IBM Capstone Project: The Battle of the Neighborhoods (Week 2)

New Town Hall, Old Town, Toruń

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

As part of the Final Capstone Project of the IBM Data Science Course, we will explore the suitability to open a new Yoga Studio in the **Toruń** neighborhoods, Poland.

Torun already has a huge gym and yoga studio, however, if we are going to open a small business, I would like to explore **any locations may still be a feasible suggestion, or not at all**.

There are a number of parameters we can assume, for example in descending priority below:

```
1. proximity to areas with more lively business around e.g. town
centre
2. proximity to residential area
3. proximity to work area
```

On top of the above search priority, we are also concern with **an area with no immediate competitor**. We also prefer locations **with complementary business around (e.g. gym but with no yoga studio, hotel, swimming pool) which may help us attract potential clients**

This business problems will possibly be interested to any furture business owners in any neightbourhoods, to follow a similar search approach and enrich with any addition search method before making a business decision.

Description of Data

To solve our business problem, we will analyze below factors of the neighborhoods before making the business decision:

- distance to town center
- any existing yoga studio in the neighborhoods
- number of and distance to lively business (e.g restaurants, markets, shopping centre), if any

Torun has 24 boroughs according to Wikipedia. We get the latitude & longitude coordinates for these districts to start exploring our candidate neighborhoods.

Required information will be extracted and generated as data sources by:

- Geospatial coordinates of town center and neighborhoods using **Geopy Library**
- Geospatial coordinates and addresses can also be retrieved by using **Google Cloud Service — Google Map API**
- Venue data using **Foursquare API**

Neighborhood Candidates

Latitude & longitude coordinates of our candidate neighborhoods is created for the 24 boroughs to use as centroids for search. The latitude & longitude is made available by using the Google Maps geocoding API.

Gathering Useful Data

Step 1: Create Map of Business Area

In this case, create the plain map of the Torun area, Poland.

```
# create map of Torun areas
address = "Torun, Poland"

geolocator = Nominatim()
```

```
(53.0102721, 18.6048094)
```

```
import getpass
# reverse geocoding
address = "Torun, Poland"

api_key = getpass.getpass() # API Key credentials is hidden
lat, lng = get_coordinates(api_key, address, verbose=False)
torun_corr = lat, lng

print(f'The geograpical coordinates of {address} are {lat}, {lng}.')

get_address = get_address(api_key, lat, lng)

print(f'The address of {lat, lng} is {get_address}.')
```

```
........
The geograpical coordinates of Torun, Poland are 53.0137902, 18.5984437.
The address of (53.0137902, 18.5984437) is DK80, 87-100 Toruń, Poland.
```

Utilize **Google Cloud Service—Google Map API**

Step 2: Get Boroughs Coordinates of the Neighborhood in the Business Area

We use a function and the Geopy library to generate the longitude and latitude coordinates of the neighborhoods and then create a Pandas dataframe.

```
# create dataframe of the 24 boroughs with their geodata
df = pd.DataFrame(geo_location(boroughs))
df
```

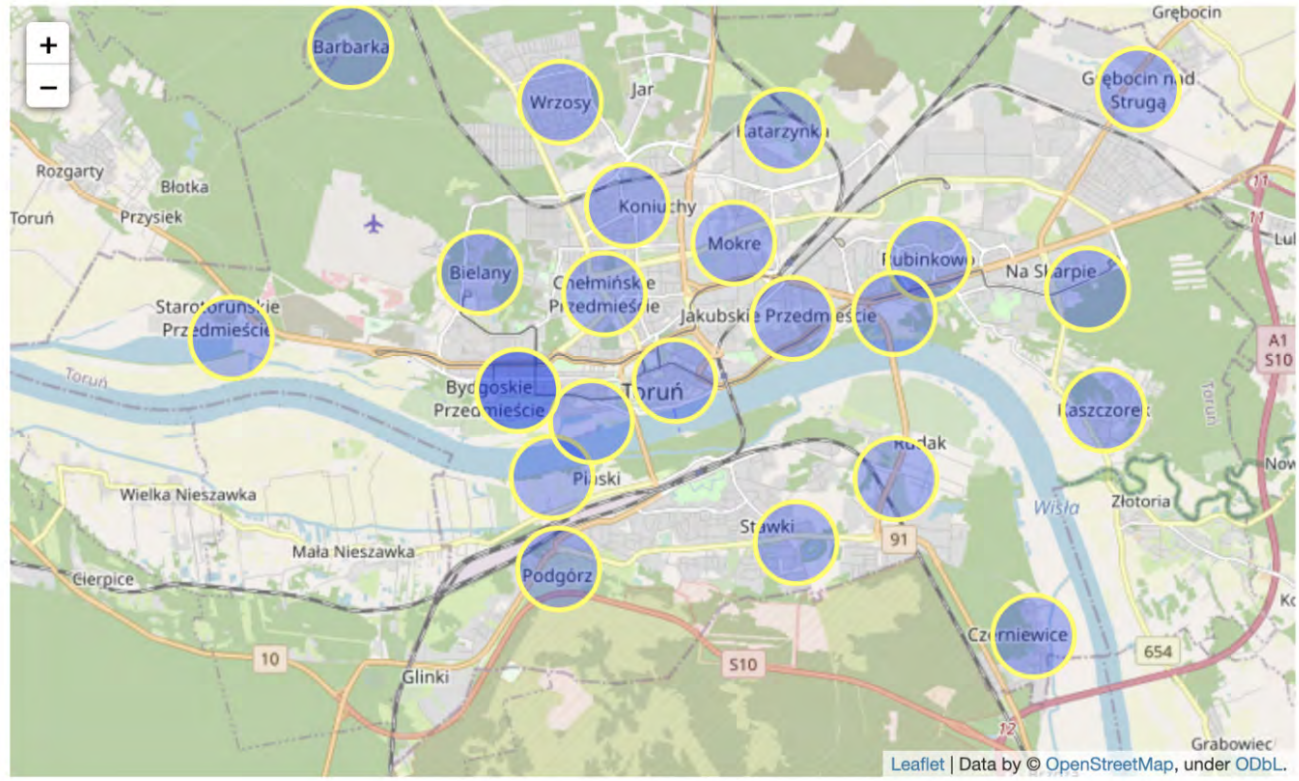
	Boroughs	Latitude	Longitude
0	Barbarka	53.053819	18.541648
1	Bielany	53.025359	18.568668
2	Bielawy	53.282226	18.740640
3	Bydgoskie Przedmieście	53.010490	18.576364
4	Chełmińskie Przedmieście	53.022616	18.594666
5	Czerniewice	52.979523	18.684492
6	Glinki	52.945260	18.340220
7	Grębocin nad Strugą	53.048392	18.706664
8	Jakubskie Przedmieście	53.019558	18.634295
9	Kaszczorek	53.007971	18.699385
10	Katarzynka	53.043213	18.632105

