Exploratory Data Analysis

Aditya Pandey
Yash Kamble

Introduction

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. The objective is to discover patterns, spot anomalies, and test hypotheses, guiding the next steps in data analysis and decision making.

01

Overview

Define Exploratory Data Analysis (EDA)

EDA is a critical step in the data analysis process, involving summarizing the main characteristics of a dataset. It emphasizes graphical representations, helping analysts understand the data's underlying structure and relationships.

Objective of the analysis

The primary objective of this EDA is to extract meaningful insights from website traffic data, identify trends and patterns, and guide strategic decisions based on the analysis.

Importance of EDA in website traffic

Conducting EDA on website traffic data helps to understand user behavior, optimize user experience, and measure engagement metrics effectively, leading to informed improvements in web strategies.

02

Dataset

Description of the dataset

The dataset consists of records captured from website visits, including unique page views, user sessions, traffic sources, and user demographics. This information is crucial for comprehensive analysis.

Data source information

The data is sourced from web analytics tools such as Google Analytics, which tracks user interactions on the website, providing essential data for analysis.

Key features to analyze

Key features in the dataset include visitor counts, session durations, bounce rates, traffic sources, and user demographics. These features will be analyzed to derive actionable insights.

03

Data Cleaning

Handling missing values

To handle missing values, various strategies were applied, such as imputation, where missing entries are replaced with the mean, median, or mode, and removal of records that had excessive missing data. It ensures that the analysis is based on complete information.

Feature engineering

Feature engineering involves creating new features from existing data to improve model performance. Techniques used may include creating categorical variables from continuous data, combining multiple features into one, or decomposing date-time features into day, month, and year to reveal trends.

Data transformation methods

Data transformation methods applied include normalization and scaling of numerical features, which help in adjusting the data to fit within a specific range without distorting differences in the ranges of values. This aids in better analysis and interpretation.

