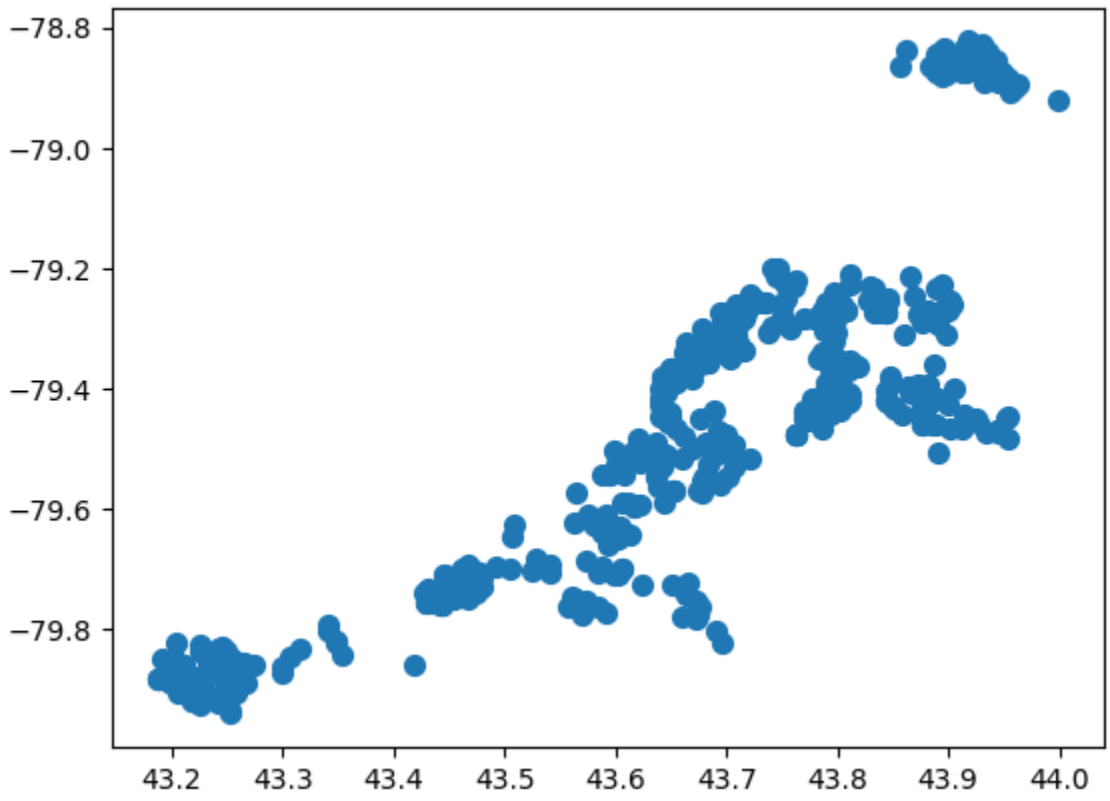
Envoi Assessment

In this assignment, I have tried to cluster all the addresses using the K-means algorithm. In this document, I have tried to explain all the aspects of my thought process while achieving the end result.

Technology Used:

- **Python** - sklearn, pandas, flask, matplotlib
- **React** - google-map-react

First of all, let's see the plot of the given dataset. In the figure below I have plotted a scattered chart of all the latitude and longitude given in the dataset using matplotlib. You will find the code in **server/src/cluster.py**

