A New Food Warehouse in Leuven (Business Recommendation)

1. Introduction: Business Problem

Leuven is the capital of the province of Flemish Brabant in Belgium. It is located about 25 kilometers east of Brussels. The municipality itself comprises the historic city and the former neighboring municipalities of Heverlee, Kessel-Lo, Wilsele and Wijgmaal. It is the eighth largest city in Belgium and the fourth in Flanders with more than 100,244 inhabitants (Federal Ministry of Home Affairs, 1/11/2016).

Leuven is home to the KU Leuven, the largest and oldest university of the Low Countries and the oldest Catholic university still in existence. The related university hospital of UZ Leuven is one of the largest hospitals in Europe. The city is also known for being the headquarters of Anheuser-Busch InBev, the world's largest brewer and one of the five largest consumer-goods companies in the world.[1]

With the knowledge of issues mentioned above over Leuven, an investor in the food sector would like to build up a new supermarket/food warehouse in Leuven (Belgium). The investor wants to deliver fresh and high-quality food products to places as Restaurants, Fast Food Stores, Bakeries, Breakfast Venues, Breweries and Cafés. The investor aims to build a food warehouse for the groceries and products bought from local villagers and farmers, in this sense he also would like to get local acceptance and support of people that will also provide more and satisfied customers. Customer satisfaction and Service quality are always the main stones in the philosophy of the company.

To get a prestigious name in the sector, besides more profit in terms of money and time, he targets with a quick and effective response to the customers. So, the location of the food warehouse has utmost importance that will provide smooth and timely delivery of products from local farmers and deliver to the related customers (venues) in the city.

The investor would also like to minimize transportation cost by building the warehouse in a close and optimized location to its customers. There are several (5) neighborhoods (Leuven Center, Heverlee, Kessel-Lo, Wilsele and Wijgmaal) in city Leuven that our investor set the focus. Determining the right neighborhood is our goal, and we will use a Machine Learning algorithm to cluster the neighborhoods and make a recommendation to the investor as a solution to this problem.

2. Data

a. Data We Need

The investor would like to invest in a food warehouse in Leuven but which neighborhood is not determined. The Postal Codes of Leuven (3000,3001,3010,3012,3018) that provides us with the necessary information to find which neighborhoods (3000 Leuven Center (Old City), 3001 Heverlee, 3010 Kessel-Lo, 3012 Wilsele and 3018 Wijgmaal) are our candidates for the location of the warehouse.

In this sense, we can get the data including postal codes and names of neighborhoods in Leuven from open source internet web pages. For this case, I got the information from the official site of Flemish Government (vlaanderen.be). [2]

We needed the geo-locational data of this specific borough (Leuven city) and the five neighborhoods for our analysis. This information can be gathered in several ways, such as using Geocoder package and by transferring pre-prepared open data source.[3] Finally, I would get the below sample data for each neighborhood:

```
PostalCode/ Borough/ Neighborhood/ Latitude/ Longitude/

3001/ Leuven/ Heverlee/ 50.851729/ 4.693131/
```

Further on, we would need data about different venues in the various neighborhoods of city Leuven. "Foursquare" application can provide us with the locational information of the sites that we needed. By locational information of the center of each neighborhood and by inputting a self-determined distance from city center we can get detailed information about the venues that lays inside the circular interest area we have described. For each venue, we can gather information including its specific location information (latitude, longitude), category and popularity of the venue.

After gathering and merging necessary information, we can do clustering to see the similarities and differences of the neighborhoods. Then we can sum up the number of food product customers per neighborhood to decide which one is potentially a better choice for investment.

Foursquare will provide us data such as the following information:

```
Neighborhood/ Neighborhood Latitude/ Neighborhood Longitude/ Venue/ Venue Latitude/ Venue Longitude/ Venue Distance/ Venue Summary/
Venue Category/

Wijgmaal 50.926428/ 4.700121/ Halte Waterstraat/ 50.922755/ 4.691789/ 713/ This spot is popular/ Bus Stop/
```

b. Data Gathering

To get the necessary data from several sources, I used different methods and libraries. I scrape the postal codes and the names of the neighborhoods from the web sites by using Beautiful Soup and Requests and transfer into data frames by using Pandas.

After building a data frame of the postal code of each neighborhood along with the borough name and neighborhood name, we should get the coordinates of each neighborhood to be able to utilize Foursquare API. I had several alternatives here.

(1) Google Maps Geocoding API (not free): In an older version of this course, we were leveraging the Google Maps Geocoding API to get the latitude and the longitude coordinates of each neighborhood. However, recently Google started charging for their API.[4] I skipped this option.

(2) Geocoder Python package[5]: The problem with this Package is that you have to be persistent sometimes to get the geographical coordinates of a given postal code.

So, you can make a call to get the latitude and longitude coordinates of a given postal code, and the result would be None, and then make the call again, and you would get the coordinates. So, to make sure that you get the coordinates for all of our neighborhoods, you can run a while loop for each postal code.

