# Date-25/10/2023
# Team ID-718
# Project Title-Website Traffic Analysis

### Import Dependencies

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
dataset = pd.read_csv('/content/drive/MyDrive/daily-website-visitors.csv')
```

### Dataset Exploration

```
dataset
```

| | 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 | 3,69 | 2,63 | 2,352 | 27 |
| 3 | 4 | Wednesday | 4 | 9-17-2014 | 8 | 0 | 2,327 | 8 |
| 4 | 5 | Thursday | 5 | 9-18-2014 | 3,66 | 2,614 | 2,130 | 28 |
| ... | ... | ... | ... | ... | 7 | 2,36 | ... | 7 |
| 2162 | 2163 | Saturday | 7 | 8-15-2020 | 3,316 | 6 ... | 1,373 | 23 |
| 2163 | 2164 | Sunday | 1 | ... | ... | 1,696 | 1,686 | 6 ... |
| 2164 | 2165 | Monday | 2 | 8-16-2020 | 2,221 | 2,03 | 2,181 | 323 |
| 2165 | 2166 | Tuesday | 3 | | 2,72 | 7 | 2,184 | 351 |
| 2166 | 2167 | Wednesday | 4 | 8-17-2020 | 4 | 2,63 | 1,297 | 45 |
| 2167 rows × 8 columns | | | | | 3,45 | 8 | | 7 |
| | | | | 8-18-2020 | 6 | 2,68 | | 49 |
| | | | | 8-19-2020 | 3,581 | 3 | | 9 |
| | | | | | 2,06 | 1,564 | | 26 |
| | | | | | 4 | | | 7 |

```
dataset.head()
```

| | 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 | 3,69 | 2,63 | 2,352 | 27 |
| 3 | 4 | Wednesday | 4 | 9-17-2014 | 8 | 0 | 2,327 | 8 |
| 4 | 5 | Thursday | 5 | 9-18-2014 | 3,66 | 2,614 | 2,130 | 28 |
| | | | | | 7 | 2,36 | | 7 |
| | | | | | 3,316 | 6 | | 23 |
| | | | | | | | | 6 |

```
dataset.shape
```

```
(2167, 8)
```

```
dataset.columns
```

```
Index(['Row', 'Day', 'Day.Of.Week', 'Date', 'Page.Loads', 'Unique.Visits',
 'First.Time.Visits', 'Returning.Visits'],
 dtype='object')
```

```
dataset.dtypes
```

```
Row int64
Day object
Day.Of.Week int64
Date object
Page.Loads object
```

```
Unique.Visits        object
First.Time.Visits    object
Returning.Visits object
dtype: object
```

### Data Preprocessing

```
dataset.isnull()
```

| | Row | Day | Day.Of.Week | Date | Page.Loads | Unique.Visits | First.Time.Visits | Returning.Visits |
|---|---|---|---|---|---|---|---|---|
| 0 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |
| 1 | False | False | e | e | e | e | e | e |
| 2 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |
| 3 | False | False | e | e | e | e | e | e |
| 4 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |
| ... | ... | ... | e | e | e | e | e | e |
| 2162 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |
| 2163 | False | False | e | e | e | e | e | e |
| 2164 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |
| 2165 | False | False | e ... | e ... | e ... | e ... | e ... | e ... |
| 2166 | False | False | Fals | Fals | Fals | Fals | Fals | Fals |

2167 rows × 8 columns

| | | | | | | |
|---|---|---|---|---|---|---|
| | Fals | Fals | Fals | Fals | Fals | Fals |
| | e | e | e | e | e | e |
| | Fals | Fals | Fals | Fals | Fals | Fals |
| | e | e | e | e | e | e |
| | Fals | Fals | Fals | Fals | Fals | Fals |
| | e | e | e | e | e | e |
| | Fals | Fals | Fals | Fals | Fals | Fals |
| | e | e | e | e | e | e |
| | Fals | Fals | Fals | Fals | Fals | Fals |
| | e | e | e | e | e | e |

```
dataset.isnull().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
```

```
dataset.isnull().sum().sum()
```

```
0
```

```
dataset.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 |

```
dataset.describe(include='all')
```

| | Row | Day | Day.Of.Week | Date | Page.Loads | Unique.Visits | First.Time.Visits | Returning.Visits |
|---|---|---|---|---|---|---|---|---|
| count | 2167.000000 | 2167 | 2167.000000 | 2167 | 2167 | 2167 | 2167 | 2167 |
| unique | NaN | 7 | NaN | 2167 | 1756 | 1658 | 1587 | 663 |
| top | NaN | Sunday | NaN | 9-14-2014 | 2,948 | 2,780 | 3,146 | 552 |
| freq | NaN | 310 | 1084.000000 | NaN | 1 | 5 | 5 | 5 | 12 |
| mean | NaN | 625.703338 | NaN | 3.997231 | NaN | NaN | NaN | NaN | NaN |
| std | 1.000000 | NaN | 2.000229 | NaN | NaN | NaN | NaN | NaN |
| min | 542.500000 | NaN | 1.000000 | NaN | NaN | NaN | NaN | NaN |
| 25% | 1084.000000 | NaN | 2.000000 | NaN | NaN | NaN | NaN | NaN |
| 50% | 1625.500000 | NaN | 4.000000 | NaN | NaN | NaN | NaN | NaN |
| 75% | 2167.000000 | NaN | 6.000000 | NaN | NaN | NaN | NaN | NaN |
| max | 2167.000000 | NaN | 7.000000 | NaN | NaN | NaN | NaN | NaN |

