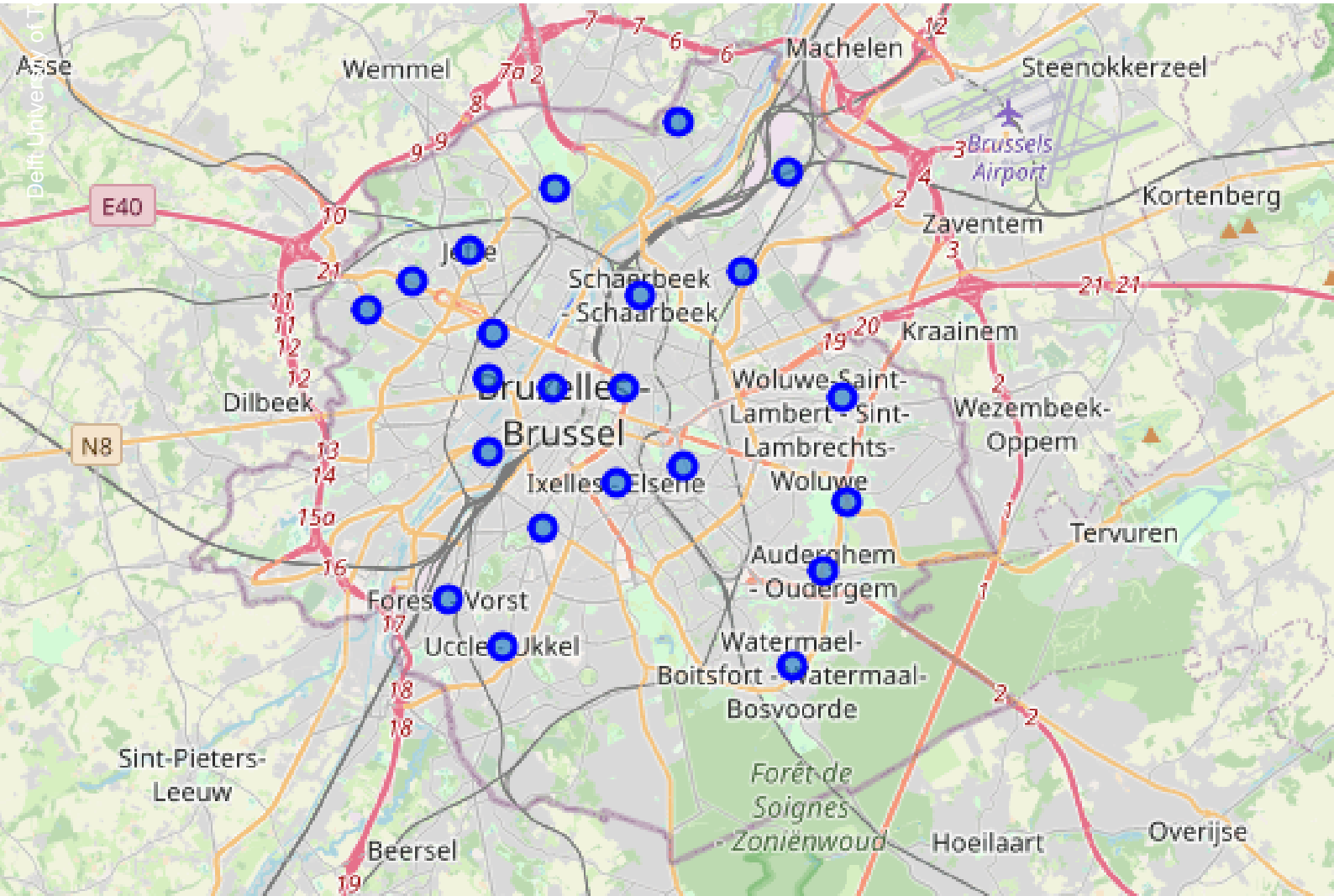
Battle of Neighbourhoods of Brussels Capital Region

Finding the best neighbourhood for relocation for a startup company owner

Nimmy Samson Attipetty
31-03-2021



Executive Summary

Relocating is a tiring process which if not planned carefully will result in series of mistakes and regrets. In this project we deal with choosing the best neighbourhood in Brussels for a partner of startup company who wishes to work in a co-working space. The neighbourhood of interest is to be selected following the criteria suggested by the client. Based on location data provided by Foursquare data, open data provided by Brussels and web scraping for postcodes of Brussels Capital Region, we obtain the venues in the respective communities. The features for the analysis are farmers market, supermarket and French Restaurant. Based on the features the municipalities are clustered using K-means clustering. Finally we find that two neighbourhoods- Etterbeek and Uccle could be the places most desired by the client.

