# **ETL Project Proposal:**

**Team members:** Solo, David Ch’ng

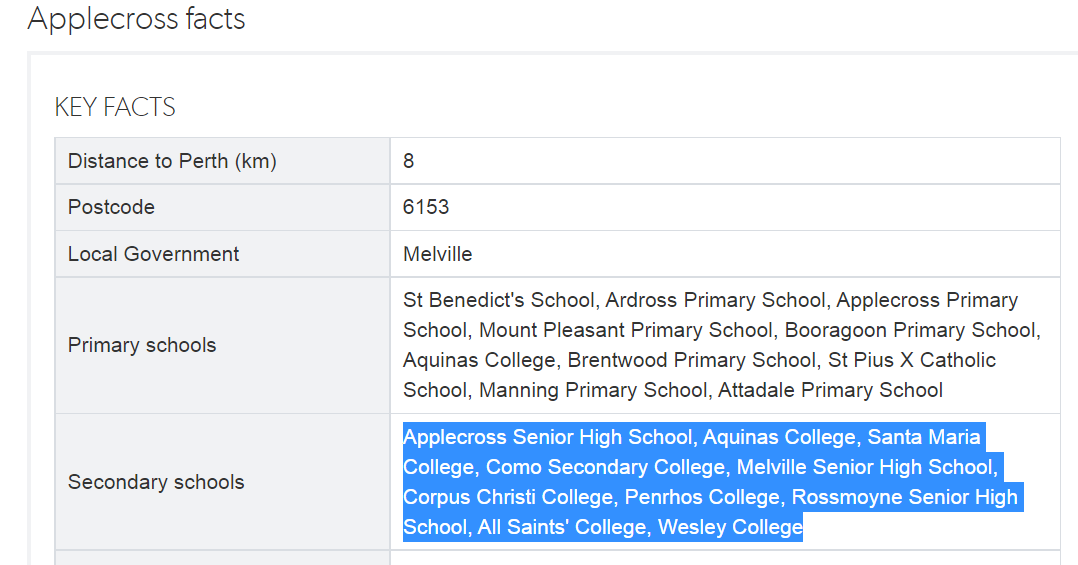
**Project:** Top performing high schools in Applecross

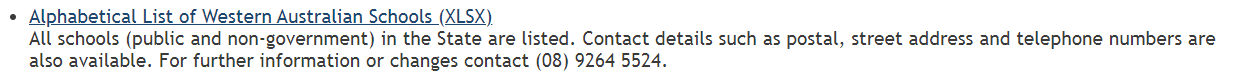
**Perceived Customer:** A family has moved into Applecross and is seeking to enrol their Year 12 child in the best performing government secondary school in the Applecross area based on its performance during the 2020 ATAR exam, and would also like to consider how many children are enrolled in Year 12 to ensure their child gets adequate attention.

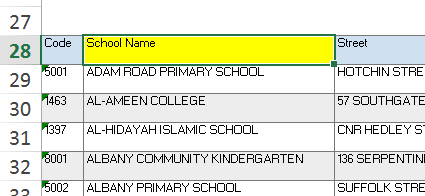
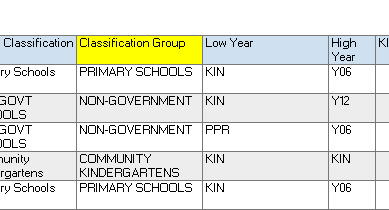
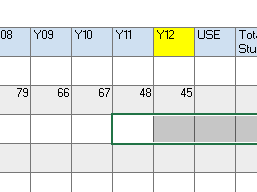
## **Project Timeline:**

