Predicting the best place to open a luxury hotel in Brussels, Belgium

Shalakha Khandelwal

November 30,2019

Introduction

- This presentation is for those who are planning to start a new hotel in the city of Brussels. I will provide a suggestion on what would be the best venue to start a new hotel in this densely populated and highly visited city.

- Brussels is a region of Belgium comprising 19 municipalities, including the City of Brussels, which is the capital of Belgium. Brussels is located in the central portion of the country and is a part of both the French Community of Belgium and the Flemish Community.

- Brussels is the most densely populated and the richest region in Belgium in terms of GDP per capita.

- Some touristic attractions like Medieval Grand-Place, is indeed grand, with many 17th-century buildings and daily flower markets.

- Reopened in 2006, the Atomium, Brussels' Eiffel Tower, provides great views, inside and out. Architecture fans should visit Musee Horta, home of Belgian master architect Victor Horta. St. Gery's clubs and bars are packed year-round. Seafood eateries abound in Ste. Catherine. Walk, rather than get snarled up in traffic, in the narrow streets.

Business Problem

- We will focus on the issue of where to open a new hotel in a city like Brussels, once one has decided to go ahead. Let's imagine Marriott Hotels are willing to open a new luxury hotel, the first and foremost important decision will be the location for its new hotel.

1. On what basis can they decide the new hotel's location?
2. While selecting the place there are key points to consider like they need to check where the most well-visited venues of the city are?
3. If in case there are already other luxury hotels which have good ratings, will it be risky to open new one near these hotels?
4. Out of scope for this project: Rent and land values in the neighbourhoods, budget for the interior decoration of the hotel, budget for opening different restaurants in the hotel etc.

Data acquisition and cleaning

- As we are doing analysis for those who want to open a new luxury hotel in Brussels,

1. The first requirement is to collect Brussels postal codes data with the name of respective neighbourhoods.
2. The second requirement would be collect data related to latitude and longitude values of the same neighbourhoods.
3. The third requirement would be clean data as required for the analysis and
4. The fourth requirement would be to merge two datasets to be used for exploratory analysis.

Data acquisition

- We will extract the data of all municipalities(municipality names and their postal codes) from below Wikipedia page using Beautiful Soup.

  List of municipalities Brussels-Capital Region

- We will extract the data related to latitude and longitude values of Neighbourhoods with a csv file (**zipcode-belgium.csv**) saved at local machine and then at the server.

- Now let's use Foursquare Location to get the amount of most visited venues per Neighbourhood. The details can be retrieved using search endpoint. Link to the dataset is:

  Foursquare Developer site

Data cleaning

- Data cleaning for neighbourhood data with their postal codes

1. Required data is in table with class – 'wikitable sortable', so we extract data and transform it into a new pandas data frame.

2. Data cleaning here will include removing extra columns, changing data type of postal codes from string to integer etc.

3. Data frame after data cleaning –

Continued…

	Neighborhood	Postal code
0	Anderlecht	1070
1	Auderghem	1160
2	Berchem-Sainte-Agathe	1082
3	Bruxelles-Ville*	1000
4	Etterbeek	1040
5	Evere	1140
6	Forest	1190
7	Ganshoren	1083
8	Ixelles	1050
9	Jette	1090
10	Koekelberg	1081
11	Molenbeek-Saint-Jean	1080
12	Saint-Gilles	1060
13	Saint-Josse-ten-Noode	1210
14	Schaerbeek	1030
15	Uccle	1180
16	Watermael-Boitsfort	1170
17	Woluwe-Saint-Lambert	1200
18	Woluwe-Saint-Pierre	1150

Data cleaning

- ## Data cleaning for Neighbourhood data with their latitude and longitude values

1. Data was extracted from a csv file named – **zipcode-belgium.csv** located at local system. This file was then saved to Watson studio server and data was extracted in notebook using pd.read_csv function to pandas data frame.

2. Top 5 rows of data in df_data_0 data frame –

	Postal code	Neighborhood	Latitude	Longitude
0	1000	Bruxelles	50.846557	4.351697
1	1020	Laeken	50.883392	4.348713
2	1030	Schaerbeek	50.867604	4.373712
3	1040	Etterbeek	50.836851	4.389510
4	1050	Ixelles	50.822285	4.381571

Data cleaning

- Merging two data frames into a single data frame for further analysis

1. The two data frames were merged using pd.merge function and below is the screenshot of top 5 rows of resulting data frame(Brussel_df) –

	Postal code	Neighborhood	Latitude	Longitude
0	1070	Anderlecht	50.838141	4.312340
1	1160	Auderghem	50.815657	4.433139
2	1082	Berchem-Sainte-Agathe	50.863984	4.292702
3	1000	Bruxelles	50.846557	4.351697
4	1040	Etterbeek	50.836851	4.389510

Exploratory Analysis

- In this phase, I worked on detailed analysis of source data to identify best locations to open a luxury hotel in Brussels.

