

Final Report

Introduction

An event organizer based in southeastern Manitoba, Canada, wants to organize a mountain bike race within the region. The organizer needs to find a suitable race venue that meets the following criteria:

1. The venue must have access to an established network of biking trails
2. There must be a hotel nearby for race participants who are traveling from farther away
3. The organizer realizes that several mountain bike races already exist in the central and western regions of Manitoba. To ensure the event is viable and provides an experience distinctive from existing events, the organizer has decided to focus on a region within 200 kilometres of Steinbach, Manitoba.

The organizer is in the data-gathering stage of developing the event. She wants to know what her options are for possible race locations, so that she can weigh the costs and benefits of each option.

Gathering Data for Analysis

This project will be based on data from the following sources: a) Foursquare, which will provide location data for the southeastern Manitoba region and amenities located within it, and b) Trail Forks, which identifies the most established mountain biking trails in a region

This project will take the geographic locations of popular cycling routes from Trail Forks as one dataset. It will take the geographic locations of trails from Trail Forks as another dataset. It will plot both on a map and show their location relative to area hotels.

The results will provide the event organizer with a set of possible venues for a mountain bike race in southeastern Manitoba. From there, the organizer can consider other factors, such as proximity of restaurants that could potentially cater a post-race meal, or features of the possible trails (such as length of the trail or nature of the terrain) that could further inform the decision-making process to select an event venue.

Methodology

To identify potential event venues, the event organizer used various analytical and mapping libraries in Python to create a map of southeastern Manitoba that showed both hotels and mountain biking trails within the region.

Querying Foursquare API generated a list of hotels around Steinbach, Manitoba, a town of approximately 15,000. The query set the search radius at 200km.

A separate search of another website that uses geographic data provided the coordinates for seven trail networks within that same region. TrailForks is a website that caters to mountain bikers, helping them find cycling routes within their region and share information on trail conditions. The seven trail networks within southeastern Manitoba had varying degrees of length and difficulty. Each trail network was composed of multiple trails, and the combined distance of these trails resulted in networks ranging from 2.5km to 32km. In a similar vein, the scope of challenges presented by these networks ranged from the St. Malo trail system, which had four trails in total and they all had the easiest rating level, to the Falcon Ride Ski Slopes, a network whose 32km of trail included 21 intermediate-level trails and two advanced-level trails.

The results were used to create three Pandas dataframes, each containing latitude and longitude data:

- A dataframe for the results of the Foursquare query
- A dataframe for the trail networks that were longer than 15km
- A dataframe for the trail networks that were shorter than 15km

Breaking the trail networks into two groups based on the total length of their respective trails allowed the trails to be colour-coded accordingly on the final map.

Results

The Foursquare API returned 30 hotels that were within 200 kilometres of Steinbach, Manitoba, but some of those hotels were too far northwest of Steinbach to be relevant to event organizer's focus on southeastern Manitoba. Strangely, the exact same query returned a different number of results from Foursquare. On the overwhelming majority of the attempts to run the query, Foursquare identified the same 30 hotels. However, on one attempt, the same query – without making any changes to the search parameters – found over 100 hotels that met the search criteria. Unfortunately, the data analyst could not reproduce this result, so the final report is based on the smaller dataset of 30 hotels. Frustratingly, this dataset did not identify a single hotel within Steinbach, despite the town having at least three.

