**WEBSITE TRAFFIC ANALYSIS**

**Project Overview:** Website Traffic Analysis Using IBM Cognos

**PROJECT OBJECTIVE:**

The primary objective of this project is to analyze website traffic data and gain valuable insights into user behavior, traffic sources, and the performance of a website. By utilizing IBM Cognos for visualization, we aim to create clear and actionable visual reports that help in making data-driven decisions to improve website performance and achieve business objectives.

**DATA SOURCES :**

The dataset for this project is sourced from Kaggle and is titled 'Daily Website Visitors.' This dataset contains a comprehensive record of daily website traffic data, providing information on visitor counts, page views, unique visitors, and more. The dataset spans a specific time frame, making it an ideal source for our website traffic analysis project. It is expected to offer valuable insights into user behavior and traffic trends, which will be leveraged to enhance our understanding of website performance.

**Dataset Link:**[**https://www.kaggle.com/datasets/bobnau/daily-website-visitors**](https://www.kaggle.com/datasets/bobnau/daily-website-visitors)

**METHODOLOGY**

**DATA COLLECTION**

For this project, we collected the daily website visitor data from the provided dataset hosted on Kaggle ("Daily Website Visitors"). This dataset includes relevant metrics such as the number of visitors, page views, and additional information.

**DATA PREPROCESSING**

**1. Data Cleaning**

We began by cleaning the dataset to ensure data quality and consistency.

This involved:

* Handling missing values by imputation or removal.
* Eliminating duplicate records to avoid redundancy.
* Standardizing data entries, such as URL formats and source names.

**2. Data Transformation**

To facilitate analysis, we carried out data transformations including:

* Converting data types where necessary, such as date formats.
* Normalizing numerical data to a common scale.
* Aggregating data into daily summaries.
* IBM Cognos for Visualization

**3. Data Loading**

The preprocessed dataset was loaded into IBM Cognos for visualization. IBM Cognos offers a user-friendly platform for creating interactive visual reports, which is crucial for translating our findings effectively.

**4. Objective Definition**

We defined specific objectives for our analysis and visualization. These objectives guided the creation of meaningful reports to address our research questions.

**DATA ANALYSIS**

**1. Exploratory Data Analysis (EDA)**

We conducted EDA to gain insights into the dataset, such as identifying key trends, patterns, and distributions. EDA provided a foundation for more in-depth analysis.

**2. Traffic Sources Analysis**

We explored different traffic sources, including organic search, direct traffic, referral traffic, and paid advertising. Our aim was to identify which sources contributed the most to website traffic and assess their quality.

**3. User Behavior Analysis**

Analysis of user behavior metrics, such as average session duration, pages per session, and conversion rates, allowed us to determine which pages or content were most engaging and which led to conversions.

**4. Traffic by Geography**

We investigated the geographical distribution of website visitors to identify which regions or countries contributed the most traffic.

**5. Time Analysis**

We analyzed how website traffic varied over time, identifying daily, weekly, or monthly trends, and looking for seasonality and notable changes in traffic patterns.

**CONCLUSION**

This methodology guided our entire project, from data collection and preprocessing to in-depth analysis and report generation. It ensures that our findings are based on a systematic and structured approach.

**SOURCE CODE**

import pandas as pd

import matplotlib.pyplot as plt

# Load data from 'cleaned\_data.csv' into a DataFrame

file\_path = 'cleaned\_data.csv'

df = pd.read\_csv(file\_path)

# Convert the 'Date' column to a datetime data type

df['Date'] = pd.to\_datetime(df['Date'])

# Total Page Loads by Day of the Week

total\_page\_loads\_by\_day = df.groupby('Day')['Page.Loads'].sum()

# Average Unique Visits by Day of the Week

average\_unique\_visits\_by\_day = df.groupby('Day')['Unique.Visits'].mean()

# Total Page Loads and Unique Visits Over Time

page\_loads\_over\_time = df.set\_index('Date')['Page.Loads']

unique\_visits\_over\_time = df.set\_index('Date')['Unique.Visits']

# Analysis of Returning Visitors

returning\_visitors\_stats = df['Returning.Visits'].describe()

# Page Load and Unique Visit Trends Over Time

plt.figure(figsize=(12, 6))

plt.plot(df['Date'], df['Page.Loads'], label='Page Loads', marker='o')

plt.plot(df['Date'], df['Unique.Visits'], label='Unique Visits', marker='o')

plt.xlabel('Date')

