 **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

 **T**ransform: what data cleaning or transformation was required.

 **L**oad: the final database, tables/collections, and why this was chosen.

Cultural diversity:

Diversity csv

* checked for datatypes. all floats (percentages) and one object (division name) as expected.
* Checked for duplicates and there are no duplicates divisions, as expected so no cleaning/wrangling required.
* checked for missing values
* Renamed columns to ensure consistency
* percentages show % of all australians that is culturally diverse? TBC
* 151 rows

Division csv

* created for purposes of merging with multiplate databases
* however did a check for duplicates of divisions (none) and datatypes (integer and objects as expected)
* merged the diversity and division df, linked on electoral division name.

Merged df

* Checked datatypes (object for division and state, float for all other)
* Division\_id changed from rounded number to number with one decimal, but with still the same datatype. Apparently, you can’t convert a float to an integer when there is NAN values. Did a len(df) check and after merge the number of rows went from 158 to 143! Odd as there are no missing values in the csvs?

Age\_df

* checked columns
* checked datatypes and are as expected
* counted rows (151), deleted NAN values and counted rows again (151 rows). no NAN were present

div\_df

* created for purposes of merging with multiplate databases
* merged the age\_df and division df, linked on electoral division name, outer join
* dropped state column TBC
* counted rows post merge (158)
* Division\_id changed from rounded number to number with one decimal, but with still the same datatype. Apparently, you can’t convert a float to an integer when there is NAN values. Did a len(df) check and after merge the number of rows went from 158 to 143! Odd as there are no missing values in the csvs?

pop\_df

* count rows (151)
* join on age\_df/div\_df merge
* renamed columns for bins to make it more readable

Merged2

* 158 rows post join
* group by division and division id -> 150 rows!