DATAFRAMES WITH PANDAS PART 1:

Read the data:

In [1]:

```
import pandas as pd
```

In [2]:

```
df = pd.read_csv("Uber Drives 2016.csv")
```

In [3]:

```
df
```

Out[3]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	1/2/2016 20:25	1/2/2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	1/5/2016 17:31	1/5/2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit
...	...	...	...	...	...	...	...
1151	12/31/2016 13:24	12/31/2016 13:42	Business	Kar?chi	Unknown Location	3.9	Temporary Site
1152	12/31/2016 15:03	12/31/2016 15:38	Business	Unknown Location	Unknown Location	16.2	Meeting
1153	12/31/2016 21:32	12/31/2016 21:50	Business	Katunayake	Gampaha	6.4	Temporary Site
1154	12/31/2016 22:08	12/31/2016 23:51	Business	Gampaha	Ilukwatta	48.2	Temporary Site
1155	Totals	NaN	NaN	NaN	NaN	12204.7	NaN

1156 rows × 7 columns

df : DataFrame

```
    A dataframe is a basic data structure inside pandas which represents data in the form
of rows and columns.
```

In [4]:

```
type(df)
```

Out[4]:

```
pandas.core.frame.DataFrame
```

To know each column datatype:

In [5]:

```
df.dtypes
```

Out[5]:

```
START_DATE*      object
END_DATE*        object
CATEGORY*        object
START*           object
STOP*            object
MILES*          float64
PURPOSE*         object
dtype: object
```

head():

```
    It is used to print only first five rows.
```

In [6]:

```
df.head()
```

Out[6]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	1/2/2016 20:25	1/2/2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	1/5/2016 17:31	1/5/2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

In [7]:

```
df.head(n = 10)
```

Out[7]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	1/2/2016 20:25	1/2/2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	1/5/2016 17:31	1/5/2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit
5	1/6/2016 17:15	1/6/2016 17:19	Business	West Palm Beach	West Palm Beach	4.3	Meal/Entertain
6	1/6/2016 17:30	1/6/2016 17:35	Business	West Palm Beach	Palm Beach	7.1	Meeting
7	1/7/2016 13:27	1/7/2016 13:33	Business	Cary	Cary	0.8	Meeting
8	1/10/2016 8:05	1/10/2016 8:25	Business	Cary	Morrisville	8.3	Meeting
9	1/10/2016 12:17	1/10/2016 12:44	Business	Jamaica	New York	16.5	Customer Visit

Shape:

```
    It is used to tell how many rows and columns are present.
```

In [9]:

```
df.shape
```

Out[9]:

```
(1156, 7)
```

tail():

```
It is used to print only last five rows.
```

In [10]:

```
df.tail()
```

Out[10]:

	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
1151	12/31/2016 13:24	12/31/2016 13:42	Business	Kar?chi	Unknown Location	3.9	Temporary Site
1152	12/31/2016 15:03	12/31/2016 15:38	Business	Unknown Location	Unknown Location	16.2	Meeting
1153	12/31/2016 21:32	12/31/2016 21:50	Business	Katunayake	Gampaha	6.4	Temporary Site
1154	12/31/2016 22:08	12/31/2016 23:51	Business	Gampaha	Ilukwatta	48.2	Temporary Site
1155	Totals	NaN	NaN	NaN	NaN	12204.7	NaN

Last row is the sum of all the values of each columns.

NaN :

```
* Is is official junk values.
* It stands for Not Any Number.
```

Changing data types:

pd.DataFrame() :

```
It is used to create a dataframe form our own definition.
```

In [14]:

```
temp = pd.DataFrame({"A" : ["1","2","3"], "B" : [11,12,13], "C" : ["12-06-2012","13-06-2015","15-06
-2017"]})
```

In [15]:

```
temp
```

Out[15]:

	A	B	C
0	1	11	12-06-2012
1	2	12	13-06-2015
2	3	13	15-06-2017

- keys : column name
- values : datas

In [16]:

```
temp.dtypes
```

Out[16]:

```
A    object
B     int64
C    object
dtype: object
```

to_datetime :

```
It converts a string into date and time.
```

In [18]:

```
date = pd.to_datetime("12-Jun-2012")
date
```

Out[18]:

```
Timestamp('2012-06-12 00:00:00')
```

In [20]:

```
type(date)
```

Out[20]:

```
pandas._libs.tslibs.timestamps.Timestamp
```

In [ ]:

```

```