1. Data acquisition and cleaning

I utilized the data from two sources for this project. The first source was the data coming from Wikipedia, where I used BeautifulSoup to scrape the information from the following site (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>).

The scraped information related to Toronto’s neighbourhoods’, boroughs, and postal codes. Figure 2 shows an image of the site with the available data.

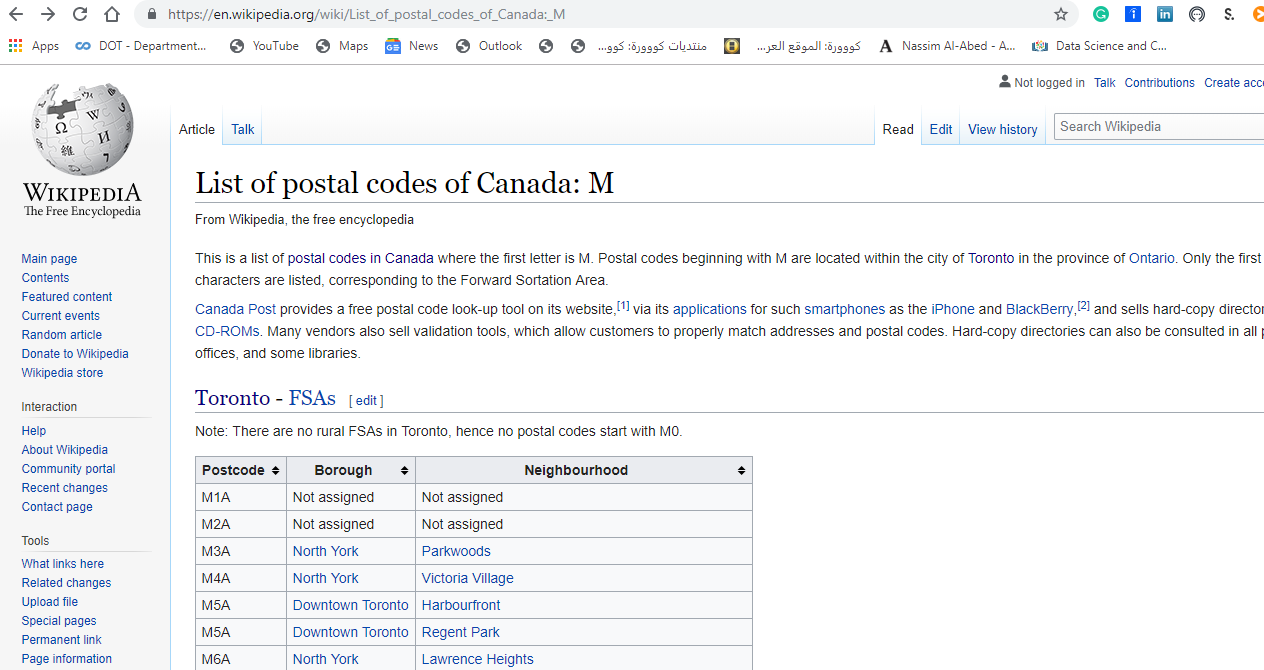


Figure 2. The website for Toronto’s districts, neighbourhoods and postal codes

Also, I scraped Toronto’s demographic information from the following page utilizing BeautifulSoup: (<https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods>).

Figure 3 shows a snapshot of the site and its content.

