Data Science using Machine learning Python

September 2, 2019

[1]:
```
#read from tsv file
import pandas as pd
```

[2]:
```
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
employee.head()
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:1:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
  """Entry point for launching an IPython kernel.
```

[2]:

	Name	Position	Office	Age	Start date	Salary
0	Airi Satou	Accountant	Tokyo	33	2008/11/28	$162,700
1	Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09	$1,200,000
2	Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12	$86,000
3	Bradley Greer	Software Engineer	London	41	2012/10/13	$132,000
4	Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07	$206,850

[4]:
```
#choose particular columns
cols=['Name','Position']
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
employee[cols]
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:3:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
  This is separate from the ipykernel package so we can avoid doing imports
until
```

[4]:

	Name	Position
0	Airi Satou	Accountant
1	Angelica Ramos	Chief Executive Officer (CEO)
2	Ashton Cox	Junior Technical Author
3	Bradley Greer	Software Engineer
4	Brenden Wagner	Software Engineer
5	Brielle Williamson	Integration Specialist
6	Bruno Nash	Software Engineer
7	Caesar Vance	Pre-Sales Support
8	Cara Stevens	Sales Assistant
9	Cedric Kelly	Senior Javascript Developer
10	Charde Marshall	Regional Director
11	Colleen Hurst	Javascript Developer
12	Dai Rios	Personnel Lead
13	Donna Snider	Customer Support
14	Doris Wilder	Sales Assistant
15	Finn Camacho	Support Engineer
16	Fiona Green	Chief Operating Officer (COO)
17	Garrett Winters	Accountant
18	Gavin Cortez	Team Leader
19	Gavin Joyce	Developer
20	Gloria Little	Systems Administrator
21	Haley Kennedy	Senior Marketing Designer
22	Hermione Butler	Regional Director
23	Herrod Chandler	Sales Assistant
24	Hope Fuentes	Secretary
25	Howard Hatfield	Office Manager
26	Jackson Bradshaw	Director
27	Jena Gaines	Office Manager
28	Jenette Caldwell	Development Lead
29	Jennifer Acosta	Junior Javascript Developer
30	Jennifer Chang	Regional Director
31	Jonas Alexander	Developer
32	Lael Greer	Systems Administrator
33	Martena Mccray	Post-Sales support
34	Michael Bruce	Javascript Developer
35	Michael Silva	Marketing Designer
36	Michelle House	Integration Specialist
37	Olivia Liang	Support Engineer
38	Paul Byrd	Chief Financial Officer (CFO)
39	Prescott Bartlett	Technical Author
40	Quinn Flynn	Support Lead
41	Rhona Davidson	Integration Specialist
42	Sakura Yamamoto	Support Engineer
43	Serge Baldwin	Data Coordinator
44	Shad Decker	Regional Director
45	Shou Itou	Regional Marketing

```
46          Sonya Frost              Software Engineer
47           Suki Burks                      Developer
48  Tatyana Fitzpatrick              Regional Director
49          Thor Walton                      Developer
50          Tiger Nixon               System Architect
51       Timothy Mooney                 Office Manager
52         Unity Butler             Marketing Designer
53       Vivian Harrell           Financial Controller
54           Yuri Berry  Chief Marketing Officer (CMO)
55        Zenaida Frank              Software Engineer
56       Zorita Serrano              Software Engineer
```

[5]:
```
employee[cols].head()
```

[5]:
```
             Name                        Position
0      Airi Satou                      Accountant
1  Angelica Ramos  Chief Executive Officer (CEO)
2      Ashton Cox         Junior Technical Author
3   Bradley Greer                Software Engineer
4  Brenden Wagner                Software Engineer
```

[6]:
```
#number of certain rows
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv",nrows=4)
employee
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:2:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
```

[6]:
```
             Name                        Position         Office  Age  \
0      Airi Satou                      Accountant          Tokyo   33
1  Angelica Ramos  Chief Executive Officer (CEO)         London   47
2      Ashton Cox         Junior Technical Author  San Francisco   66
3   Bradley Greer                Software Engineer         London   41

   Start date      Salary
0  2008/11/28    $162,700
1  2009/10/09  $1,200,000
2  2009/01/12     $86,000
3  2012/10/13    $132,000
```

[7]:
```
#displaying datatype
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
employee.dtypes
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:2:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
```

[7]:
```
Name           object
Position       object
Office         object
Age             int64
Start date     object
Salary         object
dtype: object
```

[19]:
```
#displaying integer data types
import numpy as np
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
employee.select_dtypes(include=[np.number]).dtypes
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:3:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
  This is separate from the ipykernel package so we can avoid doing imports
until
```

[19]:
```
Age    int64
dtype: object
```

[20]:
```
#Some inbuilt methods for data analysis
employee.describe()
```

[20]:
```
              Age
count   57.000000
mean    42.736842
std     14.877507
min     19.000000
25%     30.000000
50%     42.000000
75%     56.000000
max     66.000000
```