Contents

List of Figures

List of Tables

1 Introduction

1.1. The Problem of Relocation

Relocating to a new place often poses a lot of challenge for a person.Be it as part of a new job or studies or due to one's personal circumstances most of us have moved out to a new place atleast once. Finding the best place that will cater to one's personal needs is difficult. Most of the people search for a place where living conditions are similar that of previous place. These can be achieved only if there are venues (according to the persons needs nearby). Some people prefers a supermarket nearby for a quick grab of groceries and other household stuff, For others it might be the presence of a fitness center or a restaurant. Whichever be the venues, what everyone search is for a perfect place to make the living easy in a new city.

1.2. Desirable place for Pierre in Brussels

Pierre is a partner and owner of a new IT startup company. He is relocating from his native village in France to Brussels. Initially, the company will share its office space in any of the co-working spaces available in Brussels. He and his business partner are unsure about the exact location. Pierre prefers to fix the location of relocation to a place where his rented apartment and certain venues are nearby provided that area also offers a co-working space. Since his native village in France was famous for farm fresh goods, he would like to have a farmers market nearby his house. Presence of French Restaurant is also important as he also enjoys having food from outside. A supermarket is also needed as buying all household items and quick grab of food lest in case he runs late from work is easier. Pierre prefers a place which can offer all three of these venues along with an office space. He would then look out for a rented apartment in the place that offers all four of such venues. He would like to have a list of suitable sites in Brussels.

1.3. The target audience

In this report we mainly deal with analysis of municipalities in Brussels to find the best place where Pierre can relocate to, based on his personal preferences. The analysis based on location data, open data and data science analytic methods can also be extended to others with specific needs such as an exchange/international student arriving in the city or for people who are searching for a new home within Brussels itself.

1.4. The Brussels Capital Region

Known as the Capital of European Union to the World, Brussels is a dynamic city with unique history and architecture. Located at the banks of River Senne, the history of the city could be traced back to the 10th century [7]. Brussels is also the administrative center of Belgium. Though surrounded by the Dutch speaking area of Belgium, the city itself is bilingual in French and Nederlands due to the international influence.

The Capital Region of Brussels is divided into 19 municipalities each with autonomous administrative authority. The largest among the municipalities is known as the City of Bruxelles (Ville de Bruxelles in French) which also includes the European area along with six other areas. Due to the status of the city as the capital of European Union and the political and cultural changes reflected within, Brussels offers lots to explore.

2

Data

2.1. Data needed for analysis

As per the wishes of Pierre the following data should be retrieved for the analysis:

b) The postcodes and the names of municipalities
This information was derived through web-scraping the website [10] the veracity of which was verified from official sources [1, 2, 5].

c) The list of venues in all of the municipalities in Brussels
The postcodes and names of municipalities are used to derived the latitude and longitude of all the municipalities using the OpenCage Geocoder API [6]. The information thus obtained is used as an input to a location provider API called Foursquare API [4] by the Foursquare technology Company. The output is a list of venues present in the location.

a) The number of co-working spaces available in every municipality of Brussels Capital Region
The data of number of co-working spaces is obtained from the open data website of Brussels [8]. The data obtained listed the names of the cow-orking spaces, name of municipality and the corresponding post code from which the number of available co-working spaces in each municipality is derived.

2.2. Data preparation and data wrangling

To make a data worthy to be analysed it has to be cleansed and made into a format required for an analysis. Often we also need to verify the data we are working with, to ensure accuracy. In this section we would outline the steps taken to clean the data before analysing.

