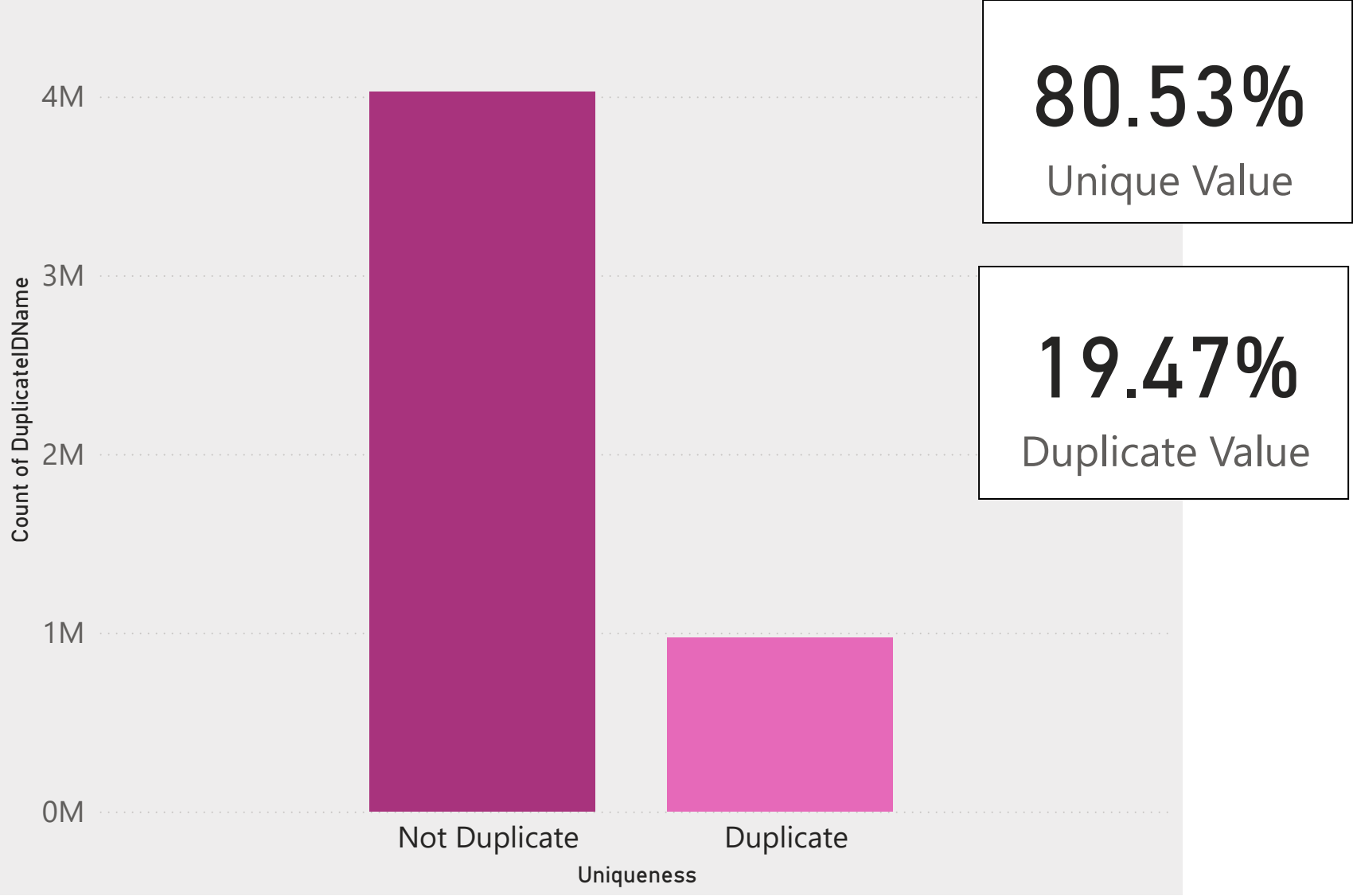


DATA QUALITY IN RAW HOSPITAL DATASET

UNIQUENESS

Duplication based on ID and Name



COMPLETENESS

Missing Value by Email

Non Missing Value Missing Value



100%

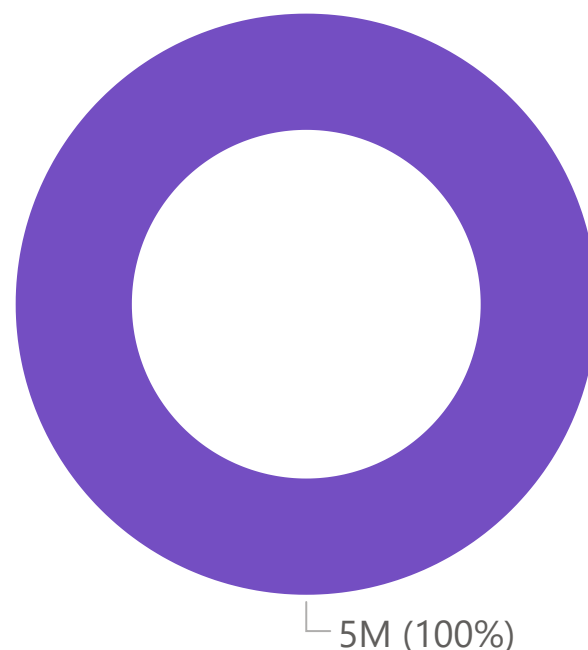
Non-Missing Value

