

Welcome to 30 Days ML

How to upload / Import Data

```
In [6]: # pandas
import pandas as pd
```

```
In [7]: pd.read_csv('hotel_bookings.csv')
```

```
Out[7]:
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_nun
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	
3	Resort Hotel	0	13	2015	July	
4	Resort Hotel	0	14	2015	July	
...
119385	City Hotel	0	23	2017	August	
119386	City Hotel	0	102	2017	August	
119387	City Hotel	0	34	2017	August	
119388	City Hotel	0	109	2017	August	
119389	City Hotel	0	205	2017	August	

119390 rows × 32 columns

Gathering Data With CSV

```
In [8]: import pandas as pd
```

```
In [9]: df = pd.read_csv('aug_train.csv')
```

```
In [10]: df
```

Out[10]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns

3. Opening a csv file from an URL

```
In [11]: import requests
from io import StringIO
url = "https://raw.githubusercontent.com/cs109/2014_data/master/countries.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)
```

Out[11]:

	Country	Region
0	Algeria	AFRICA
1	Angola	AFRICA
2	Benin	AFRICA
3	Botswana	AFRICA
4	Burkina	AFRICA
...
189	Paraguay	SOUTH AMERICA
190	Peru	SOUTH AMERICA
191	Suriname	SOUTH AMERICA
192	Uruguay	SOUTH AMERICA
193	Venezuela	SOUTH AMERICA

194 rows × 2 columns

4. Sep Parameter

In [12]: `pd.read_csv('movie_titles_metadata.tsv')`

Out[12]:

	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 × 1 columns

In [13]: `pd.read_csv('movie_titles_metadata.tsv', sep='\t')`

Out[13]:

	m0	10 things i hate about you	1999	6.90	62847	['comedy' 'romance']
0	m1	1492: conquest of paradise	1992	6.2	10421.0	['adventure' 'biography' 'drama' 'history']
1	m2	15 minutes	2001	6.1	25854.0	['action' 'crime' 'drama' 'thriller']
2	m3	2001: a space odyssey	1968	8.4	163227.0	['adventure' 'mystery' 'sci-fi']
3	m4	48 hrs.	1982	6.9	22289.0	['action' 'comedy' 'crime' 'drama' 'thriller']
4	m5	the fifth element	1997	7.5	133756.0	['action' 'adventure' 'romance' 'sci-fi' 'thri...']
...
611	m612	watchmen	2009	7.8	135229.0	['action' 'crime' 'fantasy' 'mystery' 'sci-fi'...]
612	m613	xxx	2002	5.6	53505.0	['action' 'adventure' 'crime']
613	m614	x-men	2000	7.4	122149.0	['action' 'sci-fi']
614	m615	young frankenstein	1974	8.0	57618.0	['comedy' 'sci-fi']
615	m616	zulu dawn	1979	6.4	1911.0	['action' 'adventure' 'drama' 'history' 'war']

616 rows × 6 columns

In [14]: `pd.read_csv('movie_titles_metadata.tsv', sep='\t', names=['sno', 'name', 'release_year', 'rating', 'votes', 'genres'])`

Out[14]:

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

617 rows × 6 columns

5. Index_col parameter

```
In [15]: pd.read_csv('aug_train.csv')
```

```
Out[15]:
```

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns

6. Header parameter

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

```
Out[18]:
```

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	U
0	0	enrollee_id	city	city_development_index	gender	relevent_experience	enrollme
1	1	29725	city_40	0.776	Male	No relevent experience	no.
2	2	11561	city_21	0.624	NaN	No relevent experience	Full .
3	3	33241	city_115	0.789	NaN	No relevent experience	
4	4	666	city_162	0.767	Male	Has relevent experience	no.

