

Coursera IBM Capstone Project: selecting the best city for opening a vegan or vegetarian restaurant

1 Introduction

In this report, I will describe how I combined multiple data sources to build a scoring model that predicts which of the Dutch cities is best suited for opening a vegan or vegetarian restaurant. I've combined the following sources of data for this project:

- The Foursquare API, to retrieve information on existing restaurants per city.
- The government's Central Bureau of Statistics (CBS) data, to retrieve regional demographic and economic information.
- Funda in Business, the largest real estate website, for the average retail property price.
- The National Institute for Public Health and the Environment (RIVM), for data on people's eating habits.

From this data, I've distilled 5 variables that add up to produce a suitability score for each city. This quantitative tool will not provide definitive answers as to which city is 'best' in an absolute sense. It will, however, provide future restaurant owners with a short-list of cities for more in-depth, qualitative analysis.

1.1 Target Audience

This tool is built for people in the Netherlands that are interested in opening a restaurant but are not tied to a specific geographic location. This might be a private entrepreneur looking for the most suitable city to maximize his chance of success, or a chain manager looking to add another restaurant to their holdings. The scoring model is malleable to the users' needs and preferences. With a few minor tweaks, it could be used to find the best location for other types of restaurants like Italian or Thai. The importance of each variable in the model, like price per square meter or market saturation, can easily be adjusted to the users' preferences by changing the weights associated with each variable.

1.2 Business problem

More than half of all start-ups fail within the first five years (1), and restaurant owners are not exempt from this demoralizing fate. One of the main aspects that lead to early failure is inadequate market research or a lack of market knowledge in general (1). Large businesses might have teams conducting feasibility studies, but private entrepreneurs usually just go with their gut feeling. The tool presented in the report brings market analysis to people on a budget. It is free because it uses public data. Going with gut or intuition is valuable, but not sufficient. It should be combined with quantitative, objective analysis based on independent variables. This scoring model is a useful first step to filter out unsuitable cities. As a second step, an entrepreneur might compare the top cities based on more intangible, intuitive criteria.

1.3 Business understanding: Vegetarian is booming

In the Netherlands, there is an increased awareness about the negative environmental impacts of meat and dairy consumption. According to a large survey by *Natuur en Milieu*, 61% of the Dutch are

aware of the disproportionate amount of greenhouse gases and water use associated with raising livestock (2). 25% of respondents avoid meat because of their aversion against industrial farming practices (2), which is associated with the inhumane treatment of animals, excessive use of antibiotics, soil degradation and water pollution. A third of the respondents avoided meat because of its negative impact on human health (2). This comes as no surprise after the World Health Organizations' (WHO) advice to avoid red and processed meats because of their carcinogenic effects (3).

This increased awareness has had a direct influence on peoples' eating habits. On average, the Dutch only eat meat 4.8 days a week. 37% of the survey respondents are cutting back on the amount of meat they consume each week (2). Of this sub-group, 41% have even cut their meat consumption in half (2). Moreover, the total number of vegetarians and vegans has increased from 1.1 % in 2010 to 4.4% in 2014 (4). The number of flexitarians – people who avoid meat at least 3 days a week – has increased from 14% in 2011 to an impressive 43% in 2019 (4).

According to the *Natuur en Milieu* survey, more than half of the respondents said they wanted to eat more vegetables while eating out. Only a fifth of the respondents is impressed with the quality of the vegetarian options (2). In other words, there's a lot of room for improvement in the hospitality sector.

Some entrepreneurs have responded to this trend by opening vegetarian restaurants throughout the country. But some cities have responded better than others. This creates an opportunity for entrepreneurs to open restaurants in cities where there's a mismatch between supply and demand. For example, Zaanstad and 's Hertogenbosch have the same number of inhabitants (156.000) but Zaanstad has twenty-two vegetarian venues where 's Hertogenbosch has just seven.

2 Data

2.1 Analytic Approach

From an entrepreneurs' perspective, when is a city the 'best' option for opening a restaurant? I've defined this as the city with the highest potential earnings and the least chance of foreclosure. From this perspective I started thinking about picking a model. I concluded that the model I would choose should fit the available data.

I considered building a regression model using multiple independent variables to predict a business's yearly earnings in Euros. I could have checked if there's a correlation between profit and city size, and profit and education level. Unfortunately, restaurants do not publicly share their yearly earnings for this kind of analysis. Nor do they indicated when they go bankrupt. I was forced to try another approach, and the second-best option was to build a scoring model as follows:

$$\text{Suitability Score (SS)} = v1 * w1 + v2 * w2 + v3 * w3 + v4 * w4 + v5 * w5$$

Where SS is the suitability score for a city

v stands for an independent variable

and w stands for the weight associated with the variable, which is based on the user's preference and experience.

