

# Alzheimer

```
[ ]: import numpy as np
import pandas as pd
from pandas import DataFrame
import matplotlib.pyplot as plt
```

```
[ ]: f = './Alzheimer Disease and Healthy Aging Data In US.csv'
df = pd.read_csv(f, header=0)
```

```
/var/folders/zs/_rd25w5j2ksgfxgty_jstgl00000gn/T/ipykernel_65043/865106179.py:2:
DtypeWarning: Columns (13,14) have mixed types. Specify dtype option on import
or set low_memory=False.
    df = pd.read_csv(f, header=0)
```

```
[ ]: df.dtypes
```

```
[ ]: YearStart          int64
YearEnd              int64
LocationAbbr        object
LocationDesc        object
Datasource          object
Class               object
Topic               object
Question            object
Data_Value_Unit     object
DataValueTypeID     object
Data_Value_Type     object
Data_Value          float64
Data_Value_Alt      float64
Low_Confidence_Limit  object
High_Confidence_Limit object
Sample_Size         float64
StratificationCategory1 object
Stratification1      object
StratificationCategory2 object
Stratification2      object
Geolocation         object
ClassID             object
TopicID             object
```

```

QuestionID          object
LocationID          int64
StratificationCategoryID1  object
StratificationID1    object
StratificationCategoryID2  object
StratificationID2    object
dtype: object

```

```

[ ]: #
df["Class"].value_counts() #    Class

```

```

[ ]: Overall Health          71694
Screenings and Vaccines    46867
Nutrition/Physical Activity/Obesity  24851
Cognitive Decline         19180
Caregiving                 18671
Mental Health              16600
Smoking and Alcohol Use    16599
Name: Class, dtype: int64

```

```

[ ]: nums = df["Data_Value"] #    5    Data_Value
nullnum = nums.isnull().sum()
nums = nums.dropna(axis = 0)
Minimum = min(nums)
Maximum = max(nums)
Q1 = np.percentile(nums, 25)
Median = np.median(nums)
Q3 = np.percentile(nums, 75)
print("    {}".format(nullnum))
print("    {}".format(Minimum))
print("Q1 {}".format(Q1))
print("    {}".format(Median))
print("Q3 {}".format(Q3))
print("    {}".format(Maximum))

```

```

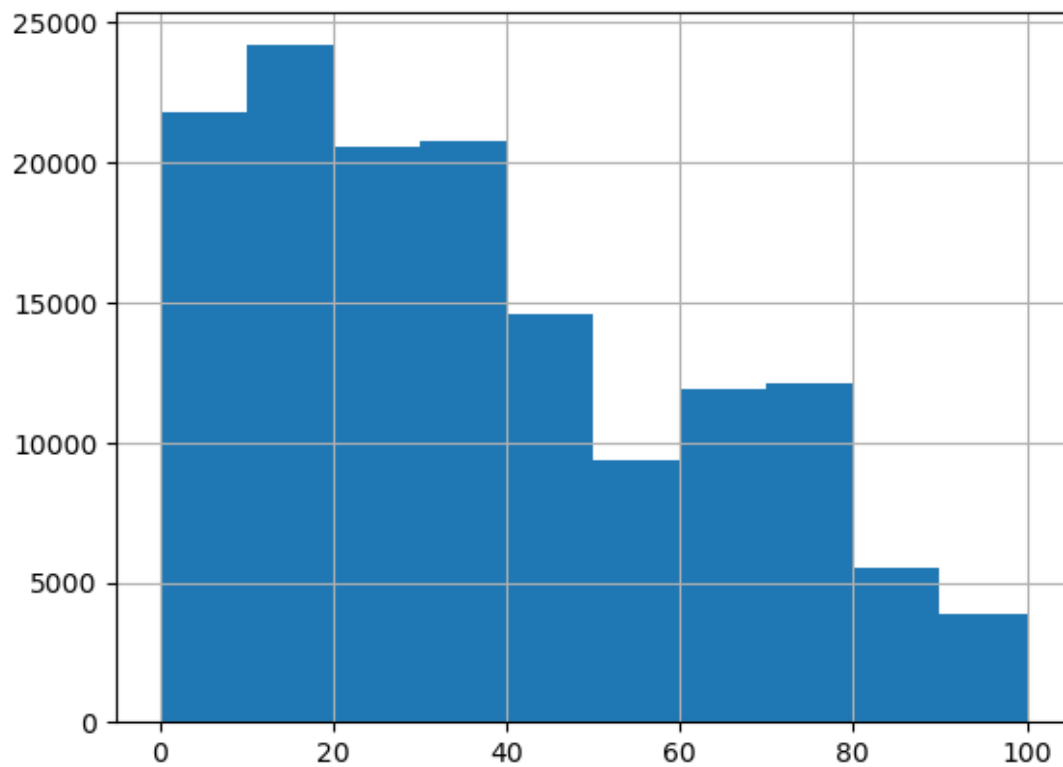
69833
0.0
Q1 15.3
32.5
Q3 56.8
100.0

```

```

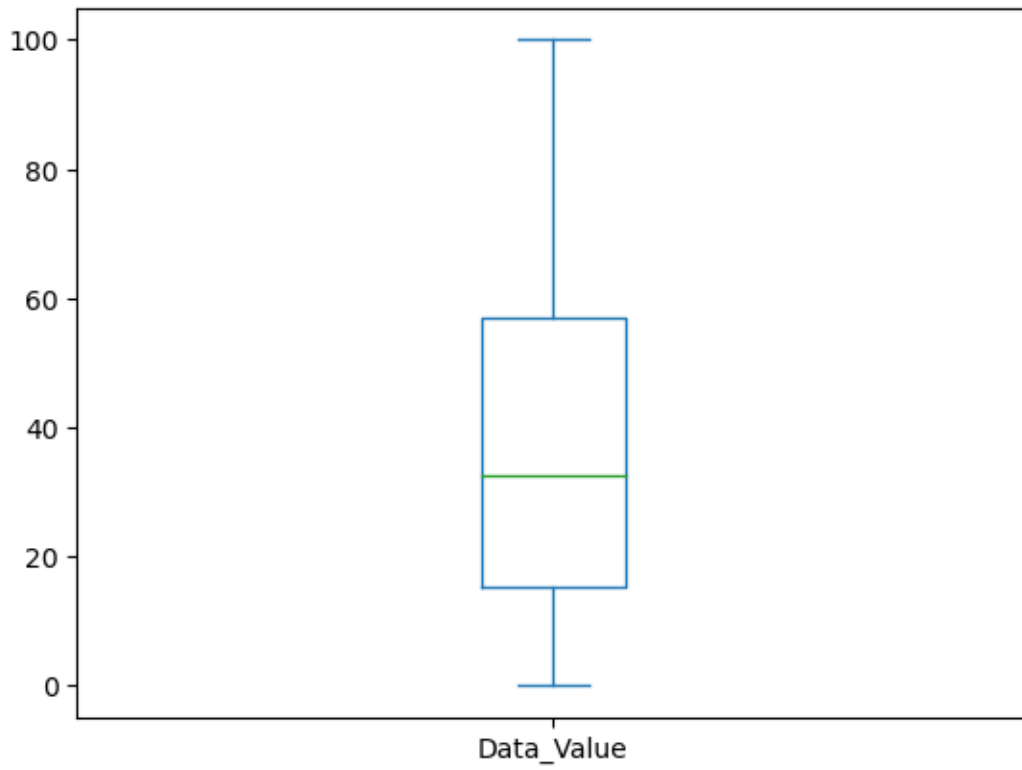
[ ]: hist = df["Data_Value"].hist() #