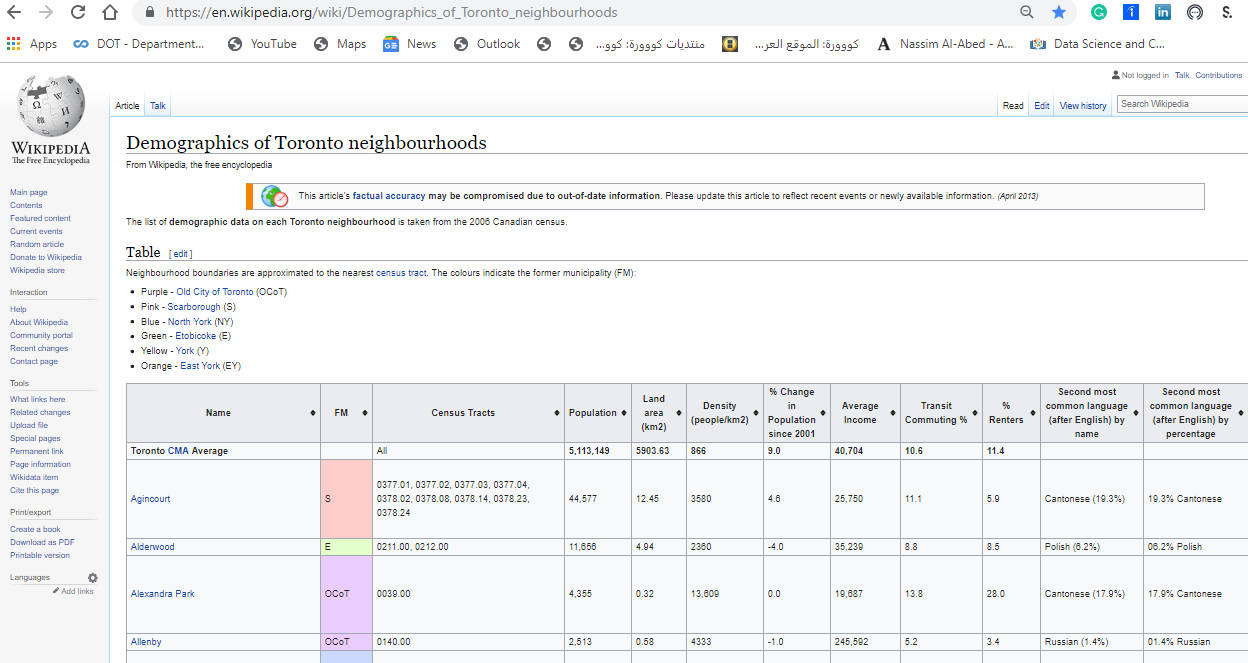


Figure 3. The website for Toronto’s demographic information

Then I used the Foursquare APIs to explore Toronto’s venues especially the restaurants to decide on the best location for a Chinese cousin restaurant based on selecting the neighbourhood with at least 10% of the residents to be of Chinese origin and with the highest population density. Figure 4 shows the Foursquare site on [www.foursquare.com](http://www.foursquare.com).

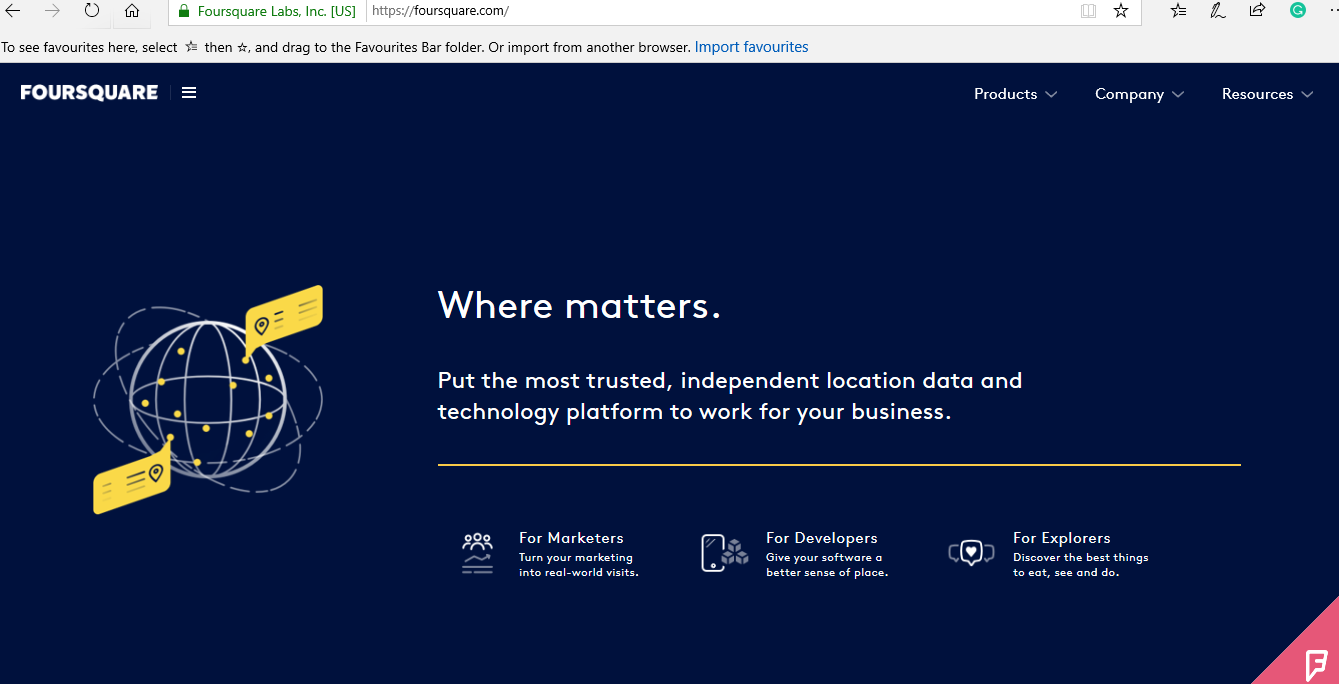


Figure 4. The Foursquare homepage

Utilizing the Foursquare website, we can see that we have around 30 Chinese restaurants in the region. Figure 5 shows the locations of the restaurants in the selected FSA (Forward Sortation Area) of North York.

