1. Introduction

This project is part of the IBM Data Science Professional Certificate course on coursera.org. Here I will be using Python's data science toolkit to demonstrate the creation of value by applying the learned skills. I will present a hypothetical business problem, discuss data and methodology used to solve the problem, share the analysis results and provide recommendations.

2. Business Opportunity

Vilnius - capital of Lithuania with a population of c 580 thousand, is among the greenest, safest and healthiest cities in the world. Attracting increasing number of tourists and investments in the recent years, Vilnius has become a go to place for anyone seeking to explore it's rich history, dynamic art and culture scene, museums as well as culinary delights.

Understanding the current city's growth potential, successful Italian Pizza Restaurant owner has set on a mission to open a family Pizza Restaurant in Vilnius, no further than 5-6km from the city center. To help choose the right location for a restaurant, I've been asked to conduct an analysis and provide recommendations.

3. Data

To perform the analysis, I will be using Vilnius Neighborhood data from Wikipedia. Then, I will use the coordinates associated with the Neighborhoods and gather most common venue data using FourSquare API. The resulting venue category data will be used to observe common patterns between each of the Neighborhoods:

- List of Vilnius Neighborhoods
- Coordinates of the Neighborhoods
- Top venues of the Neighborhoods

4. Methodology

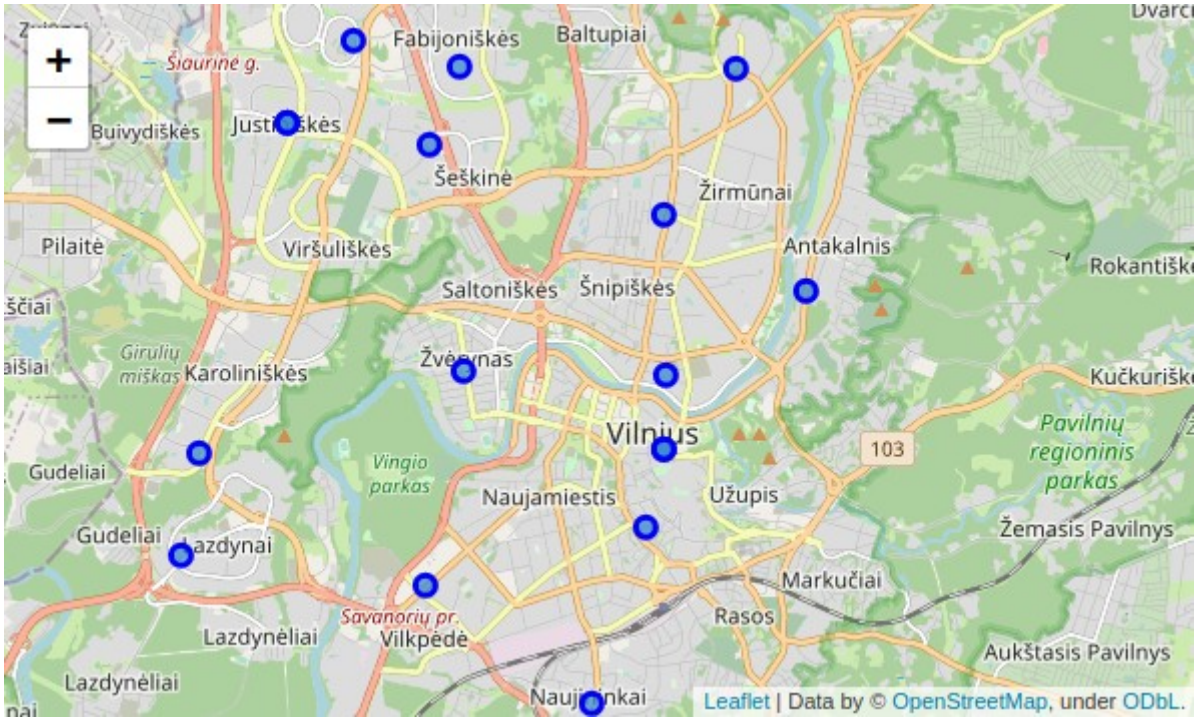
First of all, I will be gathering Neighborhood data from Wikipedia:

