

IBM Applied Data Science Capstone

Opening a restaurant in Norway

Alexander Podolski



ABSTRACT

This study utilizes the K-means clustering algorithm to derive the best location for a restaurant in Norway. The result point to Bergen due to its population and popularity of restaurants in the city.

1. Introduction

After several decades of working hard, manual labor jobs, longtime buddies Chris Bacon and Oliver Eggs are looking into opening a restaurant in Norway, so that they may retire in relaxation with a carefree life. Even before beginning promotional activities, there is an important decision that must be taken with caution, upon which lies a large part of the success of a restaurant: the choice of location. The clients requested our data science team to assist them in their decision. The task was to conduct an analysis and provide a recommendation.

The research question is:

"What is the best location to open a restaurant in Norway?"

The focus of this analysis is on finding the best city. Finding the best area or neighborhood within the target city is outside the scope of this study. The results of this research are useful to anyone interested in setting up shop in Norway.

2. Data

Clustering is a common method used by businesses in order to find opportunities within a given market. Also referred to as market segmentation, it allows for grouping of objects based on unique characteristics. In this case, in order to find the best location for a restaurant, cities were segmented according to venues.

The required geospatial information was acquired using Foursquare APIs, which provide information about different cities such as venue names and categories. Data containing city names and their coordinates were obtained from "<https://simplemaps.com/data/no-cities>".

3. Methodology

The analysis was conducted using Python programming language. Data cleaning involved removing unnecessary columns and checking for empty or misleading entries. The dataframe was grouped according to population, as shown in table 1. Considering the nature of the study, no descriptive statistics such as checking the spread and central tendency were necessary. Areas included in the dataset were plotted on the map using Folium, a Python library that creates Leaflet maps (figure 1).

	City	Latitude	Longitude	Population
0	Oslo	59.916667	10.750000	835000.0
1	Bergen	60.391111	5.324722	213585.0
2	Stavanger	58.966667	5.750000	173132.0
3	Trondheim	63.416667	10.416667	147139.0
4	Drammen	59.742500	10.211667	90722.0

Table 1. The five most populated cities.



Figure 1. A map of the populated areas of Norway.

Next, Foursquare APIs were used to gather all the venues in all the areas provided by “simplemaps.com” (table 2).

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Oslo	59.916667	10.75	Torggata Botaniske	59.916070	10.752173	Cocktail Bar
1	Oslo	59.916667	10.75	BRUS Bar	59.916625	10.753520	Beer Bar
2	Oslo	59.916667	10.75	Crowbar & Bryggeri	59.917204	10.753523	Brewery
3	Oslo	59.916667	10.75	Arakataka	59.916381	10.750594	Scandinavian Restaurant
4	Oslo	59.916667	10.75	My Ugly Baby	59.915015	10.750212	Coffee Shop

Table 2. A sample of venue information gathered using Foursquare API.

K-means clustering was chosen as the method for segmentation. In order to proceed with the analysis, venue categories needed to be converted into dummy variables, which was accomplished using One Hot Encoding (table 3).

	City	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	Automotive Shop	BBQ Joint	Bakery	Bar	Beach	Bed & Breakfast	Beer Bar	Beer Garden	Beer Store	Bistr
0	Oslo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Oslo	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
2	Oslo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Oslo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Oslo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3. Venues as dummy variables.

Next, the frequency of venues in each city was established (table 4) .

	City	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	Akkerhaugen	Hotel	Zoo	Fishing Spot	Fruit & Vegetable Store	French Restaurant	Forest	Food Truck	Food Court	Food & Drink Shop	Food
1	Alstad	Café	Market	Zoo	Fishing Store	Fruit & Vegetable Store	French Restaurant	Forest	Food Truck	Food Court	Food & Drink Shop
2	Alta	Hotel	Restaurant	Harbor / Marina	Zoo	Fishing Spot	French Restaurant	Forest	Food Truck	Food Court	Food & Drink Shop
3	Andenes	Boat or Ferry	Zoo	Gas Station	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest	Food Truck	Food Court	Food & Drink Shop
4	Arendal	Café	Hotel	Restaurant	Grocery Store	Bar	Burger Joint	Sushi Restaurant	Tapas Restaurant	Chinese Restaurant	Bookstore

Table 4. Cities and venues grouped.

Lastly, cities are given a cluster label and coordinates are added in order to map the results.