```



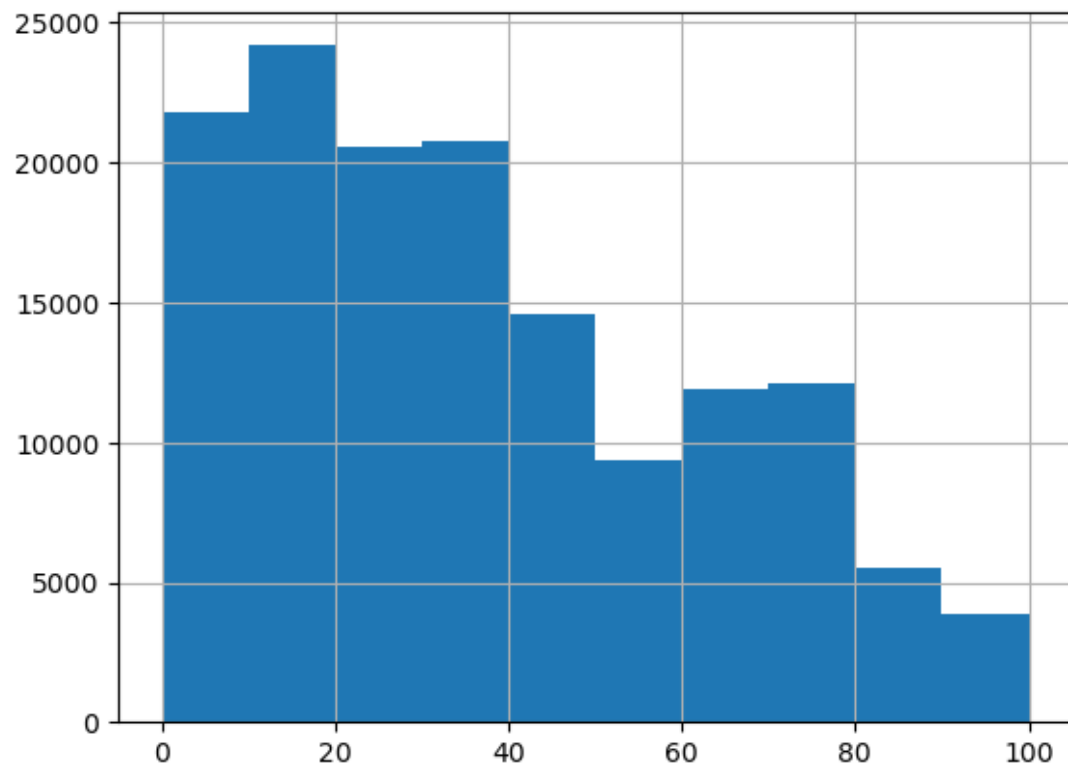
```
[ ]: df["Data_Value"].plot.box() #
```

```
[ ]: <Axes: >
```

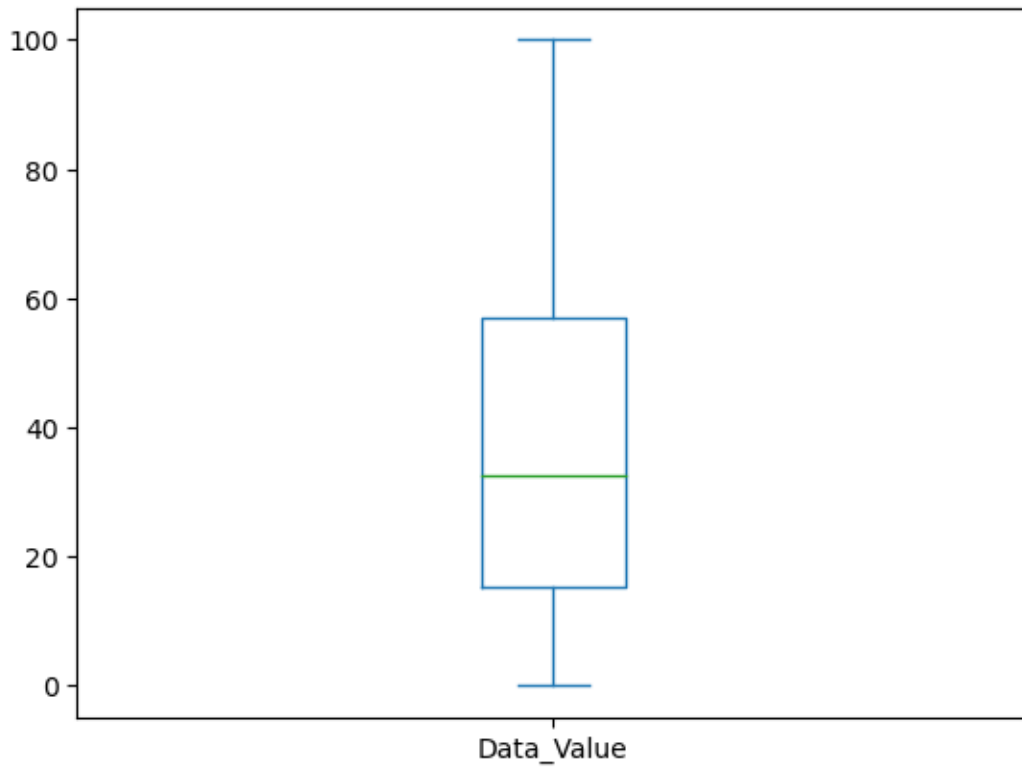


```
[ ]: #  
data_dropna = df["Data_Value"].dropna(axis = 0)  
data_dropna.hist() #
```

```
[ ]: <Axes: >
```

