HCES 2023-24

DATA QUALITY

REPORT
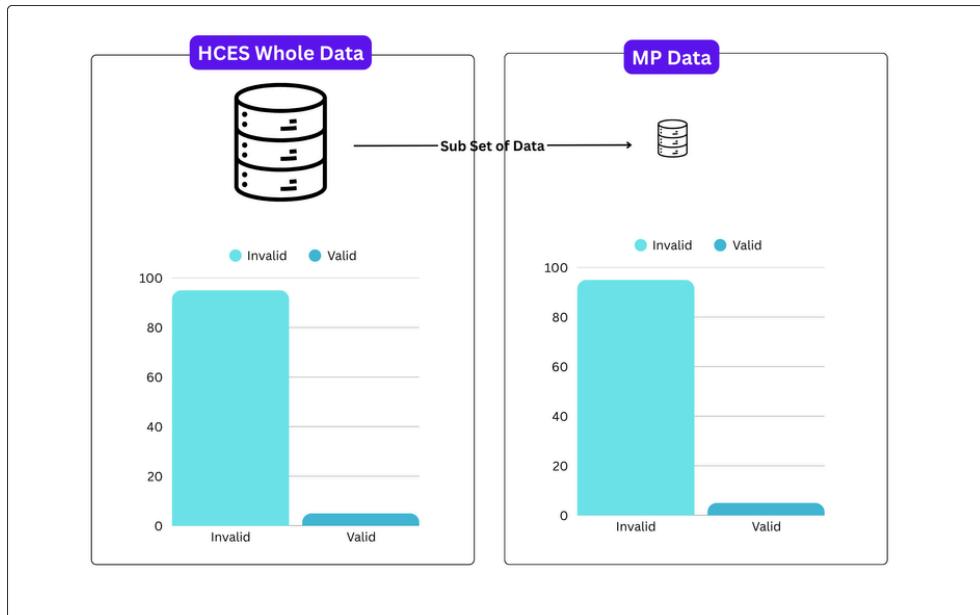
DATA IS EMPTY

Author: Rudra Prasad Bhuyan
Position: Data Scientist Intern

Organisation: SBC Labs
Guidance: Anushka Ashok, Alok Gangaramany

PREREQUIRES

FORMULA

$$\text{Official Valid record} = \frac{\text{Valid}}{\text{Total}} \times 100 \ \%$$

$$\text{Official Invalid record} = \frac{\text{Invalid}}{\text{Total}} \times 100$$

TOOLS : Pandas

FILE FORMAT: Parquet

GITHUB ISSUE: https://github.com/Rudra-G-23/rural-financial-inclusion-govt-scheme-recommendation/issues/1

NOTEBOOK LINK: https://github.com/Rudra-G-23/rural-financial-inclusion-govt-scheme-recommendation/blob/main/parquet-data/lev-07/notebooks/lev-07-04-check-null-value.ipynb

VIDEO LINK: https://youtu.be/OcsfaVl5mhA

Comparison with original and MP data for level 07

========== Total_other_items ==========

According to the official data
Total Official Recoded : 261953
Official Invalid record: 13429 --> 94.87%
Official Valid record: 248524 --> 5.13%

For our MP Data
Our MP dataset contains total record: 28210
Null Value in MP data: 27136
Null value Percentage in this MP data 96.19%

==

Total_other_items	Total no. of free other items received	^

Total no. of free other items received (Total_other_items)
Data file: **LEVEL - 07 (Section 4.2)**

Overview

Valid: **13429**	Type: **Continuous**
Invalid: **248524**	Decimal: **0**
Minimum: **1**	Width: **3**
Maximum: **50**	Range: **1 - 50**
Mean: **2.885**	Format: Numeric
Standard deviation: **2.952**	

Comparison with original and MP data for level 07

========== Num_fee_waiver_received ==========

According to the official data
Total Official Recoded : 261953
Official Invalid record: 16578 --> 93.67%
Official Valid record: 245375 --> 6.33%

For our MP Data
Our MP dataset contains total record: 28210
Null Value in MP data: 25646
Null value Percentage in this MP data 90.91%

==

Num_fee_waiver_ received	If Code 1 in Q4.2.6 Number of member received reimbursement/waiver ?	˄

