Applied Data Science Capstone

Fresh produce delivery
to the hospitality sector in city centres

Mark Jansen

Introduction

A Belgian startup company want to implement a
non-motorized supply-chain concept for delivering fresh produce
to the hospitality sector located in city centres.

Opportunity : Anticipate the growing measure of cities to reduce motorized traffic

Value Proposition

- Bulk purchase of fresh produce
- Pay for what you use (reduce waste)
- On-demand delivery within 30 minutes
- Non-motorized delivery

Problem to be Solved

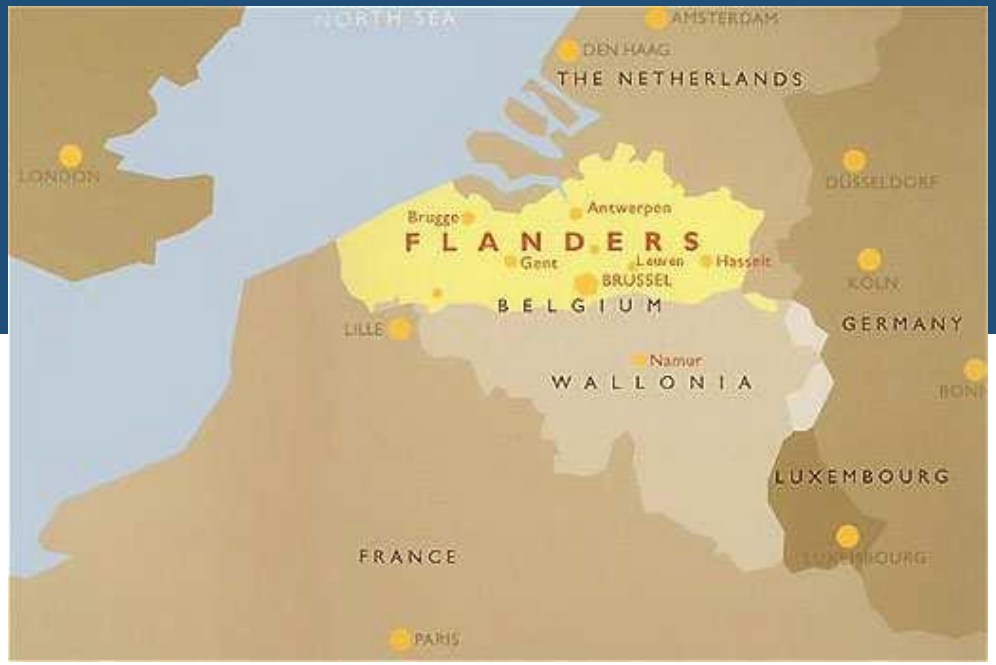
Problem : Secure a successful go-to-market strategy

Required : Identification of the cities in the Flanders region of Belgium
that have the most business potential

Criteria :

- Number of potential customers that can be reached with non-motorized vehicles in less than 30 minutes from the city centre
- Value for the business of the potential customers

Map of Flanders, Belgium



Required Data

List of Cities in Flanders (max 100)

Latitude/Longitude of the City Centre
source : http://www.gemeentezoeker.be/MonkeyProof_postcodes_developer.csv

Number of residents
(required for selection of top100 largest cities)
source : *https://www.metatopos.eu/belgcombiN.html*

Find relevant hospitality venues in a range of 2 km from the city centre

Venue information
source : FourSquare

Search Foursquare to find potential customers in the applicable 5 categories

- Hotels
- Restaurants
- Fast food places
- Breakfast places
- Café / Pub

DataFrame setup for Analysis

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

Methodology – K-Means Clustering

K-Means Clustering

- To solve our problem, we need to explore the city centres, segment them, and group them into clusters to find similarities across cities in Flanders
- Clustering should be based on the concentration of targeted hospitality venues + the value they provide for the business (e.g. a hotel provides more potential revenue than a pub, which will only require fresh produce for cocktails)
- To do so, we will use an unsupervised machine learning technique for clustering (K-Means)

Data Collection and Preparation

List of Cities in Flanders (max 100)

SOURCE A
http://www.gemeentezoeker.be/MonkeyProof_postcodes_developer.csv

SOURCE B
https://www.metatopos.eu/belgcombiN.html

Venue information
source : FourSquare

1. Geo location data for the city centres is obtained from source A

   All cities that are not part of the Flemish region are removed (post codes from the Brussels region (1000-1999) and the Walloon region (4000-7999)

2. Number of residents per city is scraped from the HTML-page from source B

   All cities that are not part of the Flemish region are removed, the list is sorted and the top 100 largest cities are selected

3. FourSquare is used to search for relevant venues in a 2km radius for the Top 100 cities

   A search query was performed for the 3 FourSquare venue category groups we are interested in : Food (id : 4d4b7105d754a06374d81259), Bar (id: 4bf58dd8d48988d11 6941735) and Hotel (id : 4bf58dd8d48988d1fa931735).

   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.

Feature Engineering

Categorising the identified venues into the required 5 categories

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).

