**Data Acquisition and Munging**

This project uses public data from Portland TriMet. TriMet provides Geo data for developers with .geojson files for boundaries, stops, bus routes and rail lines.

The raw data was downloaded from the TriMet website. (https://trimet.org/about/performance.htm)

Census Reports and Route Ridership Reports were manually downloaded for the periods Spring 2018,

Fall 2018, Spring 2019, Fall 2019. File names had parentheses “(“, “)” in them which prevented them from being downloaded and converted from pdf to csv format programmatically.

The files arrive in PDF format and were converted from PDF to XLSX format using the online service <https://www.pdftoexcel.com/> and renamed removing the parentheses.

Each file had multiple sections with headings and descriptive text in each section and some sections were offset compared to the others. VBA scripts were used to remove the top heading and name each column. Unwanted columns were removed. To do this, open each file, import and run the VB script “census.bas” or “routes.bas”. Run the first Sub in the script.

Data conversion was completed using Jupyter Notebook and Pandas. “MungingTrimet.ipynb” The notebook has three sections. Section 1 if for census data, Section 2 is for routes data, and Section 3 adds census data to “tm\_stops.json”. All files are converted to JSON format.

None: File "stop\_level\_passenger\_census\_sorted\_by\_location\_id\_weekdaySpring2018.xlsx.csv"

required some manual editing before it would work in pandas. Lines 4825 to 4871.

**Skill Sets**

Python: Jupyter Notebook

Pandas

JSON

JavaScript

Leaflet

Plotly

GeoJSON

JSON

VBA

Scripting

File type conversion

**New JavaScript module:**

dd