2hsxuferh

December 19, 2024

1 1. Importing pandas

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

2 2. Opening a local csv file

[2]:
```
df = pd.read_csv('aug_train.csv')
```

[3]:
```
df
```

[3]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
3            33241  city_115                   0.789    NaN
4              666  city_162                   0.767   Male
…              …       …                       …     …
19153         7386  city_173                   0.878   Male
19154        31398  city_103                   0.920   Male
19155        24576  city_103                   0.920   Male
19156         5756   city_65                   0.802   Male
19157        23834   city_67                   0.855    NaN

           relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                          …                   …             …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School
```

```
      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                  …          …           …              …           …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                  …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0

[19158 rows x 14 columns]
```

3 3. Opening a csv file from an URL

[4]:
```
import requests
from io import StringIO

url = "https://raw.githubusercontent.com/plotly/datasets/refs/heads/master/
 ↪solar.csv"
headers = {"User-Agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.14; rv:66.
 ↪0) Gecko/20100101 Firefox/66.0"}
req = requests.get(url, headers=headers)
data = StringIO(req.text)

pd.read_csv(data)
```

[4]:
```
            State  Number of Solar Plants  Installed Capacity (MW)  \
0      California                     289                     4395
1         Arizona                      48                     1078
2          Nevada                      11                      238
```

```
3      New Mexico                    33                    261
4        Colorado                    20                    118
5           Texas                    12                    187
6  North Carolina                   148                    669
7        New York                    13                     53

   Average MW Per Plant  Generation (GWh)
0                  15.3             10826
1                  22.5              2550
2                  21.6               557
3                   7.9               590
4                   5.9               235
5                  15.6               354
6                   4.5              1162
7                   4.1                84
```

4 4. Sep Parameter ()

[5]:
```
pd.read_csv('movie_titles_metadata.tsv')
```

[5]:
```
    m0\t10 things i hate about you\t1999\t6.90\t62847\t['comedy' 'romance']
0    m1\t1492: conquest of paradise\t1992\t6.20\t10…
1    m2\t15 minutes\t2001\t6.10\t25854\t['action' '…
2    m3\t2001: a space odyssey\t1968\t8.40\t163227\…
3    m4\t48 hrs.\t1982\t6.90\t22289\t['action' 'com…
4    m5\tthe fifth element\t1997\t7.50\t133756\t['a…
..                                                 …
611  m612\twatchmen\t2009\t7.80\t135229\t['action' …
612  m613\txxx\t2002\t5.60\t53505\t['action' 'adven…
613  m614\tx-men\t2000\t7.40\t122149\t['action' 'sc…
614  m615\tyoung frankenstein\t1974\t8.00\t57618\t[…
615  m616\tzulu dawn\t1979\t6.40\t1911\t['action' '…

[616 rows x 1 columns]
```

[6]:
```
pd.read_csv('movie_titles_metadata.tsv',sep='\t')

# In case of files other than CSV format, for example separator **\t** used for␣
↪TSV files
```

[6]:
```
       m0  10 things i hate about you  1999  6.90     62847  \
0      m1  1492: conquest of paradise  1992   6.2   10421.0
1      m2                15 minutes  2001   6.1   25854.0
2      m3      2001: a space odyssey  1968   8.4  163227.0
3      m4                   48 hrs.  1982   6.9   22289.0
4      m5          the fifth element  1997   7.5  133756.0
```

```
..    …                          …   …   …       …
611  m612                  watchmen  2009   7.8  135229.0
612  m613                       xxx  2002   5.6   53505.0
613  m614                     x-men  2000   7.4  122149.0
614  m615  young frankenstein  1974   8.0   57618.0
615  m616                zulu dawn  1979   6.4    1911.0

                                  ['comedy' 'romance']
0          ['adventure' 'biography' 'drama' 'history']
1                ['action' 'crime' 'drama' 'thriller']
2                    ['adventure' 'mystery' 'sci-fi']
3       ['action' 'comedy' 'crime' 'drama' 'thriller']
4    ['action' 'adventure' 'romance' 'sci-fi' 'thri…
..                                                 …
611  ['action' 'crime' 'fantasy' 'mystery' 'sci-fi'…
612                    ['action' 'adventure' 'crime']
613                              ['action' 'sci-fi']
614                              ['comedy' 'sci-fi']
615     ['action' 'adventure' 'drama' 'history' 'war']

[616 rows x 6 columns]
```

[7]:
```
pd.read_csv('movie_titles_metadata.
 ↪tsv',sep='\t',names=['sno','name','release_year','rating','votes','genres'])

# If column names were not present originally, the 'names' parameter has been␣
 ↪defined manually with specific column names.
```

[7]:
```
      sno                       name release_year  rating     votes  \
0      m0  10 things i hate about you         1999     6.9   62847.0
1      m1  1492: conquest of paradise         1992     6.2   10421.0
2      m2                    15 minutes         2001     6.1   25854.0
3      m3       2001: a space odyssey         1968     8.4  163227.0
4      m4                      48 hrs.         1982     6.9   22289.0
..    …                          …           …     …       …
612  m612                  watchmen         2009     7.8  135229.0
613  m613                       xxx         2002     5.6   53505.0
614  m614                     x-men         2000     7.4  122149.0
615  m615  young frankenstein         1974     8.0   57618.0
616  m616                zulu dawn         1979     6.4    1911.0

                                           genres
0                                ['comedy' 'romance']
1          ['adventure' 'biography' 'drama' 'history']
2                ['action' 'crime' 'drama' 'thriller']
3                    ['adventure' 'mystery' 'sci-fi']
4       ['action' 'comedy' 'crime' 'drama' 'thriller']
```

```
..                                                   …
612  ['action' 'crime' 'fantasy' 'mystery' 'sci-fi'…
613                    ['action' 'adventure' 'crime']
614                               ['action' 'sci-fi']
615                               ['comedy' 'sci-fi']
616     ['action' 'adventure' 'drama' 'history' 'war']

[617 rows x 6 columns]
```

5 5. Index_col parameter

[8]:
```
pd.read_csv('aug_train.csv')
```

[8]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
3            33241  city_115                   0.789    NaN
4              666  city_162                   0.767   Male
…              …       …                     …    …
19153         7386  city_173                   0.878   Male
19154        31398  city_103                   0.920   Male
19155        24576  city_103                   0.920   Male
19156         5756   city_65                   0.802   Male
19157        23834   city_67                   0.855    NaN

           relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                          …                   …             …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                  …          …          …             …          …
```

```
19153       Humanities          14          NaN              NaN            1
19154             STEM          14          NaN              NaN            4
19155             STEM         >20        50-99          Pvt Ltd            4
19156              NaN          <1      500-999          Pvt Ltd            2
19157              NaN           2          NaN              NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                 …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0

[19158 rows x 14 columns]
```

[9]:
```
pd.read_csv('aug_train.csv',index_col='enrollee_id')

# The default index column is changed to the 'enrollee_id' column, as the␣
 ↪original is irrelevant to our future analysis.
```

[9]:
```
                 city  city_development_index gender      relevent_experience  \
enrollee_id
8949         city_103                   0.920   Male  Has relevent experience
29725         city_40                   0.776   Male   No relevent experience
11561         city_21                   0.624    NaN   No relevent experience
33241        city_115                   0.789    NaN   No relevent experience
666          city_162                   0.767   Male  Has relevent experience
…                 …                     …      …                       …
7386         city_173                   0.878   Male   No relevent experience
31398        city_103                   0.920   Male  Has relevent experience
24576        city_103                   0.920   Male  Has relevent experience
5756          city_65                   0.802   Male  Has relevent experience
23834         city_67                   0.855    NaN   No relevent experience

            enrolled_university education_level major_discipline experience  \
enrollee_id
8949              no_enrollment        Graduate             STEM        >20
29725             no_enrollment        Graduate             STEM         15
11561          Full time course        Graduate             STEM          5
33241                       NaN        Graduate  Business Degree         <1
666               no_enrollment         Masters             STEM        >20
```

```
…                        …              …                 …        …
7386              no_enrollment        Graduate       Humanities         14
31398             no_enrollment        Graduate             STEM         14
24576             no_enrollment        Graduate             STEM        >20
5756              no_enrollment     High School              NaN         <1
23834             no_enrollment  Primary School              NaN          2

            company_size    company_type last_new_job  training_hours  target
enrollee_id
8949                 NaN             NaN            1              36     1.0
29725              50-99         Pvt Ltd           >4              47     0.0
11561                NaN             NaN        never              83     0.0
33241                NaN         Pvt Ltd        never              52     1.0
666                50-99  Funded Startup            4               8     0.0
…                    …              …          …              …      …
7386                 NaN             NaN            1              42     1.0
31398                NaN             NaN            4              52     1.0
24576              50-99         Pvt Ltd            4              44     0.0
5756              500-999         Pvt Ltd            2              97     0.0
23834                NaN             NaN            1             127     0.0

[19158 rows x 13 columns]
```