```
In [19]: pd.read_csv('test.csv', header=1)
```

Out[19]:

	0	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_universit
0	1	29725	city_40	0.776	Male	No relevent experience	no_enrollmen
1	2	11561	city_21	0.624	NaN	No relevent experience	Full time cours
2	3	33241	city_115	0.789	NaN	No relevent experience	NaN
3	4	666	city_162	0.767	Male	Has relevent experience	no_enrollmen

7. Use_col parameter

In [20]: `pd.read_csv('aug_train.csv',usecols=['enrollee_id','gender','education_level'])`

Out[20]:

	enrollee_id	gender	education_level
0	8949	Male	Graduate
1	29725	Male	Graduate
2	11561	NaN	Graduate
3	33241	NaN	Graduate
4	666	Male	Masters
...
19153	7386	Male	Graduate
19154	31398	Male	Graduate
19155	24576	Male	Graduate
19156	5756	Male	High School
19157	23834	NaN	Primary School

19158 rows × 3 columns

8. Squeeze parameters

In [27]: `pd.read_csv('aug_train.csv',usecols=['gender'],squeeze=True)`

```
-----
TypeError                                Traceback (most recent call last)
Cell In[27], line 1
----> 1 pd.read_csv('aug_train.csv',usecols=['gender'],squeeze=True)

TypeError: read_csv() got an unexpected keyword argument 'squeeze'
```

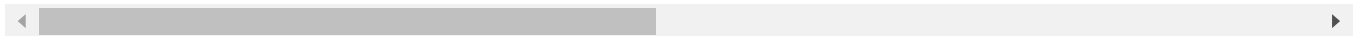
9. Skiprows/nrows Parameter

In [28]: `pd.read_csv('aug_train.csv')`

Out[28]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns



In [29]:

```
pd.read_csv('aug_train.csv',skiprows=[0,1])
```

Out[29]:

	29725	city_40	0.7759999999999999	Male	No relevent experience	no_enrollment	Graduate	ST
0	11561	city_21	0.624	NaN	No relevent experience	Full time course	Graduate	ST
1	33241	city_115	0.789	NaN	No relevent experience	NaN	Graduate	Busin Deg
2	666	city_162	0.767	Male	Has relevent experience	no_enrollment	Masters	ST
3	21651	city_176	0.764	NaN	Has relevent experience	Part time course	Graduate	ST
4	28806	city_160	0.920	Male	Has relevent experience	no_enrollment	High School	M
...
19151	7386	city_173	0.878	Male	No relevent experience	no_enrollment	Graduate	Humani
19152	31398	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	ST
19153	24576	city_103	0.920	Male	Has relevent experience	no_enrollment	Graduate	ST
19154	5756	city_65	0.802	Male	Has relevent experience	no_enrollment	High School	M
19155	23834	city_67	0.855	NaN	No relevent experience	no_enrollment	Primary School	M

19156 rows × 14 columns

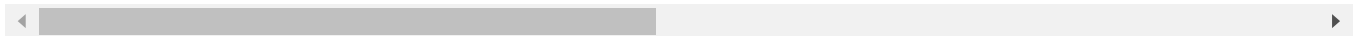


```
In [30]: pd.read_csv('aug_train.csv',skiprows=[5,6])
```


Out[30]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	28806	city_160	0.920	Male	Has relevent experience	no_enrollme
...
19151	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19152	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19153	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19154	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19155	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19156 rows × 14 columns

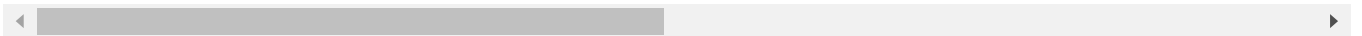


In [31]: `pd.read_csv('aug_train.csv',nrows=100)`

Out[31]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollment
1	29725	city_40	0.776	Male	No relevent experience	no_enrollment
2	11561	city_21	0.624	NaN	No relevent experience	Full time course
3	33241	city_115	0.789	NaN	No relevent experience	NaN
4	666	city_162	0.767	Male	Has relevent experience	no_enrollment
...
95	12081	city_65	0.802	Male	Has relevent experience	Full time course
96	7364	city_160	0.920	NaN	No relevent experience	Full time course
97	11184	city_74	0.579	NaN	No relevent experience	Full time course
98	7016	city_65	0.802	Male	Has relevent experience	no_enrollment
99	8695	city_11	0.550	Male	Has relevent experience	no_enrollment

100 rows × 7 columns



10. Encoding parameter

In [32]:

```
pd.read_csv('zomato.csv')
```

UnicodeDecodeError

Traceback (most recent call last)

Cell In[32], line 1

----> 1 pd.read_csv('zomato.csv')

File ~\anaconda3\Lib\site-packages\pandas\io\parsers\readers.py:912, 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)