- This analysis includes below steps –

1. Importing libraries and creating map of Brussels with Neighbourhoods

2. Creating a data frame with top 10 venues of each Neighbourhood

3. Cluster Neighbourhoods - Run k-means to cluster the neighbourhoods into 7 clusters

4. Visualizing resulting clusters

Importing libraries and creating map of Brussels with Neighbourhoods

- Libraries imported: Matplotlib and associated plotting modules, Geocoder nominatin to convert an address to latitude and longitude values, K-means for clustering stage and Folium(map rendering library).

- Used geocoder to get latitude and longitude values of Brussels and create a map including all the Neighbourhoods of Brussels.
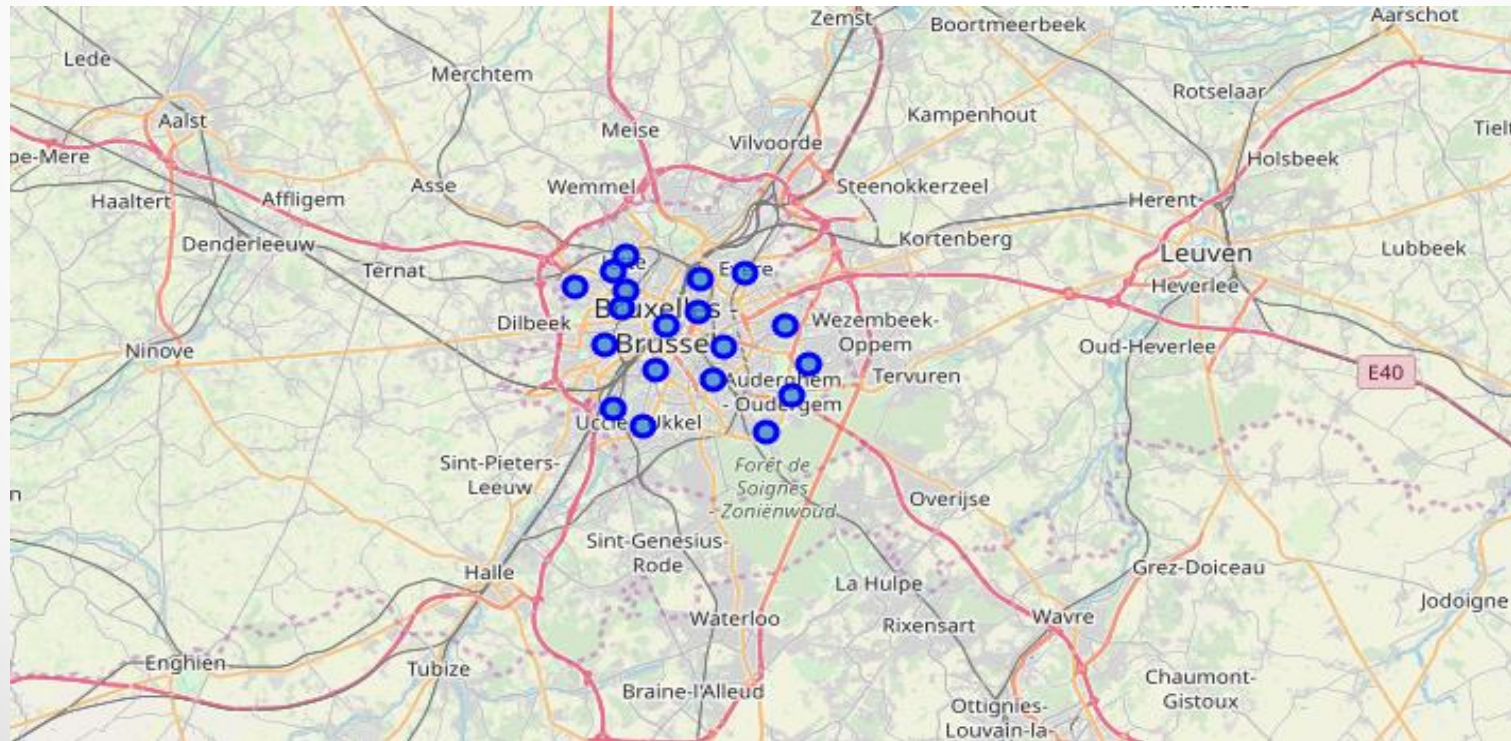
Creating a data frame with top 10 venues of each neighbourhood