The Postal codes and names of municipalities are obtained by webscraping the sourced mentioned in the previous page. To ensure accuarcy of the data, every municiplaity name and post code are cross verified from the official webpages of Brussels administration. From the above, we understand that the city of Brussels (named Ville de Bruxelles) the largest of the muncipalities are further derived into 7 districts which also hosts European Quarter. The City of Brussels have 4 official post codes.

For better analysis, in this work the city of Brussels is divided into 4 locations namely - Haren, Laeken, Neder-Over-Heembeek and Bruxelles which contains the Eurpean Quarter and adjoining areas. The other 3 areas listed are not explicitly stated as the post codes were found to overlap with other municipalities. The postcodes and name of each municiplaities are then input into a opencage geocoder API to obtain the respective latitudes and longitudes which then culminates the data preprocessing steps. A sample of the final data is shown in the screenshots 2.1.

The data for the number of co-working spaces is obtained in a .csv file from the aforementioned open data site. This is then read into a python dataframe which helps to work with data better. A list is then

	PostCode	Commune	Country	Latitude	Longitude
0	1070	Anderlecht	Belgium	50.839098	4.329653
1	1160	Auderghem	Belgium	50.817235	4.426898
2	1082	Berchem-Sainte-Agathe	Belgium	50.864923	4.294673
3	1040	Etterbeek	Belgium	50.836145	4.386174
4	1140	Evere	Belgium	50.872010	4.403418

(a) The first 5 rows of the data frame

	PostCode	Commune	Country	Latitude	Longitude
17	1150	Woluwe-Saint-Pierre	Belgium	50.830000	4.433500
18	1000	Bruxelles	Belgium	50.850450	4.348780
19	1020	Laeken	Belgium	50.886866	4.349189
20	1120	Neder-Over-Heembeek	Belgium	50.898950	4.384843
21	1130	Haren	Belgium	50.890099	4.416321

(b) The last 5 rows of the data frame

Figure 2.1: The screenshot of the data after preprocessing.

prepared by grouping the number of co-working spaces grouped by the given Postal code. A sample of the final data is shown in the screenshot 2.2 below.

	PostCode	Number of Co-working Spaces	Commune
0	1000	8	Bruxelles
1	1040	5	Etterbeek
2	1050	3	Ixelles
3	1070	1	Anderlecht
4	1080	2	Molenbeek-Saint-Jean

Figure 2.2: The first 5 rows of the data frame for co-working spaces available in Brussels.

2.3. Using the data to solve the problem at hand

This section briefly outline how the data obtained will be used to solve the problem at hand.

Every analysis starts with an exploratory data analysis. We start with visualizing the open data of the co-working spaces. From the list, the municipalities and the number of co-working spaces will be noted. To get a better feel of the data for municipalities, we also plot the municipalities data using Folium package in python [3].

The second step would be to understand the number and types of venues present in each municipality. We try to list the top 100 venues in the area in a prescribed are of 500 m, by which we will be able to understand if the venues of interest are present or not. To get a better hold of the data we prepare two files. The first list extracts the venues of interest in each municipality and the second list contains all the other venues grouped by municipality. These lists will help us to understand the areas that consist of all three of the venues as required by Pierre.

We use one hot encoding to ease our analysis as categorical data are present and cluster the municipalities according to the municipalities of interest for Pierre. This shall be done using k-means clustering algorithm. This narrowed down cluster along with the availability of co-working spaces data will enable us to finalize the municipalities most suitable for the relocation of Pierre.

3

Methodology

This section explains in detail the methodology involved in finding the most suitable location for Pierre. For this an exploratory data analysis is carried out to get the feel of the data. To analyse the data, we first carry out one hot encoding on the data. Thereafter the data is separated into venues of interest and other venues grouping by municipalities. We then find the number of municipalities with all three venues of interest. These results along with the other venues list is then used to cluster the municipalities by means of k-means clustering algorithm.

