**Extract, Transform, Load Project**

**Team Members:**

* Rajesh Nair
* Matt Vella

**Objective:**

Perform an ETL on the male, female and total populations of each state and territory in Australia from January 1983 to January 2021 in addition to each state’s total retail turnover in the same period.

**Why?**

The purpose of this task is to establish a data source which can help analysis the impact of population growth and retail spending along with the potential impact of COVID-19 on the retail sector.

We aim to have this information available in such a way that a visual analysis can be conducted to demonstrate any relationship between the two measures.

Users of this data may include small business owners who are trying to determine how a) retail trade is going both pre and post COVID-19 pandemic and b) how the Australian retail sector has developed since the 1980s. This may provide some insight to how they forecast and project retail investment and business development.

Consequently, the data should be loaded in such a way that visualisation tools such as Tableau and Power BI can be employed to create meaningful reports.

Lastly, we aim to establish a baseline in which a hypothesis can be tested in how these results compare to the impact of COVID-19 in the global retail market.

**Data Sources:**

* Australian Bureau of Statistics:
  + Population – states and territories
    - [National, state and territory population, September 2020 | Australian Bureau of Statistics (abs.gov.au)](https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release)
      * <https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/sep-2020/310104.xls>
  + Retail and Wholesale Trade:
    - [Retail Trade, Australia, April 2021 | Australian Bureau of Statistics (abs.gov.au)](https://www.abs.gov.au/statistics/industry/retail-and-wholesale-trade/retail-trade-australia/latest-release)
      * <https://www.abs.gov.au/statistics/industry/retail-and-wholesale-trade/retail-trade-australia/apr-2021/850103.xls>

**Extract & Transform**

* Downloaded .xls files were imported into Pandas through the pd.read\_excel method.
* Both datasets required a drop of eight of the initial rows as they did not contain any data.
* In the retail\_whole\_turnover dataset, only the ‘Original’ series type columns were kept.