By the assumption that the town center, usually the Old Town or New Town areas are the center of business activities.

```
# instantiate a feature group for the boroughs in the dataframe
borough = folium.map.FeatureGroup()
```

Use Folium module to create map

```
map_torun
```



Showing the 24 boroughs around Toruń

Step 3: Get Distance from the Center of Business: Stare Miasto

```
Town Center Stare Miasto (Old Town) coordinates:
(53.0116718, 18.609319116005494)
```

Out[238]:

	Boroughs	Latitude	Longitude	Distance downtown (km)
19	Stare Miasto	53.011672	18.609319	0.000000
18	Rybaki	53.006466	18.591805	1.306968
4	Chełmińskie Przedmieście	53.022616	18.594666	1.562586
8	Jakubskie Przedmieście	53.019558	18.634295	1.886896
12	Mokre	53.028967	18.621943	2.100365
14	Piaski	52.999470	18.583759	2.183068
3	Bydgoskie Przedmieście	53.010490	18.576364	2.208700
21	Przedmieście	53.010490	18.576364	2.208700
11	Koniuchy	53.033814	18.599533	2.547583

Get the distance from downtown (Old Town) and sort by descending order

Step 4: Initiate Foursquare to get Nearby Venues of the Town Center

We use the Foursquare API to look at 100 nearby venues within radius of 1000 meters of the specified neighborhoods.

```
# @hidden_cell
CLIENT_ID = # your Foursquare ID
CLIENT_SECRET = getpass.getpass() # your Foursquare Secret (hidden)
VERSION = '20190513'  # Foursquare API version
print('Foursquare initiated.')
```

```
........
Foursquare initiated.
```

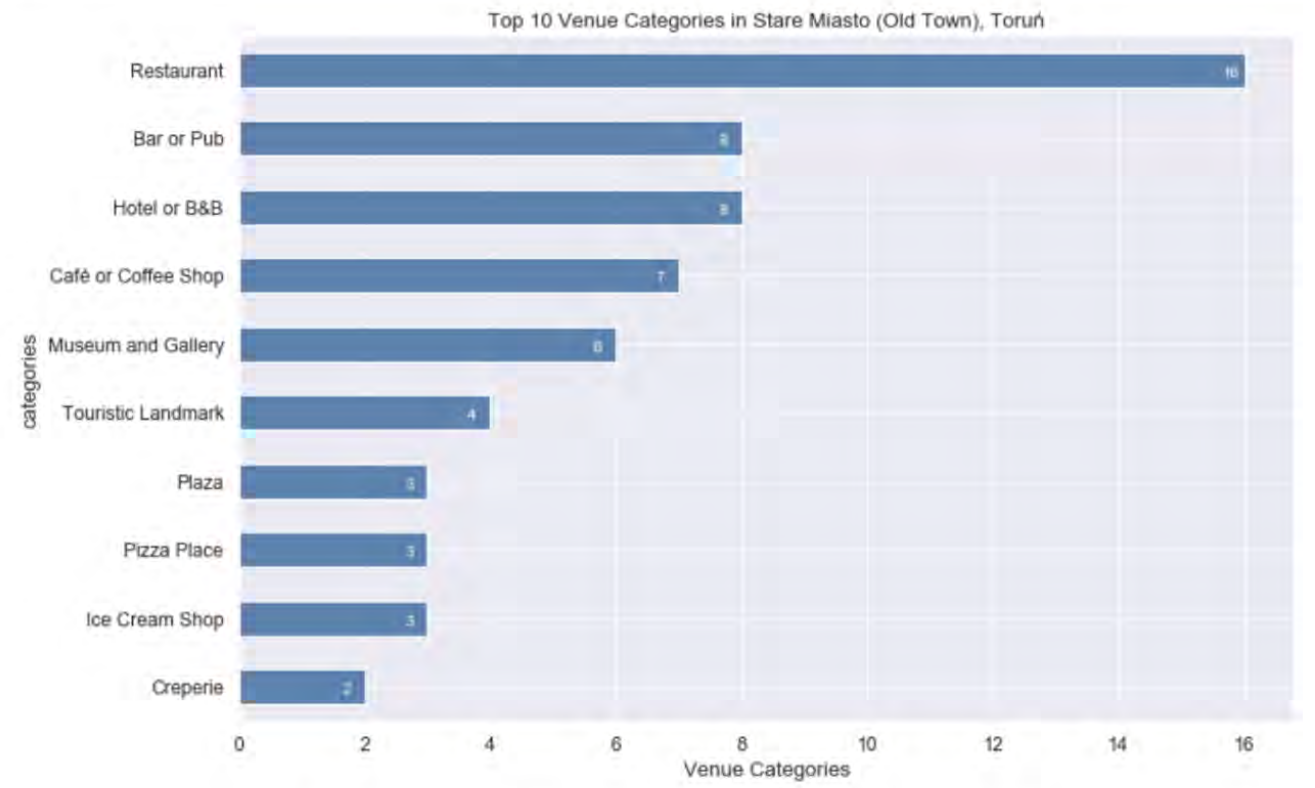
We are going to visualize and analyze nearby venues of:

1. Stare Miasto (Old Town), Toruń (which is supposed to be the Toruń centre)
2. All neighborhoods of Toruń (to look at all the business activities around Toruń)

Step 5: Venue Categories in Old Town, Toruń

We will first clean up the data, and then plot by the **value_counts** of **categories**

```
plt.show()
```



Plot Top 10 Venues Categories of Old Town by horizontal bar chart

Step 6: Get an Overview of the Overall Lively Business Areas around Toruń (Nearby Venues of ALL Neighborhoods)

To look at the distribution around all neighborhoods areas in Torun.

```
torun_venues
```

```
(263, 7)
```

