

Feature	Data Quality Issue	Handling Strategy
cdc_case_earliest_dt (datetime - months)	none	Keep as is
cdc_report_dt (datetime - months)	17% missing values	CDC reccomend to use cdc_case_earliest_dt , drop feature
pos_spec_dt (datetime - months)	69% missing values	Too many missing values, drop feature
onset_dt (datetime - months)	45% missing values	Consider rows with missing values as asymptomatic patients
current_status (category)	None	Keep as is
sex (category)	None	Keep as is
age_group (category)	None	Keep as is
race_ethnicity_combined (category)	modal value is 'unkown'	investigate rows affected, consider dropping rows
hosp_yn (category)	21% missing values, 20% 'unknown' values	Investigate rows affected, consider imputation or dropping
icu_yn (category)	74% missing values	Too many missing values, drop feature
death_yn (category)	None	Keep as is
medcond_yn (category)	72% missing values	Too many missing values, drop feature