```
    899 kwds_defaults = _refine_defaults_read(
    900     dialect,
    901     delimiter,
    (... )
    908     dtype_backend=dtype_backend,
    909 )
    910 kwds.update(kwds_defaults)
--> 912 return _read(filepath_or_buffer, kwds)
```

File ~\anaconda3\Lib\site-packages\pandas\io\parsers\readers.py:577, in _read(filepath_or_buffer, kwds)

```
    574 _validate_names(kwds.get("names", None))
    576 # Create the parser.
--> 577 parser = TextFileReader(filepath_or_buffer, **kwds)
    579 if chunksize or iterator:
    580     return parser
```

File ~\anaconda3\Lib\site-packages\pandas\io\parsers\readers.py:1407, in TextFileReader.__init__(self, f, engine, **kwds)

```
    1404 self.options["has_index_names"] = kwds["has_index_names"]
    1406 self.handles: IOHandles | None = None
-> 1407 self._engine = self._make_engine(f, self.engine)
```

File ~\anaconda3\Lib\site-packages\pandas\io\parsers\readers.py:1679, in TextFileReader._make_engine(self, f, engine)

```
    1676 raise ValueError(msg)
    1678 try:
-> 1679     return mapping[engine](f, **self.options)
    1680 except Exception:
    1681     if self.handles is not None:
```

File ~\anaconda3\Lib\site-packages\pandas\io\parsers\c_parser_wrapper.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 ~\anaconda3\Lib\site-packages\pandas_libs\parsers.pyx:550, in pandas._libs.parsers.TextReader.__cinit__()

File ~\anaconda3\Lib\site-packages\pandas_libs\parsers.pyx:639, in pandas._libs.parsers.TextReader._get_header()

File ~\anaconda3\Lib\site-packages\pandas_libs\parsers.pyx:850, in pandas._libs.parsers.TextReader._tokenize_rows()

File ~\anaconda3\Lib\site-packages\pandas_libs\parsers.pyx:861, in pandas._libs.parsers.TextReader._tokenize_rows()

```
arsers.TextReader._check_tokenize_status()
```

```
File ~\anaconda3\Lib\site-packages\pandas\_libs\parsers.pyx:2021, in pandas._libs.parsers.raise_parser_error()
```

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

```
In [33]: pd.read_csv('zomato.csv', encoding='latin-1')
```

Out[33]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Latitude
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	12.1824
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	12.1824
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	12.1824
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	12.1824
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	12.1824
...
9546	5915730	Namlı Gurme	208	Üstanbul	Kemankeş Karamustafa Paşa Mahallesi, Rıhtım...	Karaköy	Karaköy, Üstanbul	29.1026
9547	5908749	Ceviz Aca	208	Üstanbul	Koşuyolu Mahallesi, Muhittin İstinda Cadd...	Koşuyolu	Koşuyolu, Üstanbul	29.1026
9548	5915807	Huqqa	208	Üstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, Üstanbul	29.1026
9549	5916112	Afak Kahve	208	Üstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, Üstanbul	29.1026
9550	5927402	Walter's Coffee Roastery	208	Üstanbul	Cafea Üsküdar Mahallesi, Bademaltı Sokak, No 21/B,...	Moda	Moda, Üstanbul	29.1026

11. Skip bad lines

```
In [34]: pd.read_csv('zomato.csv', address=';', encoding="latin-1")
```

```
-----
TypeError                                Traceback (most recent call last)
Cell In[34], line 1
----> 1 pd.read_csv('zomato.csv', address=';', encoding="latin-1")

TypeError: read_csv() got an unexpected keyword argument 'address'
```

```
In [35]: pd.read_csv('zomato.csv', address=';', encoding="latin-1", error_bad_lines=False)
```

```
-----
TypeError                                Traceback (most recent call last)
Cell In[35], line 1
----> 1 pd.read_csv('zomato.csv', address=';', encoding="latin-1", error_bad_lines=
False)

TypeError: read_csv() got an unexpected keyword argument 'address'
```

12.dtypes parameter

```
In [37]: pd.read_csv('aug_train.csv').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  float64
dtypes: float64(2), int64(2), object(10)
memory usage: 2.0+ MB
```

```
In [38]: 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

```

13. Handling Dates

```
In [39]: pd.read_csv('IPL Matches 2008-2020.csv').info()
```

```

<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

```

```
In [40]: pd.read_csv('IPL Matches 2008-2020.csv', parse_dates=['date']).info()
```