	Neighborhood	Latitude	Longitude
0	Verkiai	54.708707	25.284686
1	Antakalnis	54.701126	25.308957
2	Pašilaičiai	54.725942	25.231328
3	Fabijoniškės	54.723397	25.249529
4	Pilaitė	54.708126	25.175803

The Neighborhood table has Latitude and Longitude columns, thus I will not need to gather Geo-Coordinates separately. Going back to the business requirement – I need to investigate locations which are not farther then 5-6km from Vilnius city center. Thus I will filter out Neighborhoods which do not fall into the analysis scope:

	Neighborhood	Latitude	Longitude	Coordinates	DistanceFromCenter
0	Verkiai	54.708707	25.284686	(54.708707, 25.284686)	2.415601
1	Antakalnis	54.701126	25.308957	(54.701126, 25.308957)	2.441722
2	Pašilaičiai	54.725942	25.231328	(54.725942, 25.231328)	5.321808
3	Fabijoniškės	54.723397	25.249529	(54.723397, 25.249529)	4.474206
4	Justiniškės	54.717905	25.220236	(54.717905, 25.220236)	5.135968

Great. I will now use Python folium library to visualize geographic details of Vilnius and it's Neighborhoods:



Let's utilize FourSquare API to explore the Neighborhoods. Here's a list of first 5 venues for a single Neighborhood obtained via API using geographical details, using 1000m radius and a limit of 100 venues:

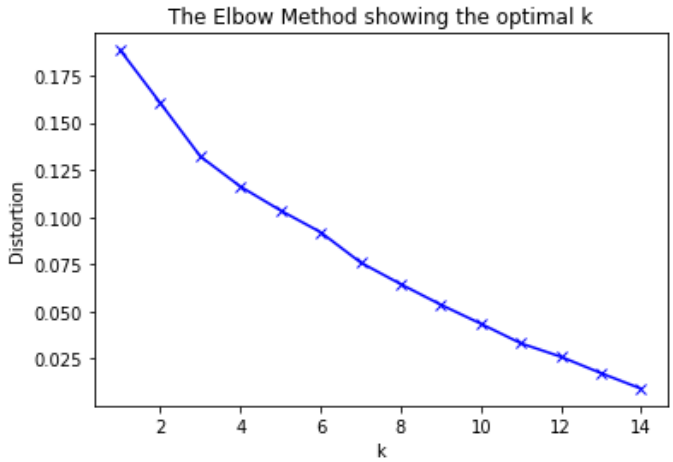
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Verkiai	54.708707	25.284686	Pizza Verde	54.707070	25.285050	Pizza Place
1	Verkiai	54.708707	25.284686	GymPlius	54.703436	25.282497	Gym
2	Verkiai	54.708707	25.284686	PANK'olis	54.711114	25.292318	Buffet
3	Verkiai	54.708707	25.284686	Alaus studija	54.707954	25.295803	Bar
4	Verkiai	54.708707	25.284686	Ulonų turgelis	54.709531	25.292245	Farmers Market

There are 166 unique Venue Categories in total. I'll group them together to see which one's are the most common in different Neighborhoods:

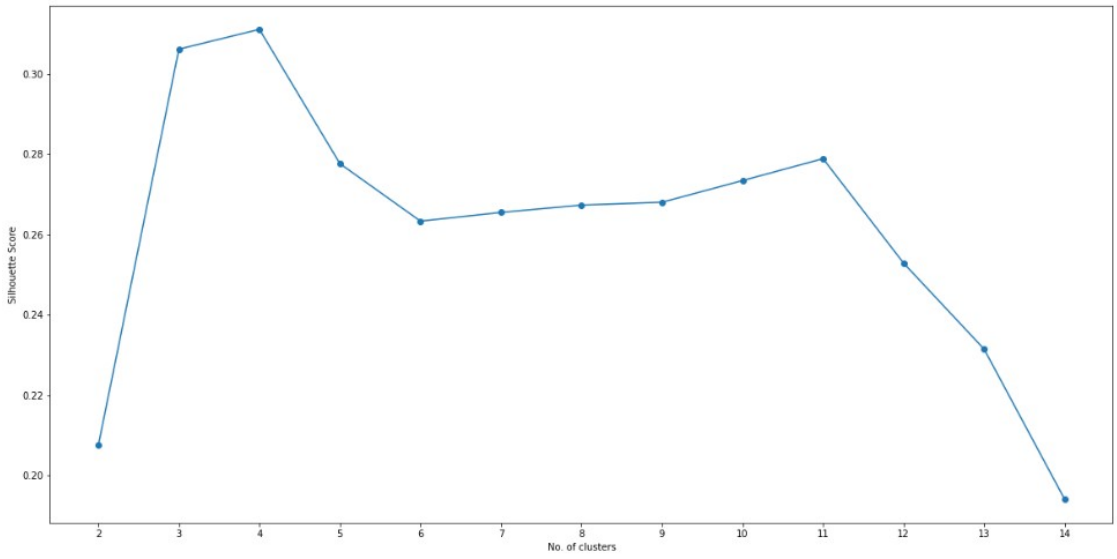
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Antakalnis	Park	Gym / Fitness Center	Electronics Store	Coffee Shop	Pizza Place	Gym	Shopping Mall	Bus Line	Bus Station	Bus Stop
1	Fabijoniškės	Pizza Place	Fast Food Restaurant	Electronics Store	Grocery Store	Food & Drink Shop	Gym / Fitness Center	Furniture / Home Store	Market	Flower Shop	Flea Market
2	Justiniškės	Convenience Store	Supermarket	Furniture / Home Store	Shopping Mall	Sporting Goods Shop	Flea Market	Market	Cosmetics Shop	Bus Station	Pizza Place
3	Karoliniškės	Grocery Store	Pool	Restaurant	Fast Food Restaurant	Sports Club	TV Station	Office	Food Truck	Bus Line	Bus Station
4	Lazdynai	Grocery Store	Bus Stop	Pizza Place	Caucasian Restaurant	Gym	Pharmacy	Fast Food Restaurant	Park	Bus Station	Food Court
5	Naujamiestis	Coffee Shop	Bar	Museum	Restaurant	Cocktail Bar	Park	Hotel	Bagel Shop	Beer Bar	Dessert Shop

As you can see, there are some common Venue Categories in the Neighborhoods, thus I will use unsupervised learning algorithm – K-means, to cluster the Neighborhoods.

Before running K-Means algorithm, I need to decide many clusters should I use to segment the data. To help me do that, I will use elbow method, which will show the most optimal cluster (k) number:



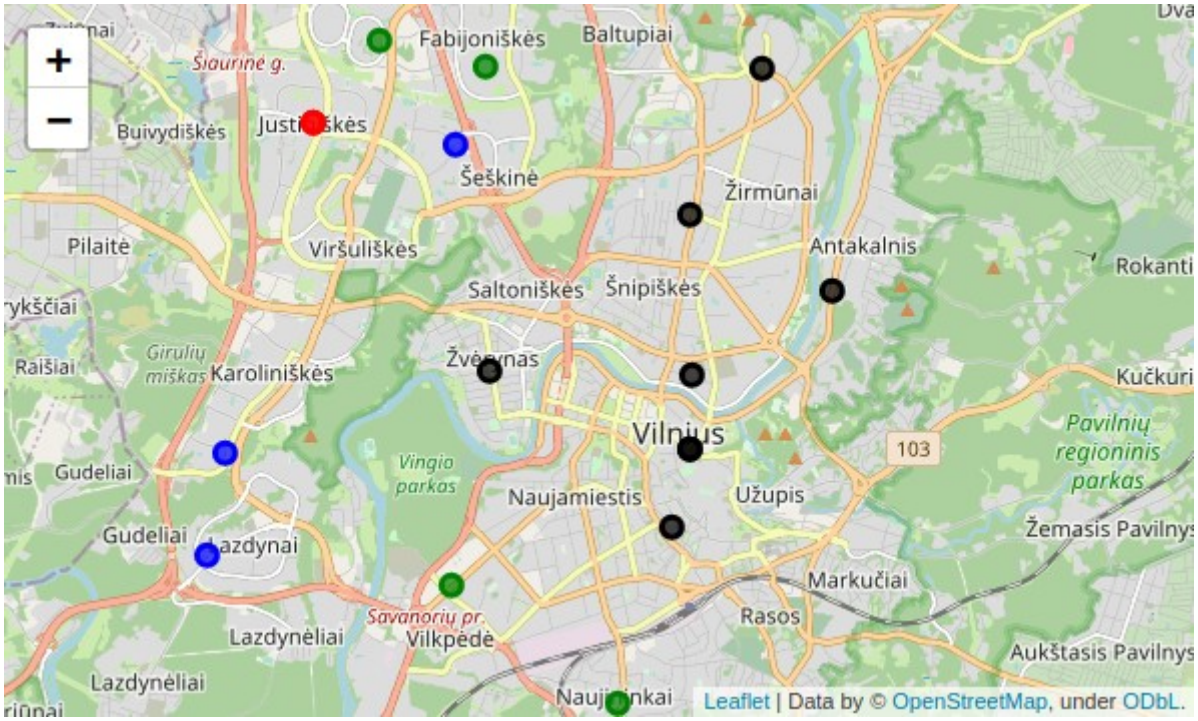
Looking at the model is not clear if 4 or 7 clusters (k) should be used, thus I will perform Silhouette analysis to help me further validate the optimal clusters (k) number:



Silhouette analysis indicates 4 clusters. Revisiting elbow method again, I can see that it breaks at the similar point, thus I will use 4 clusters in the algorithm.

5. Results

Let’s review the K-Means algorithm results by visualizing output using folium library:



Let’s examine each cluster:

Cluster 1:

	Neighborhood	DistanceFromCenter	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
4	Justiniškės	5.135968	0	Convenience Store	Supermarket	Furniture / Home Store	Shopping Mall	Sporting Goods Shop	Flea Market	Market	Cosmetics Shop	Bus Station	Pizza Place
5	Viršuliškės	5.133829	0	Convenience Store	Supermarket	Furniture / Home Store	Shopping Mall	Sporting Goods Shop	Flea Market	Market	Cosmetics Shop	Bus Station	Pizza Place

Cluster 2:

Neighborhood	DistanceFromCenter	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
Verkiai	2.415601	1	Coffee Shop	Pizza Place	Gym	Bakery	Burger Joint	Supermarket	Clothing Store	Grocery Store	Gastropub	Sushi Restaurant
Antakalnis	2.441722	1	Park	Gym / Fitness Center	Electronics Store	Coffee Shop	Pizza Place	Gym	Shopping Mall	Bus Line	Bus Station	Bus Stop
Šnipiškės	0.726413	1	Hotel	Coffee Shop	Restaurant	Plaza	Café	Italian Restaurant	Theater	Brewery	Historic Site	Dance Studio
Žirmūnai	4.168681	1	Park	Coffee Shop	Convenience Store	Beach	Hotel	Diner	Gym / Fitness Center	Department Store	Shopping Mall	Shoe Store
Žvėrynas	2.021345	1	Coffee Shop	Clothing Store	Eastern European Restaurant	Bakery	Cosmetics Shop	Restaurant	Sushi Restaurant	Chinese Restaurant	Café	Italian Restaurant
Naujamiestis	0.373576	1	Coffee Shop	Bar	Museum	Restaurant	Cocktail Bar	Park	Hotel	Bagel Shop	Beer Bar	Dessert Shop
Senamiestis	0.372794	1	Coffee Shop	Bar	Museum	Restaurant	Cocktail Bar	Park	Hotel	Bagel Shop	Beer Bar	Dessert Shop
Rasos	1.063383	1	Hotel	Coffee Shop	Bar	Cocktail Bar	Wine Bar	Museum	Restaurant	Dessert Shop	Beer Bar	Pub

Cluster 3:

	Neighborhood	DistanceFromCenter	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	Šeškinė	3.897302	2	Grocery Store	Pizza Place	Electronics Store	Bus Station	Fast Food Restaurant	Shoe Store	Food & Drink Shop	Flea Market	Motel	Sporting Goods Shop
9	Karoliniškės	4.813054	2	Grocery Store	Pool	Restaurant	Fast Food Restaurant	Sports Club	TV Station	Office	Food Truck	Bus Line	Bus Station
11	Lazdynai	5.192434	2	Grocery Store	Bus Stop	Pizza Place	Caucasian Restaurant	Gym	Pharmacy	Fast Food Restaurant	Park	Bus Station	Food Court

