
H2O.ai

# 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

```
1 nacnts = dict(zip(census_data.col_names, census_data.nacnt()))
2 print(dict((k, int(v)) for (k, v) in nacnts.items() if v > 0))
```

```
{'workclass': 1836, 'occupation': 1843, 'native-country': 583}
```

```
1 codes = census_data["native-country"].asnumeric()
2 levels = census_data["native-country"].levels()[0]
3 levels.append("Unknown")
4
5 census_data["native-country-clean"] = h2o.H2OFrame.ifelse(codes != None, codes, len(levels))
6 census_data["native-country-clean"] = census_data["native-country-clean"].asfactor()
7 census_data["native-country-clean"] = census_data["native-country-clean"].set_levels(levels)
8
9 print((census_data["native-country-clean"] == "Unknown").table())
```

| native-country-clean | Count |
|---|---|
| 0 | 31978 |
| 1 | 583 |

# 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 |