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 (1).
- Funda in Business, the largest real estate website, for the average retail property price. (2)
- The National Institute for Public Health and the Environment (RIVM), for data on people's eating habits (3).

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 (4), 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 (4). 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 (5). 25% of respondents avoid meat because of their aversion against industrial farming practices (5), 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 (5). This comes as no surprise after the World Health Organizations' (WHO) advice to avoid red and processed meats because of their carcinogenic effects (6).

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 (5). Of this sub-group, 41% have even cut their meat consumption in half (5). Moreover, the total number of vegetarians and vegans has increased from 1.1 % in 2010 to 4.4% in 2014 (7). 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 (7).

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 (5). 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:

*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. In this report, I use the word municipality and city interchangeably, because they almost mean the same thing in this context.

Figure 1: Selected municipalities with a population over 100.000.

2. Getting the number of inhabitants per municipality was straightforward from CBS Statline.

3. 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).

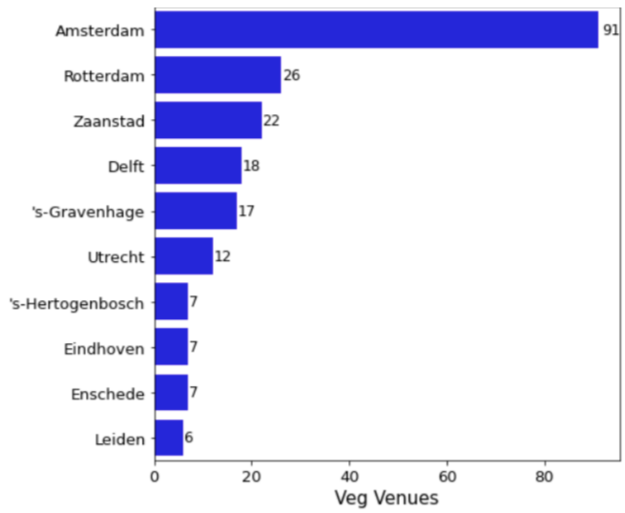


Figure 2: Top 10 cities with the most Vegetarian Venues.

4. 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).

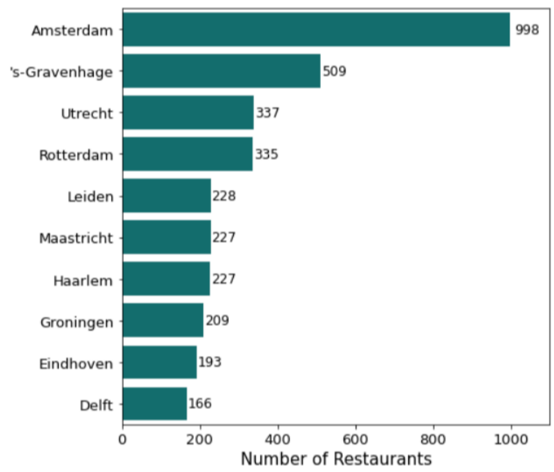


Figure 3: Top 10 cities with the most restaurants.

5. 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).

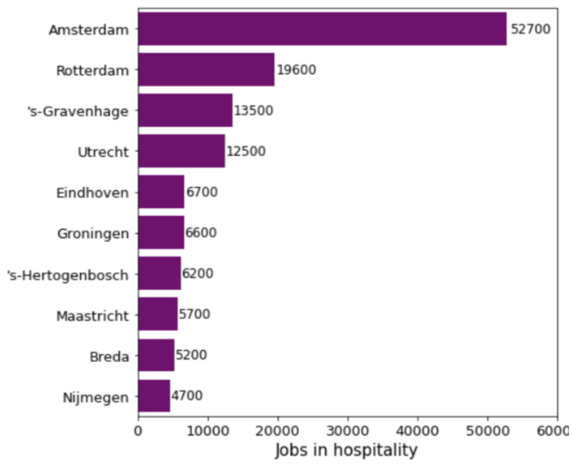


Figure 4: Top 10 cities with the most jobs in the hospitality sector.

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 per city (see fig. 5)

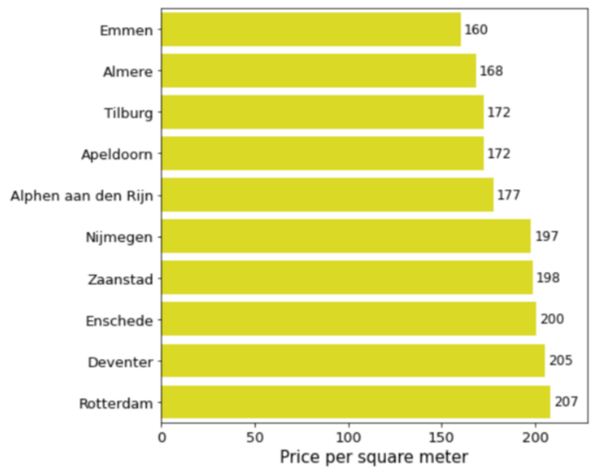


Figure 5: The cheapest with the cheapest property prices.