```
<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
```

```
In [41]: pd.read_csv('IPL Matches 2008-2020.csv', parse_dates=['date']).info()
```

```
<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
```

```
In [42]: pd.read_csv('IPL Matches 2008-2020.csv')
```

Cell In[42], line 1

```
pd.read_csv('IPL Matches 2008-2020.csv')
```

SyntaxError: unterminated string literal (detected at line 1)

14. Convertors

```
In [44]: pd.read_csv('IPL Matches 2008-2020.csv', converters={'team1': rename})
```



```

-----
NameError                                Traceback (most recent call last)
Cell In[44], line 1
----> 1 pd.read_csv('IPL Matches 2008-2020.csv',converters={'team1':rename})

NameError: name 'rename' is not defined

```

15. na_values parameter

In [45]: `pd.read_csv('aug_train.csv')`

Out[45]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

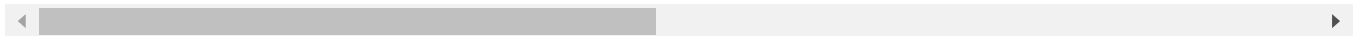
19158 rows × 14 columns

In [46]: `pd.read_csv('aug_train.csv',na_values=['Male',])`

Out[46]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	NaN	Has relevent experience	no_enrollme
1	29725	city_40	0.776	NaN	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	NaN	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	NaN	No relevent experience	no_enrollme
19154	31398	city_103	0.920	NaN	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	NaN	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	NaN	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns

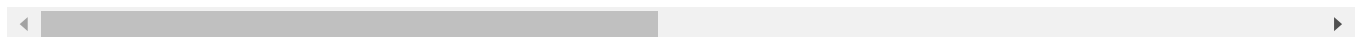


16. Loading a huge dataset in chunks

In [47]: `pd.read_csv('aug_train.csv')`

Out[47]:	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns



```
In [56]: dfs = pd.read_csv('aug_train.csv', chunksize=5000)
```

```
Cell In[56], line 2
    for chunks in dfs:
        ^
SyntaxError: incomplete input
```

```
In [58]: for chunks in dfs:
        print(chunk.shape)
```

```
-----
NameError                                Traceback (most recent call last)
Cell In[58], line 2
      1 for chunks in dfs:
----> 2     print(chunk.shape)

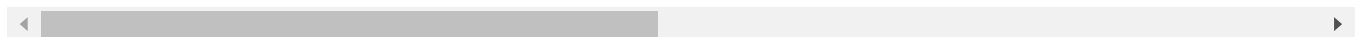
NameError: name 'chunk' is not defined
```

```
In [59]: pd.read_csv('aug_train.csv')
```

Out[59]:

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns



```
In [63]: dfs = pd.read_csv('aug_train.csv', chunksize=5000)
for chunks in dfs:
```

```
Cell In[63], line 2
    for chunks in dfs:
        ^
SyntaxError: incomplete input
```

In [62]:

```
Cell In[62], line 1
    for chunks in dfs:
        ^
SyntaxError: incomplete input
```

Working with JSON/SQL

```
In [64]: import pandas as pd
```

```
In [65]: pd.read_json('train.json')
```

Out[65]:

	id	cuisine	ingredients
0	10259	greek	[romaine lettuce, black olives, grape tomatoes...
1	25693	southern_us	[plain flour, ground pepper, salt, tomatoes, g...
2	20130	filipino	[eggs, pepper, salt, mayonaise, cooking oil, g...
3	22213	indian	[water, vegetable oil, wheat, salt]
4	13162	indian	[black pepper, shallots, cornflour, cayenne pe...
...
39769	29109	irish	[light brown sugar, granulated sugar, butter, ...
39770	11462	italian	[KRAFT Zesty Italian Dressing, purple onion, b...
39771	2238	irish	[eggs, citrus fruit, raisins, sourdough starte...
39772	41882	chinese	[boneless chicken skinless thigh, minced garli...
39773	2362	mexican	[green chile, jalapeno chilies, onions, ground...

39774 rows × 3 columns

In [66]: `pd.read_json('https://api.exchangerate-api.com/v4/latest/INR')`

Out[66]:

	provider	WARNING_UPGRADE_TO_V6		terms	base	date
AED	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
AFN	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
ALL	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
AMD	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
ANG	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
...	
XPF	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
YER	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
ZAR	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
ZMW	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	
ZWL	https://www.exchangerate-api.com	https://www.exchangerate-api.com/docs/free	https://www.exchangerate-api.com/terms	INR	2023-11-23	

162 rows × 7 columns

