

Capstone Project - GHGs emission hotspots in Edmonton

Kwabena Bediako Boateng

January 27, 2021

1 Introduction

Emission of greenhouse gases (GHGs) is increasingly becoming a major concern for large cities around the world, and the city of Edmonton is of no exception. As the Economist for the Ministry of Environment and Parks (Government of Alberta), I would like to identify areas in the city of Edmonton where emission of carbon dioxide (CO_2) is likely to be more concentrated. This will ensure that appropriate measures are taken to reduce emissions in these areas.

For the purpose of this study, I will concentrate on CO_2 emissions. This is because it is the most common GHG emitted in Canada and also the most difficult to deal with since it stays in the atmosphere for relatively longer periods. The primary sources of CO_2 emissions in Canada are: transportation, electricity production, industry, commercial and residential, agricultural, land use and forestry.

The increase in temperatures due to CO_2 emissions has several impacts. Chief among them is the human health impacts, not to mention the environmental and economic impacts. Higher

temperatures and extreme weather events may increase the risk of deaths, and of injuries from intense local weather changes. There may also be greater risk of respiratory problems.

1.1 Business Problem

Activities that result in an increase in CO₂ emissions may be concentrated in certain neighborhoods. These may include locations where industries such as oil fields, restaurants, and farmlands are located. The main objective is to find areas in the city where restaurants and other activities that result in CO₂ emissions are concentrated the most.

2 Data

2.1 Neighborhood Data

In my analysis, I will be leveraging data on the neighborhoods in Edmonton, Alberta. The data is extracted via web scrapping using `BeautifulSoup` library in Python. The data is obtained from: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T which contains a table with all the neighborhoods in the cities of Alberta. The table provides information on Postal code, Borough, Neighborhood, as well as latitude and longitude coordinates of all the neighborhoods.

2.2 Venues Data

I obtain location data which I utilize to pass the required parameters to the `FourSquare` API in order to retrieve details on the venues in each neighborhood. This data comprises venues such as restaurants , sports stadiums, bust stops etc., frequently searched using `FourSquare` API.

2.3 Data Use

Having obtained data on the venues in each neighborhood in Edmonton, I cluster neighborhoods based on their similarity in terms of the number of venues such as restaurants, oil fields, industries, office buildings etc. This facilitates the determination of hotspots for CO₂ emissions.

3 Methodology

I prepare the location data obtained from wikipedia for the analysis by making sure any Borough or neighborhood that is not assigned is removed. Afterwards, I select only the city of Edmonton with the postal code, latitude, and longitude coordinates of each neighborhood.

With my location data ready for analysis, I utilize the FourSquare API to extract data on venues in each of the 38 neighborhoods in Edmonton. I then group neighborhoods by venue to get a fair idea of how many venues exist in each neighborhood. Suffice to mention that, CO₂ emissions are likely higher in those neighborhoods with relatively more venues.

A quick inspection of the dataframe shows that North Downtown has the most venues (100) followed by West Lake District (22), and then by West Northwest Industrial, Winterburn (21). Next, I utilize a method known as one hot encoding to sort the data in order to obtain the 10 most common venues in each neighborhood.

Once again, my analysis shows that the most common venues in North Downtown are coffee Shops followed by Sandwich places and Pubs. The most common venues in West Lake District are coffee shops, fast food restaurants, and pharmacies. In West Northwest Industrial, Winterburn, Hotels, Fast Food Restaurants, and Vietnamese Restaurants are the most common venues.

I use the Silhouette Score to determine the optimal number of clusters for the neighborhoods in the dataset. The Silhouette Score measures how similar an object is to its own cluster relative to other clusters. It ranges between -1 and +1, where a high value indicates a good match and a low value indicates poor match. My analysis indicates that the Silhouette Score is highest when the number of clusters is 5. Therefore, I segment the 38 neighborhoods in my dataset into 5 clusters based on venues.

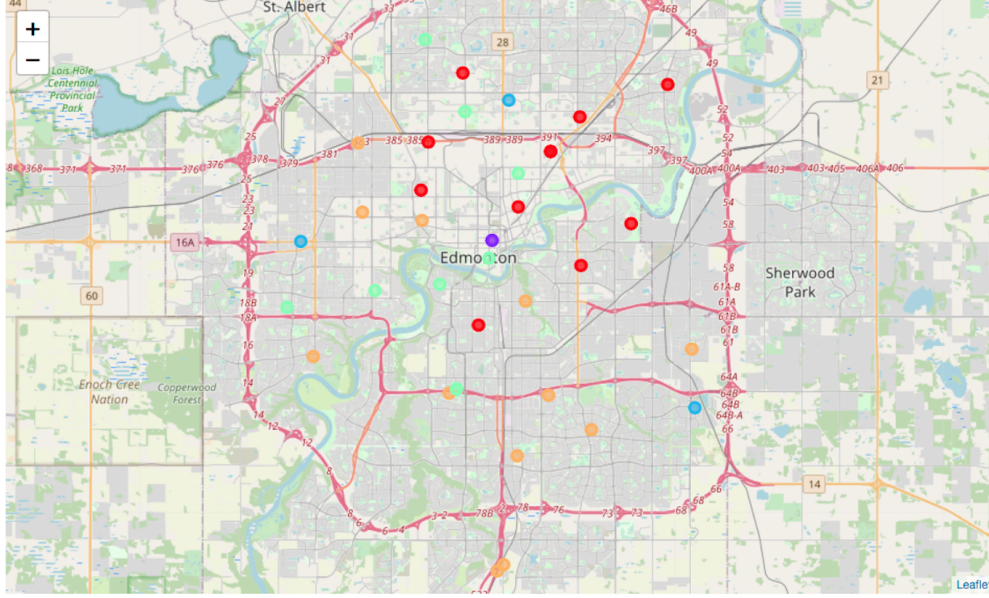
Finally, I make use of K-means clustering to train a model that segments neighborhoods in Edmonton into clusters, based on the number of venues. I choose this clustering algorithm because it is computationally faster than other algorithms, given the large number of venues in the dataset.

4 Results

The map generated from the analysis has 5 clusters and are grouped according to the number of venues in each neighborhood.

Table 1: Summary of clustered Neighborhoods

Description	Neighborhoods	Venues	Color
Cluster 0	11	4 – 6	red
Cluster 1	1	100	purple
Cluster 2	3	18 – 22	light blue
Cluster 3	9	7 – 11	light green
Cluster 4	12	1 – 3	brown



5 Discussion

The clusters from my analysis indicates the hotspots for CO₂ emissions in Edmonton. Intuitively, neighborhoods with high number of venues are hotspots for emissions. There are two very straightforward reasons for this. First, the production activities of venues such as coffee shops, restaurants, and bakeries are major sources of CO₂ emissions. Second, due to the large number of venues, consumers are more likely to travel to such neighborhoods quite frequently. As such, CO₂ emissions from vehicles running on gasoline is likely to be much more significant.

Take for instance, the purple dot on the map which represents North Downtown. This neighborhood has the highest number of venues (i.e., 100). These venues include soccer and hockey stadiums as well as most of the government offices in Alberta. For this reason, thousands of Edmontonians commute to this neighborhood on a daily basis. This suggests that North Downtown is a major hotspot for CO₂ emissions.

6 Conclusion

In this project, I have conducted analysis to determine hotspots for CO₂ emissions in the city of Edmonton. My results indicates that a major hotspot for CO₂ emission in Edmonton is north downtown, followed by west lake district, and west northwest industrial, winter-burn. With this information, the Ministry of Environment and Parks would be able to take appropriate measures to curb CO₂ emissions in Edmonton.