04

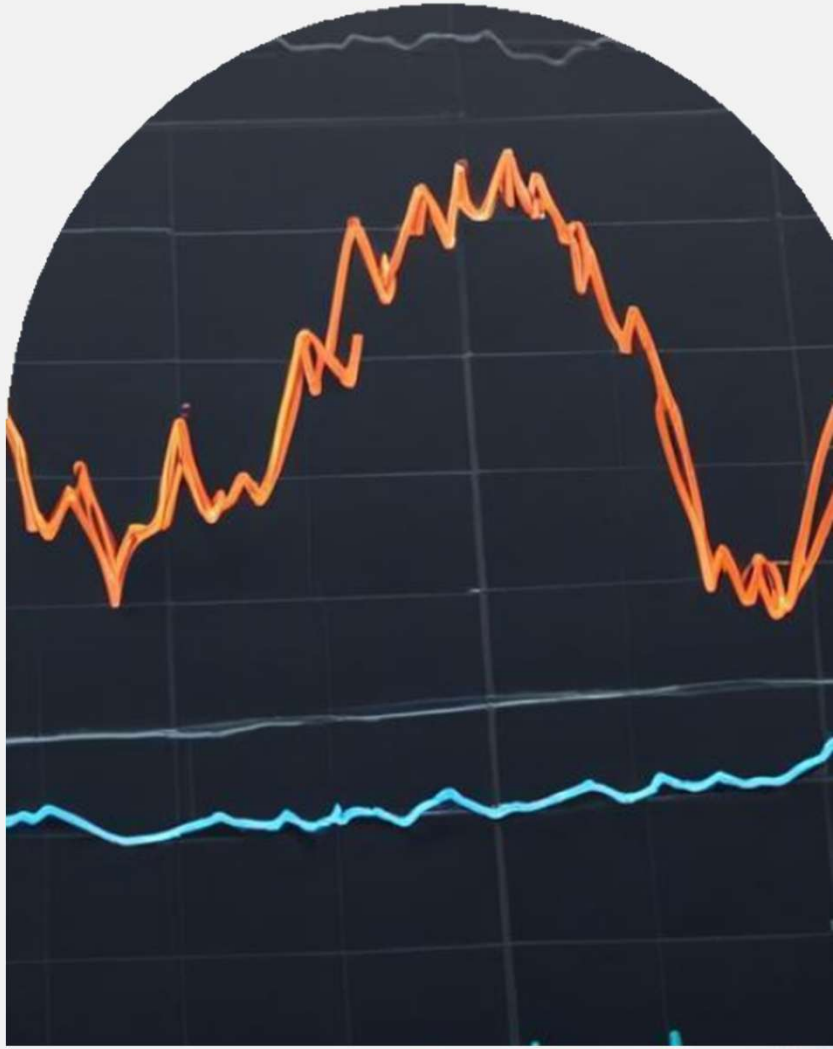
Insights

Summary statistics

Summary statistics such as mean, median, mode, and standard deviation provide a quick overview of the central tendency and dispersion of website traffic data. Understanding these metrics helps in evaluating visitor engagement and overall site performance.

Trends in website traffic

Analyzing trends in website traffic involves identifying patterns over time, such as increases during specific campaigns or seasonal shifts in visitor behavior. This analysis helps businesses anticipate user activity and optimize their strategies accordingly.



Patterns of visitor engagement

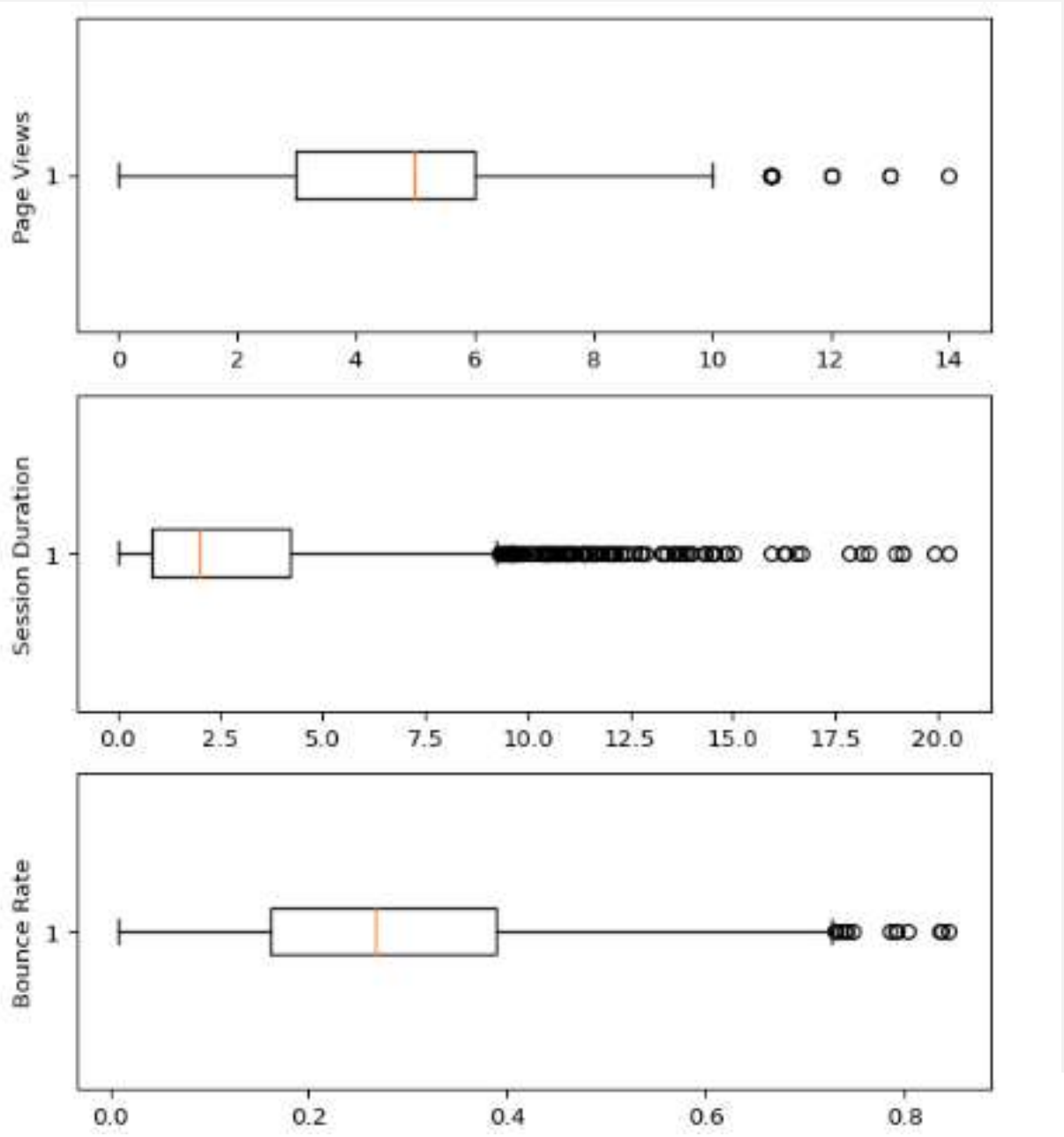
Identifying patterns of visitor engagement includes analyzing metrics such as average session duration, pages per session, and bounce rates. Understanding these patterns can guide optimizations to enhance user experience and retain visitors longer.

