



Missing Data Handling

- Missing Data Code
 - None
- Missing Data Filtering
 - `h2o_frame.filter_na_cols(frac=0.2)`
 - `h2o_frame.na_omit()`
- Missing Data Replacement
 - `h2o_frame.impute(...)`
 - `H2OGeneralizedLowRankEstimator`
- Missing Data Inclusion
 - `h2o_frame.insert_missing_values(fraction=0.1, seed=None)`
- Missing Data Summaries
 - `h2o_frame.nacnt()`

Missing Data Handling

```
h2o_frame.impute(column = 0, method = c("mean", "median",  
  "mode"), combine_method c("interpolate", "average", "lo", "hi"),  
  by = NULL, groupByFrame = NULL, values = NULL)
```

Arguments

| | |
|-----------------------|---|
| h2o_frame | The dataset containing the column to impute. |
| column | The column to impute. |
| method | "mean" replaces NAs with the column mean; "median" replaces NAs with the column median; "mode" replaces with the most common factor (for factor columns only) |
| combine_method | If method is "median", then choose how to combine quantiles on even sample sizes. This parameter is ignored in all other cases |
| by | Group by columns |
| groupByFrame | Impute the column col with this pre-computed grouped frame. |
| values | A vector of impute values (one per column). NaN indicates to skip the column |

Missing Data Handling

- Missing Data Code
 - None
- Missing Data Filtering
 - `h2o_frame.filter_na_cols(frac=0.2)`
 - `h2o_frame.na_omit()`
- Missing Data Replacement
 - `h2o_frame.impute(...)`
 - `H2OGeneralizedLowRankEstimator`
- Missing Data Inclusion
 - `h2o_frame.insert_missing_values(fraction=0.1, seed=None)`
- Missing Data Summaries
 - `h2o_frame.nacnt()`