- Created a function to get nearby venues of each neighbourhood.
- Grouped the data based on 'Neighborhood' column.
- Used onehot encoding to create a new data frame(Brussel_grouped) with frequency of each venue in neighbourhoods.
- Calculated top 10 venues of each neighbourhood.
- Resultant data frame (neighbourhoods_venues_sorted) –

|	Neighborhood	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	Anderlecht	Bar	Restaurant	Convenience Store	Grocery Store	Greek Restaurant	Italian Restaurant	Belgian Restaurant	Discount Store	Middle Eastern Restaurant	Pharmacy
1	Auderghem	French Restaurant	Belgian Restaurant	Middle Eastern Restaurant	Snack Place	Grocery Store	Sushi Restaurant	Italian Restaurant	Garden	Restaurant	Fast Food Restaurant
2	Berchem-Sainte-Agathe	Greek Restaurant	Plaza	Gym	French Restaurant	Supermarket	Café	Furniture / Home Store	Sandwich Place	Restaurant	Burger Joint
3	Bruxelles	Chocolate Shop	Bar	Beer Bar	Sandwich Place	Belgian Restaurant	Greek Restaurant	Gastropub	Bookstore	Plaza	Thai Restaurant
4	Etterbeek	Pizza Place	Italian Restaurant	Brasserie	Lounge	Plaza	Bar	French Restaurant	Greek Restaurant	Monument / Landmark	Museum

Cluster Neighbourhoods

- Run K-means to cluster neighbourhoods into 7 clusters from Brussel_df data frame and then merge it with neighbourhoods_venues_sorted data frame into a new data frame named Brussel_merged.

	Postal code	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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1070	Anderlecht	50.838141	4.312340	3	Bar	Restaurant	Convenience Store	Grocery Store	Greek Restaurant	Italian Restaurant	Belgian Restaurant	Discount Store	Middle Eastern Restaurant	Pharmacy
1	1160	Auderghem	50.815657	4.433139	0	French Restaurant	Belgian Restaurant	Middle Eastern Restaurant	Snack Place	Grocery Store	Sushi Restaurant	Italian Restaurant	Garden	Restaurant	Fast Food Restaurant
2	1082	Berchem-Sainte-Agathe	50.863984	4.292702	3	Greek Restaurant	Plaza	Gym	French Restaurant	Supermarket	Café	Furniture / Home Store	Sandwich Place	Restaurant	Burger Joint
3	1000	Bruxelles	50.846557	4.351697	0	Chocolate Shop	Bar	Beer Bar	Sandwich Place	Belgian Restaurant	Greek Restaurant	Gastropub	Bookstore	Plaza	Thai Restaurant
4	1040	Etterbeek	50.836851	4.389510	0	Pizza Place	Italian Restaurant	Brasserie	Lounge	Plaza	Bar	French Restaurant	Greek Restaurant	Monument / Landmark	Museum

