02032021 - Read _ Write Files in Pandas

February 27, 2021

[1]:
```
import pandas as pd
```

1 File Read and Write Support

1.0.1 Reading text file

[3]:
```
data = pd.read_csv('hills.txt')
```

[4]:
```
display (data)
```

```
   Race\t      Distance\tClimb\tTime
0      Greenmantle\t2.5\t650\t16.083
1           Carnethy\t6\t2500\t48.35
2         CraigDunain\t6\t900\t33.65
3             BenRha\t7.5\t800\t45.6
4         BenLomond\t8\t3070\t62.267
5          Goatfell\t8\t2866\t73.217
6      BensofJura\t16\t7500\t204.617
7        Cairnpapple\t6\t800\t36.367
8              Scolty\t5\t800\t29.75
9            Traprain\t6\t650\t39.75
10     LairigGhru\t28\t2100\t192.667
11            Dollar\t5\t2000\t43.05
12            Lomonds\t9.5\t2200\t65
13        CairnTable\t6\t500\t44.133
14      EildonTwo\t4.5\t1500\t26.933
15        Cairngorm\t10\t3000\t72.25
16      SevenHills\t14\t2200\t98.417
17          KnockHill\t3\t350\t78.65
18      BlackHill\t4.5\t1000\t17.417
19       CreagBeag\t5.5\t600\t32.567
20        KildconHill\t3\t300\t15.95
21  MeallAnt-Suidhe\t3.5\t1500\t27.9
22     HalfBenNevis\t6\t2200\t47.633
23           CowHill\t2\t900\t17.933
24       NBerwickLaw\t3\t600\t18.683
25        CreagDubh\t4\t2000\t26.217
```

```
26         Burnswark\t6\t800\t34.433
27          LargoLaw\t5\t950\t28.567
28          Criffel\t6.5\t1750\t50.5
29             Acmony\t5\t500\t20.95
30        BenNevis\t10\t4400\t85.583
31       Knockfarrel\t6\t600\t32.383
32    TwoBreweries\t18\t5200\t170.25
33         Cockleroi\t4.5\t850\t28.1
34    MoffatChase\t20\t5000\t159.833
```

[5]:
```
data = pd.read_csv('hills.txt', sep = '\t')

# '\t' - tab sep, when you have space " ", comma ","
```

[6]:
```
display (data)
```

	Race	Distance	Climb	Time
0	Greenmantle	2.5	650	16.083
1	Carnethy	6.0	2500	48.350
2	CraigDunain	6.0	900	33.650
3	BenRha	7.5	800	45.600
4	BenLomond	8.0	3070	62.267
5	Goatfell	8.0	2866	73.217
6	BensofJura	16.0	7500	204.617
7	Cairnpapple	6.0	800	36.367
8	Scolty	5.0	800	29.750
9	Traprain	6.0	650	39.750
10	LairigGhru	28.0	2100	192.667
11	Dollar	5.0	2000	43.050
12	Lomonds	9.5	2200	65.000
13	CairnTable	6.0	500	44.133
14	EildonTwo	4.5	1500	26.933
15	Cairngorm	10.0	3000	72.250
16	SevenHills	14.0	2200	98.417
17	KnockHill	3.0	350	78.650
18	BlackHill	4.5	1000	17.417
19	CreagBeag	5.5	600	32.567
20	KildconHill	3.0	300	15.950
21	MeallAnt-Suidhe	3.5	1500	27.900
22	HalfBenNevis	6.0	2200	47.633
23	CowHill	2.0	900	17.933
24	NBerwickLaw	3.0	600	18.683
25	CreagDubh	4.0	2000	26.217
26	Burnswark	6.0	800	34.433
27	LargoLaw	5.0	950	28.567
28	Criffel	6.5	1750	50.500
29	Acmony	5.0	500	20.950

```
30          BenNevis            10.0   4400   85.583
31       Knockfarrel             6.0    600   32.383
32      TwoBreweries            18.0   5200  170.250
33         Cockleroi             4.5    850   28.100
34       MoffatChase            20.0   5000  159.833
```

[11]:
```
display (data.head()) # first 5 rows
# display (data.head(10))

# display (data.tail()) # last 5 rows
# display (data.tail(10))
```

```
          Race        Distance  Climb    Time
0  Greenmantle             2.5    650  16.083
1     Carnethy             6.0   2500  48.350
2  CraigDunain             6.0    900  33.650
3       BenRha             7.5    800  45.600
4    BenLomond             8.0   3070  62.267
```