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. No data was 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 (3). There was a clear correlation between eating habits and education. This was not so evident with age or other variables. Fortunately, education level data was available on the municipal level through CBS Statline (see fig. 6).

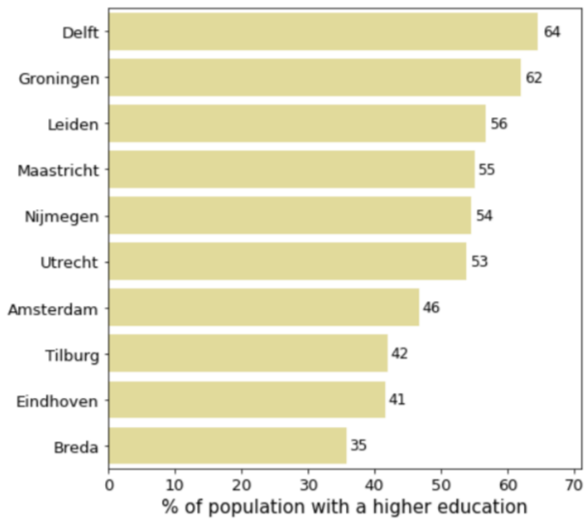


Figure 6: Top 10 cities with the highest levels of education.

3 Methodology

3.1 Data Preparation

To add the data my data frame in general, I had to make some adjustment to it by either transforming or cleaning it. Throughout the project, I had to rename variables from Dutch to English. Sometimes, I had to drop rows with missing values, and rows with data on municipalities that were not part of my selection (28 largest by population). Sometimes I had to change a column's data type from string to integer or vice versa. Sometimes I hat to strip away the empty spaces in the data or column names.

3.2 Preparing the variables

<u>Municipalities and Population.</u>

I started with a CSV file from CBS Statline that included the names of all municipalities (dependent variable) in the Netherlands, and their population in 2019. I created a filter to drop all municipalities with less than 100.000 inhabitants. Investing in a city that is too small is a liability to an investor. There might simply not be enough customers for a restaurant with a very specific niche. That is not to say that starting a restaurant in a smaller city is doomed per se. It's simply a matter of risk avoidance.

The municipalities Westland and Haarlemmermeer did not have a city within their boundaries. I studied the regions on Google Maps. They contained an agglomeration of villages near the larger cities of Amsterdam and Rotterdam. I removed these outliers from the dataset. I also renamed some municipalities to match the names used by ArcGIS.

<u>Vegetarian Venues</u>

Before I could make a query from the Foursquare API, I needed longitude and latitude data for each city. I accessed ArcGIS location data through the geocoder library and added coordinates for each city to my data frame.

Subsequently, I defined a function that takes all municipalities with their location data as inputs and outputs a data frame called *nearby_venues.* This data frame contains – for each city – all the vegetarian or vegan restaurants in a 10 km radius from the city centre *(287 in total).* I grouped all restaurants by their municipality and counted them. I added this count to my main data frame(see fig 2).

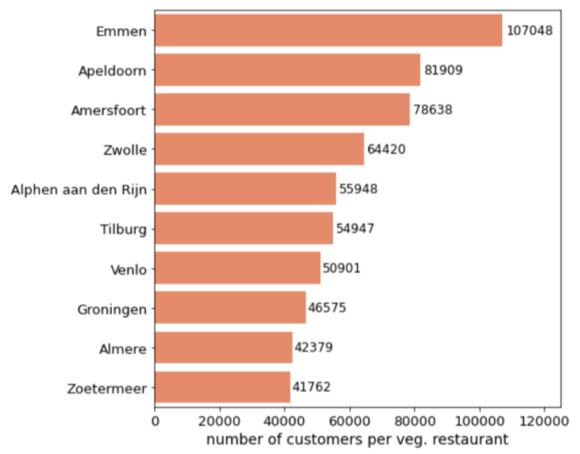


Figure 7: Cities with the least direct competition

Cities like Amsterdam, with a lot of vegetarian venues, are not great for a new investor trying to break into the market. These venues are the direct competition, after all. It would be natural to select a city with the least competition. But before I could use this data in my model, I had to correct for population size. Cities with large population naturally have more restaurants. To get the relative variable, I divided the population by the number of vegetarian venues. I saved the result in a new column as [VegRestaurantDensity]. For Zwolle for example, the value is 64420, which means that one restaurant has a potential customer base of 64420. The higher this number, the more customers per restaurant, the less competition (see fig 7).

