

Profile report – Combined dataset

Toxic comment classification - wrangle – Job ID: 18267268

All Data

7 columns 223,125 rows 2 data types



100% valid values 0% mismatching values 0% missing values

Output destinations

Name	Location	Status
combined_data	bigquery://toxicc-kagglex/kagglex_tcc/combined_data	-

Data sources

Name	Location	Schema validation
train – 2.csv	gs://tcc-bucket-lynnx/train.csv	✓ No schema changes found
test – 3.csv	gs://tcc-bucket-lynnx/test.csv	✓ No schema changes found
test_labels – 2.csv	gs://tcc-bucket-lynnx/test_labels.csv	✓ No schema changes found

Execution summary

Job Status
Completed

Job Type
Manual

Environment
Photon

Start time
Thu, Mar 9, 2023 8:36 PM +02:00

End time
Thu, Mar 9, 2023 8:38 PM +02:00

Duration
2 min

comment_text		toxic		severe_toxic		obscene	
Type	String	Type	Boolean	Type	Boolean	Type	Boolean
Valid	223,125	Valid	223,125	Valid	223,125	Valid	223,125
Mismatched	0	Mismatched	0	Mismatched	0	Mismatched	0
Empty	0	Empty	0	Empty	0	Empty	0
Top 20 values		Top 2 values		Top 2 values		Top 2 values	
please do not add nonsense to wiki...	40	0	201,743	0	221,164	0	210,987
preceding unsigned comment added b...	37	1	21,382	1	1,961	1	12,138
thanks	36						
january	35						
please stop your disruptive editin...	35						
thank you for experimenting with ...	34						
jun utc	33						
thanks for experimenting with wiki...	33						
february	31						
please stop adding nonsense to wik...	31						
october	30						
thank you	30						
august	29						
no problem	29						
please stop your disruptive editin...	29						
please stop if you continue to van...	28						
preceding unsigned comment added b...	27						
december	26						
jan utc	25						
please stop if you continue to va...	25						

threat		insult		identity_hate	
Type	Boolean	Type	Boolean	Type	Boolean
Valid	223,125	Valid	223,125	Valid	223,125
Mismatched	0	Mismatched	0	Mismatched	0
Empty	0	Empty	0	Empty	0
Top 2 values		Top 2 values		Top 2 values	
0	222,436	0	211,823	0	221,008
1	689	1	11,302	1	2,117