(3) Open Data source as Json/csv: Given the fact that Geocoder package can be very unreliable, I tried the third alternative. I searched and found an open source json and a csv file that has the geographical coordinates of each postal code in Belgium[6]. I set the necessary parts out of this data into another data frame. After some preprocessing my Leuven Data Frame was ready for further steps.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	3000	Leuven	Leuven	50.881253	4.692990
1	3001	Leuven	Heverlee	50.851729	4.693131
2	3010	Leuven	Kessel-Lo	50.889915	4.730761
3	3012	Leuven	Wilsele	50.909536	4.713629
4	3018	Leuven	Wijgmaal	50.926428	4.700121

Figure 1- Leuven Neighborhoods Location Data

I continued data gathering here to explore, analysis and cluster the neighborhoods in Leuven. Firstly, I created a map of Leuven with neighborhoods superimposed on top by using Folium.

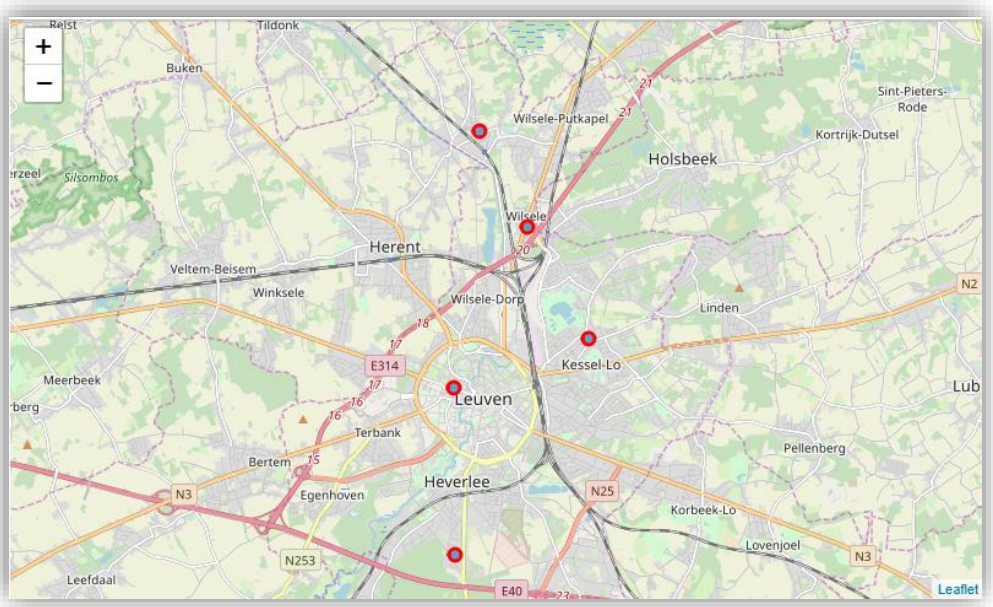

Figure 2-Neighborhoods in Leuven

In this step, I would like to show the map of Leuven to get an overall idea how could we get venues data from Foursquare and how could I determine optimum values of the radius values from the center points of neighborhoods. With the information from Leuven Data Frame and the intuition from the overview map of Leuven, I was ready to utilize Foursquare API to get venues data.

