WEB TRAFFIC ANALYSIS

Phase 3:Visualization and Preprocessing

INTRODUCTION:

The amount of data can be overwhelming at first. That's why it's important to identify a few key metrics, particularly as you're getting started. For example, you might start by focusing on the bounce rate for a few key pages on your site. If visitors are quickly bouncing from your homepage, then that indicates they're not finding the information they're looking for quickly or easily enough. From there, you can identify possible next steps, like redesigning your website navigation.

Increase in Daily Visitors:

The most basic objective is to increase the number of daily visitors to your website. This can be a key indicator of your site's overall reach and visibility.

Unique Visitors:

Monitor the number of unique visitors who come to your site each day. This helps you understand how many different individuals are engaging with your content.

Returning Visitors:

Encourage repeat visits by setting objectives to increase the number of returning visitors. This indicates that your content is engaging and valuable.

Time on Site:

Set objectives to increase the average time visitors spend on your site. This is a good indicator of how engaging and relevant your content is.

Pageviews per Visitor:

Aim to increase the number of pages a visitor views during their session. This can indicate that visitors are exploring your content.

Bounce Rate Reduction:

Set objectives to reduce your website's bounce rate. A lower bounce rate indicates that visitors are finding what they're looking for and engaging with your site.

Conversion Rate:

If your website's main goal is conversions (e.g., sales, sign-ups, or downloads), set objectives for daily conversion rates.

Traffic Sources:

Monitor the sources of your traffic (e.g., organic search, social media, direct visits) and set objectives to increase traffic from specific sources that align with your marketing strategy.

Geographic or Demographic Objectives:

Depending on your target audience, you may set objectives to increase visitors from specific geographic regions or demographics.

Engagement Metrics:

Set objectives for specific engagement metrics, such as the number of comments, shares, or likes on your content if your website encourages user interaction.

Mobile vs. Desktop:

Monitor the percentage of visitors accessing your site from mobile devices and desktops. Set objectives to align with your site's mobile-friendliness.

Page Load Time:

If you're focused on user experience, set objectives to improve page load times to keep visitors engaged and reduce bounce rates.

Search Engine Rankings:

Monitor daily changes in search engine rankings for your key pages and keywords. The objective may be to improve your rankings over time.

Newsletter or Email Sign-ups:

If email marketing is a priority, set objectives for daily newsletter or email sign-ups.

A/B Testing:

If you're conducting A/B tests to optimize your site, set objectives to achieve specific improvements based on the test results.

Increase Traffic:

This is a fundamental goal for most websites. You might aim to increase daily visitor numbers as a primary objective. This could be to a specific number or percentage increase from your current traffic.

Unique Visitors:

Monitor the number of unique visitors to get a sense of how many individuals are interacting with your site daily.

Pageviews:

Set goals for the total number of pages viewed on your site per day. High pageviews can indicate visitor engagement.

Bounce Rate:

Aim to reduce your website's bounce rate, which measures the percentage of visitors who leave after viewing only one page. Lowering the bounce rate typically means visitors are exploring your site more.

Time on Site:

Set a goal for the average time visitors spend on your website daily. This can indicate engagement and the quality of your content.

Conversion Goals:

If your website's primary purpose is to drive conversions (e.g., sales, sign-ups, downloads), track and set daily objectives for these specific actions.

Traffic Sources:

Analyze where your traffic is coming from, such as organic search, social media, referral, or direct traffic. Set goals to increase traffic from specific sources.

PROCESS:

Process and clean the collected data to ensure its accuracy and reliability .We have.

```
[5] data.fillna(data.mean(), inplace=True)

<ipython-input-5-a20208f658ce>:1: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is deprecated. Select only va
  data.fillna(data.mean(), inplace=True)
```

```
from google.colab import files
uploaded=files.upload()

Choose Files daily-website-visitors.csv
• daily-website-visitors.csv(text/csv) - 115811 bytes, last modified: 18/10/2023 - 100% done
Saving daily-website-visitors.csv to daily-website-visitors.csv
```

First, you'll need to download the dataset from the provided Kaggle link. Once you have it, make sure to load the necessary libraries in your Python environment. The usual suspects are pandas, numpy, and matplotlib or seaborn for visualization.