2.3 Data Description

The variables below were used in building the scoring model. I will discuss why these variables have been chosen in the methodology chapter.

Variable per city	Source
1. List of city names	CBS Statline
2. Inhabitants	CBS Statline
3. Number veg. or vegan restaurants	Foursquare API
4. Total number of restaurants	CBS Statline
5. Number of tourists	CBS Statline
6. Average rent price per M ²	Funda in Business
7. Number of vegetarians or vegans	RIVM

2.3 Data Collection

Even though I could eventually find all the required data, it was rarely straightforward. I had to use a lot of workarounds to get to the information, or I had to use a similar variable as a proxy for what I truly wanted but could not acquire. The numbers below correspond to the variables in the table above.

1. Getting a list of cities wasn't as straight forward as I had imagined. CBS did not collect data on a city level but on a municipal level. Municipalities are often bigger than cities and can also contain villages. Because I was only looking at cities with more than 100.000 inhabitants, there was a good match between the names of the municipalities and names of the cities. For example, the city of Amsterdam was inside the municipality which was also called Amsterdam. Out of the 30 municipalities with a population of over 100.00, only two did not have a large city within its borders. I removed these outliers from the dataset.
2. Getting the number of inhabitants per municipality was straightforward from CBS Statline.



Figure 1: Selected municipalities with a population over 100.000.

- The first step in getting the number of vegetarian restaurants per city was adding location data to my data frame (DF). I imported latitude and longitude coordinates through the ArcGIS API using the Geocoder library in Python. The second step was connecting to the Foursquare API with a request for vegetarian venues within a 10-kilometre radius from the city centre. I summed up all the vegetarian venues per city and added them to the data frame (see fig.2).

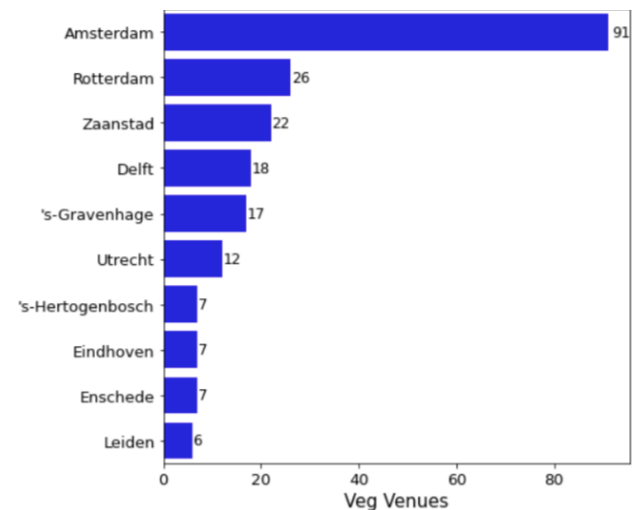


Figure 2: Top 10 cities with the most Vegetarian Venues.

- The Foursquare API was of limited use when trying to figure out the total number of restaurants in a city because a query was limited to a maximum of 100 results only. Fortunately, the Central Bureau of Statistics had counted the total number of restaurants in a radius of 5 kilometres from the city centre (see fig 3).
- I was unable to find the number of tourists on a city or municipality level. CBS counted the number of tourists on the province level. I had to find a workaround for this problem, with data available on the appropriate level. After looking through CBS's database for an alternative data source, I found a variable that showed the number of people working in the hospitality sector. This sector corresponds with tourism quite well as it covers jobs in: Hotels, restaurants and cafés. Municipalities with a lot of tourism should have relatively more jobs in hospitality (see fig. 4).

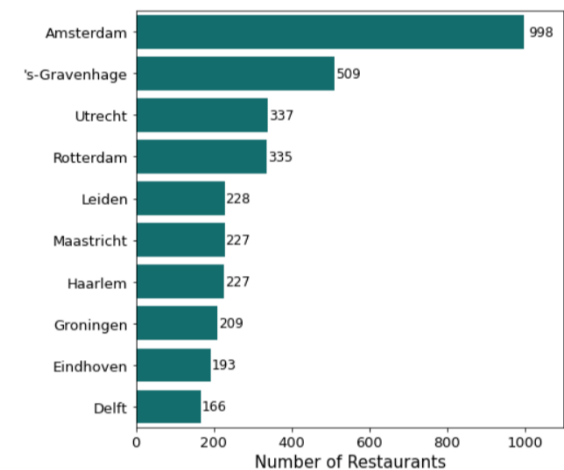


Figure 3: Top 10 cities with the most restaurants.