After connecting to the Foursquare API to gather information about venues in each neighborhood, I applied the radius value as 1500 meter. It means that we have asked Foursquare to find venues that are at most 1500 meter far from the center of the neighborhood. And I saved the information into a data frame. I got as a result comprising 294 venues in the data set. I had 110 unique venues in the data.

```
# Let's check the size of the resulting dataframe
print(leuven_venues.shape)
leuven_venues.tail()
```

```
(294, 9)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Distance	Venue Summary	Venue Category
289	Wijgmaal	50.926428	4.700121	Sportlokaal	50.924938	4.686548	966	This spot is popular	Bar
290	Wijgmaal	50.926428	4.700121	Lijnloperspad	50.917226	4.706344	1113	This spot is popular	Bike Trail
291	Wijgmaal	50.926428	4.700121	Apotheek Haegemans	50.917074	4.685834	1445	This spot is popular	Pharmacy
292	Wijgmaal	50.926428	4.700121	Apotheek Adriaens	50.922932	4.720527	1483	This spot is popular	Pharmacy
293	Wijgmaal	50.926428	4.700121	Nachtwinkel Euro	50.924335	4.721062	1487	This spot is popular	Convenience Store

Figure 3- Venues per Neighborhoods Data Frame

3. Methodology

a. General Overview:

In this part, I explored the data with using visualization and basic data analysis approach to understand which parts of the data is more meaningful and try to understand which variables can help to clarify the differences between candidate neighborhoods for the investment. Descriptive statistics and visualization are followed by machine learning.

With the use of gathered data, I applied K-means Clustering Machine learning algorithm to group the neighborhoods and get some insight which neighborhoods that are more suitable for our food warehouse, in terms of minimizing the delivery distances and costs while improving the revenues on the other side.

b. Explore and Analyze Data Set/ Neighborhoods in Leuven:

I will in this part give some summary information over the neighborhoods that I gathered, try to explore and later to analyze data to get insight over the solution of our business problem. Since we would like to build a new food warehouse, then the number of venues and especially food consuming or selling venues in each neighborhood and their total numbers would be valuable.