6 6. Header parameter

[10]:
```
pd.read_csv('test.csv')
```

[10]:
```
   Unnamed: 0   Unnamed: 1 Unnamed: 2             Unnamed: 3 Unnamed: 4  \
0           0  enrollee_id       city  city_development_index     gender
1           1        29725    city_40                   0.776       Male
2           2        11561    city_21                   0.624        NaN
3           3        33241   city_115                   0.789        NaN
4           4          666   city_162                   0.767       Male

                Unnamed: 5           Unnamed: 6       Unnamed: 7  \
0      relevent_experience  enrolled_university  education_level
1   No relevent experience        no_enrollment         Graduate
2   No relevent experience     Full time course         Graduate
3   No relevent experience                  NaN         Graduate
4  Has relevent experience        no_enrollment          Masters

         Unnamed: 8  Unnamed: 9   Unnamed: 10     Unnamed: 11   Unnamed: 12  \
0  major_discipline  experience  company_size    company_type  last_new_job
1              STEM          15         50-99         Pvt Ltd            >4
2              STEM           5           NaN             NaN         never
3   Business Degree          <1           NaN         Pvt Ltd         never
```

```
4              STEM         >20         50-99  Funded Startup             4

      Unnamed: 13 Unnamed: 14
0  training_hours      target
1              47           0
2              83           0
3              52           1
4               8           0
```

[11]:
```
pd.read_csv('test.csv',header=1)

# we can specify the header number to a number where our cloumn headers starts␣
 ↪(other than topmost row)
```

[11]:
```
   0  enrollee_id      city  city_development_index gender  \
0  1        29725   city_40                   0.776   Male
1  2        11561   city_21                   0.624    NaN
2  3        33241  city_115                   0.789    NaN
3  4          666  city_162                   0.767   Male

       relevent_experience enrolled_university education_level  \
0   No relevent experience       no_enrollment        Graduate
1   No relevent experience    Full time course        Graduate
2   No relevent experience                 NaN        Graduate
3  Has relevent experience       no_enrollment         Masters

  major_discipline experience company_size    company_type last_new_job  \
0             STEM         15        50-99         Pvt Ltd           >4
1             STEM          5          NaN             NaN        never
2  Business Degree         <1          NaN         Pvt Ltd        never
3             STEM        >20        50-99  Funded Startup            4

   training_hours  target
0              47       0
1              83       0
2              52       1
3               8       0
```

7 7. use_cols parameter

[12]:
```
pd.read_csv('aug_train.csv')
```

[12]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
```

```
3           33241  city_115                  0.789    NaN
4             666  city_162                  0.767   Male
…             …       …                    …    …
19153        7386  city_173                  0.878   Male
19154       31398  city_103                  0.920   Male
19155       24576  city_103                  0.920   Male
19156        5756   city_65                  0.802   Male
19157       23834   city_67                  0.855    NaN

          relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                        …                 …             …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                 …          …          …             …          …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                 …     …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0
```

```
[19158 rows x 14 columns]
```

[13]:
```
pd.read_csv('aug_train.csv',usecols=['enrollee_id','gender','experience'])

# to use only specific columns for our analysis.
```

[13]:
```
       enrollee_id gender experience
0             8949   Male        >20
1            29725   Male         15
2            11561    NaN          5
3            33241    NaN         <1
4              666   Male        >20
…              …    …          …
19153         7386   Male         14
19154        31398   Male         14
19155        24576   Male        >20
19156         5756   Male         <1
19157        23834    NaN          2

[19158 rows x 3 columns]
```

8 9. Skiprows Parameter

[14]:
```
pd.read_csv('aug_train.csv')
```

[14]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
3            33241  city_115                   0.789    NaN
4              666  city_162                   0.767   Male
…              …       …                     …    …
19153         7386  city_173                   0.878   Male
19154        31398  city_103                   0.920   Male
19155        24576  city_103                   0.920   Male
19156         5756   city_65                   0.802   Male
19157        23834   city_67                   0.855    NaN

           relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                          …                   …             …
```

```
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                  …          …          …              …           …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                  …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0

[19158 rows x 14 columns]
```

[15]:
```
pd.read_csv('aug_train.csv',skiprows=[1,4])

# 1st and 3rd row is skipped.
```

[15]:
```
       enrollee_id      city  city_development_index gender  \
0            29725   city_40                   0.776   Male
1            11561   city_21                   0.624    NaN
2              666  city_162                   0.767   Male
3            21651  city_176                   0.764    NaN
4            28806  city_160                   0.920   Male
…            …       …                        …     …
19151         7386  city_173                   0.878   Male
19152        31398  city_103                   0.920   Male
```

```
19153        24576  city_103                  0.920   Male
19154         5756   city_65                  0.802   Male
19155        23834   city_67                  0.855    NaN

           relevent_experience enrolled_university education_level  \
0       No relevent experience       no_enrollment        Graduate
1       No relevent experience    Full time course        Graduate
2      Has relevent experience       no_enrollment         Masters
3      Has relevent experience    Part time course        Graduate
4      Has relevent experience       no_enrollment     High School
…                          …                  …             …
19151   No relevent experience       no_enrollment        Graduate
19152  Has relevent experience       no_enrollment        Graduate
19153  Has relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment     High School
19155   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM         15        50-99         Pvt Ltd           >4
1                 STEM          5          NaN             NaN        never
2                 STEM        >20        50-99  Funded Startup            4
3                 STEM         11          NaN             NaN            1
4                  NaN          5        50-99  Funded Startup            1
…                  …          …           …               …           …
19151       Humanities         14          NaN             NaN            1
19152             STEM         14          NaN             NaN            4
19153             STEM        >20        50-99         Pvt Ltd            4
19154              NaN         <1      500-999         Pvt Ltd            2
19155              NaN          2          NaN             NaN            1

       training_hours  target
0                  47     0.0
1                  83     0.0
2                   8     0.0
3                  24     1.0
4                  24     0.0
…                 …      …
19151              42     1.0
19152              52     1.0
19153              44     0.0
19154              97     0.0
19155             127     0.0

[19156 rows x 14 columns]
```

[16]:
```
pd.read_csv('aug_train.csv',skiprows=4)
```

```
# 4 number of lines is skipped at the start of the file.
```

[16]:
```
       33241  city_115  0.789 Unnamed: 3   No relevent experience  \
0        666  city_162  0.767       Male  Has relevent experience
1      21651  city_176  0.764        NaN  Has relevent experience
2      28806  city_160  0.920       Male  Has relevent experience
3        402   city_46  0.762       Male  Has relevent experience
4      27107  city_103  0.920       Male  Has relevent experience
…      …       …    …        …                      …
19149   7386  city_173  0.878       Male   No relevent experience
19150  31398  city_103  0.920       Male  Has relevent experience
19151  24576  city_103  0.920       Male  Has relevent experience
19152   5756   city_65  0.802       Male  Has relevent experience
19153  23834   city_67  0.855        NaN   No relevent experience

             Unnamed: 5        Graduate Business Degree   <1 Unnamed: 9  \
0         no_enrollment         Masters             STEM  >20      50-99
1      Part time course        Graduate             STEM   11        NaN
2         no_enrollment     High School              NaN    5      50-99
3         no_enrollment        Graduate             STEM   13        <10
4         no_enrollment        Graduate             STEM    7      50-99
…                 …             …              …  …        …
19149     no_enrollment        Graduate      Humanities   14        NaN
19150     no_enrollment        Graduate             STEM   14        NaN
19151     no_enrollment        Graduate             STEM  >20      50-99
19152     no_enrollment     High School              NaN   <1    500-999
19153     no_enrollment  Primary School              NaN    2        NaN

             Pvt Ltd never   52  1.0
0      Funded Startup     4    8  0.0
1                 NaN     1   24  1.0
2      Funded Startup     1   24  0.0
3             Pvt Ltd    >4   18  1.0
4             Pvt Ltd     1   46  1.0
…               …   …  …  …
19149            NaN     1   42  1.0
19150            NaN     4   52  1.0
19151         Pvt Ltd     4   44  0.0
19152         Pvt Ltd     2   97  0.0
19153            NaN     1  127  0.0

[19154 rows x 14 columns]
```

9 10. nrows Parameter

[17]:
```
pd.read_csv('aug_train.csv',nrows=15)

# To read only a specific Number of rows of file
```