[12]:
```
pwd() # current working directory
```

[12]:
```
'/home/labsuser/DSwPython'
```

- to know where your jupyter notebook is located - pwd
- store files in the same location of the jupyter notebook

1.0.2 In the different directory than the jupyter notebook

[13]:
```
data = pd.read_csv('hills_temp.txt', sep = '\t')
display (data.head())

# FileNotFoundError: [Errno 2] File hills_temp.txt does not exist: 'hills_temp.
 ↪txt'
```

```
␣
↪---------------------------------------------------------------------------

FileNotFoundError                         Traceback (most recent call␣
↪last)

<ipython-input-13-eb58d2e1e560> in <module>
----> 1 data = pd.read_csv('hills_temp.txt', sep = '\t')
      2 display (data.head())
```

```
      /usr/local/lib/python3.7/site-packages/pandas/io/parsers.py in␣
↪parser_f(filepath_or_buffer, sep, delimiter, header, names, index_col,␣
↪usecols, squeeze, prefix, mangle_dupe_cols, dtype, engine, converters,␣
↪true_values, false_values, skipinitialspace, skiprows, skipfooter, nrows,␣
↪na_values, keep_default_na, na_filter, verbose, skip_blank_lines, parse_dates,␣
↪infer_datetime_format, keep_date_col, date_parser, dayfirst, cache_dates,␣
↪iterator, chunksize, compression, thousands, decimal, lineterminator,␣
↪quotechar, quoting, doublequote, escapechar, comment, encoding, dialect,␣
↪error_bad_lines, warn_bad_lines, delim_whitespace, low_memory, memory_map,␣
↪float_precision)
      674         )
      675
  --> 676         return _read(filepath_or_buffer, kwds)
      677
      678     parser_f.__name__ = name


      /usr/local/lib/python3.7/site-packages/pandas/io/parsers.py in␣
↪_read(filepath_or_buffer, kwds)
      446
      447     # Create the parser.
  --> 448     parser = TextFileReader(fp_or_buf, **kwds)
      449
      450     if chunksize or iterator:


      /usr/local/lib/python3.7/site-packages/pandas/io/parsers.py in␣
↪__init__(self, f, engine, **kwds)
      878             self.options["has_index_names"] = kwds["has_index_names"]
      879
  --> 880         self._make_engine(self.engine)
      881
      882     def close(self):


      /usr/local/lib/python3.7/site-packages/pandas/io/parsers.py in␣
↪_make_engine(self, engine)
     1112     def _make_engine(self, engine="c"):
     1113         if engine == "c":
  -> 1114             self._engine = CParserWrapper(self.f, **self.options)
     1115         else:
     1116             if engine == "python":


      /usr/local/lib/python3.7/site-packages/pandas/io/parsers.py in␣
↪__init__(self, src, **kwds)
     1889         kwds["usecols"] = self.usecols
```

```
   1890
-> 1891         self._reader = parsers.TextReader(src, **kwds)
   1892         self.unnamed_cols = self._reader.unnamed_cols
   1893

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader.__cinit__()

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader.
↪_setup_parser_source()

FileNotFoundError: [Errno 2] File hills_temp.txt does not exist:␣
↪'hills_temp.txt'
```

[14]:
```
data = pd.read_csv('/home/labsuser/Loading_Files/hills_temp.txt', sep = '\t')
display (data.head())
```

	Race	Distance	Climb	Time
0	Greenmantle	2.5	650	16.083
1	Carnethy	6.0	2500	48.350
2	CraigDunain	6.0	900	33.650
3	BenRha	7.5	800	45.600
4	BenLomond	8.0	3070	62.267

Alternatively, to grab files from any location on your computer, simply pass in the entire file path.