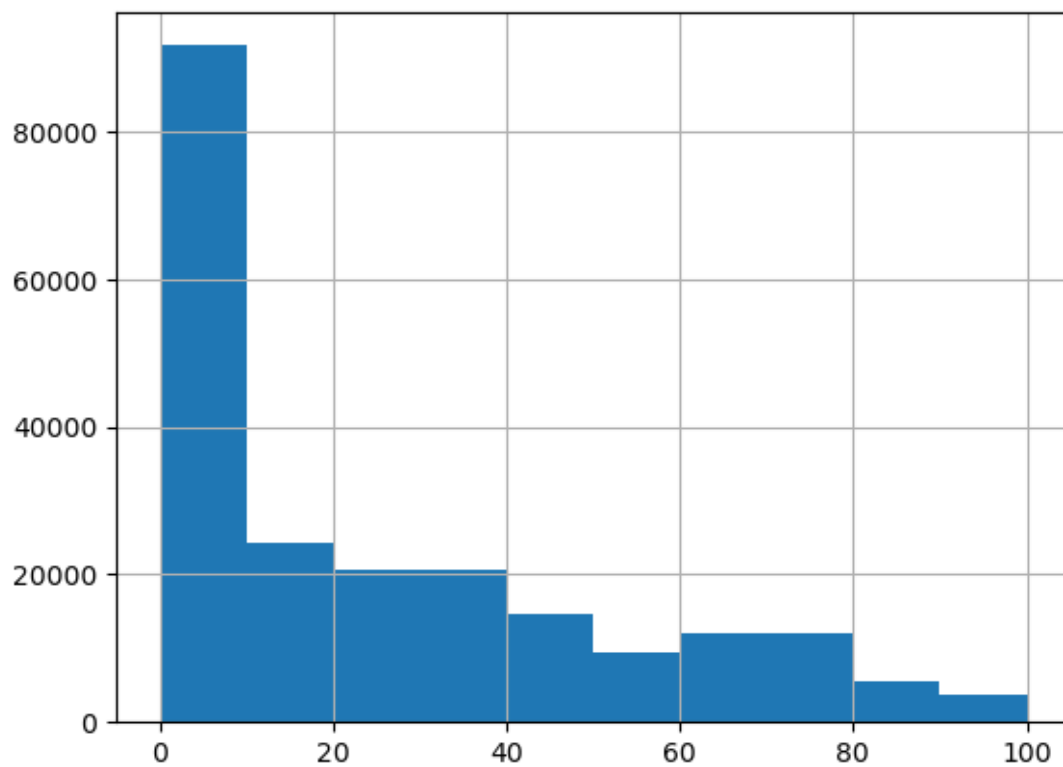


```
[ ]: data_dropna.plot.box()  
plt.show() #
```

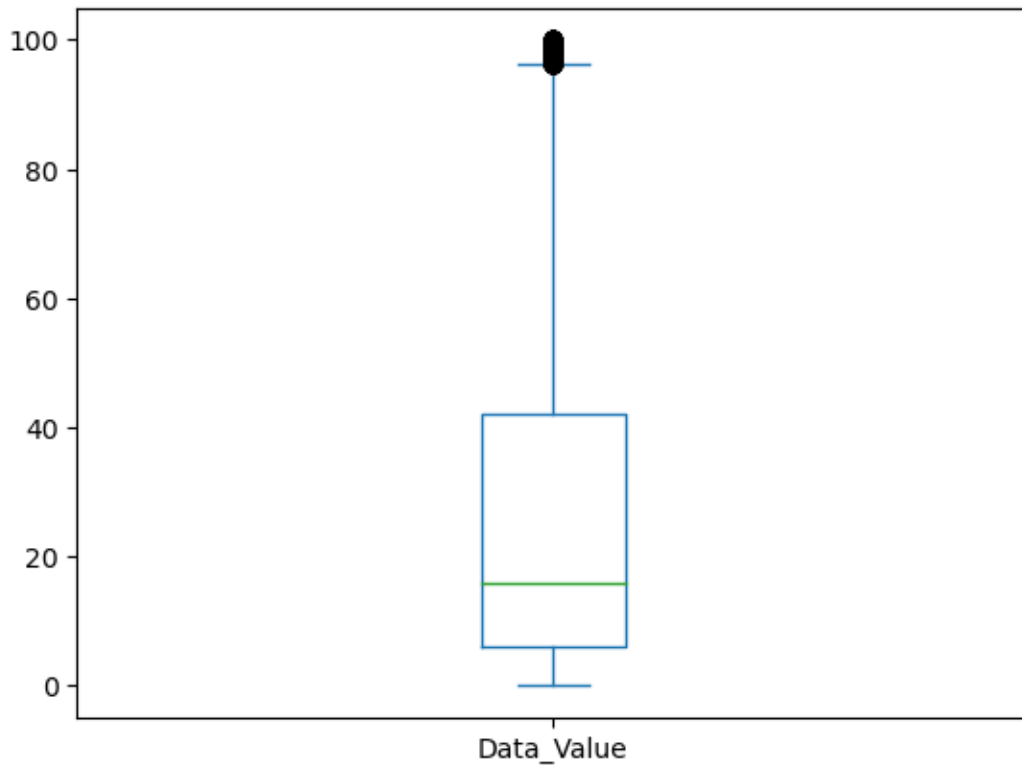


```
[ ]: #  
data_fillna=df["Data_Value"].fillna(df["Data_Value"].mode()[0])  
data_fillna.hist() #
```

```
[ ]: <Axes: >
```



```
[ ]: data_fillna.plot.box()  
plt.show() #
```



```
[ ]: df["Data_Value"].compare(data_fillna)
```

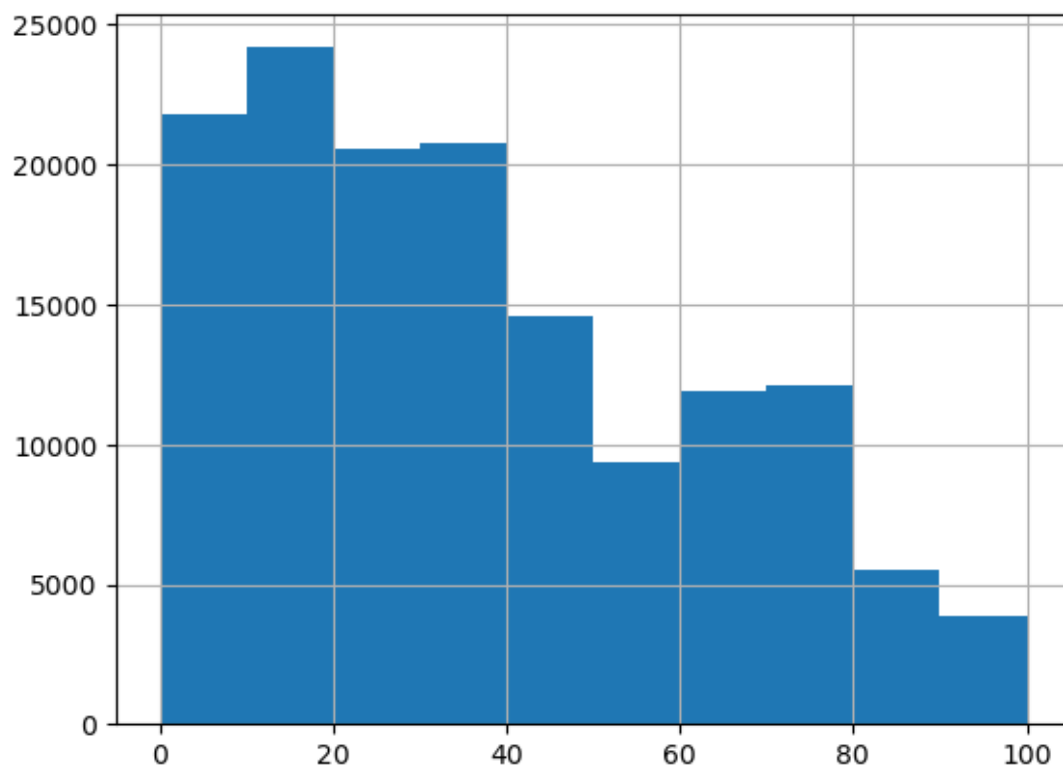
```
[ ]:      self  other
32      NaN    6.0
33      NaN    6.0
42      NaN    6.0
47      NaN    6.0
48      NaN    6.0
...      ...    ...
214456   NaN    6.0
214457   NaN    6.0
214458   NaN    6.0
214459   NaN    6.0
214461   NaN    6.0
```

```
[69833 rows x 2 columns]
```

