Reading Files (.csv)

Import Python Library

We will use a library: Pandas

```
#Reading a CSV File
import pandas as pd
```

Reading a .csv file.

The csv file is a text file in which the values in the columns are separated by a comma. Let's consider the following data present in the file named **input.csv**.

The **read_csv** function of the pandas library is used read the content of a CSV file into the python environment as a pandas DataFrame. The function can read the files from the OS by using proper path to the file.

```
data = pd.read_csv("E:/input.csv")
data
```

	EmpID	FirstName	LastName	Gender	Birthdate	Age	Joiningdate	Salary	Region
0	677509	Lois	Walker	F	3/29/1981	36	11/24/2003	168251	West
1	940761	Brenda	Robinson	F	7/31/1970	47	7/27/2008	51063	South
2	428945	Joe	Robinson	M	6/16/1963	54	8/3/2016	50155	Midwest
3	408351	Diane	Evans	F	12/4/1977	67	4/16/1999	180294	Northeast
4	193819	Benjamin	Russell	M	4/17/1977	74	7/25/2013	117642	Midwest
5	499687	Patrick	Bailey	M	9/27/1982	85	7/22/2005	72305	Midwest
6	539712	Nancy	Baker	F	6/13/1995	22	9/14/2016	98189	South
7	380086	Carol	Murphy	F	6/30/1958	15	1/28/1983	60918	Midwest
8	477616	Frances	Young	F	6/9/1959	10	4/27/1994	121587	South

When we execute the above code, it produces the following result. Please note how an additional column starting with zero as an index has been created by the function.

Reading Specific Rows

```python
#Reading Specific Rows
print(data[0:6]['Birthdate'])
```

```
0    3/29/1981
1    7/31/1970
2    6/16/1963
3    12/4/1977
4    4/17/1977
5    9/27/1982
Name: Birthdate, dtype: object
```

Reading Specific Columns

```python
#Reading Specific Coloumns
print(data.loc[:,['FirstName','Salary']])
```

```
    FirstName  Salary
0        Lois  168251
1      Brenda   51063
2         Joe   50155
3       Diane  180294
4    Benjamin  117642
5     Patrick   72305
6       Nancy   98189
7       Carol   60918
8     Frances  121587
```

Reading Specific Columns and Rows

```python
#Reading Specific Coloumns and Rows
print(data.loc[[1,3,5],['FirstName','Salary']])
```

```
   FirstName  Salary
1     Brenda   51063
3      Diane  180294
5    Patrick   72305
```

Reading Specific Columns for a Range of Rows

```
#Reading Specific Columns for a Range of Rows
print(data.loc[2:6,['FirstName','Salary']])
```

```
   FirstName  Salary
2        Joe   50155
3      Diane  180294
4   Benjamin  117642
5    Patrick   72305
6      Nancy   98189
```

Printing first 5 rows

```
# Prints the first 5 rows
data.head()
```

	EmpID	FirstName	LastName	Gender	Birthdate	Age	Joiningdate	Salary	Region
0	677509	Lois	Walker	F	3/29/1981	36	11/24/2003	168251	West
1	940761	Brenda	Robinson	F	7/31/1970	47	7/27/2008	51063	South
2	428945	Joe	Robinson	M	6/16/1963	54	8/3/2016	50155	Midwest
3	408351	Diane	Evans	F	12/4/1977	67	4/16/1999	180294	Northeast
4	193819	Benjamin	Russell	M	4/17/1977	74	7/25/2013	117642	Midwest

Printing last 5 rows

```
#Prints the last 5 rows
data.tail()
```

	EmpID	FirstName	LastName	Gender	Birthdate	Age	Joiningdate	Salary	Region
4	193819	Benjamin	Russell	M	4/17/1977	74	7/25/2013	117642	Midwest
5	499687	Patrick	Bailey	M	9/27/1982	85	7/22/2005	72305	Midwest
6	539712	Nancy	Baker	F	6/13/1995	22	9/14/2016	98189	South
7	380086	Carol	Murphy	F	6/30/1958	15	1/28/1983	60918	Midwest
8	477616	Frances	Young	F	6/9/1959	10	4/27/1994	121587	South