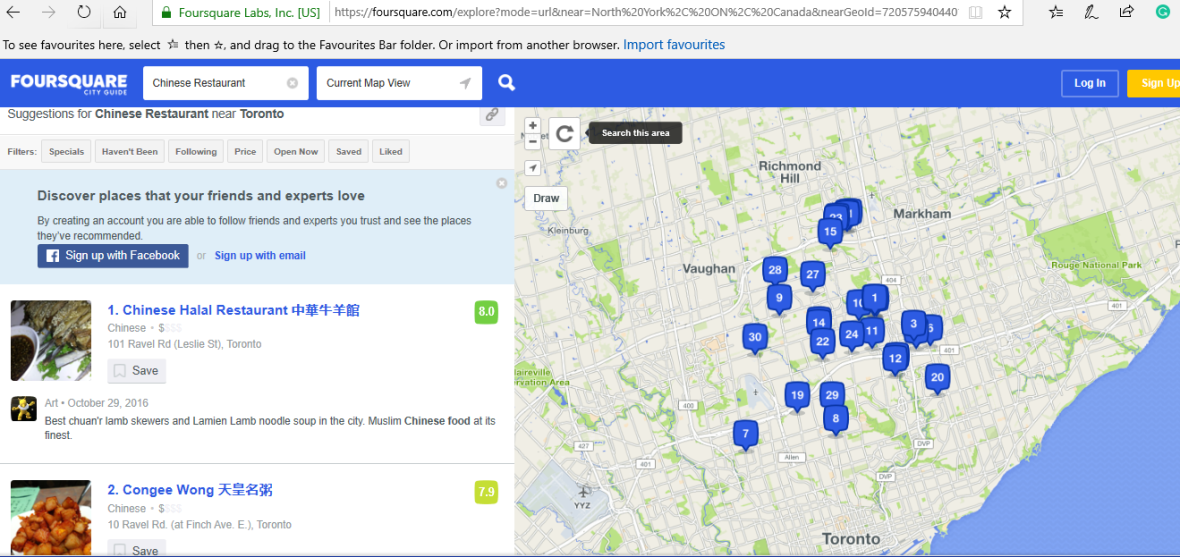


Figure 5. The Foursquare.com homepage displaying the Chinese restaurants

In this project, I selected only the FSA with the highest population density utilizing Folium mapping capability. Choosing the exact location inside the FSA requires more analysis that can be done either by ArcGIS or QGIS software which has full spatial analysis capabilities, which are outside the scope of this project.

In this project and to use Folium capabilities for creating a choropleth map of the population density utilizing the demographic data I had to download statistical data from the StatsCan website: (<https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm>) based on the FSA (Forward Sortation Area — the first three digits of the Canadian Postal Code). A forward sortation area (FSA) is a way to designate a geographical unit based on the first three characters in a Canadian postal code. All postal codes that start with the same three characters—for example, K1A—are together considered an FSA. The use of Folium for creating a choropleth map requires a GeoJSON file for Toronto as an input. This file is not readily available online, or from other sources, so I had to create this file utilizing QGIS, which is free GIS software. I had to download QGIS and have it installed before I can start this process. Figure 6 shows the QGIS software.