	City	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	Oslo	59.916667	10.750000	2.0	Bar	Coffee Shop	Beer Bar	Hotel	French Restaurant	Cocktail Bar	Japanese Restaurant	Burrito Place	Bus Station	Speakeasy
1	Bergen	60.391111	5.324722	2.0	Bar	Restaurant	Café	Pub	Hotel	Scandinavian Restaurant	Bakery	Coffee Shop	Sushi Restaurant	American Restaurant
2	Stavanger	58.966667	5.750000	2.0	Grocery Store	Italian Restaurant	Diner	Food Court	Pizza Place	Watch Shop	Market	Coffee Shop	Fishing Store	Forest
3	Trondheim	63.416667	10.416667	2.0	Pizza Place	Grocery Store	Train Station	Zoo	Fishing Store	French Restaurant	Forest	Food Truck	Food Court	Food & Drink Shop
4	Drammen	59.742500	10.211667	2.0	American Restaurant	Bar	Café	Tapas Restaurant	Gas Station	Movie Theater	Chinese Restaurant	Pizza Place	Sports Bar	Lounge

Table 5. Final dataframe.

Now the data may be visualized and grouped by individual clusters.

4. Results

Cities were grouped into 4 clusters according to venue categories. Figure 2 visualizes the clusters on a map of the entire country, while tables 6 through 9 show each cluster and provide us with the insight needed to answer the research question.



Figure 2. K-means clusters visualized.

	City	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	Narvik	Hotel		Bar	Café	Ski Area	Zoo	Food	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant
22	Alta	Hotel	Restaurant	Harbor / Marina		Zoo	Food	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest
40	Rosendal	Park		Hotel	Food	Gas Station	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest
111	Brattvåg	Hotel	Pizza Place	Zoo		Food	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest

Table 6. Cluster 1.

	City	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
201	Fauske	Bakery		Zoo	Food	Gas Station	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest
219	Rødberg	Bakery		Zoo	Food	Gas Station	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest
293	Tonstad	Bakery		Zoo	Food	Gas Station	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest
388	Vinje	Bakery	History Museum		Zoo	Food	Garden Center	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant	Forest

Table 7. Cluster 2.

	City	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
2	Stavanger	Grocery Store		Diner	Watch Shop	Market	Pizza Place	Coffee Shop	Italian Restaurant	Auto Workshop	Food Service
3	Trondheim	Grocery Store		Pizza Place	Train Station		Zoo	Gas Station	Furniture / Home Store	Fruit & Vegetable Store	French Restaurant
7	Tomasjord	Grocery Store		Boat or Ferry	Food Truck		Zoo	Food & Drink Shop	Garden Center	Furniture / Home Store	Fruit & Vegetable Store
21	Molde	Sporting Goods Shop		Grocery Store	Shopping Mall		Zoo	Flower Shop	Fruit & Vegetable Store	French Restaurant	Forest

Table 8. Cluster 3.

	City	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	Oslo		Bar	Beer Bar	Coffee Shop	Japanese Restaurant	Hotel	French Restaurant	Cocktail Bar	Mini Golf	Gym / Fitness Center
1	Bergen		Bar	Restaurant	Café		Hotel	Scandinavian Restaurant	Pub	Bakery	Sushi Restaurant
4	Drammen	American Restaurant		Plaza	Pizza Place	Movie Theater	Tapas Restaurant	Coffee Shop	Seafood Restaurant	Lounge	Bar
6	Kristiansand	Brewery		Lake	Music Venue	Café	Pub	Farm	Food Service	Garden Center	Furniture / Home Store

Table 9. Cluster 4.

5. Discussion

When evaluating a location, the first thing to consider is the population of a city. In order to attract as many people as possible, it is important to be positioned in an area of high pedestrian traffic. A large number of potential clients passing by your restaurant will reduce the cost of marketing. Furthermore, gauging the popularity of restaurants within a city is perhaps as important. One may question the attractiveness of such business if restaurants are not even in the top 10 of the most popular venues within an area. Once a city has been chosen, further analysis into surrounding areas and neighborhoods will be useful, as the clientele can vary depending on whether the business is in a popular tourist destination, in a residential area or surrounded by offices.

Cluster 4 features the two most populated cities of Norway. At first glance, Oslo seemed like the best location for a restaurant, being the most populated city and the capital of Norway. However, the results show that restaurants are not even in the top 10 of the most popular venues in Oslo. Bergen, however, ranks them as number 2. Bergen is the second most populated city of Norway, and based on this analysis is the best location for a restaurant.

While providing a solid foundation for decision-making, it is important to note that such decision cannot be based on only one analysis. There are many other variables to consider. For example, while it is reasonable to assume that rent prices in Oslo are higher than in Bergen, one cannot know without doing proper research. Such questions, however, are beyond the scope of this study.

6. Conclusion

The purpose of this study was to find the best location to open a restaurant in Norway. Using K-means clustering algorithm, it is found that Bergen is the best location, being the second most populated city of Norway and ranking restaurants as the second most popular type of venue.

Further studies may look into other factors such as rent prices and cost of living in each city as well as conducting an in-depth analyzing of local surroundings and neighborhoods.