

## Data Quality Plan

Variable	Identified Issue	Proposed Solution
State	Constant value	This column has already been dropped in Part 1 – no variability in the data
Hazard	Constant value	This column has already been dropped – no variability in the data
OBJECTID	None	Do Nothing
Damage	None	Do Nothing
CAL FIRE Unit	None	Do Nothing
County	None	Do Nothing
Structure Type	None	Do Nothing
Structure Category	None	Do Nothing
Latitude	None	Do Nothing
Longitude	None	Do Nothing
Deck/Porch on Grade	Missing values	Fill in the missing values with the mode of the column
Deck/Porch Elevated	Missing values	Fill in the missing values with the mode of the column
Street Number	Missing values	Drop missing values
Street Name	Missing values	Find matching addresses with non-missing Street Names: If found, use those Street Names to fill in the missing values; If no matching Street Names are found, Site Address is used.
Street Type	Missing values	Fill in the missing values with the mode of the column
City	Missing values	Use data in other columns to impute missing values
Zip Code	Missing values	Drop the column – other columns contain address information, so locations can be determined effectively
Incident Start Date	Needs to be converted to datetime format	This has already been converted to datetime format in Part 1
Structure Defense Actions Taken	Missing values	Drop the column – too many missing values as above the 60/40 threshold
Community	Missing values	Drop this column – to reduce noise
# Units in Structure (if multi unit)	Missing values	Drop this column – too many missing values as above the 60/40 threshold
Roof Construction	Missing values	Fill the missing values with the mode of the column

Eaves	Missing values	Fill the missing values with the mode of the column
Vent Screen	Missing values	Fill the missing values with the mode of the column
Exterior Sliding	Missing values	Fill the missing values with the mode of the column
Window Pane	Missing values	Fill the missing values with the mode of the column
Patio Cover/Carport Attached to Structure	Missing values	Fill the missing values with the mode of the column
Fence Attached to Structure	Missing values	Fill the missing values with the mode of the column
Distance – Propane Tank to Structure	Missing values	Drop the column – too many missing values as above the 60/40 threshold
Distance – Residence to Utility/Misc Structure > 120 SQFT	Missing values	Drop the column – too many missing values as above the 60/40 threshold
Site Address (parcel)	Missing values	Remove the rows with missing values
Year Built (parcel)	Missing values and outliers present	Median imputation of missing values. Remove outliers using the Interquartile range (IQR) method – prevents distortion in the prediction model
Assessed Improved Value (parcel)	Outliers present	Remove outliers using the Interquartile range (IQR) method – prevents distortion in the prediction model