Total Venues

The total number of restaurants is also a relevant metric, because restaurants belonging to another category are also part of the competition. I call this segment the indirect competition. Customers might, for example, choose to go to an Indian or Thai restaurant for their vegetarian meals, instead of to our hypothetical restaurant. Before adding this category to the scoring model, I had to correct for population size, in the same way as I did with the vegetarian restaurants.

I created a new column called [TotRestaurantDensity] where population is divided by the total number of restaurants. For Emmen for example, the value is 4198, which means that one restaurant has a potential customer base of 4198. The higher this number, the more customers per restaurant, the less competition (see fig 8). Note that population will not be used directly in our scoring model. It is used as a part of other variables.

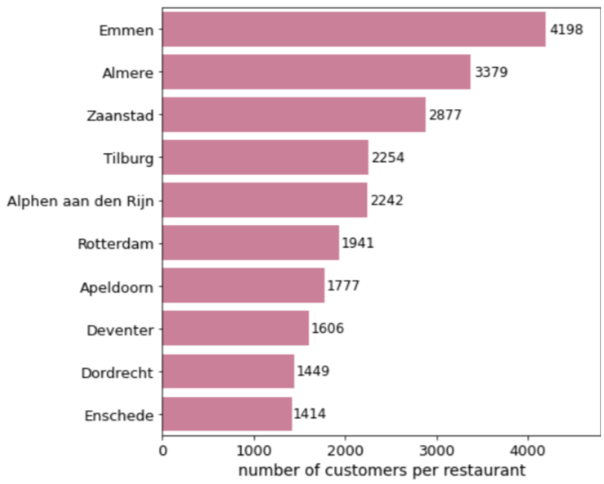


Figure 8: Cities with the least indirect competition

Tourism

Tourism can be a major source of additional income for a restaurant owner. Tourists are an additional group of seasonal customers, next to the steadier local population. As described in chapter 2, I could not find data on the number of tourists per municipality. I chose jobs in hospitality as a reliable proxy for tourism.

Like before, a correction for population size had to be made, because larger cities can have more jobs in hospitality due to population size alone. I, therefore, divided the population size by the number of people working in this sector. I saved this data in a new column called [TourismIndex]. For example, Amsterdam has a population of 872757 and 52700 jobs in hospitality 872.757 / 52.700 = 16.6. This means

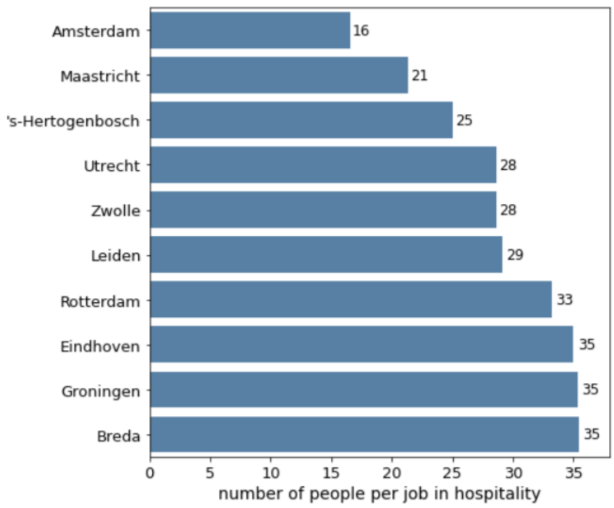


Figure 9: Top 10 most touristic cities

that 1 job serves 16.6 people. The lower this number, the more touristic an area is (see fig 9).

Retail price

The amount a restaurant owner pays in rent per square meter is perhaps the most important variable to be considered. Rent is often the biggest cost for a starting entrepreneur. For a new restaurant owner, a low price per square meter means that she has a longer time frame in which to set up her successful business. Low rent can also lead to lower prices on the menu. This will increase the price/quality ratio and in turn attract more customers.

Before I started web scraping with the Beautiful Soup library, I created a new data frame called retail_prices. I added a municipalities column and an empty column for the average price per square meter. I made some corrections the names of my municipalities to match the names of Funda In Business site.

I created a for loop that looped through 2 pages per city on fundainbusiness.nl (30 entries per city), took the average from these entries, and added these values to my data frame(see fig. 5).

Education Level

As stated in chapter 2, education level is a proxy vegetarianism, veganism and flexitarianism. People with a higher education eat less meat and more vegetables. The differences are significant, but not that large, so this variable is the least important of all five in the predictive model. I expected a difference in meat consumption between different age groups, but this was not evident in the data (3).