0%

Missing Value

Missing Value by Name

Non Missing Value Missing Value



100%

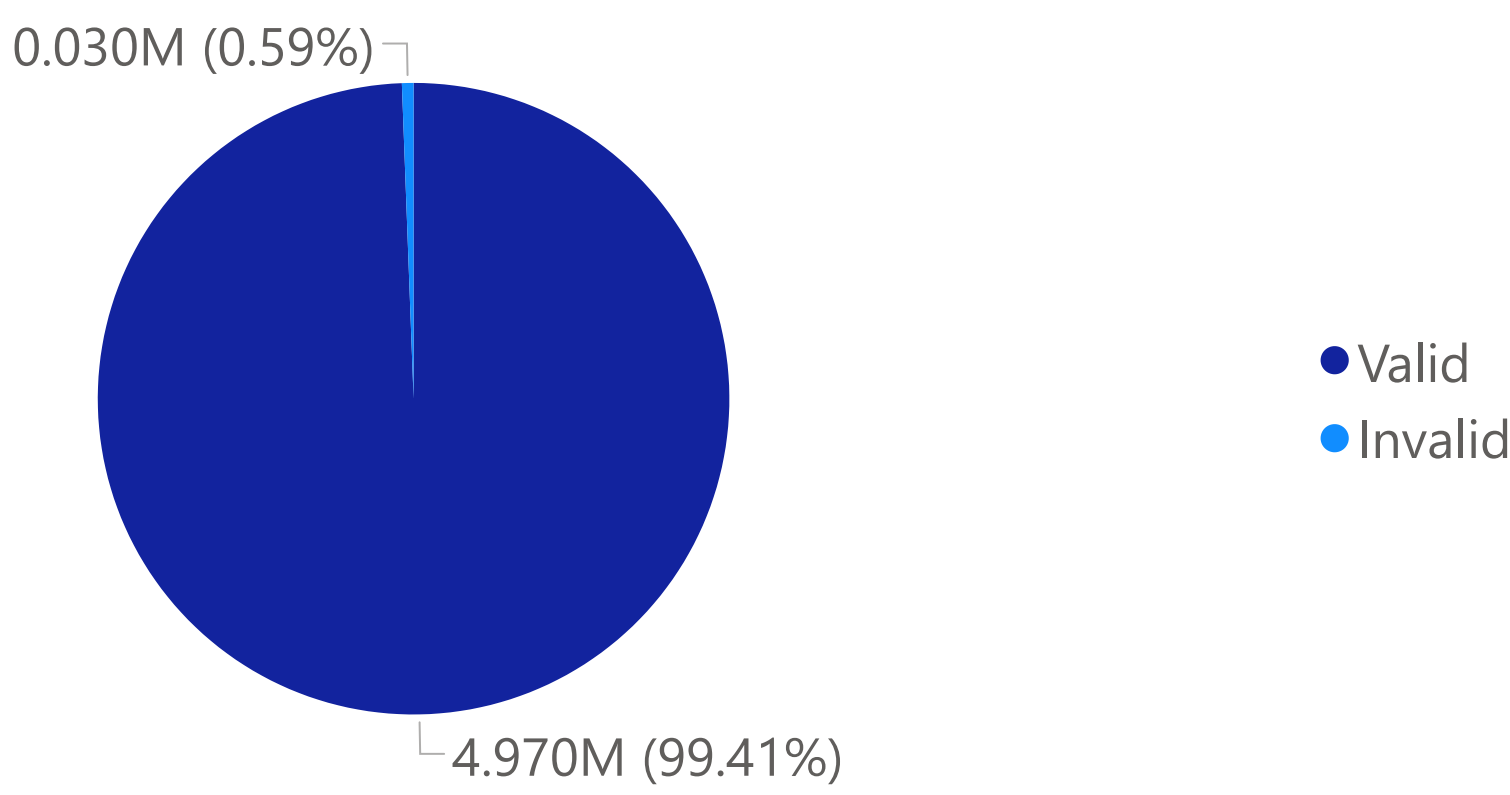
Non-Missing Value

0%

Missing Value

VALIDITY

Validity based on Age and Title



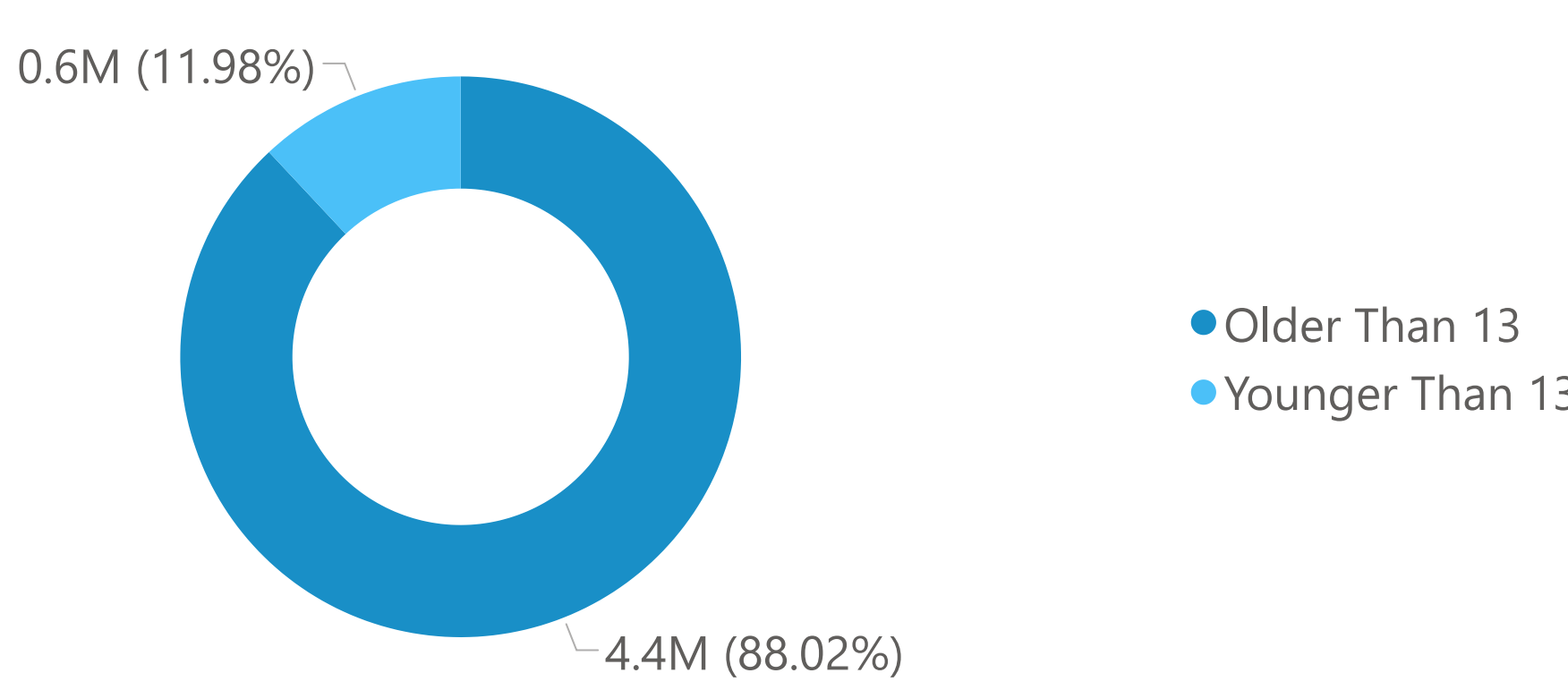
Value :

