Capstone Project - GHGs emission hotspots in Edmonton

Kwabena Bediako Boateng

Data Science with Python Certificate, IBM

January 27, 2021

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.

Introduction

- GHGs results in warm temperatures by trapping heat in the atmosphere. Studies show that human activities are the main cause of increase in these GHGs in the atmosphere.

Introduction

- For the purposes 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.

Introduction

- The primary sources of CO_2 emissions in Canada are: transportation, electricity production, industry, commercial and residential, agricultural, land use and forestry.

Business Problem

- 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.

Business Problem

- Activities that results in an increase in CO_2 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 CO_2 emissions are concentrated the most.

Data

Neighborhood Data

In my analysis, I will be leveraging data on the neighborhoods in Edmonton, Alberta. This data would be extracted by web scrapping using `BeautifulSoup` library in Python.

Data

Neighborhood Data

The data will be 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 cordinates of all the neighborhoods.

Data

Venues Data

Next, the location data will be used to pass the required parameters to the `FourSquare API` in order to retrieve details on the different venues in each neighborhood.

Data

Data Use

Having obtained data on the venues in each neighborhood in Edmonton, I will cluster neighborhoods based on their similarity in terms of venues such as restaurants, oil fields, industries, office buildings etc. This will allow for determining hotspots for CO_2 emissions.

Methodology

Data Cleaning

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 cordinates of each neighborhood.

FourSquare API

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.

Methodology

Most common 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 Noth Downtown are coffee Shops followed by Sandwitch 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.

Methodology

Silhouette Score

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 3 (i.e., 0.388). Therefore, I segment the 38 neighborhoods in my dataset into 3 clusters based on venues.

Methodology

K-means clustering

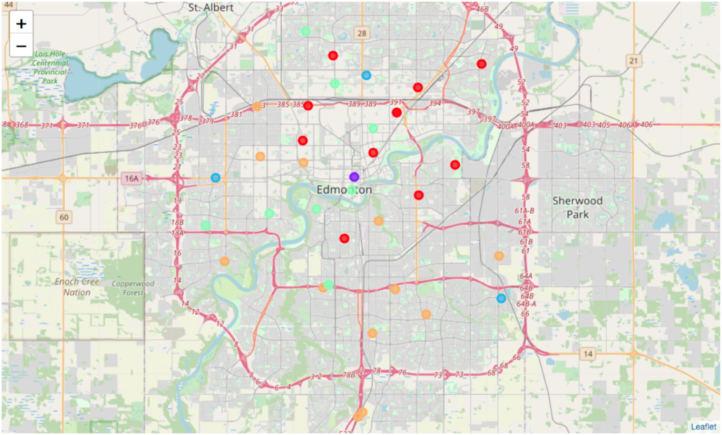
Finally, I make use of K-means clustering to train a model that segments neighborhoods in Edmonton into 3 clusters based on similarity in venues. I choose this clustering algorithm because is computationally faster than other algorithms, given the large number of venue categories in my dataset.

Results

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

Table: 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


+
−
St. Albert
Lois Hole Centennial Provincial Park
Edmonton
Sherwood Park
Enoch Cree Nation
Copperwood Forest
16A
60
28
21
14
44
Leaflet

Discussion

The neighborhoods in my dataset are divided into 3 clusters and are visualized in a map using different colors in order to allow for distinction between neighborhoods.

Conclusion

The neighborhoods in my dataset are divided into 5 clusters and are visualized in a map using different colors in order to allow for distinction between neighborhoods.