If Code 1 in Q4.2.6 Number of member received reimbursement/waiver ? (Num_fee_waiver_received)

Data file: **LEVEL - 07 (Section 4.2)**

Overview

Valid: **16578**
Invalid: **245375**

Type: **Discrete**
Decimal: **0**
Width: **2**
Range: **1 - 9**
Format: Numeric

Questions and instructions

CATEGORIES

| Value	Category	Cases	
1		8990	54.2%
2		5577	33.6%
3		1556	9.4%

Comparison with original and MP data for level 07

========== Num_medical_beneficiaries ==========

According to the official data
Total Official Recoded : 261953
Official Invalid record: 4668 --> 98.22%
Official Valid record: 257285 --> 1.78%

For our MP Data
Our MP dataset contains total record: 28210
Null Value in MP data: 27782
Null value Percentage in this MP data 98.48%

===

Num_medical_be neficiaries	If yes in Q4.2.9 number of member received benifit ?	^

If yes in Q4.2.9 number of member received benifit ?
(Num_medical_beneficiaries)
Data file: **LEVEL - 07 (Section 4.2)**

Overview

Valid: **4668**	Type: **Discrete**
Invalid: **257285**	Decimal: **0**
	Width: **2**
	Range: **1 - 9**
	Format: **Numeric**

Questions and instructions

CATEGORIES

Value	Category	Cases	
1		4434	95%
2		180	3.9%
3		18	0.4%
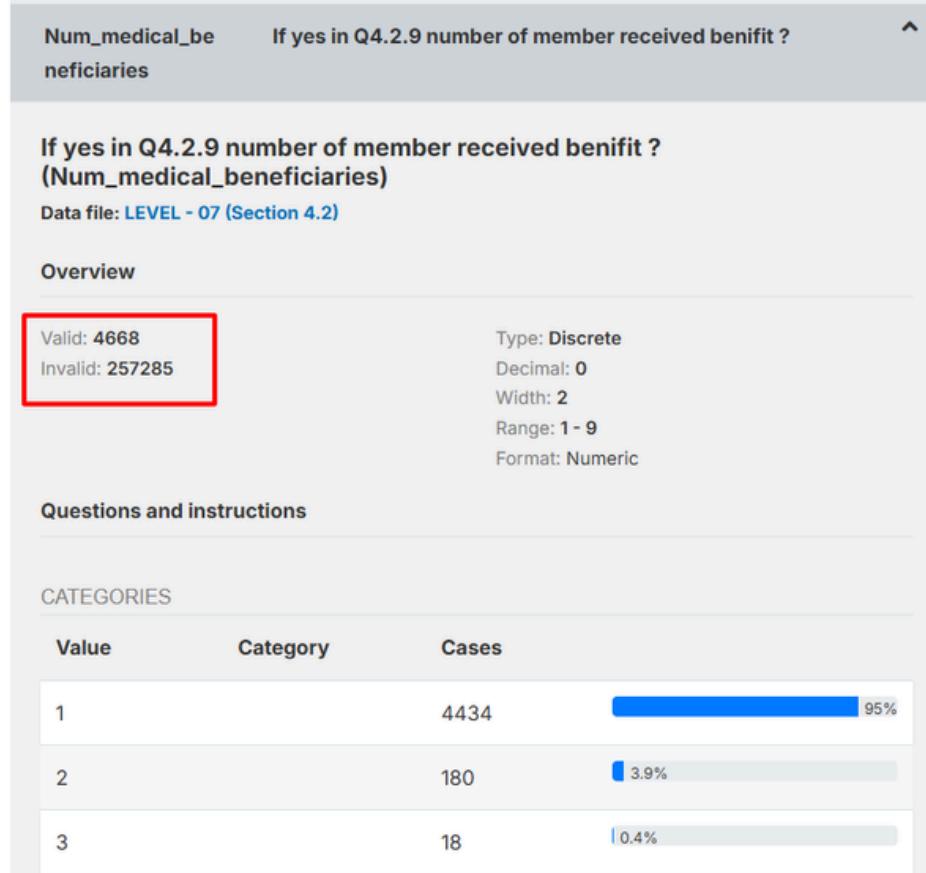
Comparison with original and MP data for level 07