Invalid	Valid	Total
0.03M	4.97M	5M

Percentage :

Invalid	Valid	Total
0.59%	99.41%	100%

Validity based on Age and Email



Value :

Younger Than 13	Older Than 13	Total
0.60M	4.40M	5M

Percentage :

Younger Than 13	Older Than 13	Total
11.98%	88.02%	100%

ACCURACY

id	name	email	TrueCount
3283256	Amanda Phillips	amandaphillips@example.net	1
3814702	Andrew Moore	andrewmoore@example.com	1
2520629	Angela Williams	angelawilliams@example.com	1
2218632	Brandon Wilson	brandonwilson@example.net	2
379156	Cynthia Johnson	cynthiajohnson@example.com	1
3564905	Holly Dixon	hollydixon@example.net	1
938029	James Sanchez	james sanzhez@example.org	2
3825729	Jennifer Williams	jenniferwilliams@example.com	2
444892	Matthew Johnson	matthewjohnson@example.org	2
934077	Patrick Warren	patrickwarren@example.org	2
2584258	Robert Lewis	robertlewis@example.net	1
3645070	Robert Miller	robertmiller@example.net	2
Total			19

19

Accurate Count

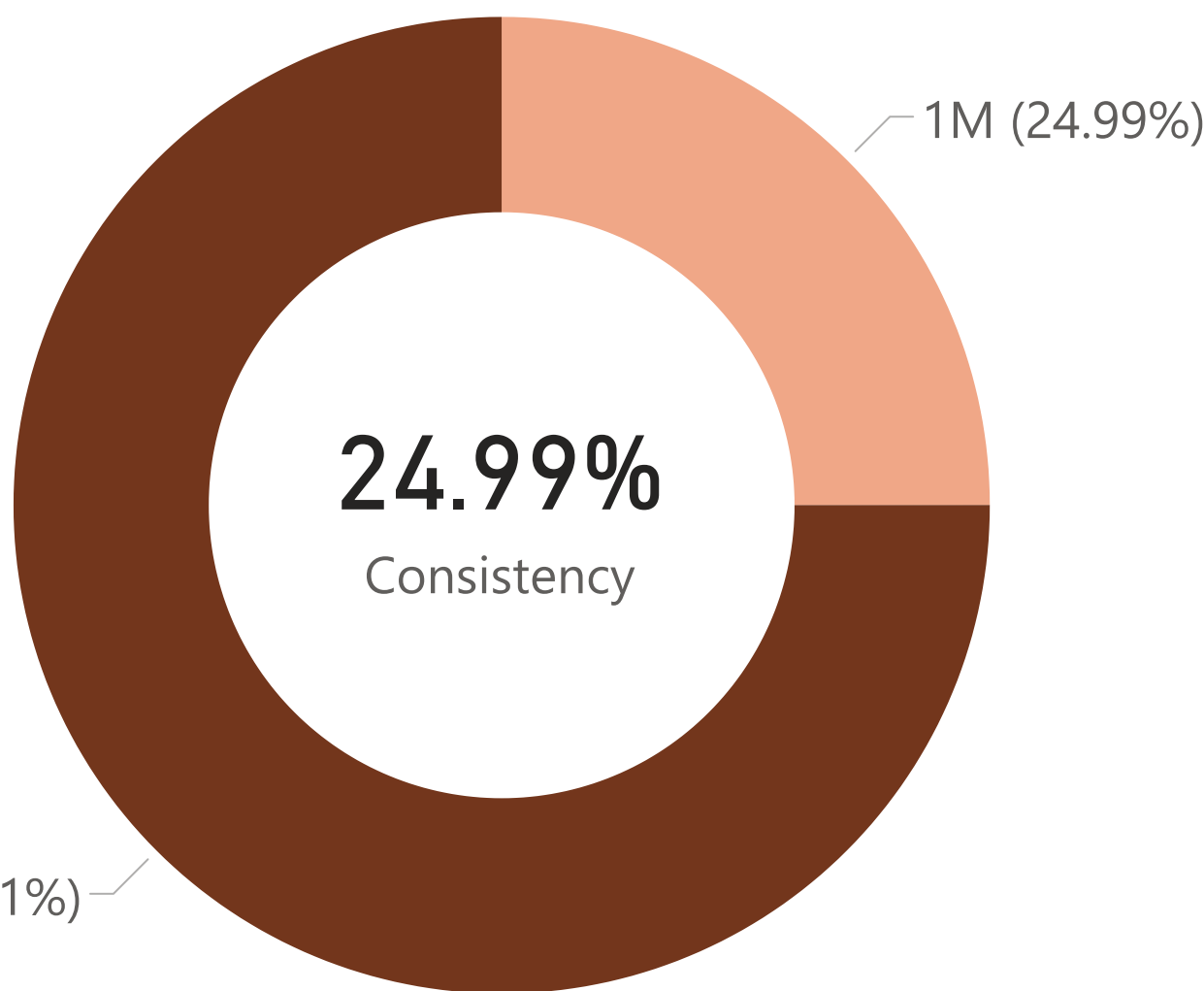
0.00038%

Accurate Percentage

CONSISTENCY

Consistency of Email Username

Alphanumeric Non-Alphanumeric



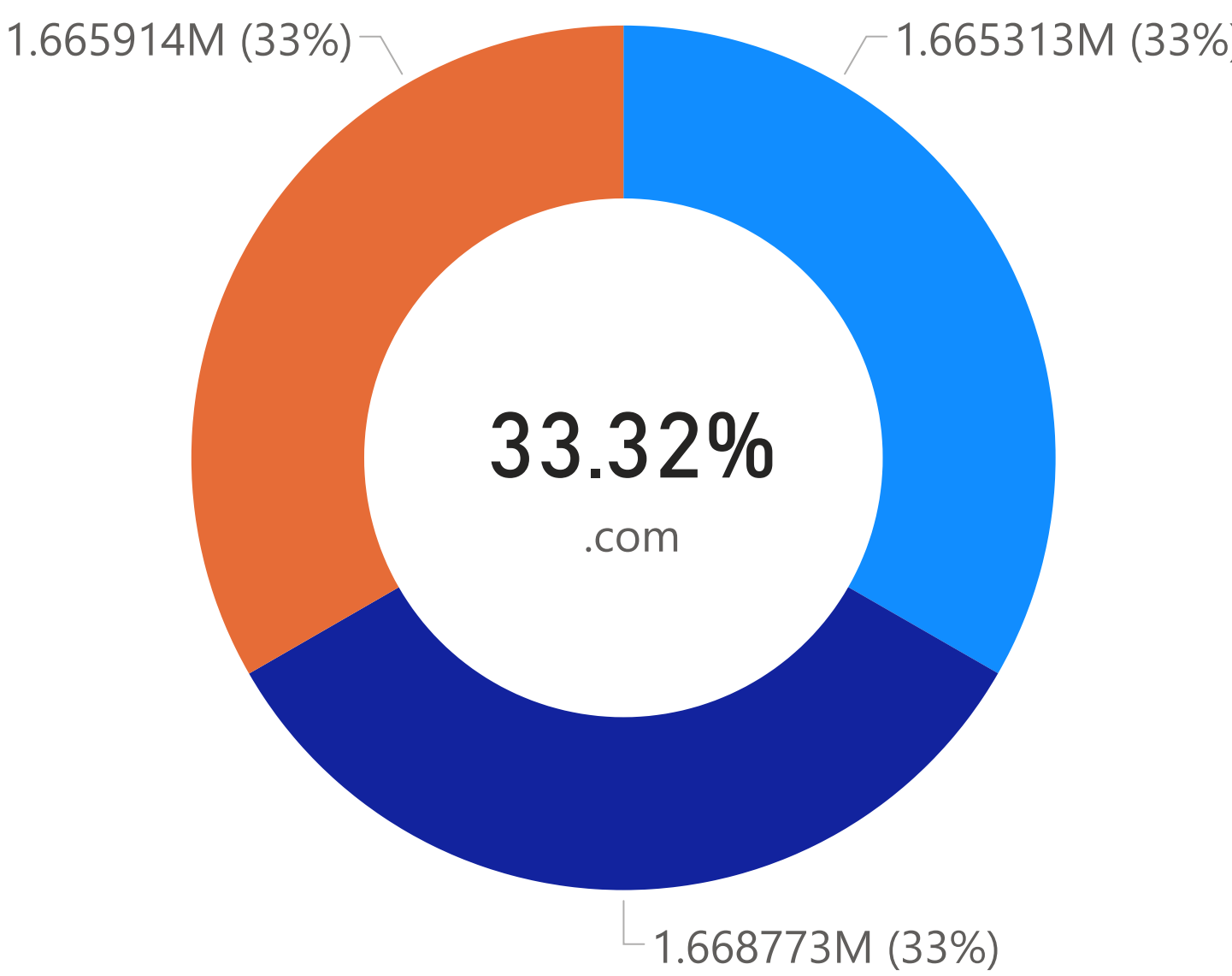
3.75M

Non-Alphanumeric

1.25M

Alphanumeric

Consistency of email domain



33.31%

.org

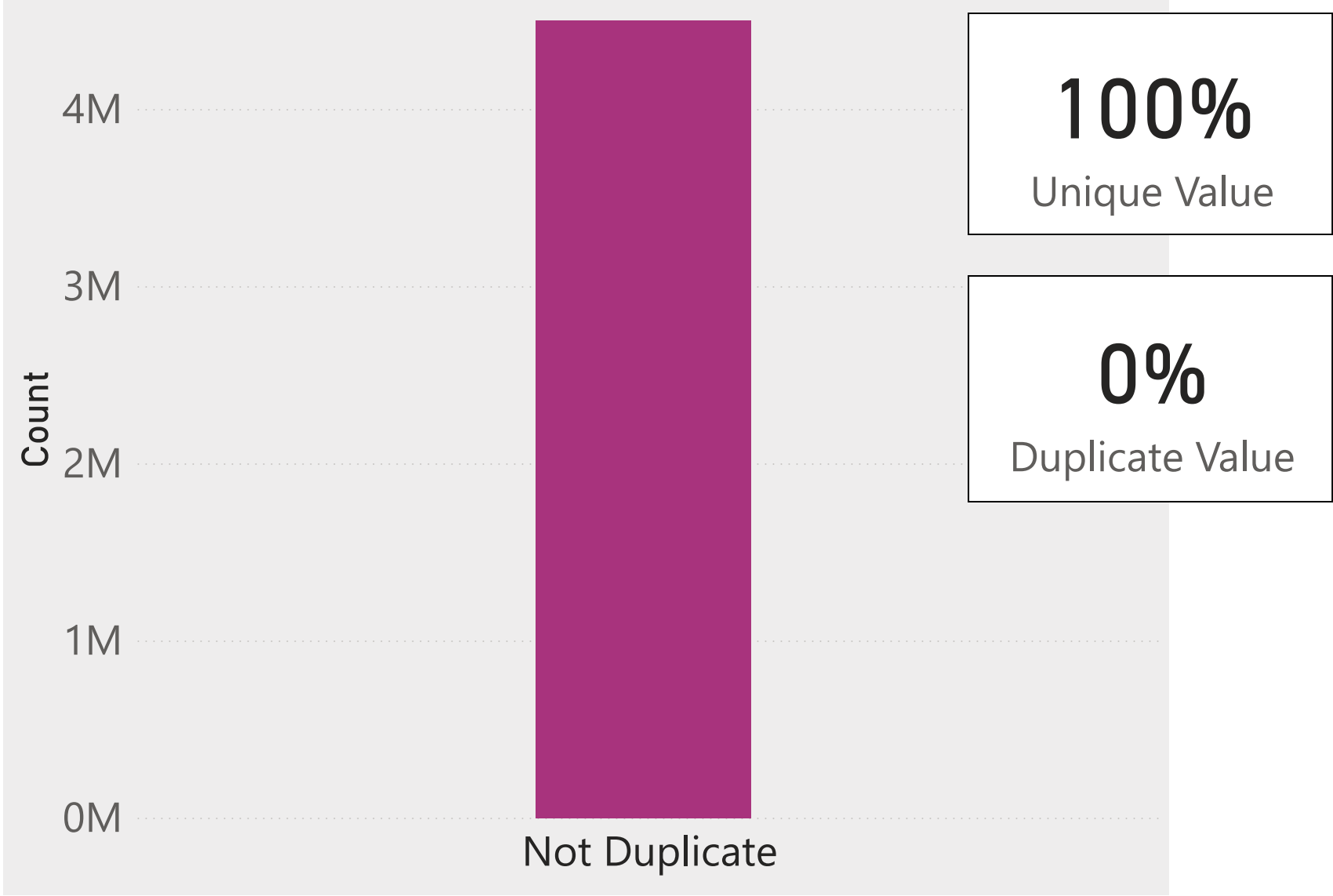
33.38%

.net