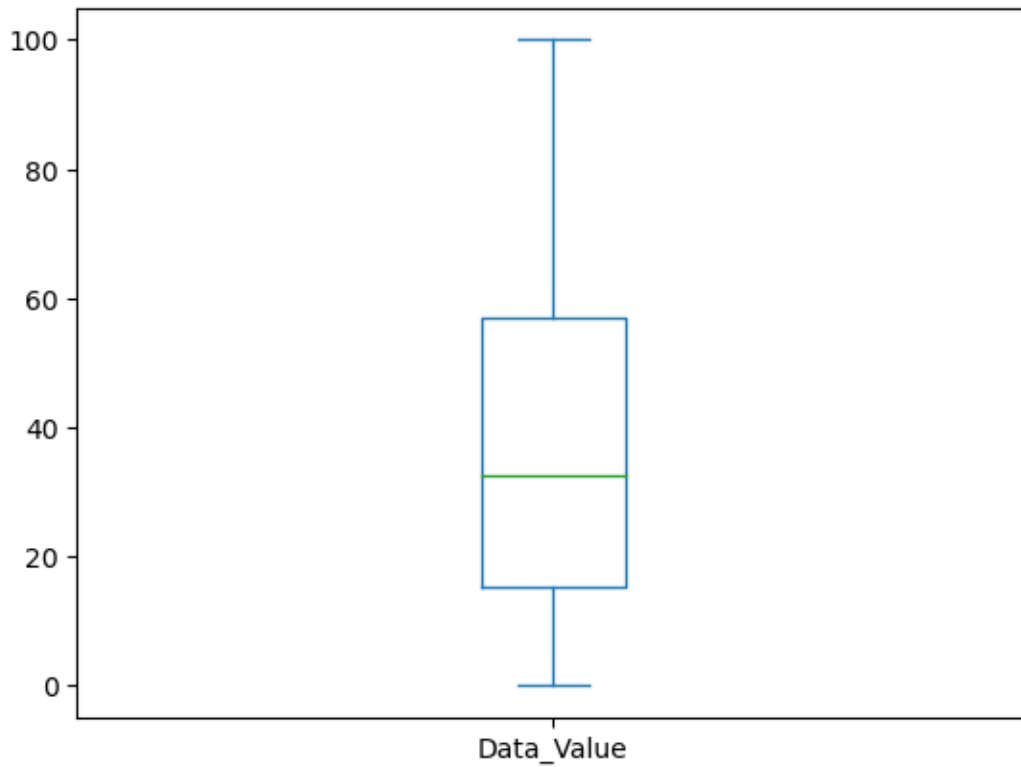
```
[ ]: #
data_fillna=df["Data_Value"].fillna(df["Data_Value_Alt"].corr(df["Data_Value"]))
↳*
df["Data_Value_Alt"]
data_fillna.hist() #
```



```
[ ]: <Axes: >
```



```
[ ]: data_fillna.plot.box()  
plt.show() #
```

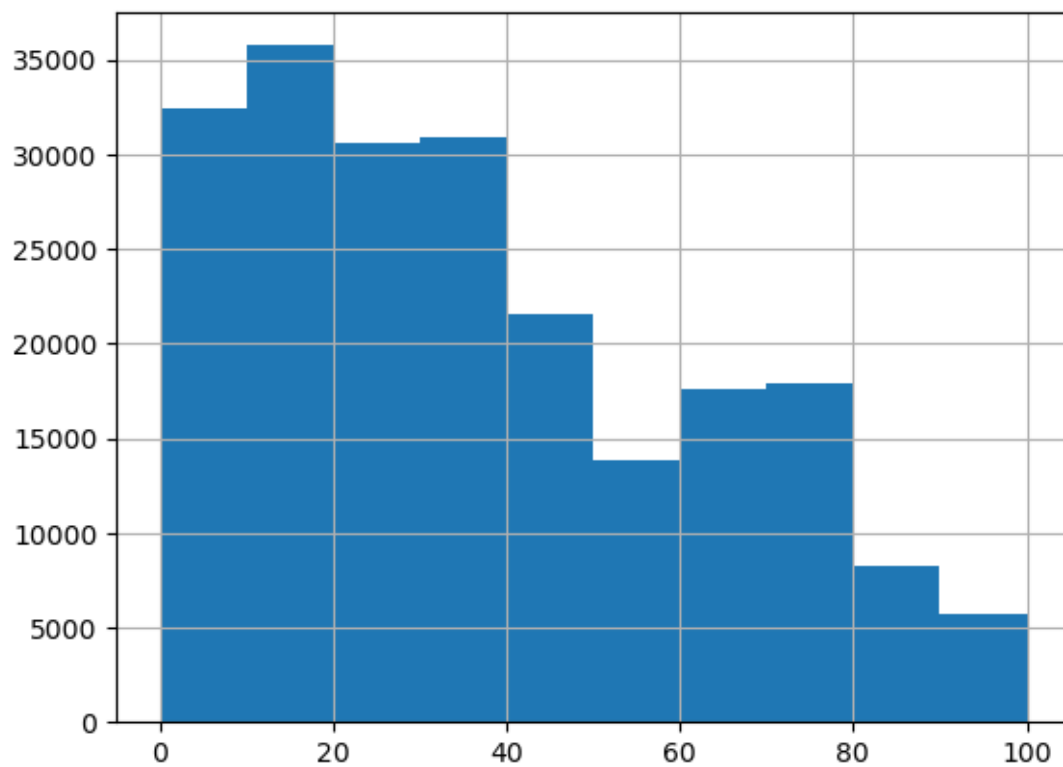


```
[ ]: df["Data_Value"].compare(data_fillna) #
```