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

# Load the dataset
dataset_url = "https://www.kaggle.com/datasets/bobnau/daily-website-visitors"
# Assuming you've already downloaded the dataset as 'water_quality.csv'
data = pd.read_csv("daily-website-visitors.csv")

# Display the first few rows
print(data.head())
```

```
   Row        Day  Day.Of.Week       Date Page.Loads Unique.Visits  \
0    1     Sunday            1  9/14/2014      2,146        1,582
1    2     Monday            2  9/15/2014      3,621        2,528
2    3    Tuesday            3  9/16/2014      3,698        2,630
3    4  Wednesday            4  9/17/2014      3,667        2,614
4    5   Thursday            5  9/18/2014      3,316        2,366

  First.Time.Visits Returning.Visits
0             1,430              152
1             2,297              231
2             2,352              278
3             2,327              287
4             2,130              236
```

Make sure to replace 'water_quality.csv' with the actual filename if it's different.

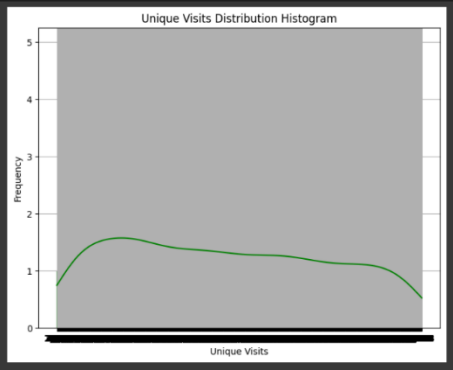
Next, check for missing values and handle them accordingly. You can use the isnull() function to identify missing values, and methods like fillna() or dropna() to handle them.

```
[4]  data=data.dropna()
     data.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
```

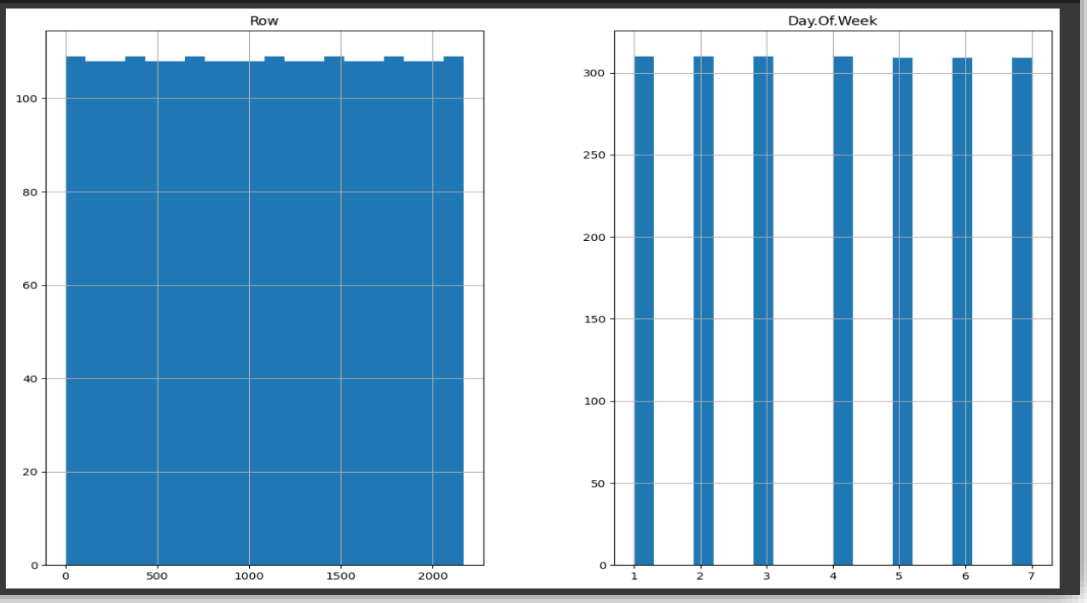
Handle missing values (for example, fill with mean or drop) Assuming you want to fill missing values with the mean

