### Purpose:

Analysing and visualising NSW Public School data using Power BI.

### Data Source:

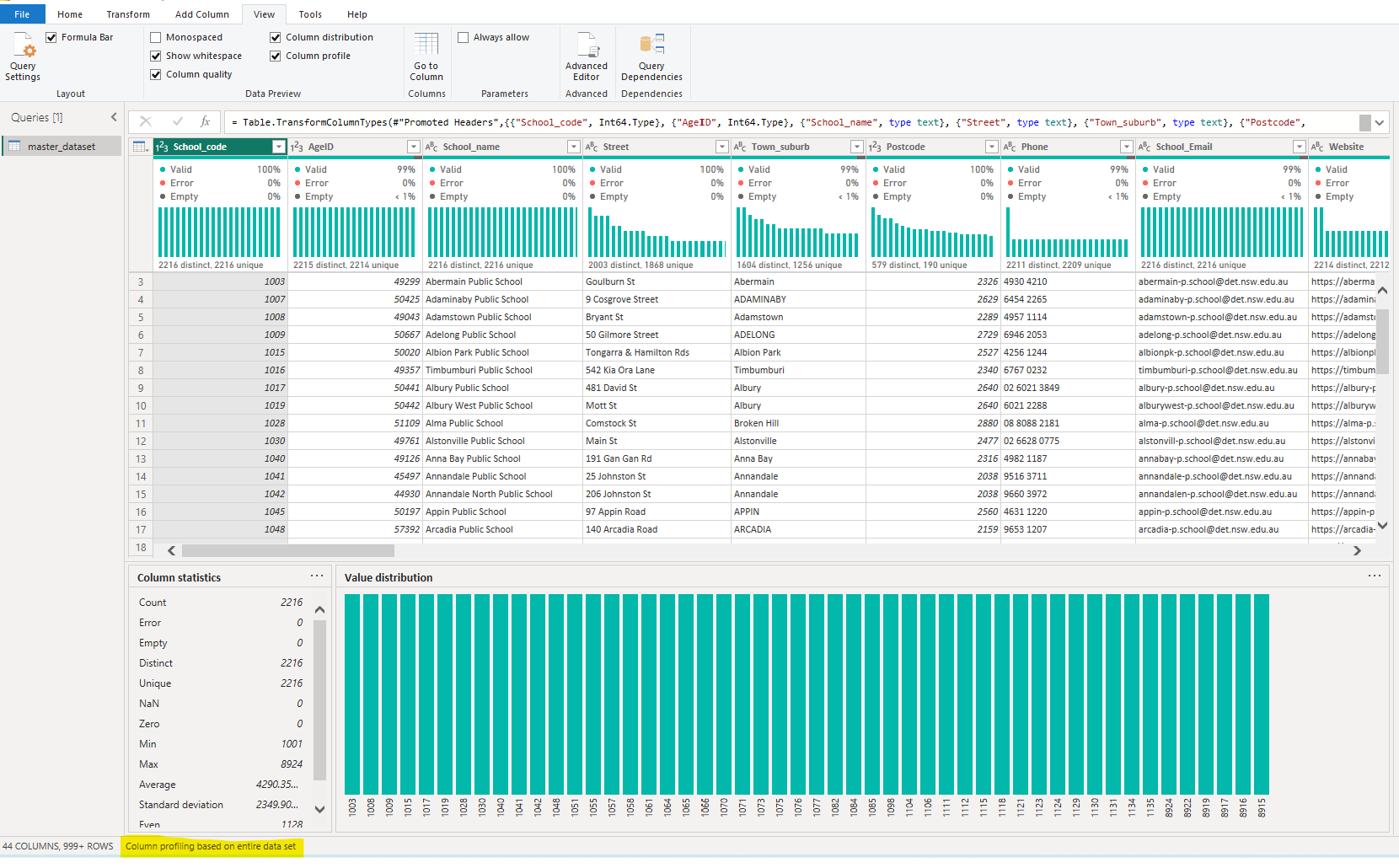
|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset Files** | **File name** | **Location** | **Comments** |
| NSW public schools master dataset in csv format | master\_dataset.csv | <https://data.cese.nsw.gov.au/data/dataset/nsw-public-schools-master-dataset/resource/2ac19870-44f6-443d-a0c3-4c867f04c305> | Total number of records in the csv = **2216 records** |
| Data dictionary for this dataset in csv | master-dataset-data-dictionary.csv | <https://data.cese.nsw.gov.au/data/dataset/027493b2-33ad-3f5b-8ed9-37cdca2b8650/resource/31b7ba05-03fa-45af-a47e-8efabe261e5b/download/master-dataset-data-dictionary.csv> |  |

