Virtual trip and planning advisor

IBM Data Science Professional Certificate: Applied Data Science Capstone

By Ihor Kulmatytskyy

May 04, 2020

Table of Contents:

Introduction

Hello, I would like to present to you a problem in the context of current related to pandemic self-isolation - all we are a bit bored while sitting at home and a lot of us are far from nature or fresh impressions. It would be nice to get some joy from photos of wild nature or nice landscapes now and make a plan to visit natural landmarks near as soon as self-isolation ends.

We would like to determine beautiful places, landmarks, and compilations to look at for now and probably to visit in the future. Kind of virtual trip and planning advisor.

This project it more social-oriented, but some business holders might be interested in it also in order to develop nearest infrastructure and attract more customers.

Data Requirements

Firstly, we need a list of interesting outdoor places, after that we need places near first ones in relatively close distance to build a compilation of such and related places.
We will use Foursquare API to obtain needed information:

- Venue name
- Venue categories
- Venue latitude, and longitude location

We will perform **trending** query to get a list popular outdoor places and **search** query to obtain lists of closest venues to first ones.

Methodology

I used data from Wikipedia and Foursquare to make a clustering of some regions of Ukrainian part of Carpathian Mountains based on outdoor places. Next data were used:

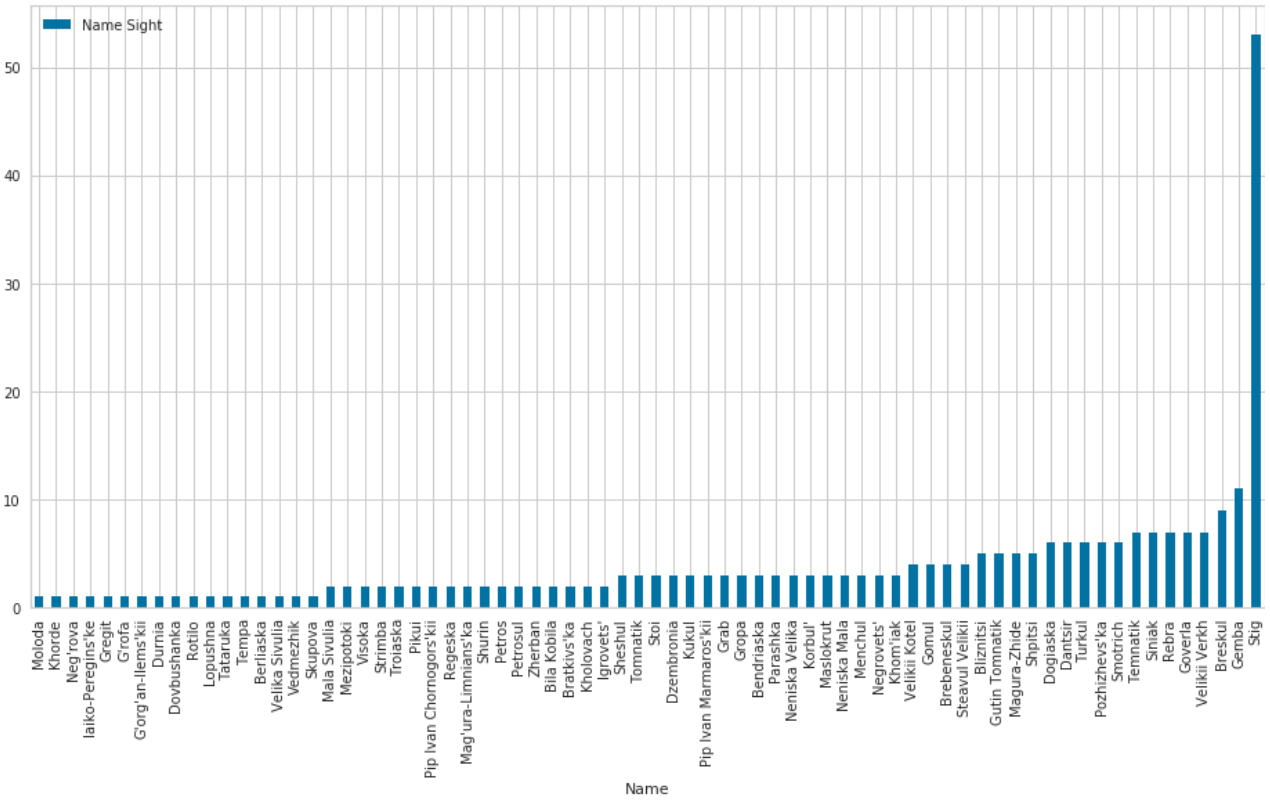
- list of mountain peaks with its locations and coordinates (122 rows),