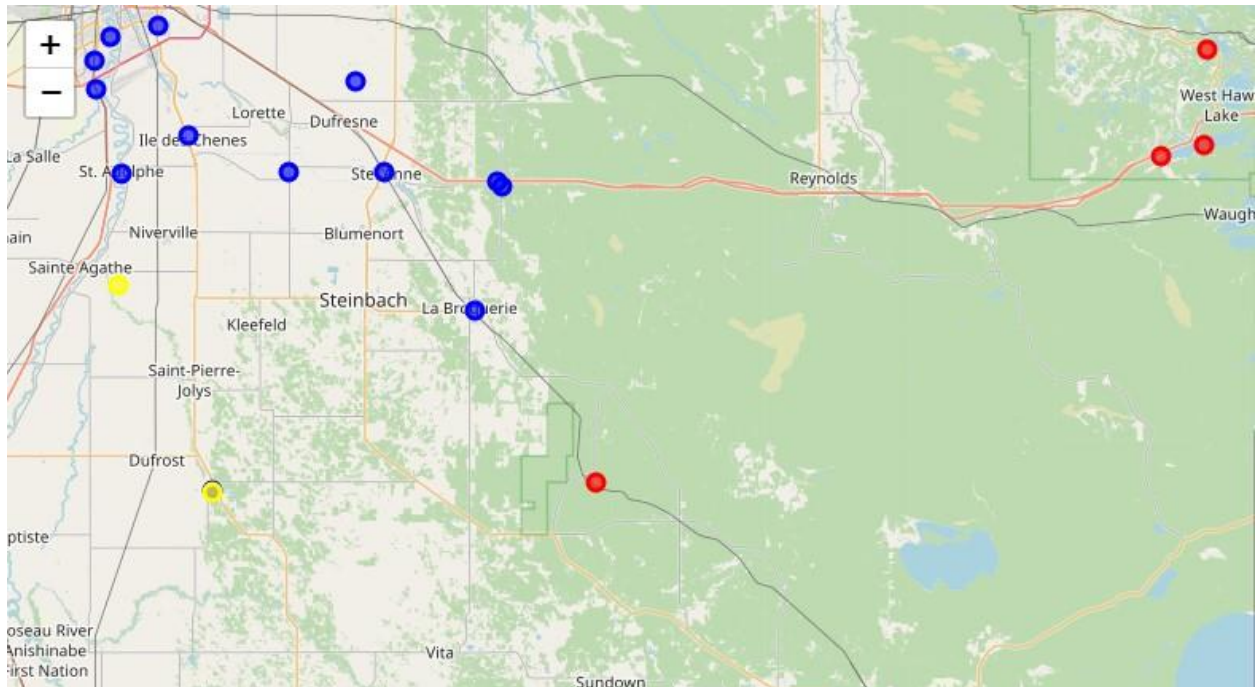


Figure 3: A map of southeastern Manitoba showing mountain bike trail networks of 15km or longer as red dots, and shorter trail networks as yellow dots. The blue dots represent hotel venues.

Plotting the seven networks of mountain biking trails in southeastern Manitoba onto the same map as the 30 hotels revealed that none of the hotels Foursquare found were located close to the longest trail networks. For the event organizer's purposes, this would suggest that her goal of hosting a mountain bike race in southeastern Manitoba somewhere that had a hotel located close to the trails is not possible. The event organizer would be required to redefine the concept behind the event, for example by:

- Organizing a "gravel grind" race that could be held on dirt and gravel roads instead of mountain bike trails in order to take place within the vicinity of a hotel, or
- Organizing a mountain bike race that would exclusively draw local riders because it would be difficult for participants from farther away to travel to and from the event within a single day, or
- Conduct further venue research to see if it may be possible to arrange for camping at or near one of the trail networks (three of the trail networks are located within a provincial park that has campsites)

Discussion

It is noteworthy that both of Coursera's examples showcasing Foursquare's value used large urban centres, specifically New York and Toronto. While Foursquare was, for example, able to find multiple options for nearby Italian restaurants in New York, it

yielded a paucity of venue results when asked for information on a region with a population of less than 200,000.

As a result, although this analysis of possible venues for a mountain bike race complied with the instructions of the Coursera capstone project, it is not a good example of the benefits of a scientifically-grounded approach to data analysis. The data sample was too small to allow for complex data analysis, and the information available from Foursquare was smaller than the number of results that would have been returned simply by searching “hotel” in Google Maps.

In fact, the results of the Foursquare API would likely have discouraged the event organizer by suggesting it was impossible to host a mountain bike race within this region on a trail network longer than 15km and with at least one hotel nearby. A comparison search of Google Maps demonstrates that is not the case, and that multiple hotel venues exist within the vicinity of the region’s longest and most challenging trail networks.

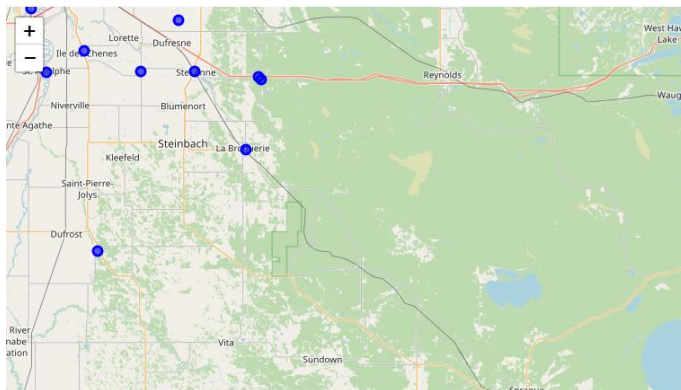


Figure 1: A map generated in Python, showing hotels identified by Foursquare.

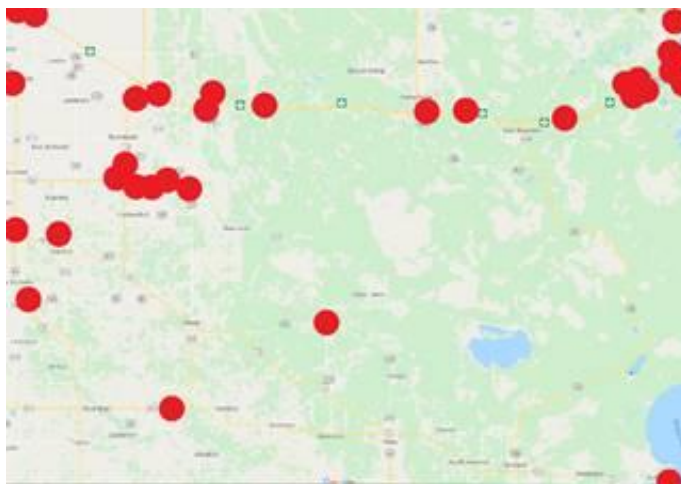


Figure 2: A screenshot from Google Maps, with highlights of hotel locations.

Conclusion

While this analysis qualifies as an example of using Foursquare to elicit insight and attempt to solve a business problem, the final results would be of limited benefit to an actual business. The variability of results from the same Foursquare query when run multiple times raises questions about the reliability of it as a tool for a novice-level data analyst. It also begs the question of whether, since this Coursera capstone course was developed in 2018, other online tools such as Google Maps have improved to an extent that, for a project of this scope, using Foursquare is no more reliable than the kind of online search a layperson would complete with freely available search tools.