

CSV File

CSV (comma-separated value) files are a common file format for transferring and storing data.

The ability to read, manipulate, and write data to and from CSV files using Python is a key skill to master for any data scientist or business analysis.

Pandas is the most popular data manipulation package in Python.

Data Frame are the Pandas data type for storing tabular 2D data.

Creating a csv file

To create a csv file we use 'x' mode in file handling

```
open('books.csv','x')
```

Pandas

- Pip install pandas
- `python -m pip install --upgrade pip`

Read different rows

```
import pandas as pd  
df = pd.read_csv('url_csv Or file_name', nrows=5)  
Df
```

```
import pandas as pd  
df = pd.read_csv('url_csv Or file_name', index_col=0)  
df
```

Importing and reading

```
import pandas as pd
url_csv = 'URL name'
df = pd.read_csv(url_csv, index_col=0)
df.head()
df.tail()
```

```
import pandas as pd
df = pd.read_csv('boooks.csv', index_col=0)
df.head()
df.tail()
```

Read different cols

```
import pandas as pd  
cols = ['id','name','address']  
df = pd.read_csv('boooks.csv', index_col=0, usecols=cols)  
df.head()
```

Read by index

```
import pandas as pd
csv_url = 'boooks.csv'
df = pd.read_csv(csv_url, index_col='id',
                  usecols=['id','name', 'address', 'phone', 'salary'])
df.head()
```

Selecting Rows

Selecting first 10 row:

```
data.iloc[0:4, 0:2]
```

Select the rows where “Name” value is ‘Ram’.

```
data[data["Name"] == "Ram"]
```


Write On CSV (One By One data)

```
import csv
```

```
fields=['Kamal','28','Baluwatar']
```

```
with open('boooks.csv', 'a') as f:
```

```
    x = csv.writer(f)
```

```
    x.writerow(fields)
```

DictReader

```
import csv  
with open('videos.csv') as csvfile:  
    reader = csv.DictReader(csvfile)  
    for row in reader:  
        print(row)
```

DictWriter

```
import csv
csv_columns = ['Name', 'Age', 'Address']
dict = [
    {'Name':'Aayush','Age':23,'Address':'Katmandu'},
    {'Name':'Shyam','Age':33,'Address':'Patan'}
]
try:
    with open("data.csv", 'w') as csvfile:
        writer = csv.DictWriter(csvfile, fieldnames=csv_columns)
        writer.writeheader()
        for data in dict:
            writer.writerow(data)
except IOError:
    print("I/O error")
```

Write csv files (Using Dataframe)

```
import pandas as pd
df = pd.DataFrame({'Names':['Ram', 'Shyam', 'Hari','Bikash', 'Samrat'],
'Address':['Kathmandu','Palpa','Pyuthan','Bhaktapur','Patan'],
'Age':[21, 22, 20, 19, 18]})
df.to_csv('boooks.csv')
df.head()
```

Rename multiple columns in one go

```
df.rename(columns={"Physics": "physics", "Year": "year"}, inplace=True)
```

Deleting columns and rows

Delete the “Name” column from the dataframe

```
data = data.drop(“Name”, axis=1)
```

alternatively, delete columns using the columns parameter of drop

```
data = data.drop(columns=“Name”)
```

Delete the Area column from the dataframe in place

Note that the original 'data' object is changed when inplace=True

```
data.drop(“Name”, axis=1, inplace=True).
```

Delete multiple columns from the dataframe

```
data = data.drop([“Name”, “Age”, “Address”], axis=1)
```

Delete the rows with labels 0,1,5

```
data = data.drop([0,1,2], axis=0)
```

Read / Write Multiple dataframe

```
import pandas as pd

df1 = pd.DataFrame({'Names': ['Andreas', 'George', 'Steve',
                              'Sarah', 'Joanna', 'Hanna'],
                   'Age': [21, 22, 20, 19, 18, 23]})

df2 = pd.DataFrame({'Names': ['Pete', 'Jordan', 'Gustaf',
                              'Sophie', 'Sally', 'Simone'],
                   'Age': [22, 21, 19, 19, 29, 21]})

df3 = pd.DataFrame({'Names': ['Ulrich', 'Donald', 'Jon',
                              'Jessica', 'Elisabeth', 'Diana'],
                   'Age': [21, 21, 20, 19, 19, 22]})

df = pd.concat([df1, df2, df3], keys=['Group1', 'Group2', 'Group3'],
              names=['Group', 'Row Num']).reset_index()

df.to_csv('boooks.csv', index=False)
```