

Data Cleansing Record

1. Data Sources

The dataset is the combination of 03 datasets:

- **Population of Australia:** <https://www.abs.gov.au/statistics/people/population>
- **Crash Data Queensland (Data file 1 and e):**
<https://www.kaggle.com/datasets/joebeachcapital/crash-data-queensland-australia-2001-2022>

The join between these datasets was conducted using the Year column as the primary key.

2. Exploratory Data Analysis (EDA)

Initially, an fast Exploratory Data Analysis (EDA) was conducted using Tableau Prep Builder. During the process, missing data and potential insights were discovered.

3. Data Cleansing Process

Removal of "Property Damage Only" Cases : The crash data initially contained records classified as "Property damage only" for the years between 2001 and 2010. However, due to a policy change after 2010, these cases were no longer recorded. This category represents minor incidents that only involve with vehicle damage, low force crashes, more than a decade old and not significant to the public safety. Therefore, to maintain consistency of the analysis, all "Property damage only" cases will be drop from the dataset. Even though the number of such cases was high, their exclusion is not affecting the overall findings.

Tools Used: For data cleansing, Tableau Desktop was used. Specifically, the filtering and grouping features of Tableau were utilized for:

- **The filter feature** was used to remove the "Property damage only" severity cases so the impact of those cases on the analysis can be observed.
- **The group feature** is to categorize and organize crash conditions for analysis and visualization.