Firstly, I plotted the venues in the Leuven Neighborhoods and to get some visual insight and tried to see the overall distribution of the venues.

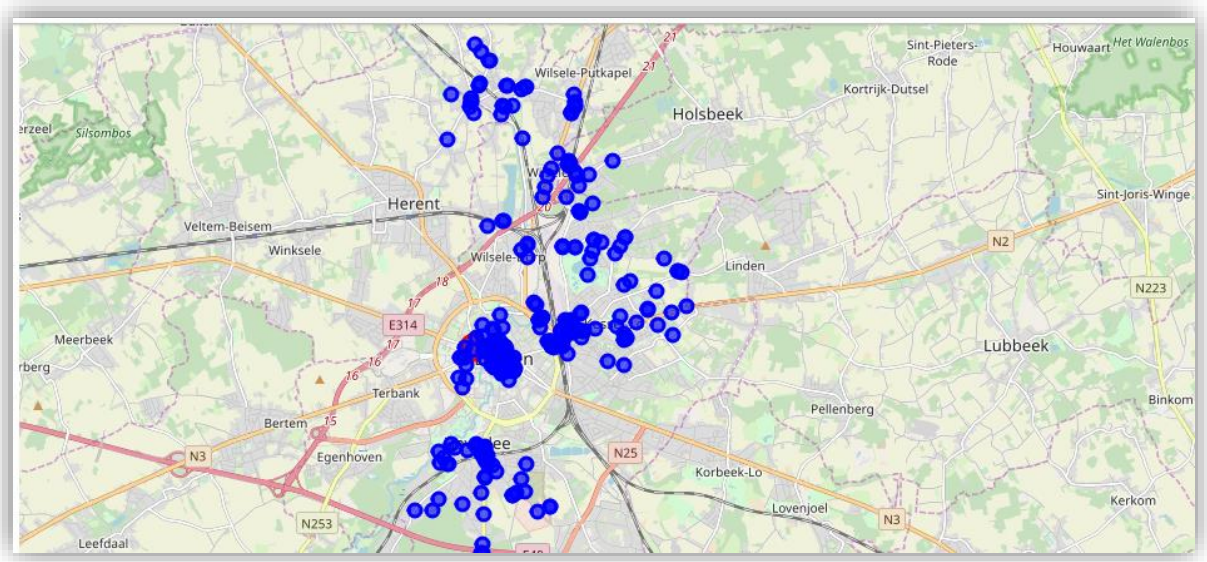

Figure 4-Venues in the Neighborhoods

The other proper remedy was to group the venues into different clusters. Each cluster would be then represented by the number of venues in each neighborhood. These clusters can be thought of as pockets of Leuven which we can then analyze separately. To implement this, I started by instantiating a *MarkerCluster* object and adding all the data points in the data frame to this object.

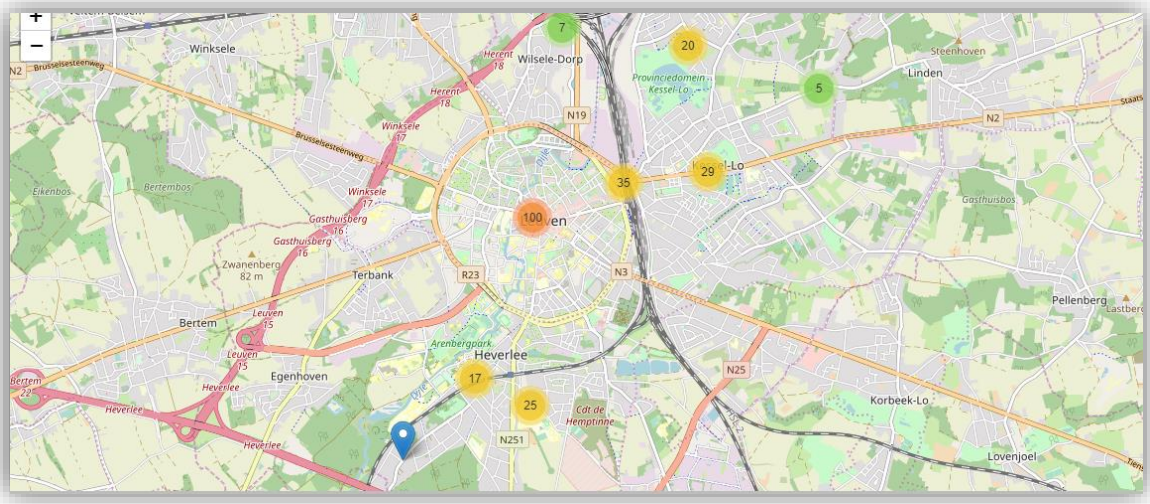

Figure 5- Grouped Venues in Leuven

As a result I got the map in Figure 5 which is an interactive one. When I zoom out all the way; all markers are grouped into one cluster, *the global cluster*, of 100 markers or venues which is the total number of venues in our data frame. Once you start zooming in, the *global cluster* will begin breaking up into smaller clusters. Zooming in all the way will result in individual markers. In this manner, we can see where the venues are located and geographically grouped as a simple inside out, we can see that the old city center is the most popular place by 100 venues.

I would like to see differences between neighborhoods and the distribution of venues over the neighborhoods I prepared the following pie and bar charts.

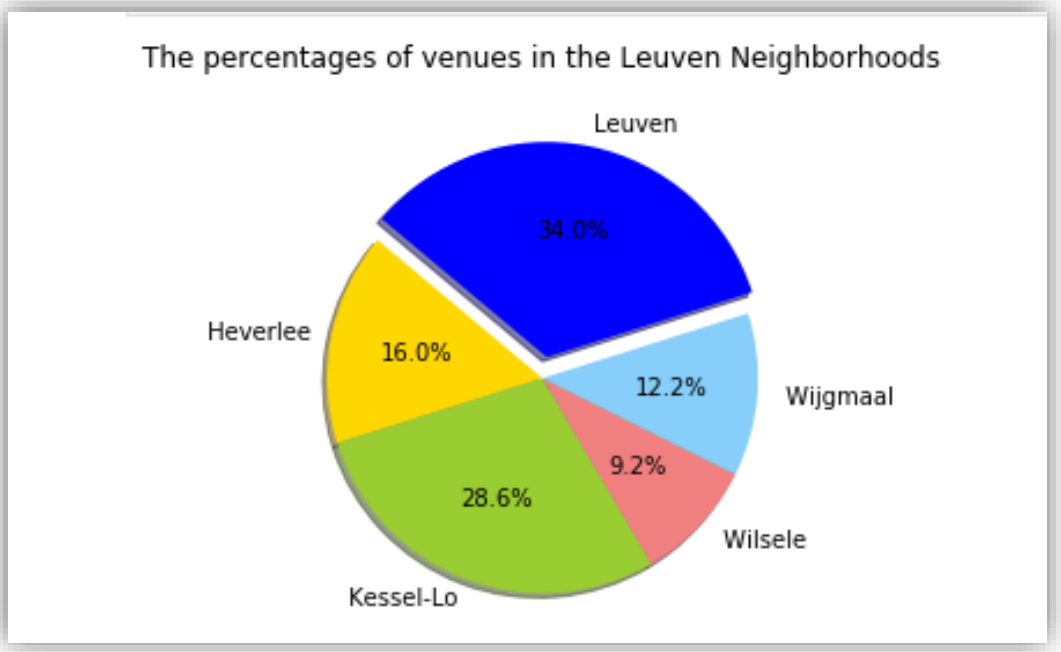

Figure 6-Venues Distribution in the Neighborhoods

By using charts, that would be easier to see that 34 procent of venues were located in Leuven City Center, and this corresponds to 100 venues in the old city center.

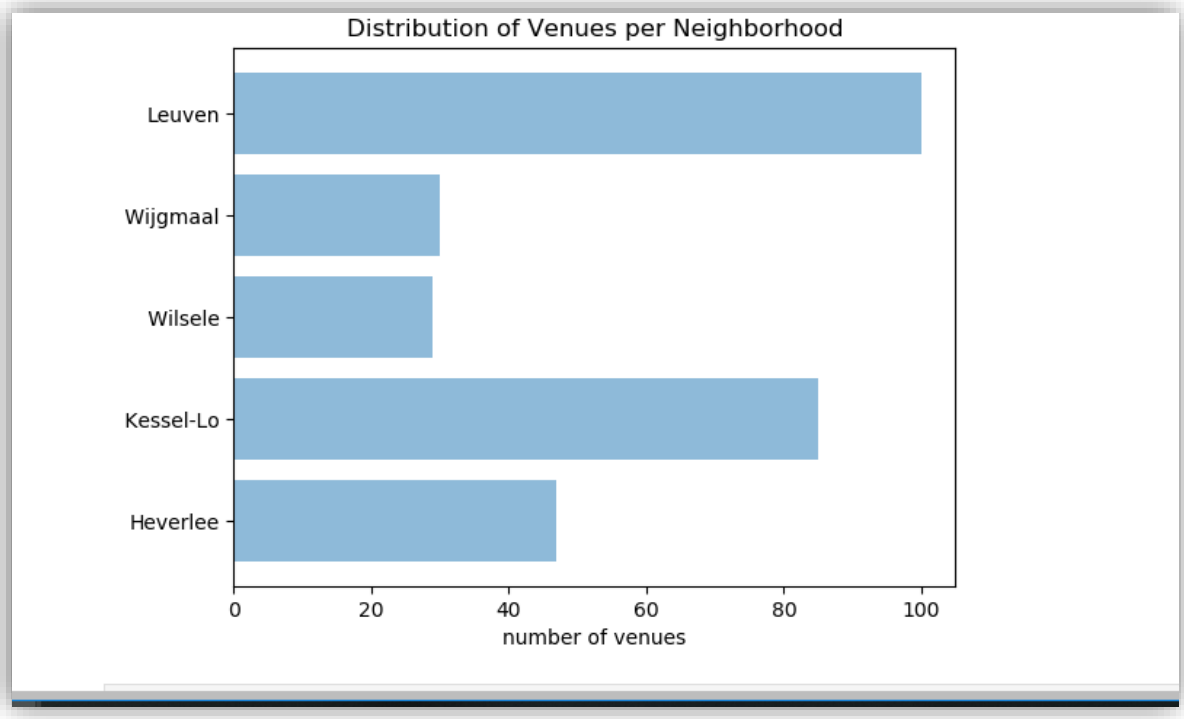

Figure 7-Venues per Neighborhood Bar Chart

I took a deeper look into the descriptive analytics and the following statistics over the distance values of venues to the center of each neighborhood. It is shown out that the minimum average distance value is in Leuven City Center and, here the venues are more centrally distributed. Our histogram of the distances of venues in Leuven City Center exposes a more intuitive representation of this result. 100 venues in the old city center have been spread as depicted in Figure 9. Most of the venues are clustered in a range of 400-600 meters from the central point of Leuven City Center.