[17]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_size	company_type	last_new_job
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	>20	NaN	NaN	1
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment	Graduate	STEM	15	50-99	Pvt Ltd	>4
2	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	STEM	5	NaN	NaN	never
3	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Business Degree	<1	NaN	Pvt Ltd	never
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	STEM	>20	50-99	Funded Startup	4
5	21651	city_176	0.764	NaN	Has relevent experience	Part time course	Graduate	STEM	11	NaN	NaN	1
6	28806	city_160	0.920	Male	Has relevent experience	no_enrollment	High School	NaN	NaN	NaN	NaN	NaN
7	402	city_46	0.762	Male	Has relevent experience	no_enrollment	Graduate	Arts	>20	NaN	NaN	1
8	27107	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	5	NaN	NaN	1
9	699	city_103	0.920	NaN	Has relevent experience	no_enrollment	Graduate	STEM	13	500-999	Pvt Ltd	2
10	29452	city_21	0.624	NaN	No relevent experience	Full time course	High School	NaN	<1	NaN	NaN	never
11	23853	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	STEM	13	50-99	Pvt Ltd	1
12	25619	city_61	0.913	Male	Has relevent experience	no_enrollment	Graduate	STEM	7	50-99	Pvt Ltd	1
13	5826	city_21	0.624	Male	No relevent experience	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	8722	city_21	0.624	NaN	No relevent experience	Full time course	High School	NaN	<1	NaN	NaN	never

```
6                NaN          5        50-99  Funded Startup            1
7               STEM         13          <10         Pvt Ltd           >4
8               STEM          7        50-99         Pvt Ltd            1
9               STEM         17       10000+         Pvt Ltd           >4
10               NaN          2          NaN             NaN        never
11              STEM          5    5000-9999         Pvt Ltd            1
12              STEM        >20    1000-4999         Pvt Ltd            3
13               NaN          2          NaN             NaN        never
14               NaN          5          NaN             NaN        never

    training_hours  target
0               36     1.0
1               47     0.0
2               83     0.0
3               52     1.0
4                8     0.0
5               24     1.0
6               24     0.0
7               18     1.0
8               46     1.0
9              123     0.0
10              32     1.0
11             108     0.0
12              23     0.0
13              24     0.0
14              26     0.0
```

10 11. Encoding parameter

[18]:
```
pd.read_csv('zomato.csv')

# If encoding format is not'utf-8'
```

```
---------------------------------------------------------------------------
UnicodeDecodeError                        Traceback (most recent call last)
Cell In[18], line 1
----> 1 pd.read_csv('zomato.csv')
      3 # If encoding format is not'utf-8'
```

```
File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1026, in read_csv(filepath_or_buffer, sep, delimiter, header, names,␣
 ↪index_col, usecols, 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, date_format, dayfirst, cache_dates, iterator,␣
 ↪chunksize, compression, thousands, decimal, lineterminator, quotechar,␣
 ↪quoting, doublequote, escapechar, comment, encoding, encoding_errors, dialect␣
 ↪on_bad_lines, delim_whitespace, low_memory, memory_map, float_precision,␣
 ↪storage_options, dtype_backend)
   1013 kwds_defaults = _refine_defaults_read(
   1014     dialect,
   1015     delimiter,
   (…)
   1022     dtype_backend=dtype_backend,
   1023 )
   1024 kwds.update(kwds_defaults)
-> 1026 return _read(filepath_or_buffer, kwds)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:620, in _read(filepath_or_buffer, kwds)
    617 _validate_names(kwds.get("names", None))
    619 # Create the parser.
--> 620 parser = TextFileReader(filepath_or_buffer, **kwds)
    622 if chunksize or iterator:
    623     return parser

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1620, in TextFileReader.__init__(self, f, engine, **kwds)
   1617     self.options["has_index_names"] = kwds["has_index_names"]
   1619 self.handles: IOHandles | None = None
-> 1620 self._engine = self._make_engine(f, self.engine)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1898, in TextFileReader._make_engine(self, f, engine)
   1895     raise ValueError(msg)
   1897 try:
-> 1898     return mapping[engine](f, **self.options)
   1899 except Exception:
   1900     if self.handles is not None:

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\c_parser_wrap
 ↪py:93, in CParserWrapper.__init__(self, src, **kwds)
     90 if kwds["dtype_backend"] == "pyarrow":
     91     # Fail here loudly instead of in cython after reading
     92     import_optional_dependency("pyarrow")
```

```
---> 93 self._reader = parsers.TextReader(src, **kwds)
     95 self.unnamed_cols = self._reader.unnamed_cols
     97 # error: Cannot determine type of 'names'

File parsers.pyx:574, in pandas._libs.parsers.TextReader.__cinit__()

File parsers.pyx:663, in pandas._libs.parsers.TextReader._get_header()

File parsers.pyx:874, in pandas._libs.parsers.TextReader._tokenize_rows()

File parsers.pyx:891, in pandas._libs.parsers.TextReader._check_tokenize_status )

File parsers.pyx:2053, in pandas._libs.parsers.raise_parser_error()

File <frozen codecs>:322, in decode(self, input, final)

UnicodeDecodeError: 'utf-8' codec can't decode byte 0xed in position 7044:␣
 ↪invalid continuation byte
```

[19]:
```
pd.read_csv('zomato.csv',encoding='latin-1')
```

[19]:

	Restaurant ID	Restaurant Name	Country Code	City	\
0	6317637	Le Petit Souffle	162	Makati City	
1	6304287	Izakaya Kikufuji	162	Makati City	
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	
3	6318506	Ooma	162	Mandaluyong City	
4	6314302	Sambo Kojin	162	Mandaluyong City	
…	…	…	…	…	
9546	5915730	NamlÛ± Gurme	208	ÛÁstanbul	
9547	5908749	Ceviz AÛôacÛ±	208	ÛÁstanbul	
9548	5915807	Huqqa	208	ÛÁstanbul	
9549	5916112	A ô ôk Kahve	208	ÛÁstanbul	
9550	5927402	Walter's Coffee Roastery	208	ÛÁstanbul	

	Address	\
0	Third Floor, Century City Mall, Kalayaan Avenu…	
1	Little Tokyo, 2277 Chino Roces Avenue, Legaspi…	
2	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal…	
3	Third Floor, Mega Fashion Hall, SM Megamall, O…	
4	Third Floor, Mega Atrium, SM Megamall, Ortigas…	
…	…	
9546	Kemanke ô Karamustafa Pa ôa Mahallesi, RÛ±htÛ±…	
9547	Ko ôuyolu Mahallesi, Muhittin íistí_ndaÛô Cadd…	
9548	Kuruí_e ôme Mahallesi, Muallim Naci Caddesi, N…	
9549	Kuruí_e ôme Mahallesi, Muallim Naci Caddesi, N…	
9550	CafeaÛôa Mahallesi, BademaltÛ± Sokak, No 21/B,…	

```
                                       Locality  \
0      Century City Mall, Poblacion, Makati City
1     Little Tokyo, Legaspi Village, Makati City
2     Edsa Shangri-La, Ortigas, Mandaluyong City
3         SM Megamall, Ortigas, Mandaluyong City
4         SM Megamall, Ortigas, Mandaluyong City
…                                           …
9546                                   Karakí_y
9547                                  Ko ôuyolu
9548                                 Kuruí_e ôme
9549                                 Kuruí_e ôme
9550                                        Moda

                                     Locality Verbose   Longitude  \
0     Century City Mall, Poblacion, Makati City, Mak…  121.027535
1     Little Tokyo, Legaspi Village, Makati City, Ma…  121.014101
2     Edsa Shangri-La, Ortigas, Mandaluyong City, Ma…  121.056831
3     SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.056475
4     SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.057508
…                                                  …         …
9546                                Karakí_y, ÛÁstanbul   28.977392
9547                               Ko ôuyolu, ÛÁstanbul   29.041297
9548                              Kuruí_e ôme, ÛÁstanbul   29.034640
9549                              Kuruí_e ôme, ÛÁstanbul   29.036019
9550                                     Moda, ÛÁstanbul   29.026016

       Latitude                          Cuisines  …          Currency  \
0     14.565443       French, Japanese, Desserts  …  Botswana Pula(P)
1     14.553708                          Japanese  …  Botswana Pula(P)
2     14.581404  Seafood, Asian, Filipino, Indian  …  Botswana Pula(P)
3     14.585318                  Japanese, Sushi  …  Botswana Pula(P)
4     14.584450                 Japanese, Korean  …  Botswana Pula(P)
…           …                                …  …             …
9546  41.022793                          Turkish  …  Turkish Lira(TL)
9547  41.009847   World Cuisine, Patisserie, Cafe  …  Turkish Lira(TL)
9548  41.055817            Italian, World Cuisine  …  Turkish Lira(TL)
9549  41.057979                  Restaurant Cafe  …  Turkish Lira(TL)
9550  40.984776                             Cafe  …  Turkish Lira(TL)

     Has Table booking Has Online delivery Is delivering now  \
0                  Yes                  No                No
1                  Yes                  No                No
2                  Yes                  No                No
3                   No                  No                No
4                  Yes                  No                No
…                    …                   …                 …
9546                No                  No                No
```

```
9547                No                  No                No
9548                No                  No                No
9549                No                  No                No
9550                No                  No                No

     Switch to order menu Price range  Aggregate rating  Rating color  \
0                      No           3               4.8    Dark Green
1                      No           3               4.5    Dark Green
2                      No           4               4.4         Green
3                      No           4               4.9    Dark Green
4                      No           4               4.8    Dark Green
…                     …           …                 …           …
9546                   No           3               4.1         Green
9547                   No           3               4.2         Green
9548                   No           4               3.7        Yellow
9549                   No           4               4.0         Green
9550                   No           2               4.0         Green

     Rating text Votes
0      Excellent   314
1      Excellent   591
2      Very Good   270
3      Excellent   365
4      Excellent   229
…          …   …
9546   Very Good   788
9547   Very Good  1034
9548        Good   661
9549   Very Good   901
9550   Very Good   591

[9551 rows x 21 columns]
```

