

# Fresh produce delivery to the hospitality sector in city centres

## 1. Introduction

### Background information

A Belgian startup company want to implement a new supply-chain concept for delivering fresh produce to the hospitality sector located in city centres. The company wants to anticipate the growing measures of cities to reduce motorised traffic in the city centres and provide a future-safe alternative for hotels, restaurants and others to get their fresh produce delivered to their location in the city centre.

Their business model is based on creating value for their customer by providing the following services :

- Bulk purchase of fresh produce on the early markets

*The customer will pay a lower price for the produce due to the ability of the start-up to purchase in bulk. In addition, the customer will save on employee cost as they no longer need to send staff to the early market.*

- Pay for what you use (reduce waste)

*A major benefit for the customer is that they no longer have to estimate the amount of produce they need each day and can eliminate the potential cost associated with wasted produce. The startup company will use modern data analysis techniques to predict the expected amount of produce to buy based on historical data, season, special events in the area, etc.*

- On-demand delivery within 30 minutes

*The start-up company will only deliver the agreed (minimal) amount of produce in the morning (or when requested). Additional produce can be delivered on-demand and will be delivered within a 30-minute time frame.*

- Delivery of the fresh produce with non-motorized vehicles

*The start-up company will deliver the produce with non-motorised vehicles. More and more cities in Belgium are implementing traffic restriction measures to improve air quality. It is becoming more and more difficult (and expensive) for the hospitality sector in the city centres to manage the delivery of goods to their facilities.*

### Problem description

The start-up company wants to deploy their service in selected cities in the Flemish region of Belgium. The problem they need to solve is to determine which cities to select first to secure the success of their go-to-market.

This means that they need to be able to cluster the potential cities based on the following criteria:

- The number of potential customers (Restaurants, Hotels, Pubs) in the city centre that can be serviced via non-motorized vehicles in 30 minutes (estimated radius of 2000 meters)
- Keep in consideration the value of the potential customers for the start-up company. The start-up company has determined that the difference in potential revenue can be defined as follows : Fastfood place (base reference), Hotel (1.5 times base), Restaurant (1.3 times base), Breakfast place (0.7 times base), Café/Pub (0.3 times base). This means that e.g. a typical Hotel will generate 1.5 times the revenue that a typical Fastfood restaurant will generate.

## 2. Description of the data

In order to solve the problem, the following data is required :

A list of the cities in the Flemish region of Belgian and their geo-data. This information can be derived from an online source that contains this info for Belgium ([http://www.gemeentezoeker.be/MonkeyProof\\_postcodes\\_developer.csv](http://www.gemeentezoeker.be/MonkeyProof_postcodes_developer.csv)) (source A).

A list of the number of residents for each of the cities. We want to limit are research to only the top 100 largest cities due to restriction with FourSquare usage and the limited potential of smaller cities for the start-up company. This information can be obtained online via the following source : (<https://www.metatopos.eu/belgcombiN.html>)(source B).

The location of selected venues in the top 100 largest cities in the Flemish region of Belgium. The venues we need to find are all types of hotels, restaurants, Fast food places, breakfast places and entertainment venues where food and drinks are served. FourSquare will be used to obtain this data (source C).

In order to process the information, the collected data will be stored in a dataframe with the following layout :

Variable	Type
Index	int64
City	String
Latitude	Float
Longitude	Float
Venue Name	String
Venue Latitude	Float
Venue Longitude	Float
Venue Category	String

### 3. Methodology

#### K-means clustering approach

The target of this project is to find the cities in the Flemish region of Belgium with a high concentration of specific hospitality facilities in a range of 2 km from the city centre. For this reason, the K-means clustering machine learning methodology is used to group the cities with a similar structure in terms of hospitality concentration. With this information we can then select the cluster(s) that are the best target for the startup company to deploy their services.

#### Data collection

First step is to collect the geo-data for the cities in Belgium from **source A**. The required data (Postcode, Name, Latitude, Longitude) is stored in dataframe '**PostCodeList**'. As we need to limit our research to the top 100 largest cities in the Flemish region of Belgium, we need access to data about the number of residents for each of the cities. This information is captured by scraping the HTML page from **source B** and is then merged with the geo-data in '**PostCodeList**' and stored in a new dataframe, called '**SelectedCities**'. The last data to collect is the venue information for selected hospitality facilities by using the FourSquare API (**source C**), based on the collected geo-data of the cities. A search query was performed for the 3 venue category groups we are interested in : Food (id : 4d4b7105d754a06374d81259), Bar (id: 4bf58dd8d48988d116941735) and Hotel (id : 4bf58dd8d48988d1fa931735).

#### Data Preparation

In order to make the collected data usable for the analysis, the following data correction was required.

From both the collected data from **source A** and **source B**, the information for cities that do not belong to the Flemish region are removed (cities with a postcode between 1000 and 1999 or between 4000 and 7999 are dropped).

From **source A** and **source B** only the top 100 cities with the largest population are selected.

The collected 'Food', 'Bar' and 'Hotel' venue data from FourSquare is merged into one dataframe '**AllVenues**'.

All collected venues are inspected to secure that they are relevant for our research. Additionally, incorrect data – e.g. wrongly assigned venue types to FourSquare categories – is identified. Not relevant venues and erroneous data is removed from the "**AllVenues**" dataframe.