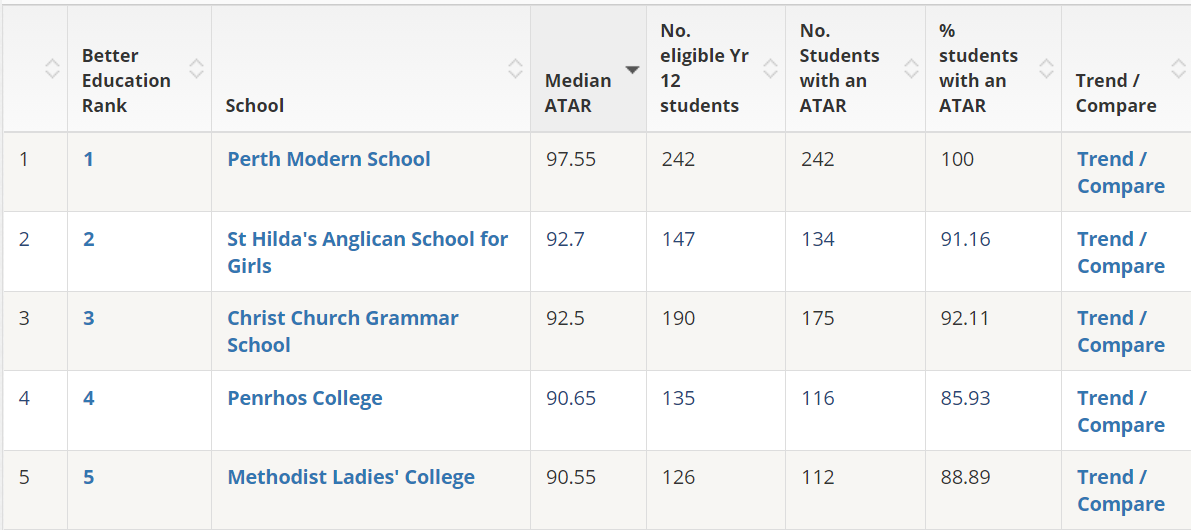
* **Tue 31/8:**
  + Determine project proposal and data sources
  + Commence EXTRACTION
    - Data source 2 - download CSV file, cleanse data and stage
* **Thu 2/9:**
  + Complete EXTRACTION
    - Data source 1 - Screen scrape secondary schools
    - Data source 3 - Screen scrape school and ranking
  + Commence TRANSFORMATION
    - Dataset 1
    - Dataset 2
* **Mon 6/9:**
  + Complete TRANSFORMATION
    - Dataset 3
    - Dataset 4
  + Complete LOAD
    - SQL Schema

## **EXTRACT:**

1. **Data source 1 –** REIWA Applecross suburb profile, secondary schools
   1. **Method** - Screen scrape
   2. **Purpose** - Extract a list of all the secondary schools in the Applecross area
   3. [**reiwa.com - Suburb profile for Applecross**](https://reiwa.com.au/suburb/applecross/)
   4. 
2. **Data source 2** – Alphabetical List of Western Australian Schools (XLSX)
   1. **Method** – CSV download
   2. **Purpose** – obtain a list of WA schools and a flag to see whether it is government or non-government, as well as a flag to determine how many children are in the Yr 12 cohort
   3. <http://det.wa.edu.au/redirect/?oid=SiteProxy-id-19622149&title=Alphabetical+List+of+WA+Schools+xlsx&skip=true&launch=true>



1. **Data source 3** – WA School Ranking – 2020
   1. **Method** – Screen scrape
   2. **Purpose** – obtain a list of WA School Rankings so the user can merge on the school information to determine the top ranking public schools
   3. <https://bettereducation.com.au/results/wa/wace.aspx>
   4. 

## **TRANSFORM:**

1. **Dataset 1:** WA\_Schools\_Y12
   1. Data source: 2
   2. Columns: School Name, Y12, #ID for Classification Group
2. **Dataset 2:** WA\_Schools\_classification
   1. Data source: 2
   2. Columns: ID, Classification Group
3. **Dataset 3:** Applecross\_secondary\_schools
   1. Data source : 1
   2. Columns: Secondary School, Suburb
      1. Note – Suburb is hardcoded as ‘Applecross’
4. **Dataset 4:** Secondary\_schools\_ranking
   1. Data source: 3
   2. Columns: School, Better Education Rank, Median ATAR

## **LOAD:**

1. **SQL Databases**
   1. Schools\_db
      1. Datasets x 5