www.kiet.edu
Delhi-NCR, Ghaziabad

KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

Analyzing Website Traffic Data Prepared

by: **Pratham Mishra**

Date: 11 March 2025

Course: INTRODUCTION TO AI

1. Introduction

This report provides an analysis of website traffic data to understand patterns, trends, and key insights. By leveraging the traffic_data.csv dataset, the analysis utilizes Python's Pandas and Matplotlib libraries for efficient data handling and visualization. The goal is to uncover valuable insights such as peak traffic periods, relationships between metrics, and overall website performance.

2.Methodology for Analyzing Traffic Data with AI

1. Problem Definition

- The primary goal is to analyze and extract meaningful insights from website traffic data and identify patterns, trends, and anomalies.
- AI plays a role in automating the analysis process, enhancing predictive accuracy, and identifying underlying relationships in the data.

2. Data Collection

- The data is sourced from the CSV file traffic_data.csv, which contains metrics like Date, PageViews, UniqueVisitors, and BounceRate.
- AI-enabled tools can be used to regularly fetch or update traffic data from web analytics platforms through APIs (e.g., Google Analytics API).

3. Data Preprocessing

- Handling Missing Values: AI algorithms such as KNN Imputation or regression-based methods can fill missing data points.
- Feature Engineering:
  - Convert the Date column into a machine-readable format (datetime object).
  - Extract additional temporal features (e.g., day of the week, month) to enhance model performance.
- Scaling and Normalization:
  - Standardize data (e.g., Min-Max Scaling) to ensure consistent input for AI models.
  - Use AI preprocessing frameworks like Scikit-learn to automate these transformations.

4. Exploratory Data Analysis (EDA)

- EDA leverages AI-assisted tools to generate:

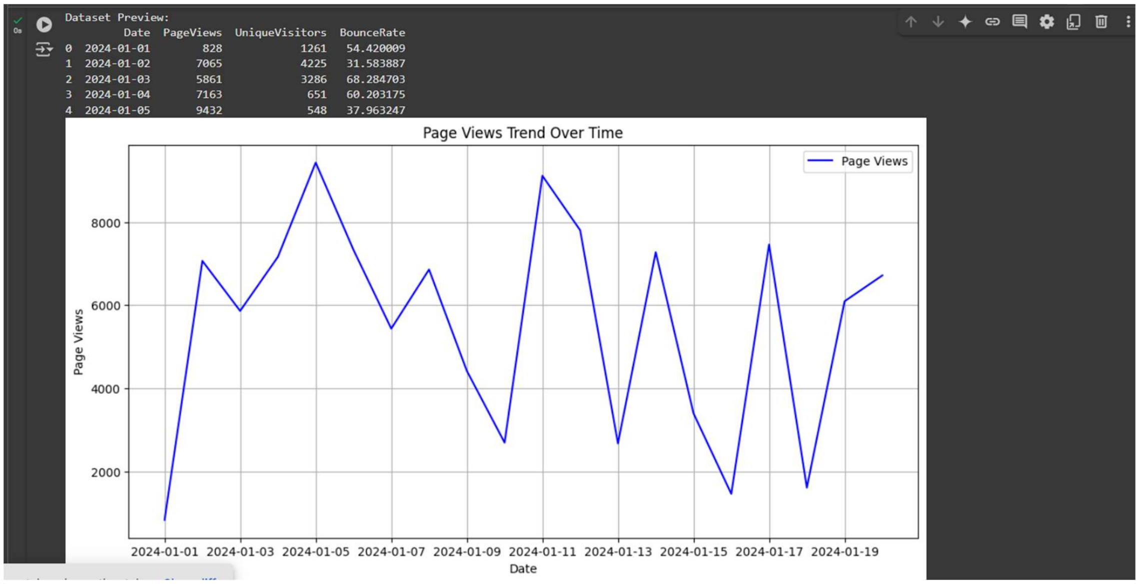
- Automated data visualizations for trends and patterns (e.g., AI visualization platforms like Tableau with ML plugins).
- Clustering techniques (e.g., K-Means or DBSCAN) to group similar traffic patterns.

