

DAC Phase 3:

Problem Statement: Website Traffic Analysis

Loading and Pre-processing of data:

```
from google.colab import drive
drive.mount('/content/drive')
```

Loading the data

```
import pandas as pd
import numpy as np
data = pd.read_csv('/content/drive/MyDrive/daily-website-visitors.csv')
data.head(5)
```

	Row	Day	Day.Of.Week	Date	Page.Loads	Unique.Visits	First.Time.Visits	Returning.Visits
0	1	Sunday	1	9/14/2014	2,146	1,582	1,430	152
1	2	Monday	2	9/15/2014	3,621	2,528	2,297	231
2	3	Tuesday	3	9/16/2014	3,698	2,630	2,352	278
3	4	Wednesday	4	9/17/2014	3,667	2,614	2,327	287
4	5	Thursday	5	9/18/2014	3,316	2,366	2,130	236

```
data.describe()
```

	Row	Day.Of.Week
count	2167.000000	2167.000000
mean	1084.000000	3.997231
std	625.703338	2.000229
min	1.000000	1.000000
25%	542.500000	2.000000
50%	1084.000000	4.000000
75%	1625.500000	6.000000
max	2167.000000	7.000000

Insights:

This command is used to view the brief summary of the dataset. We can see the mathematical parameters such as percentiles, standard deviation, mean, minimum and maximum values and count of each column.

```
data.shape  
(2167, 8)
```

Insights:

It shows the shape whereas rows and columns of the dataset.

```
data.info()  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2167 entries, 0 to 2166  
Data columns (total 8 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   Row                   2167 non-null  int64  
1   Day                   2167 non-null  object  
2   Day.Of.Week           2167 non-null  int64  
3   Date                  2167 non-null  object  
4   Page.Loads            2167 non-null  object  
5   Unique.Visits         2167 non-null  object  
6   First.Time.Visits     2167 non-null  object  
7   Returning.Visits      2167 non-null  object  
dtypes: int64(2), object(6)  
memory usage: 135.6+ KB
```

Insights:

Info command is used to check the datatype of every column and the count of each column. The difference between the describe() and info() is that describe command will give the mathematical parameters but info command will not give the mathematical parameters such as mean and standard deviation.

```
data.isna().sum()  
Row                0  
Day                0  
Day.Of.Week        0  
Date               0  
Page.Loads         0  
Unique.Visits      0  
First.Time.Visits  0  
Returning.Visits    0  
dtype: int64
```

The above command is used to check for null values in each column.

Converting the date column to date format from object

```
data['Date'] = pd.to_datetime(data['Date'])  
data['Date'].dtype
```

```
dtype('<M8[ns]')
```