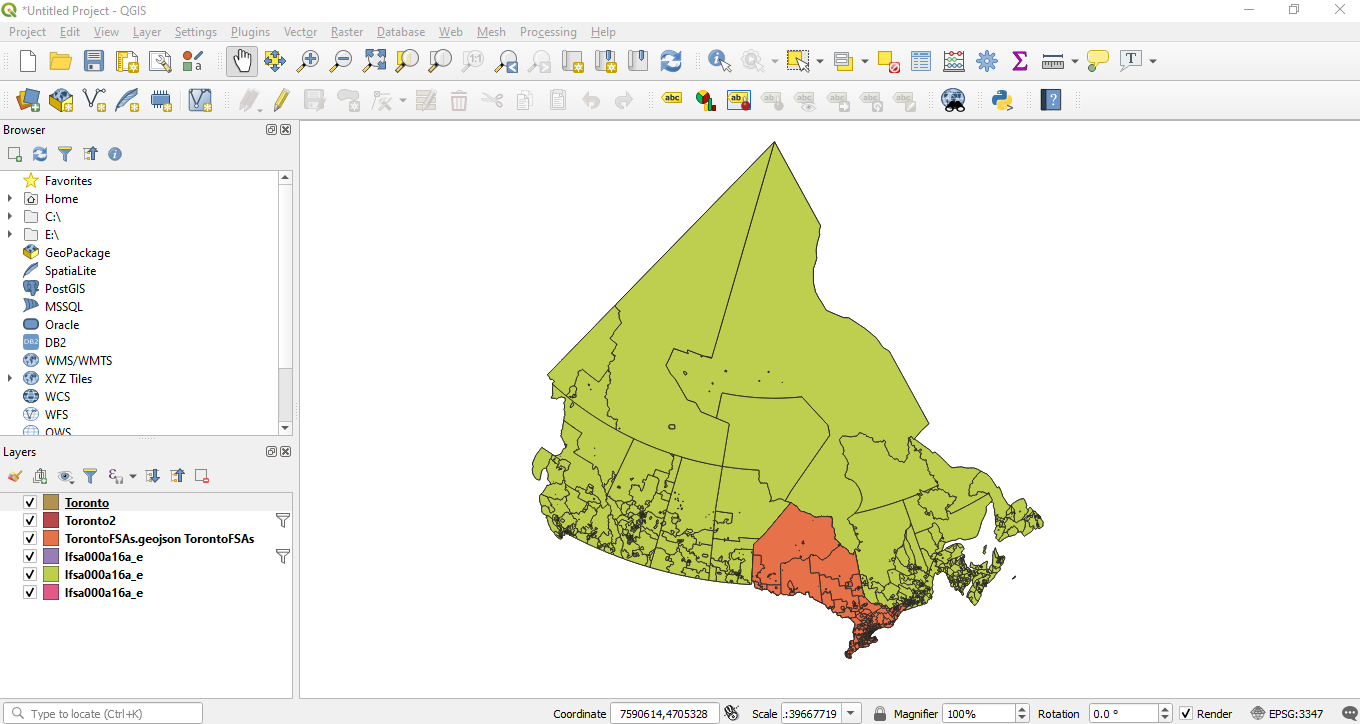


Figure 6. QGIS Software used to open and create a GeoJSON file for Toronto

To download the data from StatsCan website, we need to select the format as ArcGIS shapefile format as shown in figure 7 and to choose the data type as forward sortation area (FSA) as shown in figure 8.