CODE TYPED:

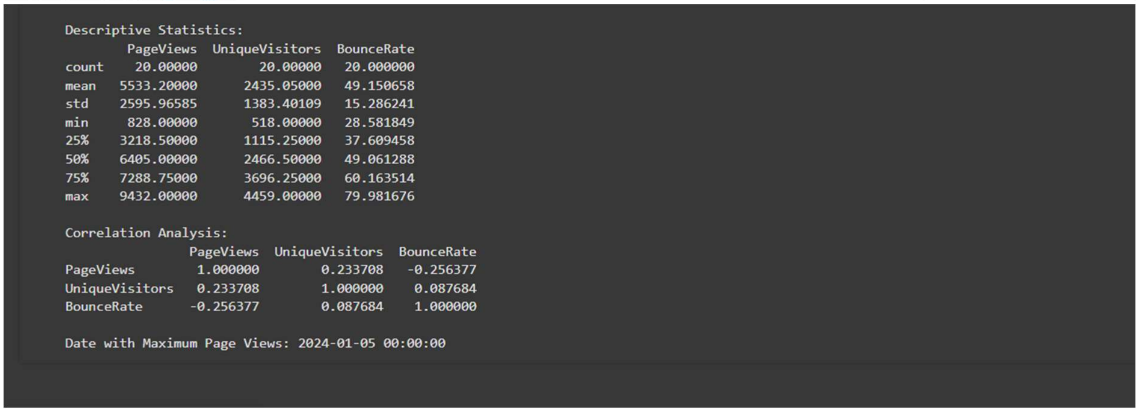
```
import pandas as pd
import matplotlib.pyplot as plt

# Read data from the CSV file
# This assumes the data is stored in a file named 'traffic_data.csv'
filename = "traffic_data.csv"
df = pd.read_csv(filename)

# Display the first few rows of the dataset to get an overview
print("Dataset Preview:")
print(df.head())

# Convert the 'Date' column from string format to a datetime object
# This makes it easier to perform date-based operations
df['Date'] = pd.to_datetime(df['Date'])

# Set the 'Date' column as the index for the DataFrame
# This is useful for time-series analysis and plotting
df.set_index('Date', inplace=True)
```

```python
# Plot the trend of Page Views over time
# This visualization helps analyze the traffic pattern
plt.figure(figsize=(12,6))  # Set the figure size for better readability
plt.plot(df.index, df['PageViews'], label='Page Views', color='blue')
plt.xlabel('Date')  # Label for the x-axis
plt.ylabel('Page Views')  # Label for the y-axis
plt.title('Page Views Trend Over Time')  # Title of the plot
plt.legend()  # Show the legend for the plot
plt.grid(True)  # Add a grid to the plot
plt.show()

# Calculate and display key statistical metrics for the dataset
# This provides insights into the data's distribution and variance
print("\nDescriptive Statistics:")
print(df.describe())

# Compute the correlation matrix for selected columns
# This shows the relationship between Page Views, Unique Visitors, and Bounce Rate
correlations = df[['PageViews', 'UniqueVisitors', 'BounceRate']].corr()
print("\nCorrelation Analysis:")
print(correlations)

# Find the date with the maximum number of Page Views
# This highlights the day with peak website traffic
highest_page_views_date = df['PageViews'].idxmax()
print(f"\nDate with Maximum Page Views: {highest_page_views_date}")
```

SCREENSHOTS:



```
Dataset Preview:
         Date  PageViews  UniqueVisitors  BounceRate
0  2024-01-01        828            1261   54.420009
1  2024-01-02       7065            4225   31.583887
2  2024-01-03       5861            3286   68.284703
3  2024-01-04       7163             651   60.203175
4  2024-01-05       9432             548   37.963247
```



```
Descriptive Statistics:
        PageViews  UniqueVisitors  BounceRate
count    20.00000        20.00000   20.000000
mean   5533.20000      2435.05000   49.150658
std    2595.96585      1383.40109   15.286241
min     828.00000       518.00000   28.581849
25%    3218.50000      1115.25000   37.609458
50%    6405.00000      2466.50000   49.061288
75%    7288.75000      3696.25000   60.163514
max    9432.00000      4459.00000   79.981676

Correlation Analysis:
                PageViews  UniqueVisitors  BounceRate
PageViews        1.000000        0.233708   -0.256377
UniqueVisitors   0.233708        1.000000    0.087684
BounceRate      -0.256377        0.087684    1.000000

Date with Maximum Page Views: 2024-01-05 00:00:00
```