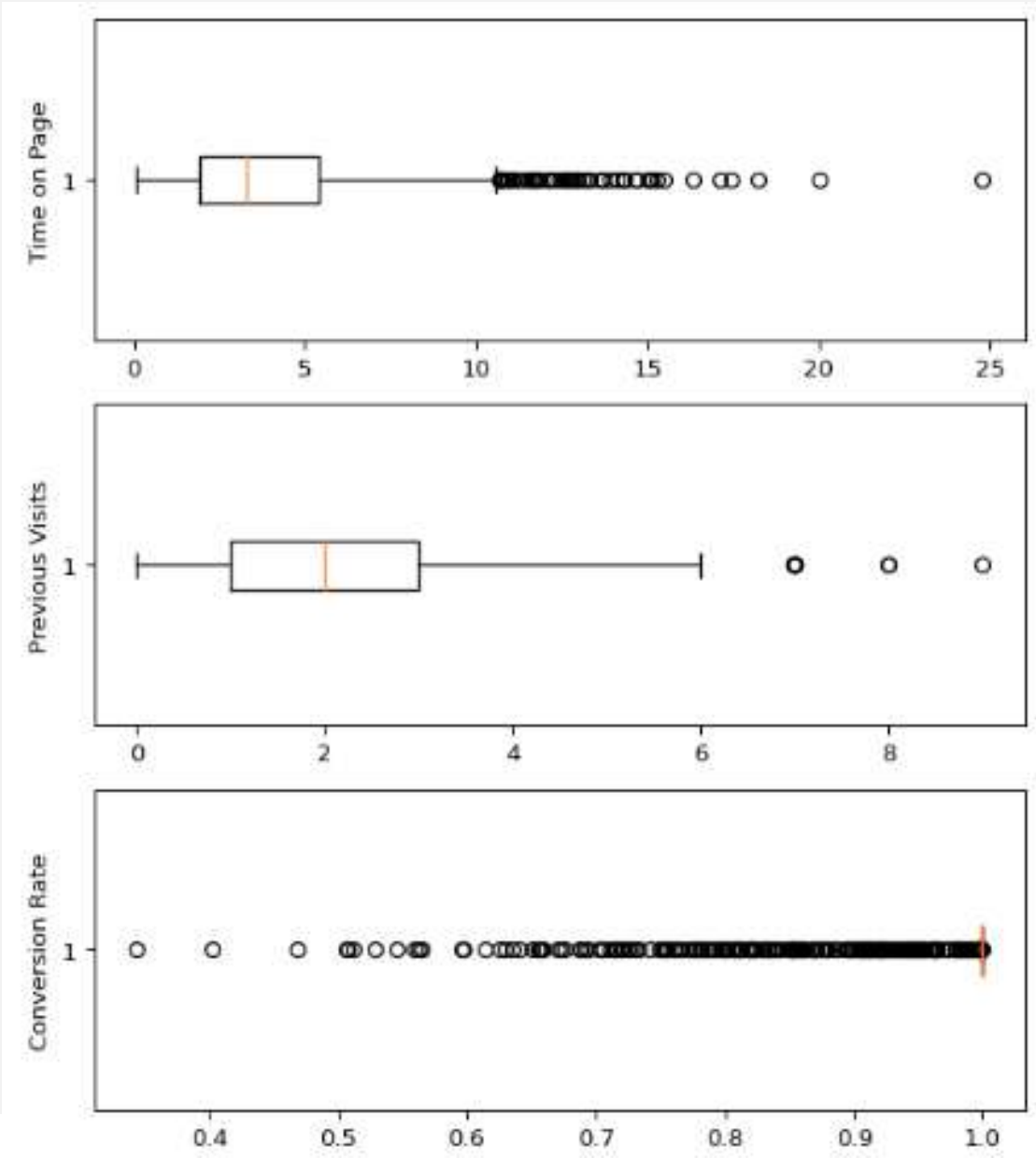
05

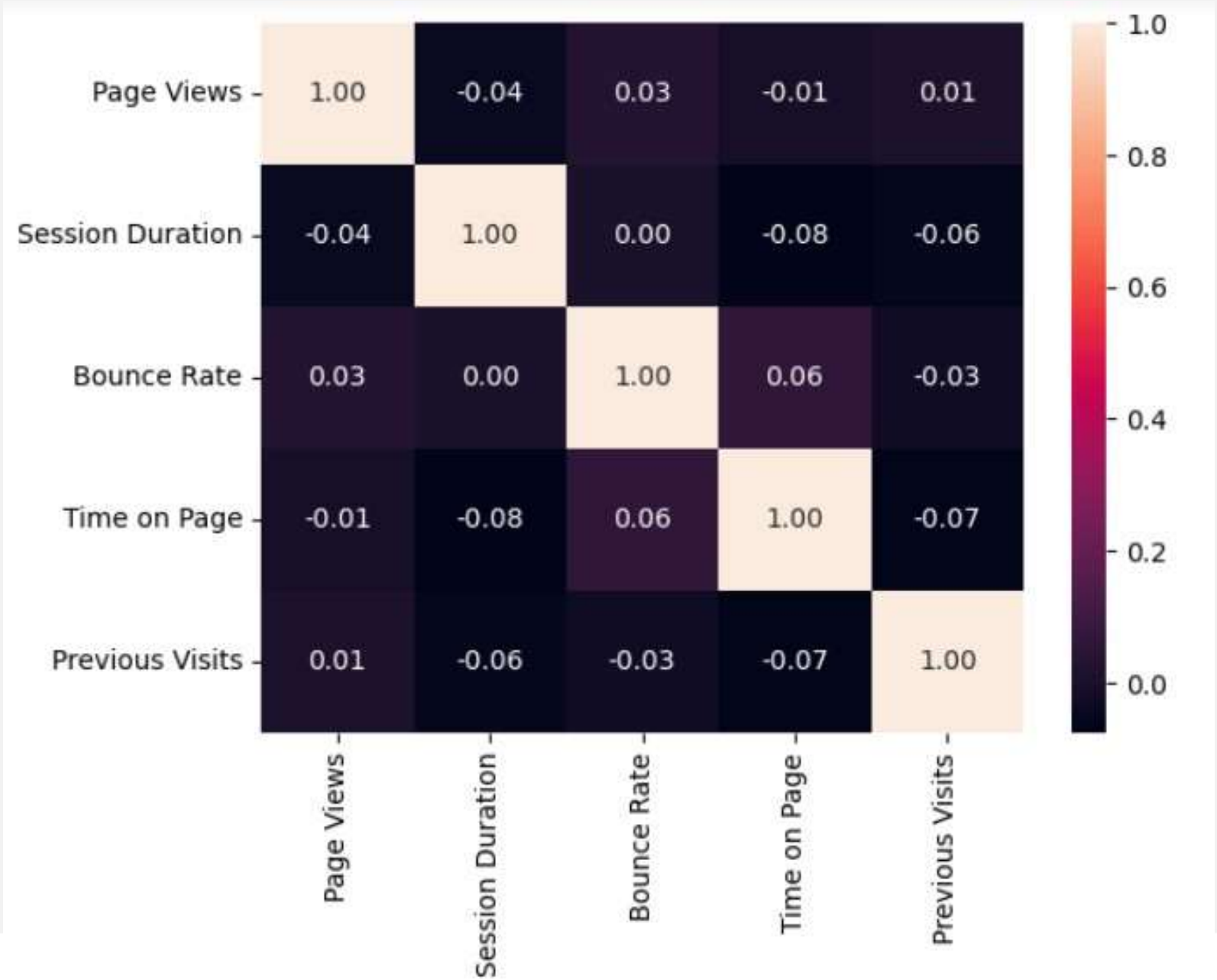
Visuals