```
[7]  # Create a histogram of Unique Visits distribution
     plt.figure(figsize=(8, 6))
     sns.histplot(data['Unique.Visits'], bins=30, kde=True, color='g')
     plt.xlabel('Unique Visits')
     plt.ylabel('Frequency')
     plt.title('Unique Visits Distribution Histogram')
     plt.grid(True)
     plt.show()
```

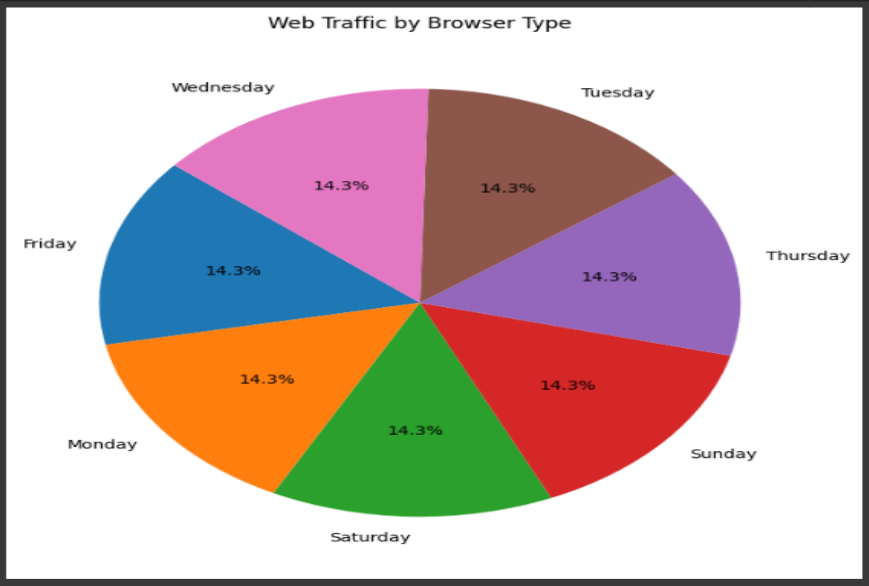


Now, let's explore the data visually through some EDA. Create histograms to visualize the distribution of each parameter.

```
[8]  # Visualize parameter distributions
     data.hist(bins=20, figsize=(15, 10))
     plt.show()
```



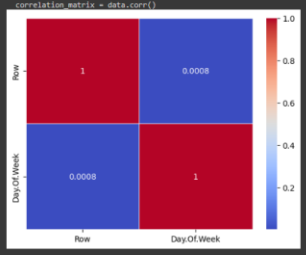

```
# Pie chart Vizualization
grouped_data = data.groupby('Day').size()
plt.figure(figsize=(8, 8))
plt.pie(grouped_data, labels=grouped_data.index, autopct='%1.1f%%', startangle=140)
plt.title('Web Traffic by Browser Type')