For Windows you need to use double so python doesn't treat the second as an escape character, a file path is in the form:

```
myfile = pd.read_csv("C:\\Users\\YourUserName\\Home\\Folder\\myfile.txt")
```

For MacOS and Linux you use slashes in the opposite direction:

```
myfile = pd.read_csv("/Users/YouUserName/Folder/myfile.txt")
```

1.0.3 Reading csv file

[15]:
```
data = pd.read_csv("winequality.csv")
display (data.head())
```

	ID	fixed acidity	volatile acidity	citric acid	residual sugar	\
0	W0001	7.0	0.27	0.36	20.7	
1	W0002	6.3	0.30	0.34	1.6	
2	W0003	8.1	0.28	0.40	6.9	
3	W0004	7.2	0.23	0.32	8.5	

```
4  W0005            7.2              0.23         0.32             8.5

   chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
0      0.045                 45.0                 170.0   1.0010  3.00
1      0.049                 14.0                 132.0   0.9940  3.30
2      0.050                 30.0                  97.0   0.9951  3.26
3      0.058                 47.0                 186.0   0.9956  3.19
4      0.058                 47.0                 186.0   0.9956  3.19

   sulphates  alcohol  quality
0       0.45      8.8      2.0
1        NaN      9.5      2.0
2        NaN     10.1      2.0
3       0.40      9.9      2.0
4       0.40      9.9      2.0
```

1.0.4 Reading excel file

[16]:
```
data = pd.read_excel('winequality.xlsx')
display (data.head())
```

```
      ID  fixed acidity  volatile acidity  citric acid  residual sugar  \
0  W0001            7.0              0.27         0.36            20.7
1  W0002            6.3              0.30         0.34             1.6
2  W0003            8.1              0.28         0.40             6.9
3  W0004            7.2              0.23         0.32             8.5
4  W0005            7.2              0.23         0.32             8.5

   chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
0      0.045                 45.0                 170.0   1.0010  3.00
1      0.049                 14.0                 132.0   0.9940  3.30
2      0.050                 30.0                  97.0   0.9951  3.26
3      0.058                 47.0                 186.0   0.9956  3.19
4      0.058                 47.0                 186.0   0.9956  3.19

   sulphates  alcohol  quality
0       0.45      8.8      2.0
1        NaN      9.5      2.0
2        NaN     10.1      2.0
3       0.40      9.9      2.0
4       0.40      9.9      2.0
```

[17]:
```
data = pd.read_excel('winequality.xlsx', 'Sheet2')
display (data.head())
```

```
   Gender Married Dependents     Education Self_Employed  ApplicantIncome  \
```

```
0    Male     Yes          2      Graduate            No             6250
1    Male      No          2      Graduate            No             5532
2  Female     Yes          0  Not Graduate            No             4100
3    Male     Yes          1      Graduate            No             4945
4  Female      No          0  Not Graduate            No             2165

   CoapplicantIncome  Loan_Amount_Term  Credit_History Property_Area  \
0               5654             180.0             1.0     Semiurban
1               4648             360.0             1.0         Rural
2                  0             360.0             NaN         Rural
3                  0             360.0             0.0         Rural
4                  0             360.0             1.0     Semiurban

  Loan_Status
0           Y
1           Y
2           Y
3           N
4           Y
```

[18]:
```
data = pd.read_excel('winequality.xlsx', 'Sheet3')
display (data.head())
```

```
      ␣
↪---------------------------------------------------------------------------

       ValueError                                Traceback (most recent call␣
↪last)

       /usr/local/lib/python3.7/site-packages/xlrd/book.py in␣
↪sheet_by_name(self, sheet_name)
       473         try:
   --> 474             sheetx = self._sheet_names.index(sheet_name)
       475         except ValueError:


       ValueError: 'Sheet3' is not in list


   During handling of the above exception, another exception occurred:


       XLRDError                                 Traceback (most recent call␣
↪last)

       <ipython-input-18-b03a663043f7> in <module>
```

```
----> 1 data = pd.read_excel('winequality.xlsx', 'Sheet3')
      2 display (data.head())


/usr/local/lib/python3.7/site-packages/pandas/io/excel/_base.py in␣
↪read_excel(io, sheet_name, header, names, index_col, usecols, squeeze, dtype,␣
↪engine, converters, true_values, false_values, skiprows, nrows, na_values,␣
↪keep_default_na, verbose, parse_dates, date_parser, thousands, comment,␣
↪skipfooter, convert_float, mangle_dupe_cols, **kwds)
      332         convert_float=convert_float,
      333         mangle_dupe_cols=mangle_dupe_cols,
--> 334         **kwds,
      335     )
      336


/usr/local/lib/python3.7/site-packages/pandas/io/excel/_base.py in␣
↪parse(self, sheet_name, header, names, index_col, usecols, squeeze,␣
↪converters, true_values, false_values, skiprows, nrows, na_values,␣
↪parse_dates, date_parser, thousands, comment, skipfooter, convert_float,␣
↪mangle_dupe_cols, **kwds)
      886             convert_float=convert_float,
      887             mangle_dupe_cols=mangle_dupe_cols,
--> 888             **kwds,
      889         )
      890


/usr/local/lib/python3.7/site-packages/pandas/io/excel/_base.py in␣
↪parse(self, sheet_name, header, names, index_col, usecols, squeeze, dtype,␣
↪true_values, false_values, skiprows, nrows, na_values, verbose, parse_dates,␣
↪date_parser, thousands, comment, skipfooter, convert_float, mangle_dupe_cols,␣
↪**kwds)
      437
      438             if isinstance(asheetname, str):
--> 439                 sheet = self.get_sheet_by_name(asheetname)
      440             else:  # assume an integer if not a string
      441                 sheet = self.get_sheet_by_index(asheetname)


/usr/local/lib/python3.7/site-packages/pandas/io/excel/_xlrd.py in␣
↪get_sheet_by_name(self, name)
       41
       42     def get_sheet_by_name(self, name):
---> 43         return self.book.sheet_by_name(name)
       44
       45     def get_sheet_by_index(self, index):
```

```
        /usr/local/lib/python3.7/site-packages/xlrd/book.py in␣
↪sheet_by_name(self, sheet_name)
        474             sheetx = self._sheet_names.index(sheet_name)
        475         except ValueError:
    --> 476             raise XLRDError('No sheet named <%r>' % sheet_name)
        477         return self.sheet_by_index(sheetx)
        478

        XLRDError: No sheet named <'Sheet3'>
```