```
In [67]: !pip install mysql.connector
```

Collecting mysql.connector

Downloading mysql-connector-2.2.9.tar.gz (11.9 MB)

```
----- 0.0/11.9 MB ? eta -:--:--
----- 0.0/11.9 MB ? eta -:--:--
----- 0.1/11.9 MB 1.0 MB/s eta 0:00:12
----- 0.3/11.9 MB 1.5 MB/s eta 0:00:08
----- 0.4/11.9 MB 1.9 MB/s eta 0:00:06
----- 0.6/11.9 MB 2.3 MB/s eta 0:00:05
----- 0.9/11.9 MB 2.7 MB/s eta 0:00:04
----- 1.4/11.9 MB 3.7 MB/s eta 0:00:03
----- 1.9/11.9 MB 4.3 MB/s eta 0:00:03
----- 2.2/11.9 MB 4.6 MB/s eta 0:00:03
----- 2.7/11.9 MB 4.8 MB/s eta 0:00:02
----- 3.0/11.9 MB 5.1 MB/s eta 0:00:02
----- 3.4/11.9 MB 5.2 MB/s eta 0:00:02
----- 4.0/11.9 MB 5.6 MB/s eta 0:00:02
----- 4.3/11.9 MB 5.7 MB/s eta 0:00:02
----- 4.6/11.9 MB 5.7 MB/s eta 0:00:02
----- 4.7/11.9 MB 5.5 MB/s eta 0:00:02
----- 5.2/11.9 MB 5.3 MB/s eta 0:00:02
----- 5.4/11.9 MB 5.1 MB/s eta 0:00:02
----- 6.2/11.9 MB 5.6 MB/s eta 0:00:02
----- 6.8/11.9 MB 5.6 MB/s eta 0:00:01
----- 7.1/11.9 MB 5.8 MB/s eta 0:00:01
----- 7.4/11.9 MB 5.7 MB/s eta 0:00:01
----- 7.6/11.9 MB 5.6 MB/s eta 0:00:01
----- 8.0/11.9 MB 5.6 MB/s eta 0:00:01
----- 8.4/11.9 MB 5.7 MB/s eta 0:00:01
----- 9.0/11.9 MB 5.8 MB/s eta 0:00:01
----- 9.5/11.9 MB 5.9 MB/s eta 0:00:01
----- 9.8/11.9 MB 6.0 MB/s eta 0:00:01
----- 10.4/11.9 MB 6.4 MB/s eta 0:00:01
----- 10.8/11.9 MB 6.8 MB/s eta 0:00:01
----- 11.2/11.9 MB 7.0 MB/s eta 0:00:01
----- 11.6/11.9 MB 6.9 MB/s eta 0:00:01
----- 11.8/11.9 MB 7.0 MB/s eta 0:00:01
----- 11.8/11.9 MB 7.0 MB/s eta 0:00:01
----- 11.9/11.9 MB 6.4 MB/s eta 0:00:00
```

Preparing metadata (setup.py): started

Preparing metadata (setup.py): finished with status 'done'

Building wheels for collected packages: mysql.connector

Building wheel for mysql.connector (setup.py): started

Building wheel for mysql.connector (setup.py): finished with status 'done'

Created wheel for mysql.connector: filename=mysql_connector-2.2.9-cp311-cp311-win_amd64.whl size=247958 sha256=7cb80c9a2740fd25dc6d73e14c769a1f045366bc132788cf3206feca09ff8a69

Stored in directory: c:\users\asus\appdata\local\pip\cache\wheels\17\cd\ed\2d49e9bac69cf09382e4c7cc20a2511202b48324b87db26019

Successfully built mysql.connector

Installing collected packages: mysql.connector

Successfully installed mysql.connector-2.2.9

```
In [68]: import mysql.connector
```

```
In [72]: conn = mysql.connector.connect(host='localhost',user='root',password='',database='v
```

ProgrammingError

Traceback (most recent call last)

Cell In[72], line 1

