Repository: <https://github.com/chrisp26/18-Project/tree/main/Resources/Code>

Resources are:

Postcodes & suburbs:

This file has a SQL script.

<https://github.com/chrisp26/18-Project/blob/main/Resources/Code/creating_postcode_database.ipynb>

Vehicle statistics

Code to create statistics

<https://github.com/chrisp26/18-Project/blob/main/Resources/Code/creating_state_statistics.ipynb>

Raw data

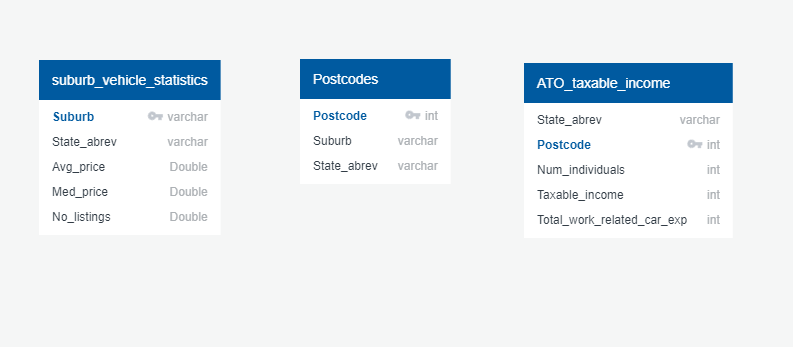
<https://github.com/chrisp26/18-Project/blob/main/Resources/Data/scraped_cars_2021-10-12.json>

ATO income data

Still in an xlsx format

<https://github.com/chrisp26/18-Project/blob/main/Resources/Data/ts19individual06taxablestatusstateterritorypostcode.xlsx>

Plan was to create a SQL database:



I have not actually put them into SQL tables yet. That is the data I have.

My end goal is to create a JSON file with the following fields:

* Postcode
* State\_abrev
* Avg\_price
* Med\_price
* No\_listings
* Num\_individuals
* Taxable\_income
* Total\_work\_related\_car\_exp

My plan was create the tables above in a SQL DB, import it to Python/Pandas and do a join to get *Postcode* into a dataframe with the variables from *suburb\_vehicle\_statistics*. As a suburb may fall into multiple suburbs, I was planning to take the first one and drop the rest.

Then I’d do another join to load in the variables from *ATO\_taxable\_income* and export the whole thing as JSON.