**01/05/2021 Minutes**

Did the following -

* Processed the teacher supply and demand tables in the Word Document using Python. We figured out how to put it in a data frame and now need to finalise it and figure out a few issues mentioned below.
* Manually filtered out the AEDC Excel Spreadsheet to only have Victorian data and be all in one sheet.
* Narrowed our scope - decided to work mostly with Early Childhood (<13 years)
* Found more data we believe would be relevant e.g proportion of children’s parents on health-care cards.
* Deleted some redundant data.

Issues –

* Some local government areas in the TSDR data do not match up with the ones in the AEDC data set. Only data for Catholic schools has this issue.
  + Decided to remove any rows in the dataframe with missing values as we cannot calculate any student:teacher ratio this way.
* Some preprocessing on the numbers need to be done as we found anomalies like
  + ‘ < 5 ‘
  + Commas
* We had to manually count the index for the tables in the Word document
* Dealing with LGAs vs DHS areas.

Actions before Next Meeting

* Figuring out how to deal with LGAs vs DHS areas – **Sen**
* Putting VCAMS data into pd dataframes
  + Bullying, connectedness, family stress - **Steven**
  + Food, support - **Lina**