Cluster 4:

Neighborhood	DistanceFromCenter	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
Pašilaičiai	5.321808	3	Pizza Place	Electronics Store	Furniture / Home Store	Fast Food Restaurant	Pharmacy	Bed & Breakfast	Shoe Store	Soccer Field	Pet Store	Creperie
Fabijoniškės	4.474206	3	Pizza Place	Fast Food Restaurant	Electronics Store	Grocery Store	Food & Drink Shop	Gym / Fitness Center	Furniture / Home Store	Market	Flower Shop	Flea Market
Vilkpėdė	2.861739	3	Gym	Grocery Store	Hotel	American Restaurant	Convenience Store	Shoe Store	Cafeteria	Café	Gym / Fitness Center	Pharmacy
Naujininkai	3.018341	3	Fast Food Restaurant	Pizza Place	Grocery Store	Supermarket	Paintball Field	Clothing Store	Coffee Shop	Motel	Eastern European Restaurant	Electronics Store

6. Discussion

I will evaluate algorithm results/clusters based on a 3 factors – Distance to City Center, Variety of Cuisines/Restaurants, Number of existing Pizza Places. This will help us to address the primary concern of the stakeholder (how far is the cluster from the City Center), understand the demand for non traditional cuisines as well as evaluate the potential competition between Pizza Places in the clusters.

- Cluster 1, on average, seems to be located ~5.1km away from the City Center, has no presence of restaurants and has relatively few Pizza Places. Comparing to other Clusters, on average, Cluster 1 seems to be farthest away from the City Center. Results also indicates that restaurants and Pizza Places are not so popular in the Cluster thus it may require time and larger marketing budget to grow the demand for a Pizza Restaurant.

- Cluster 2, on average, seems to be located ~1.9km away from the City Center, has good presence of international restaurants and relatively few Pizza Places. Comparing to other Clusters, on average, Cluster 2 seems to be located closest to the City Center. Results indicate that Cluster 2 may be a good option for opening a Pizza Restaurant, since the Cluster has a good variety of restaurants and competition between Pizza Places is relatively low.

- Cluster 3, on average, seems to be located ~4.6km away from the City Center, has low presence of restaurants and has relatively good presence of Pizza Places. Comparing to other Clusters, on average, Cluster 3 seems to take third place in terms of the distance to the City Center. Results also indicates that the restaurant market is mostly dominated by the fast food players, whilst demand for Pizza Places seems to be booming. If the importance of distance to the City Center is re-prioritized, Cluster 3 may be a good option for a Pizza Restaurant. However the food menu may need to be adopted to a more conservative/local taste.

- Cluster 4, on average, seems to be located ~3.9km away from the City center, has low presence of restaurants but high presence of Pizza Places. Comparing to other Clusters, on average, Cluster 4 seems to be located at the second place in terms of the distance to the City Center. Restaurant market is mostly dominated by the fast food players, whilst Pizza Places seems to be one of the top common venues in the Cluster. This indicates that a new Pizza Restaurant in Cluster 4 may face a high competition, thus it may require time and larger marketing budget to build a customer base.

7. Conclusion

In conclusion, few of the Clusters seems to be a good option for opening a new restaurant, whilst others may have a better potential for an expansion. Depending on the marketing budget, eagerness to compete and versatility of the restaurant menu, I would recommend the following:

- Cluster 1 – to be considered for an expansion rather then the first location.

- Cluster 2 – is a good option for a Pizza Restaurant where food menu is predominantly international.

- Cluster 3 – is a good option for a Pizza Restaurant, if the importance of distance to City Center is re-prioritized and food menu is adopted to a more local/conservative style.

- Cluster 4 – depending on willingness to compete and available marketing budget, Cluster 4 may be a considerable option for opening a Pizza Restaurant. However, the food menu may may need to be adjusted to a more local/conservative style.

This concludes the analysis.

8. References

- Tourism in Vilnius
- Investments in Vilnius
- Coursera
- FourSquare
- Wikipedia