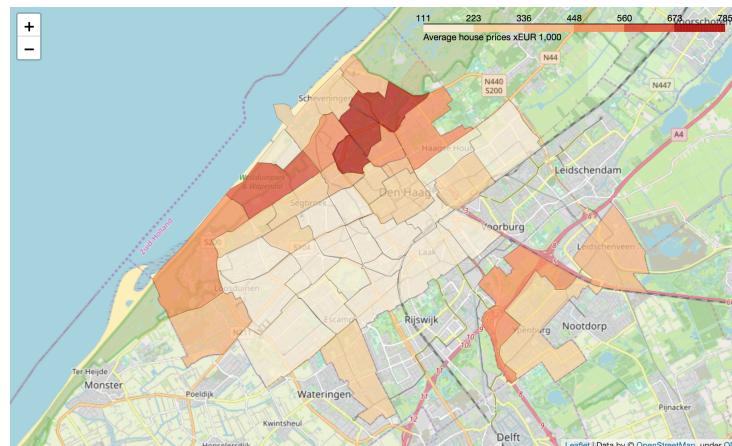
- It is subjective
- It may be biased (only very satisfied & very dissatisfied client submit a rating, as may the owner's friends)
- There may be a lack of data point (only a few ratings)
- The price of a product may not be properly taken into account by a customer (e.g. what can you expect from a EUR 5 pizza vs a EUR 20 one?)

Nonetheless, in the absence of objective quality data, subjective data is all we have, so we have to work with it. But keep its limitations & pitfalls in mind!

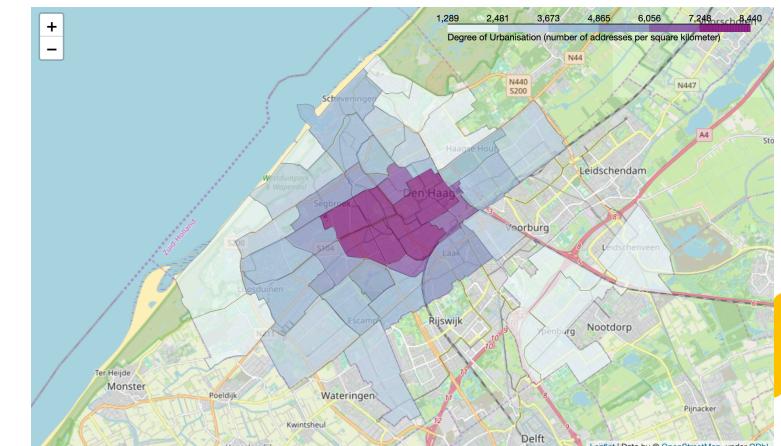
# Methodology (1)

- The Dutch Statistics Bureau CBS publishes a wealth of socio-economic information on each neighborhood in The Hague.
- After assessing the type, quality and completeness of information, it was decided to proceed with 2 key characteristics:
  - Average house price – this serves as a proxy for the wealth & income of the neighborhood
  - Degree of urbanization – this serves as a proxy for “sophistication”. The working assumption is that the more urban the area, the more demanding its residents are when it comes to quality of amenities.

House prices:

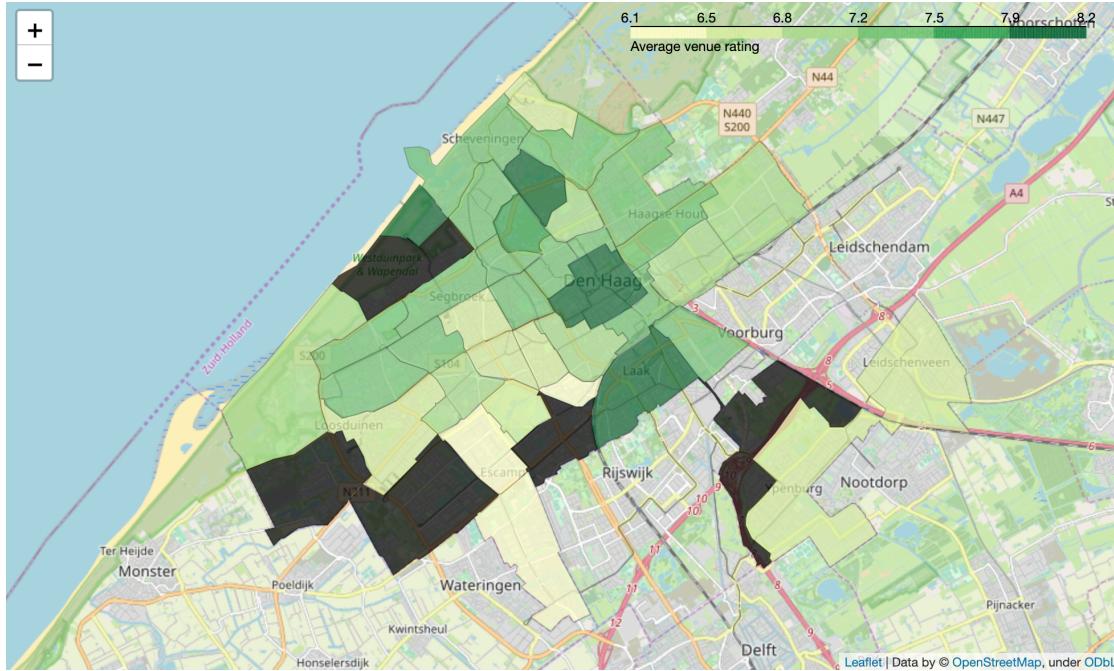


Degree of urbanization:



# Methodology (2)

- Next, we retrieve the venue detail information for each of the venues from Foursquare. This contains a wealth of information, yet we only use ratings and discard the rest. There is a substantial difference in the average venue quality across the neighborhoods in The Hague:



- Finally, we are able to look for correlation and statistical significance.

# Results & conclusion

- A regression of average venue rating against the degree of urbanization shows an almost completely random distribution:
- For the average house price, there does appear to be a modest positive correlation:
  - The Pearson Correlation Coefficient is 0.30 with a P-value of 0.07
  - This means there could well be a correlation between neighborhood wealth and the (perceived) quality of amenities, yet at best a weak one.