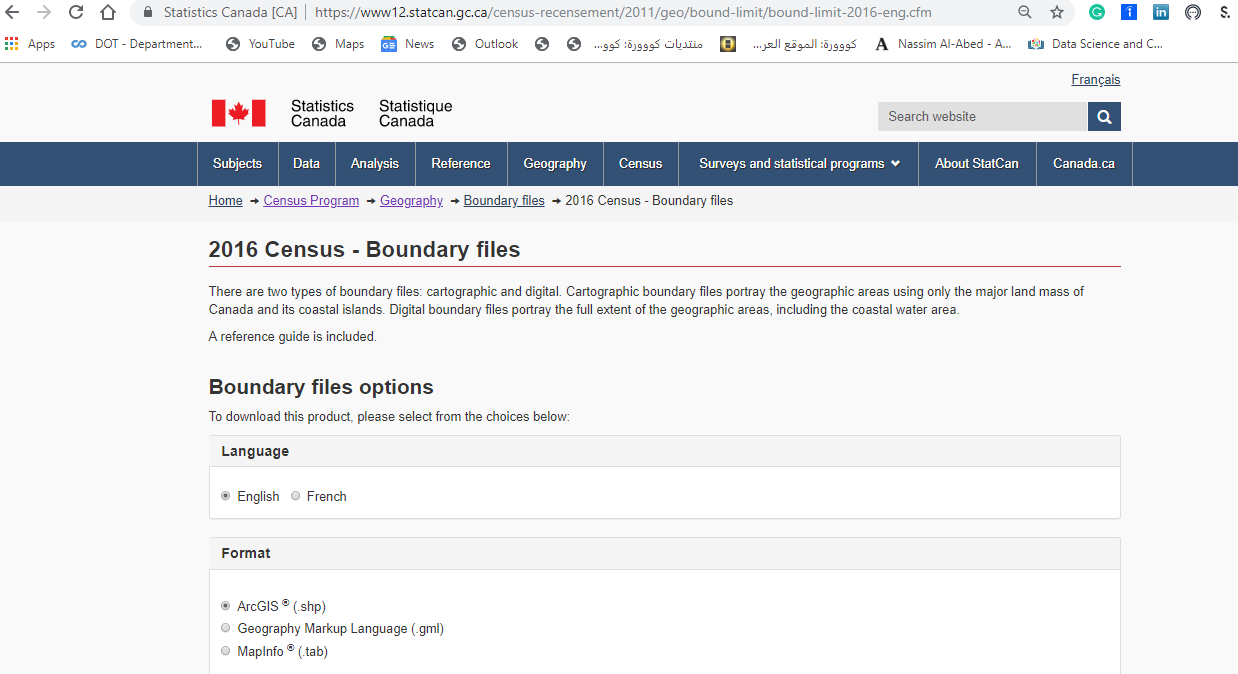


Figure 7. Downloading the 2016 census data

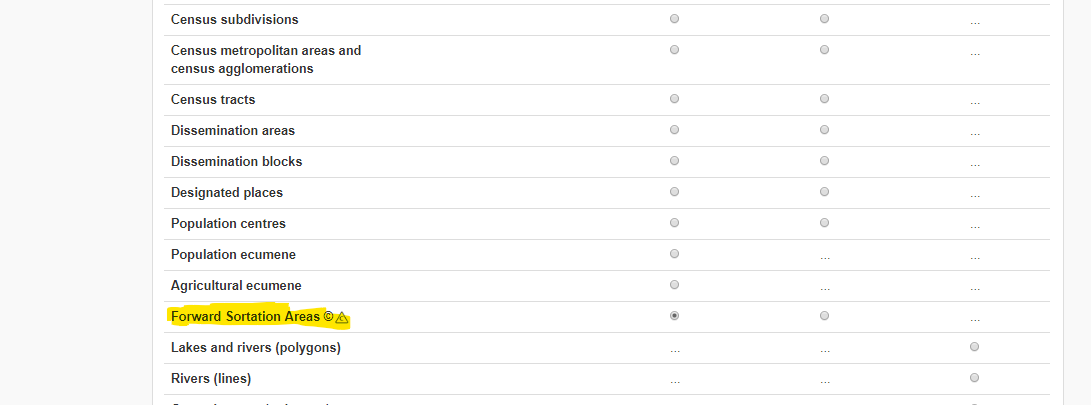


Figure 8. Selecting the boundary file as a forward sortation area (FSA)

The boundary file download will create a zip file that contains several files, and one of them is the ArcGIS shapefile. ArcGIS software does not support the conversion of the shapefile to Geojson format, but QGIS software can convert the shapefile to Geojson format. In QGIS, I added the ArcGIS shapefile using add layer, then add as vector layer as shown in figure 9.

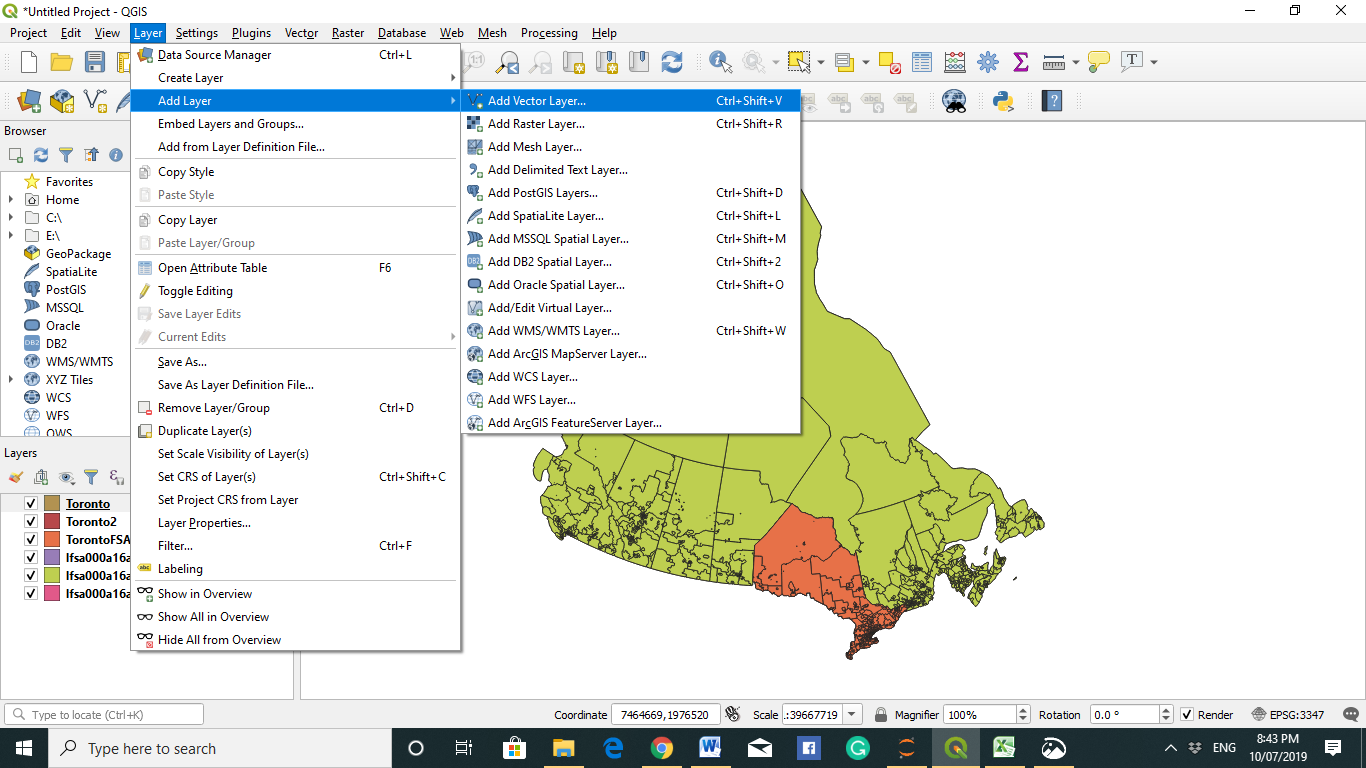


Figure 9. Adding the shapefile to QGIS as a vector layer

Then we need to specify the shapefile name and the directory where it was saved as shown in Figure 10, and then we click add. The file is vast and covers the whole of Canada, so we need to filter only Ontario province and then Toronto city. The size reduction will be done utilizing the Filter option in QGIS as shown in Figure 11.

