

Introduction

"Hello everyone! In this tutorial, I'll explain you how to download census tract boundaries for Canada. Let's get started!"

Step-by-Step Guide

"First, open your browser and search for '*Canada census boundary download*' on Google. Click on the first link that appears in the search results.

<https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limités/index2021-eng.cfm?year=21>

Once the page loads, look for the section called 'Statistical boundaries.' Here, select '*Census Tract*' from the dropdown menu.

Next, find the 'Format' bar and choose '*Shapefile*' as the format. After that, click on '*Continue*.'

A download link will appear. Simply click on it to download the shapefile. The file will be downloaded as a zip file to your computer."

Importing the File into QGIS

"Now let's import the file into QGIS.

1. Open your QGIS application.
2. Go to the top menu and click on '*Layer*,' then hover over '*Add Layer*' and select '*Add Vector Layer*.'
3. In the dialog box that appears, click on '*Browse*,' navigate to the folder where you saved the zip file, and select it.
4. Finally, click '*Add*.'

Your census tract boundaries will now appear on the map in QGIS."

Filtering for Toronto Census Tracts

"The shapefile covers census tracts for several cities and regions. To focus on Toronto:

1. Open the Query Builder for the layer by right-clicking on the layer and selecting '*Filter*.'
2. In the 'Provider Specific Filter Expression' box, use this template:

```
DGUID in ('2021S05075350128.04', '2021S05075350363.06', ...)
```

3. Click *'Test'* to check your expression, and then click *'OK.'*

Only the census tracts for Toronto will now be selected."

Exporting the Filtered Data as GeoJSON

"Now let's export the filtered data:

1. Right-click on the layer and select *'Export,'* then *'Save Features As.'*
2. In the dialog box:
 - Set the format to *GeoJSON.*
 - Adjust the CRS (coordinate reference system) as needed.
 - Choose a location and file name for the export.
3. Click *'OK'* to save the file.

Your Toronto census tracts are now exported as a GeoJSON file!"

Creating a Polygon Boundary for Your Area

"If you don't have a list of census IDs, here's an alternative:

1. Go to the website *Geojson.io.* <https://geojson.io/#map=2/0/20>
2. Draw a polygon that covers your area of interest.
3. Save the polygon as a GeoJSON file.

Now import the GeoJSON file into QGIS alongside the census shapefile:

1. Click on *'Layer,'* then *'Add Layer,'* and select *'Add Vector Layer.'*
2. Browse to your GeoJSON file and click *'Add.'*

The polygon for your area of interest will now appear on the map."

Selecting Census Tracts by Location

"To filter for your area:

1. Go to the menu and click *'Vector,'* then *'Research Tools,'* and select *'Select by Location.'*
2. In the dialog box:
 - Set *'Select features from'* to your census shapefile.

- Tick *'Intersect'* and *'Are within'* options.
 - Under *'By comparing to features from,'* choose your GeoJSON polygon.
- 3. Click *'Run.'*

The census tracts within or overlapping your polygon will now be selected."

Exporting the Filtered Data

"To save the filtered data:

1. Right-click on the modified shapefile in the *'Layers'* section.
2. Select *'Export,'* then *'Save Selected Features As.'*
3. In the dialog box:
 - Set the file format to *GeoJSON*.
 - Adjust the CRS as needed.
 - Set the file name and location.
4. Click *'OK.'*

This will create a file with your desired census boundaries."

Polishing the File in Geojson.io

"Note that some additional census tracts may be included. To remove them:

1. Go back to *Geojson.io* and import your file.
 2. Click on any unwanted census tracts and delete them.
 3. Once you've finished, save the polished file as GeoJSON."
-

In this section, I'll show you how to download and preprocess the 2021 Canadian Census of Population data, focusing only on the census tracts you need. Let's get started!"

Downloading the Data

"First, search Google for '2021 Canadian census.' You'll find a link to Statistics Canada with this address:

<https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E>

Open the page, and in the top-right corner, click on the 'Download Data' button.

Next, under the *Comprehensive download file* dropdown, select:

'Canada, provinces, territories, census divisions (CDs), census subdivisions (CSDs), and dissemination areas (DAs) - Ontario only.'

This file contains the detailed census data we need. Remember, other files might not have the level of detail required or may cover regions too large for our work."

Handling the Downloaded File

"Once the file is downloaded, you'll notice it's a large CSV file—more than 2 GB! To extract just the census tract data relevant to your area, you'll need the *census boundaries GeoJSON file* we created earlier.

Place the GeoJSON file and the downloaded folder in the same directory, then run the following Python code:"

Python Code Explanation

"Here's what the code does:

1. It reads the large CSV file and your GeoJSON file.
2. It filters the census data to include only the rows corresponding to the census tracts in your GeoJSON file.
3. The filtered data is saved as a smaller CSV file for easier handling.

Here's the code:

```
import pandas as pd
import json

# Provide the correct path to your CSV file
file_path = "98-401-X2021007_eng_CSV/98-401-X2021007_English_CSV_data.csv"
```

```

# Load the data
df = pd.read_csv(file_path, encoding='ISO-8859-1')

# Load your GeoJSON file
with open('census_Boundaries.geojson', 'r') as f:
    geojson_data = json.load(f)

# Extract the unique DGUID values
dguids = set()
for feature in geojson_data['features']:
    dguids.add(feature['properties']['DGUID'])

# Filter the DataFrame for relevant DGUIDs
filtered_df = df[df['DGUID'].isin(dguids)]
filtered_df.to_csv('censusdataunclean.csv', index=False)

```

After running this, you'll have a new file named *censusdataunclean.csv* containing only the relevant data."

Splitting Large Files

"If the filtered file is still too large for uploading or sharing, you can split it into smaller parts using this code:

```

import math

# Define the input file and output prefix
input_file = "censusdataunclean.csv"
output_prefix = "Raw_census_part_"
num_parts = 6 # Number of parts to split the file into

# Read the entire CSV to determine the number of rows
try:
    df = pd.read_csv(input_file)
    total_rows = len(df)
    rows_per_part = math.ceil(total_rows / num_parts)

    # Split the DataFrame into parts and save each as a CSV
    for i in range(num_parts):
        start_row = i * rows_per_part
        end_row = min((i + 1) * rows_per_part, total_rows) # Ensure we don't
go out of bounds
        chunk = df.iloc[start_row:end_row]
        output_file = f"{output_prefix}{i+1}.csv"
        chunk.to_csv(output_file, index=False)
        print(f"Part {i+1} saved to {output_file}, rows {start_row} to
{end_row - 1}")

except Exception as e:
    print(f"An error occurred: {e}")

```

This will create smaller CSV files like *Raw_census_part_1.csv*, making it easier to manage your data."

Conclusion

"And that's it! You now have all the census data for the census tracts you're working with. Once the preprocessing is complete, you can move on to analyzing your area using the main Python code for your project.

If you found this helpful, please like and share. See you in the next tutorial!"