1.0.5 Write the file

[19]:
```
data1 = pd.read_csv('hills.txt', sep = '\t')
display (data1.head())
```

	Race	Distance	Climb	Time
0	Greenmantle	2.5	650	16.083
1	Carnethy	6.0	2500	48.350
2	CraigDunain	6.0	900	33.650
3	BenRha	7.5	800	45.600
4	BenLomond	8.0	3070	62.267

[20]:
```
data1.to_csv('hills_clean.txt')
```

[21]:
```
data = pd.read_csv("winequality.csv")
display (data.head())
```

	ID	fixed acidity	volatile acidity	citric acid	residual sugar	\
0	W0001	7.0	0.27	0.36	20.7	
1	W0002	6.3	0.30	0.34	1.6	
2	W0003	8.1	0.28	0.40	6.9	
3	W0004	7.2	0.23	0.32	8.5	
4	W0005	7.2	0.23	0.32	8.5	

	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	\
0	0.045	45.0	170.0	1.0010	3.00	
1	0.049	14.0	132.0	0.9940	3.30	
2	0.050	30.0	97.0	0.9951	3.26	
3	0.058	47.0	186.0	0.9956	3.19	
4	0.058	47.0	186.0	0.9956	3.19	

	sulphates	alcohol	quality
0	0.45	8.8	2.0

```
1        NaN      9.5      2.0
2        NaN     10.1      2.0
3       0.40      9.9      2.0
4       0.40      9.9      2.0
```

[25]:
```
data.to_csv('winequality_clean.csv')
```

[27]:
```
data_clean = pd.read_csv("winequality_clean.csv")
display (data_clean.head())
```

```
   Unnamed: 0     ID  fixed acidity  volatile acidity  citric acid  \
0           0  W0001            7.0              0.27         0.36
1           1  W0002            6.3              0.30         0.34
2           2  W0003            8.1              0.28         0.40
3           3  W0004            7.2              0.23         0.32
4           4  W0005            7.2              0.23         0.32

   residual sugar  chlorides  free sulfur dioxide  total sulfur dioxide  \
0            20.7      0.045                 45.0                 170.0
1             1.6      0.049                 14.0                 132.0
2             6.9      0.050                 30.0                  97.0
3             8.5      0.058                 47.0                 186.0
4             8.5      0.058                 47.0                 186.0

   density    pH  sulphates  alcohol  quality
0   1.0010  3.00       0.45      8.8      2.0
1   0.9940  3.30        NaN      9.5      2.0
2   0.9951  3.26        NaN     10.1      2.0
3   0.9956  3.19       0.40      9.9      2.0
4   0.9956  3.19       0.40      9.9      2.0
```