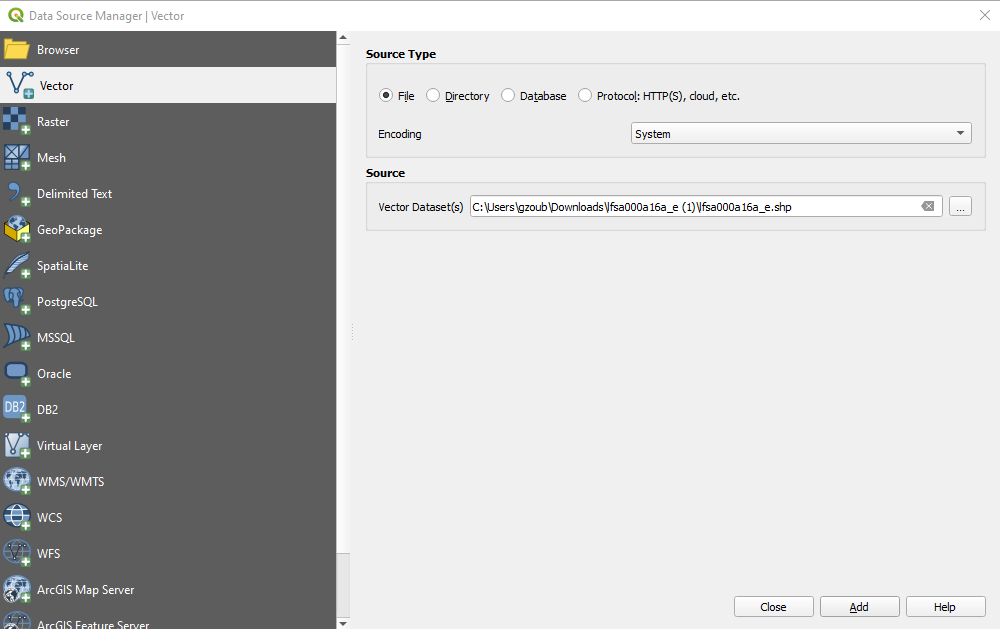


Figure 10. Specifying the shapefile name and its location in QGIS

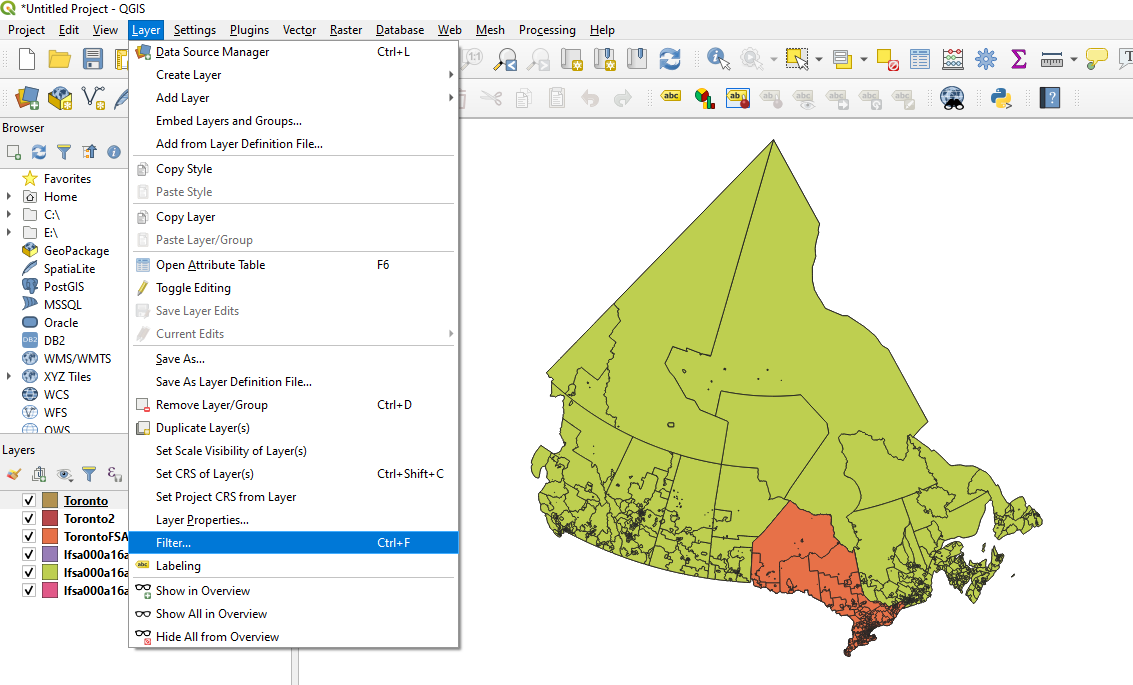


Figure 11. Using the Filter option in QGIS to select Ontario and then Toronto

As shown in Figure 12, we filtered based on province name Ontario, and based on CFSAUID of Toronto.

