

Clustering Canadian Ski Resorts

Abstract

This paper describes the process of using K-means clustering to create groups of Canadian Ski Resorts, using location, venue categories and characteristics of each cluster. This is intended to help resort managers or marketing teams to decide on the activities and amenities that are available, or identify the potential target market to whom they could be advertising. I have created five clusters, based on the types of venues in the vicinity.

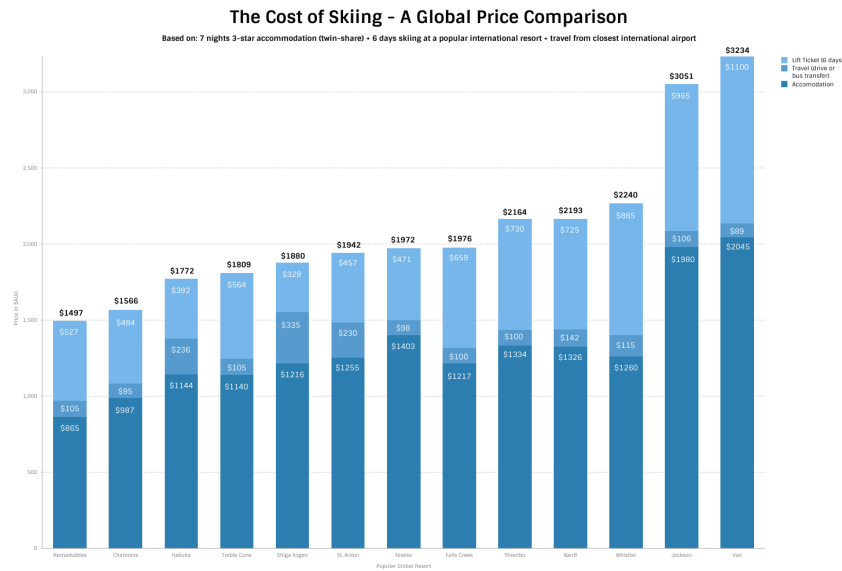
For example, it could help managers identify that their resort is similar to others who offer their customers a coffee shop, and prompt them to open one too. Or, marketing teams could understand their audience would be families looking for long-term stays, and advertise focused on these specific groups.

1. Introduction

In this notebook, I will be using Python and Machine Learning to analyse and segment Canadian Ski Resorts. I thought it would be interesting to locate different resorts, then use the Foursquare API to analyse bars, hotels and restaurants in the area to segment resorts. Both Resort Managers and skiers would be able to view resorts and their respective clusters, and make decisions.

1.1 Background

The demand for snowsports is at an all time high. More people than ever are skiing, snowboarding and exploring the wonderful world of winter sports. However, the cost of a ski holiday has also majorly increased. People are paying more for their holidays, and as such, demand the best possible experience.



As we can see above, the total cost of a ski holiday at Vail could set you back on average over \$3200. So, ski resorts must cater to this increased demand and deliver quality.

1.2 Problem

With the added pressure of climate change on the amount of snow, ski resorts are trying to improve customer experience within resorts. In addition to skiing, they are offering a range of other activities and facilities.

Their ultimate goals are to :

1. Attract as many tourists as possible through targeted marketing
2. Enhance customer experience
3. Increase the revenue per customer within the resort

Therefore, ski resort managers are looking to enhance customer experience, improve their marketing whilst minimising costs. This is the problem we are trying to solve using clustering. If resorts can identify their niche, or relevant competitors, they can gain insights into their marketable strengths or use ideas from similar resorts.