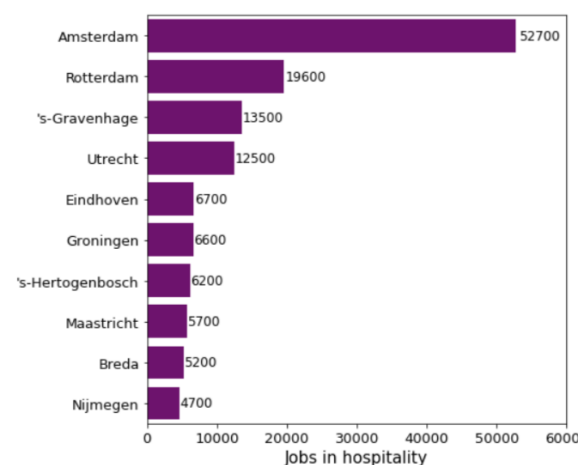


Figure 4: Top 10 cities with the most jobs in the hospitality sector.

6. After exploring public databases and reports, I could not find the average retail price per square meter for each municipality. I decided to use the Beautiful Soup library to scrape the prices from fundainbusiness.nl – the largest retail broker in the Netherlands. Although this was technically challenging, it provided me with accurate and up-to-date information on the price of property (see fig. 5)
7. Although there was a lot of data on vegetarians and vegans on a national level (see chapter 1.3), there wasn't any data available on a local level. There was no data showing the percentage of vegetarians and vegans per city. I decided to use education as a proxy for vegetarianism. According to The National Institute for Public Health and the Environment, adults with a high education level eat less meat, and more vegetables than average (5). There was to be a clear correlation between eating habits and education. This was not so evident with age. Education level data was available on the municipal level through CBS Statline (see fig. 6).

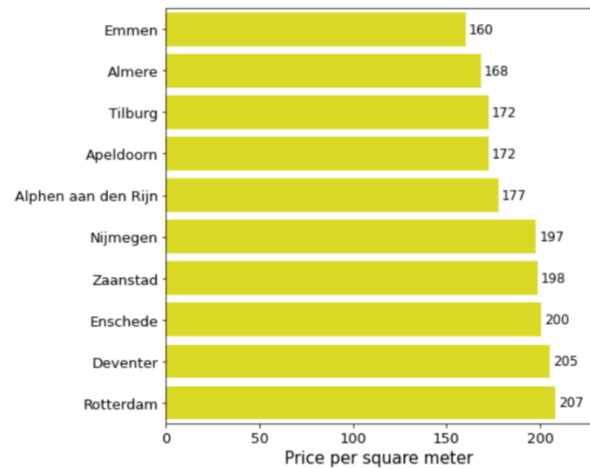


Figure 5: The cheapest with the cheapest property prices.

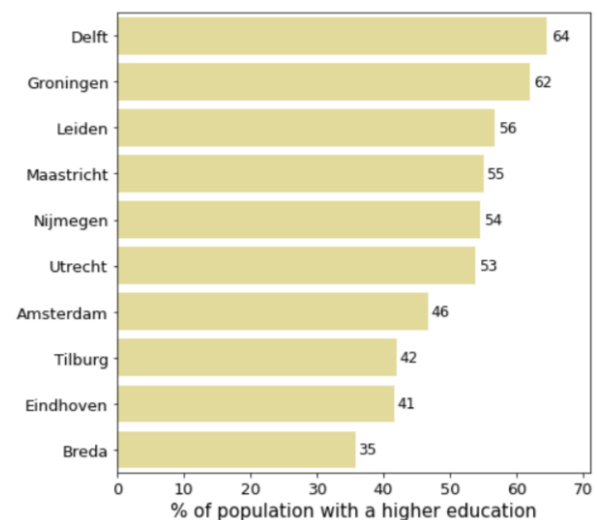


Figure 6: Top 10 cities with the highest levels of education.

References

1. **Bakker, P.** Business Partners. [Online] <https://businesspartners.nl/startups-de-echte-cijfers-en-hoe-je-wel-kunt-slagen/>.
2. **Onderzoek Natuur & Milieu. s.l. : Natuur & Milieu, 2017.**
3. **Organisation, World Health.** *Cancer: Carcinogenicity of the consumption of red meat and processed meat.* 2015.
4. **Waart, S. de.** *Factsheet 1: Consumptiecijfers en aantallen vegetariërs.* s.l. : Nederlandse Vegetariërsbond, 2020.
5. **Environment, National Institute for Health and the.** *Consumptie van voedingsmiddelen.* [Online] 2016. <https://www.waateetnederland.nl/resultaten/voedingsmiddelen/consumptie>.