I imported data on education from CBS Statline into a new data frame called education. There were four levels of education in the data. Level 1: secondary education or lower, level 2: vocational college, level 3: higher vocational education, and level 4: university. Higher education as defined by the RIVM (3) is education on level 3 and 4. I converted the numbers per municipality from absolute to relative (%). I added up the percentages of level 3 and 4 to get a percentage per municipality (see fig 6). In Delft, for example, 64% of all inhabitants are highly educated. This unusually high number is because Delft is a student town, with a large technical university. I saved this data in a new column called [HigherEdu].

I combined all five variables (columns) to a new data frame(df2) and I was finally ready to build the scoring model.

	Municipalities	VegRestaurantDensity	TotRestaurantDensity	TourismIndex	RetailPriceM2	HigherEdu
0	Amsterdam	9591	874	16.6	342.10	46.7
18	Maastricht	30394	533	21.3	325.95	55.0
6	's-Hertogenbosch	22159	1317	25.0	286.70	26.9
5	Utrecht	29800	1059	28.6	333.30	53.8
27	Zwolle	64420	1224	28.6	226.00	27.6
9	Leiden	20850	548	29.1	239.80	56.7
1	Rotterdam	25044	1941	33.2	207.85	32.5
7	Eindhoven	33485	1213	35.0	243.00	41.7
10	Groningen	46575	1112	35.3	289.35	62.0
13	Breda	36814	1388	35.4	231.55	35.8

Figure 10: a snapshot of data frame2 to be used for building the scoring model.

3.3 The Scoring Model

Now that all the data is prepared and ready, it is finally time to create the scoring model.

The model looks like this:

*Suitability Score = - [RetailPriceM2] * w1 + [VegRestaurantDensity] * w2 + [TotRestaurantDensity] * w3 - [TourismIndex] * w4 + [HigherEdu] * w5*

Where wN are the weights associated with each variable.

Retail Price and Tourism index are deducted from the score because a high value (like rent) is considered negative. The other variables have a positive influence on the score and are therefore added.

The model, as it is now, would not work well because the variables are not normalized or scaled. If we look at Maastricht's vegetarian restaurant density the number is 30394. Compare this to its higher education which is only 55. This would mean that restaurant density would influence the suitability score way more than higher education would. This is the reason we need to normalize – i.e., scale all values to a number between zero and one. Zero would correspond with the lowest value and 1 with the largest. For example, retail price is 0 in Emmen as it is the cheapest place and 1 in Amsterdam because it is the most expensive. Using this method, only the weights in the formula will determine how important a variable is.

I used Sikit-Learn's MinMaxScaler to normalize the data and saved it to a new data frame called Normalized. I decided on the weights as shown below, based on my personal preference and reasoning.

*Suitability Budget = - [RetailPriceM2] * 0.4 + [VegRestaurantDensity] * 0.2 + [TotRestaurantDensity] * 0.2 - [TourismIndex] * 0.12 + [HigherEdu] * 0.08*

In my opinion, retail price would be the most important variable (0.4), followed by the direct (0.2) and indirect (0.2) competition (total 0.4), followed by tourism (0.12) and higher education (0.08). I called this option Budget because it weighs heavily towards a low rent.

Because choosing weights is subjective, I've also built a model where all weights are equal.

*Suitability Equal = - [RetailPriceM2] * 0.2 + [VegRestaurantDensity] * 0.2 + [TotRestaurantDensity] * 0.2 - [TourismIndex] * 0.2 + [HigherEdu] * 0.2*

And a third model called Balance, which lies somewhere in between the Budget two Equal model.

*Suitability Balance = - [RetailPriceM2] * 0.3 + [VegRestaurantDensity] * 0.2 + [TotRestaurantDensity] * 0.2 - [TourismIndex] * 0.17 + [HigherEdu] * 0.13*

After adding the results to the Normalized data frame, I applied a transformation, so that there would be no negative score. For example, the lowest score in the Suitability Budget was -0.32 for Amsterdam. I added 0.32 to the whole column. Now Amsterdam has a value of 0 and the highest-scoring city (Emmen) has a score of 0.6.

Because I wanted to know which cities performed well in all three versions of the model, I added the scores up and saved it in a new column called [CityScore].

4 Results

4.1 Exploratory Analysis

Before looking at the results of the model, I studied the five variables separately to get a feeling for the data and the differences between the cities. I started with the most important variable [RetailPriceM2] (see fig. 5). The average retail price for all cities is 229€ per square meter. Emmen is the cheapest at 160€ per square meter and Amsterdam is the most expensive, more than twice the price of Emmen, at 342€ per square meter. Rotterdam stands out when looking at the data. While it is the second-largest city by population, it is still in the top ten of the cheapest retail prices at 207€ per square meter.

The mean [VegRestaurantDensity] is 38774 which means that on average a Dutch vegetarian restaurant has a theoretical customer base of 38774 people. Emmen has the largest potential customer base of 107048. This number is so high because currently there's only one restaurant labelled vegetarian. It's no surprise that Emmen little vegetarian venues. It's a border town near Germany, with little tourism, far away from the cultural centre of the Netherlands. Innovation and cultural change, start in the large cities, and smaller, more rural places like Emmen change more slowly. Delft has the highest density of 5755 because it is a student town. Students like to eat out. Plus, there is a correlation between education and vegetarianism.

The mean [TotRestaurantDensity] is 1513. Maastricht has the most restaurants relative to its population at 533 people per restaurant. This is because the province of Limburg has a fine dining and eating out culture and because Maastricht is a tourist hotspot. Again, Emmen has the least amount of restaurants relative to its population (4198). Rotterdam scores above average at 1941, which is very good for a large city.

The mean [TourismIndex] is 40 and ranges from the highest tourism density in Amsterdam (16.6) to the lowest in Emmen (71.4). Looking at tourism in isolation, Amsterdam scores well. The tourism index is not perfect because it's a proxy for tourism based on jobs in hospitality. I expected Rotterdam to have a higher score. Student towns also score a bit better than average because they have a lot of cafés and bars.

The mean for [HigherEdu] is 32 % with the lowest value in Emmen (11%) and the highest in Delft, where the people with a higher education is six times higher at 66%.

4.2 Model Expectations

Based on the exploratory analysis (chapter 4.1), I expect Emmen to do well in the scoring models. Because it scores best in three out of five categories. Emmen's scores will be dragged down a bit because it preforms worst in the other two categories.

Even though Amsterdam has the most tourism, I expect it to do poorly because it has the highest rent prices. It also has a very high direct and indirect competition. In other words, Amsterdam has a saturated market, with high rent and little room for new entrepreneurs.

Of the five largest cities, I expect Rotterdam to perform best due to its low retail price and below-average restaurant density.

4.3 Model Scores

Now, let's look at the top ten cities for opening a vegetarian restaurant according to the budget model (see fig. 11), the equal model (see fig. 12), and the balance model (see fig 13).

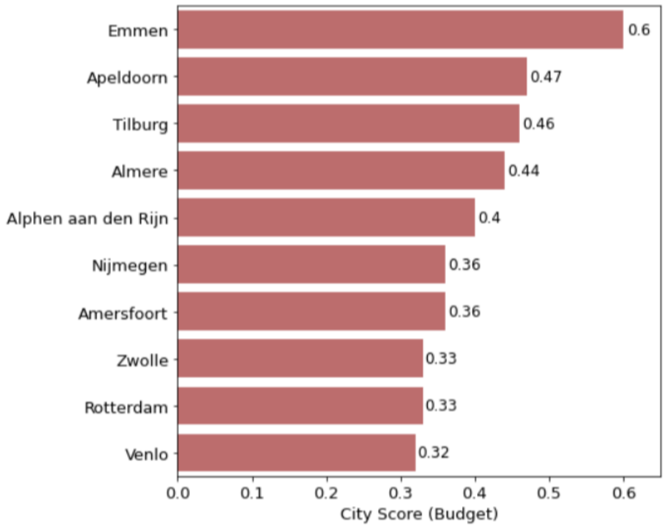


Figure 12: Ten best suitable cities according to the budget model.

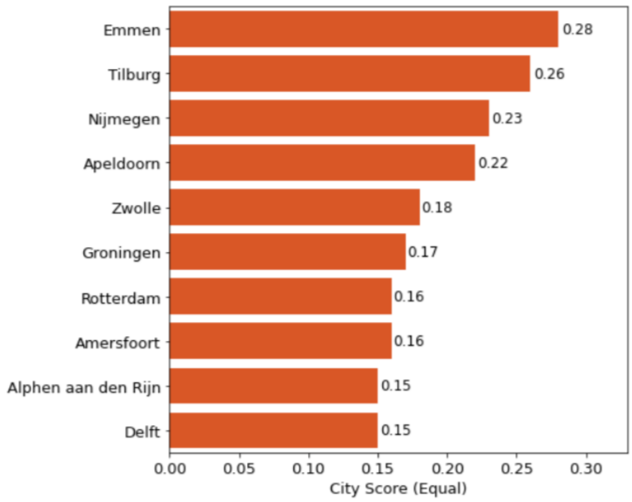


Figure 13: Ten best suitable cities according to the equal model.

As you can see, Emmen comes out as the best choice in all three models. I expected it to do well, but not this well. It performs especially well for investors on a budget. Tilburg comes in on the second spot with scores always in the top 3. Apeldoorn comes in third place with scores in the top 4. All top-scoring cities have one thing in common: low rent price per square meter.

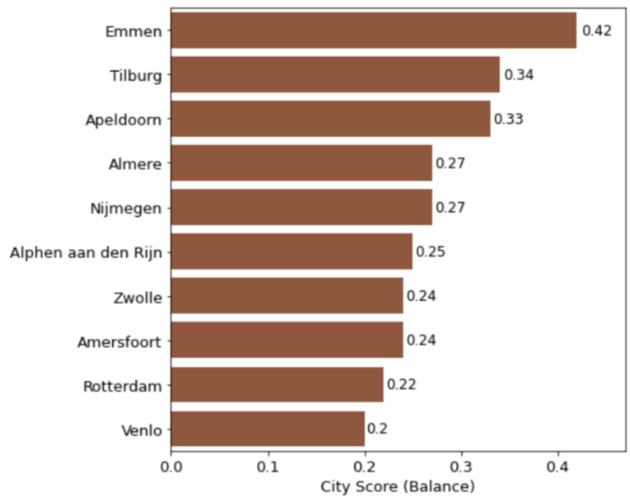


Figure 11: Ten best suitable cities according to the balance model.

The results of all three models combined are shown in figure 14 below. Large cities perform badly with 3 of the five largest cities at the bottom of the bar plot. The exception being, Rotterdam, which takes an impressive ninth place.

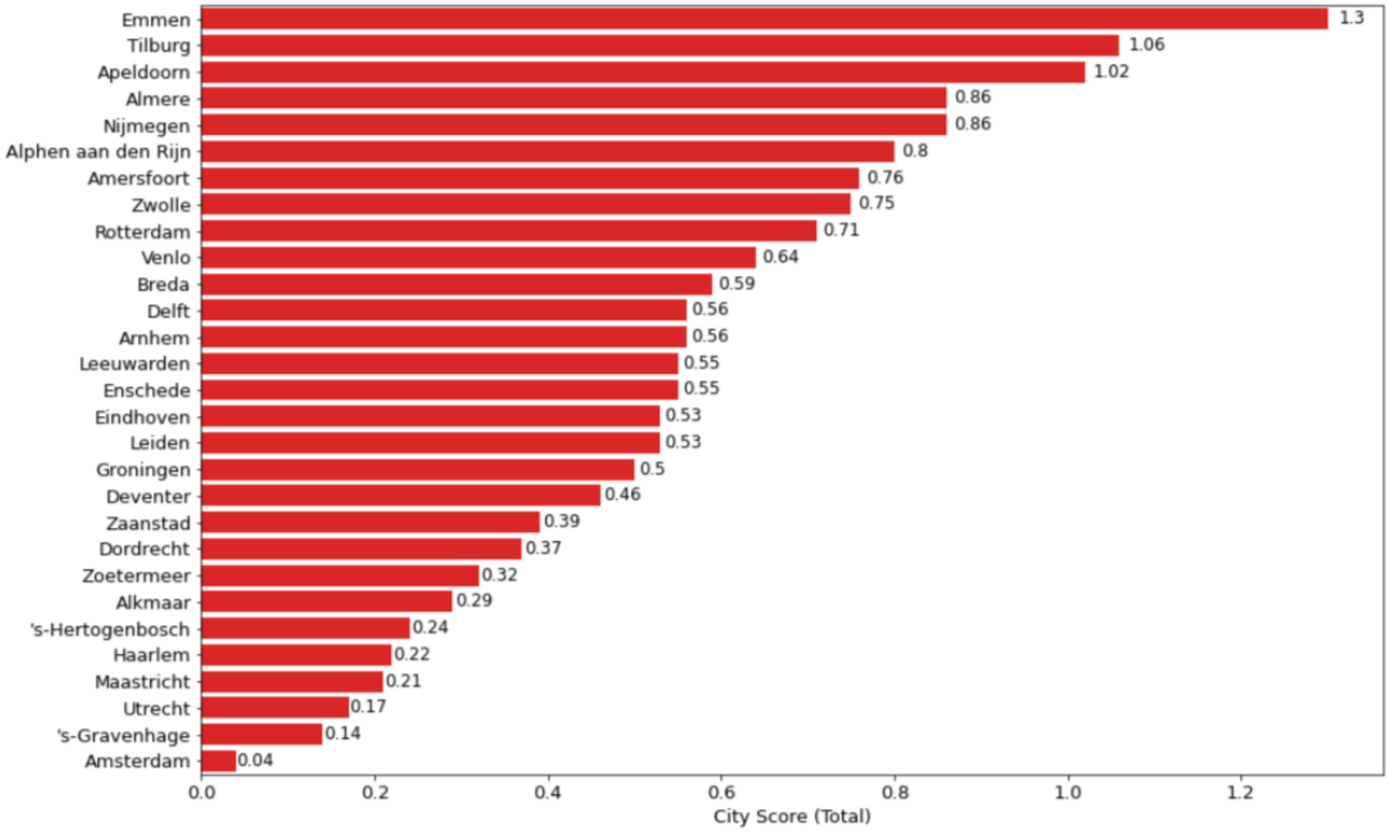
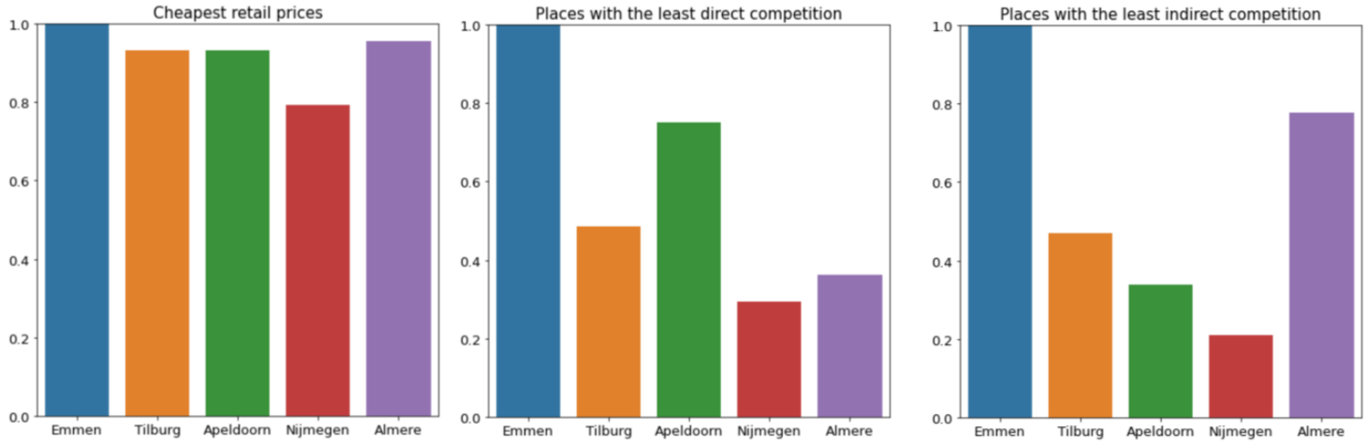


