

Data Quality Plan

Variable Names	Data Quality Issue	Handling Strategy
county_fibs_code	Outliers	Drop the entire column
res_county	Null Values or empty cells	5% so Do Nothing
sex	Null Values or empty cells	2% so Do Nothing
sex	Has both Missing and Unknown	Replace Unknown to Missing
race	Has both Missing and Unknown	Replace Unknown to Missing
race	Null Values or empty cells	Do Nothing
ethnicity	Null Values or empty cells	Do Nothing
process	Null Values or empty cells	Do Nothing
process	Outliers	Replace with null
process	Has both Missing and Unknown	Replace Unknown to Missing
case_positive_specimen_interval	Null Values or empty cells	Replace with interpolated mean
case_positive_specimen_interval	Negative values	Replace with null
case_onset_interval	Null Values or empty cells	Replace with interpolated mean
case_onset_interval	Negative values	Replace with 0
case_onset_interval	Outliers	Replace with interpolated mean
exposure_yn	Null Values or empty cells	Do Nothing
exposure_yn	Outliers	Remove Outliers
exposure_yn	Has both Missing and Unknown	Replace Unknown to Missing
hosp_yn	Outliers	Replace with null
hosp_yn	Null Values or empty cells	Do Nothing
hosp_yn	Has both Missing and Unknown	Replace Unknown to Missing
icu_yn	Outliers	Do Nothing
icu_yn	Outliers	Do Nothing
icu_yn	Null Values or empty cells	Replace Unknown to Missing
current_status	Outliers	Replace with null
symptom_status	Outliers	Do Nothing
symptom_status	Null Values or empty cells	Do Nothing
symptom_status	Has both Missing and Unknown	Replace Unknown to Missing
underlying_conditions_yn	Null Values or empty cells	Do Nothing