	Name	Height	Location	id	Latitude	Longitude
0	Goverla	2061.0	Chornogora	1	48.15111	24.50000
1	Brebeneskul	2035.8	Chornogora	2	48.09833	24.58056
2	Pip Ivan Chornogors'kii	2028.5	Chornogora	3	48.04778	24.62778
3	Petros	2022.5	Chornogora	4	48.17194	24.42111
4	Gutin Tomnatik	2016.4	Chornogora	5	48.10000	24.55667

- lists of nearest to mountain peak venues with its names, categories and coordinates (189 rows)

	id	Name	Category	Latitude	Longitude
0	1	Говерла (2061 м) / Hoverla	Mountain	48.16039	24.50037
1	2	Брескул (1911 м) / Breskul	Mountain	48.15037	24.51103
2	3	Гірське Крісло	Other Great Outdoors	48.16272	24.50573
3	4	Говерлянка	Mountain	48.16615	24.50546
4	5	Пожижевська (1822 м)	Mountain	48.14430	24.52357

To look at counts of venues near each mountain peak, we may choose bar plot as follows:

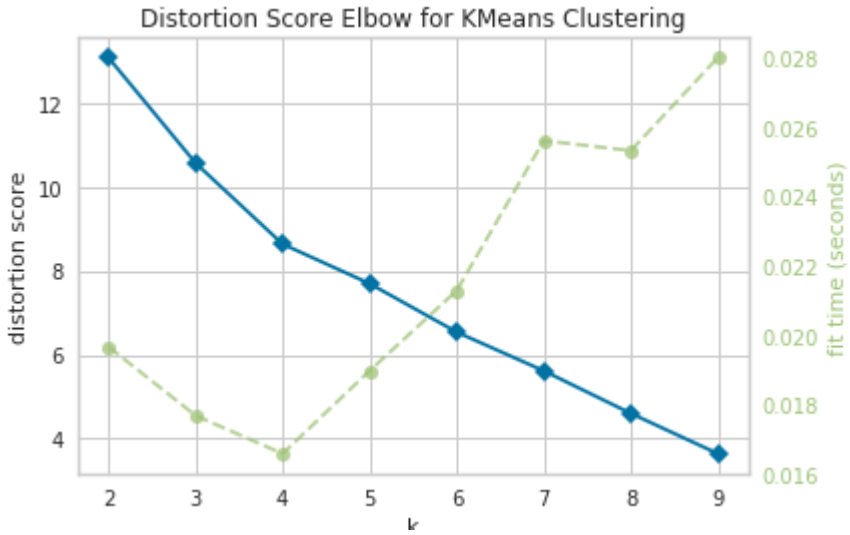


Sorting by categories of venues near each mountain peak will show to us next picture:

	Name	1st	2nd	3rd	4th	5th
3	Bliznitsi	Category_Lake	Category_Mountain	Category_Other Great Outdoors	Category_Well	Category_River
4	Bratkivs'ka	Category_Apres Ski Bar	Category_Ski Chalet	Category_River	Category_Athletics & Sports	Category_Campground
5	Brebeneskul	Category_Mountain	Category_Lake	Category_Well	Category_River	Category_Athletics & Sports
6	Breskul	Category_Mountain	Category_Scenic Lookout	Category_Well	Category_River	Category_Athletics & Sports
7	Dantsir	Category_Mountain	Category_Field	Category_Well	Category_River	Category_Athletics & Sports