11 12. dtypes parameter

[20]:
```
pd.read_csv('aug_train.csv')
```

[20]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
3            33241  city_115                   0.789    NaN
4              666  city_162                   0.767   Male
…              …       …                     …    …
19153         7386  city_173                   0.878   Male
19154        31398  city_103                   0.920   Male
```

```
19155        24576  city_103                  0.920   Male
19156         5756   city_65                  0.802   Male
19157        23834   city_67                  0.855    NaN

          relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                          …                   …             …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                    …          …           …               …           …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                 …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0

[19158 rows x 14 columns]
```

[21]:
```
pd.read_csv('aug_train.csv').info()
```

```
# here 'target' column is float datatype which can be changed to 'integer' to␣
 ↪save some memory
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19158 entries, 0 to 19157
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   enrollee_id             19158 non-null  int64
 1   city                    19158 non-null  object
 2   city_development_index  19158 non-null  float64
 3   gender                  14650 non-null  object
 4   relevent_experience     19158 non-null  object
 5   enrolled_university     18772 non-null  object
 6   education_level         18698 non-null  object
 7   major_discipline        16345 non-null  object
 8   experience              19093 non-null  object
 9   company_size            13220 non-null  object
 10  company_type            13018 non-null  object
 11  last_new_job            18735 non-null  object
 12  training_hours          19158 non-null  int64
 13  target                  19158 non-null  float64
dtypes: float64(2), int64(2), object(10)
memory usage: 2.0+ MB
```

[22]:
```
pd.read_csv('aug_train.csv',dtype={'target':int})

# 'target' column data type changed to 'integer' from 'float'
```

[22]:

	enrollee_id	city	city_development_index	gender	\
0	8949	city_103	0.920	Male	
1	29725	city_40	0.776	Male	
2	11561	city_21	0.624	NaN	
3	33241	city_115	0.789	NaN	
4	666	city_162	0.767	Male	
…	…	…	…	…	
19153	7386	city_173	0.878	Male	
19154	31398	city_103	0.920	Male	
19155	24576	city_103	0.920	Male	
19156	5756	city_65	0.802	Male	
19157	23834	city_67	0.855	NaN	

	relevent_experience	enrolled_university	education_level	\
0	Has relevent experience	no_enrollment	Graduate	
1	No relevent experience	no_enrollment	Graduate	
2	No relevent experience	Full time course	Graduate	
3	No relevent experience	NaN	Graduate	

```
4      Has relevent experience       no_enrollment         Masters
…                        …                 …              …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                  …          …           …              …          …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36       1
1                  47       0
2                  83       0
3                  52       1
4                   8       0
…                 …      …
19153              42       1
19154              52       1
19155              44       0
19156              97       0
19157             127       0

[19158 rows x 14 columns]
```

[23]:
```
pd.read_csv('aug_train.csv',dtype={'target':int}).info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19158 entries, 0 to 19157
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   enrollee_id             19158 non-null  int64
 1   city                    19158 non-null  object
 2   city_development_index  19158 non-null  float64
 3   gender                  14650 non-null  object
```

```
 4   relevent_experience     19158 non-null  object
 5   enrolled_university     18772 non-null  object
 6   education_level         18698 non-null  object
 7   major_discipline        16345 non-null  object
 8   experience              19093 non-null  object
 9   company_size            13220 non-null  object
 10  company_type            13018 non-null  object
 11  last_new_job            18735 non-null  object
 12  training_hours          19158 non-null  int64
 13  target                  19158 non-null  int32
dtypes: float64(1), int32(1), int64(2), object(10)
memory usage: 2.0+ MB
```

12 13. Handling Dates

[24]:
```
pd.read_csv('IPL Matches 2008-2020.csv')
```

[24]:

	id	city	date	player_of_match	\
0	335982	Bangalore	18-04-2008	BB McCullum	
1	335983	Chandigarh	19-04-2008	MEK Hussey	
2	335984	Delhi	19-04-2008	MF Maharoof	
3	335985	Mumbai	20-04-2008	MV Boucher	
4	335986	Kolkata	20-04-2008	DJ Hussey	
..	…	…	…	…	
811	1216547	Dubai	28-09-2020	AB de Villiers	
812	1237177	Dubai	05-11-2020	JJ Bumrah	
813	1237178	Abu Dhabi	06-11-2020	KS Williamson	
814	1237180	Abu Dhabi	08-11-2020	MP Stoinis	
815	1237181	Dubai	10-11-2020	TA Boult	

	venue	neutral_venue	\
0	M Chinnaswamy Stadium	0	
1	Punjab Cricket Association Stadium, Mohali	0	
2	Feroz Shah Kotla	0	
3	Wankhede Stadium	0	
4	Eden Gardens	0	
..	…	…	
811	Dubai International Cricket Stadium	0	
812	Dubai International Cricket Stadium	0	
813	Sheikh Zayed Stadium	0	
814	Sheikh Zayed Stadium	0	
815	Dubai International Cricket Stadium	0	

	team1	team2	\
0	Royal Challengers Bangalore	Kolkata Knight Riders	
1	Kings XI Punjab	Chennai Super Kings	
2	Delhi Daredevils	Rajasthan Royals	

```
3                Mumbai Indians  Royal Challengers Bangalore
4          Kolkata Knight Riders              Deccan Chargers
..                           …                           …
811  Royal Challengers Bangalore               Mumbai Indians
812              Mumbai Indians               Delhi Capitals
813  Royal Challengers Bangalore          Sunrisers Hyderabad
814              Delhi Capitals          Sunrisers Hyderabad
815              Delhi Capitals               Mumbai Indians

                    toss_winner toss_decision                       winner  \
0    Royal Challengers Bangalore         field        Kolkata Knight Riders
1            Chennai Super Kings           bat          Chennai Super Kings
2               Rajasthan Royals           bat             Delhi Daredevils
3                Mumbai Indians           bat  Royal Challengers Bangalore
4                Deccan Chargers           bat        Kolkata Knight Riders
..                           …           …                           …
811              Mumbai Indians         field  Royal Challengers Bangalore
812              Delhi Capitals         field                Mumbai Indians
813          Sunrisers Hyderabad         field          Sunrisers Hyderabad
814              Delhi Capitals           bat                Delhi Capitals
815              Delhi Capitals           bat                Mumbai Indians

      result  result_margin eliminator method      umpire1         umpire2
0       runs          140.0          N    NaN    Asad Rauf     RE Koertzen
1       runs           33.0          N    NaN    MR Benson      SL Shastri
2    wickets            9.0          N    NaN    Aleem Dar  GA Pratapkumar
3    wickets            5.0          N    NaN     SJ Davis       DJ Harper
4    wickets            5.0          N    NaN    BF Bowden     K Hariharan
..       …              …        …    …          …               …
811      tie            NaN          Y    NaN  Nitin Menon      PR Reiffel
812     runs           57.0          N    NaN  CB Gaffaney     Nitin Menon
813  wickets            6.0          N    NaN   PR Reiffel          S Ravi
814     runs           17.0          N    NaN   PR Reiffel          S Ravi
815  wickets            5.0          N    NaN  CB Gaffaney     Nitin Menon

[816 rows x 17 columns]
```

[25]:
```
pd.read_csv('IPL Matches 2008-2020.csv').info()

# 'date' column has been passed as string object by Pandas
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               816 non-null    int64
```

```
 1   city             803 non-null    object
 2   date             816 non-null    object
 3   player_of_match  812 non-null    object
 4   venue            816 non-null    object
 5   neutral_venue    816 non-null    int64
 6   team1            816 non-null    object
 7   team2            816 non-null    object
 8   toss_winner      816 non-null    object
 9   toss_decision    816 non-null    object
 10  winner           812 non-null    object
 11  result           812 non-null    object
 12  result_margin    799 non-null    float64
 13  eliminator       812 non-null    object
 14  method           19 non-null     object
 15  umpire1          816 non-null    object
 16  umpire2          816 non-null    object
dtypes: float64(1), int64(2), object(14)
memory usage: 108.5+ KB
```

[26]:
```
pd.read_csv('IPL Matches 2008-2020.csv',parse_dates=['date']).info()

# data type of 'date' column has been changed to 'datetime' which will be␣
 ↪useful for date related operations.
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               816 non-null    int64
 1   city             803 non-null    object
 2   date             816 non-null    datetime64[ns]
 3   player_of_match  812 non-null    object
 4   venue            816 non-null    object
 5   neutral_venue    816 non-null    int64
 6   team1            816 non-null    object
 7   team2            816 non-null    object
 8   toss_winner      816 non-null    object
 9   toss_decision    816 non-null    object
 10  winner           812 non-null    object
 11  result           812 non-null    object
 12  result_margin    799 non-null    float64
 13  eliminator       812 non-null    object
 14  method           19 non-null     object
 15  umpire1          816 non-null    object
 16  umpire2          816 non-null    object
dtypes: datetime64[ns](1), float64(1), int64(2), object(13)
memory usage: 108.5+ KB
```