Figure 14: cumulative city scores based on the three scoring models.

5 Discussion

5.1 Interpreting the results: Post model evaluation

Choosing the best city is not as straightforward as figure 14 might indicate. To choose the best location, I will compare the normalized scores of the top five cities (see fig. 15). The cities with the highest City Scores are, in descending order: Emmen, Tilburg, Apeldoorn, Nijmegen and Almere. The variables [RetailPriceM2] and [TourismIndex] are negatively correlated to the City Score (subtracted). The other variables are positively correlated. This difference is confusing for the scores of the different cities. I took the inverse of [RetailPriceM2] and [TourismIndex] (1 – value) as input for figure 15. Now all values have a positive influence on the Suitability Score. A high bar in figure 15 means that a city scores well that particular category.



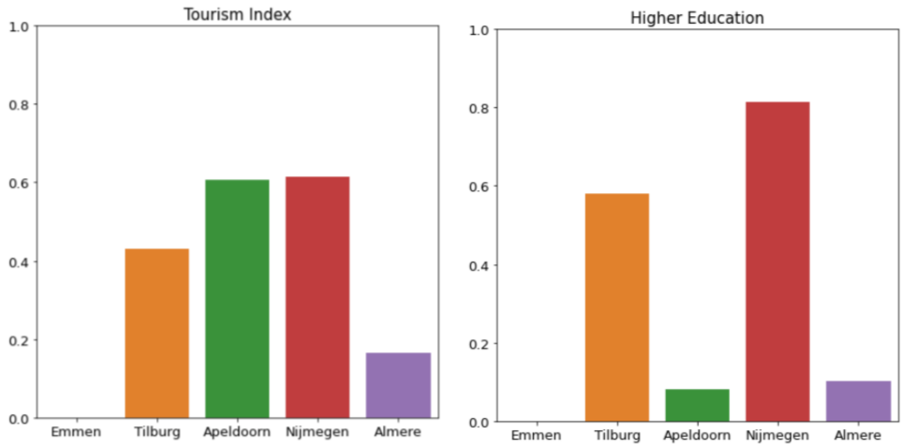


Figure 15: Normalized scores for the top 5 cities according to all three models.