========== Medical_benefit_amount ==========

According to the official data
Total Official Recoded : 261953
Official Invalid record: 4668 --> 98.22%
Official Valid record: 257285 --> 1.78%

For our MP Data
Our MP dataset contains total record: 28210
Null Value in MP data: 27782
Null value Percentage in this MP data 98.48%

===

Medical_benefit_ amount	If yes in Q4.2.9 Amount ?	^

If yes in Q4.2.9 Amount ? (Medical_benefit_amount)
Data file: **LEVEL - 07 (Section 4.2)**

Overview

Valid: **4668**	Type: **Continuous**
Invalid: **257285**	Decimal: **0**
Minimum: **0**	Width: **8**
Maximum: **1000000**	Range: **0 - 1000000**
Mean: **28576.168**	Format: Numeric
Standard deviation: **54939.066**	

Comparison with original and MP data for level 07

========== Online_purchase_education ==========

According to the official data
Total Official Recoded : 261953
Official Invalid record: 4600 --> 98.24%
Official Valid record: 257353 --> 1.76%
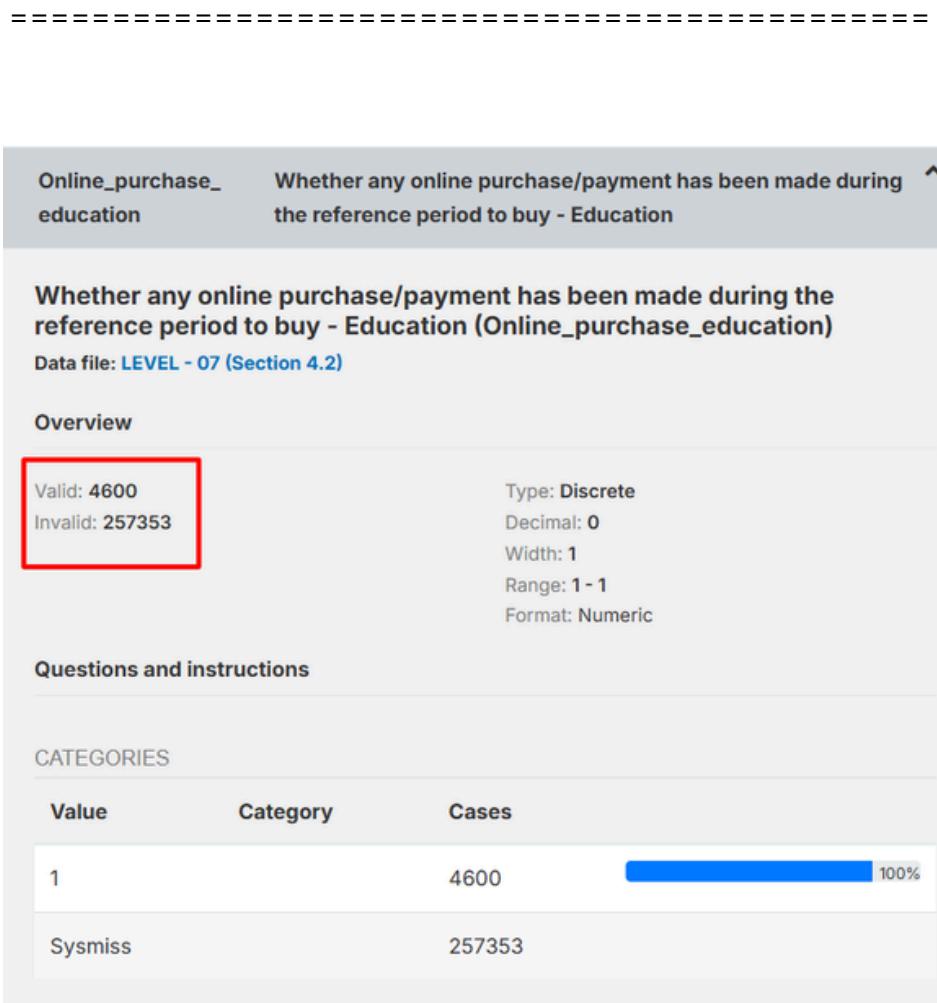
For our MP Data
Our MP dataset contains total record: 28210
Null Value in MP data: 27894
Null value Percentage in this MP data 98.88%

==

THANK YOU