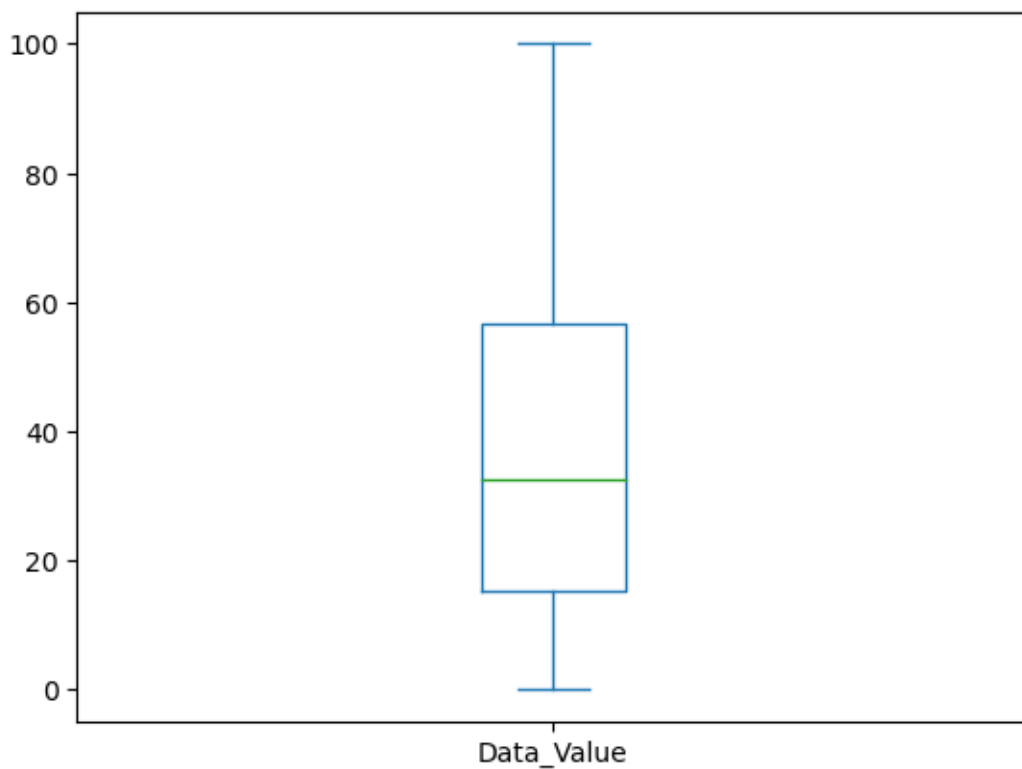
```
[ ]: Empty DataFrame  
      Columns: [self, other]  
      Index: []
```

```
[ ]: #  
      data_fillna=df["Data_Value"].interpolate(method='nearest') #  
      data_fillna.hist() #
```

```
[ ]: <Axes: >
```



```
[ ]: data_fillna.plot.box()  
plt.show() #
```



```
[ ]: df["Data_Value"].compare(data_fillna) #
```

```
[ ]:
      self  other
32      NaN   68.5
33      NaN   44.0
42      NaN    7.6
47      NaN   41.5
48      NaN   41.5
...      ...   ...
214453  NaN   80.4
214456  NaN   18.7
214457  NaN   18.7
214458  NaN   10.6
214459  NaN   10.6
```

```
[69832 rows x 2 columns]
```

# Stock

```
[ ]: import numpy as np
import pandas as pd
from pandas import DataFrame
import matplotlib.pyplot as plt
```

```
[ ]: f = './reduced_dataset-release.csv'
df = pd.read_csv(f, header=0)
```

```
/var/folders/zs/_rd25w5j2ksgfxgty_jstgl00000gn/T/ipykernel_67436/4148426267.py:2
: DtypeWarning: Columns (13) have mixed types. Specify dtype option on import or
set low_memory=False.
df = pd.read_csv(f, header=0)
```

```
[ ]: df.dtypes
```

```
[ ]: Unnamed: 0      object
TWEET              object
STOCK              object
DATE              object
LAST_PRICE        float64
1_DAY_RETURN      float64
2_DAY_RETURN      float64
3_DAY_RETURN      float64
7_DAY_RETURN      float64
PX_VOLUME         float64
VOLATILITY_10D    float64
VOLATILITY_30D    float64
LSTM_POLARITY     object
TEXTBLOB_POLARITY object
MENTION           object
dtype: object
```

```
[ ]: #
df["STOCK"].value_counts() #    Class
```

```
[ ]: Nike      3797
79.6         2710
Reuters      2482
```

```

Apple      2238
eBay       2063
...
37.52      1
1201.26    1
413.5      1
108.25     1
81.86      1
Name: STOCK, Length: 2696, dtype: int64

```

```

[ ]: nums = df["VOLATILITY_30D"] # 5 VOLATILITY_30D
nullnum = nums.isnull().sum()
nums = nums.dropna(axis = 0)
Minimum = min(nums)
Maximum = max(nums)
Q1 = np.percentile(nums, 25)
Median = np.median(nums)
Q3 = np.percentile(nums, 75)
print(" {} ".format(nullnum))
print(" {} ".format(Minimum))
print("Q1 {} ".format(Q1))
print(" {} ".format(Median))
print("Q3 {} ".format(Q3))
print(" {} ".format(Maximum))

```

```

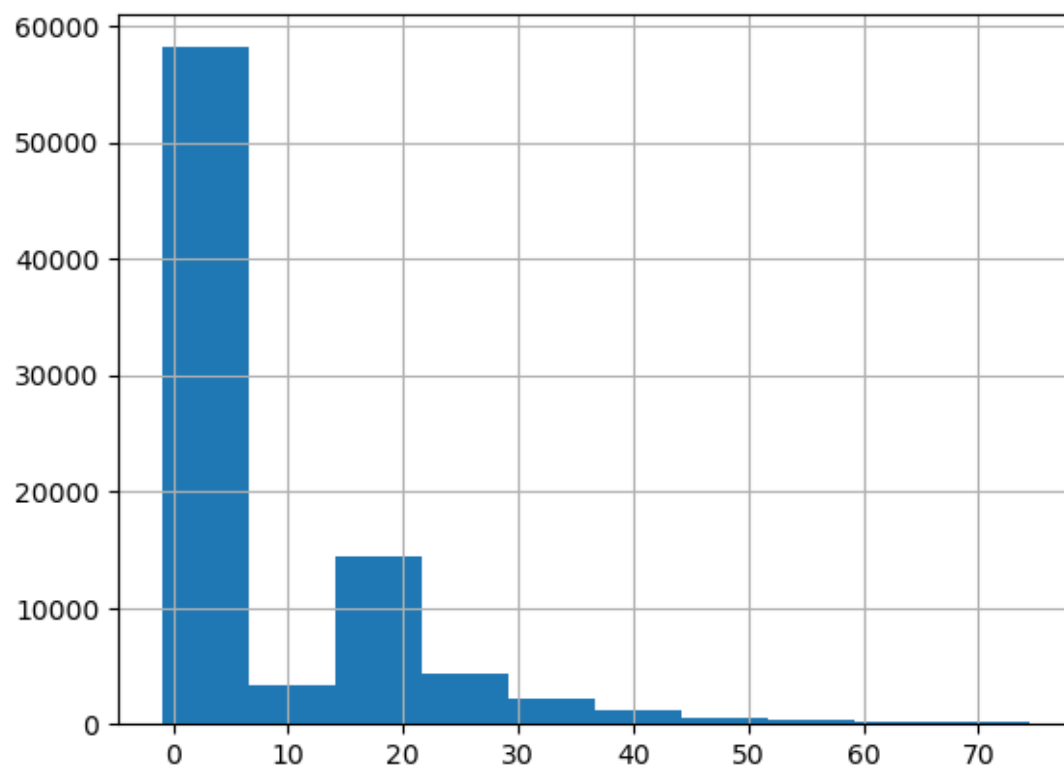
58117
-1.0
Q1 0.0
0.3
Q3 16.026
74.355

