

Website Traffic Analysis Report

Name: Vaishnavi Rai

Class: CSE AIML C

Roll No:

202401100400205

Introduction

Website traffic data refers to the comprehensive collection of metrics and information about the interactions and activity on a website. This data tracks how users visit, navigate, and engage with a website, offering valuable insights into its performance, audience behavior, and areas for improvement. It is an essential component of website analytics and is widely used in marketing, business development, user experience design, and decision-making processes

Methodology

- 1. The dataset was uploaded and read into a DataFrame.**
- 2. The 'Date' column was converted to datetime format.**
- 3. Data was checked for missing values and cleaned.**
- 4. Summary statistics and correlation were calculated.**
- 5. Page Views over time were plotted.**
- 6. A heatmap was used to visualize correlations**

CODE

```
# Import necessary libraries

import pandas as pd # For data manipulation and analysis

import matplotlib.pyplot as plt # For data visualization

import seaborn as sns # For creating attractive statistical graphics


# Upload the dataset

from google.colab import files

uploaded = files.upload() # Opens a file upload dialog to upload a CSV file


# Read the dataset into a DataFrame

filename = list(uploaded.keys())[0] # Extract the filename from the uploaded files

data = pd.read_csv(filename) # Load the CSV file into a pandas DataFrame


# Display the first few rows of the dataset to understand its structure

print("Preview of the dataset:")

print(data.head())


# Convert the 'Date' column to datetime format for easier analysis

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


# Set the 'Date' column as the index to make time-series analysis easier

data.set_index('Date', inplace=True)
```

```
# Plot overall page views over time

plt.figure(figsize=(10, 5)) # Set the figure size

plt.plot(data.index, data['Page Views'], label='Page Views', color='blue') # Line plot

plt.xlabel('Date') # Label for the X-axis

plt.ylabel('Page Views') # Label for the Y-axis

plt.title('Website Traffic - Page Views Over Time') # Title of the plot

plt.legend() # Add a legend to the plot

plt.grid(True) # Add grid lines for better readability

plt.show() # Display the plot
```

```
# Analyze the correlation between different metrics in the dataset

correlation = data[['Page Views', 'Unique Visitors', 'Bounce Rate']].corr()

print("Correlation between metrics:")

print(correlation)
```

```
# Visualize the correlation matrix using a heatmap

plt.figure(figsize=(8, 6)) # Set the figure size

sns.heatmap(correlation, annot=True, cmap='coolwarm', fmt=".2f") # Create a heatmap

plt.title('Correlation Heatmap') # Title of the heatmap

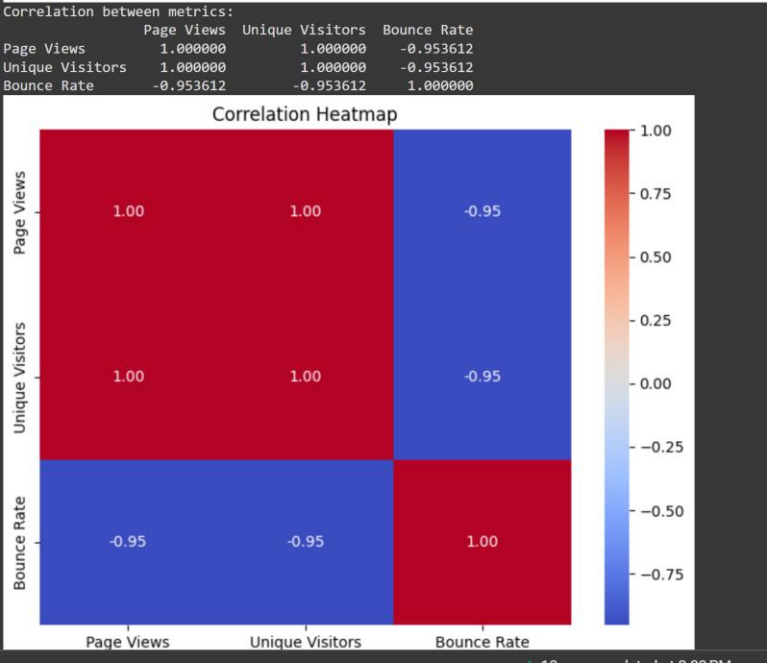
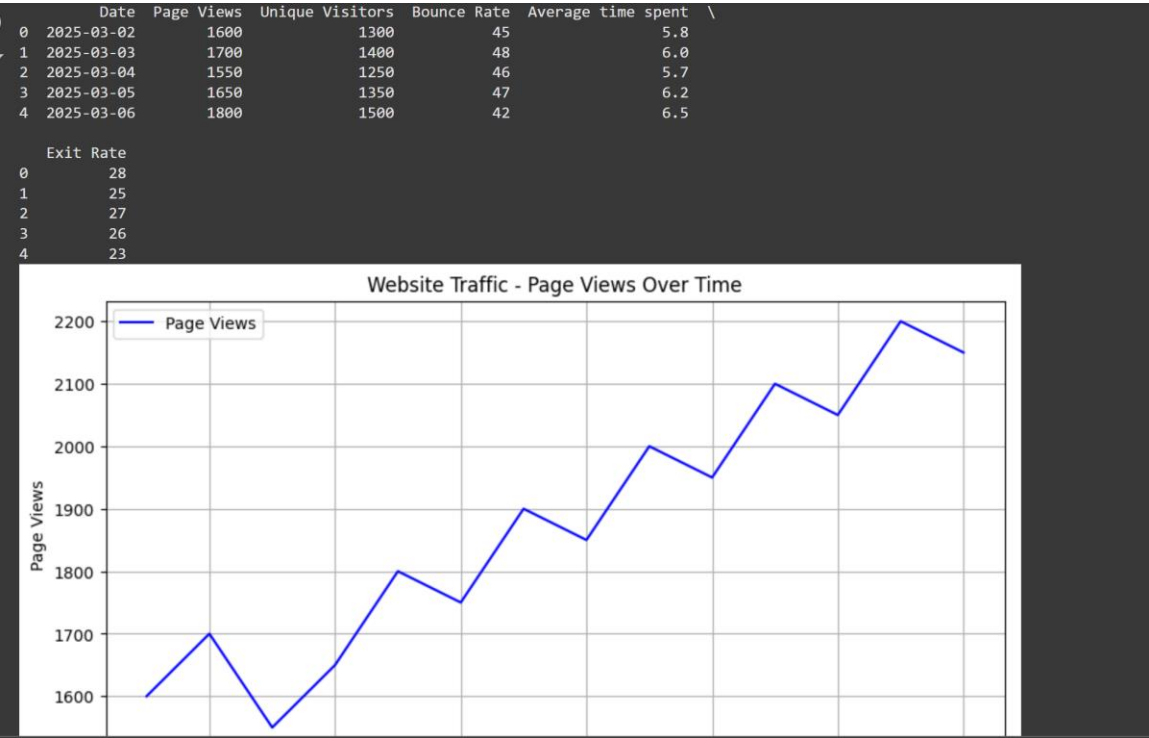
plt.show() # Display the heatmap
```

```
# Find and print summary statistics for each column in the dataset

print("Summary Statistics:")

print(data.describe())
```

OUTPUT/RESULT



References/Credits

Data source: Provided dataset from chatgpt

Libraries used: Pandas, Matplotlib, Seaborn