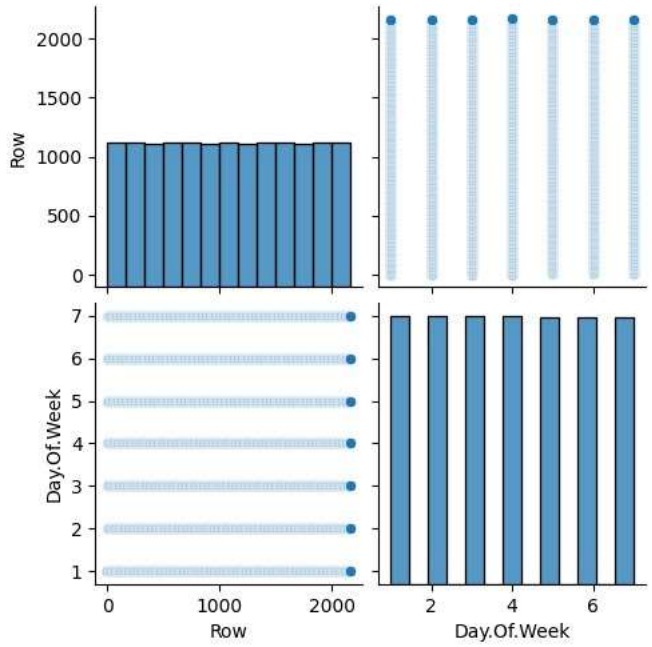
```
dataset.info()
```

```
<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
First.Time.Visits  2167  non-null  object
Returning.Visits 2167 non-null object dtypes:
int64(2), object(6)
memory usage: 135.6+ KB
```

### Data Visualization

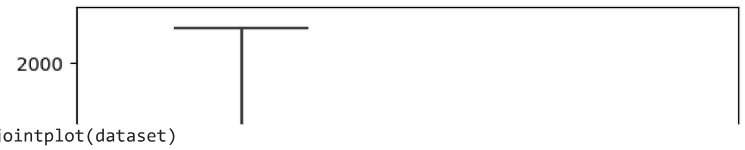
```
plt.figure(figsize=(20,20))
sns.pairplot(dataset)
```

```
<seaborn.axisgrid.PairGrid at 0x7d9b3aee9060>
<Figure size 2000x2000 with 0 Axes>
```



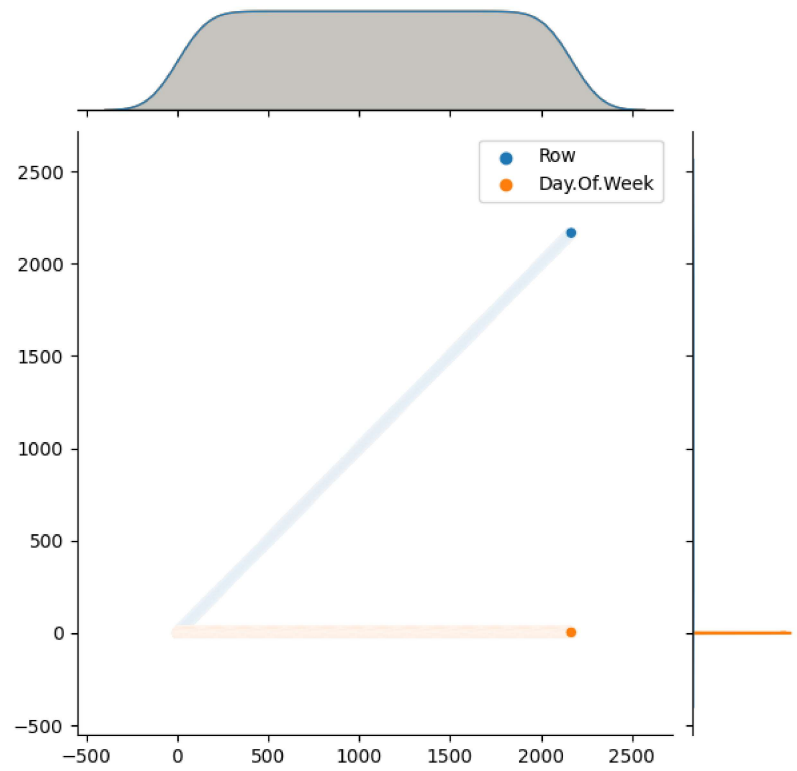

```
sns.boxplot(dataset)
```

```
<Axes: >
```




```
sns.jointplot(dataset)
```

```
<seaborn.axisgrid.JointGrid at 0x7d9b3ad0c190>
```




**Correlation Visualization**

```
dataset.corr()
```

```
<ipython-input-18-c187c74d1e71>:1: FutureWarning: The default value of numeric_only i
 dataset.corr()
```

| | Row | Day.Of.Week |
|---|---|---|
| Row | 1.0000 | 0.0008 |
| Day.Of.Week | 0.0008 | 1.0000 |

```
sns.heatmap(dataset.corr(),annot=True)
```

```
<ipython-input-19-9d3fd451b567>:1: FutureWarning: The default value of numeric_only i
sns.heatmap(dataset.corr(),annot=True)
<Axes: >
```