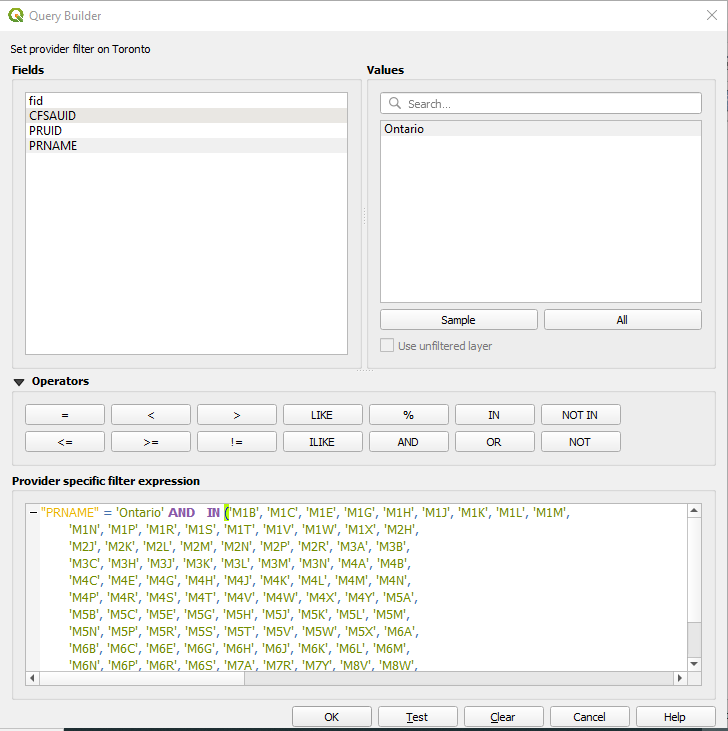


Figure 12. Filtering the data of Toronto city in QGIS

To filter Toronto city data, I had to write a python code to find the unique postal codes of Toronto, as shown in figure 13.

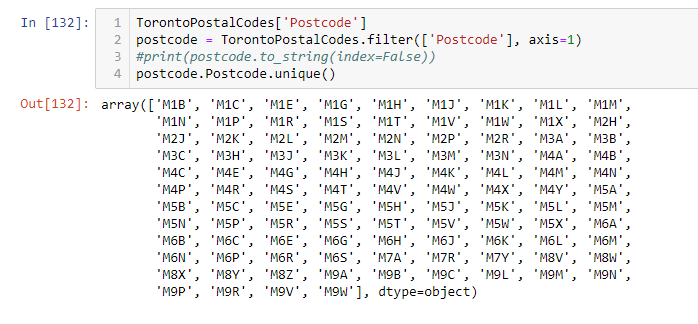


Figure 13. Finding the exclusive postal codes for Toronto

After clicking on OK in QGIS Query builder, the data for Toronto will be selected as shown in figure 14.

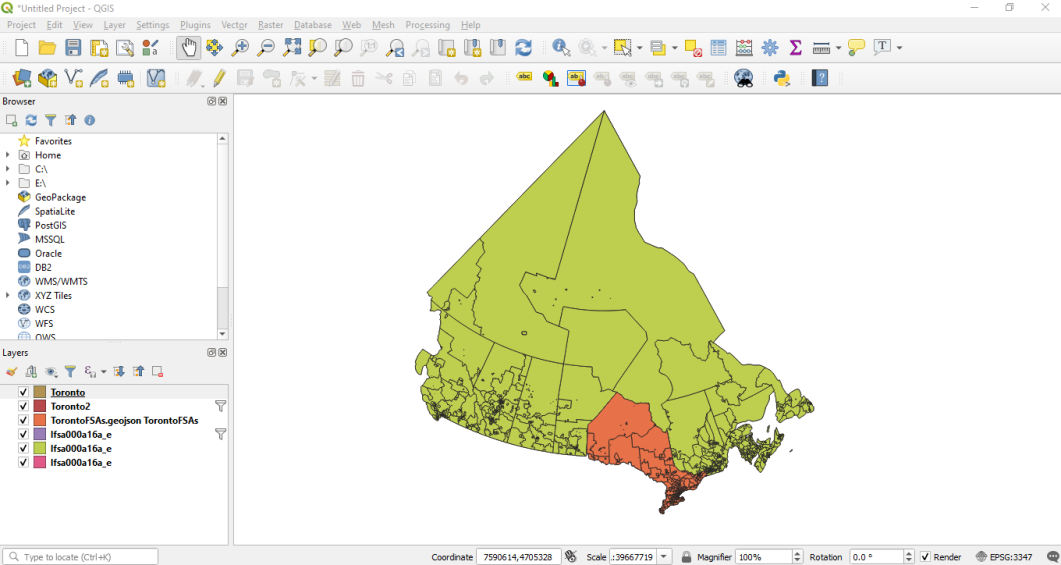


Figure 14. Saving the filtered layer in QGIS

Then after filtering the data, I selected save as option in QGIS as shown in figure 15. Then we need to specify in the save vector layer as the format that we need to save the layer into and the folder where to save it, and we need to select EPSG:4326 – WGS 84 which specify that the data will be in geographic projection as shown in figure 16.