Whenever we write a file out, we will get an index column

How to write the file without the index column

[28]:
```
data.to_csv('winequality_wo_index.csv', index = False)
```

[29]:
```
data2 = pd.read_csv('winequality_wo_index.csv')
display(data2.head())
```

```
      ID  fixed acidity  volatile acidity  citric acid  residual sugar  \
0  W0001            7.0              0.27         0.36            20.7
1  W0002            6.3              0.30         0.34             1.6
2  W0003            8.1              0.28         0.40             6.9
3  W0004            7.2              0.23         0.32             8.5
4  W0005            7.2              0.23         0.32             8.5
```

```
   chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
0      0.045                 45.0                 170.0   1.0010  3.00
1      0.049                 14.0                 132.0   0.9940  3.30
2      0.050                 30.0                  97.0   0.9951  3.26
3      0.058                 47.0                 186.0   0.9956  3.19
4      0.058                 47.0                 186.0   0.9956  3.19

   sulphates  alcohol  quality
0       0.45      8.8      2.0
1        NaN      9.5      2.0
2        NaN     10.1      2.0
3       0.40      9.9      2.0
4       0.40      9.9      2.0
```

How to set Index

[30]:
```
data2.set_index('ID')
```

[30]:
```
        fixed acidity  volatile acidity  citric acid  residual sugar  \
ID
W0001             7.0              0.27         0.36           20.70
W0002             6.3              0.30         0.34            1.60
W0003             8.1              0.28         0.40            6.90
W0004             7.2              0.23         0.32            8.50
W0005             7.2              0.23         0.32            8.50
…                 …                …            …               …
W2300             7.0              0.32         0.31            6.40
W2301             7.3              0.30          NaN            2.30
W2302             6.6              0.22         0.28           12.05
W2303             6.0              0.26         0.18            7.00
W2304             6.9              0.44         0.18           11.80

        chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
ID
W0001       0.045                 45.0                 170.0  1.00100  3.00
W0002       0.049                 14.0                 132.0  0.99400  3.30
W0003       0.050                 30.0                  97.0  0.99510  3.26
W0004       0.058                 47.0                 186.0  0.99560  3.19
W0005       0.058                 47.0                 186.0  0.99560  3.19
…           …                     …                    …      …        …
W2300       0.031                 38.0                 115.0  0.99235  3.38
W2301       0.043                 28.0                 125.0  0.99084   NaN
W2302       0.058                 25.0                 125.0  0.99856  3.45
W2303       0.055                 50.0                 194.0  0.99591  3.21
W2304       0.051                 26.0                 126.0  0.99750  3.23

        sulphates  alcohol  quality
```

```
ID
W0001       0.45      8.8      2.0
W0002        NaN      9.5      2.0
W0003        NaN     10.1      2.0
W0004       0.40      9.9      2.0
W0005       0.40      9.9      2.0
…            …        …        …
W2300       0.58     12.2      2.0
W2301       0.44     12.6      2.0
W2302       0.45      9.4      1.0
W2303       0.43      9.0      1.0
W2304       0.00      NaN      NaN

[2304 rows x 12 columns]
```

[31]:
```
data2 = pd.read_csv('winequality_wo_index.csv', index_col = 'ID')
display(data2.head())
```

```
      fixed acidity  volatile acidity  citric acid  residual sugar  \
ID
W0001           7.0              0.27         0.36            20.7
W0002           6.3              0.30         0.34             1.6
W0003           8.1              0.28         0.40             6.9
W0004           7.2              0.23         0.32             8.5
W0005           7.2              0.23         0.32             8.5

      chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
ID
W0001     0.045                 45.0                 170.0   1.0010  3.00
W0002     0.049                 14.0                 132.0   0.9940  3.30
W0003     0.050                 30.0                  97.0   0.9951  3.26
W0004     0.058                 47.0                 186.0   0.9956  3.19
W0005     0.058                 47.0                 186.0   0.9956  3.19

      sulphates  alcohol  quality
ID
W0001      0.45      8.8      2.0
W0002       NaN      9.5      2.0
W0003       NaN     10.1      2.0
W0004      0.40      9.9      2.0
W0005      0.40      9.9      2.0
```