2. Data

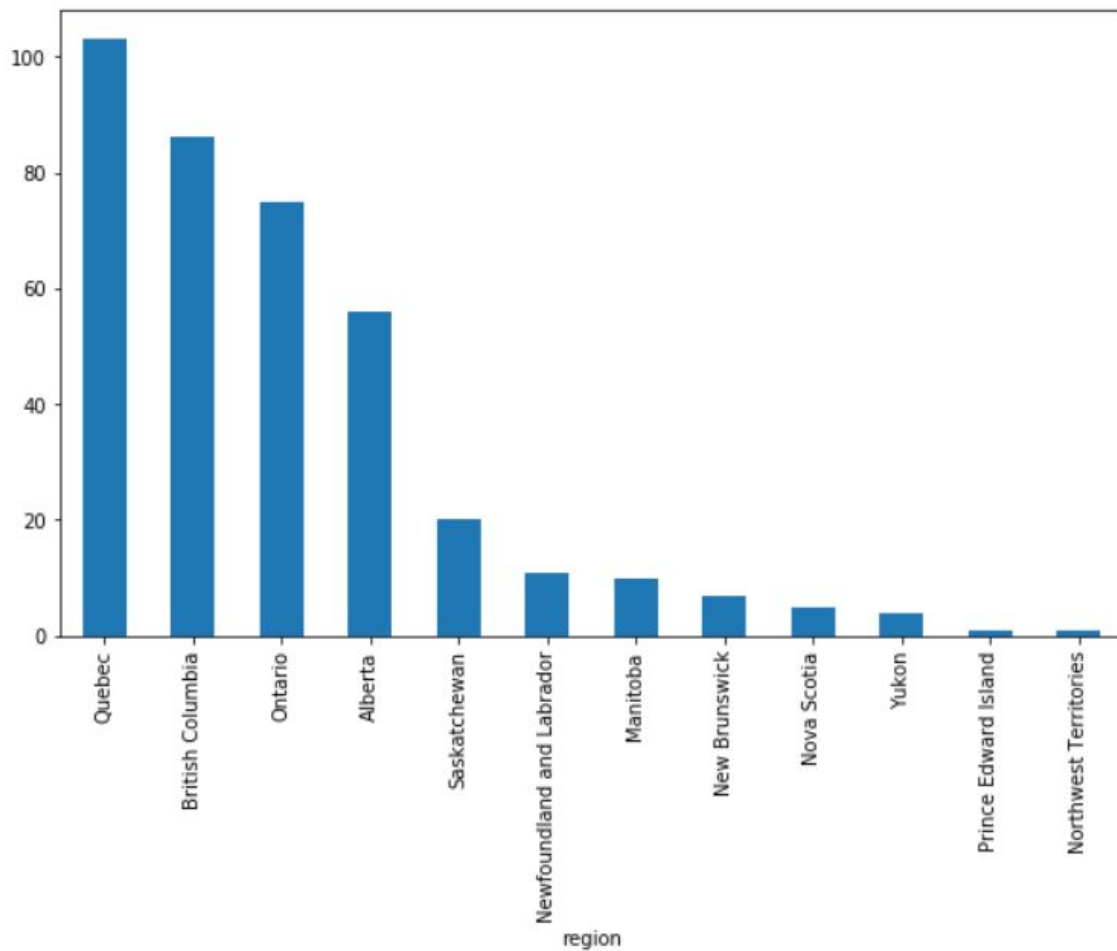
The data I will be using is a combination of the Foursquare API, and the OpenStreetMapski resort data.

2.1 Data Gathering

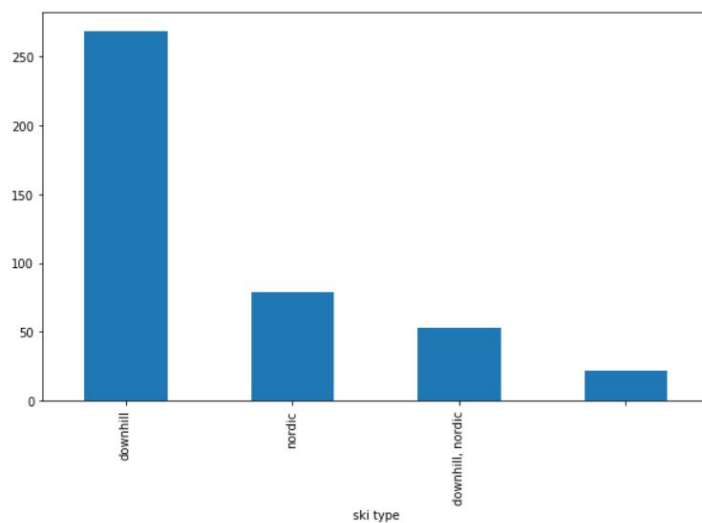
Data will be gathered firstly from OpenSkiMap, which gives the names and regions of each ski resort. Restricting this to Canada, we can then view the data for Canadian Ski Resorts and the type of skiing. This will then be geolocated, and then combined with relevant venue data from Foursquare API. This will demonstrate the types of venues surrounding our ski resorts.