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Stare Miasto	53.011672	18.609319	Koi Sushi	53.011778	18.608490	Sushi Restaurant
1	Stare Miasto	53.011672	18.609319	Róże i Zen	53.010368	18.608665	Café
2	Stare Miasto	53.011672	18.609319	Szeroka No 9	53.010660	18.608163	Restaurant
3	Stare Miasto	53.011672	18.609319	Central Coffee Perks	53.010978	18.608622	Coffee Shop
4	Stare Miasto	53.011672	18.609319	Klub NRD	53.011052	18.612019	Pub
5	Stare Miasto	53.011672	18.609319	Lenkiewicz Lody	53.010852	18.609500	Ice Cream Shop
6	Stare Miasto	53.011672	18.609319	Rynek Staromiejski	53.010576	18.603919	Plaza
7	Stare Miasto	53.011672	18.609319	Kredens	53.011325	18.605119	Pub
8	Stare Miasto	53.011672	18.609319	Ratusz	53.010754	18.604300	History Museum

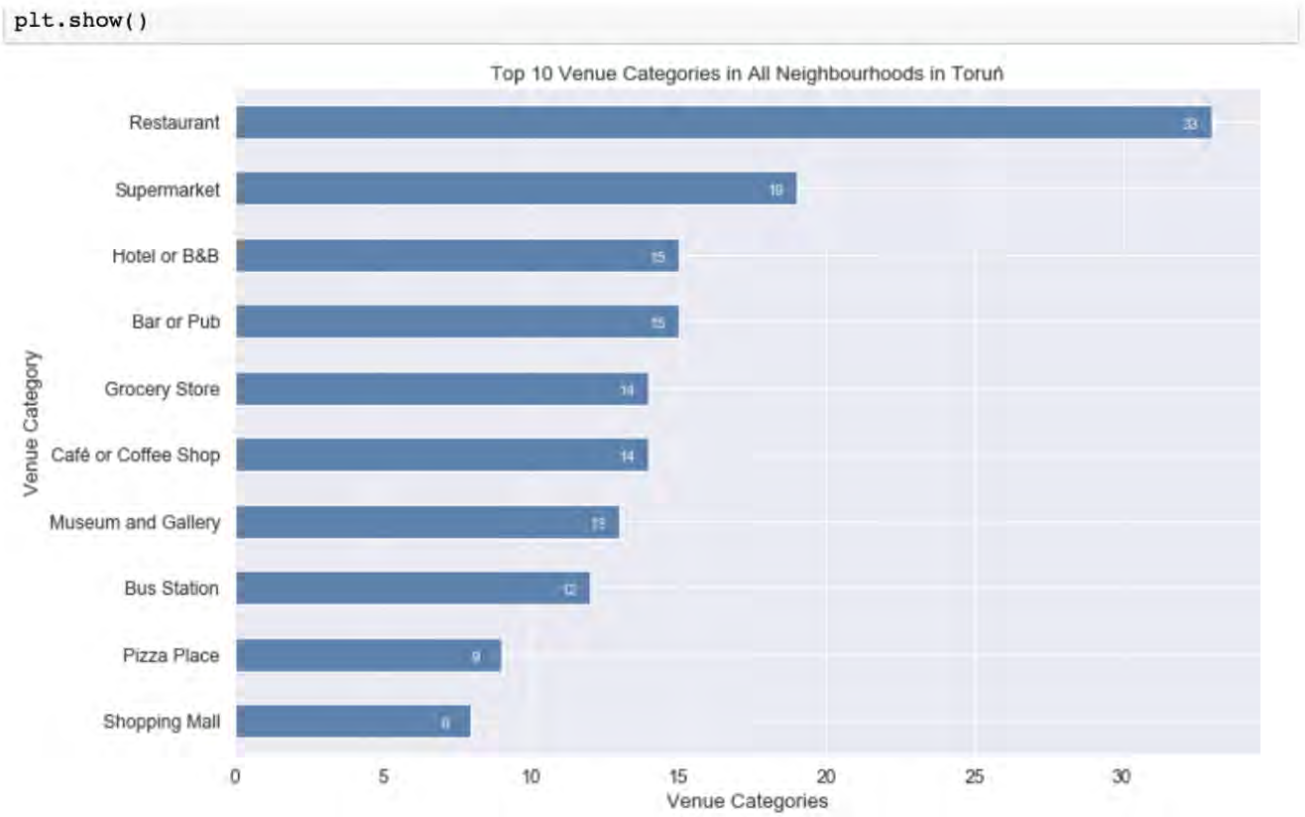
```
torun_venues
```

233	Wrzosy	53.046837	18.585489	Sklep spożywczo-przemysłowy u Gajesia	53.048003	18.586842	Food & Drink Shop
234	Wrzosy	53.046837	18.585489	Kuflewscy	53.050233	18.582860	Bakery
235	Wrzosy	53.046837	18.585489	Wymienialnia Ksiazek AMMA	53.042770	18.581673	Bookstore
236	Wrzosy	53.046837	18.585489	Intermarche	53.042882	18.581200	Market
237	Wrzosy	53.046837	18.585489	Angels Toruń	53.042768	18.580488	Athletics & Sports
238	Wrzosy	53.046837	18.585489	Optimetria Karczewski	53.040546	18.583097	Optical Shop
239	Wrzosy	53.046837	18.585489	Krągłe Placki	53.050478	18.574042	Pizza Place
240	Wrzosy	53.046837	18.585489	Sklep Spożywczy	53.050224	18.573713	Food & Drink Shop
241	Na Skarpie	53.023280	18.696140	Carrefour Market	53.021929	18.694168	Supermarket
242	Na Skarpie	53.023280	18.696140	Piotr i Paweł	53.026035	18.694782	Grocery Store
243	Na Skarpie	53.023280	18.696140	Olimpijska (pętla)	53.027398	18.698626	Light Rail Station
244	Na Skarpie	53.023280	18.696140	Kolankowskiego	53.028547	18.691970	Bus Station

Step 7: The Overall Top 10 Business Category Around Torun

We plot by the top10 of **Venue Category** count in ALL neighborhoods.

```
plt.show()
```



We also compare preliminary the data according to a ascending order to distance downtown (proximity to business center).

By now we understand our data better.

This concludes the data gathering and cleaning. As it goes, we will further crunch our data to provide different visualizations as our analysis proceed.

Methodology

In this project we will approach the business problem of opening a yoga studio around Toruń by understanding the business livelihood around the area, identify the type and robustness of buesiness activity and analyze potential locations that are favorable to our target business, a yoga studio in this case.

We adopt a methodology in three stages:

1. Identify the location, category and distribution of all business venues around the targeted business aread, Toruń. This enables us to have a quick overview of business activities category distributions and identify the most robust business areas.
2. Use heatmaps and frequency analysis to analyze the density of Old Town business activities and identify a few potential good locations for the yoga studio
3. Use K-means clustering to refine the search down to specific requirements, for
   - example,
   - within 500 meters to hotels,

     at least 500 meters away from bars or clubs which may be too loud or attract alcoholics

Analysis

In Step 1 we want to confirm our understanding the robustness of business is related to the distance of town center (old town), which is supposed to be the center of business activity.