# Display Pie
plt.show()
```



For correlations, you can use a heatmap:

```
[10] # Visualize correlations
     correlation_matrix = data.corr()
     sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
     plt.show()
```

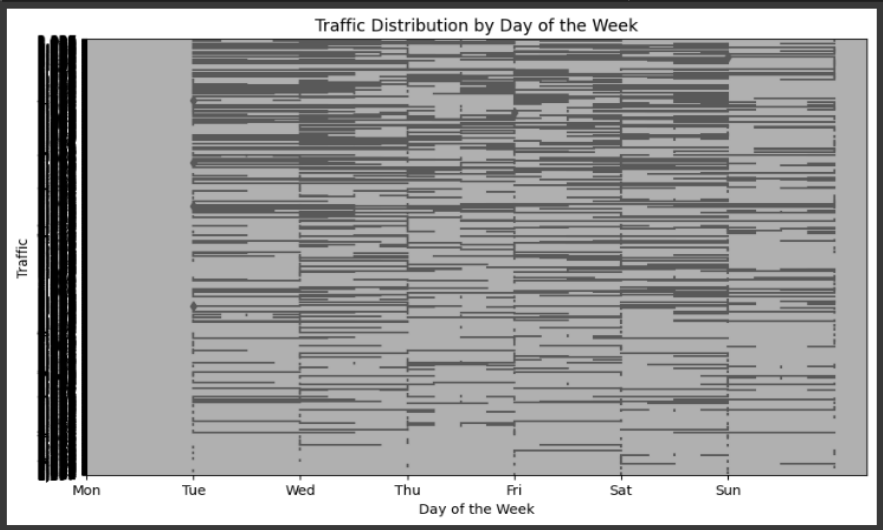
```
<ipython-input-10-274f8ae092af>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  correlation_matrix = data.corr()
```



Define Standards: Identify the standards or guidelines for each water quality parameter. This information can be obtained from regulatory bodies, environmental agencies, or specific standards for drinking water.

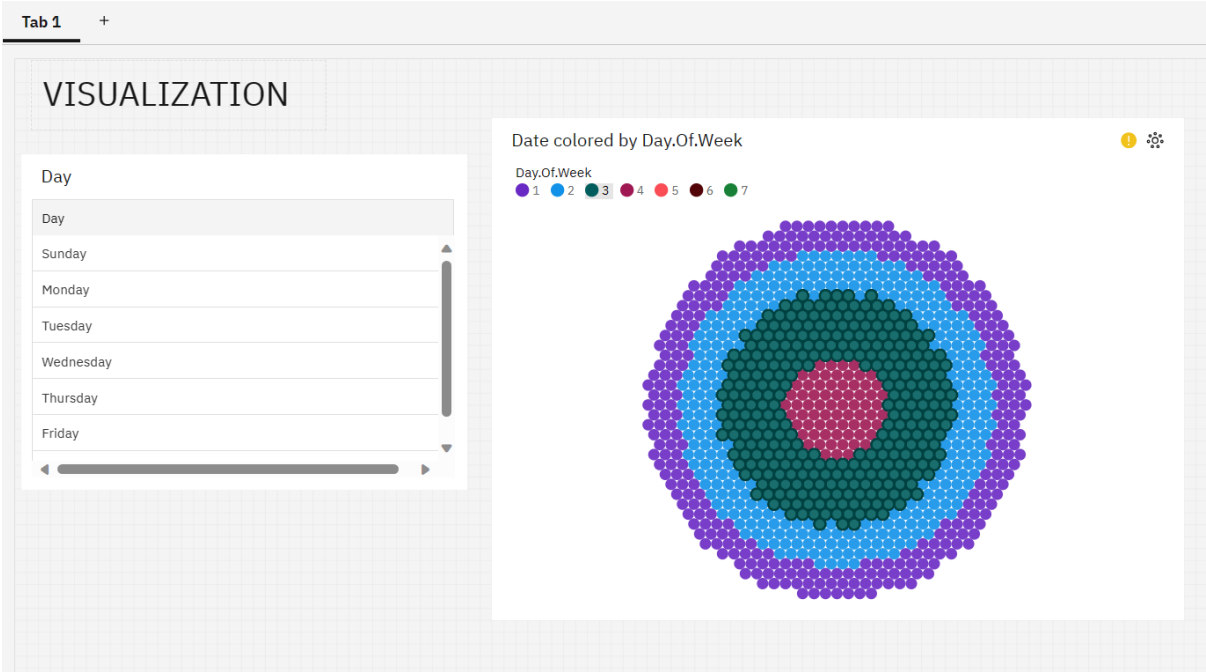
Compare Against Standards: For each parameter, compare the values in your dataset against the established standards. You can create a function to flag instances where the values deviate from the standards.

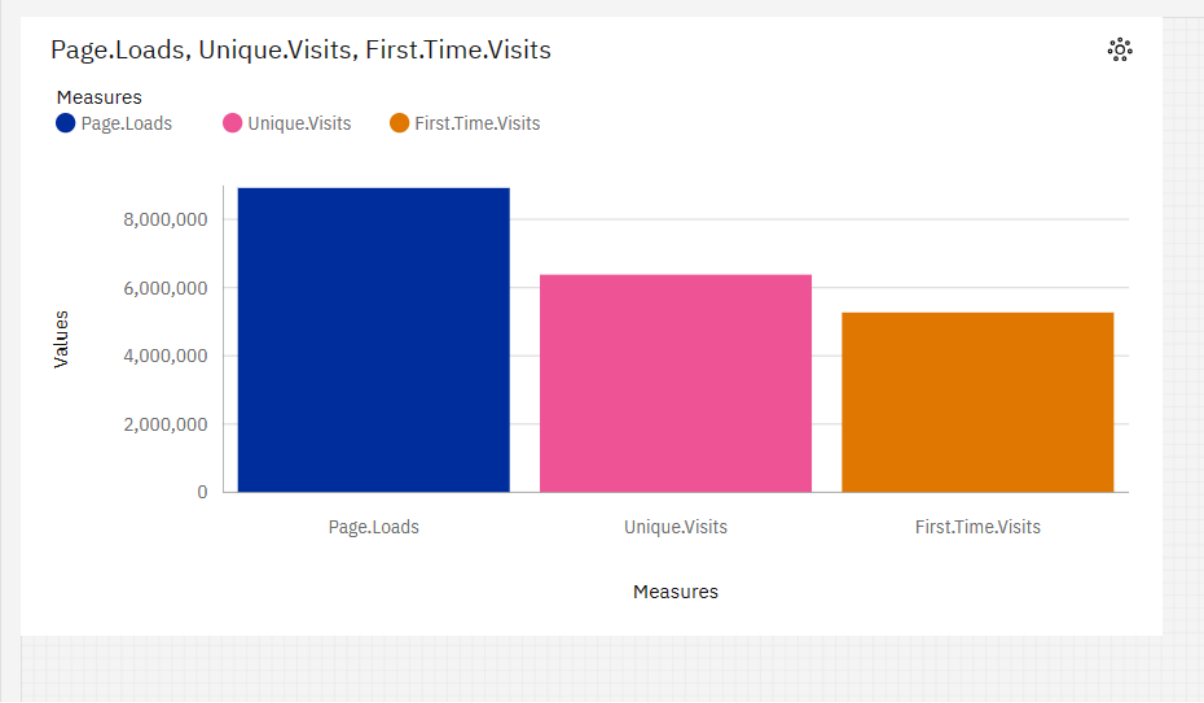
```
# Create a box plot to visualize traffic distribution by day of the week
plt.figure(figsize=(10, 6))
sns.boxplot(x=data['Day.Of.Week'], y=data['First.Time.Visits'], palette='Set2')
plt.xlabel('Day of the Week')
plt.ylabel('Traffic')
plt.title('Traffic Distribution by Day of the Week')
plt.xticks(range(7), ['Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun'])
plt.grid(True)
plt.show()