[32]:
```
data.to_csv('/home/labsuser/Loading_Files/winequality_clean.csv')
```

How to drop a column

[33]:
```
data_clean = pd.read_csv('winequality_clean.csv')
display (data_clean.head())
```

```
   Unnamed: 0     ID  fixed acidity  volatile acidity  citric acid  \
0           0  W0001            7.0              0.27         0.36
1           1  W0002            6.3              0.30         0.34
2           2  W0003            8.1              0.28         0.40
3           3  W0004            7.2              0.23         0.32
4           4  W0005            7.2              0.23         0.32

   residual sugar  chlorides  free sulfur dioxide  total sulfur dioxide  \
0            20.7      0.045                 45.0                 170.0
1             1.6      0.049                 14.0                 132.0
2             6.9      0.050                 30.0                  97.0
3             8.5      0.058                 47.0                 186.0
4             8.5      0.058                 47.0                 186.0

   density    pH  sulphates  alcohol  quality
0   1.0010  3.00       0.45      8.8      2.0
1   0.9940  3.30        NaN      9.5      2.0
2   0.9951  3.26        NaN     10.1      2.0
3   0.9956  3.19       0.40      9.9      2.0
4   0.9956  3.19       0.40      9.9      2.0
```

[34]:
```
data_clean.drop('Unnamed: 0', axis = 1)

# axis = 1 == delete ea row for thsi col
# does not update the dataframe == the cols will still be in the orignal df
```

[34]:
```
          ID  fixed acidity  volatile acidity  citric acid  residual sugar  \
0     W0001            7.0              0.27         0.36           20.70
1     W0002            6.3              0.30         0.34            1.60
2     W0003            8.1              0.28         0.40            6.90
3     W0004            7.2              0.23         0.32            8.50
4     W0005            7.2              0.23         0.32            8.50
…     …              …                …            …              …
2299  W2300            7.0              0.32         0.31            6.40
2300  W2301            7.3              0.30          NaN            2.30
2301  W2302            6.6              0.22         0.28           12.05
2302  W2303            6.0              0.26         0.18            7.00
2303  W2304            6.9              0.44         0.18           11.80

      chlorides  free sulfur dioxide  total sulfur dioxide  density    pH  \
0         0.045                 45.0                 170.0  1.00100  3.00
1         0.049                 14.0                 132.0  0.99400  3.30
2         0.050                 30.0                  97.0  0.99510  3.26
```

```
3          0.058                47.0                 186.0  0.99560  3.19
4          0.058                47.0                 186.0  0.99560  3.19
…          …                    …                    …      …        …
2299       0.031                38.0                 115.0  0.99235  3.38
2300       0.043                28.0                 125.0  0.99084   NaN
2301       0.058                25.0                 125.0  0.99856  3.45
2302       0.055                50.0                 194.0  0.99591  3.21
2303       0.051                26.0                 126.0  0.99750  3.23

      sulphates  alcohol  quality
0          0.45      8.8      2.0
1           NaN      9.5      2.0
2           NaN     10.1      2.0
3          0.40      9.9      2.0
4          0.40      9.9      2.0
…          …       …        …
2299       0.58     12.2      2.0
2300       0.44     12.6      2.0
2301       0.45      9.4      1.0
2302       0.43      9.0      1.0
2303       0.00      NaN      NaN

[2304 rows x 13 columns]
```

[35]:
```
display (data_clean.head())
```

```
   Unnamed: 0     ID  fixed acidity  volatile acidity  citric acid  \
0           0  W0001            7.0              0.27         0.36
1           1  W0002            6.3              0.30         0.34
2           2  W0003            8.1              0.28         0.40
3           3  W0004            7.2              0.23         0.32
4           4  W0005            7.2              0.23         0.32

   residual sugar  chlorides  free sulfur dioxide  total sulfur dioxide  \
0            20.7      0.045                 45.0                 170.0
1             1.6      0.049                 14.0                 132.0
2             6.9      0.050                 30.0                  97.0
3             8.5      0.058                 47.0                 186.0
4             8.5      0.058                 47.0                 186.0

   density    pH  sulphates  alcohol  quality
0   1.0010  3.00       0.45      8.8      2.0
1   0.9940  3.30        NaN      9.5      2.0
2   0.9951  3.26        NaN     10.1      2.0
3   0.9956  3.19       0.40      9.9      2.0
4   0.9956  3.19       0.40      9.9      2.0
```

[36]:
```
data_clean.drop('Unnamed: 0', axis = 1, inplace = True)

# inplace = True saya apply these changes to the orignal df
```

[37]:
```
display (data_clean.head())
```

	ID	fixed acidity	volatile acidity	citric acid	residual sugar	\
0	W0001	7.0	0.27	0.36	20.7	
1	W0002	6.3	0.30	0.34	1.6	
2	W0003	8.1	0.28	0.40	6.9	
3	W0004	7.2	0.23	0.32	8.5	
4	W0005	7.2	0.23	0.32	8.5	