DATA QUALITY IN PROCESSED HOSPITAL DATASET

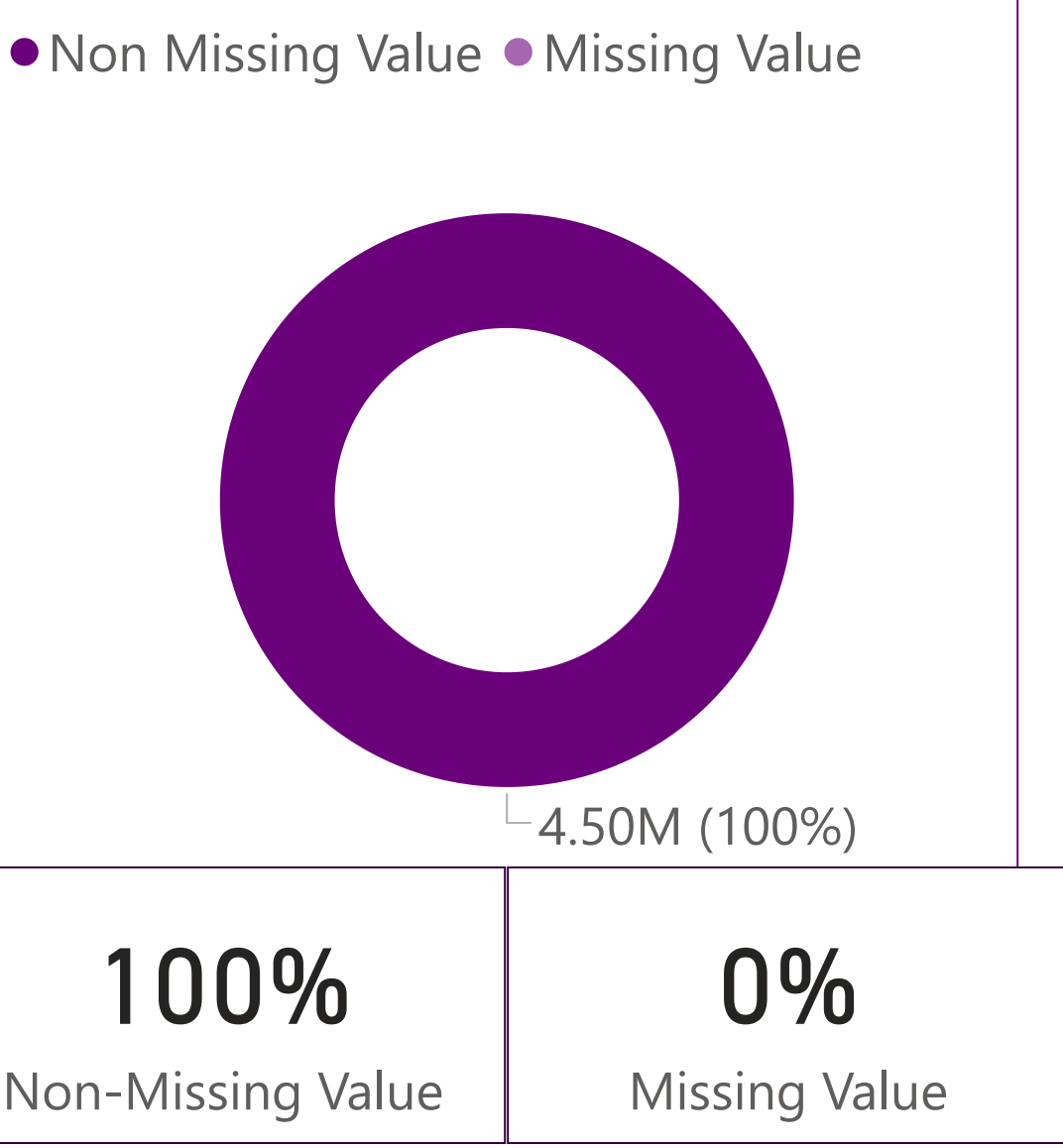
UNIQUENESS

Duplication based on ID and Name

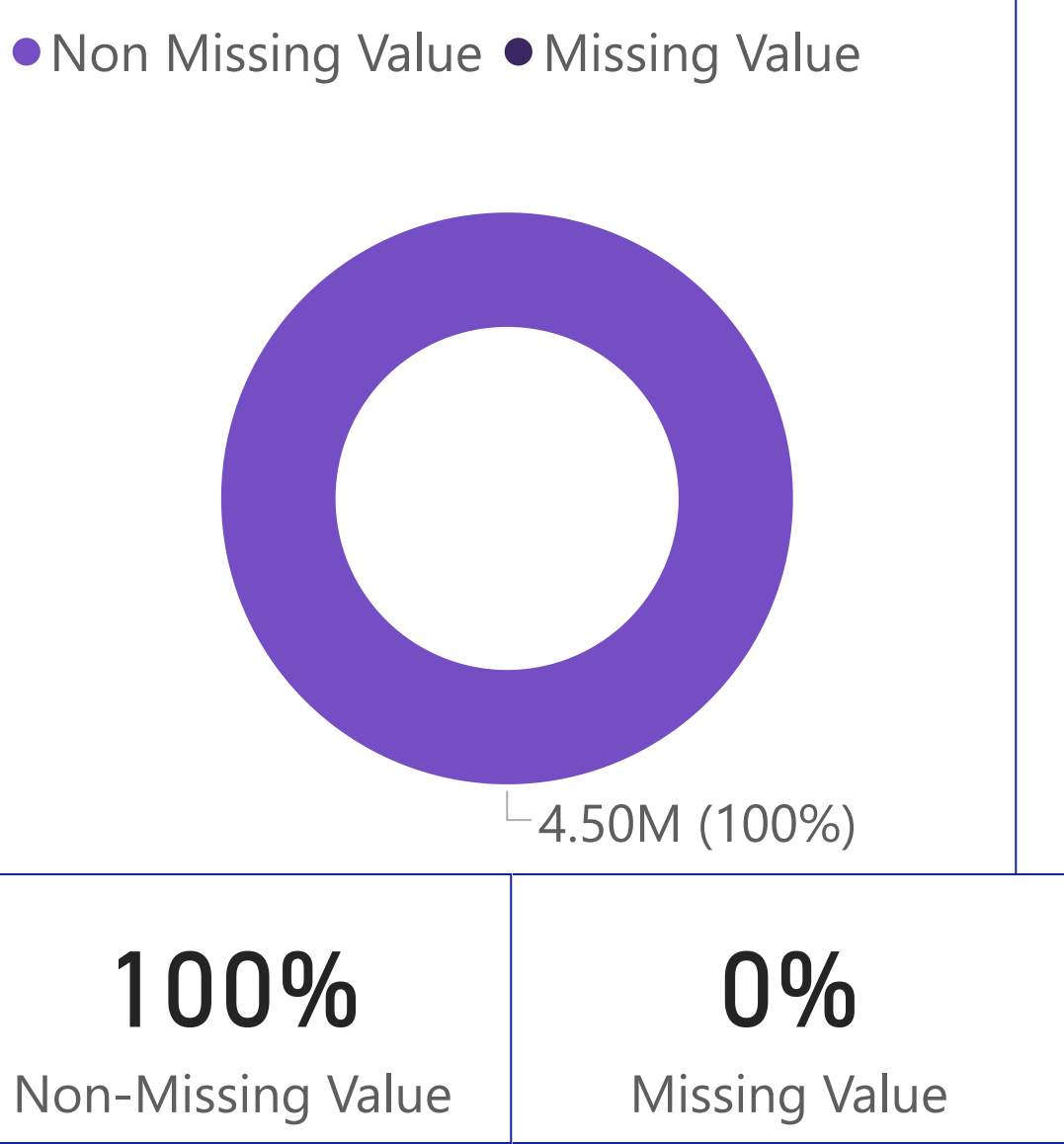


COMPLETENESS

Missing Value by Email

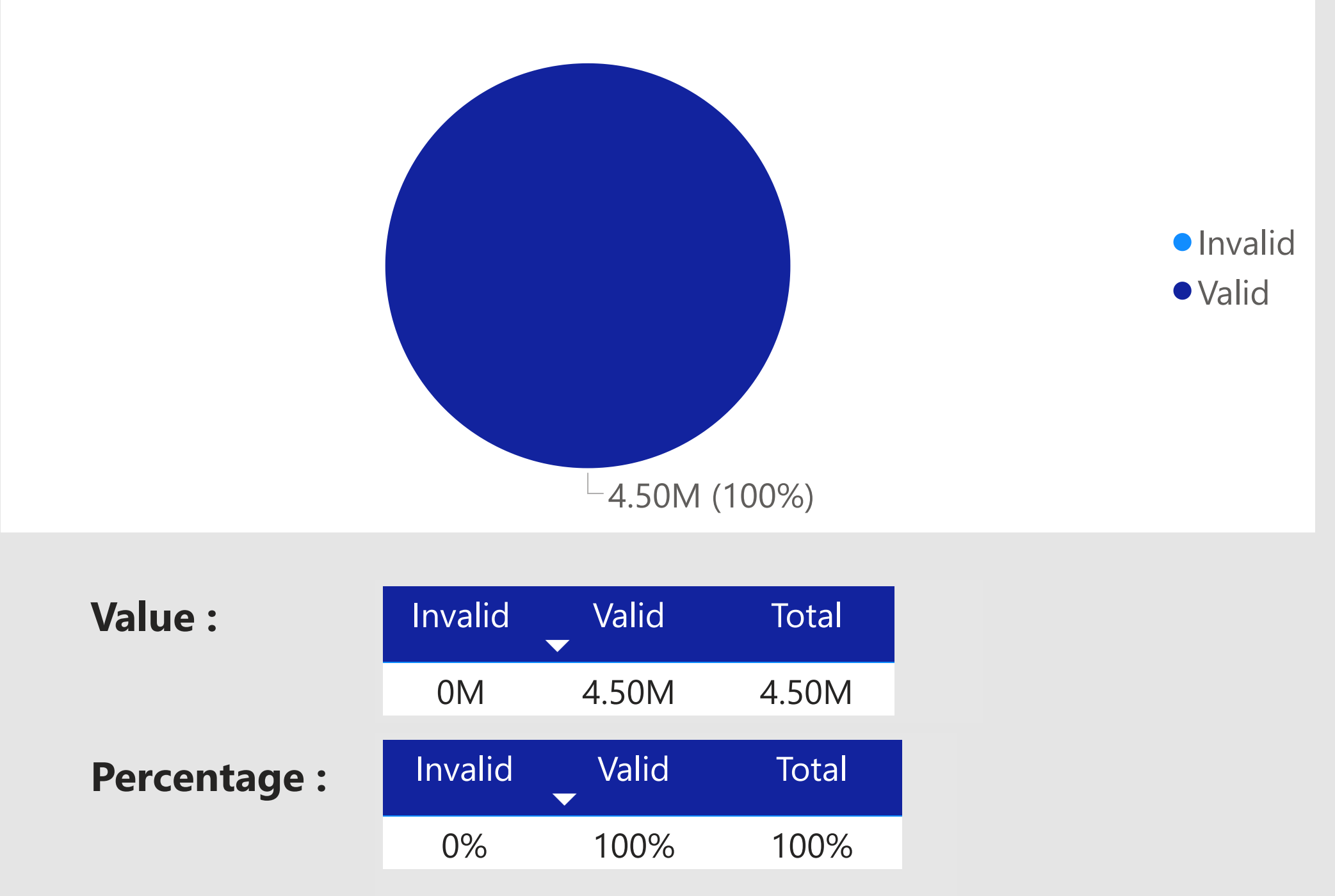


Missing Value by Name

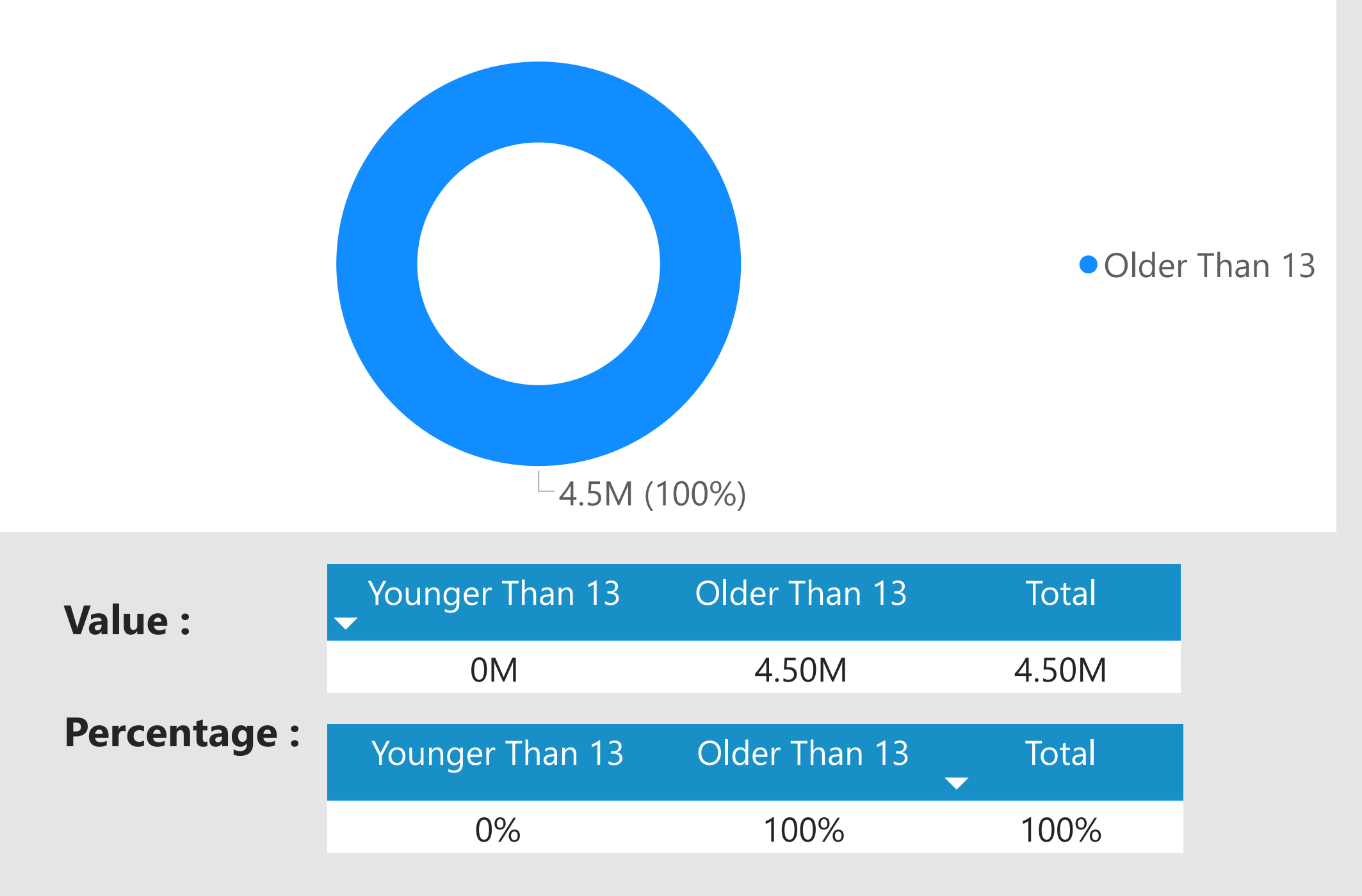


VALIDITY

Validity based on Age and Title



Validity based on Age and Email



ACCURACY

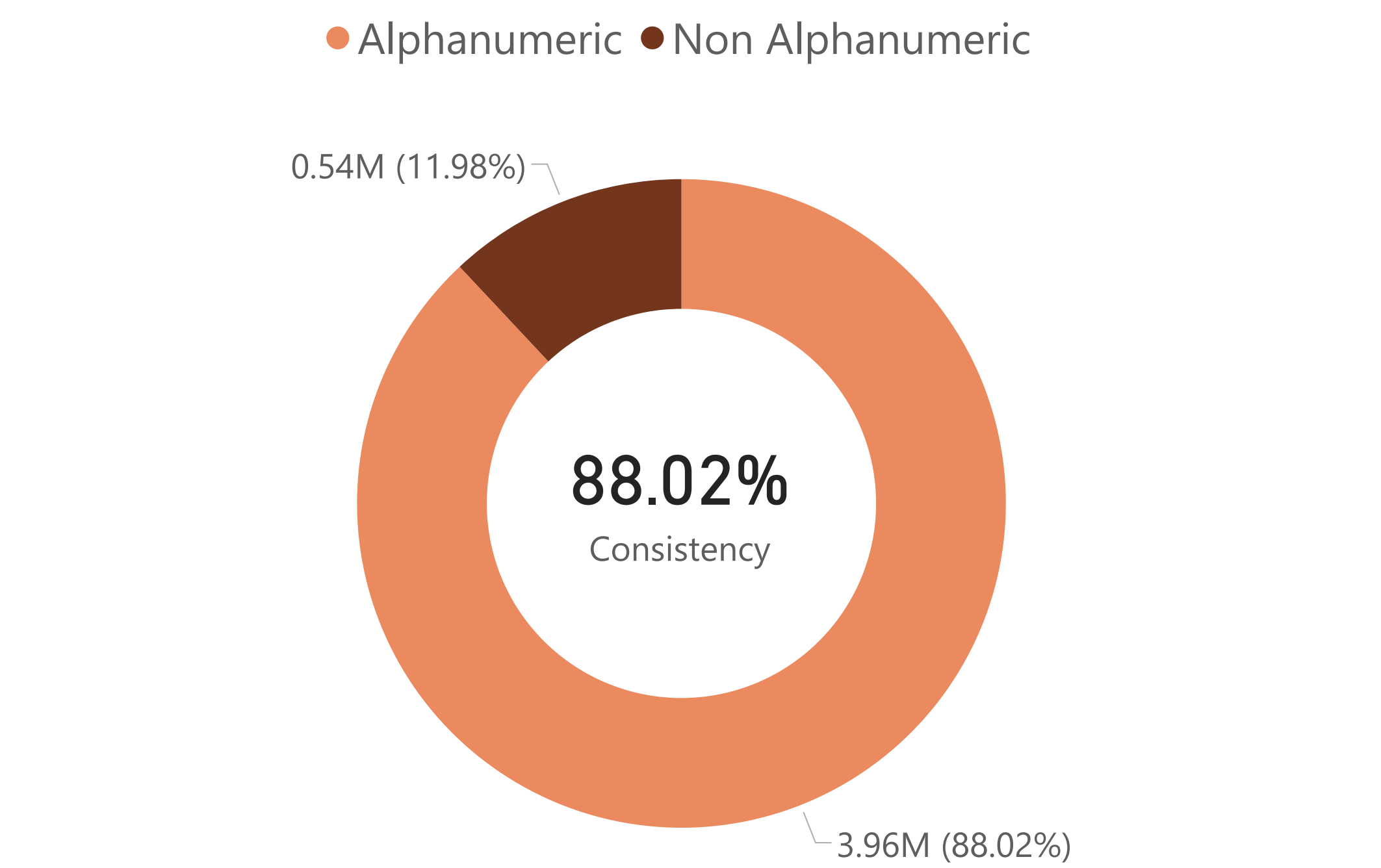
id	name	email	True Count
2249065	Aaron Abbott	aaronabbott_2249065@example.com	1
2806995	Aaron Abbott	aaronabbott_2806995@example.com	1
2869977	Aaron Abbott	aaronabbott_2869977@example.com	1
3427490	Aaron Abbott	aaronabbott_3427490@example.com	1
347775	Aaron Abbott	aaronabbott_347775@example.com	1
3913802	Aaron Abbott	aaronabbott_3913802@example.com	1
874227	Aaron Abbott	aaronabbott_874227@example.com	1
1019566	Aaron Acevedo	aaronacevedo_1019566@example.com	1
1526338	Aaron Acevedo	aaronacevedo_1526338@example.com	1
3364462	Aaron Acevedo	aaronacevedo_3364462@example.com	1
3443589	Aaron Acevedo	aaronacevedo_3443589@example.com	1
4075621	Aaron Acevedo	aaronacevedo_4075621@example.com	1
Total			3960918

3.96M
Accurate Count

88.02%
Accurate Percentage

CONSISTENCY

Consistency of Email Username



Consistency of email domain

