

Capstone Project — The Battle of Neighborhoods in Warsaw: New Hotel location

Introduction

a. Background

This report is for those who are planning to start a new hotel in the city of Warsaw. It provides a suggestion on what would be the best venue to start a new hotel in a highly visited city with already many good hotels.

Warsaw, is one of the most popular tourist destinations. Currently ranked as 74 of most visited cities worldwide. Warsaw is definitely one of the best places to start up a new hotel business. In 2018, Warsaw received over 2,8 million visitors, tourist areas in Warsaw provide huge opportunities for hotels. We will go through the benefits and pitfalls of opening a new Hotel in a highly visited city with already many hotels offering their services. The core of Warsaw is made of 18 districts but, We will concentrate on districts, where the busiest venues of Warsaw can be found to target the tourists visiting the city. With that in mind, I will be able to find the Top 3 districts to open a brand new hotel in the city.

b. Business Problem

This report focusses on the issue of where to open a new hotel in a city like Warsaw, once one has decided to go ahead. Let's imagine investment company Mariette willing to open a new luxury hotel, a first and foremost important decision will be the location for its new hotel.

c. Interest

- On what basis can Mariette decide its new hotel's location?
- While selecting the place there are key points to consider like they need to check out like where the most well-visited venues of the city are?
- If incase there are already other luxury hotels which have good ratings, will it be risky to open new one near these hotels?

Data Preparation

Data 1: Warsaw has a total of 18 districts. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 10 districts as well as their latitude and longitude coordinates.

This dataset exists for free on the web.

Link to the dataset is : https://en.wikipedia.org/wiki/Districts_of_Warsaw

Data 2: Now once we got the districts' latitude and longitude, let's use Foursquare Location to get the amount of Restaurants venues per district, which will give us an idea of where the tourist are moving when visiting the city. This will already show us the best districts to start a hotel. The Restaurants details can be retrieved using search endpoint. For my project I need restaurant data, and in search endpoint there is a attribute called Rating, i.e for each Rating foursquare has a defined Rating which will help us to get the desired data.

Link to the dataset is: <https://developer.foursquare.com/docs/data>

a. Scrapping Warsaw Districts Table from Wikipedia

I first make use of Districts of Warsaw page from Wiki to scrap the table to create a data-frame. For this, I used requests and BeautifulSoup4 library to create a dataframe containing name of the 18 districts of Warsaw Area.

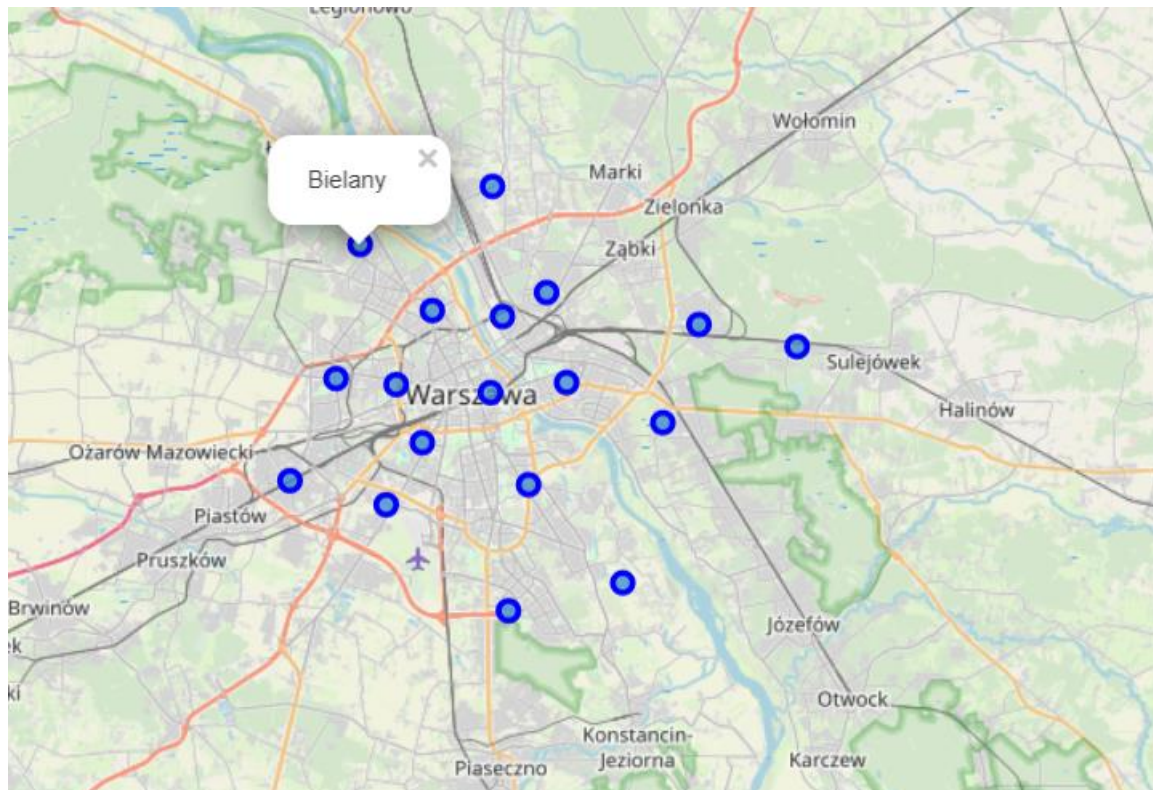
b. Getting Coordinates of Major Districts : Geopy Client