	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	\
0	0.045	45.0	170.0	1.0010	3.00	
1	0.049	14.0	132.0	0.9940	3.30	
2	0.050	30.0	97.0	0.9951	3.26	
3	0.058	47.0	186.0	0.9956	3.19	
4	0.058	47.0	186.0	0.9956	3.19	

	sulphates	alcohol	quality
0	0.45	8.8	2.0
1	NaN	9.5	2.0
2	NaN	10.1	2.0
3	0.40	9.9	2.0
4	0.40	9.9	2.0

[38]:
```
data_clean.drop(['fixed acidity', 'free sulfur dioxide'], axis = 1, inplace =␣
↪True)
```

[39]:
```
display (data_clean.head())
```

	ID	volatile acidity	citric acid	residual sugar	chlorides	\
0	W0001	0.27	0.36	20.7	0.045	
1	W0002	0.30	0.34	1.6	0.049	
2	W0003	0.28	0.40	6.9	0.050	
3	W0004	0.23	0.32	8.5	0.058	
4	W0005	0.23	0.32	8.5	0.058	

	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	170.0	1.0010	3.00	0.45	8.8	2.0
1	132.0	0.9940	3.30	NaN	9.5	2.0
2	97.0	0.9951	3.26	NaN	10.1	2.0
3	186.0	0.9956	3.19	0.40	9.9	2.0
4	186.0	0.9956	3.19	0.40	9.9	2.0

[41]:
```
data_clean.drop(['density'], inplace = True) # default axis will 0
# KeyError: "['density'] not found in axis"
```

```
␣
↪---------------------------------------------------------------------------

KeyError                                  Traceback (most recent call␣
↪last)

<ipython-input-41-e5962dc96f3b> in <module>
----> 1 data_clean.drop(['density'], inplace = True) # default axis will 0


/usr/local/lib/python3.7/site-packages/pandas/core/frame.py in␣
↪drop(self, labels, axis, index, columns, level, inplace, errors)
   3995             level=level,
   3996             inplace=inplace,
-> 3997             errors=errors,
   3998         )
   3999


/usr/local/lib/python3.7/site-packages/pandas/core/generic.py in␣
↪drop(self, labels, axis, index, columns, level, inplace, errors)
   3934         for axis, labels in axes.items():
   3935             if labels is not None:
-> 3936                 obj = obj._drop_axis(labels, axis, level=level,␣
↪errors=errors)
   3937
   3938         if inplace:


/usr/local/lib/python3.7/site-packages/pandas/core/generic.py in␣
↪_drop_axis(self, labels, axis, level, errors)
   3968                 new_axis = axis.drop(labels, level=level,␣
↪errors=errors)
   3969             else:
-> 3970                 new_axis = axis.drop(labels, errors=errors)
   3971             result = self.reindex(**{axis_name: new_axis})
   3972


/usr/local/lib/python3.7/site-packages/pandas/core/indexes/base.py in␣
↪drop(self, labels, errors)
   5015         if mask.any():
   5016             if errors != "ignore":
```

```
-> 5017                 raise KeyError(f"{labels[mask]} not found in axis")
   5018             indexer = indexer[~mask]
   5019         return self.delete(indexer)


KeyError: "['density'] not found in axis"
```

How to drop a row?

[40]:
```
data_clean.drop(data_clean.index[0:5], axis = 0, inplace = True) #default axis
↪will 0
display (data_clean.head())
```

	ID	volatile acidity	citric acid	residual sugar	chlorides	\
5	W0006	0.28	0.40	6.9	0.050	
6	W0007	0.32	NaN	7.0	0.045	
7	W0008	0.27	0.36	20.7	0.045	
8	W0009	0.30	0.34	1.6	0.049	
9	W0010	0.22	0.43	1.5	0.044	

	total sulfur dioxide	density	pH	sulphates	alcohol	quality
5	97.0	0.9951	3.26	0.44	10.1	2.0
6	136.0	0.9949	NaN	0.47	9.6	2.0
7	170.0	1.0010	3.00	0.45	8.8	2.0
8	132.0	0.9940	3.30	NaN	9.5	2.0
9	129.0	0.9938	3.22	0.45	11.0	2.0

[ ]:
```

```