### Data View:

* **Schools (table)**

**Step1:** Imported master\_dataset.csv and created a new table called Schools(table)

**Step2:** Data profiling (column profiling) was performed against the ‘**entire data set’** to provide more accurate and comprehensive view of the data.



**Step3:** Based on the Power BI data profiling, following discoveries were ascertained against the main dataset.

Note: For all columns, Error % value were 0%

|  |  |  |  |
| --- | --- | --- | --- |
| **field\_name** | **Valid %** | **Empty %** | **Review comments** |
| School\_code | 100 | 0 | * 2216 distinct, 2216 unique.   Therefore, this field can be used as a primary key for the main dataset. |
| AgeID | 99 | 1 | * 2215 distinct, 2215 unique, 2 empty.   Empty values may require an additional investigation |
| School\_name | 100 | 0 |  |
| Street | 100 | 0 |  |
| Town\_suburb | 99 | 1 | * 1064 distinct, 1256 unique, 20 empty.   Empty values may require an additional investigation.   * Some column values are in proper case and some column values are all capitalised. May require further review/data update. |
| Postcode | 100 | 0 |  |
| Phone | 99 | 1 | * 1064 distinct, 1256 unique, 20 empty.   Empty values may require an additional investigation. |
| School\_Email | 99 | 1 | * 1 Empty string.   Empty values may require an additional investigation. |
| Website | 100 | 0 |  |
| Fax | 98 | 2 | * 99 Empty String.   Empty values may require an additional investigation. |
| latest\_year\_enrolment\_FTE | 100 | 0 | * 50 Empty.   Empty values may require an additional investigation.   * Also, some values were in decimal point. E.g. 73.4. This needs to e reviewed and updated. |
| Indigenous\_pct | 100 | 0 | * Includes double type numbers and string ‘np’ as values. May need to convert from varchar to double/integer and clean up ‘np’ values to 0. Additional investigation is required as it should hold percentage values. |
| LBOTE\_pct | 97 | 3 | * Includes double type numbers and string ‘np’ as values. Imported as Text, however should be type double/integer and clean up ‘np’ values to 0. Additional investigation is required as it should hold percentage values. |
| ICSEA\_value | 97 | 3 | * 7 Empty.   Empty values may require an additional investigation. |
| Level\_of\_schooling | 100 | 0 |  |
| Selective\_school | 100 | 0 |  |
| Opportunity\_class | 100 | 0 |  |
| School\_specialty\_type | 100 | 0 |  |
| School\_subtype | 100 | 0 |  |
| Support\_classes | 0 | 100 | * Based on the master-data-set-dictionary: Currently not available. |
| Preschool\_ind | 100 | 0 |  |
| Distance\_education | 100 | 0 |  |
| Intensive\_english\_centre | 100 | 0 |  |
| School\_gender | 100 | 0 |  |
| Late\_opening\_school | 100 | 0 |  |
| Date\_1st\_teacher | 99 | < 0 | * 1 Empty String.   Empty values may require an additional investigation. |
| LGA | 100 | 0 |  |
| Electorate\_from\_2023 | 99 | < 1 | * Unable to find this field in the master-data-set-dictionary. Only ‘Electorate’ field is found in the master-data-set-dictionary. Assume that this is an additional column and may need to confirm. * 1 Empty string. Empty values may require an additional investigation. |
| Electorate\_2015\_2022 | 99 | < 1 | * Unable to find this field in the master-data-set-dictionary. Only ‘Electorate’ field is found in the master-data-set-dictionary. Assume that this is same as ‘Electorate’, but may need to confirm. * 8 Empty String: Empty values may require an additional investigation. |
| Fed\_electorate | 99 | < 1 | * 8 Empty String: Empty values may require an additional investigation. |
| Operational\_directorate | 99 | < 1 | * 1 Empty String: Empty values may require an additional investigation. |
| Principal\_network | 100 | 0 |  |
| Operational\_directorate\_office | 99 | < 1 | * 1 Empty String: Empty values may require an additional investigation. |
| Operational\_directorate\_office\_phone | 99 | < 1 | * 1 Empty String: Empty values may require an additional investigation |
| Operational\_directorate\_office\_address | 99 | < 1 | * 1 Empty String: Empty values may require an additional investigation |
| FACS\_district | 99 | < 1 | * 10 Empty String: Empty values may require an additional investigation |
| Local\_health\_district | 99 | < 1 | * 10 Empty String: Empty values may require an additional investigation |
| AECG\_region | 99 | < 1 | * 24 Empty String: Empty values may require an additional investigation |
| ASGS\_remoteness | 100 | 0 |  |
| Latitude | 100 | 0 |  |
| Longitude | 100 | 0 |  |
| Assets unit | 98 | 2 | * 38 Empty String: Empty values may require an additional investigation |
| SA4 | 99 | < 1 | * 1 Empty String: Empty values may require an additional investigation |
| Healthy canteen |  |  | * Unable to find this column in the master-data-set. Only found in master-data-set-dictionary. |
| Date\_extracted | 100 | 0 |  |