```
----> 1 conn = mysql.connector.connect(host='localhost',user='root',password='',da  
tabase='world')
```

File ~\anaconda3\Lib\site-packages\mysql\connector__init__.py:179, in connect(*args, **kwargs)

```
177     return CMySQLConnection(*args, **kwargs)
178 else:
--> 179     return MySQLConnection(*args, **kwargs)
```

File ~\anaconda3\Lib\site-packages\mysql\connector\connection.py:95, in MySQLConnection.__init__(self, *args, **kwargs)

```
92 self._pool_config_version = None
94 if len(kwargs) > 0:
--> 95     self.connect(**kwargs)
```

File ~\anaconda3\Lib\site-packages\mysql\connector\abstracts.py:716, in MySQLConnectionAbstract.connect(self, **kwargs)

```
713     self.config(**kwargs)
715 self.disconnect()
--> 716 self._open_connection()
717 self._post_connection()
```

File ~\anaconda3\Lib\site-packages\mysql\connector\connection.py:208, in MySQLConnection._open_connection(self)

```
206 self._socket.open_connection()
207 self._do_handshake()
--> 208 self._do_auth(self._user, self._password,
209               self._database, self._client_flags, self._charset_id,
210               self._ssl)
211 self.set_converter_class(self._converter_class)
212 if self._client_flags & ClientFlag.COMPRESS:
```

File ~\anaconda3\Lib\site-packages\mysql\connector\connection.py:144, in MySQLConnection._do_auth(self, username, password, database, client_flags, charset, ssl_options)

```
137 packet = self._protocol.make_auth(
138     handshake=self._handshake,
139     username=username, password=password, database=database,
140     charset=charset, client_flags=client_flags,
141     ssl_enabled=self._ssl_active,
142     auth_plugin=self._auth_plugin)
143 self._socket.send(packet)
--> 144 self._auth_switch_request(username, password)
146 if not (client_flags & ClientFlag.CONNECT_WITH_DB) and database:
147     self.cmd_init_db(database)
```

File ~\anaconda3\Lib\site-packages\mysql\connector\connection.py:173, in MySQLConnection._auth_switch_request(self, username, password)

```
171 packet = self._socket.recv()
172 if packet[4] != 1:
--> 173     return self._handle_ok(packet)
174 else:
175     auth_data = self._protocol.parse_auth_more_data(packet)
```

File ~\anaconda3\Lib\site-packages\mysql\connector\connection.py:331, in MySQLConnection._handle_ok(self, packet)

```
329     return ok_pkt
330 elif packet[4] == 255:
--> 331     raise errors.get_exception(packet)
332 raise errors.InterfaceError('Expected OK packet')
```

```
ProgrammingError: 1045 (28000): Access denied for user 'root'@'localhost' (using password: NO)
```

```
In [73]: df = pd.read_sql_query("SELECT * FROM countrylanguage",conn)
```

```
-----
NameError                                Traceback (most recent call last)
Cell In[73], line 1
----> 1 df = pd.read_sql_query("SELECT * FROM countrylanguage",conn)

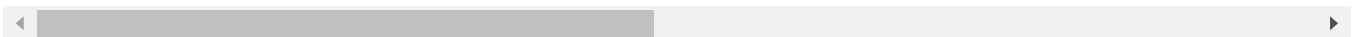
NameError: name 'conn' is not defined
```

```
In [74]: df
```

```
Out[74]:
```

	enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_univers
0	8949	city_103	0.920	Male	Has relevent experience	no_enrollme
1	29725	city_40	0.776	Male	No relevent experience	no_enrollme
2	11561	city_21	0.624	NaN	No relevent experience	Full time cou
3	33241	city_115	0.789	NaN	No relevent experience	N
4	666	city_162	0.767	Male	Has relevent experience	no_enrollme
...
19153	7386	city_173	0.878	Male	No relevent experience	no_enrollme
19154	31398	city_103	0.920	Male	Has relevent experience	no_enrollme
19155	24576	city_103	0.920	Male	Has relevent experience	no_enrollme
19156	5756	city_65	0.802	Male	Has relevent experience	no_enrollme
19157	23834	city_67	0.855	NaN	No relevent experience	no_enrollme

19158 rows × 14 columns



```
In [ ]:
```