**PROJECT PHASE-3**

**TOPIC: Website Traffic Analysis**

**Introduction:**

Data cleaning and data visualization are essential components of the website traffic analysis process. They play a vital role in transforming raw, often messy data into meaningful insights that help website owners and businesses understand user behavior, optimize their websites, and make data-driven decisions. Here is an introduction to the data cleaning and data visualization phases of website traffic analysis:

**Data Cleaning for Website Traffic Analysis:**

Website traffic data can be complex, containing various issues such as missing values, duplicates, inconsistencies, and outliers. Data cleaning is the initial step in the analysis process, focused on preparing the data for meaningful insights. It involves:

1.**Handling Missing Values**: Identifying and dealing with data points that are incomplete or missing, which can affect the accuracy of analysis.

2.**Removing Duplicates**: Identifying and eliminating duplicate entries in the dataset to avoid skewing results.

3.**Outlier Detection**: Identifying and addressing outliers that can distort patterns and trends in the data.

4**.Data Transformation**: Converting data types, standardizing formats, and making data consistent to ensure accurate analysis.

5.**Data Validation**: Ensuring data quality by validating entries, checking for inaccuracies, and addressing inconsistencies.

The goal of data cleaning is to create a clean, reliable dataset that can serve as a foundation for accurate analysis.

PROGRAM:

import pandas as pd

import matplotlib.pyplot as plt

import plotly.express as px

import plotly.graph\_objects as go

from statsmodels.tsa.seasonal import seasonal\_decompose

from statsmodels.graphics.tsaplots import plot\_pacf

from statsmodels.tsa.arima\_model import ARIMA

import statsmodels.api as sm

#read the csv file

data = pd.read\_csv("daily-website-visitors.csv", \

index\_col = 'Date', thousands = ',', parse\_dates=True)

print(data.head())

print(data.info())

#1.handle duplicates

data.dropna(inplace=True)

# 2. Removing Duplicates

data.drop\_duplicates(inplace=True)

#3.conversion of data types

data["Date"] = pd.to\_datetime(data["Date"],

format="%m/%d/%Y")

print(data.info())

#4.checking missing values

print(data.isnull().sum())

#5.Drop rows with missing values

data=data.dropna()

#6.remove the duplicate values

data=data.drop\_duplicates()

**OUTPUT:**

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

<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

None

<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 datetime64[ns]

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: datetime64[ns](1), int64(2), object(5)

memory usage: 135.6+ KB

None

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

**Data Visualization for Website Traffic Analysis:**

Data visualization is the art of representing data graphically to make it more accessible and understandable. In website traffic analysis, data visualization serves several critical purposes:

**1.Pattern Identification:** Visualizations help identify patterns and trends in website traffic data. Line charts, for instance, can reveal traffic fluctuations over time.

**2.User Behavior Analysis:** Visualizations like heatmaps and click maps provide insights into how users interact with a website, indicating which elements receive the most attention and which paths they follow.

**3.Source and Channel Analysis:** Visualizations can help website owners understand where their traffic comes from, which sources and channels are most effective, and where they should allocate their marketing resources.

**4.Conversion Funnel Analysis:** By creating conversion funnels, website owners can track the user journey, identify drop-off points, and optimize the conversion process.

**5.Performance Monitoring:** Visualizing page load times, error rates, and other performance metrics allows for the identification of issues affecting the user experience.

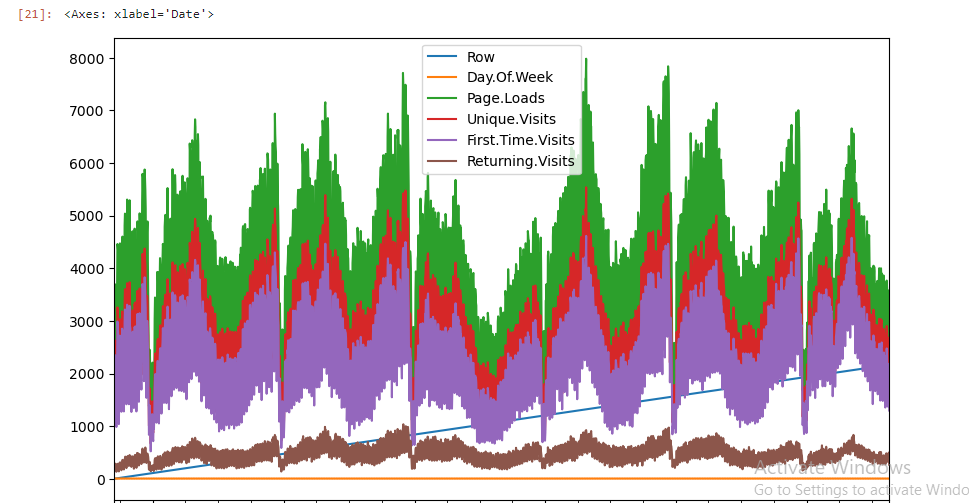
Data visualization not only helps in understanding the current state of the website but also assists in making data-driven decisions to improve user experience and achieve website objectives.

In conclusion, data cleaning and data visualization are integral to website traffic analysis. They enable website owners and businesses to make sense of complex data, uncover insights, and take informed actions to optimize their websites, enhance user experiences, and achieve their goals. These phases are essential for the success of any data-driven online presence.

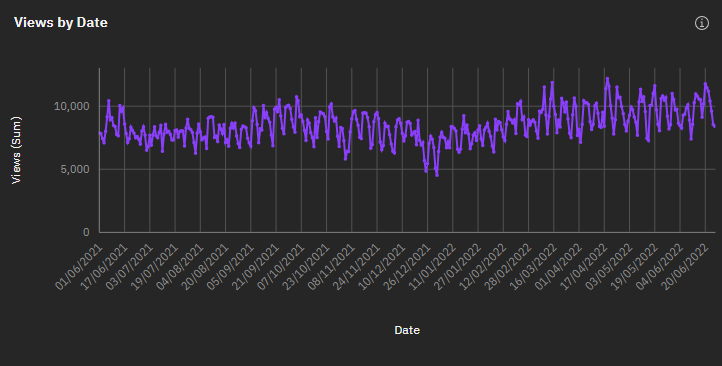
**PROGRAM:**

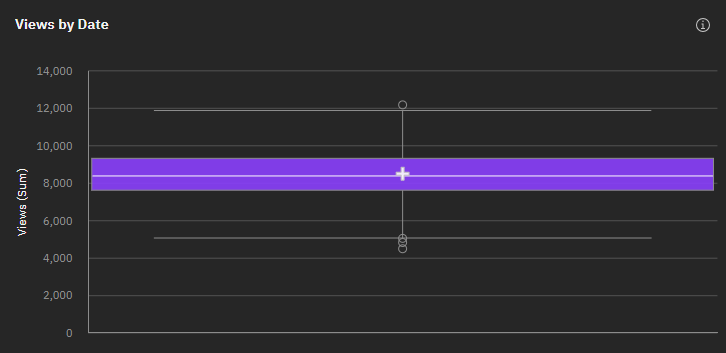
|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  df = pd.read\_csv("daily-website-visitors.csv", \  index\_col = 'Date', thousands = ',', parse\_dates=True)  df.head()  #visualize the figure  df.plot(figsize=(10,6)) |

**OUTPUT:**



**Using IBM Cognos For Data Visualization**

****

****