

# EDA\_IYCF\_12

Group K

2025-09-09

## 1. Loading the dataset

Display the first 5 rows

Table 1: First 5 rows of the dataset

ISO3	DataId	Indicator	Value	CountryName	SurveyYear
ZAF	795971	Children ever breastfed	87.4	South Africa	1998
ZAF	795973	Children who started breastfeeding within 1 hour of birth	38.9	South Africa	1998
ZAF	621666	Children exclusively breastfed	6.9	South Africa	1998
ZAF	621667	Children breastfeeding and consuming plain water only	6.3	South Africa	1998
ZAF	621670	Children breastfeeding and consuming complementary foods	40.9	South Africa	1998

## 2. Data Overview

Summary statistics

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

Variable	Statistic	Value
CIHigh	mean	NaN
CIHigh	sd	NA
CIHigh	min	Inf
CIHigh	max	-Inf
CILow	mean	NaN
CILow	sd	NA
CILow	min	Inf
CILow	max	-Inf
DenominatorUnweighted	mean	755.70000
DenominatorUnweighted	sd	539.48992
DenominatorUnweighted	min	346.00000

Variable	Statistic	Value
DenominatorUnweighted	max	2041.00000
DenominatorWeighted	mean	750.40000
DenominatorWeighted	sd	533.80239
DenominatorWeighted	min	345.00000
DenominatorWeighted	max	2010.00000
Value	mean	41.19545
Value	sd	27.13407
Value	min	6.30000
Value	max	87.40000

### 3. Data Quality

#### Missing values per column

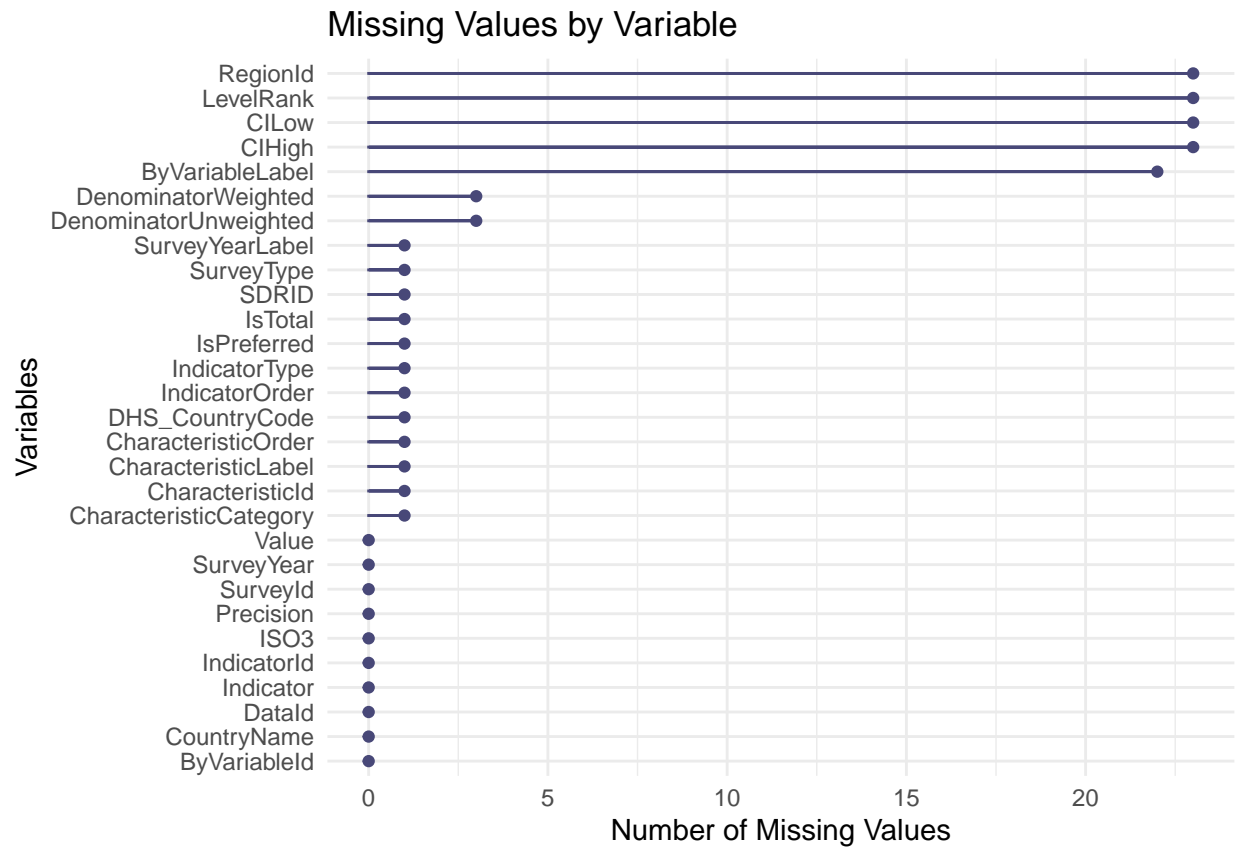
```
##          ISO3          DataId          Indicator
##          0          0          0
##          Value          Precision          DHS_CountryCode
##          0          0          1
##          CountryName          SurveyYear          SurveyId
##          0          0          0
##          IndicatorId          IndicatorOrder          IndicatorType
##          0          1          1
##          CharacteristicId          CharacteristicOrder          CharacteristicCategory
##          1          1          1
##          CharacteristicLabel          ByVariableId          ByVariableLabel
##          1          0          22
##          IsTotal          IsPreferred          SDRID
##          1          1          1
##          RegionId          SurveyYearLabel          SurveyType
##          23          1          1
##          DenominatorWeighted          DenominatorUnweighted          CILow
##          3          3          23
##          CIHigh          LevelRank
##          23          23
```