3.1. Exploratory Data Analysis

We plot a bar diagram 3.1 depicting the number of available work spaces and the corresponding Municipalities. Thus only the municipalities which has a working space could be considered. We then

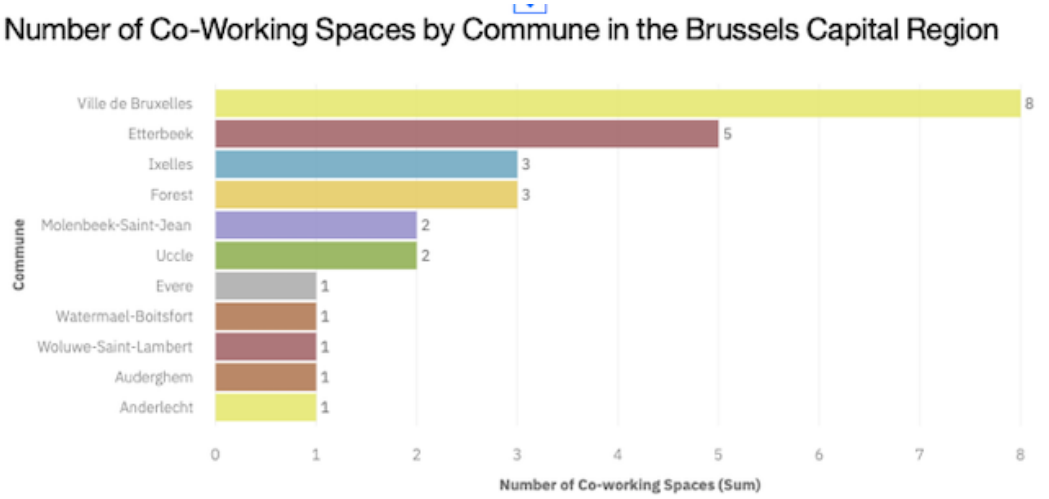


Figure 3.1: The number of available work spaces and the corresponding municipalities.

try to check if the venues of interest are also present in these areas. For this, we need to modify the dataframe using one-hot encoding.

3.2. One-hot encoding of data

One-hot encoding is used to represent the categorical data in a binary form. Thus the value is returned 1 only for the integer index else a 0 is returned. The following plots 3.2 shows the data before and after one-hot encoding.

	Commune	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
698	Haren	50.890099	4.416321	Snack Verdun	50.891731	4.417738	Snack Place
699	Haren	50.890099	4.416321	Kortenbach (MIVB)	50.892075	4.417218	Bus Stop
700	Haren	50.890099	4.416321	Kortenbachplein / Place Cortenbach	50.892389	4.417450	Plaza
701	Haren	50.890099	4.416321	Spoor / Voie 1 (Spoor 1)	50.888499	4.419596	Platform
702	Haren	50.890099	4.416321	Dienstpad Schaarbeek Group R - Haren-Zuid	50.890915	4.410647	Trail

(a) The first 5 rows of the data frame before one-hot encoding

	Commune	African Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bagel Shop	Bakery	Bar	Bath House	Beer Bar	Beer Store	Belgian Restaurant	Big Box Store	Bike Rental / Bike Share	Bistro
0	Anderlecht	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	Auderghem	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0
2	Berchem-Sainte-Agathe	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	Bruxelles	0	0	0	1	1	0	0	1	1	0	0	0	1	0	0	0
4	Etterbeek	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	1

(b) The first 5 rows of the data frame after one-hot encoding

Figure 3.2: The screenshot of the data after before and after one-hot encoding.

3.2.1. Extracting venues of interest

After one-hot encoding we extract the venues of interest along with the municipalities into a new dataframe. This enables us to shortlist the municipalities of interest. We can see that many municipalities hosts only one or two of the specified venues, see figure 3.3. Only 4 hosts all three of the venues of interest.

	Commune	Farmers Market	French Restaurant	Supermarket
0	Anderlecht	0	0	1
1	Auderghem	0	1	0
2	Berchem-Sainte-Agathe	0	1	1
3	Bruxelles	0	1	1
4	Etterbeek	1	1	1
5	Evere	0	0	1
6	Forest	0	1	1
7	Ganshoren	0	0	1
8	Haren	0	0	0
9	Ixelles	0	1	0
10	Jette	0	0	1

Figure 3.3: An example of the extracted dataframe with venues of interest.