```
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\2088694883.py:1: UserWarning:
Parsing dates in %d-%m-%Y format when dayfirst=False (the default) was
specified. Pass `dayfirst=True` or specify a format to silence this warning.
  pd.read_csv('IPL Matches 2008-2020.csv',parse_dates=['date']).info()
```

13 14. Convertors

[27]:
```
pd.read_csv('IPL Matches 2008-2020.csv')
```

[27]:

	id	city	date	player_of_match	\
0	335982	Bangalore	18-04-2008	BB McCullum	
1	335983	Chandigarh	19-04-2008	MEK Hussey	
2	335984	Delhi	19-04-2008	MF Maharoof	
3	335985	Mumbai	20-04-2008	MV Boucher	
4	335986	Kolkata	20-04-2008	DJ Hussey	
..	…	…	…	…	
811	1216547	Dubai	28-09-2020	AB de Villiers	
812	1237177	Dubai	05-11-2020	JJ Bumrah	
813	1237178	Abu Dhabi	06-11-2020	KS Williamson	
814	1237180	Abu Dhabi	08-11-2020	MP Stoinis	
815	1237181	Dubai	10-11-2020	TA Boult	

	venue	neutral_venue	\
0	M Chinnaswamy Stadium	0	
1	Punjab Cricket Association Stadium, Mohali	0	
2	Feroz Shah Kotla	0	
3	Wankhede Stadium	0	
4	Eden Gardens	0	
..	…	…	
811	Dubai International Cricket Stadium	0	
812	Dubai International Cricket Stadium	0	
813	Sheikh Zayed Stadium	0	
814	Sheikh Zayed Stadium	0	
815	Dubai International Cricket Stadium	0	

	team1	team2	\
0	Royal Challengers Bangalore	Kolkata Knight Riders	
1	Kings XI Punjab	Chennai Super Kings	
2	Delhi Daredevils	Rajasthan Royals	
3	Mumbai Indians	Royal Challengers Bangalore	
4	Kolkata Knight Riders	Deccan Chargers	
..	…	…	
811	Royal Challengers Bangalore	Mumbai Indians	
812	Mumbai Indians	Delhi Capitals	
813	Royal Challengers Bangalore	Sunrisers Hyderabad	
814	Delhi Capitals	Sunrisers Hyderabad	
815	Delhi Capitals	Mumbai Indians	

```
                    toss_winner toss_decision                     winner  \
0    Royal Challengers Bangalore         field        Kolkata Knight Riders
1            Chennai Super Kings           bat          Chennai Super Kings
2               Rajasthan Royals           bat             Delhi Daredevils
3                Mumbai Indians           bat  Royal Challengers Bangalore
4                Deccan Chargers           bat        Kolkata Knight Riders
..                           …           …                           …
811              Mumbai Indians         field  Royal Challengers Bangalore
812              Delhi Capitals         field               Mumbai Indians
813          Sunrisers Hyderabad         field         Sunrisers Hyderabad
814              Delhi Capitals           bat               Delhi Capitals
815              Delhi Capitals           bat               Mumbai Indians

      result  result_margin eliminator method      umpire1         umpire2
0       runs          140.0          N    NaN    Asad Rauf     RE Koertzen
1       runs           33.0          N    NaN    MR Benson      SL Shastri
2    wickets            9.0          N    NaN    Aleem Dar  GA Pratapkumar
3    wickets            5.0          N    NaN     SJ Davis       DJ Harper
4    wickets            5.0          N    NaN    BF Bowden     K Hariharan
..       …              …          …    …          …               …
811      tie            NaN          Y    NaN  Nitin Menon      PR Reiffel
812     runs           57.0          N    NaN  CB Gaffaney     Nitin Menon
813  wickets            6.0          N    NaN   PR Reiffel           S Ravi
814     runs           17.0          N    NaN   PR Reiffel           S Ravi
815  wickets            5.0          N    NaN  CB Gaffaney     Nitin Menon

[816 rows x 17 columns]
```

[28]:
```python
def rename(name):
    if name=='Kolkata Knight Riders':
        return 'KKR'
    elif name=='Royal Challengers Bangalore':
        return 'RCB'
    elif name=='Kings XI Punjab':
        return 'KXIP'
    elif name=='Mumbai Indians':
        return 'MI'
    elif name=='Chennai Super Kings':
        return 'CSK'
    elif name=='Rajasthan Royals':
        return 'RR'
    else:
        return name
```

[29]:
```python
rename('Kolkata Knight Riders')
```

[29]:
```
'KKR'
```

[30]:
```
pd.read_csv('IPL Matches 2008-2020.csv',converters={'team1':rename,'team2':
 ↪rename, 'toss_winner':rename, 'winner':rename})

# To apply transformation on any column values
```

[30]:
```
          id        city        date player_of_match  \
0     335982   Bangalore  18-04-2008     BB McCullum
1     335983  Chandigarh  19-04-2008      MEK Hussey
2     335984       Delhi  19-04-2008     MF Maharoof
3     335985      Mumbai  20-04-2008      MV Boucher
4     335986     Kolkata  20-04-2008       DJ Hussey
..       …          …          …              …
811  1216547       Dubai  28-09-2020  AB de Villiers
812  1237177       Dubai  05-11-2020       JJ Bumrah
813  1237178   Abu Dhabi  06-11-2020   KS Williamson
814  1237180   Abu Dhabi  08-11-2020      MP Stoinis
815  1237181       Dubai  10-11-2020        TA Boult

                                      venue  neutral_venue  \
0                     M Chinnaswamy Stadium              0
1    Punjab Cricket Association Stadium, Mohali              0
2                          Feroz Shah Kotla              0
3                          Wankhede Stadium              0
4                              Eden Gardens              0
..                                      …            …
811        Dubai International Cricket Stadium              0
812        Dubai International Cricket Stadium              0
813                    Sheikh Zayed Stadium              0
814                    Sheikh Zayed Stadium              0
815        Dubai International Cricket Stadium              0

               team1                team2          toss_winner toss_decision  \
0                RCB                  KKR                  RCB         field
1               KXIP                  CSK                  CSK           bat
2    Delhi Daredevils                   RR                   RR           bat
3                 MI                  RCB                   MI           bat
4                KKR      Deccan Chargers      Deccan Chargers           bat
..               …                    …                    …           …
811              RCB                   MI                   MI         field
812               MI       Delhi Capitals       Delhi Capitals         field
813              RCB  Sunrisers Hyderabad  Sunrisers Hyderabad         field
814    Delhi Capitals  Sunrisers Hyderabad       Delhi Capitals           bat
815    Delhi Capitals                   MI       Delhi Capitals           bat

                winner   result  result_margin eliminator method  \
```

```
0                    KKR     runs          140.0          N    NaN
1                    CSK     runs           33.0          N    NaN
2       Delhi Daredevils  wickets            9.0          N    NaN
3                    RCB  wickets            5.0          N    NaN
4                    KKR  wickets            5.0          N    NaN
..                   …      …             …         …    …
811                  RCB      tie            NaN          Y    NaN
812                   MI     runs           57.0          N    NaN
813  Sunrisers Hyderabad  wickets            6.0          N    NaN
814       Delhi Capitals     runs           17.0          N    NaN
815                   MI  wickets            5.0          N    NaN

         umpire1         umpire2
0      Asad Rauf     RE Koertzen
1      MR Benson      SL Shastri
2      Aleem Dar  GA Pratapkumar
3       SJ Davis       DJ Harper
4      BF Bowden     K Hariharan
..           …             …
811  Nitin Menon      PR Reiffel
812  CB Gaffaney     Nitin Menon
813   PR Reiffel          S Ravi
814   PR Reiffel          S Ravi
815  CB Gaffaney     Nitin Menon

[816 rows x 17 columns]
```

14 15. na_values parameter

[31]:
```
pd.read_csv('aug_train.csv').head(100)
```

[31]:
```
    enrollee_id      city  city_development_index gender  \
0          8949  city_103                   0.920   Male
1         29725   city_40                   0.776   Male
2         11561   city_21                   0.624    NaN
3         33241  city_115                   0.789    NaN
4           666  city_162                   0.767   Male
..          …       …                     …    …
95        12081   city_65                   0.802   Male
96         7364  city_160                   0.920    NaN
97        11184   city_74                   0.579    NaN
98         7016   city_65                   0.802   Male
99         8695   city_11                   0.550   Male

        relevent_experience enrolled_university education_level  \
0   Has relevent experience       no_enrollment        Graduate
```

```
1    No relevent experience       no_enrollment        Graduate
2    No relevent experience    Full time course        Graduate
3    No relevent experience                 NaN        Graduate
4   Has relevent experience       no_enrollment         Masters
..                       …                   …           …
95  Has relevent experience    Full time course        Graduate
96   No relevent experience    Full time course     High School
97   No relevent experience    Full time course        Graduate
98  Has relevent experience       no_enrollment        Graduate
99  Has relevent experience       no_enrollment        Graduate

   major_discipline experience company_size    company_type last_new_job  \
0              STEM        >20          NaN             NaN            1
1              STEM         15        50-99         Pvt Ltd           >4
2              STEM          5          NaN             NaN        never
3   Business Degree         <1          NaN         Pvt Ltd        never
4              STEM        >20        50-99  Funded Startup            4
..              …         …           …              …          …
95             STEM          9        50-99         Pvt Ltd            1
96              NaN          2      100-500         Pvt Ltd            1
97             STEM          2      100-500         Pvt Ltd            1
98             STEM          6        50-99         Pvt Ltd            2
99             STEM          6        10/49         Pvt Ltd            2

    training_hours  target
0               36     1.0
1               47     0.0
2               83     0.0
3               52     1.0
4                8     0.0
..             …     …
95              33     0.0
96             142     0.0
97              34     0.0
98              14     1.0
99              27     1.0