region	locality	resort	ski type	latitude	longitude
Quebec	Laval	Bois Duvernay	nordic	45.55201	-73.74433
None	Mansfield-et-Pontefract	Mont Chilly	downhill	45.84979	-76.72708
Quebec	Val-David	Parc régional de Val-David - Val-Morin	nordic	46.03258	-74.20374
None	None	None	downhill	60.10867	-113.64258
Saskatchewan	None	Prairie West Terminal Snow Park (Twin Towers)	downhill	60.10867	-113.64258

As above, this is the ski resort data we will be working with.



Here's an overview of the number of resorts in each province. According to the chart above, Quebec has the largest number of resorts, closely followed by British Columbia.



We can also see an overview of the types of ski resorts in Canada. The majority of resorts have downhill skiing, followed by nordic skiing. Around 50 offer both downhill and nordic skiing.

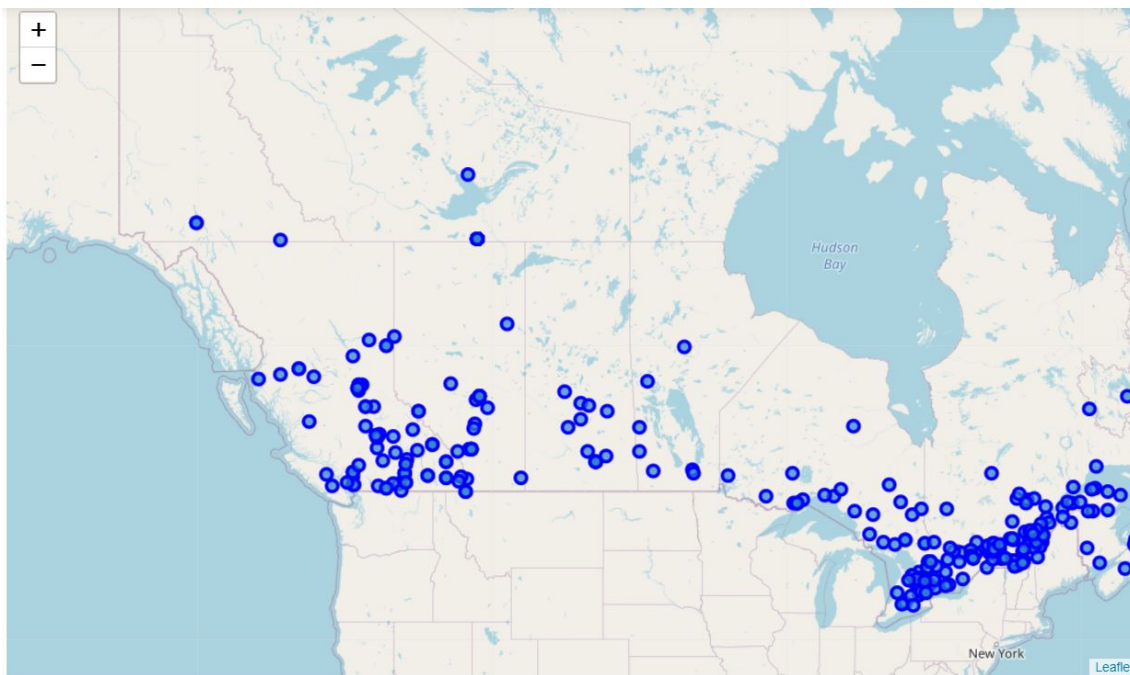
2.2 Solving the Problem

With our ski resort and venue data, we are able to categorise resorts based on the amenities they offer to guests. This will allow us to cluster and segment resorts, so resort managers can compare their resort with similar resorts to understand potential revenue building strategies, and more finely tune their marketing based on their target market.

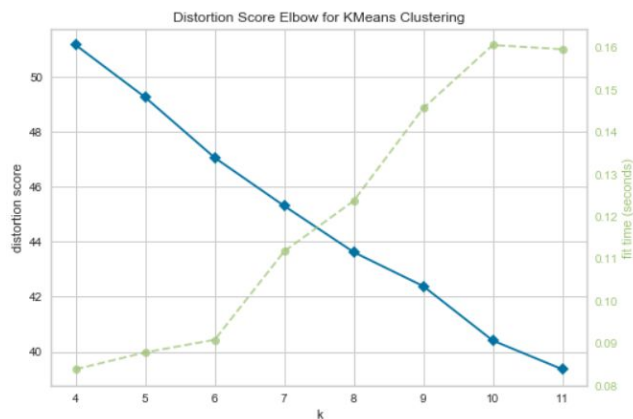
3. Methodology

3.1 Geolocating Data

Using the Foursquare API, I gathered the nearest venues to each ski resort, within 2KM. These venues were then categorized based on the services they offer.



After cleaning the data and preparing it for cluster analysis, I ran the Elbow Method to determine the optimal value for K.

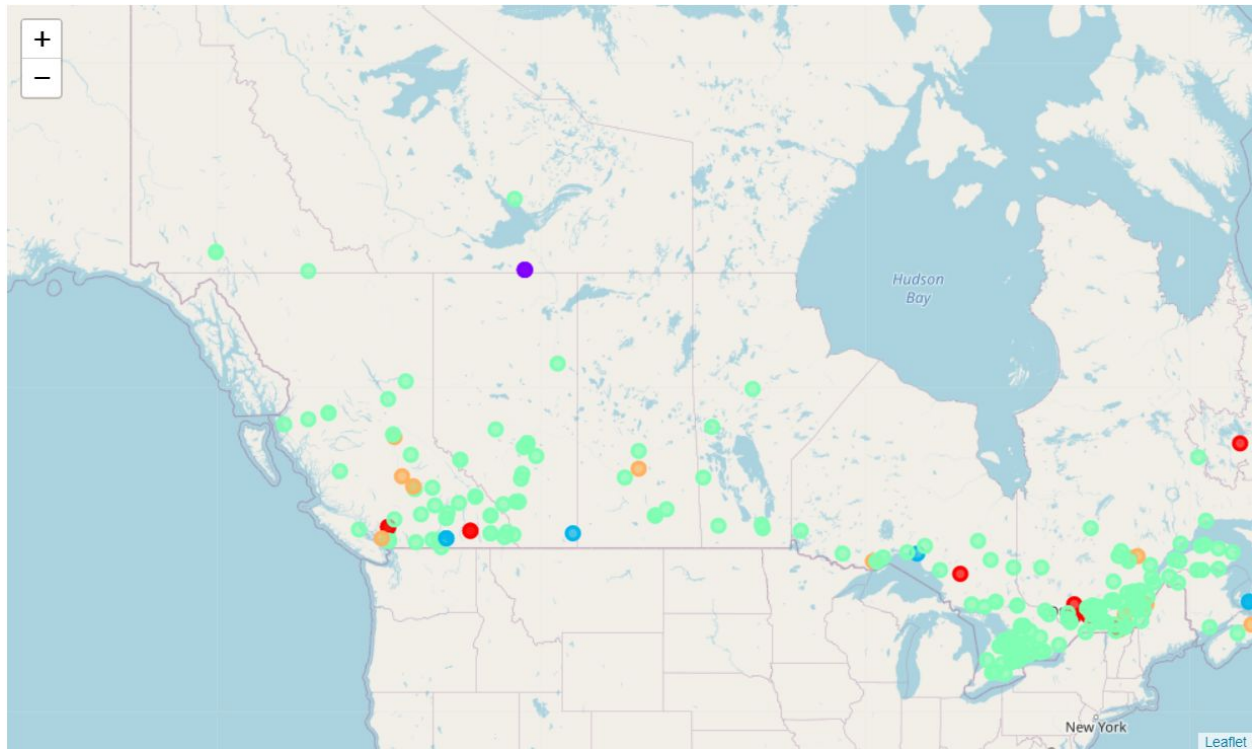


Unfortunately, the figure shows that there is no obvious elbow point. It is important to remember that the “elbow” method does not work well if the data is not very clustered. In this case, you might see a smooth curve and the optimal value of K is therefore unclear.

As K is unclear, we will set the number to 5. This is because we are using the top-10 venues, and having this number of attributes lends itself well to a K of 5.

4. Results

The algorithm produced five separate clusters. I have visualised the clusters on a map of Canada, which are grouped together according to the types of venues in the vicinity.



I then looked at the types of venues contained within each cluster, and named them accordingly.

	Cluster	Name	Count
0	0.0	Day Resorts	10
1	1.0	Geographic	3
2	2.0	Outdoor Resorts	7
3	3.0	Vaction Resorts	253
4	4.0	Local Resorts	14

5. Discussion

Given the different venues included in each cluster, I have named them based on domain knowledge of ski resorts.

The different types of resorts are as follows:

Day Resorts

All of these boast restaurants, lakes or zoos. They are most likely day-use ski resorts.

Geographic

These resorts are all located close by, and therefore have the same amenities. These are geographically similar.

Outdoor Resorts

These resorts are most outdoor-focused, with campgrounds, trails and zoos being the most frequent venues.

Day Resorts

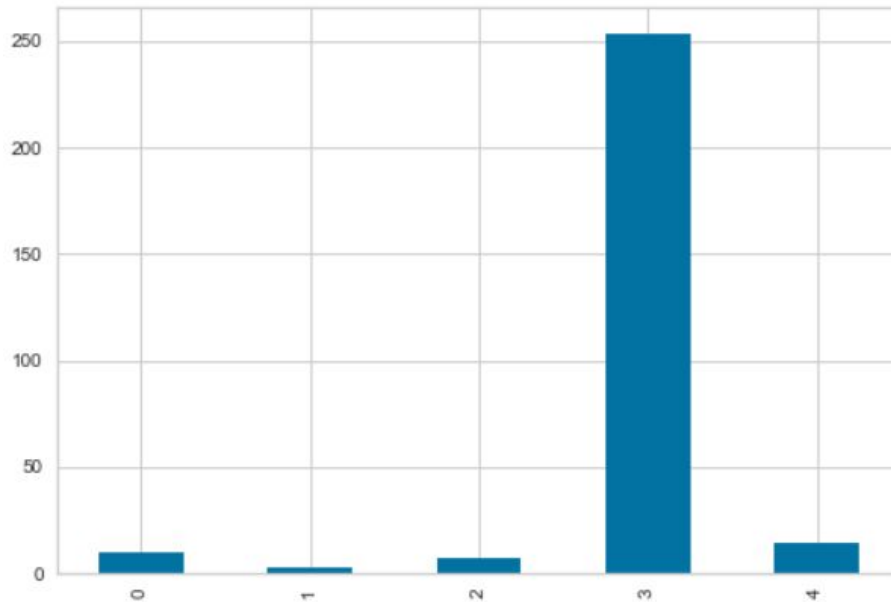
All of these boast restaurants, lakes or liquor stores. They are most likely day-use ski resorts.

Vacation Resorts

This is the majority of resorts. They have lots of amenities close by, with the majority boasting coffee shops, grocery stores and hotels. These would be classed as the vacation resorts. People choose to come here for leisure and longer amounts of time.

Local Resorts

It seems like these ski areas are located in industrial areas. They are probably unlikely to host tourists, due to the lack of amenities. These will be local resorts.



The majority of resorts are included in the Vacation Resort cluster, over 250 of the total number of resorts.

6. Conclusion

In this study, I analysed the different types of ski resort in Canada, based on the venue and geographic data collected. I identified locations, venues, venue categories and the characteristics of each cluster. I built a K-means clustering model, to predict the type of group in which a resort might be located. This model can be useful in helping resort managers and marketing teams decide on the activities and amenities that are available, as well as the potential target market to whom they could be advertising. For example, it could help managers identify that their resort is similar to others who offer their customers a coffee shop, and prompt them to open one too. Or, marketing teams could understand their audience would be families looking for long-term stays, and advertise focused on these specific groups.

To continue the analysis, I would like to further cut down the data and perhaps focus solely on the downhill resorts, and try other types of clustering algorithms to see if the results differ.