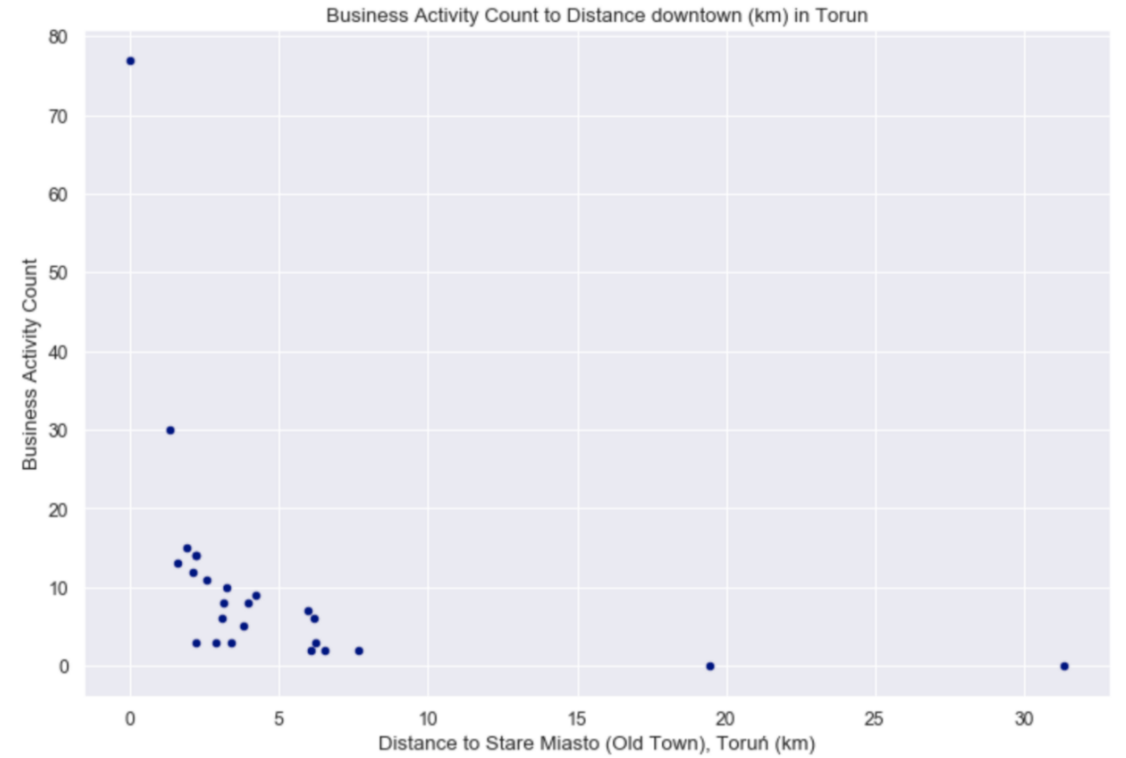
We make educated assumption that because this time the area chosen is a smaller town and we want to assume that business is better at areas where human activities are very active.

The direction of our business decision takes off from here.

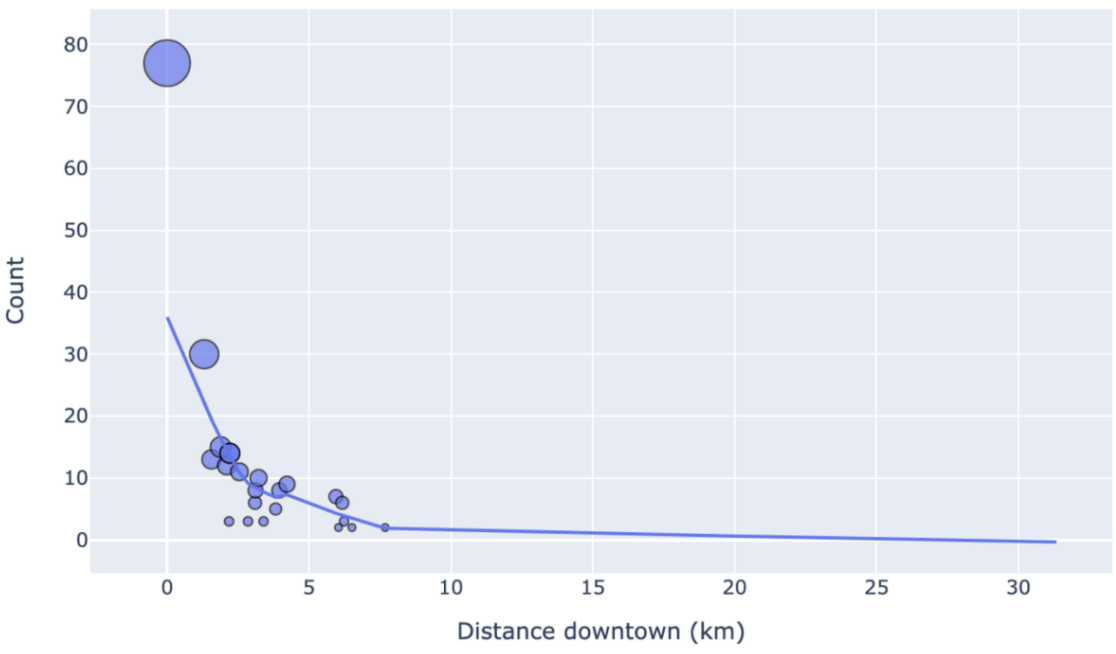
Step 1: Data Visualization — Confirm the Relationship of Overall Business Activity (Robustness) against Distance to Old Town

We compose a Scatter Plot by **Venue Categories Count** and **Distance downtown (km)**.