```

```

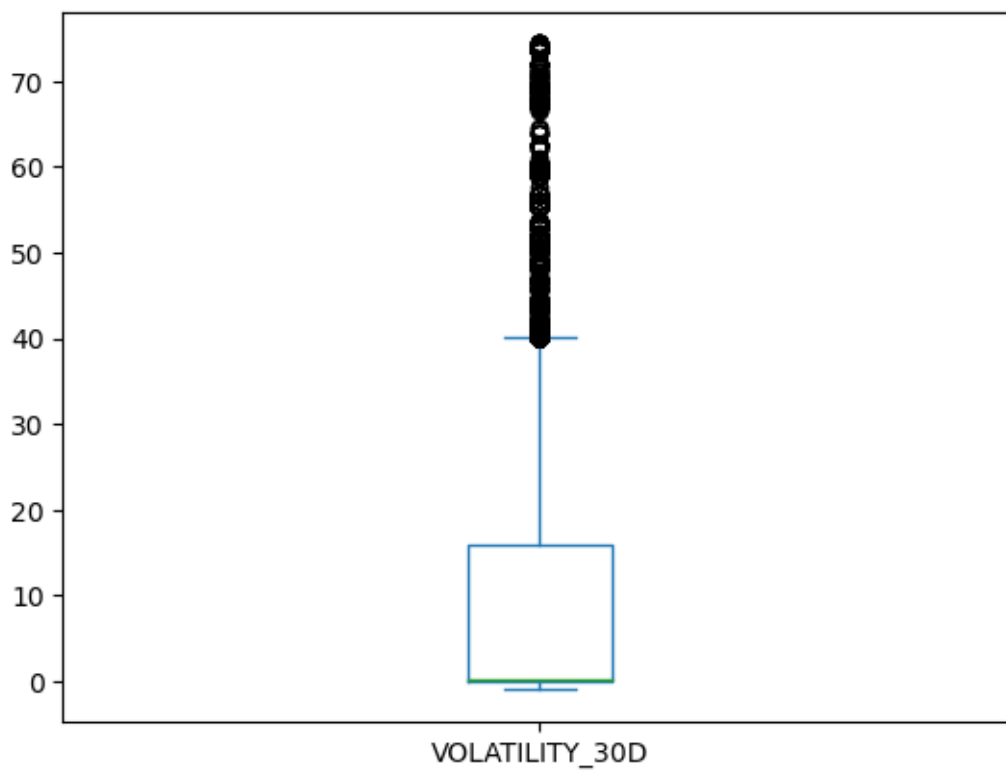
[ ]: hist = df["VOLATILITY_30D"].hist() #