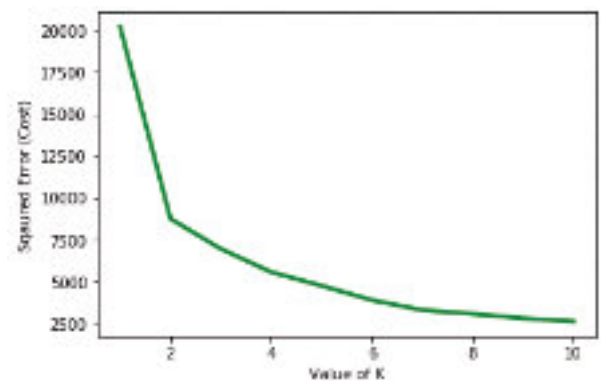
#### Feature engineering

As we want to use the K-means clustering methodology to get insights on the concentration of certain hospitality venues within a range of 2 km from the city centre, we need to further manipulate the data to incorporate the specific requirements from the problem to be solved. First of all we need to categorise the various FourSquare venue types into the 5 categories we defined in the problem description : Restaurants, Fast food places, Breakfast places, Hotels and Cafe/Pub. This will secure that the K-means clustering model

will not be influenced by venue details that are not relevant to our research (e.g. different types of restaurants). Second, we need to perform clustering based on the total number of venues in the 2 km range for each of our 5 venue categories. This is important because we are interested in clustering based on the density of the venues. The last manipulation required is to incorporate the different value of each of the venue categories in the clustering approach. As indicated in the problem description, not all venues are of equal value. For example, a restaurant will create 1.3 times more revenue for the start-up company than a Fast food joint. To incorporate this information in our clustering approach, we multiply the total number of venues in the 2 km range from the city centre with the multipliers provided in the problem description.

#### Modelling

To get the best result with K-means clustering, the Squared Error is calculated for K-values from 1 to 10. Based on the results, it was decided to perform K-means clustering for 2 clusters (K=2).



### 4. Results

The analysis resulted in 2 very distinct clusters. The first cluster (indicated in red in figure 1 available on the next page) contains 11 of the total analysed cities and is characterised by a high concentration of hotels - the most valuable venue type – and a good concentration of restaurants and Fast food places.

As expected, some of the largest cities (Antwerpen, Gent, Leuven, Mechelen, Hasselt, Kortrijk) are part of this cluster. However, 2 of the largest cities (Brugge, Aalst) are not part. After inspecting the data, we can conclude that it makes sense. Although Brugge is one of the most touristic cities in Belgium, it has very few of our targeted hospitality venues in the city centre. The reason for this is that a large part of the city centre is an historical protected area containing many medieval buildings. Aalst - on the other hand - has limited touristic appeal compared to the other larger cities and therefore less hotels and restaurants.

The other 5 cities in the first clusters are relatively small cities. The reason why they made it to the first cluster is because they are very touristic places with many of our targeted hospitality venues, focused on delivering their services to tourists. Four of them - Oostende, Knokke-Heist, Blankenberge and Middelkerke

– are popular seaside town. Ieper – on the other hand – is an important place where many foreign soldiers from world war II are buried. Ieper therefore has many hospitality facilities to cater for the many – mainly British – visitors.

The second cluster is significantly larger than the first cluster. Beside Brugge and Aalst – which are discussed above - this cluster contain the typical mid-size and limited touristic cities of the Flanders region of Belgium. For most of them, the largest concentration of hospitality venues in the city centre are Fast food places. In Flanders we love our Friteries!

5. Discussion

The start-up company should focus their go-to-market strategy on the 11 cities from the first cluster (see table below) for the reasons mentioned in the results section. Highest priority should be given to the cities of Antwerpen, Gent, Leuven, Mechelen, Hasselt and Kortrijk as they are not only touristic places, but they also have a large population. For the seaside towns of Knokke-Heist, Blankenberge and Middelkerke, it should be considered that these have relatively few permanent residents and occupation of the hospitality venues is highly seasonal and focused on the summer months. This makes them a little bit less interesting as priority locations for the go-to-market strategy.

Table 1 : Cities in the first cluster

City	1st most applicable venue type	2nd most applicable venue type	3rd most applicable venue type	4th most applicable venue type	5th most applicable venue type
Antwerpen	Hotel	Fast food	Restaurant	Pub	Breakfast
Gent	Hotel	Restaurant	Fast food	Breakfast	Pub
Leuven	Hotel	Fast food	Restaurant	Breakfast	Pub
Mechelen	Hotel	Restaurant	Fast food	Breakfast	Pub
Hasselt	Fast food	Restaurant	Hotel	Breakfast	Pub
Kortrijk	Hotel	Restaurant	Fast food	Pub	Breakfast
Oostende	Hotel	Fast food	Restaurant	Pub	Breakfast
Ieper	Hotel	Restaurant	Fast food	Breakfast	Pub
Knokke-Heist	Hotel	Fast food	Restaurant	Pub	Breakfast
Blankenbrge	Hotel	Restaurant	Fast food	Pub	Breakfast
Middelkerke	Hotel	Restaurant	Fast food	Pub	Breakfast

6. Conclusion

As the results of this analysis shows, using an unsupervised machine learning approach can provide valuable information for a company to support them in making educated business decisions. It provides insights that are not straightforward and - most likely - will not be detected with traditional analysis.

This report also shows that it makes sense to conduct an even deeper analysis of the available data. Brugge - the most touristic city of Flanders - was not indicated as a potential candidate to open a fresh produce delivery service. Maybe there is a potential business case for some of the suburbs that are close to the historical city centre. Why was Aalst - one of the larger cities - not a candidate? What are the characteristics of this city that result in a less developed hospitality concentration in the city centre? Diving into these details will secure an even better go-to-market strategy with all potential risks analysed and controlled.

Figure 1 : Mapping of the 2 clusters on the map of Flanders