Visualizing resulting Clusters

- From the resulting data frame – Brussel_merged, I then created a map showing all the clusters created.
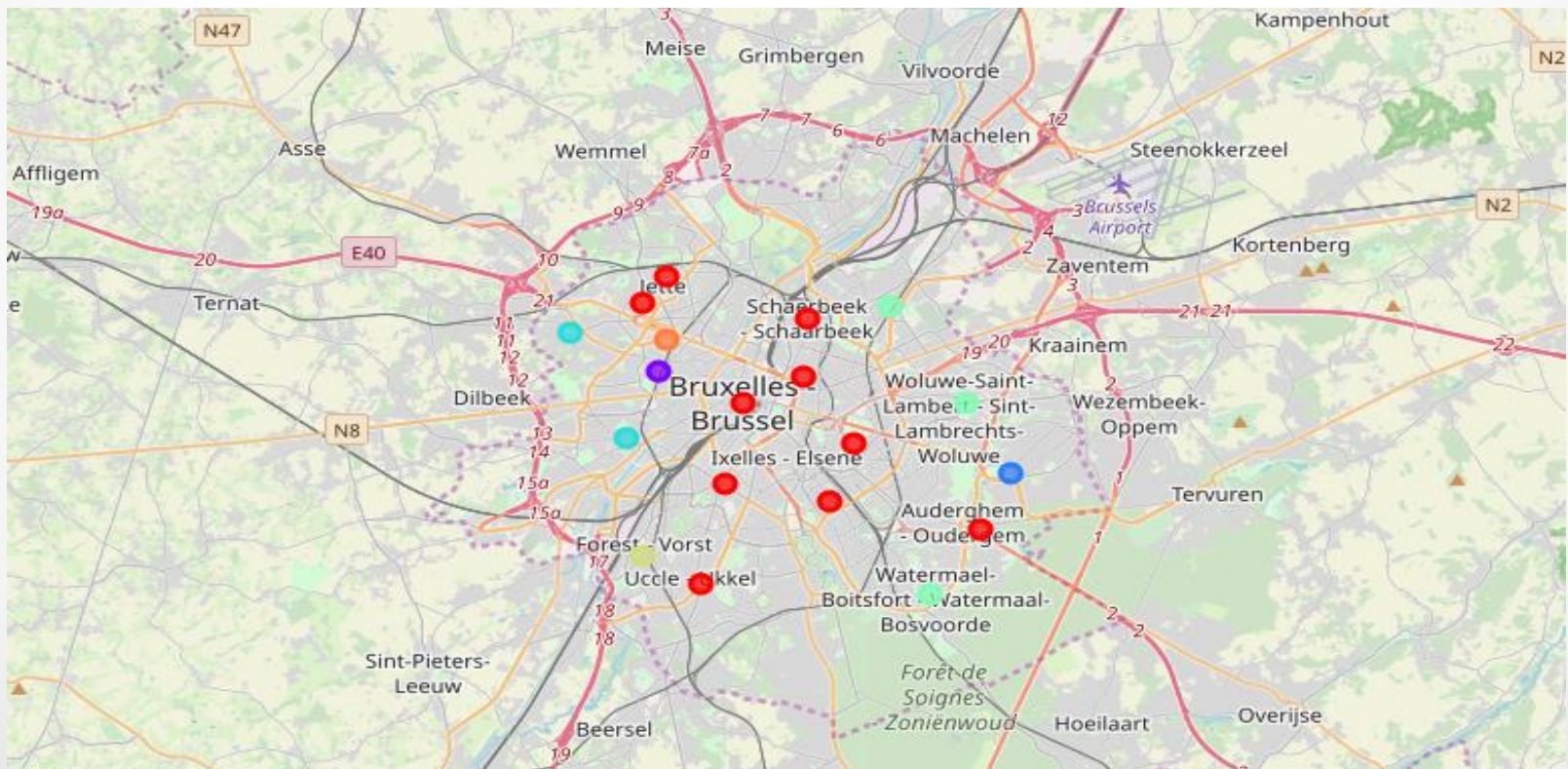
Visualizing resulting Clusters

- ## 1ˢᵗ Cluster

1. Most of the municipalities situated in centre of Brussels consisting of lot of venues and some of the locations tourists would be interested in visiting.

2. Neighbourhoods which we would be most interested in are 1000 Brussels, 1040 Etterbeek and 1050 Ixelles.

3. These municipalities are located in the centre of the city and have most common venues like bars, restaurants and chocolate shops. Many of the touristic attractions like grand place, maneeken pis are located in centre of the city.

Continued…

```
In [34]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 0, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[34]:

	Postal code	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	1160	French Restaurant	Belgian Restaurant	Middle Eastern Restaurant	Snack Place	Grocery Store	Sushi Restaurant	Italian Restaurant	Garden	Restaurant	Fast Food Restaurant
3	1000	Chocolate Shop	Bar	Beer Bar	Sandwich Place	Belgian Restaurant	Greek Restaurant	Gastropub	Bookstore	Plaza	Thai Restaurant
4	1040	Pizza Place	Italian Restaurant	Brasserie	Lounge	Plaza	Bar	French Restaurant	Greek Restaurant	Monument / Landmark	Museum
7	1083	Bar	Sandwich Place	Italian Restaurant	Burger Joint	Flower Shop	Steakhouse	Supermarket	Creperie	Pool	Chinese Restaurant
8	1050	Bar	Vietnamese Restaurant	French Restaurant	Plaza	Thai Restaurant	Paper / Office Supplies Store	Pizza Place	Burger Joint	Lebanese Restaurant	Sporting Goods Shop
9	1090	Bar	Supermarket	Bakery	Gastropub	Grocery Store	Tram Station	Snack Place	Platform	Italian Restaurant	Brasserie
12	1060	Bar	Brasserie	Plaza	French Restaurant	Italian Restaurant	Grocery Store	Pizza Place	Bakery	Moroccan Restaurant	Friterie
13	1210	Sandwich Place	Italian Restaurant	Restaurant	Snack Place	Kebab Restaurant	Pizza Place	Supermarket	Plaza	Hotel	Lebanese Restaurant
14	1030	Plaza	Tram Station	Bakery	Supermarket	Snack Place	Italian Restaurant	Gastropub	Beer Store	Pub	French Restaurant
15	1180	Plaza	Cosmetics Shop	French Restaurant	Supermarket	Sandwich Place	Bar	Italian Restaurant	Tram Station	Food & Drink Shop	Bakery

Visualizing resulting Clusters

- **2nd Cluster**

1. The 2nd cluster consists of only one neighbourhood with most common venues like supermarket, spa and performing arts venue.

Visualizing resulting Clusters

- 3rd Cluster

1. The 3rd cluster as well has one neighbourhood with most common venues like Sports club, hockey field and park.