```



```
[ ]: df["VOLATILITY_30D"].plot.box() #
```

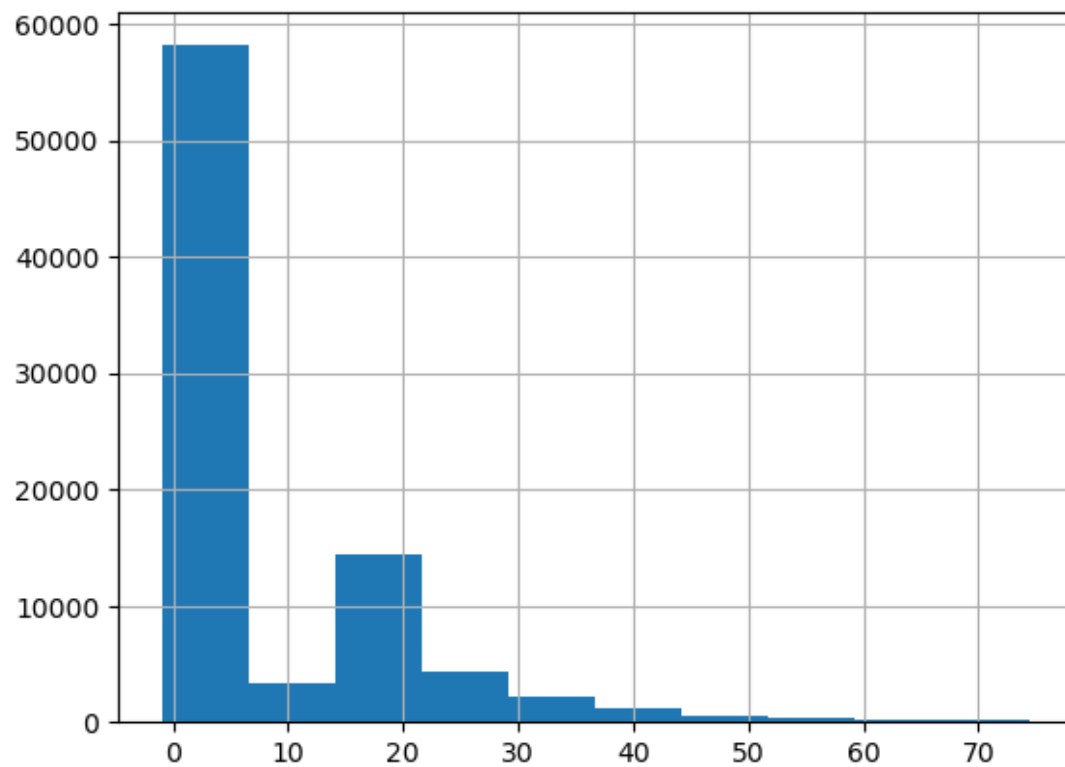
```
[ ]: <Axes: >
```



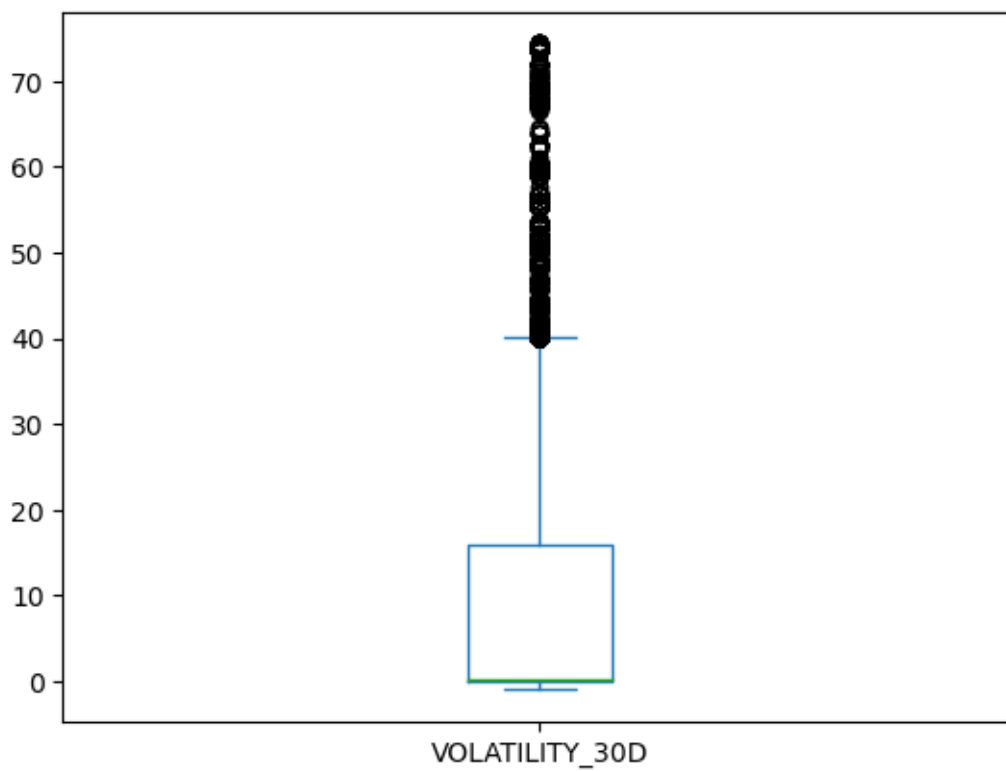
```
[ ]: #  
data_dropna = df["VOLATILITY_30D"].dropna(axis = 0)  
data_dropna.hist() #
```

```
[ ]: <Axes: >
```



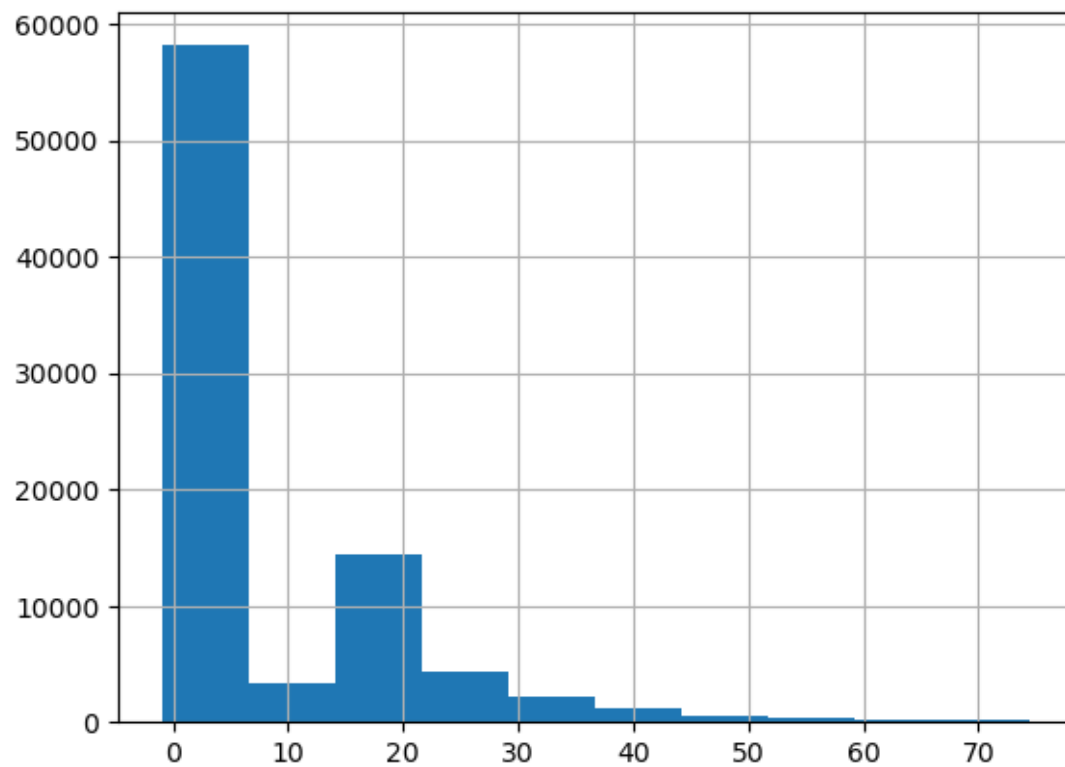


```
[ ]: data_dropna.plot.box()  
plt.show() #
```

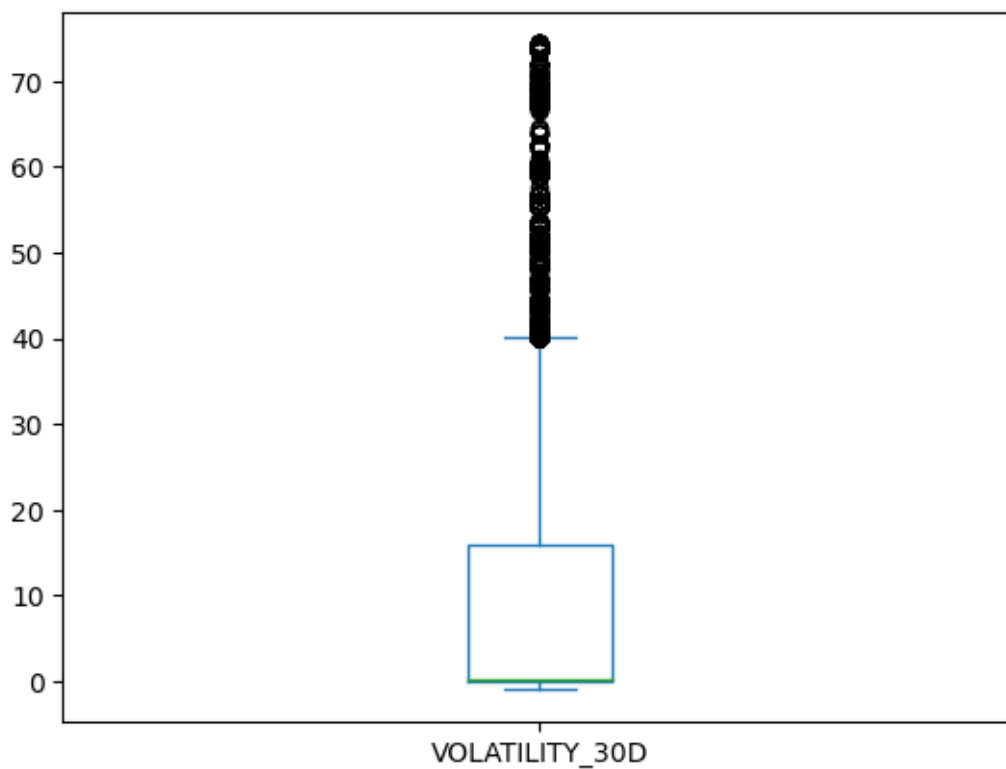


```
[ ]: #  
data_fillna=df["VOLATILITY_30D"].fillna(df["VOLATILITY_10D"].  
    ↪corr(df["VOLATILITY_30D"]) * df["VOLATILITY_10D"])  
data_fillna.hist() #
```

```
[ ]: <Axes: >
```



```
[ ]: data_fillna.plot.box()  
plt.show() #
```