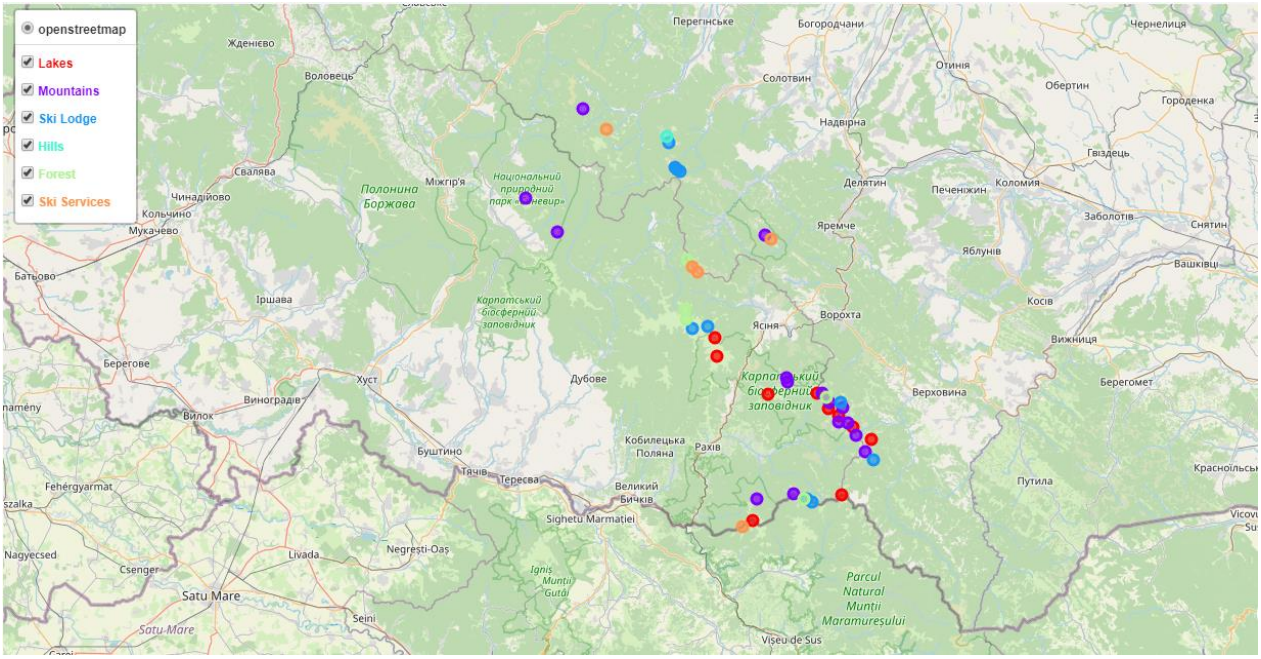
In this study I was using k-Means algorithm to make such kind of clustering: mountain peaks were chosen as item to clustering and nearest outdoor venues by categories were chosen as features to be processed with help of k-Means algorithm. As venues have common categories, the k-Means algorithm suits well and is common to perform such kind of task.

With help of elbow method was determined optimal number of clusters k = 6.



After clustering each mountain peak was assigned to a related cluster and for each cluster was picked up proper name to display its relevance from observation:

- Lakes – 10 mountain peaks
- Mountains – 15 mountain peaks
- Ski Lodges – 10 mountain peaks
- Hills – 1 mountain peak
- Forest – 5 mountain peaks
- Ski Services – 5 mountain peaks



Results and Discussion

As result of this work we have discovered 6 clusters: Lakes, Mountains, Ski Lodges, Hills, Forest, Ski Services and mark related mountain peaks in order of belonging to cluster.

More precision estimation of cluster names may be needed. Also, datasets were small, but algorithms perform well and we will need larger amounts of data to get better and more accurate results.

Conclusion

As a result, we can build different kind of web services or applications using a result of our work, to determine type of nearest outdoor places and group by these features. People can use this data to make a better plan for their trips and achieve better satisfaction, or enjoy performing searches of pictures based on this result, business owners can develop their businesses and use these data to determine proper customer's needs.

References

- [1] Wikipedia: Ukrainian Carpathian Peaks

  https://uk.wikipedia.org/wiki/%D0%92%D0%B5%D1%80%D1%88%D0%B8%D0%BD%D0%B8_%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%81%D1%8C%D0%BA%D0%B8%D1%85_%D0%9A%D0%B0%D1%80%D0%BF%D0%B0%D1%82

- [2] Foursquare API

  https://developer.foursquare.com/docs/

- [3] IBM Data Science Lab: Segmenting and Clustering Neighborhoods in New York City

https://www.coursera.org/learn/applied-data-science-capstone/ungradedLti/f0QY7/segmenting-and-clustering-neighborhoods-in-new-york-city)

- [4] Folium Map Legend

https://stackoverflow.com/a/58077924