Next objective is to get the coordinates of these 18 districts using geocoder class of Geopy client.

Resulting Data 1 set: Warsaw districts with attributes

	District	Population	Size	Latitude	Longitude
0	Mokotów	220682	35.4	52.193987	21.045781
1	Praga Południe	178665	22.4	52.237396	21.071258
2	Ursynów	145938	48.6	52.141039	21.032321
3	Wola	137519	19.26	52.236238	20.954781
4	Bielany	132683	32.3	52.294652	20.929980
5	Targówek	123278	24.37	52.275192	21.058085
6	Śródmieście	122646	15.57	52.232810	21.019067
7	Bemowo	115873	24.95	52.238974	20.913288
8	Białołęka	96588	73.04	52.319665	21.021177
9	Ochota	84990	09.7	52.212225	20.972630
10	Wawer	69896	79.71	52.220358	21.137083
11	Praga Północ	69510	11.4	52.264884	21.027344
12	Ursus	53755	09.35	52.196098	20.882899
13	Żoliborz	48342	08.5	52.267594	20.979698
14	Włochy	38075	28.63	52.186109	20.948438
15	Wilanów	23960	36.73	52.153083	21.110441
16	Rembertów	23280	19.30	52.261415	21.162819
17	Wesoła	22811	22.6	52.251794	21.229276

I used python **folium** library to visualize geographic details of Warsaw and its 18 districts



Exploratory Data Analysis:

We exploratory data analysis to uncover hidden properties of data and provide useful insights to the investor.

Finally, we used of Foursquare API and get the top 100 venues that are in Warsaw districts within a radius of 2000 meters.

The number of unique venue categories was 103.

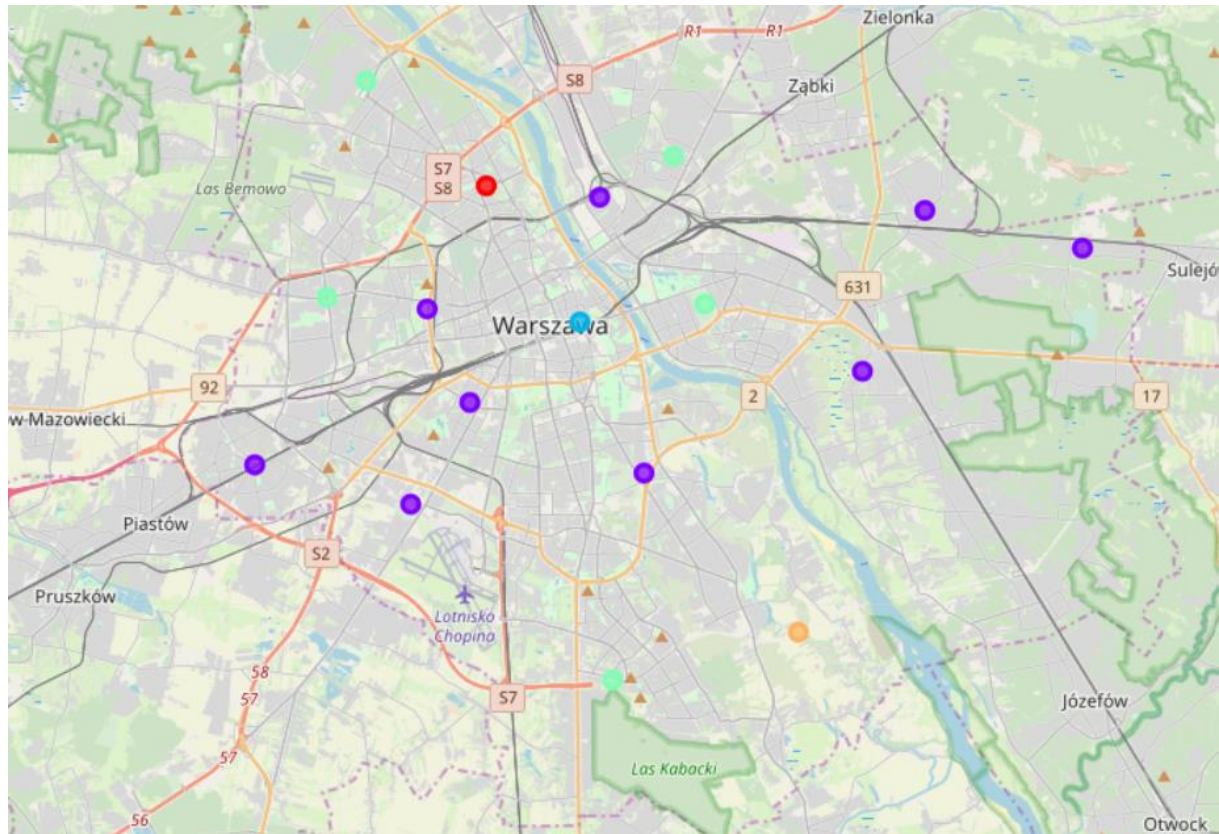
We identified Top 10 venues in each district:

	District	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	Bemowo	Food & Drink Shop	Italian Restaurant	Donut Shop	Coffee Shop	Bus Station	Sandwich Place	Japanese Restaurant	Sporting Goods Shop	Supermarket	Café
1	Bielany	Clothing Store	Shopping Mall	Bus Station	Bus Line	Ice Cream Shop	Burger Joint	Metro Station	Coffee Shop	Comedy Club	Convenience Store
2	Mokotów	Skate Park	Diner	Arcade	Bus Station	Racetrack	Intersection	Lake	Coffee Shop	Comedy Club	Convenience Store
3	Ochota	Hotel	Pharmacy	Yoga Studio	Electronics Store	Dessert Shop	Italian Restaurant	Diner	Park	Basketball Court	Skating Rink
4	Praga Południe	Café	Bus Line	Racetrack	Road	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
5	Praga Północ	Bike Rental / Bike Share	Amphitheater	Comedy Club	Convenience Store	Bus Station	Plaza	Light Rail Station	Diner	Coffee Shop	Cupcake Shop
6	Rembertów	Café	Pizza Place	Discount Store	Gay Bar	Dim Sum Restaurant	Cocktail Bar	Grocery Store	Coffee Shop	Comedy Club	Convenience Store
7	Targówek	Bus Station	Gym / Fitness Center	Pet Store	Cupcake Shop	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Deli / Bodega
8	Ursus	Park	Train Station	Train	Hotel	Supermarket	Yoga Studio	Dim Sum Restaurant	Clothing Store	Cocktail Bar	Coffee Shop
9	Ursynów	Italian Restaurant	Clothing Store	Food Court	Food & Drink Shop	Theme Park Ride / Attraction	Supermarket	Sporting Goods Shop	Dim Sum Restaurant	Cocktail Bar	Coffee Shop
10	Wawer	Gun Range	Athletics & Sports	Food & Drink Shop	Falafel Restaurant	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop
11	Wesoła	Train Station	Pizza Place	Plaza	Tennis Court	Yoga Studio	Dim Sum Restaurant	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club
12	Wilanów	Baby Store	Yoga Studio	Discount Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store
13	Wola	Bus Station	Grocery Store	Café	Skating Rink	Flea Market	Falafel Restaurant	Italian Restaurant	Motorcycle Shop	Music Store	Park
14	Włochy	Accessories Store	Tram Station	Hotel	Bed & Breakfast	Gastropub	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
15	Śródmieście	Coffee Shop	Café	Beer Bar	Cocktail Bar	Sushi Restaurant	Bakery	Nightclub	Italian Restaurant	Ice Cream Shop	Mediterranean Restaurant

Resulting clusters

We used prescriptive analytics to help the investor to decide a location of new hotel. We used clustering (KMeans). Our expectation would be based on the similarities of venue categories, these districts will be clustered.

We found 5 clusters, marked in different colors.