Clustering based on total number of venues

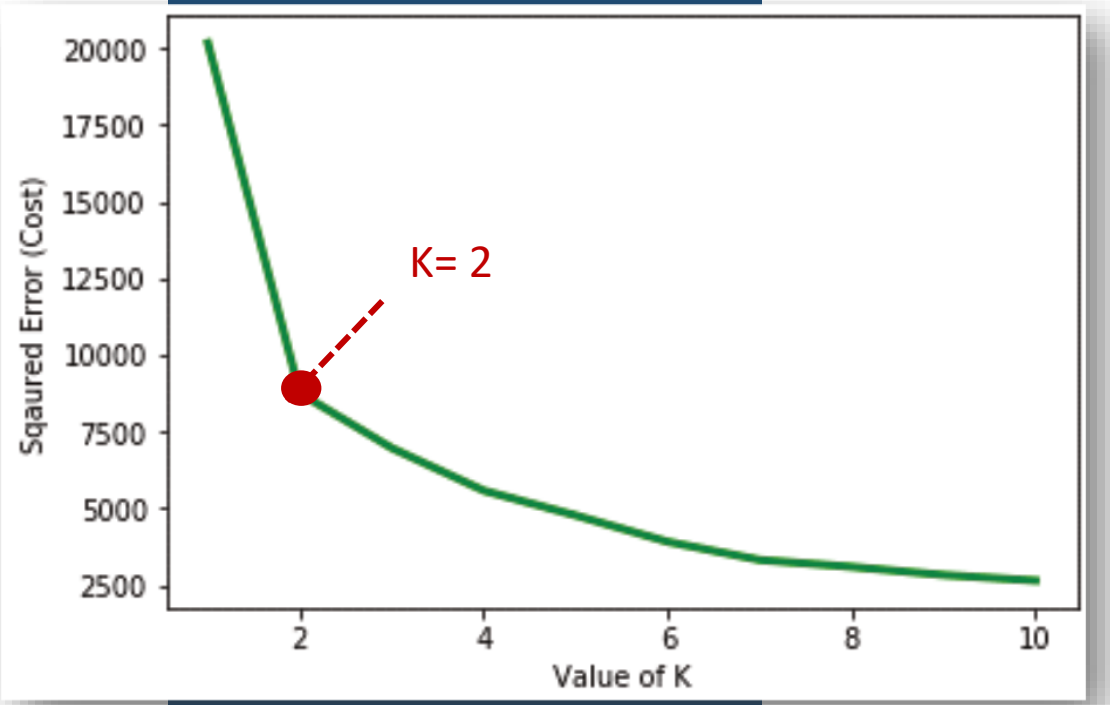
Second, we need to perform clustering based on the total number of venues in the 2 km range for each of our 5 venue categories. This is important because we are interested in clustering based on the density of the venues.

Considering the business value of each venue category

The last manipulation required is to incorporate the different value of each of the venue categories in the clustering approach. 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 indicating the potential business value.

Modelling

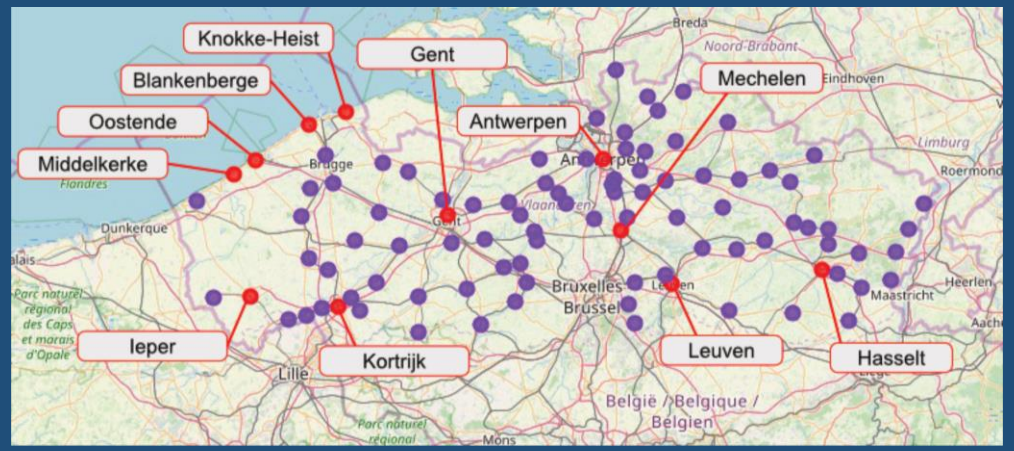
Squared Error calculation for K= 1 to 10



- K-means clustering was performed for different K-values 1 to 10
- The Square Error clearly shows that limited gain can be achieved from selecting more than 2 clusters

Results

Results from K-means clustering



CLUSTER 1

High concentration of hotels - the most valuable venue type – and a good concentration of restaurants and fast food places

CLUSTER 2

This cluster contain the typical mid-size and limited touristic cities. For most of them, the largest concentration of hospitality venues in the city centre are fast food places.

Observations

- As expected, some of the largest cities are part of cluster 1. However, 2 of the largest cities are not part. 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 therefor 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 therefor has many hospitality facilities to cater for the many – mainly British – visitors

Discussion

- The start-up company should focus their go-to-market strategy on the 11 cities from cluster 1
- 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 then a little bit less interesting as priority locations for the go-to-market strategy

Prioritising the cities to focus on

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

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.