[22]:
```
#Show the number of columns and rows
employee.shape
```

[22]:
```
(57, 6)
```

[23]:
```
#type of Object
type(employee)
```

[23]:
```
pandas.core.frame.DataFrame
```

Data Massaging and Filtering

[26]:
```
#concatenating 2 columns
employee["Name Salary"]=employee["Name"]+employee["Salary"]
employee.head()
#Nama Salary is a new column
```

[26]:
```
             Name                         Position         Office  Age  \
0      Airi Satou                       Accountant          Tokyo   33
1  Angelica Ramos  Chief Executive Officer (CEO)         London   47
2      Ashton Cox         Junior Technical Author  San Francisco   66
3   Bradley Greer                Software Engineer         London   41
4  Brenden Wagner                Software Engineer  San Francisco   28

   Start date      Salary                Name Salary
0  2008/11/28    $162,700         Airi Satou$162,700
1  2009/10/09  $1,200,000   Angelica Ramos$1,200,000
2  2009/01/12     $86,000          Ashton Cox$86,000
3  2012/10/13    $132,000      Bradley Greer$132,000
4  2011/06/07    $206,850     Brenden Wagner$206,850
```

[29]:
```
#deleting column
employee.drop('Office',axis=1,inplace=True)
employee.head()
#here Office column is deleted
```

[29]:
```
             Name                         Position  Age  Start date      Salary  \
0      Airi Satou                       Accountant   33  2008/11/28    $162,700
1  Angelica Ramos  Chief Executive Officer (CEO)   47  2009/10/09  $1,200,000
2      Ashton Cox         Junior Technical Author   66  2009/01/12     $86,000
3   Bradley Greer                Software Engineer   41  2012/10/13    $132,000
4  Brenden Wagner                Software Engineer   28  2011/06/07    $206,850

                Name Salary
0        Airi Satou$162,700
1  Angelica Ramos$1,200,000
2         Ashton Cox$86,000
3      Bradley Greer$132,000
4     Brenden Wagner$206,850
```

[34]:
```
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
cols=['Name','Pos','Office','Age','Start Date','Sal']
employee.columns=cols
employee.head()
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:1:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
  """Entry point for launching an IPython kernel.
```

[34]:
```
             Name                          Pos         Office  Age  \
0      Airi Satou                       Accountant          Tokyo   33
1  Angelica Ramos  Chief Executive Officer (CEO)         London   47
2      Ashton Cox         Junior Technical Author  San Francisco   66
3   Bradley Greer                Software Engineer         London   41
4  Brenden Wagner                Software Engineer  San Francisco   28
```

```
   Start Date         Sal
0  2008/11/28    $162,700
1  2009/10/09  $1,200,000
2  2009/01/12     $86,000
3  2012/10/13    $132,000
4  2011/06/07    $206,850
```

[40]:
```
#sorting the data by Order clause
employee=pd.read_table("/home/sakil/Desktop/DataScience/Udemy/Module2/
 ↪tab_seperated_values.tsv")
employee.sort_values(by='Age',ascending=True)
employee.head()
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/ipykernel_launcher.py:2:
FutureWarning: read_table is deprecated, use read_csv instead, passing sep='\t'.
```

[40]:
```
              Name                        Position         Office  Age  \
0      Airi Satou                      Accountant          Tokyo   33
1  Angelica Ramos  Chief Executive Officer (CEO)         London   47
2      Ashton Cox        Junior Technical Author  San Francisco   66
3   Bradley Greer              Software Engineer         London   41
4  Brenden Wagner              Software Engineer  San Francisco   28

   Start date      Salary
0  2008/11/28    $162,700
1  2009/10/09  $1,200,000
2  2009/01/12     $86,000
3  2012/10/13    $132,000
4  2011/06/07    $206,850
```

[43]:
```
#sorting a series
employee['Name'].sort_values().head()
```

[43]:
```
0        Airi Satou
1    Angelica Ramos
2        Ashton Cox
3     Bradley Greer
4    Brenden Wagner
Name: Name, dtype: object
```

[45]:
```
#select all the rows having age <40
employee[employee.Age<40].head()
```

[45]:
```
              Name                    Position         Office  Age  \
0      Airi Satou                   Accountant          Tokyo   33
4  Brenden Wagner           Software Engineer  San Francisco   28
6      Bruno Nash           Software Engineer         London   38
7    Caesar Vance           Pre-Sales Support       New York   21
```

```
9    Cedric Kelly  Senior Javascript Developer      Edinburgh   22

   Start date    Salary
0  2008/11/28  $162,700
4  2011/06/07  $206,850
6  2011/05/03  $163,500
7  2011/12/12  $106,450
9  2012/03/29  $433,060
```

[47]:
```
#using multiple condition
employee[(employee.Age<40)&(employee.Name=="Airi Satou")]
```

[47]:
```
          Name    Position Office  Age  Start date    Salary
0  Airi Satou  Accountant  Tokyo   33  2008/11/28  $162,700
```

[48]:
```
#selecting particular columns
cols=["Name","Position"]
employee[employee.Age<40][cols].head()
```

[48]:
```
             Name                     Position
0      Airi Satou                   Accountant
4  Brenden Wagner            Software Engineer
6      Bruno Nash            Software Engineer
7    Caesar Vance            Pre-Sales Support
9    Cedric Kelly  Senior Javascript Developer
```