[100 rows x 14 columns]
```

[32]:
```
pd.read_csv('aug_train.csv',na_values=['no_enrollment','never'])

# To specify some particular values as NaN (Not a Number)
```

[32]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
```

```
3            33241  city_115                  0.789    NaN
4              666  city_162                  0.767   Male
…              …       …                     …     …
19153         7386  city_173                  0.878   Male
19154        31398  city_103                  0.920   Male
19155        24576  city_103                  0.920   Male
19156         5756   city_65                  0.802   Male
19157        23834   city_67                  0.855    NaN

           relevent_experience enrolled_university education_level  \
0      Has relevent experience                NaN        Graduate
1       No relevent experience                NaN        Graduate
2       No relevent experience   Full time course        Graduate
3       No relevent experience                NaN        Graduate
4      Has relevent experience                NaN         Masters
…                         …                    …             …
19153   No relevent experience                NaN        Graduate
19154  Has relevent experience                NaN        Graduate
19155  Has relevent experience                NaN        Graduate
19156  Has relevent experience                NaN     High School
19157   No relevent experience                NaN  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN          NaN
3      Business Degree         <1          NaN         Pvt Ltd          NaN
4                 STEM        >20        50-99  Funded Startup            4
…                    …          …            …               …           …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1

       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                 …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0
```

```
[19158 rows x 14 columns]
```

15 16. Loading a huge dataset in chunks

[33]:
```
pd.read_csv('aug_train.csv')
```

[33]:
```
       enrollee_id      city  city_development_index gender  \
0             8949  city_103                   0.920   Male
1            29725   city_40                   0.776   Male
2            11561   city_21                   0.624    NaN
3            33241  city_115                   0.789    NaN
4              666  city_162                   0.767   Male
…              …       …                       …     …
19153         7386  city_173                   0.878   Male
19154        31398  city_103                   0.920   Male
19155        24576  city_103                   0.920   Male
19156         5756   city_65                   0.802   Male
19157        23834   city_67                   0.855    NaN

           relevent_experience enrolled_university education_level  \
0      Has relevent experience       no_enrollment        Graduate
1       No relevent experience       no_enrollment        Graduate
2       No relevent experience    Full time course        Graduate
3       No relevent experience                 NaN        Graduate
4      Has relevent experience       no_enrollment         Masters
…                          …                   …             …
19153   No relevent experience       no_enrollment        Graduate
19154  Has relevent experience       no_enrollment        Graduate
19155  Has relevent experience       no_enrollment        Graduate
19156  Has relevent experience       no_enrollment     High School
19157   No relevent experience       no_enrollment  Primary School

      major_discipline experience company_size    company_type last_new_job  \
0                 STEM        >20          NaN             NaN            1
1                 STEM         15        50-99         Pvt Ltd           >4
2                 STEM          5          NaN             NaN        never
3      Business Degree         <1          NaN         Pvt Ltd        never
4                 STEM        >20        50-99  Funded Startup            4
…                    …          …           …               …           …
19153       Humanities         14          NaN             NaN            1
19154             STEM         14          NaN             NaN            4
19155             STEM        >20        50-99         Pvt Ltd            4
19156              NaN         <1      500-999         Pvt Ltd            2
19157              NaN          2          NaN             NaN            1
```

```
       training_hours  target
0                  36     1.0
1                  47     0.0
2                  83     0.0
3                  52     1.0
4                   8     0.0
…                 …      …
19153              42     1.0
19154              52     1.0
19155              44     0.0
19156              97     0.0
19157             127     0.0

[19158 rows x 14 columns]
```

[34]:
```
# Reading in chunks of 7000 rows at a time

dfnew = pd.read_csv('aug_train.csv',chunksize=7000)

# To process each chunk in a loop
for chunk in dfnew:
    # Performing operations on the chunk
    print(chunk.head(5)) # # Display shape of each chunk
```

```
   enrollee_id      city  city_development_index gender  \
0         8949  city_103                   0.920   Male
1        29725   city_40                   0.776   Male
2        11561   city_21                   0.624    NaN
3        33241  city_115                   0.789    NaN
4          666  city_162                   0.767   Male

       relevent_experience enrolled_university education_level  \
0  Has relevent experience       no_enrollment        Graduate
1   No relevent experience       no_enrollment        Graduate
2   No relevent experience    Full time course        Graduate
3   No relevent experience                 NaN        Graduate
4  Has relevent experience       no_enrollment         Masters

  major_discipline experience company_size    company_type last_new_job  \
0             STEM        >20          NaN             NaN            1
1             STEM         15        50-99         Pvt Ltd           >4
2             STEM          5          NaN             NaN        never
3  Business Degree         <1          NaN         Pvt Ltd        never
4             STEM        >20        50-99  Funded Startup            4

   training_hours  target
0              36     1.0
```

```
1              47     0.0
2              83     0.0
3              52     1.0
4               8     0.0
      enrollee_id      city  city_development_index gender  \
7000         1262  city_160                   0.920   Male
7001        29574   city_41                   0.827   Male
7002        31531   city_16                   0.910   Male
7003        13841   city_83                   0.923  Other
7004        14759  city_136                   0.897    NaN

          relevent_experience enrolled_university education_level  \
7000   No relevent experience       no_enrollment        Graduate
7001  Has relevent experience    Part time course        Graduate
7002  Has relevent experience       no_enrollment        Graduate
7003  Has relevent experience       no_enrollment        Graduate
7004  Has relevent experience    Part time course         Masters

     major_discipline experience company_size    company_type last_new_job  \
7000             STEM         17          NaN             NaN            2
7001             STEM          7          <10  Funded Startup            3
7002             STEM         12    1000-4999         Pvt Ltd            4
7003             STEM         10          NaN             NaN           >4
7004             STEM          4          NaN             NaN            2

      training_hours  target
7000              12     0.0
7001              48     0.0
7002             108     0.0
7003             196     1.0
7004             160     0.0
       enrollee_id      city  city_development_index  gender  \
14000        29120  city_103                   0.920    Male
14001        11420  city_103                   0.920  Female
14002        28511   city_16                   0.910     NaN
14003        19119  city_103                   0.920     NaN
14004        33273  city_102                   0.804    Male

          relevent_experience enrolled_university education_level  \
14000  Has relevent experience       no_enrollment        Graduate
14001  Has relevent experience       no_enrollment        Graduate
14002  Has relevent experience       no_enrollment        Graduate
14003  Has relevent experience       no_enrollment        Graduate
14004  Has relevent experience    Part time course     High School

      major_discipline experience company_size company_type last_new_job  \
14000             STEM         12      500-999      Pvt Ltd            1
14001             STEM          3       10000+      Pvt Ltd            1
```

```
14002             STEM         10        10/49      Pvt Ltd            4
14003             STEM          8      100-500         NaN            1
14004              NaN         17        50-99      Pvt Ltd            4

       training_hours  target
14000             204     0.0
14001              99     0.0
14002              12     0.0
14003              72     0.0
14004              20     0.0
```

[35]:
```
# Reading in chunks of 7000 rows at a time

dfnew = pd.read_csv('aug_train.csv',chunksize=7000)

# To process each chunk in a loop
for chunk in dfnew:
    # Performing operations on the chunk
    print(chunk.shape) # # Display shape of each chunk
```

```
(7000, 14)
(7000, 14)
(5158, 14)
```

16 17. To Skip bad lines using *on_bad_lines*

[36]:
```
pd.read_csv('BX-Books.csv')
```

```
---------------------------------------------------------------------------
ParserError                               Traceback (most recent call last)
Cell In[36], line 1
----> 1 pd.read_csv('BX-Books.csv')

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1026, in read_csv(filepath_or_buffer, sep, delimiter, header, names,␣
 ↪index_col, usecols, 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, date_format, dayfirst, cache_dates, iterator,␣
 ↪chunksize, compression, thousands, decimal, lineterminator, quotechar,␣
 ↪quoting, doublequote, escapechar, comment, encoding, encoding_errors, dialect␣
 ↪on_bad_lines, delim_whitespace, low_memory, memory_map, float_precision,␣
 ↪storage_options, dtype_backend)
   1013 kwds_defaults = _refine_defaults_read(
   1014     dialect,
   1015     delimiter,
   (…)
   1022     dtype_backend=dtype_backend,
   1023 )
```

```
   1024 kwds.update(kwds_defaults)
-> 1026 return _read(filepath_or_buffer, kwds)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:626, in _read(filepath_or_buffer, kwds)
    623     return parser
    625 with parser:
--> 626     return parser.read(nrows)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1923, in TextFileReader.read(self, nrows)
   1916 nrows = validate_integer("nrows", nrows)
   1917 try:
   1918     # error: "ParserBase" has no attribute "read"
   1919     (
   1920         index,
   1921         columns,
   1922         col_dict,
-> 1923     ) = self._engine.read(  # type: ignore[attr-defined]
   1924         nrows
   1925     )
   1926 except Exception:
   1927     self.close()

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\c_parser_wra
 ↪py:234, in CParserWrapper.read(self, nrows)
    232 try:
    233     if self.low_memory:
--> 234         chunks = self._reader.read_low_memory(nrows)
    235         # destructive to chunks
    236         data = _concatenate_chunks(chunks)

File parsers.pyx:838, in pandas._libs.parsers.TextReader.read_low_memory()

File parsers.pyx:905, in pandas._libs.parsers.TextReader._read_rows()

File parsers.pyx:874, in pandas._libs.parsers.TextReader._tokenize_rows()

File parsers.pyx:891, in pandas._libs.parsers.TextReader._check_tokenize_status )

File parsers.pyx:2061, in pandas._libs.parsers.raise_parser_error()

ParserError: Error tokenizing data. C error: Expected 1 fields in line 54, saw 
```