#### Dupilcates

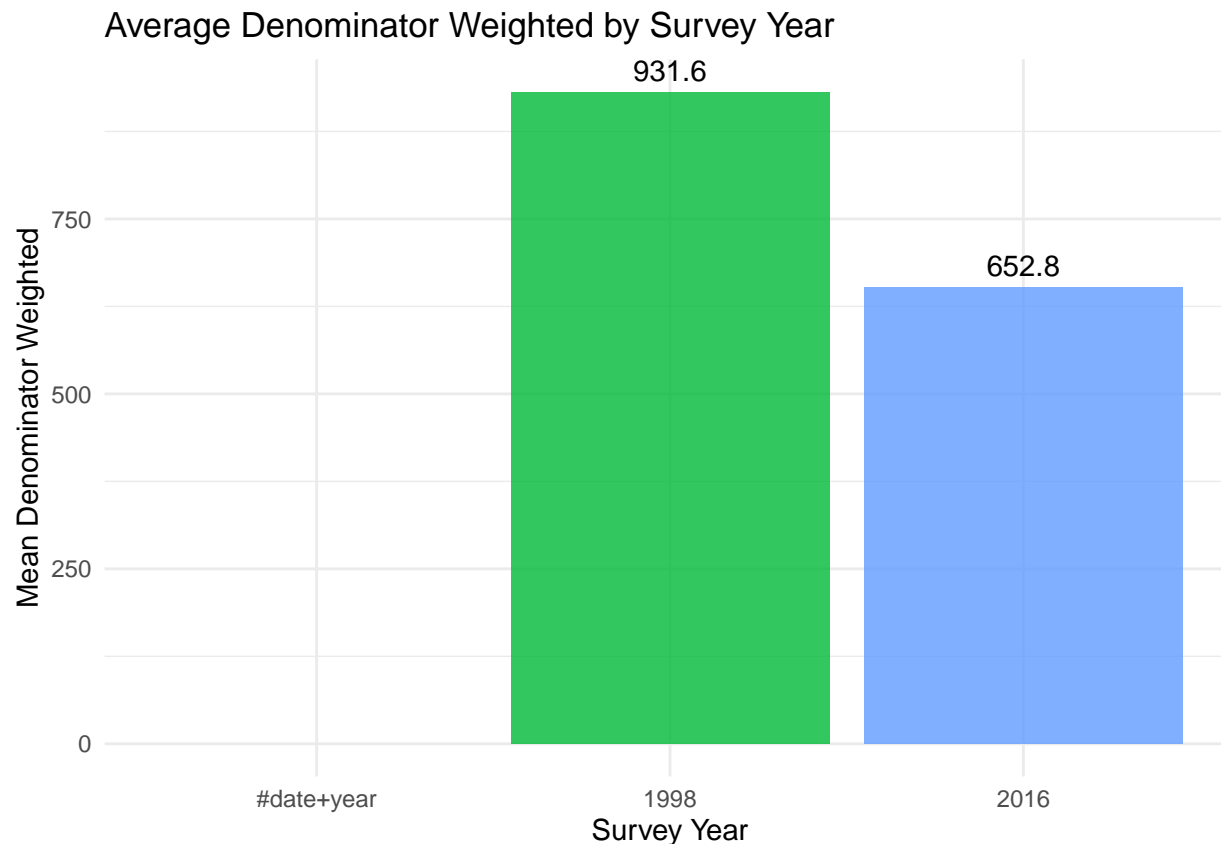
```
## [1] 0
```

## 4. Visualizations

Plot showing which columns have missing values across the 23 rows



Bar Plot to compare overall survey size between 1998 and 2016.



## 6. Data dictionary of all Datasets

Table 3: Data Dictionary for the DHS Maternal Mortality Dataset

Column	DataType	Description
ISO3	character	ISO3 country code
DataId	character	Unique ID for the dataset row
Indicator	character	Name of the health indicator
Value	character	Numeric value of the indicator (currently character)
Precision	character	Number of decimal places or precision
DHS_CountryCode	character	Country code from DHS dataset
CountryName	character	Full country name
SurveyYear	character	Year survey was conducted
SurveyId	character	Unique ID for the survey
IndicatorId	character	DHS indicator code
IndicatorOrder	numeric	Order of indicator in dataset
IndicatorType	character	Type of indicator (I,N,D,U,C)
CharacteristicId	numeric	ID for the population characteristic
CharacteristicOrder	numeric	Order of the characteristic
CharacteristicCategory	character	Category of the population (e.g., Total 15-49)
CharacteristicLabel	character	Label describing the population characteristic

Column	DataType	Description
ByVariableId	character	ID for any breakdown variable
ByVariableLabel	character	Label for breakdown variable
IsTotal	numeric	1 if row represents a total, 0 otherwise
IsPreferred	numeric	1 if this is the preferred row for this indicator, 0 otherwise
SDRID	character	DHS short code for the indicator
RegionId	logical	Region identifier (mostly NA)
SurveyYearLabel	numeric	Year label for survey (redundant with SurveyYear)
SurveyType	character	Type of survey (e.g., DHS)
DenominatorWeighted	numeric	Weighted denominator used to calculate indicator
DenominatorUnweighted	numeric	Unweighted denominator used to calculate indicator
CILow	numeric	Lower bound of the confidence interval
CIHigh	numeric	Upper bound of the confidence interval
LevelRank	logical	Level rank (mostly NA; optional hierarchy indicator)