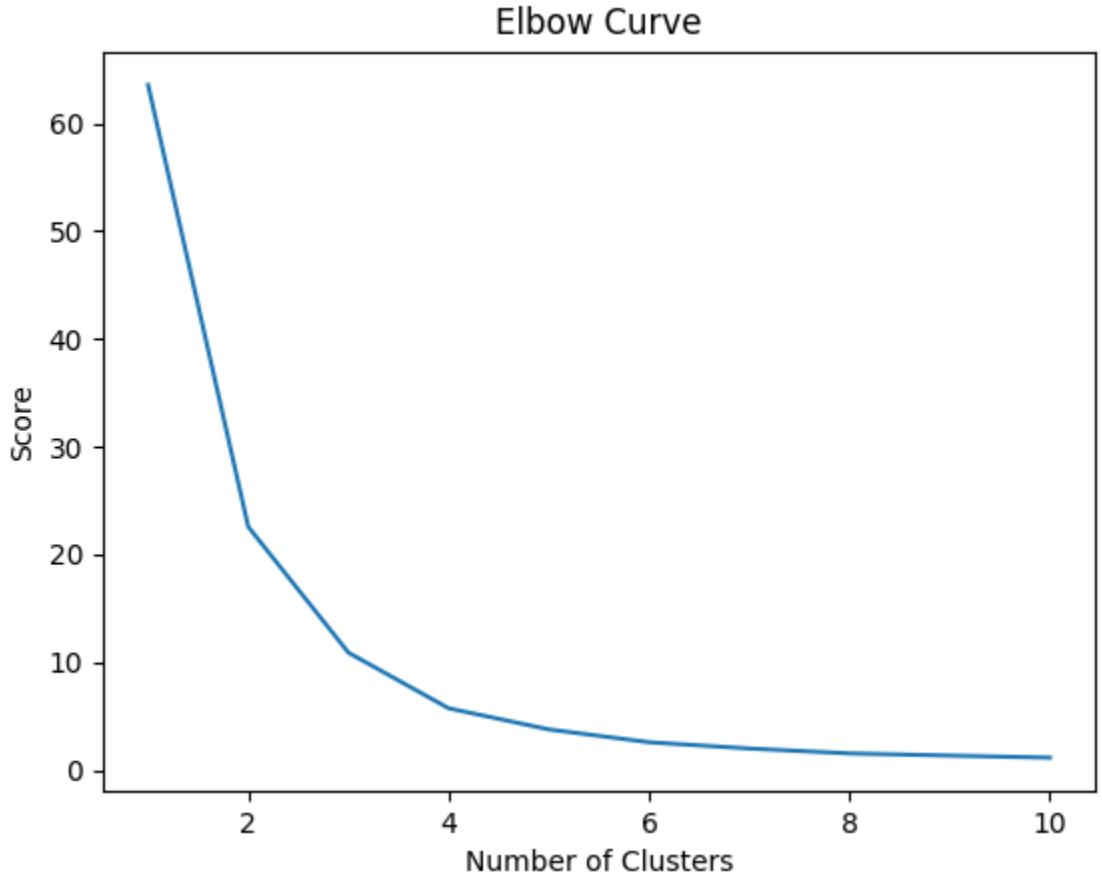


In k-means it is crucial to find the accurate value of K in order to get a proper cluster.
To find the value of K I have analyzed 2 algorithms namely:

- Elbow Method
- Silouhette Method

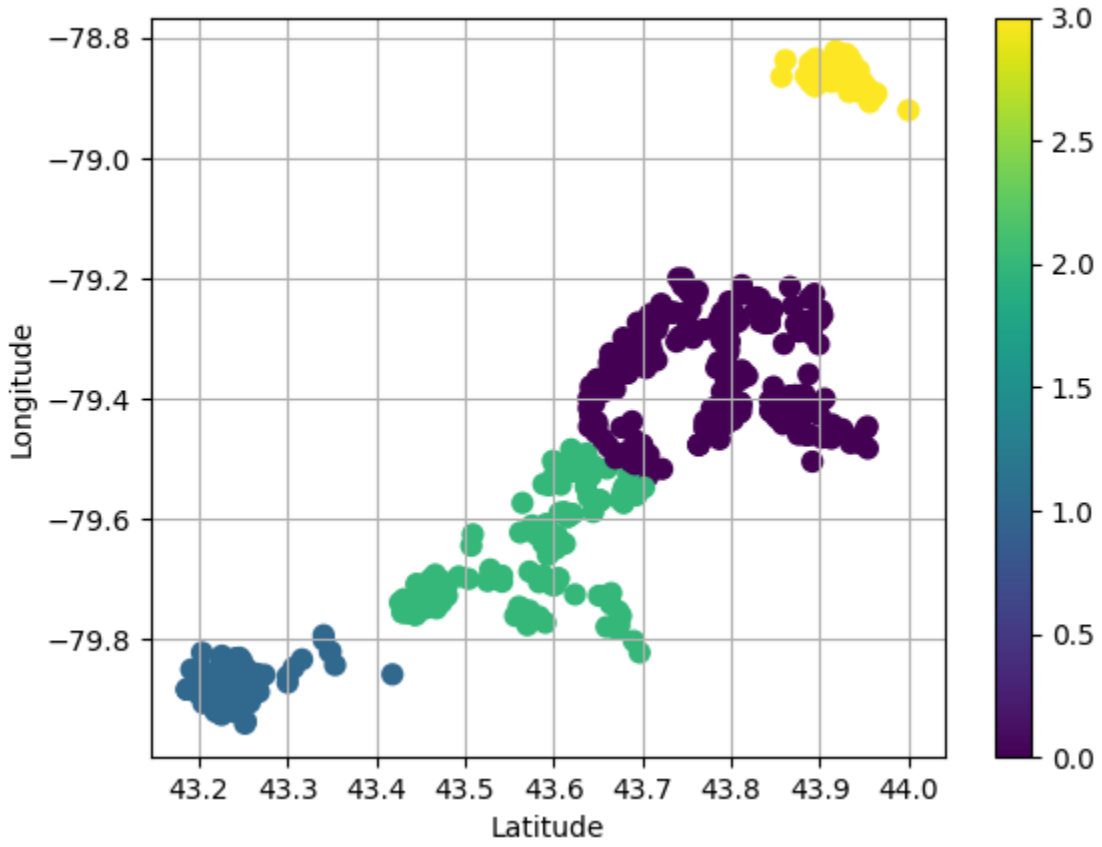
Elbow Method:

The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use.
From the elbow curve, I decided to pick a value of 4 because that is when the line is getting flat.



Final Result:

After using the K-means algorithm with K=4, I got the following result.



This can further be improved if we know the minimum distance of 2 extreme points in each cluster.