```
count     294.000000
mean      863.448980
std       394.473723
min        68.000000
25%       531.000000
50%       805.000000
75%      1230.500000
max      1501.000000
Name: Venue Distance, dtype: float64
```

```
Neighborhood
 Heverlee     1077.148936
 Kessel-Lo    1137.619048
 Wijgmaal      743.259259
 Wilsele       902.444444
Leuven         551.120000
Name: Venue Distance, dtype: float64
```

Figure 8- Venue Distances to the Neighborhood Centre

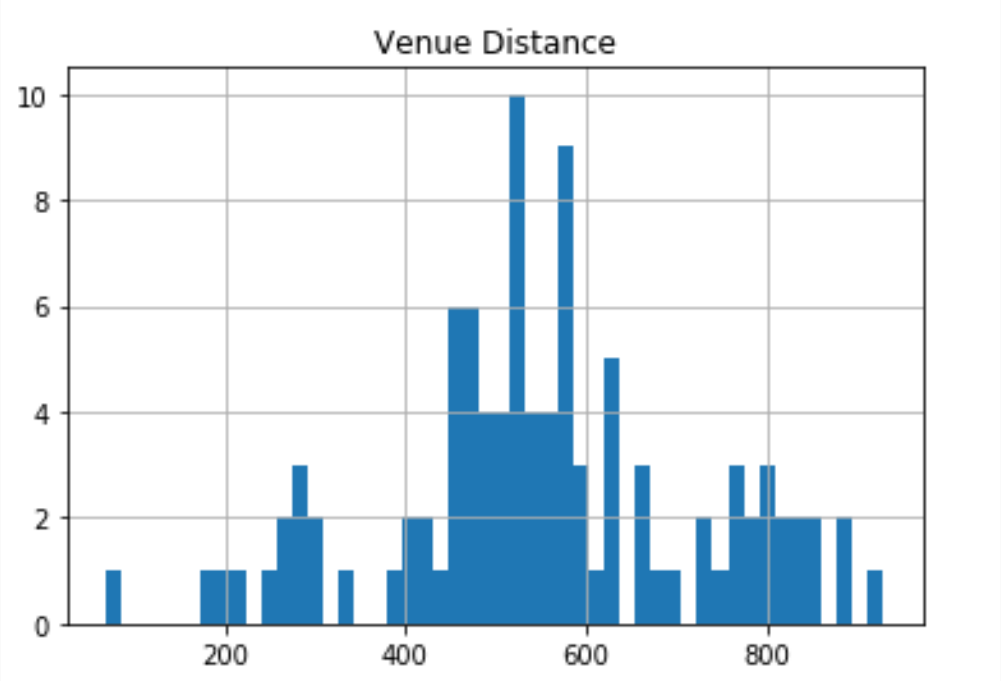

Figure 9-Leuven City Center Histogram

Moreover, I cleaned the unrelated venues out of my venue list by cleansing the non-food related venue categories out. Then I had the following data frame as a summary list of food-related venues per neighborhood which are also our future potential customers.

	Neighborhood	Total Food Customers
0	Heverlee	24
1	Kessel-Lo	57
2	Wijgmaal	11
3	Wilsele	15
4	Leuven	81

Figure 10-Number of Food Venues

I had 188 food product consumer venues out of 294 as our potential customers. 81 of them are based in central Neighborhood. The following table shows us the first five most common venues per neighborhood in our scope.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Heverlee	Bar	Bakery	Brasserie	Belgian Restaurant	Chinese Restaurant
1	Kessel-Lo	Bar	Bakery	Friterie	Hotel	Coffee Shop
2	Wijgmaal	Bar	Friterie	Supermarket	Sandwich Place	Bakery
3	Wilsele	Supermarket	Fruit & Vegetable Store	Liquor Store	Sandwich Place	Friterie
4	Leuven	Bar	Coffee Shop	Italian Restaurant	Restaurant	Cocktail Bar

Figure 11-Most common venues per Neighborhood