Cluster 0 - Most common venues: Cafe, Coffee Shop and Beer Bars

	District	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
13	Żoliborz	0	Café	Coffee Shop	Beer Bar	Park	Burger Joint	Italian Restaurant	Cocktail Bar	Deli / Bodega	Bus Station	Kebab Restaurant

Cluster 1 - Most common venues: Park, Pizza Place, Bike Rental, Train Station

	District	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	Mokotów	1	Skate Park	Diner	Arcade	Bus Station	Racetrack	Intersection	Lake	Coffee Shop	Comedy Club	Convenience Store
3	Wola	1	Bus Station	Grocery Store	Café	Skating Rink	Flea Market	Falafel Restaurant	Italian Restaurant	Motorcycle Shop	Music Store	Park
9	Ochota	1	Hotel	Pharmacy	Yoga Studio	Electronics Store	Dessert Shop	Italian Restaurant	Diner	Park	Basketball Court	Skating Rink
10	Wawer	1	Gun Range	Athletics & Sports	Food & Drink Shop	Falafel Restaurant	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop
11	Praga Północ	1	Bike Rental / Bike Share	Amphitheater	Comedy Club	Convenience Store	Bus Station	Plaza	Light Rail Station	Diner	Coffee Shop	Cupcake Shop
12	Ursus	1	Park	Train Station	Train	Hotel	Supermarket	Yoga Studio	Dim Sum Restaurant	Clothing Store	Cocktail Bar	Coffee Shop
14	Włochy	1	Accessories Store	Tram Station	Hotel	Bed & Breakfast	Gastropub	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
16	Rembertów	1	Café	Pizza Place	Discount Store	Gay Bar	Dim Sum Restaurant	Cocktail Bar	Grocery Store	Coffee Shop	Comedy Club	Convenience Store
17	Wesoła	1	Train Station	Pizza Place	Plaza	Tennis Court	Yoga Studio	Dim Sum Restaurant	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club

Cluster 2 - Most common venues: Cofee Shop, Beer Bar, Coctail Bar (No Hotels in Top 10 venues)

	District	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
6	Śródmieście	2	Coffee Shop	Café	Beer Bar	Cocktail Bar	Sushi Restaurant	Bakery	Nightclub	Italian Restaurant	Ice Cream Shop	Mediterranean Restaurant

Cluster 3 - Most common venues: Italian Restaurant, Food Shop, Shopping Mall, Gym (No Hotels in Top 10 venues)

	District	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	Praga Południe	3	Café	Bus Line	Racetrack	Road	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store
2	Ursynów	3	Italian Restaurant	Clothing Store	Food Court	Food & Drink Shop	Theme Park Ride / Attraction	Supermarket	Sporting Goods Shop	Dim Sum Restaurant	Cocktail Bar	Coffee Shop
4	Bielany	3	Clothing Store	Shopping Mall	Bus Station	Bus Line	Ice Cream Shop	Burger Joint	Metro Station	Coffee Shop	Comedy Club	Convenience Store
5	Targówek	3	Bus Station	Gym / Fitness Center	Pet Store	Cupcake Shop	Diner	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Deli / Bodega
7	Bemowo	3	Food & Drink Shop	Italian Restaurant	Donut Shop	Coffee Shop	Bus Station	Sandwich Place	Japanese Restaurant	Sporting Goods Shop	Supermarket	Café

Cluster 4 - Most common venues: Baby Store, Yoga Studio, Discount Store (No Hotels in Top 10 venues)

	District	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
15	Wilanów	4	Baby Store	Yoga Studio	Discount Store	Cocktail Bar	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	Deli / Bodega	Department Store

Results

The following are the observation about the 5 clusters above:

1. Hotels are located only in Cluster 1 (Ochota). It looks strange is contrast with local observation.
Conclusion: most of hotels data is missing in analysed data set.
2. Entertainment, and shopping venues are popular in Cluster 1
3. Italian Restaurants, Food Shops, Shopping Mall and Gyms are popular in Cluster 3
4. Most common venues in Cluster 2: Cofee Shop, Beer Bar, Coctail Bar
5. Wilanów and Śródmieście districts have different venues characteristics than other Warsaw districts.

Discussion and Conclusion

Above analysis shows that decision about hotel location based on venues data from foursquare.com could be risky.

Data about hotels in Warsaw looks incomplete. Additional data source should be used.

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 transportantion access and allowance of more venues exploration with the Foursquare (limited venues for free calls).

It is recommended to make more detailed analysis based on transportation access and business offices location.