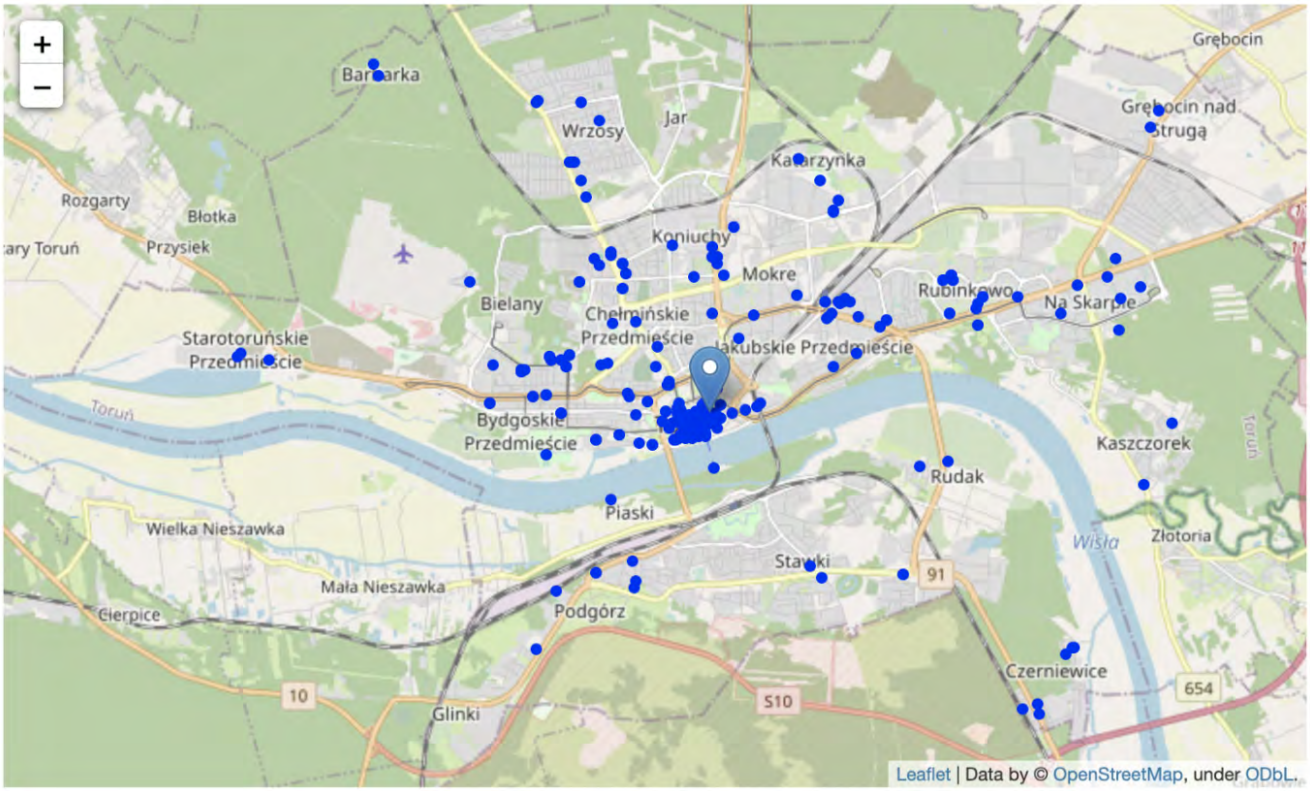
```
plt.show()
```



Venue Categories Count to Distance downtown (km) in Torun



A Map Visualization of the Overall Business Activities in Toruń



Upon Step 1, we set the direction of our business decision.

Step 2: Extract Meaningful Data from the Overall Business Activity

Relating to our business problem, we want to look at potential venues to open a yoga studio or conduct yoga related activities.

For this we identify:

1. **Related Venue** that may encourage potential customers to our business, and
2. **Complementary Venues** which may bring existing customers of complementary business into our offerings.

1. Frequency Analysis: Top 10 neighborhood boroughs by business activity Robustness

One hot encoding to generate the frequency of top venues in each neighborhood.

```
torun_df = torun_onehot.groupby('Neighbourhood').mean().reset_index()
torun_df.head()
```

	Neighbourhood	Airport	Athletics & Sports	Auto Garage	BBQ Joint	Bakery	Bar or Pub	Bistro	Boarding House	Bookstore	...	Skating Rink	Snack Place
0	Barbarka	0.000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.000000	0.0
1	Bielany	0.125	0.0	0.0	0.000000	0.0	0.125000	0.0	0.0	0.0	...	0.000000	0.0
2	Bydgoskie Przedmieście	0.000	0.0	0.0	0.071429	0.0	0.071429	0.0	0.0	0.0	...	0.000000	0.0
3	Chełmińskie Przedmieście	0.000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.076923	0.0
4	Czerniewice	0.000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.000000	0.0

5 rows × 61 columns

```
----Barbarka----
                  venue  freq
0          Hotel or B&B   0.5
1  Other Great Outdoors   0.5
2               Airport   0.0
3    Athletics & Sports   0.0
4        Ice Cream Shop   0.0


----Bielany----
           venue  freq
0    Bus Station  0.25
1        Airport  0.12
2       Creperie  0.12
3         Market  0.12
4  Grocery Store  0.12


----Bydgoskie Przedmieście----
           venue  freq
0  Grocery Store  0.14
1    Supermarket  0.14
2            Zoo  0.07
3     Restaurant  0.07
4    Pizza Place  0.07


----Chełmińskie Przedmieście----
                 venue  freq
0         Skating Rink  0.08
1            Multiplex  0.08
2           Restaurant  0.08
3         Hotel or B&B  0.08
4  Café or Coffee Shop  0.08
```