```
---- Heverlee----
                venue  freq
0                 Bar   3.0
1          Brasserie   2.0
2             Bakery   2.0
3  Belgian Restaurant   2.0
4  Chinese Restaurant   2.0


---- Kessel-Lo----
         venue  freq
0          Bar  10.0
1       Bakery   7.0
2     Friterie   6.0
3        Hotel   4.0
4  Coffee Shop   3.0


---- Wijgmaal----
                venue  freq
0                 Bar   3.0
1          Restaurant   1.0
2            Friterie   1.0
3              Bakery   1.0
4  Chinese Restaurant   1.0
```

```
---- Wilsele----
                     venue  freq
0              Supermarket   4.0
1  Fruit & Vegetable Store   1.0
2        Indian Restaurant   1.0
3             Gourmet Shop   1.0
4           Sandwich Place   1.0


----Leuven----
                venue  freq
0                 Bar  14.0
1         Coffee Shop  10.0
2  Italian Restaurant   7.0
3            Friterie   3.0
4          Restaurant   3.0
```

Figure 12-Frequencies of Most Common Venues

c. Machine Learning/ Cluster Neighborhoods

In this part of our project, I would apply K-means clustering algorithm, and in this sense, I practiced Elbow method to determine which k (number of clusters) has been better for an optimum solution.

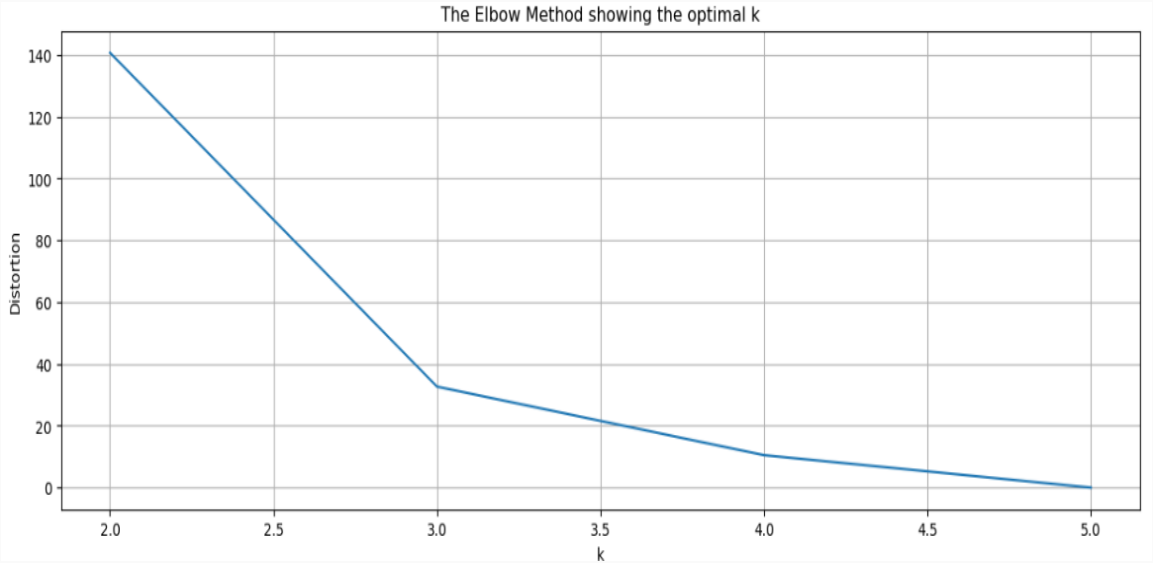

Figure 13-Elbow Method

When k increases, the centroids are closer to the cluster centroids. Here the distortion, mean sum of squared distances to centers, decreases to the optimum point. The improvements will decline, at some point rapidly, creating the elbow shape. That point is the optimal value for k in the image above, k=3.

Within the light of this, I applied a machine learning algorithm with k=3(number of clusters) to get the clustered neighborhoods. In this part, I grouped the clusters to get the total food-related venues in each cluster. Naturally, the cluster with the highest number of potential food product customers would be our first choice. Then we c try to get more insight with additional graphics and maps. The following is the merged table with clusters, neighborhoods and most common venues.

PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3000	Leuven	Leuven	50.881253	4.692990	0	Bar	Coffee Shop	Italian Restaurant	Restaurant	Cocktail Bar
3001	Leuven	Heverlee	50.851729	4.693131	1	Bar	Chinese Restaurant	Bakery	Brasserie	Belgian Restaurant
3010	Leuven	Kessel-Lo	50.889915	4.730761	2	Bar	Bakery	Friterie	Hotel	Coffee Shop
3012	Leuven	Wilsele	50.909536	4.713629	1	Supermarket	Pizza Place	Sandwich Place	Club House	Gourmet Shop
3018	Leuven	Wijgmaal	50.926428	4.700121	1	Bar	Chinese Restaurant	Sandwich Place	Convenience Store	Friterie

Figure 14-Clustered Neighborhoods and Common Venues

4. Results

When we examine the results in Figure above, we can see the neighborhood grouping in the clusters as follows:

- Cluster 0: “Leuven(City Center)”
- Cluster 1: “Heverlee, Wilsele, Wijgmaal”
- Cluster 2: “Kessel-Lo”

We can also examine what is the frequency of food venues in different clusters as follows.