3.2.2. Clustering

Clustering helps us to sort the data into similar structures based on a given criteria. Thus data points within a cluster are similar while data points in other clusters are different. In this work our main aim is to cluster the data based on number of venues present in each cluster. For the clustering, we make use of the machine learning algorithm called k-means clustering. This algorithm iteratively tries to group the data into predefined k number of non-overlapping clusters with one dataset unique to one cluster only.

The important factor in the k-means algorithm is the predefined number (k) for classification. To determine the right value of k, we use the Elbow Method. We define the term inertia as the sum of squared distances of samples to their closest cluster center. To determine the optimal number of clusters, we have to select the value of k at the “elbow” ie the point after which the distortion/inertia start decreasing in a linear fashion.

In this work we find that the optimal number of clusters could be 5 3.4. To verify if the k-value is

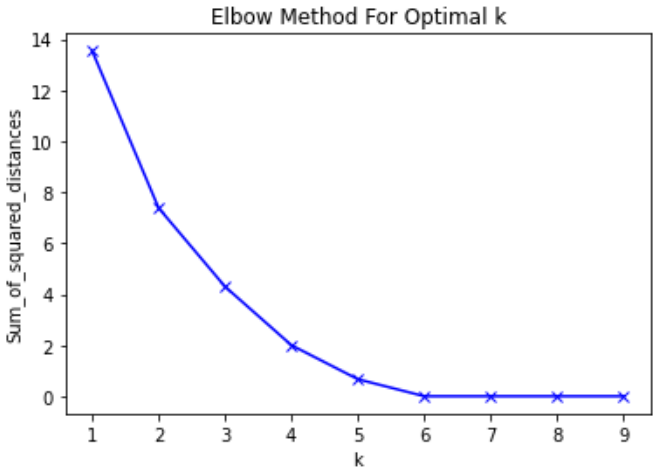


Figure 3.4: Optimal value of k from Elbow Method.

correct or not we also use Silhouette Score which is superior technique to Elbow method. This score measures the goodness of the cluster and has values ranging between -1 and 1. If the clustering is good enough, the Silhouette score should be closer to 1. In our case, the Silhouette score value was highest at k=5 with a flat curve for higher k values. Thus the number of clusters chosen is 5 The cluster

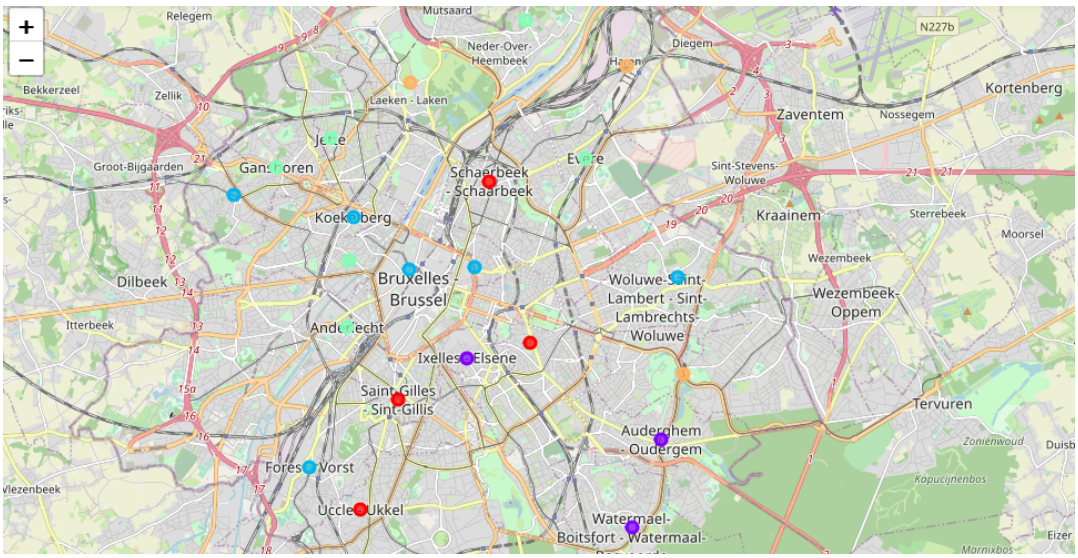


Figure 3.5: Clustered municipalities based on the number of venues.