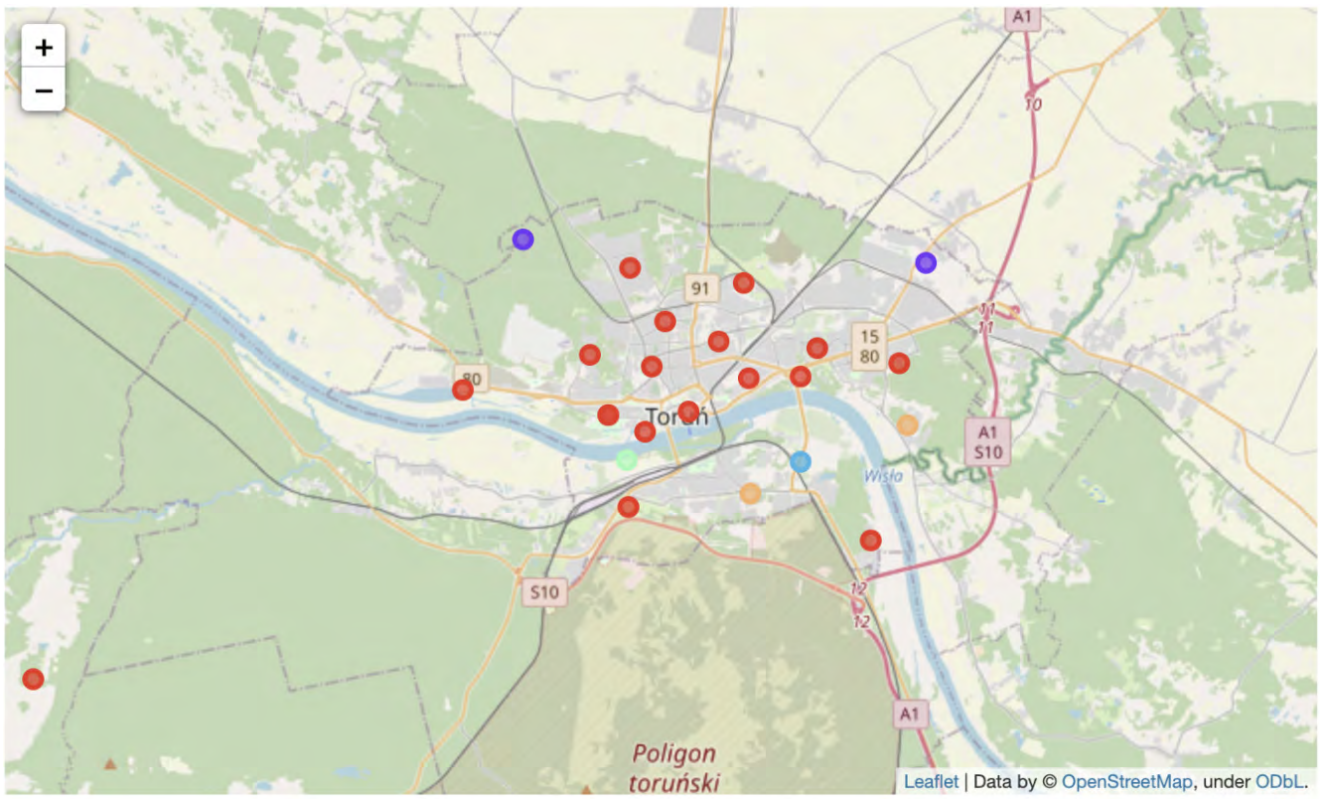
Top 10 Venues for Top 10 boroughs in Torun area

Next we look at the Top 10 Venues. Due to simplicity of data, Top 10 Venues of all neighborhoods are analyzed to get an overall full picture of the Torun area.

If the business activity around an area is enormous, it is suggested to perform Top 10 Venues for the Top 10 neighborhoods.

Neighbourhood	Latitude	Longitude	Distance downtown (km)	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
Stare Miasto	53.011672	18.609319	0.000000	0	Restaurant	Bar or Pub	Hotel or B&B	Café or Coffee Shop	Museum and Gallery	Touristic Landmark
Rybaki	53.006466	18.591805	1.306968	0	Museum and Gallery	Bar or Pub	Café or Coffee Shop	Hotel or B&B	Restaurant	Touristic Landmark
Chełmińskie Przedmieście	53.022616	18.594666	1.562586	0	Food & Drink Shop	Skating Rink	Bus Station	Pizza Place	Light Rail Station	Restaurant
Jakubskie Przedmieście	53.019558	18.634295	1.886896	0	Bus Station	Gym	Electronics Store	Clothing Store	Restaurant	Shoe Store
Mokre	53.028967	18.621943	2.100365	0	Restaurant	Gas Station	Supermarket	Shopping Mall	Museum and Gallery	Train Station

By use of K-means, the business activities are clustered into 5 clusters by rainbow colors shown on map.



Examine each Cluster and determine the discriminating venue categories that distinguish that cluster.

```
# Cluster 1
torun_merged2.loc[torun_merged2['Cluster Labels'] == 0, torun_merged2.columns[[1] + list(range(
```

	Latitude	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	
0	53.011672	Restaurant	Bar or Pub	Hotel or B&B	Café or Coffee Shop	Museum and Gallery	Touristic Landmark	Plaza	Ice Cream Shop	Pizza Place	
1	53.006466	Museum and Gallery	Bar or Pub	Café or Coffee Shop	Hotel or B&B	Restaurant	Touristic Landmark	Zoo	Food Court	Grocery Store	
2	53.022616	Food & Drink Shop	Skating Rink	Bus Station	Pizza Place	Light Rail Station	Restaurant	Hotel or B&B	Café or Coffee Shop	Multiplex	E
3	53.019558	Bus Station	Gym	Electronics Store	Clothing Store	Restaurant	Shoe Store	Museum and Gallery	Shopping Mall	Sporting Goods Shop	Tra
4	53.028967	Restaurant	Gas Station	Supermarket	Shopping Mall	Museum and Gallery	Train Station	Bakery	Diner	Comedy Club	Co

```
# Cluster 2
torun_merged2.loc[torun_merged2['Cluster Labels'] == 1, torun_merged2.columns[[1] + list(range(
```

	Latitude	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
21	53.053819	Hotel or B&B	Other Great Outdoors	Zoo	Candy Store	Food Court	Food & Drink Shop	Flea Market	Farmers Market	Factory	Electronics Store
22	53.048392	Hotel or B&B	Market	Zoo	Candy Store	Food Court	Food & Drink Shop	Flea Market	Farmers Market	Factory	Electronics Store

```
# Cluster 3
torun_merged2.loc[torun_merged2['Cluster Labels'] == 2, torun_merged2.columns[[1] + list(range(
```

	Latitude	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	52.999273	Food & Drink Shop	Auto Garage	Zoo	Candy Store	Gas Station	Food Court	Flea Market	Farmers Market	Factory	Electronics Store

```
# Cluster 4
torun_merged2.loc[torun_merged2['Cluster Labels'] == 3, torun_merged2.columns[[1] + list(range(
```

	Latitude	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	52.99947	Bus Stop	Convenience Store	River	Café or Coffee Shop	Food Court	Food & Drink Shop	Flea Market	Farmers Market	Factory	Electronics Store