BX-Books.csv file is separated by semi colon (;) and encoded (latin-1)

[37]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1")
```

```
---------------------------------------------------------------------------
ParserError                               Traceback (most recent call last)
Cell In[37], line 1
----> 1 pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1")

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1026, in read_csv(filepath_or_buffer, sep, delimiter, header, names,␣
 ↪index_col, usecols, 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, date_format, dayfirst, cache_dates, iterator,␣
 ↪chunksize, compression, thousands, decimal, lineterminator, quotechar,␣
 ↪quoting, doublequote, escapechar, comment, encoding, encoding_errors, dialect␣
 ↪on_bad_lines, delim_whitespace, low_memory, memory_map, float_precision,␣
 ↪storage_options, dtype_backend)
   1013 kwds_defaults = _refine_defaults_read(
   1014     dialect,
   1015     delimiter,
   (…)
   1022     dtype_backend=dtype_backend,
   1023 )
   1024 kwds.update(kwds_defaults)
-> 1026 return _read(filepath_or_buffer, kwds)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:626, in _read(filepath_or_buffer, kwds)
    623     return parser
    625 with parser:
--> 626     return parser.read(nrows)

File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\readers.
 ↪py:1923, in TextFileReader.read(self, nrows)
   1916 nrows = validate_integer("nrows", nrows)
   1917 try:
   1918     # error: "ParserBase" has no attribute "read"
   1919     (
   1920         index,
   1921         columns,
   1922         col_dict,
-> 1923     ) = self._engine.read(  # type: ignore[attr-defined]
   1924         nrows
   1925     )
   1926 except Exception:
   1927     self.close()
```

```
File␣
 ↪~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\io\parsers\c_parser_wra
 ↪py:234, in CParserWrapper.read(self, nrows)
    232 try:
    233     if self.low_memory:
--> 234         chunks = self._reader.read_low_memory(nrows)
    235         # destructive to chunks
    236         data = _concatenate_chunks(chunks)

File parsers.pyx:838, in pandas._libs.parsers.TextReader.read_low_memory()

File parsers.pyx:905, in pandas._libs.parsers.TextReader._read_rows()

File parsers.pyx:874, in pandas._libs.parsers.TextReader._tokenize_rows()

File parsers.pyx:891, in pandas._libs.parsers.TextReader._check_tokenize_status )

File parsers.pyx:2061, in pandas._libs.parsers.raise_parser_error()

ParserError: Error tokenizing data. C error: Expected 8 fields in line 6451, sa␣
 ↪9
```

on_bad_lines{'error', 'warn', 'skip'} or Callable, default 'error' Specifies what to do upon encountering a bad line (a line with too many fields). Allowed values are :

error, raise an Exception when a bad line is encountered.

warn, raise a warning when a bad line is encountered and skip that line.

skip, skip bad lines without raising or warning when they are encountered

Reference link

and few rows have more values than the column

[38]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
```

```
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\3888888656.py:1:
ParserWarning: Skipping line 6451: expected 8 fields, saw 9
Skipping line 43666: expected 8 fields, saw 10
Skipping line 51750: expected 8 fields, saw 9

  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\3888888656.py:1:
ParserWarning: Skipping line 92037: expected 8 fields, saw 9
Skipping line 104318: expected 8 fields, saw 9
Skipping line 121767: expected 8 fields, saw 9
```

```
  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\3888888656.py:1:
ParserWarning: Skipping line 144057: expected 8 fields, saw 9
Skipping line 150788: expected 8 fields, saw 9
Skipping line 157127: expected 8 fields, saw 9
Skipping line 180188: expected 8 fields, saw 9
Skipping line 185737: expected 8 fields, saw 9

  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\3888888656.py:1:
ParserWarning: Skipping line 209387: expected 8 fields, saw 9
Skipping line 220625: expected 8 fields, saw 9
Skipping line 227932: expected 8 fields, saw 11
Skipping line 228956: expected 8 fields, saw 10
Skipping line 245932: expected 8 fields, saw 9
Skipping line 251295: expected 8 fields, saw 9
Skipping line 259940: expected 8 fields, saw 9
Skipping line 261528: expected 8 fields, saw 9

  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\3888888656.py:1:
DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or
set low_memory=False.
  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='warn')
```

[38]:

	0195153448	Classical Mythology	\
0	0002005018	Clara Callan	
1	0060973129	Decision in Normandy	
2	0374157065	Flu: The Story of the Great Influenza Pandemic…	
3	0393045218	The Mummies of Urumchi	
4	0399135782	The Kitchen God's Wife	
…	…	…	
271354	0440400988	There's a Bat in Bunk Five	
271355	0525447644	From One to One Hundred	
271356	006008667X	Lily Dale : The True Story of the Town that Ta…	
271357	0192126040	Republic (World's Classics)	
271358	0767409752	A Guided Tour of Rene Descartes' Meditations o…	

	Mark P. O. Morford	2002	\
0	Richard Bruce Wright	2001	
1	Carlo D'Este	1991	
2	Gina Bari Kolata	1999	
3	E. J. W. Barber	1999	
4	Amy Tan	1991	
…	…	…	
271354	Paula Danziger	1988	
271355	Teri Sloat	1991	

```
271356      Christine Wicker  2004
271357                Plato  1996
271358   Christopher  Biffle  2000

                                Oxford University Press  \
0                                 HarperFlamingo Canada
1                                      HarperPerennial
2                                 Farrar Straus Giroux
3                            W. W. Norton &amp;amp; Company
4                                     Putnam Pub Group
…                                                    …
271354                  Random House Childrens Pub (Mm)
271355                                      Dutton Books
271356                                 HarperSanFrancisco
271357                              Oxford University Press
271358  McGraw-Hill Humanities/Social Sciences/Languages

       http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
```

```
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

[271359 rows x 8 columns]
```

[39]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='skip')
```

```
C:\Users\Sandeep\AppData\Local\Temp\ipykernel_5216\986491662.py:1: DtypeWarning:
Columns (3) have mixed types. Specify dtype option on import or set
low_memory=False.
  pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1", on_bad_lines='skip')
```

[39]:
```
        0195153448                                Classical Mythology  \
0       0002005018                                       Clara Callan
1       0060973129                                Decision in Normandy
2       0374157065  Flu: The Story of the Great Influenza Pandemic…
3       0393045218                              The Mummies of Urumchi
4       0399135782                              The Kitchen God's Wife
…              …                                                   …
271354  0440400988                          There's a Bat in Bunk Five
271355  0525447644                             From One to One Hundred
271356  006008667X  Lily Dale : The True Story of the Town that Ta…
271357  0192126040                        Republic (World's Classics)
271358  0767409752  A Guided Tour of Rene Descartes' Meditations o…

          Mark P. O. Morford  2002  \
0       Richard Bruce Wright  2001
1              Carlo D'Este  1991
2          Gina Bari Kolata  1999
3           E. J. W. Barber  1999
4                  Amy Tan  1991
…                       …   …
271354       Paula Danziger  1988
271355           Teri Sloat  1991
271356     Christine Wicker  2004
271357                Plato  1996
271358   Christopher  Biffle  2000

                                Oxford University Press  \
0                                 HarperFlamingo Canada
1                                       HarperPerennial
```

```
2                                   Farrar Straus Giroux
3                             W. W. Norton &amp;amp; Company
4                                       Putnam Pub Group
…                                                     …
271354                    Random House Childrens Pub (Mm)
271355                                       Dutton Books
271356                                  HarperSanFrancisco
271357                               Oxford University Press
271358  McGraw-Hill Humanities/Social Sciences/Languages

       http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                     …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
```

```
271358  http://images.amazon.com/images/P/0767409752.0…

[271359 rows x 8 columns]
```

low_memory: bool, default True

Internally process the file in chunks, resulting in lower memory use while parsing, but *possibly mixed type inference*. To ensure no mixed types either set False, or specify the type with the dtype parameter.

[40]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1",␣
↪on_bad_lines='skip',low_memory=False)
```

