

Working with CSV file.

read CSV from
csv file

read CSV from
url

(i) Separator → CSV file can have different separator like {, ; tab} so we need to specify it.

```
pd.read_csv('name.csv', sep = '\t')
```

(ii) When our CSV file do not have names of column we can specify it by using names parameter.

```
pd.read_csv('name.csv', names = ["first_name",  
                                   "last_name", "city"])
```

(iii) index_cols ⇒ when we do not want default index but the index will be counted on basis of column.

```
pd.read_csv('name.csv', index_cols = ['serno'])  
                                   col. name
```

(iv) Header = 1 ⇒ when we want our dataset is readed without header and header name is taken as row then we use this

```
pd.read_csv('name.csv', header = 1)
```

(v) When we are having n number of columns of which few are of use then we can specify them

```
pd.read_csv('names.csv', usecols = ['name', 'gender'])
```



```
pd.read_csv('names.csv', usecols=['gender'],  
            squeeze=True)
```

It is done when we want only one column from a large number of columns.

(vi) Skip rows / n rows

```
pd.read_csv('file', skiprows=[1, 2])
```

```
pd.read_csv('file', nrows=100)
```

(vii) Encoding parameter

```
pd.read_csv(____, encoding='')
```

Encoding can be → utf-8, latin-1

UNICODEDECODE
Error
↓
Some need to
try diff encodings.

(viii) Parser-Error :- It occurs when pandas is not able to read data because of irregularities in data for 5 columns we have 6 values in such case this error occurs.

```
pd.read_csv('file', error_bad_lines=False)
```

(ix) dtype → change data type of column.

```
pd.read_csv('file', dtype={'target': int})
```

(x) Handling dates.

```
pd.read_csv('file', parse_dates=['dates'])
```


(xi) Converters: When we want to apply function on column we use converters.

```
def renamer(x):
```

```
pd.read_csv('file', converters = {'team1': renamer})
```

(xii) Handling null values: Many often null values appear as /, -, other instead of NaN so we can specify them as NaN

```
pd.read_csv('file', na_values = ['-', '/'])
```

(xiii) Loading data into chunks

```
df = pd.read_csv('file', chunksize = 5000)
```