If we look at Emmen (blue bars), we might conclude that investing in Emmen is risky. It scores maximum points on three categories but no points on the other two. Compare this to Tilburg (orange bars), a city that scores points across all five categories. We can say that Emmen has an unbalanced profile while Tilburg has a more balanced profile. Investing in Emmen is risky because we're not sure how important tourism and higher education are owning a successful restaurant. There might also be other factors, that fall outside the scope of this model, that explain the low retail prices and the unsaturated market for restaurants. Perhaps, the local culture is not supportive of eating out very often. Or, the population might be moving to other parts of the Netherlands, creating less demand over a longer timeframe. Income, which is not part of this model, might influence eating out habits. Investing in Emmen is the equivalent of investing in a highly volatile stock – high risk and a potentially high reward. Almere's profile (purple bars) looks a lot like Emmen's and could also be described as unbalanced.

Tilburg (orange bars) and Nijmegen (red bars) are similar. They are both cities with balanced profiles and with a high score for education. Both are student cities. Investing in one of the student towns, is a safer bet, because of the balanced profiles. This could be compared to investing in bonds on the stock market. Apeldoorn (green bars) is somewhere in between Emmen and the student towns. It scores poorly on education, but unlike Emmen, it scores well on tourism.

If I had to recommend a city to a potential investor, it would be Tilburg. Tilburg has a large population (219.000) of potential customers. It comes in second with a City Score of 1.06 and has a balanced profile. Direct and indirect competition are below average and rent prices are among the cheapest. Moreover, Tilburg is a student city, which tends to be progressive, and more open to alternative lifestyles and alternative sources of nutrition.

This analysis is done on a national scale. A more qualitative, regional study should be carried out before choosing a city. Things like the atmosphere of a neighbourhood, the feeling of safety, a regions' acceptance of outsiders, and its eating habits aren't things that could easily be quantified in a model.

6 Conclusion

In general, the scoring model holds up to my expectations. It's a useful tool for selecting a few cities if you want to open a vegetarian or vegan restaurant. It is especially useful for people with time or money constraints. This model uses readily available data that is accessible to everyone. The City scores themselves are not the only useful feature of this model. By using the model, the user goes through a process of selecting important variables and then deciding on how to weigh them. This process can help a user gain insight into his values and preferences. The model offers a systematic way to look at data and decide on more than just a hunch. Without the model, you will think of one variable and forget about the others. Our minds can't hold a lot of variables at the same time, let alone weigh them simultaneously.

<u>Model Weakness</u>

This model is far from perfect. Is has a few weaknesses, which are:

- [TourismIndex] is a proxy for the number of tourists, based on the jobs in hospitality corrected for population size. This is not ideal, because jobs in hospitality are only partly explained by tourists. For example, student cities have a lot of bars and people in the southern provinces tend to eat out more.
- [HigherEdu] is also a proxy, but for the amount of meat and vegetables people eat. Although Education influences meat and vegetable consumption, it is by far not the only variable that is of importance. I compensated for this weak link by assigning a low weight to the variable [HigherEdu].
- This model depends on assigning weights (which is arbitrary) instead of stronger methods like linear regression. Unfortunately, data was lacking for more advanced forms of analysis.

<u>Future improvements</u>

If I'd ever build a second version of this scoring model, I would change the following things:

- I would change from Foursquare to Google API for the venue data. I believe the Google database has more entries and is more up to date in the Netherlands. I would also spend more time checking the integrity of the vegetarian venues. Are the venues correctly labelled as vegetarian or are there a lot of false positives? I would also investigate the false negatives. How many venues are vegetarian or vegan but not labelled as such by the API?
- I would not allow negative values in the formula after normalization. I would deduct these values from 1 and turn them into positive scores. It makes for a more intuitive comparison of the different variables.
- I would research if income level influences the number of times people eat out. There's probably a correlation, and I would add income as a variable to the scoring model.

- I would try to find data on regional differences in the frequency restaurant visits and incorporate this into the model. Are there cultural differences between regions? I suspect that people in the south and west of the Netherlands eat out more often.

References

1. **CBS.** CBS Statline. [Online] opendata.cbs.n.

2. **Funda in Business. [Online] https://www.fundainbusiness.nl/winkel//.**

3. National Institute for Health and the Environment. Consumptie van voedingsmiddelen. [Online] 2016. https://www.wateetnederland.nl/resultaten/voedingsmiddelen/consumptie.

4. Bakker, P. Business Partners. [Online] https://businesspartners.nl/startups-de-echte-cijfers-en-hoe-je-wel-kunt-slagen/.

5. *Onderzoek Natuur & Milieu.* s.l. : Natuur & Milieu, 2017.

6. Organisation, World Health. *Cancer: Carcinogenicity of the consumption of red meat and processed meat.* 2015.

7. Waart, S. de. *Factsheet 1: Consumptiecijfers en aantallen vegetariërs.* s.l. : Nederlandse Vegetariersbond, 2020.