EDA_Covid_19_Prevention_National_04

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	795844	Population using an improved water source	83.5	South Africa	1998
ZAF	795750	Population using water piped into dwelling	36	South Africa	1998
ZAF	795755	Population using water piped into yard/plot	23.1	South Africa	1998
ZAF	795740	Population using a public tap/standpipe	19.3	South Africa	1998
ZAF	795744	Population with improved water source on the premises	60.3	South Africa	1998

2. Data Overview

Summary statistics

Table 2: Summary Statistics for Numeric Variables in the Dataset

Variable	Statistic	Value
CharacteristicId	mean	1.000000e+03
CharacteristicId	sd	0.000000e+00
CharacteristicId	min	1.000000e+03
CharacteristicId	max	1.000000e+03
CharacteristicOrder	mean	0.000000e+00
CharacteristicOrder	sd	0.000000e+00
CharacteristicOrder	min	0.000000e+00
CharacteristicOrder	max	0.000000e+00
DenominatorUnweighted	mean	3.935287e+04
DenominatorUnweighted	sd	1.276580e+04
DenominatorUnweighted	min	1.106600e+04
DenominatorUnweighted	max	5.246500e+04
DenominatorWeighted	mean	3.881516e+04
DenominatorWeighted	sd	1.265890e+04

Variable	Statistic	Value
DenominatorWeighted	min	1.106600e+04
DenominatorWeighted	max	5.200700e+04
IndicatorOrder	mean	2.520402e+08
IndicatorOrder	sd	4.011155e+06
IndicatorOrder	min	2.501620e+08
IndicatorOrder	max	2.608311e+08
IsPreferred	mean	1.000000e+00
IsPreferred	sd	0.000000e+00
IsPreferred	min	1.000000e+00
IsPreferred	max	1.000000e+00
IsTotal	mean	1.000000e+00
IsTotal	sd	0.000000e+00
IsTotal	min	1.000000e+00
IsTotal	max	1.000000e+00
SurveyYearLabel	mean	2.008588e+03
SurveyYearLabel	sd	8.991975e+00
SurveyYearLabel	min	1.998000e+03
SurveyYearLabel	max	2.016000e+03

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
##                    1                     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                      34
##              IsTotal           IsPreferred                   SDRID
##                    1                     1                       1
##             RegionId       SurveyYearLabel              SurveyType
##                   35                     1                       1
## DenominatorWeighted DenominatorUnweighted                   CILow
##                    3                     3                      35
##               CIHigh             LevelRank
##                   35                    35


##                 ISO3                DataId               Indicator
```

2

```
##                  0.00                0.00                0.00
##                 Value           Precision     DHS_CountryCode
##                  2.86                0.00                2.86
##           CountryName          SurveyYear            SurveyId
##                  0.00                0.00                0.00
##           IndicatorId       IndicatorOrder       IndicatorType
##                  0.00                2.86                2.86
##        CharacteristicId CharacteristicOrder CharacteristicCategory
##                  2.86                2.86                2.86
##       CharacteristicLabel         ByVariableId       ByVariableLabel
##                  2.86                0.00               97.14
##                IsTotal         IsPreferred               SDRID
##                  2.86                2.86                2.86
##              RegionId      SurveyYearLabel          SurveyType
##                100.00                2.86                2.86
##     DenominatorWeighted DenominatorUnweighted               CILow
##                  8.57                8.57              100.00
##                CIHigh           LevelRank
##                100.00              100.00
```
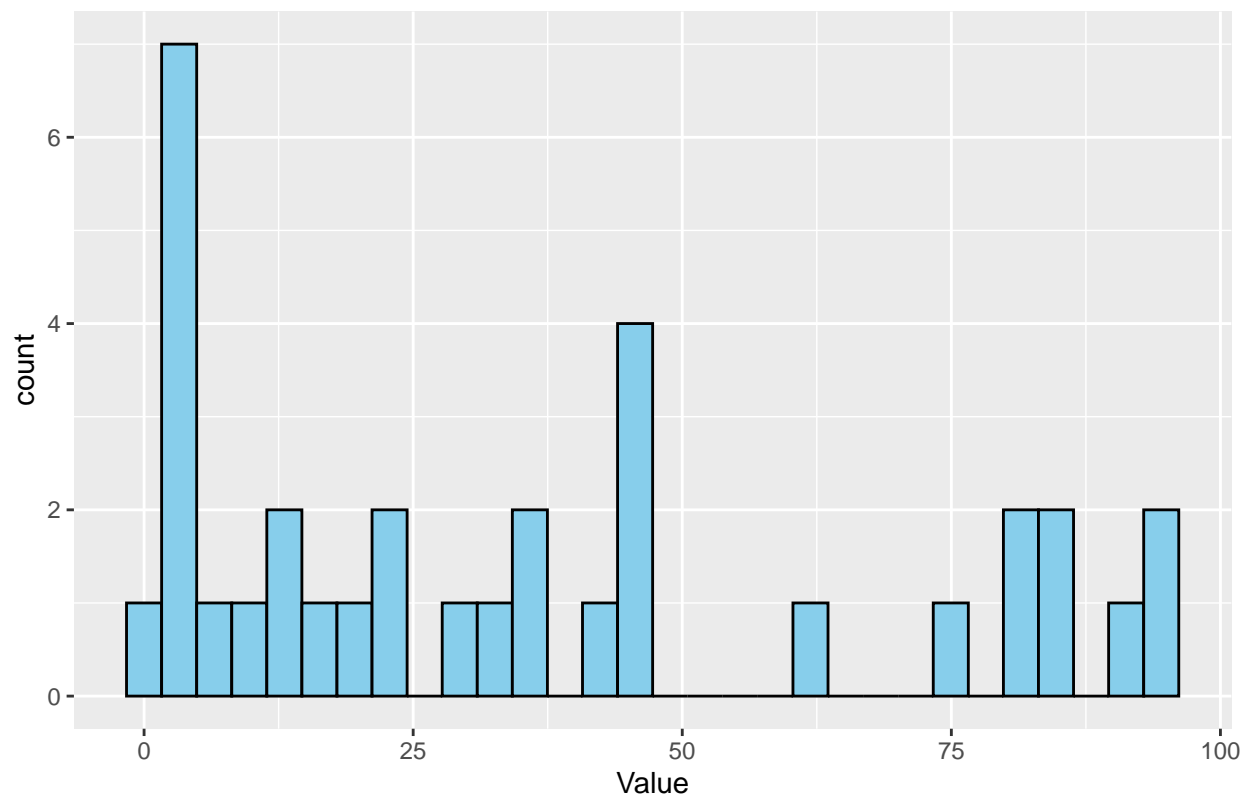
Check for unique values

```
##                   ISO3              DataId           Indicator
##                      2                  35                  21
##                  Value           Precision     DHS_CountryCode
##                     35                   3                   2
##            CountryName          SurveyYear            SurveyId
##                      2                   3                   3
##            IndicatorId      IndicatorOrder       IndicatorType
##                     21                  21                   2
##        CharacteristicId CharacteristicOrder CharacteristicCategory
##                      2                   2                   2
##       CharacteristicLabel        ByVariableId       ByVariableLabel
##                      2                   2                   2
##                IsTotal         IsPreferred               SDRID
##                      2                   2                  21
##               RegionId     SurveyYearLabel          SurveyType
##                      1                   3                   2
##     DenominatorWeighted DenominatorUnweighted               CILow
##                      9                   9                   1
##                 CIHigh           LevelRank
##                      1                   1
```

4. Visualizations

Distribution of Indicator Values

Outlier check for Values