* **Dates (table)**

**Step1:** Manually created a **Dates** table (date range from year 1910 to year 2023) to support report requirements. Only Year column was added to support the visuals.

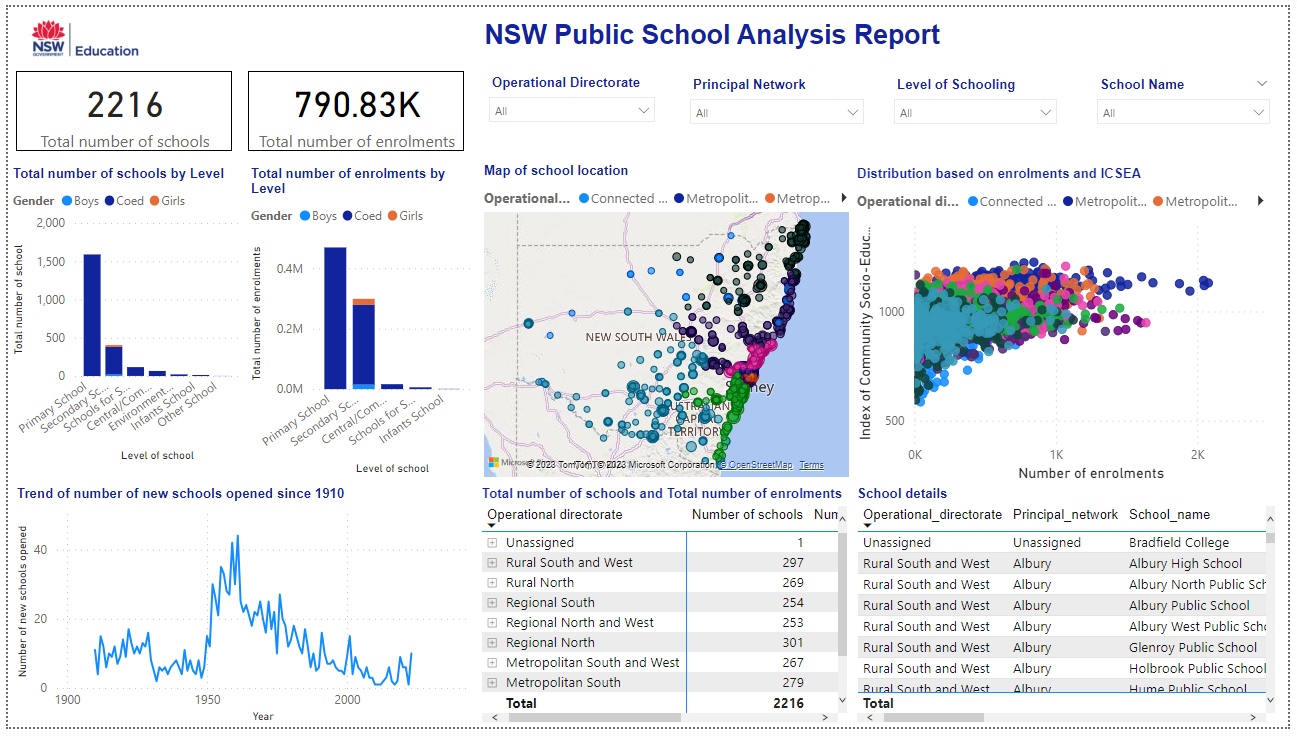
If required, other date related columns can be added at a later state and also, the date range can be modified in the table.

