**README File for CORI Data Analyst Candidate Exercise**

The purpose of this project is to visualize maximum advertised upload speeds by census county subdivision for New Hampshire and Vermont, weighting upload speeds by population. To showcase the programming skills, the project used R, PostGreSQL and Python codes, separately. The project is finished by following steps:

1. Download the required data (or use the data in /Data directory)
2. Inspect the broadband data by R and subset to the study region (pre\_process.R)
3. Create table in PostGreSQL (sqlCreateTable) and import broadband data to PostGreSQL (Tools-Import/Export)
4. Process the broadband data by Python (after\_process.py or after\_process.ipynb) and prepare the gpkg file (/Result/broadBandData\_county.gpkg)
5. Upload gpkg file to CARTO and create the map

**Step 1 Download the required data from the links below or use the data in /Data directory**

Broadband Data: <https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477>

Population data: <https://www.census.gov/data/datasets/2017/demo/popest/counties-total.html>

County Information: <https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html>

**Step 2 Inspect the broadband data by R and subset to the study region**

Please find the R code: pre\_process.R

The propose of this code is to check the data types and subset to only Vermont and New Hampshire data. It greatly reduces the 11 GB file to 74 MB.

The output csv file is /Result/bbdata\_vt\_nh.csv

**Step 3 Create table in PostGreSQL and import broadband data to PostGreSQL**

In pgAdmin 4, create a new table by the script sqlCreateTable

Then, import the “bbdata\_vt\_nh.csv” file to the created table

**Step 4 Process the broadband data by Python**

Please find the python code: after\_process.py or after\_process.ipynb

The main propose of this code is to export data from PostGreSQL to python and prepare the desired gpkg file for CARTO map making.

The output csv file is /Result/broadBandData\_county.gpkg

**Step 5 Upload gpkg file to CARTO and create the map**

In CARTO, upload the prepared gpkg file to the CARTO datasets.

Then, in data, select all fields, and round the *weighted\_maxadup* to integer saving as a new column *round\_wmaup*, by:

SELECT \*, round(weighted\_maxadup) AS round\_wmaup

FROM chenchen.broadbanddata\_county

In the style, visualize the weighted maximum advertised upload speeds with color ramps from light to dark, and also separate two states with individual color ramps.

#layer [stname='New Hampshire']{

polygon-fill: ramp([round\_wmaup], (#fde0c5, #facba6, #f8b58b, #f59e72, #f2855d, #ef6a4c, #eb4a40), jenks);

}

[stname='Vermont']{

polygon-fill: ramp([round\_wmaup], (#eff3ff, #c6dbef, #9ecae1, #6baed6, #4292c6, #2171b5, #084594), jenks);

}

#layer::outline {

line-width: 1;

line-color: #FFFFFF;

line-opacity: 0.5;

}

#layer::labels {

text-name: [name];

text-face-name: 'DejaVu Sans Book';

text-size: 10;

text-fill: #FFFFFF;

text-label-position-tolerance: 0;

text-halo-radius: 1;

text-halo-fill: #6F808D;

text-dy: -10;

text-allow-overlap: true;

text-placement: point;

text-placement-type: dummy;

}

Create two widgets for *Provider Counts* and *Average MAUS by States* to show

1. The dominant broadband providers in two states
2. The average weighted maximum advertised upload speeds in two states

Create popup window for each county to show State, County, 2017 Population, Maximum Advertised Upload Speeds, Weighted Maximum Advertised Upload Speeds, Provider Name, Holding Company Name