```
Cluster Labels
0    81.000000
2    57.000000
1    16.666667
Name: Total Customer Sum, dtype: float64
```

Figure 15- Venues per Cluster

We can see the clustered neighborhoods in a map for better insight.

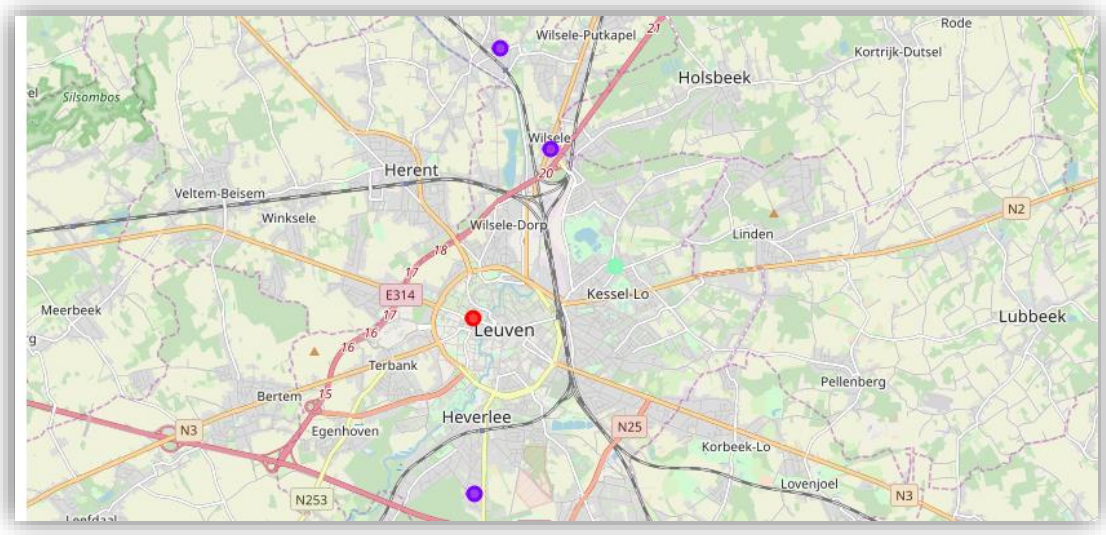

Figure 16-Clustered Neighborhoods

Pie Chart says that the most favourite place of food venues is Cluster 0, old city center.

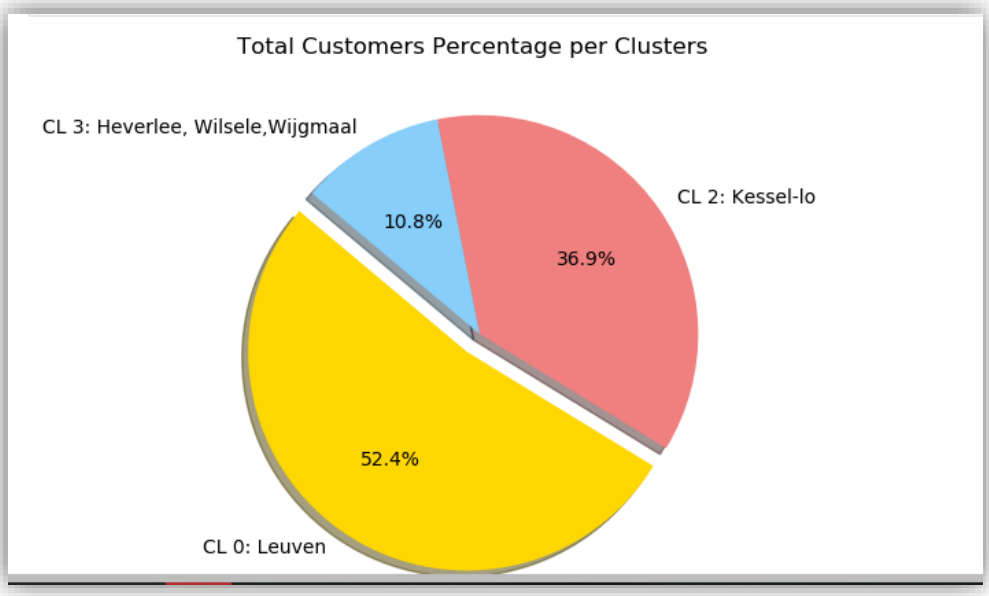

Figure 17-Food-Related Venues per Cluster

After examining each cluster one by one, we can verify that the most popular cluster is the same place in our first exploration in the data which has been Leuven Center Neighborhood. So, I can conclude that cluster number 0 has 81 in total customer potential. Leuven City centrum is the best cluster (and here is the best neighborhood for our investment). Following Kessel-Lo is the second-best alternative.

As a more specific step, while determining the location of warehouse, we may consider the fact that most of the venues are clustered in a range of 400-600 meters (as previously mentioned in this report) from the central point of Leuven City Center.

Examine Clusters

Best Neighborhood to build a Food warehouse : Cluster 0: Leuven City Center