### Model View

1 – M relationship is created between the **Dates(table)** and **Schools(table)**

### Report View

Below is a screen image of the NSW Public School Analysis Report (created using Power BI desktop).



1. Slicers are included, allowing the visuals to be filtered based on the following fields:

(Dropdown type slicers are added to save screen pace):

* 1. Operational\_directorate
  2. Principal\_network
  3. School\_name
  4. Level\_of\_schooling

1. Card visualisation is included:
   1. Total number of schools
   2. Total number of enrolments (used latest\_year\_enrolment\_FTE)
2. Break-up of Total number of schools and Total number of enrolments by
   1. Level\_of\_schooling
   2. School\_gender (Included as Legend)
3. Trend of number of new schools opened each year since 1910 (as a Line chart)
4. Map view of the school’s locations with
   1. School’s enrolment numbers determining the bubble size.
   2. Operational\_directorate on colour
   3. Date\_1st\_teacher on tooltip
5. Scatter chart showing distribution of schools displaying relationship between enrolments and school’s ICSEA\_value.
   1. Operational\_directorate on colour
   2. Date\_1st\_teacher on tooltip
6. Matrix visualisation for Total number of schools and Total number of enrolments with the below fields in the rows in the order mentioned.
   1. Operational\_directorate
   2. Principal\_network
   3. School\_name
7. Table view of schools with the following fields in the order mentioned.
   1. Operational\_directorate
   2. Principal\_network
   3. School\_name
   4. Website
   5. School code
   6. Level of schooling
   7. School gender
   8. Date\_1st\_teacher (or simplicity, Year column from Dates table is displayed instead)
   9. ICSEA\_value
   10. latest\_year\_enrolment\_FTE

### Assumptions

* Single fact table called ‘Schools’ was created from the master\_dataset csv file. If the table size needs to be reduced to improve the report performance, then ‘Schools’ table can be normalised to create separate dimension tables E.g. lookup table for Operational\_directorate , Principal\_network, Operational\_directorate\_office …etc
* All empty columns values and incorrect values are remained unchanged at this stage. If required, these values can be reviewed and updated with the correct values via the ‘Power Query’ editor in Power BI. Data corrections should provide more accurate values and comprehensive reports.
* All columns from the master\_dataset csv file is included for now. If required, some of the columns that are not required in the visuals can be removed. E.g. School Phone, Date\_Extracted …etc
* All column names from the master\_dataset csv file is left unchanged. If required, some of the column names can be replaced with more intuitive names.
* For simplicity, all visuals are included in a single report tab. If required, the layout of the visuals can be rearranged and spread across multiple report tabs.