values are as follows:

- 0 - Only French Restaurant or Farmers market is present or both (Violet)
- 1 - Two venues are present including supermarket (Light Blue)
- 2 - Only supermarket is present (Light Green)

- 3 - None of the venues are present (Orange)
- 4 - All three venues are present (red)

3.2.3. Finding the ideal locations

We now try to find out the number of co-working spaces and number of interesting venues in each municipality. The figure 3.6 summarizes all the findings of this analysis. From the stacked column

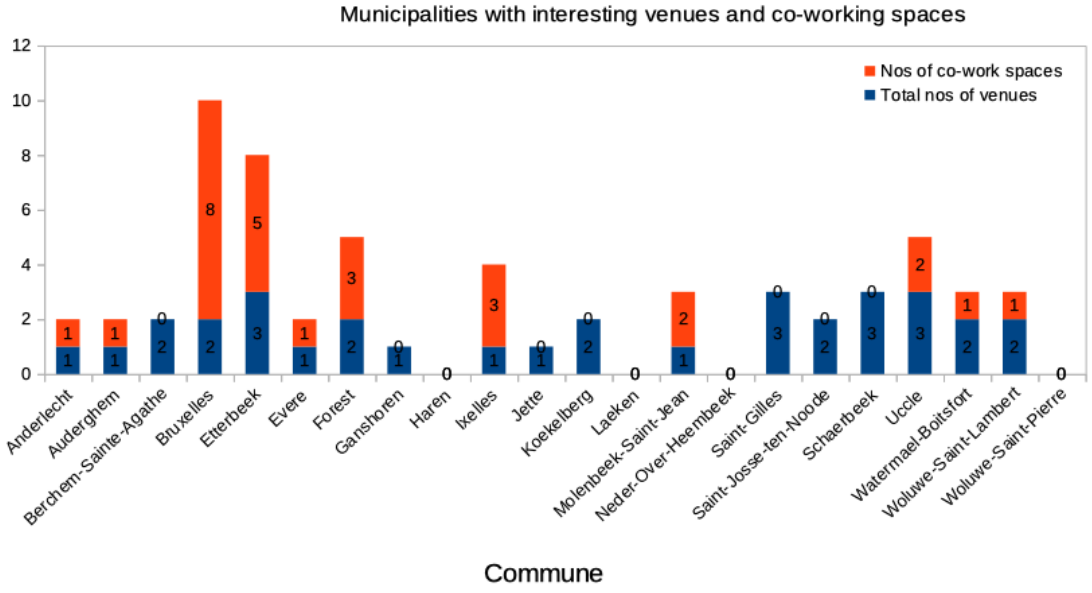


Figure 3.6: Number of venues and co-working spaces in each municipality

chart we clearly see that there are some locations which are ideally suitable for Pierre while some locations can be considered as secondary options.

4

Results and Discussion

4.1. Results

Considering the number of venues present in each municipality, the deciding factor is the presence of three venues as specified by Pierre. We tried to list out the municipalities with all three venues present. From the list we could see that 4 places had all three venues listed. These places are:

- Etterbeek
- Saint-Gilles
- Schaerbeek
- Uccle

Since Saint-Gilles and Schaerbeek do not have co-working spaces, these can be effectively removed from the most desired locations. There are also 7 municipalities where any of the two specified venues are present. These venues could be used as secondary options for Pierre provided co-working spaces are available. Out of the 7 municipalities 4 of them are suitable as the next best options.

Comparing the data from the number of co-working spaces, we see that **Etterbeek** and **Uccle** are most suitable for Pierre with 5 and 1 co-working spaces respectively. The table 4.1 shows all the ideal neighbourhoods and secondary options available for choosing.

Commune	**No of Venues**	**No of co-working spaces**
Etterbeek	3	5
Uccle	3	1
Bruxelles	2	8
Forest	2	3
Watermael Boistfort	2	1
Woluwe Saint Lambert	2	1

Table 4.1: The ideal locations (in green) along with secondary options.

