# EDA_toilet_facilities_09

## Group K

### 2025-09-09

## 1. Loading the dataset

### Dislay the first 5 rows

Table 1: First 5 rows of the Access to Health Care Dataset (excluding first row)

| ISO3 | DataId | Indicator | Value | CountryName | SurveyYear |
|------|--------|-----------|-------|-------------|------------|
| ZAF | 795762 | Households with an improved sanitation facility | 50.1 | South Africa | 1998 |
| ZAF | 795768 | Households with an unimproved sanitation facility | 38.3 | South Africa | 1998 |
| ZAF | 795760 | Households with a pit latrine without a slab or an open pit | 31.2 | South Africa | 1998 |
| ZAF | 795764 | Households with a bucket toilet | 6 | South Africa | 1998 |
| ZAF | 795765 | Households using open defecation | 11.6 | South Africa | 1998 |

## 2. Data Overview

### Summary statistics

```
## Warning: There were 4 warnings in `summarise()`.
## The first warning was:
## i In argument: `across(...)`.
## Caused by warning in `min()`:
## ! no non-missing arguments to min; returning Inf
## i Run `dplyr::last_dplyr_warnings()` to see the 3 remaining warnings.
```

Table 2: Summary Statistics for Key Numeric Variables in the toilet_facilities Dataset

| Variable | Statistic | Value |
|----------|-----------|-------|
| CIHigh | mean | NaN |
| CIHigh | sd | NA |

| Variable | Statistic | Value |
|---|---|---|
| CIHigh | min | Inf |
| CIHigh | max | -Inf |
| CILow | mean | NaN |
| CILow | sd | NA |
| CILow | min | Inf |
| CILow | max | -Inf |
| DenominatorUnweighted | mean | 27495.238 |
| DenominatorUnweighted | sd | 16938.786 |
| DenominatorUnweighted | min | 11083.000 |
| DenominatorUnweighted | max | 52465.000 |
| DenominatorWeighted | mean | 27185.143 |
| DenominatorWeighted | sd | 16667.463 |
| DenominatorWeighted | min | 11083.000 |
| DenominatorWeighted | max | 52007.000 |
| Value | mean | 4942.057 |
| Value | sd | 13082.679 |
| Value | min | 0.400 |
| Value | max | 52465.000 |

# 3. Data Quality

## Missing values per column

```
##                  ISO3                DataId              Indicator
##                     0                     0                      0
##                 Value             Precision        DHS_CountryCode
##                     0                     0                      1
##           CountryName            SurveyYear               SurveyId
##                     0                     0                      0
##           IndicatorId        IndicatorOrder          IndicatorType
##                     0                     1                      1
##        CharacteristicId    CharacteristicOrder CharacteristicCategory
##                     1                     1                      1
##     CharacteristicLabel          ByVariableId         ByVariableLabel
##                     1                     0                     46
##               IsTotal           IsPreferred                  SDRID
##                     1                     1                      1
##              RegionId        SurveyYearLabel             SurveyType
##                    47                     1                      1
##   DenominatorWeighted DenominatorUnweighted                  CILow
##                     5                     5                     47
##                CIHigh             LevelRank
##                    47                    47
```

Dupilcates

# 4. Visualization

**Toilet Facilities Indicators in South Africa (2016)**