

# EDA\_Symptoms\_Of\_Acute\_Respiratory\_Infection\_11

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	598577	Children with symptoms of ARI	21.9	South Africa	1998
ZAF	397915	Children with symptoms of ARI	19.3	South Africa	1998
ZAF	598578	Number of children born in the last five (or three) years	2912	South Africa	1998
ZAF	384931	Number of children born in the last five (or three) years	4740	South Africa	1998
ZAF	139860	Number of children born in the last five (or three) years (unweighted)	2958	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	1654.722
DenominatorUnweighted	sd	1680.379
DenominatorUnweighted	min	59.000

Variable	Statistic	Value
DenominatorUnweighted	max	4797.000
DenominatorWeighted	mean	1659.278
DenominatorWeighted	sd	1658.977
DenominatorWeighted	min	68.000
DenominatorWeighted	max	4740.000
Value	mean	1159.346
Value	sd	1573.418
Value	min	3.100
Value	max	4797.000

### 3. Data Quality

Find duplicated values

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

Inspect columns for missing/empty data as values and pecentages respectively

```
##          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              0
##          IsTotal          IsPreferred          SDRID
##          1              1              1
##          RegionId        SurveyYearLabel      SurveyType
##          27             1              1
##          DenominatorWeighted DenominatorUnweighted CILow
##          9              9              27
##          CIHigh          LevelRank
##          27             27
```

```
##          ISO3          DataId          Indicator
##          0.00          0.00          0.00
##          Value          Precision        DHS_CountryCode
##          0.00          0.00          3.70
##          CountryName      SurveyYear      SurveyId
##          0.00          0.00          0.00
##          IndicatorId      IndicatorOrder    IndicatorType
##          0.00          3.70          3.70
##          CharacteristicId  CharacteristicOrder CharacteristicCategory
##          3.70          3.70          3.70
```

##	CharacteristicLabel	ByVariableId	ByVariableLabel
##	3.70	0.00	0.00
##	IsTotal	IsPreferred	SDRID
##	3.70	3.70	3.70
##	RegionId	SurveyYearLabel	SurveyType
##	100.00	3.70	3.70
##	DenominatorWeighted	DenominatorUnweighted	CILow
##	33.33	33.33	100.00
##	CIHigh	LevelRank	
##	100.00	100.00	

### Check for unique values

##	IS03	DataId	Indicator
##	2	27	8
##	Value	Precision	DHS_CountryCode
##	27	3	2
##	CountryName	SurveyYear	SurveyId
##	2	3	3
##	IndicatorId	IndicatorOrder	IndicatorType
##	8	8	4
##	CharacteristicId	CharacteristicOrder	CharacteristicCategory
##	2	2	2
##	CharacteristicLabel	ByVariableId	ByVariableLabel
##	2	3	3
##	IsTotal	IsPreferred	SDRID
##	2	3	8
##	RegionId	SurveyYearLabel	SurveyType
##	1	3	2
##	DenominatorWeighted	DenominatorUnweighted	CILow
##	9	9	1
##	CIHigh	LevelRank	
##	1	1	