4.2. Discussion

Though the Etterbeek and Uccle have all three venues and also have available work spaces, there are also several deciding factors that determines if the place is most suitable or not. Some of these factors could be availability of apartments in the regions, rental prices of apartments and studio flats in those areas, noise in the region and so on.

Apartment prices in Uccle has seen a 15.4 percent increase from 2018 to 2019. Compared to Uccle, Etterbeek apartment prices have risen only 3.3 percent from 2018 to 2019 [9]. Based on the CIB Rent Barometer 2020 provided by the real estate organisation in Flanders, Belgium, the average rental prices

for apartments in Etterbeek is listed at around €1076 while in Uccle it is about €1206 **??**. Thus Etterbeek could be seen as a slightly better option when compared to Uccle. The rentals for apartments in the municipalities listed in the table 4.1 are seen to be above €1000 as recorded by CIB rent barometer **??**. Since the rental prices are similar, the availability of apartments will play a huge role in the location that will be chosen by Pierre.

5

Conclusion

The main aim of this project was to identify the ideal place in Brussels to relocate and start a new start up business for Pierre. With relocation comes the challenge of acclimatising the city which could be difficult since there are always certain areas near your home that a person will wish he had in the new place. Hence to make this relocation go smoothly as possible Pierre suggested certain venues he wish to have near his ideal location in addition to the presence of a co-working space. This will help Pierre to choose a studio in the location selected.

From the list of venues grouped by the name of municipality we could extract three venues which are required to be present in the ideal location according to Pierre. A total of five municipalities have the desired venues as suggested by Pierre. The locations are then clustered into five based on the k-means clustered algorithm. The list of ideal location is narrowed down by comparing with the number of co-working spaces available. This then results in two ideal locations- Etterbeek and Uccle which have 5 and 1 co-working spaces respectively.

We then try to suggest a better option of the two based on demography and rental prices per apartment in each commune. Considering all the above factors , Etterbeek is considered to be the ideal choice for Pierre for his relocation and start up company.

References

[1] artwhere.be. *Les 19 communes et centres publics d'action sociale de la Région de Bruxelles-Capitale*. url: https://www.avcb-vsgb.be/fr/Communes/anderlecht.html (visited on 03/21/2021).

[2] cirb.brussels. *Les communes*. url: http://pouvoirs-locaux.brussels/fr/acteurs/les-communes (visited on 03/21/2021).

[3] *folium 0.12.1*. url: https://pypi.org/project/folium/ (visited on 03/22/2021).

[4] Foursquare 2021. *OpenCage Geocoding API*. url: https://foursquare.comm (visited on 03/22/2021).

[5] i-CITY. *Newcomer*. url: https://www.brussels.be/newcomer (visited on 03/21/2021).

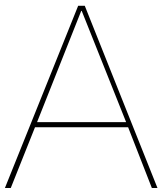
[6] OpenCage GmbH. *OpenCage Geocoding API*. url: https://opencagedata.com (visited on 03/22/2021).

[7] Papadopoulos, A. , Verniers, . Louis and Hermans, Theo Jozef. *Brussels.Encyclopedia Britannica*. 2019,October 20. url: https://www.britannica.com/place/Brussels (visited on 03/28/2021).

[8] U Talk Freelance. *Liste des espaces de coworking à Bruxelles*. 2016,September 19. url: https://data.gov.be/fr/dataset/310f1955-2326-47f4-96f3-5da9899e54bc (visited on 03/20/2021).

[9] *Update on housing prices in Brussels in 2020*. url: https://www.kbcbrussels.be/retail/en/home/the-cost-of-housing-in-brussels.html (visited on 03/28/2021).

[10] Wiki. *List of Munciplaities of the Brussels Capital Region*. url: https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Brussels-Capital_Region (visited on 03/21/2021).



Title

B

Task Division

Table B.1: Distribution of the workload

	Task	Student Name(s)
	Summary	
Chapter 1	Introduction	
Chapter 2		
Chapter 3		
Chapter *		
Chapter *	Conclusion	
	Editors	
	CAD and Figures	
	Document Design and Layout	