**Segmenting and Clustering Mining Towns in Canada**

Andi An

# Introduction

## Background

Minerals and metals are the building blocks of our modern society, and the mining industry is one of the highest-paid industries within Canada. However, since mining jobs are only found where there is a mine, mining professionals usually must move to a mining town/city, which is usually located in remote regions. Despite the huge progress made to grant miners fair living conditions, mining towns or mining cities are not always fun.

## Problem

In this project, I will explore and categorize the condition of the mining towns in Canada. Each town will be categorized based on its amenities. This will help mining professionals to make an informed choice on their work location.

# Data acquisition and cleaning

## Data requirement

To fully explore our problem, we must gather data on the location of Mines and its near by mining town. We will also need the amenities data that is related to said town.

## Data extraction

After identifying the data required for my project, I started by trying to gather the location of each mine. Through searching the web, I was able to find an ArcGIS layer that is maintained by the Canadian open government project. Through the ArcGIS REST API, I was able to pull all 200 plus operating Mine data. This data includes quarry, metal mine, non-metal mine, and oil sand mines. In the chart below, we can see all 4 types of Mines and how they distribute throughout the country. Yellow - Metal Mine, Green - Nonmetal Mines, Red - Oil Sand Mines, Blue - Coal Mine.

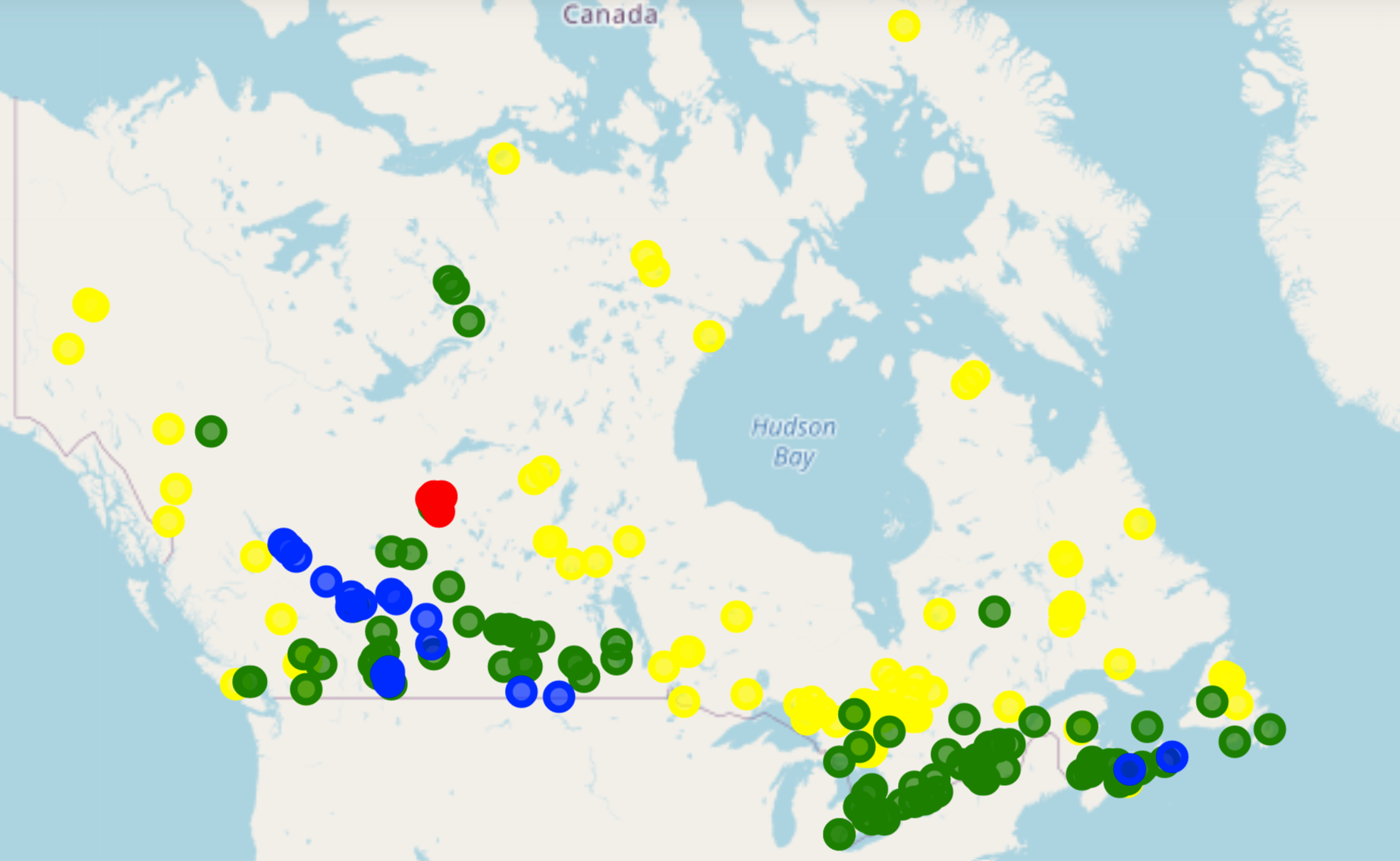


Figure : All Operating Mine in Canada

I then called the Nominatim object with the latitude and the longitude of each mine to find out the address. This result was an JSON file, which I then parsed to obtain the town name information. At this point, I have the data for the location of mines and their town.

