

Data Understanding: Data Quality Plan

Feature	Data Quality Issue	Handling Strategy
case_month (category)	None	Keep as is
res_state (category)	2 Null Rows	Drop affected rows, rows contain mostly null values
state_fips_code (category)	2 Null Rows	Same as res_state. Combine res_state and state_fips_code
res_county (category)	Null Values (5.6%), Cardinality doesnt match county_fips_code	Null values changed to 'Surpressed', Combine with county_fips_code
county_fips_code (category)	Null Values (5.6%), Cardinality doesnt match county_fips_code	Null values changed to 'Surpressed', Combine with county_fips_code
age_group (category)	Null values (0.8%)	Changed Null values to 'Missing'
sex (category)	Null values (2.2%)	Changed Null values to 'Missing'
race (category)	Null values (12.24%)	Changed Null values to 'Missing'
ethnicity (category)	Null values (13.34%)	Changed Null values to 'Missing'
case_positive_specimen_interval (float64)	Null Values (47%), 0 values (47%), High and Low Outliers	Too many Null Values, Drop Feature
case_onset_interval (float64)	Null values (57%), 0 value (41%), High and Low Outliers	Too many Null Values, Drop Feature
process (category)	Missing Values (90%)	Keep as is
exposure_yn (category)	Missing Values (86%)	Too many Missing Values, Drop Feature
current_status (category)	None	Nothing
symptom_status (category)	Unknown + Missing Values (52%)	Too many Missing or Unknown Values, Drop Feature
hosp_yn (category)	unknown + Missing values (35%)	Keep as is
icu_yn (category)	Missing Values (78%)	Too many Missing Values, Drop Feature
death_yn (category)	None	Keep as is
underlying_conditions_yn	Null values (91.7%), 98.6% of Non Null are value 'Yes'	Too many Null values, Drop Feature