Data Quality Plan for Cleaned CSV file

Feature	Data Quality Issue	Handling Strategy
case_positive_specimen_interval	Large amount of null values (47%), Failure of Logical integrity checks	Drop column as this feature has little effect on target outcome
case_onset_interval	Large amount of null values (57%), Failure of Logical integrity checks	Drop column as this feature has little effect on target outcome
case_month	No data quality issues	Keep as is
res_state	1 null value	Drop row as it is only 1 row
state_fips_code	1 null value	Drop feature as it is like a duplicate column to res_state
res_county	6% null values	Convert null to missing, and take missing values into account when analysing data
county_fips_code	6% null values	Drop feature as its like a duplicate column to res_county
age_group	0.9% missing/unknown values	Convert unknown to missing, and take missing values into account when analysing data
sex	2.4% null/missing/unknown values	Convert unknown/null to missing, and take missing values into account when analysing data
race	24% null/missing/unknown values	Convert unknown/null to missing, and take missing values into account when analysing data
ethnicity	31.5% null/missing/unknown values	Convert unknown/null to missing, and take missing values into account when analysing data
process	92% missing/unknown values	Drop column as this feature has little effect on target outcome
exposure_yn	90% missing/unknown values	Drop column as this feature has little effect on target outcome
current_status	No data quality issues	Keep as is
symptom_status	53% missing/unknown values	Drop column as this feature has little effect on target outcome and is missing >50% of values
hosp_yn	33% missing/unknown values, 2 rows failed Logical Intergrity Check	Convert unknown to missing, and take missing values into account when analysing data
icu_yn	91% missing/unknown values	Convert unknown/null to missing, and take missing values into account when analysing data
death_yn	No data quality issues	Keep as is
underlying_conditions_yn	91% null values	Convert null to missing, and take missing values into account when analysing data