

EDA_Access_To_Healthcare_National_01

Group K

2025-09-09

1. Loading the dataset

Display 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	751751	Antenatal care provider: Doctor	28.5	South Africa	1998
ZAF	567476	Antenatal care provider: Doctor	30	South Africa	1998
ZAF	205488	Antenatal care provider: Doctor	27.3	South Africa	1998
ZAF	751748	Antenatal care provider: Nurse/midwife	66.6	South Africa	1998
ZAF	567472	Antenatal care provider: Nurse/midwife	65	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	2.062166e+03
DenominatorUnweighted	sd	1.445142e+03
DenominatorUnweighted	min	5.900000e+01
DenominatorUnweighted	max	5.066000e+03
DenominatorWeighted	mean	2.048647e+03
DenominatorWeighted	sd	1.428928e+03
DenominatorWeighted	min	6.800000e+01

Variable	Statistic	Value
DenominatorWeighted	max	4.992000e+03
IndicatorOrder	mean	8.712642e+07
IndicatorOrder	sd	5.002221e+06
IndicatorOrder	min	8.336301e+07
IndicatorOrder	max	9.409617e+07
IsPreferred	mean	4.218182e-01
IsPreferred	sd	4.947501e-01
IsPreferred	min	0.000000e+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	2.007622e+03
SurveyYearLabel	sd	8.994862e+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
##          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              13
##          IsTotal          IsPreferred          SDRID
##          1              1              1
##          RegionId          SurveyYearLabel          SurveyType
##          276              1              1
##          DenominatorWeighted  DenominatorUnweighted          CILow
##          35              35              276
##          CIHigh          LevelRank
##          276              276

##          ISO3          DataId          Indicator
##          0.00          0.00          0.00
```

##	Value	Precision	DHS_CountryCode
##	0.00	0.00	0.36
##	CountryName	SurveyYear	SurveyId
##	0.00	0.00	0.00
##	IndicatorId	IndicatorOrder	IndicatorType
##	0.00	0.36	0.36
##	CharacteristicId	CharacteristicOrder	CharacteristicCategory
##	0.36	0.36	0.36
##	CharacteristicLabel	ByVariableId	ByVariableLabel
##	0.36	0.00	4.71
##	IsTotal	IsPreferred	SDRID
##	0.36	0.36	0.36
##	RegionId	SurveyYearLabel	SurveyType
##	100.00	0.36	0.36
##	DenominatorWeighted	DenominatorUnweighted	CILow
##	12.68	12.68	100.00
##	CIHigh	LevelRank	
##	100.00	100.00	

Check for unique values

##	ISO3	DataId	Indicator
##	2	276	69
##	Value	Precision	DHS_CountryCode
##	191	3	2
##	CountryName	SurveyYear	SurveyId
##	2	3	3
##	IndicatorId	IndicatorOrder	IndicatorType
##	73	73	6
##	CharacteristicId	CharacteristicOrder	CharacteristicCategory
##	2	2	2
##	CharacteristicLabel	ByVariableId	ByVariableLabel
##	2	5	5
##	IsTotal	IsPreferred	SDRID
##	2	3	73
##	RegionId	SurveyYearLabel	SurveyType
##	1	3	2
##	DenominatorWeighted	DenominatorUnweighted	CILow
##	27	27	1
##	CIHigh	LevelRank	
##	1	1	

4. Visualizations

Plot: Distribution of Values by Survey Year

```
## # A tibble: 6 x 29
##   ISO3 DataId Indicator Value Precision DHS_CountryCode CountryName SurveyYear
##   <chr> <chr> <chr>   <dbl> <chr>    <chr>          <chr>    <fct>
## 1 ZAF  751751 Antenatal~ 28.5 1      ZA          South Afri~ 1998
## 2 ZAF  567476 Antenatal~ 30    1      ZA          South Afri~ 1998
```

```
## 3 ZAF 205488 Antenatal~ 27.3 1 ZA South Afri~ 1998
## 4 ZAF 751748 Antenatal~ 66.6 1 ZA South Afri~ 1998
## 5 ZAF 567472 Antenatal~ 65 1 ZA South Afri~ 1998
## 6 ZAF 205485 Antenatal~ 68.4 1 ZA South Afri~ 1998
## # i 21 more variables: SurveyId <chr>, IndicatorId <chr>, IndicatorOrder <dbl>,
## # IndicatorType <chr>, CharacteristicId <dbl>, CharacteristicOrder <dbl>,
## # CharacteristicCategory <chr>, CharacteristicLabel <chr>,
## # ByVariableId <chr>, ByVariableLabel <chr>, IsTotal <dbl>,
## # IsPreferred <dbl>, SDRID <chr>, RegionId <lgl>, SurveyYearLabel <dbl>,
## # SurveyType <chr>, DenominatorWeighted <dbl>, DenominatorUnweighted <dbl>,
## # CILow <lgl>, CIHigh <lgl>, LevelRank <lgl>
```