[40]:
```
         0195153448                                Classical Mythology  \
0       0002005018                                       Clara Callan
1       0060973129                                Decision in Normandy
2       0374157065  Flu: The Story of the Great Influenza Pandemic…
3       0393045218                              The Mummies of Urumchi
4       0399135782                              The Kitchen God's Wife
…              …                                                   …
271354  0440400988                          There's a Bat in Bunk Five
271355  0525447644                             From One to One Hundred
271356  006008667X  Lily Dale : The True Story of the Town that Ta…
271357  0192126040                        Republic (World's Classics)
271358  0767409752  A Guided Tour of Rene Descartes' Meditations o…

           Mark P. O. Morford  2002  \
0       Richard Bruce Wright  2001
1               Carlo D'Este  1991
2          Gina Bari Kolata  1999
3           E. J. W. Barber  1999
4                   Amy Tan  1991
…                        …   …
271354        Paula Danziger  1988
271355             Teri Sloat  1991
271356      Christine Wicker  2004
271357                 Plato  1996
271358   Christopher  Biffle  2000

                                  Oxford University Press  \
0                                   HarperFlamingo Canada
1                                         HarperPerennial
2                                    Farrar Straus Giroux
3                               W. W. Norton &amp; Company
4                                        Putnam Pub Group
…                                                      …
271354                  Random House Childrens Pub (Mm)
271355                                     Dutton Books
```

```
271356                                HarperSanFrancisco
271357                            Oxford University Press
271358  McGraw-Hill Humanities/Social Sciences/Languages

       http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                      …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg  \
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                      …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

       http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg
0       http://images.amazon.com/images/P/0002005018.0…
1       http://images.amazon.com/images/P/0060973129.0…
2       http://images.amazon.com/images/P/0374157065.0…
3       http://images.amazon.com/images/P/0393045218.0…
4       http://images.amazon.com/images/P/0399135782.0…
…                                                      …
271354  http://images.amazon.com/images/P/0440400988.0…
271355  http://images.amazon.com/images/P/0525447644.0…
271356  http://images.amazon.com/images/P/006008667X.0…
271357  http://images.amazon.com/images/P/0192126040.0…
271358  http://images.amazon.com/images/P/0767409752.0…

[271359 rows x 8 columns]
```

headers names are not correct.

[41]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1",␣
↪on_bad_lines='skip',low_memory=False).info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271359 entries, 0 to 271358
Data columns (total 8 columns):
 #   Column                                                Non-Null
Count   Dtype
---  ------
--------------   -----
 0   0195153448                                            271359 non-
null  object
 1   Classical Mythology                                   271359 non-
null  object
 2   Mark P. O. Morford                                    271357 non-
null  object
 3   2002                                                  271359 non-
null  object
 4   Oxford University Press                               271357 non-
null  object
 5   http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg  271359 non-
null  object
 6   http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg  271359 non-
null  object
 7   http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg  271356 non-
null  object
dtypes: object(8)
memory usage: 16.6+ MB
```

[45]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1",␣
↪on_bad_lines='skip',low_memory=False,
            names=['ISBN','Book-Title','Book-Author','Year of Publication',␣
↪'Publisher', 'Image URL'])
```

[45]:
```
ISBN  \
0195153448 Classical Mythology                                Mark P. O.
Morford
0002005018 Clara Callan                                     Richard Bruce
Wright
0060973129 Decision in Normandy                                   Carlo
D'Este
0374157065 Flu: The Story of the Great Influenza Pandemic …      Gina Bari
Kolata
0393045218 The Mummies of Urumchi                              E. J. W.
Barber
…
…
```

```
0440400988 There's a Bat in Bunk Five                                    Paula
Danziger
0525447644 From One to One Hundred                                          Teri
Sloat
006008667X Lily Dale : The True Story of the Town that Tal…      Christine
Wicker
0192126040 Republic (World's Classics)
Plato
0767409752 A Guided Tour of Rene Descartes' Meditations on…   Christopher
Biffle

                                                              Book-Title  \
0195153448 Classical Mythology                                      2002
0002005018 Clara Callan                                             2001
0060973129 Decision in Normandy                                     1991
0374157065 Flu: The Story of the Great Influenza Pandemic …        1999
0393045218 The Mummies of Urumchi                                   1999
…                                                                    …
0440400988 There's a Bat in Bunk Five                               1988
0525447644 From One to One Hundred                                  1991
006008667X Lily Dale : The True Story of the Town that Tal…        2004
0192126040 Republic (World's Classics)                              1996
0767409752 A Guided Tour of Rene Descartes' Meditations on…        2000

                    Book-Author  \
0195153448 Classical Mythology
Oxford University Press
0002005018 Clara Callan
HarperFlamingo Canada
0060973129 Decision in Normandy
HarperPerennial
0374157065 Flu: The Story of the Great Influenza Pandemic …
Farrar Straus Giroux
0393045218 The Mummies of Urumchi
W. W. Norton &amp;amp; Company
…
…
0440400988 There's a Bat in Bunk Five
Random House Childrens Pub (Mm)
0525447644 From One to One Hundred
Dutton Books
006008667X Lily Dale : The True Story of the Town that Tal…
HarperSanFrancisco
0192126040 Republic (World's Classics)
Oxford University Press
0767409752 A Guided Tour of Rene Descartes' Meditations on…  McGraw-Hill
Humanities/Social Sciences/Languages
```

```
             Year of Publication  \
0195153448 Classical Mythology
http://images.amazon.com/images/P/0195153448.0…
0002005018 Clara Callan
http://images.amazon.com/images/P/0002005018.0…
0060973129 Decision in Normandy
http://images.amazon.com/images/P/0060973129.0…
0374157065 Flu: The Story of the Great Influenza Pandemic …
http://images.amazon.com/images/P/0374157065.0…
0393045218 The Mummies of Urumchi
http://images.amazon.com/images/P/0393045218.0…
…
…
0440400988 There's a Bat in Bunk Five
http://images.amazon.com/images/P/0440400988.0…
0525447644 From One to One Hundred
http://images.amazon.com/images/P/0525447644.0…
006008667X Lily Dale : The True Story of the Town that Tal…
http://images.amazon.com/images/P/006008667X.0…
0192126040 Republic (World's Classics)
http://images.amazon.com/images/P/0192126040.0…
0767409752 A Guided Tour of Rene Descartes' Meditations on…
http://images.amazon.com/images/P/0767409752.0…

                       Publisher  \
0195153448 Classical Mythology
http://images.amazon.com/images/P/0195153448.0…
0002005018 Clara Callan
http://images.amazon.com/images/P/0002005018.0…
0060973129 Decision in Normandy
http://images.amazon.com/images/P/0060973129.0…
0374157065 Flu: The Story of the Great Influenza Pandemic …
http://images.amazon.com/images/P/0374157065.0…
0393045218 The Mummies of Urumchi
http://images.amazon.com/images/P/0393045218.0…
…
…
0440400988 There's a Bat in Bunk Five
http://images.amazon.com/images/P/0440400988.0…
0525447644 From One to One Hundred
http://images.amazon.com/images/P/0525447644.0…
006008667X Lily Dale : The True Story of the Town that Tal…
http://images.amazon.com/images/P/006008667X.0…
0192126040 Republic (World's Classics)
http://images.amazon.com/images/P/0192126040.0…
0767409752 A Guided Tour of Rene Descartes' Meditations on…
```

```
http://images.amazon.com/images/P/0767409752.0…

                        Image URL
0195153448 Classical Mythology
http://images.amazon.com/images/P/0195153448.0…
0002005018 Clara Callan
http://images.amazon.com/images/P/0002005018.0…
0060973129 Decision in Normandy
http://images.amazon.com/images/P/0060973129.0…
0374157065 Flu: The Story of the Great Influenza Pandemic …
http://images.amazon.com/images/P/0374157065.0…
0393045218 The Mummies of Urumchi
http://images.amazon.com/images/P/0393045218.0…
…
…
0440400988 There's a Bat in Bunk Five
http://images.amazon.com/images/P/0440400988.0…
0525447644 From One to One Hundred
http://images.amazon.com/images/P/0525447644.0…
006008667X Lily Dale : The True Story of the Town that Tal…
http://images.amazon.com/images/P/006008667X.0…
0192126040 Republic (World's Classics)
http://images.amazon.com/images/P/0192126040.0…
0767409752 A Guided Tour of Rene Descartes' Meditations on…
http://images.amazon.com/images/P/0767409752.0…

[271360 rows x 6 columns]
```

[44]:
```
pd.read_csv('BX-Books.csv', sep=';', encoding="latin-1",␣
 ↪on_bad_lines='skip',low_memory=False,
            names=['ISBN','Book-Title','Book-Author','Year of Publication',␣
 ↪'Publisher', 'Image URL']).info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 271360 entries, ('0195153448', 'Classical Mythology') to
('0767409752', "A Guided Tour of Rene Descartes' Meditations on First Philosophy
with Complete Translations of the Meditations by Ronald Rubin")
Data columns (total 6 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   ISBN                 271358 non-null  object
 1   Book-Title           271360 non-null  object
 2   Book-Author          271358 non-null  object
 3   Year of Publication  271360 non-null  object
 4   Publisher            271360 non-null  object
 5   Image URL            271357 non-null  object
dtypes: object(6)
```

```
memory usage: 34.5+ MB
```

[ ]: