```
import pandas as pd
import numpy as np

heart = pd.read_csv("heart.csv")

heart.head()
```


```
{"summary":"{\n  \"name\": \"heart\",\n  \"rows\": 12000,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 3464,\n
\"min\": 1,\n        \"max\": 12000,\n        \"num_unique_values\":
12000,\n        \"samples\": [\n          1936,\n          6495,\n
1721\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 13,\n        \"min\": 18,\n        \"max\": 64,\n
\"num_unique_values\": 47,\n        \"samples\": [\n          24,\n
48,\n          54\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Gender\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          \"Female\",\n          \"Male\",\n          \"Other\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Occupation\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 6,\n        \"samples\": [\n
\"Other\",\n          \"Teacher\",\n          \"Engineer\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Sleep Hours\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2.014885180605652,\n        \"min\":
3.0,\n        \"max\": 10.0,\n        \"num_unique_values\": 71,\n
\"samples\": [\n          4.1,\n          9.6,\n          7.4\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Physical Activity (hrs/week)\",\
n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 2.889000184337612,\n        \"min\": 0.0,\n        \"max\":
10.0,\n        \"num_unique_values\": 101,\n        \"samples\": [\n
7.7,\n          6.7,\n          0.6\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Caffeine Intake (mg/day)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
144,\n        \"min\": 0,\n        \"max\": 499,\n
\"num_unique_values\": 500,\n        \"samples\": [\n          276,\n
351,\n          82\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Alcohol Consumption (drinks/week)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 0,\n
\"max\": 19,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          6,\n          17,\n          5\n        ],\n
```

```
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Smoking\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
2,\n        \"samples\": [\n          \"Yes\",\n          \"No\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Family History of Anxiety\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Stress Level (1-10)\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          3\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Heart Rate (bpm during attack)\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 34,\n
\"min\": 60,\n        \"max\": 179,\n        \"num_unique_values\":
120,\n        \"samples\": [\n          84,\n          116\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Breathing Rate (breaths/min)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8,\n        \"min\": 12,\n
\"max\": 39,\n        \"num_unique_values\": 28,\n        \"samples\":
[\n          12,\n          21\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Sweating Level (1-5)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 1,\n
\"max\": 5,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          5,\n          2\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Dizziness\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Yes\",\n          \"No\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Medication\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Therapy Sessions (per month)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 0,\n
\"max\": 9,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          0\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Recent Major Life Event\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"No\",\n          \"Yes\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \"Diet Quality (1-10)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          8,\n
10\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Severity of Anxiety Attack (1-10)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          8\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"heart"}
```


print true false for null values

```
heart.isnull()
```


```
{"summary":"{\n  \"name\": \"heart\",\n  \"rows\": 12000,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\
n        \"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Age\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Gender\",\n      \"properties\":
{\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\
n        \"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Occupation\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Sleep Hours\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Physical Activity (hrs/week)\",\
n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Caffeine Intake (mg/day)\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Alcohol Consumption
(drinks/week)\",\n      \"properties\": {\n        \"dtype\":
\"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\":
```

```
[\n          false\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Smoking\",\n      \"properties\": {\n        \"dtype\":
\"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\":
[\n          false\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Family History of Anxiety\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          false\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Stress Level (1-10)\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Heart Rate (bpm during
attack)\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Breathing Rate (breaths/min)\",\
n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Sweating Level (1-5)\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Dizziness\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Medication\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Therapy Sessions (per month)\",\
n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Recent Major Life Event\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Diet Quality (1-10)\",\n
\"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Severity of Anxiety Attack (1-
10)\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          false\n
```

```
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe"}
```

check the number of null values

```
heart.isnull().sum()
```

```
ID                                    0
Age                                   0
Gender                                0
Occupation                            0
Sleep Hours                           0
Physical Activity (hrs/week)          0
Caffeine Intake (mg/day)              0
Alcohol Consumption (drinks/week)     0
Smoking                               0
Family History of Anxiety             0
Stress Level (1-10)                   0
Heart Rate (bpm during attack)        0
Breathing Rate (breaths/min)          0
Sweating Level (1-5)                  0
Dizziness                             0
Medication                            0
Therapy Sessions (per month)          0
Recent Major Life Event               0
Diet Quality (1-10)                   0
Severity of Anxiety Attack (1-10)     0
dtype: int64
```

add null value

```
heart.loc[1,"Age"] = np.nan
```

check again number of null values

```
heart.isnull().sum()
```

```
ID                                    0
Age                                   1
Gender                                0
Occupation                            0
Sleep Hours                           0
Physical Activity (hrs/week)          0
Caffeine Intake (mg/day)              0
Alcohol Consumption (drinks/week)     0
Smoking                               0
Family History of Anxiety             0
Stress Level (1-10)                   0
Heart Rate (bpm during attack)        0
```

```
Breathing Rate (breaths/min)          0
Sweating Level (1-5)                  0
Dizziness                             0
Medication                            0
Therapy Sessions (per month)          0
Recent Major Life Event               0
Diet Quality (1-10)                   0
Severity of Anxiety Attack (1-10)     0
dtype: int64
```

```
heart.loc[4,"Age"] = np.nan
heart.loc[5,"Age"] = np.nan

heart.head()
```


```
{"summary":"{\n  \"name\": \"heart\",\n  \"rows\": 12000,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 3464,\n
\"min\": 1,\n        \"max\": 12000,\n        \"num_unique_values\":
12000,\n        \"samples\": [\n          1936,\n          6495,\n
1721\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 13.474070305313091,\n        \"min\": 18.0,\n        \"max\":
64.0,\n        \"num_unique_values\": 47,\n        \"samples\": [\n
35.0,\n          48.0,\n          26.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Gender\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
3,\n        \"samples\": [\n          \"Female\",\n
\"Male\",\n          \"Other\"\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Occupation\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 6,\n
\"samples\": [\n          \"Other\",\n          \"Teacher\",\n
\"Engineer\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Sleep Hours\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2.014885180605652,\n        \"min\":
3.0,\n        \"max\": 10.0,\n        \"num_unique_values\": 71,\n
\"samples\": [\n          4.1,\n          9.6,\n          7.4\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Physical Activity (hrs/week)\",\
n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 2.889000184337612,\n        \"min\": 0.0,\n        \"max\":
10.0,\n        \"num_unique_values\": 101,\n        \"samples\": [\n
7.7,\n          6.7,\n          0.6\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Caffeine Intake (mg/day)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
```

```
144,\n        \"min\": 0,\n        \"max\": 499,\n
\"num_unique_values\": 500,\n        \"samples\": [\n          276,\n
351,\n          82\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Alcohol Consumption (drinks/week)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 0,\n
\"max\": 19,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          6,\n          17,\n          5\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Smoking\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
2,\n        \"samples\": [\n          \"Yes\",\n          \"No\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Family History of Anxiety\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Stress Level (1-10)\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          3\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Heart Rate (bpm during attack)\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 34,\n
\"min\": 60,\n        \"max\": 179,\n        \"num_unique_values\":
120,\n        \"samples\": [\n          84,\n          116\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Breathing Rate (breaths/min)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8,\n        \"min\": 12,\n
\"max\": 39,\n        \"num_unique_values\": 28,\n        \"samples\":
[\n          12,\n          21\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Sweating Level (1-5)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 1,\n
\"max\": 5,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          5,\n          2\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Dizziness\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Yes\",\n          \"No\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Medication\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Therapy Sessions (per month)\",\n      \"properties\": {\n
```

```
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 0,\n
\"max\": 9,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          0\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Recent Major Life Event\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"No\",\n          \"Yes\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Diet Quality (1-10)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          8,\n
10\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Severity of Anxiety Attack (1-10)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          8\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"heart"}
```


```
heart.shape
```

```
(12000, 20)
```

check not null values

```
heart.notnull().sum()
```

```
ID                                  12000
Age                                 11997
Gender                              12000
Occupation                          12000
Sleep Hours                         12000
Physical Activity (hrs/week)        12000
Caffeine Intake (mg/day)            12000
Alcohol Consumption (drinks/week)   12000
Smoking                             12000
Family History of Anxiety           12000
Stress Level (1-10)                 12000
Heart Rate (bpm during attack)      12000
Breathing Rate (breaths/min)        12000
Sweating Level (1-5)                12000
Dizziness                           12000
Medication                          12000
Therapy Sessions (per month)        12000
Recent Major Life Event             12000
Diet Quality (1-10)                 12000
Severity of Anxiety Attack (1-10)   12000
dtype: int64
```

drop row containing na values

```
heart_drop = heart.dropna(axis=0)
heart_drop.shape
```

```
(11997, 20)
```

drop columns containing na values

```
heart_drop = heart.dropna(axis=1)
heart_drop.shape
```

```
(12000, 19)
```

drop rows with atleast one na value

```
heart_drop = heart.dropna(how = "any")
heart_drop.shape
```

```
(11997, 20)
```

drop rows if and only if every element is na

```
heart_drop = heart.dropna(how = "all")
heart_drop.shape
```

```
(12000, 20)
```

if there are not more than 5 notnull values row will drop.

```
heart_drop = heart.dropna(thresh = 5)
heart_drop.shape
```

```
(12000, 20)
```

```
heart_drop = heart.dropna(thresh = 20)
heart_drop.shape
```

```
(11997, 20)
```

```
heart_drop = heart.dropna(subset = ["Age" , "ID"])
heart_drop.shape
```

```
(11997, 20)
```

```
heart_drop = heart.dropna(subset = ["Gender" , "ID"])
heart_drop.shape
```

```
(12000, 20)
```

fill na values with -1.

```
heart.fillna(-1)
```


```
{"summary":"{\n  \"name\": \"heart\",\n  \"rows\": 12000,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 3464,\n
\"min\": 1,\n        \"max\": 12000,\n        \"num_unique_values\":
12000,\n        \"samples\": [\n          1936,\n          6495,\n
1721\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 13.488714991147173,\n        \"min\": -1.0,\n        \"max\":
64.0,\n        \"num_unique_values\": 48,\n        \"samples\": [\n
26.0,\n          48.0,\n          24.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Gender\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
3,\n        \"samples\": [\n          \"Female\",\n
\"Male\",\n          \"Other\"\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Occupation\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 6,\n
\"samples\": [\n          \"Other\",\n          \"Teacher\",\n
\"Engineer\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Sleep Hours\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2.014885180605652,\n        \"min\":
3.0,\n        \"max\": 10.0,\n        \"num_unique_values\": 71,\n
\"samples\": [\n          4.1,\n          9.6,\n          7.4\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Physical Activity (hrs/week)\",\
n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 2.889000184337612,\n        \"min\": 0.0,\n        \"max\":
10.0,\n        \"num_unique_values\": 101,\n        \"samples\": [\n
7.7,\n          6.7,\n          0.6\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Caffeine Intake (mg/day)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
144,\n        \"min\": 0,\n        \"max\": 499,\n
\"num_unique_values\": 500,\n        \"samples\": [\n          276,\n
351,\n          82\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Alcohol Consumption (drinks/week)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 0,\n
\"max\": 19,\n        \"num_unique_values\": 20,\n        \"samples\":
[\n          6,\n          17,\n          5\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Smoking\",\n      \"properties\":
{\n        \"dtype\": \"category\",\n        \"num_unique_values\":
2,\n        \"samples\": [\n          \"Yes\",\n          \"No\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
```

```
}\n    },\n    {\n      \"column\": \"Family History of Anxiety\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Stress Level (1-10)\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          3\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Heart Rate (bpm during attack)\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 34,\n
\"min\": 60,\n        \"max\": 179,\n        \"num_unique_values\":
120,\n        \"samples\": [\n          84,\n          116\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Breathing Rate (breaths/min)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8,\n        \"min\": 12,\n
\"max\": 39,\n        \"num_unique_values\": 28,\n        \"samples\":
[\n          12,\n          21\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Sweating Level (1-5)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 1,\n
\"max\": 5,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          5,\n          2\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Dizziness\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Yes\",\n          \"No\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Medication\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"Yes\",\
n          \"No\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Therapy Sessions (per month)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 0,\n
\"max\": 9,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          0\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Recent Major Life Event\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"No\",\n          \"Yes\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Diet Quality (1-10)\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
2,\n        \"min\": 1,\n        \"max\": 10,\n
\"num_unique_values\": 10,\n        \"samples\": [\n          8,\n
10\n        ],\n        \"semantic_type\": \"\",\n
```

```
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Severity of Anxiety Attack (1-10)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 10,\n        \"num_unique_values\": 10,\n        \"samples\":
[\n          9,\n          8\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe"}
```


```
new_data = pd.read_csv("fillna.csv")
new_data.head()
```


```
{"summary":"{\n  \"name\": \"new_data\",\n  \"rows\": 8,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 2,\n
\"min\": 1,\n        \"max\": 8,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          2,\n          6,\n          1\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 7,\n
\"samples\": [\n          \"Aarav\",\n          \"Aditi\",\n
\"Rohan\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 4.956957592256421,\n        \"min\": 25.0,\n        \"max\":
40.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n
25.0,\n          30.0,\n          40.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"State\",\n      \"properties\": {\
n        \"dtype\": \"string\",\n        \"num_unique_values\": 6,\n
\"samples\": [\n          \"Karnataka\",\n          \"Maharashtra\",\n
\"Uttar Pradesh\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Department\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"HR\",\n          \"Marketing\",\n
\"Engineering\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Salary\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 9416.29792788369,\n        \"min\": 50000.0,\n
\"max\": 75000.0,\n        \"num_unique_values\": 6,\n
\"samples\": [\n          50000.0,\n          55000.0,\n
75000.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Joining_Date\",\n      \"properties\": {\n        \"dtype\":
\"object\",\n        \"num_unique_values\": 7,\n        \"samples\":
[\n          \"15-01-2020\",\n          \"20-05-2019\",\n
\"11-12-2017\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"new_data"}
```

```
new_data.fillna(-1)
```


```
{"summary":"{\n  \"name\": \"new_data\",\n  \"rows\": 8,\n
\"fields\": [\n    {\n      \"column\": \"ID\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 2,\n
\"min\": 1,\n        \"max\": 8,\n        \"num_unique_values\": 8,\n
\"samples\": [\n          2,\n          6,\n          1\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 8,\n
\"samples\": [\n          \"Aditi\",\n          -1,\n
\"Aarav\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 12.302729081677075,\n        \"min\": -1.0,\n        \"max\":
40.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n
-1.0,\n          29.0,\n          25.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"State\",\n      \"properties\": {\
n        \"dtype\": \"string\",\n        \"num_unique_values\": 7,\n
\"samples\": [\n          \"Karnataka\",\n          \"Maharashtra\",\n
\"Delhi\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Department\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"HR\",\n          \"Marketing\",\n
\"Engineering\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Salary\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 29486.51992618225,\n        \"min\": -1.0,\n        \"max\":
75000.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n
50000.0,\n          55000.0,\n          58000.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Joining_Date\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 8,\n        \"samples\": [\n          \"20-05-
2019\",\n          \"05-03-2022\",\n          \"15-01-2020\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```


mean of age for new data

```
meanofa = new_data["Age"].mean()
print(meanofa)
```

```
31.285714285714285
```

add mean of age of new data to clo age in na values.

```
new_data["Age"].fillna(new_data["Age"].mean())
```

```
0    25.000000
1    31.285714
2    30.000000
3    28.000000
4    35.000000
5    29.000000
6    40.000000
7    32.000000
Name: Age, dtype: float64
```

cal salary median

```
medianofs = new_data["Salary"].median()
print(medianofs)
```

```
59000.0
```

add salary median to na values of salary.

```
new_data["Salary"].fillna(new_data["Salary"].median())
```

```
0    50000.0
1    55000.0
2    59000.0
3    60000.0
4    70000.0
5    58000.0
6    59000.0
7    75000.0
Name: Salary, dtype: float64
```

add unknown to na values in state column.

```
new_data["State"].fillna("Unknown")
```

```
0        Karnataka
1      Maharashtra
2        Karnataka
3          Unknown
4           Kerala
5       Tamil Nadu
6            Delhi
7    Uttar Pradesh
Name: State, dtype: object
```

backward fill joining date.

```
new_data["Joining_Date"].bfill()
```

```
0    15-01-2020
1    20-05-2019
2    10-06-2021
3    01-08-2018
4    05-03-2022
5    05-03-2022
6    11-12-2017
7    20-10-2019
Name: Joining_Date, dtype: object
```

forward fill joining data.

```
new_data["Joining_Date"].ffill()
```

```
0    15-01-2020
1    20-05-2019
2    10-06-2021
3    01-08-2018
4    01-08-2018
5    05-03-2022
6    11-12-2017
7    20-10-2019
Name: Joining_Date, dtype: object
```

```
new_data["Salary"].ffill()
```

```
0    50000.0
1    55000.0
2    55000.0
3    60000.0
4    70000.0
5    58000.0
6    58000.0
7    75000.0
Name: Salary, dtype: float64
```

check mean after forward fill.

```
new_data["Salary"].ffill().mean()
```

```
60125.0
```

```
new_data["Salary"].bfill()
```

```
0    50000.0
1    55000.0
2    60000.0
3    60000.0
```

```
4    70000.0
5    58000.0
6    75000.0
7    75000.0
Name: Salary, dtype: float64
```

cal mean after backward fill.

```
new_data["Salary"].bfill().mean()
```

```
62875.0
```