```

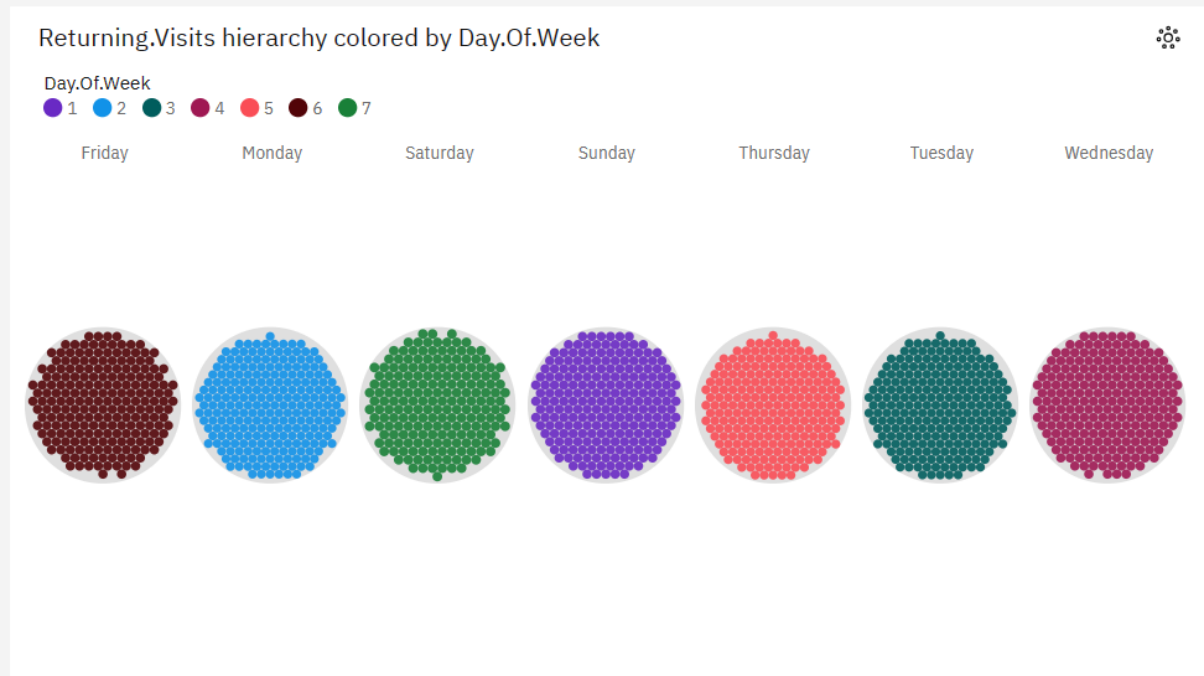


From the above steps we can conclude the cleaning process.

VISUALIZATION USING COGNOS:







CONCLUSION:

The more you know about your traffic analytics, the more effective and impactful your website can be. Equipped with the right website traffic analysis tools, you'll identify your top site pages, track visitor trends, calculate conversion rates, and ensure your marketing spend translates into an increase in conversions and sales.

There are many people out there who run their websites without keeping an eye on their metrics, and while that might have a limited success rate, when you have so many analytics tools at your disposable, it would be foolish to not use them. For example, even if you just keep an eye on the traffic to your site alone, you can monitor whether or not your content strategy is working properly.With web analytics, it can seem quite daunting when you see the amount of data that is created. These numbers could end up putting you off – however, as we've outlined in this article, you only need to focus on the relevant metrics. This will give you the tools for success.

DATASET LINK: https://www.kaggle.com/datasets/bobnau/daily-website-visitors

---

SAKTHIVELAN.M 18:10:2023

NANDHA COLLEGE OF TECHNOLOGY