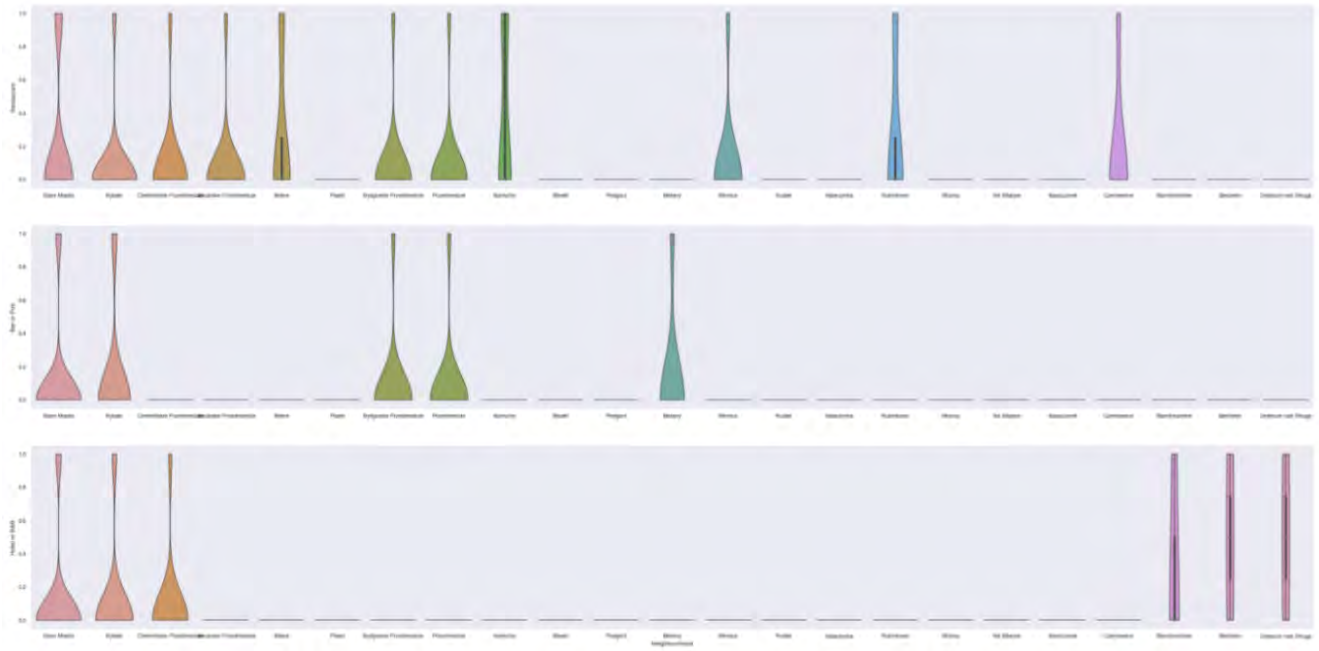
```
# Cluster 5
torun_merged2.loc[torun_merged2['Cluster Labels'] == 4, torun_merged2.columns[[1] + list(range(
```

	Latitude	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
9	52.991229	Supermarket	Auto Garage	Zoo	Café or Coffee Shop	Food Court	Food & Drink Shop	Flea Market	Farmers Market	Factory	Electronics Store
18	53.007971	Bus Stop	Supermarket	Café or Coffee Shop	Food Court	Food & Drink Shop	Flea Market	Farmers Market	Factory	Electronics Store	Diner

From the 5 clusters, we can safely arrive to an early analysis that below business activities form the most robust business activities around Toruń **business clusters**:

1. Restaurants
2. Hotels
3. Bars, pubs & clubs
4. Food & drink Shop
5. Supermarket

On the other hand, the violin plot below visualize the density estimation by use of the frequency distributions of business activities in **each neighborhood**.