```
leuven_merged.loc[leuven_merged['Cluster Labels'] == 0, leuven_merged.columns[[2] + list(range(5, leuven_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Leuven	0	Bar	Coffee Shop	Italian Restaurant	Restaurant	Cocktail Bar	Friterie	Vegetarian / Vegan Restaurant	Gourmet Shop	Bakery	Sushi Restaurant

Second Best Neighborhood: ## Cluster 2: Kessel-Lo

```
leuven_merged.loc[leuven_merged['Cluster Labels'] == 2, leuven_merged.columns[[2] + list(range(5, leuven_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Kessel-Lo	2	Bar	Bakery	Friterie	Hotel	Sandwich Place	Coffee Shop	French Restaurant	Pizza Place	Brasserie	Brewery

Third alternatif is among Cluster 1: Heverlee, Wilsele, Wijgmaal

```
leuven_merged.loc[leuven_merged['Cluster Labels'] == 1, leuven_merged.columns[[2] + list(range(5, leuven_merged.shape[1]))]]
```

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Heverlee	1	Bar	Restaurant	Bakery	Chinese Restaurant	Brasserie	Belgian Restaurant	Sandwich Place	Farmers Market	Sushi Restaurant	Gastropub
3	Wilsele	1	Supermarket	Fruit & Vegetable Store	Brewery	Friterie	Pizza Place	Indian Restaurant	Club House	Gourmet Shop	Sandwich Place	Liquor Store
4	Wijgmaal	1	Sandwich Place	Market	Supermarket	Bar	Bakery	Restaurant	Kebab Restaurant	Sports Bar	Pizza Place	Chinese Restaurant

Figure 18-Final Examination of Clusters

5. Discussion

As Leuven is a nice and comparatively a small city with five neighborhoods, we get relatively small data set it may be thought that it is with intuition also easy to choose where to invest. But the important thing is the methodology and scientific approach we show here that can be re-performed and re-used for similar problems in other cities or countries.

Although our clustering algorithm recommends us Leuven City Center (cluster 0 with 81 food consumer venues) as the first alternative to build a new food warehouse, there may be additional factors to be taken into attention in this decision process. These can be such as the limitations of laws in the region (old city center), comparison of the rental prices, incomes, populations are specifically not included in this project but it is an open improvement area for future engagements.

I used the k-means clustering algorithm as part of this solution process. As a result of the Elbow method, I used the optimum k value as 3. Although we used this algorithm for only five neighborhood and it can also easily be used for larger cities as Brussels or Antwerp (or any other city in the world).

I also performed data analysis and visual exploration to see the differences between neighborhoods and try to find in which areas food-related venues are summoned.

I ended the study by with 3 clusters that provide three alternatives for the investment. The investors will give the final decision by adding his market experience into statistical results.

6. Conclusion

In conclusion, we recommend as the first alternative Leuven City Center neighborhood (post code: 3000) which is very touristic, a high student population in the greater part of the year. Second alternative neighborhood for investment is Kessel-lo. The third alternative is the worst one in accordance with our results, but we should also consider that the city is not so large in size, and the distance of neighborhoods to each other is relatively small. In this sense, the profit we will get by choosing the first alternative will be higher but may be negligible if there are other important challenging aspects beyond the coverage of this study or in the pocket of higher-management.

The methodology can be easily and smoothly applied to similar problems in much bigger cities and data sets. The research can be supported by analysis of rental prices, population and income data of the neighborhoods.

References:

1 https://en.wikipedia.org/wiki/Leuven
2 https://www.vlaanderen.be/gemeenten-en-provincies/provincie-vlaams-brabant/leuven
3 https://raw.githubusercontent.com/jief/zipcode-belgium/master/zipcode-belgium.csv
4 http://geoawesomeness.com/developers-up-in-arms-over-google-maps-api-insane-price-hike/
5 https://geocoder.readthedocs.io/index.html
6 https://raw.githubusercontent.com/jief/zipcode-belgium/master/zipcode-belgium.json

Levent BINGOL

levobelgium@gmail.com
https://www.linkedin.com/in/levent-bingöl-957160139/
https://github.com/LeventBINGOL