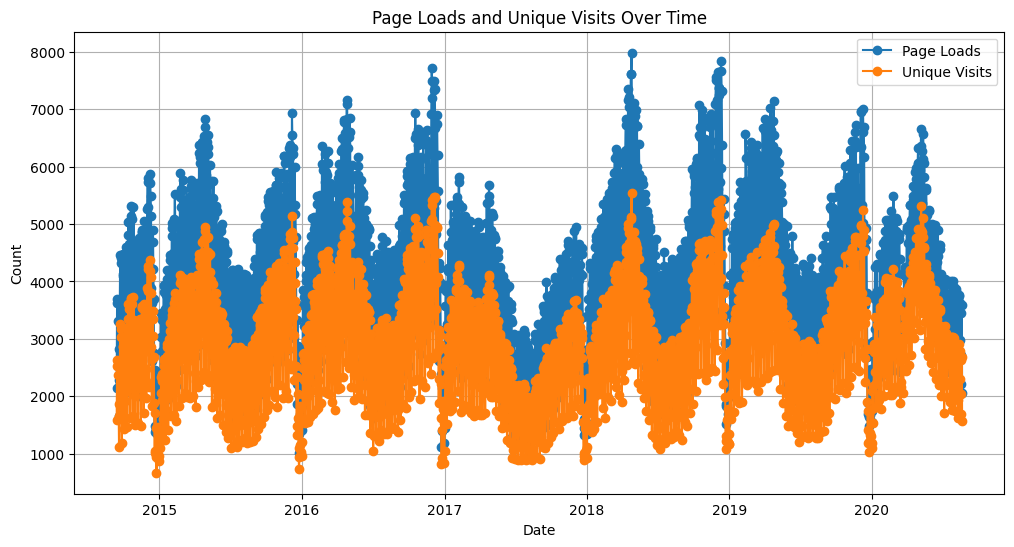
plt.ylabel('Count')

plt.title('Page Loads and Unique Visits Over Time')

plt.legend()

plt.grid()

plt.show()



import seaborn as sns

total\_page\_loads\_by\_day = df.groupby('Day')['Page.Loads'].sum()

# Create an attractive bar chart using Seaborn

plt.figure(figsize=(10, 6))

sns.set\_style("whitegrid")

sns.barplot(x=total\_page\_loads\_by\_day.index, y=total\_page\_loads\_by\_day.values, palette="viridis")

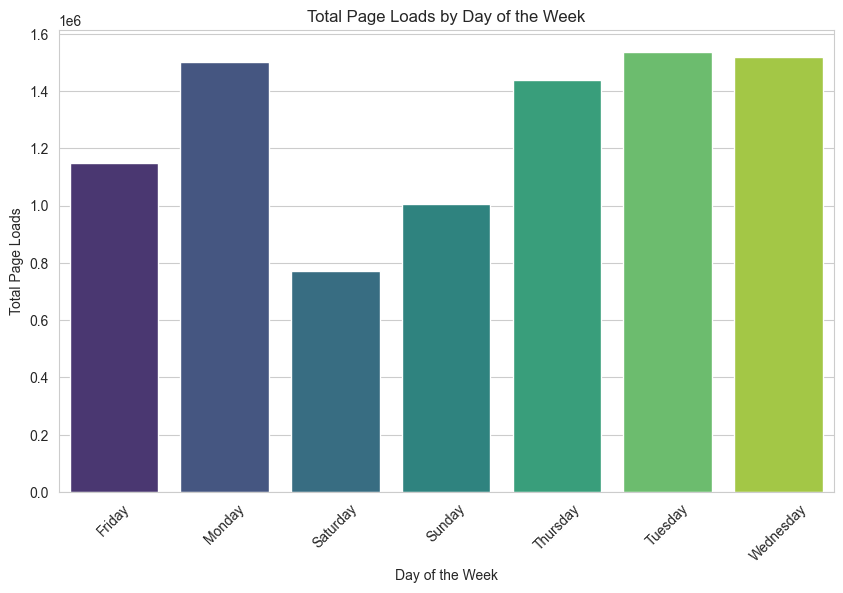
plt.title("Total Page Loads by Day of the Week")

plt.xlabel("Day of the Week")

plt.ylabel("Total Page Loads")

plt.xticks(rotation=45)

plt.show()



**IBM COGNOS**