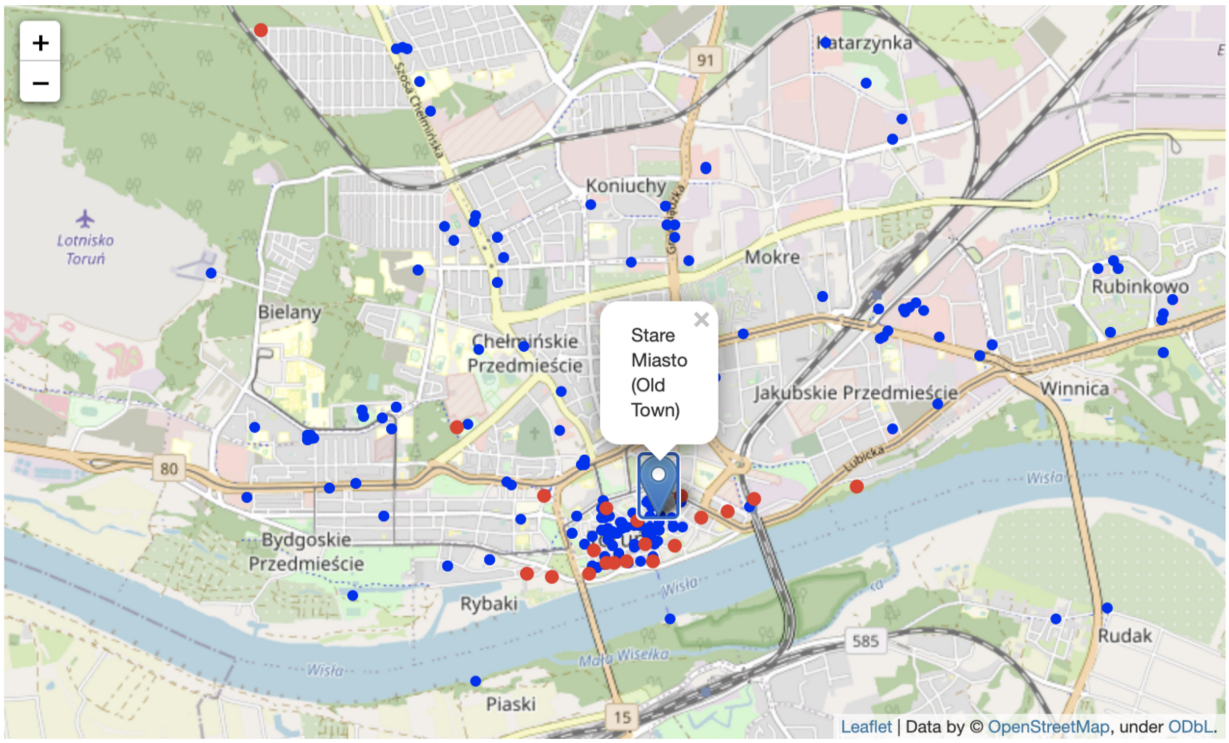


2. Heatmap: Identify all yoga related business activities around Toruń (if at all), and visualize

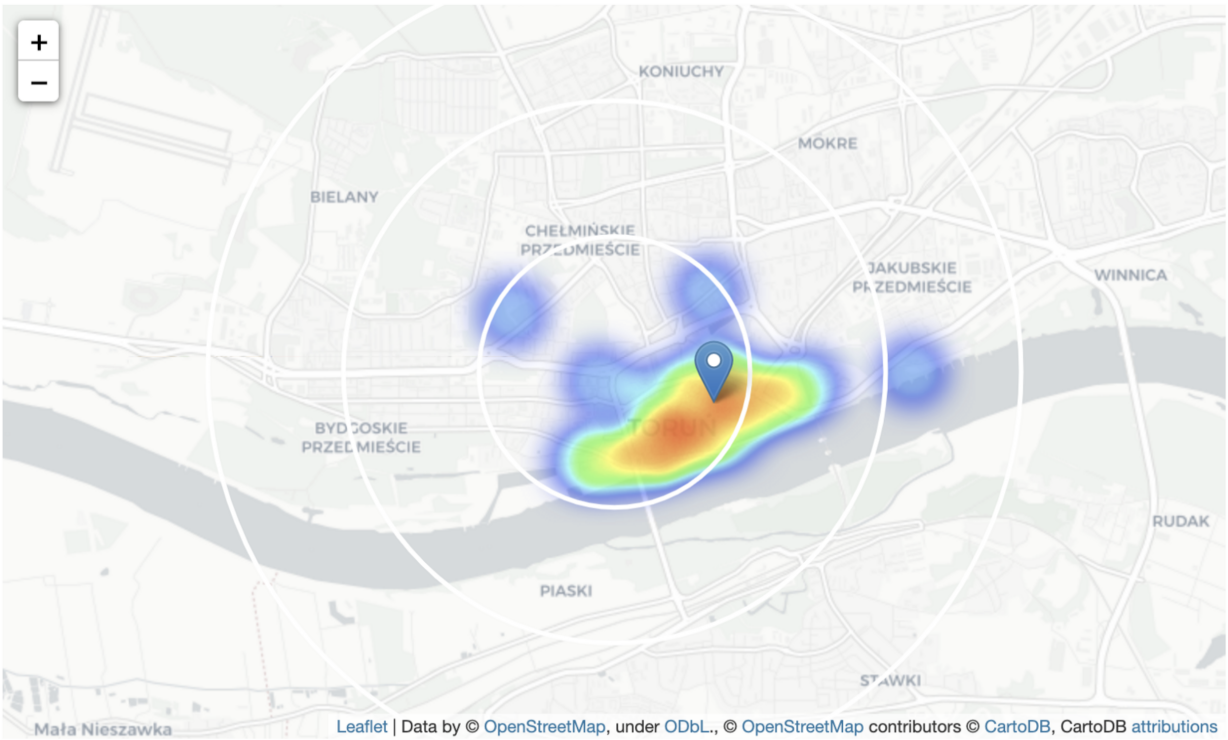
Now we zoom in to the kind of business activities relating to our business problem.

Highlight the Relevant Business Venues (hotels in this case)

On top of the full scenario of business activities on the Torun map, below highlight in red are relevant business activity to give potential locations to open a yoga studio or offer yoga.



Identify the related business activity as indicated by red dots (hotels)



Density of business activity by heatmap

Note that now we have obtained a visualization of our data and assign meaning relating to business activities.

We get an idea of the overall distribution of business activities around the Torun area.

We also visualize the business related venues of hotels or any potential yoga related venues to open our business nearby.

Upon Step 2, we refine our business decision.

Step 3: Refine Search Down to Specific Requirements

We want to look at possible business venues location to satisfy more specific requirements. For example

1. Location that has a minimal distances from the hotels, which is potential relevant business venue.
2. Exclude bars, pubs or clubs, or see if possible to stay as far away from these alcoholic areas as possible

Our code eliminates the venues that are hotels or bars & pubs themselves.

```
old_town_sort = old_town_sort.sort_values(['Distance to Hotels', 'Distance to alcohol'],
                                          ascending=[True, False])
old_town_sort
```

	name	categories	lat	lng	Distance to Hotels	Distance to alcohol
71	Vistula Apartments	Boarding House	53.012137	18.616955	0.000000	0.013960
16	Karczma Spichrz	Restaurant	53.008763	18.608807	0.000086	0.006276
17	Żywe Muzeum Piernika	Museum and Gallery	53.008669	18.604709	0.000225	0.007688
25	Stare Metropolis	Restaurant	53.011169	18.607271	0.000328	0.004749
22	Karczma Gęsia Szyja	Diner	53.011158	18.607333	0.000372	0.004687
12	Luizjana	Restaurant	53.009148	18.608837	0.000417	0.006100
69	Centrum Gier Feniks	Hobby Shop	53.012693	18.611880	0.000453	0.009366
27	Beza-krówka - naturalne lody rzemieślnicze	Ice Cream Shop	53.011214	18.606715	0.000459	0.005306
56	Teatr im. Wilama Horzycy	Theater	53.012275	18.603031	0.000553	0.009070
64	Plac Teatralny	Plaza	53.012859	18.603353	0.000575	0.008852

Finally, we make reference to different existing venues with regard to their

1. Closest distance to any hotels
2. Furthest distance to alcoholic places

as shown in the above table for making our final choices of references to the venue location as to which areas in the Old Town of Torun is in the best potential and most favorable to open up a yoga studio or offer yoga.

Results and Discussion

Our analysis shows that most lively business activities in Toruń cluster around the Old Town area.

As Torun is a medium size town with a population of around two hundred thousand, we decide to view the data on a whole for the full picture as the data size is not huge.

The most robust business activity are 1) Restaurants 2) Bar, pubs & clubs 3) Hotels

We believe related business venues or activities like 1) Hotels & Resort 2) Gym & sports center 3) Wellness centers e.g. spa & sauna are **complementary business activities** relating to yoga which will benefit our business if we decide to open a yoga studio close to these places.

Last but not the least, we refine our requirements specifically to 1) close to hotels which helps our business, and 2) further away from alcoholic places e.g. bars & clubs which conflicts with or may harm our business in terms of the business nature.

Finally, we will expect to find an optimal location for our yoga studio by referencing potential venue candidates with regard to existing favorable businesses venues in the Old Town. We expect to look for possible venues vacancy near these candidates, which we believe to be beneficial to the nature of our business.

Conclusion

In conclusion, the purpose of this capstone project is to showcase the ability and potential of wielding data science to give solid support and evidence in the business decision making process.

It should be well aware that when approaching different types of business decisions, the business nature, industry information, product knowledge, partners / competitors and any relevant contributing factors to the success of the business itself should be well researched and studied, so that the business problem or question is well defined, analyses go in the right directions and finally lead to good business solutions.