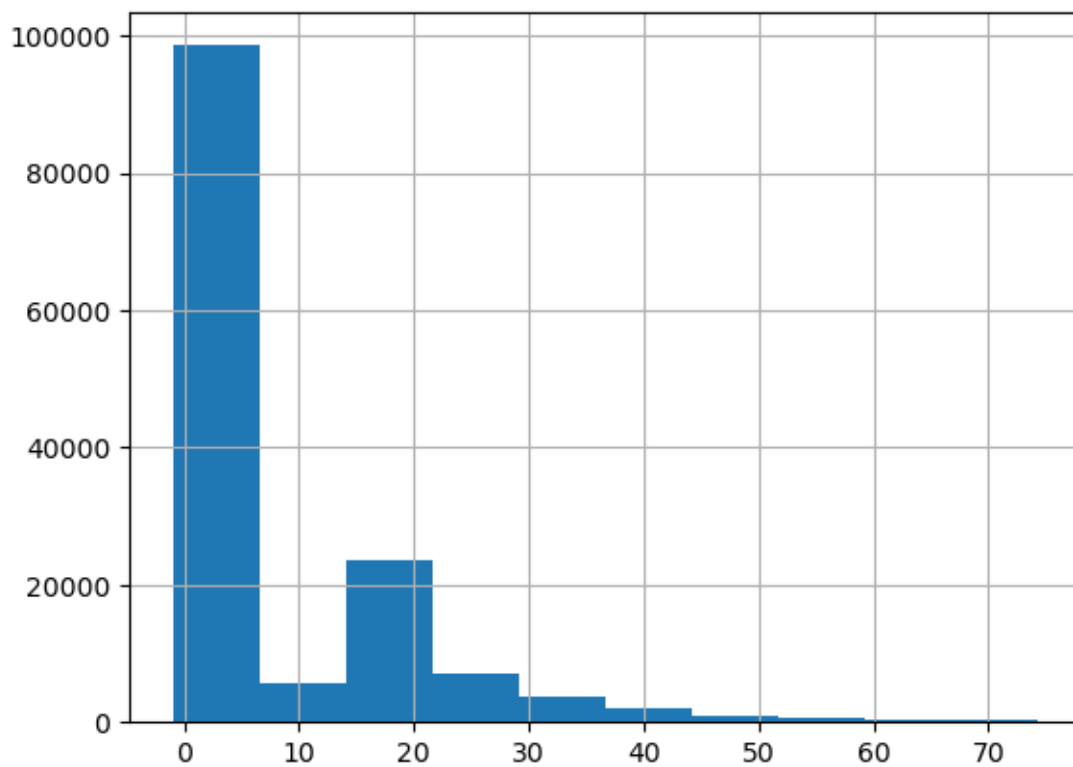


```
[ ]: df["VOLATILITY_30D"].compare(data_fillna)
```

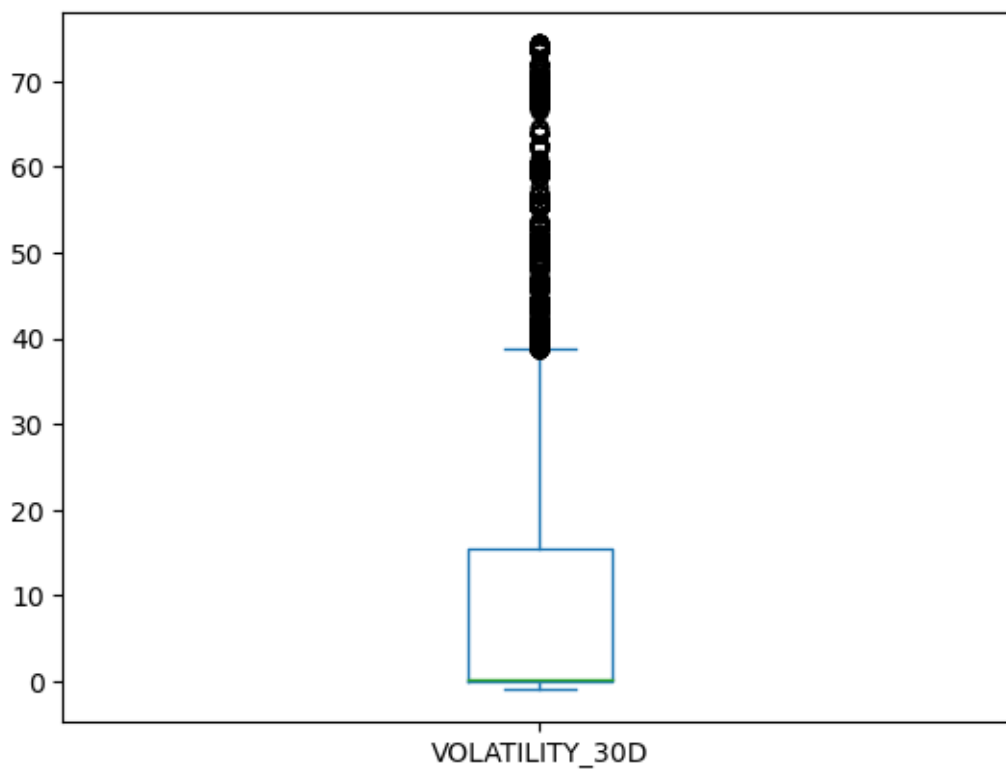
```
[ ]:      self      other
674      NaN  62.629640
3320      NaN  62.629640
4147      NaN  62.629640
5493      NaN  67.870481
5738      NaN  67.870481
5888      NaN  67.870481
5933      NaN  67.870481
6011      NaN  67.870481
6561      NaN  67.870481
61165     NaN  -0.735763
125613    NaN   0.735763
```

```
[ ]: #
data_fillna=df["VOLATILITY_30D"].interpolate(method='nearest') #
data_fillna.hist() #
```

```
[ ]: <Axes: >
```



```
[ ]: data_fillna.plot.box()  
plt.show() #
```



```
[ ]: df["VOLATILITY_30D"].compare(data_fillna) #
```

```
[ ]:      self  other
5      NaN  16.099
7      NaN  -1.000
9      NaN   1.000
12     NaN  17.298
14     NaN   1.000
...
143270  NaN  -1.000
143272  NaN   1.000
143274  NaN  -1.000
143277  NaN  22.946
143280  NaN  17.588
```

```
[58116 rows x 2 columns]
```