To complete my data extraction, I started looking at the Foursquare API. Using the latitude and longitude of the town address, I made a search query with a 50km radius and 100 Venus limits. This data was then parsed to extract venue information for each mine. The results were a 200 by 300 matrix with mine name as row name and venue type as columns. The values were the number of venues for a specific mine and venue type.

# Methodology

## Explore trends through visualized the mine town venues matrix

Each mine can have a unique operating environment. Some mines can be fly-in-fly-out where there are no venues nearby. Other mines can be located near a city where it has all the normal venues a large city can provide. Most mines are likely somewhere in between these two. To better understand the situation, I generated an interactive heat map using clustergrammer2 package. This allowed for quick exploration and understanding of the data. Figures 2 and 3 show an example of an exploratory process. I was able to identify a point of interest and understand what venues were most common.

|  |  |
| --- | --- |
| Figure : Heat Map of the Data (High Zoom) | Figure :Heat Map of the Data (High Zoom) |

## Explore trends through sorting by most common venues

To characterize each mine, I looked at which top 5 places were most common in its town. This was achieved through a simple sorting and printing out the results. A very interesting trend is that coffee shop are almost always in the top 5 places. Another trends that were identified quickly was that some mines have an airport and hotels as top five most common place. This indicate a lack of venues overall and it is likely due to the fly-in-fly-out nature of those mine. Workers mostly fly home on there days off, which could explain the lack of venues.

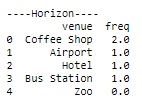


Figure : Imperial Oil Horizon Mine (Fly-in-Fly-Out)

## Explore trends through sorting by visited venues

By looking at the most visited places in each mine, we can quickly understand the nature of the mining towns. A quick exercise I did was looking at the most visited place in each mine and them summarizing them. This yields figure 5, which is a 201 by 10 matrix showing which places are the most visited.



Figure

# Results

As shown in the section 3.2, the final results of the data extraction process was the mine town venues matrix. This matrix records all 201 mine and how many venues they each have. Venues were categorized into 300 different categories.

## 4.1 Segmenting and Clustering

To better understand the different kind of mining town and look pass the noise in the data, Kmean was used to cluster the mines into 5 different clusters.

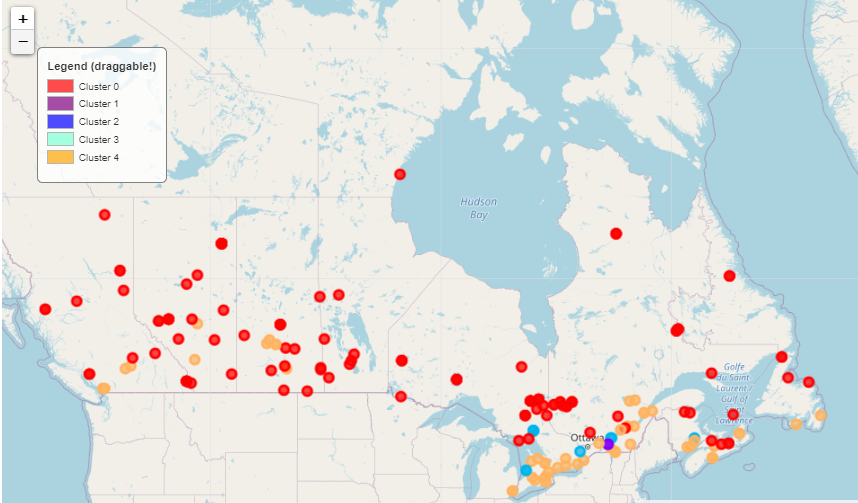


Figure : Clustered Mine Map (Canada)

# Discussion

By leveraging the k means algorithm, each of the mining towns are separated into 5 clusters. By examining each cluster, I noticed some interesting traits about each cluster.

* Cluster 0: These are towns with limited amenities and often have a busy airport.
* Cluster 1: City Center (This cluster is straight in the city of Montreal).
* Cluster 2: Town with good varieties of amenities and high number of amenities.
* Cluster 3: There are some errors in the data. This cluster is out of the map.
* Cluster 4: Town with medium varieties of amenities and medium number of amenities.

If we are grouping the mining towns by tier, I would conclude the following ranking.

1. Cluster 1
2. Cluster 2
3. Cluster 4
4. Cluster 0

# Conclusions

In this study, I analyzed the venues data for all mining towns in Canada. Through clustering, I identified 4 clusters of mining towns. They each have different level of amenities. If possible, as a mining professional might want to consider the cluster the mining town is in before relocating.