```
In [36]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 2, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[36]:

	Postal code	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
18	1150	Sports Club	Hockey Field	Park	Tram Station	Trail	Event Service	Flower Shop	Fish & Chips Shop	Fast Food Restaurant	Farmers Market

Visualizing resulting Clusters

- 4th Cluster

1. This cluster consists of 2 neighbourhoods with most common venues like Bar, plaza, convenience store and gym.


```
In [37]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 3, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[37]:

	Postal code	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	1070	Bar	Restaurant	Convenience Store	Grocery Store	Greek Restaurant	Italian Restaurant	Belgian Restaurant	Discount Store	Middle Eastern Restaurant	Pharmacy
2	1082	Greek Restaurant	Plaza	Gym	French Restaurant	Supermarket	Café	Furniture / Home Store	Sandwich Place	Restaurant	Burger Joint

Visualizing resulting Clusters

- ## 5th Cluster

1. 5th cluster consists of three neighbourhoods with most common venues like restaurant, park, bakery and pharmacy.

```
In [38]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 4, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[38]:

	Postal code	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
5	1140	Restaurant	Pharmacy	Bakery	Pizza Place	Sushi Restaurant	Park	Pool	Chinese Restaurant	Sandwich Place	Bus Stop
16	1170	Italian Restaurant	Restaurant	Chinese Restaurant	Park	Thai Restaurant	Bus Stop	Soccer Field	Sandwich Place	French Restaurant	Bookstore
17	1200	Park	Restaurant	French Restaurant	Supermarket	Grocery Store	Skating Rink	Chinese Restaurant	Basketball Court	Italian Restaurant	Bus Station

Visualizing resulting Clusters

- 6th Cluster

1. 6th cluster consists of one neighbourhood with most common venues like supermarket, pharmacy and health food store.

```
In [39]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 5, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[39]:

	Postal code	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
6	1190	Supermarket	Pharmacy	Health Food Store	Men's Store	Halal Restaurant	Snack Place	Athletics & Sports	Plaza	Park	Cafeteria

Visualizing resulting Clusters

- 7th Cluster

1. 7th cluster consists of one neighbourhood with most common venues like gym, hostel and breakfast joint.


```
In [40]: Brussel_merged.loc[Brussel_merged['Cluster Labels'] == 6, Brussel_merged.columns[[0] + list(range(5, Brussel_merged.shape[1]))]]
```

Out[40]:

	Postal code	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
10	1081	Gym / Fitness Center	Hostel	Bed & Breakfast	Indonesian Restaurant	French Restaurant	Falafel Restaurant	Park	Piano Bar	History Museum	Sandwich Place

Results

The following are the highlights of the 7 clusters above:

- The most common venues are clearly located in 1st Cluster (centre of Brussels), which makes the choice of the final location very easy.

- As for restaurants, bars, coffee shops and chocolate shops are very popular also in 1st Cluster (centre of Brussels), Especially in 1000 Brussels, 1040 Etterbeek and 1050 Ixelles.

- Although, the Clusters have variations, a very visible presence is the predominance of chocolate shops, tourists locations and bars and restaurants, so new hotel can be opened in those places.

Discussion and Conclusion

- It is noticeable that 1st Cluster is the most viable clusters to build a new luxury hotel with guarantees. The proximity to a big number of Restaurants (lunch and dinner venues for guests), Coffee shops and other amenities and accessibility to station are also very important points to take into account when making the right choice.

- The municipalities like 1000 Brussels, 1040 Etterbeek and 1050 Ixelles, lies in the centre of Brussels and have proximity to all kind of most common venues visited by locals as well as tourists. These neighbourhoods could be the best places to open new luxury hotel in the city.

- In conclusion, this project would have had better results if there were more available data in terms of actual land pricing data within the area, public transportation access and allowance of more venues exploration with the Foursquare (limited venues for free calls). However, based on the available data, my advice to Marriott group would be to focus on 1000 Brussels, 1040 Etterbeek and 1050 Ixelles when investing on a new luxury hotel.