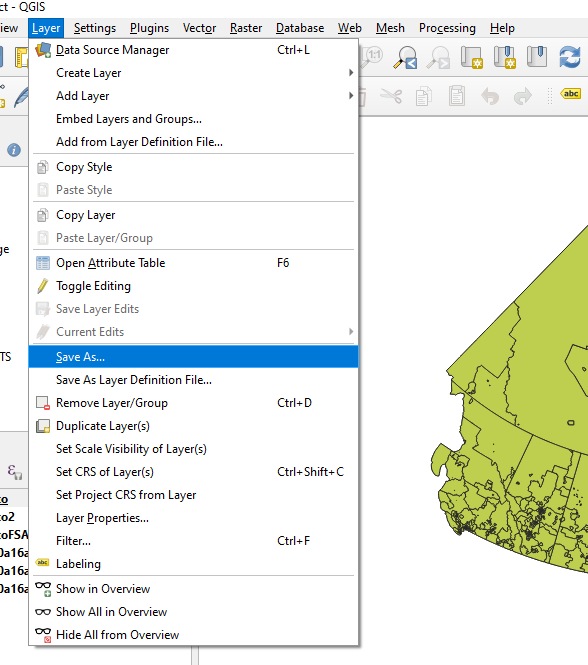


Figure 15. Saving the filtered layer in QGIS

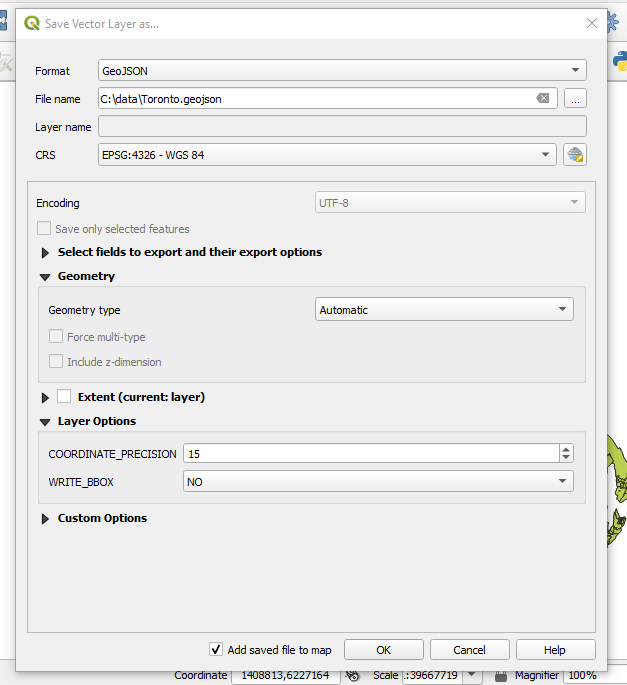


Figure 16. Selecting the Geojson file format in QGIS

Then using Folium and specifying the Geojson file as shown in the code in figure 17, the choropleth map of Toronto’s population density drawn in the chart.

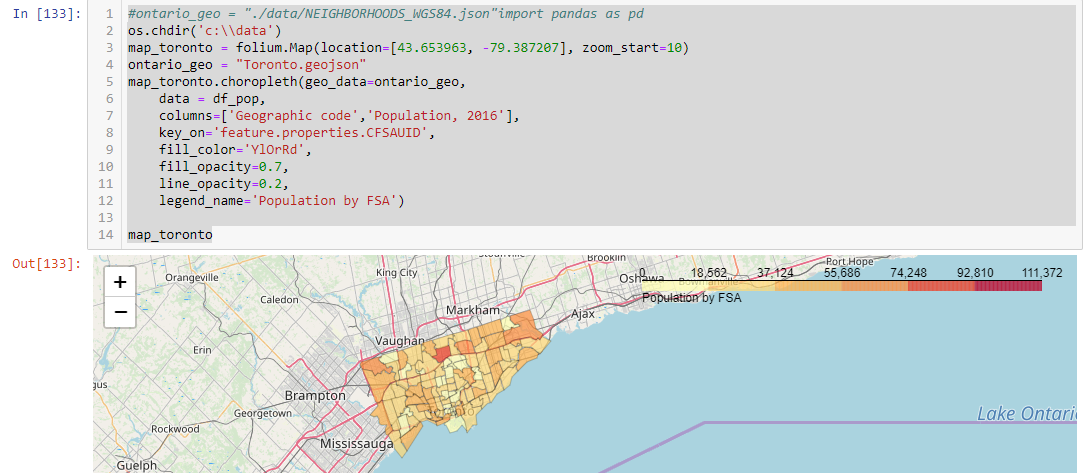


Figure 17. Drawing the population density for Toronto

The demographic data analysis of Toronto showed that the Chines language was spoken by more than 10% in North York. Using the spatial analysis showed that North York had the highest population density in Toronto as shown in figure 17, even though there are around 30 Chinese restaurants in the area, North York fulfilled the conditions set at the start of the analysis so the new restaurant shall be located there.