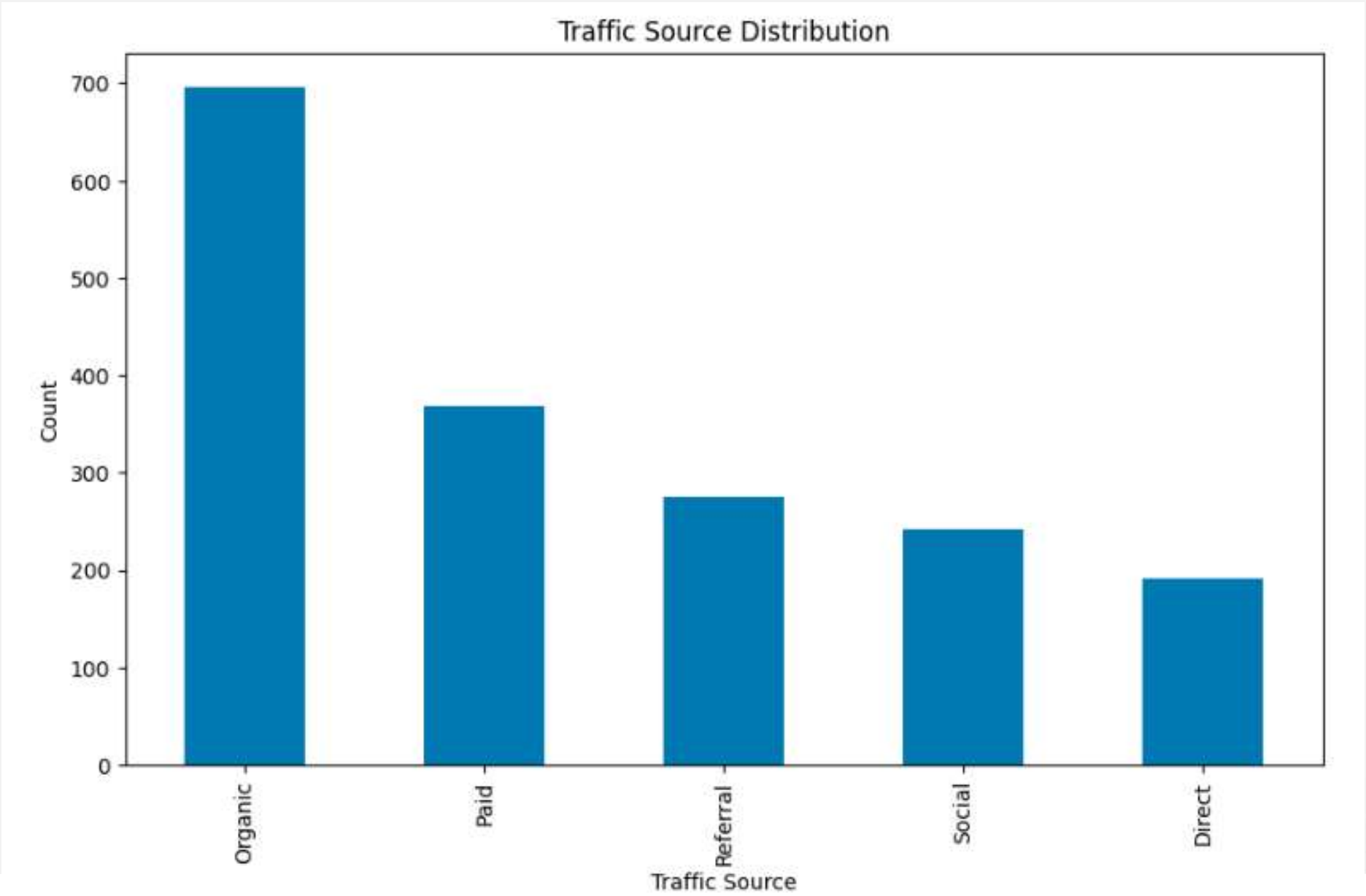
Graphs and charts for data analysis

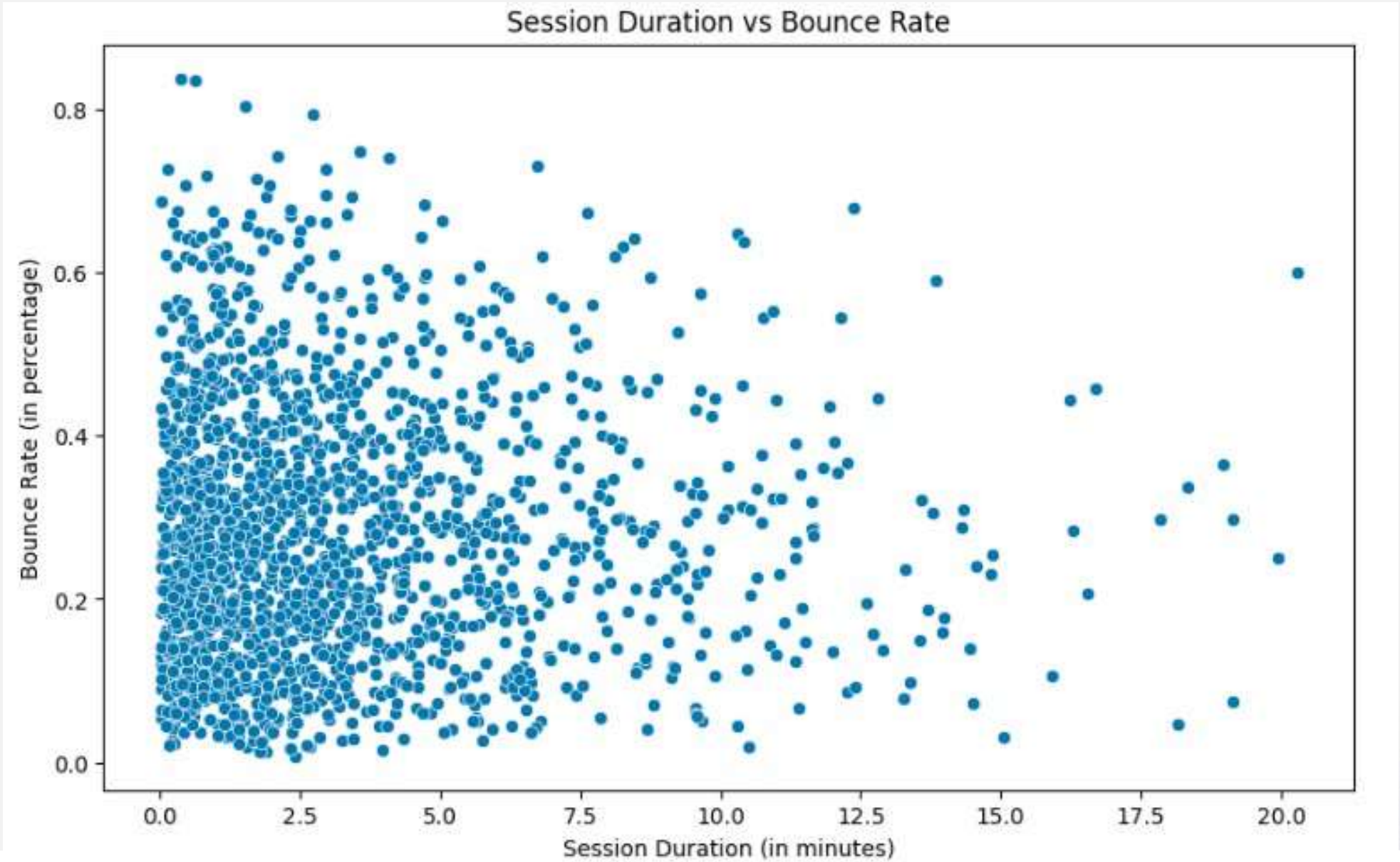
Graphs and charts, such as bar charts, pie charts, and histograms, are essential tools for visualizing data analysis results. They provide an easy-to-understand representation of complex data, enabling stakeholders to grasp insights at a glance.

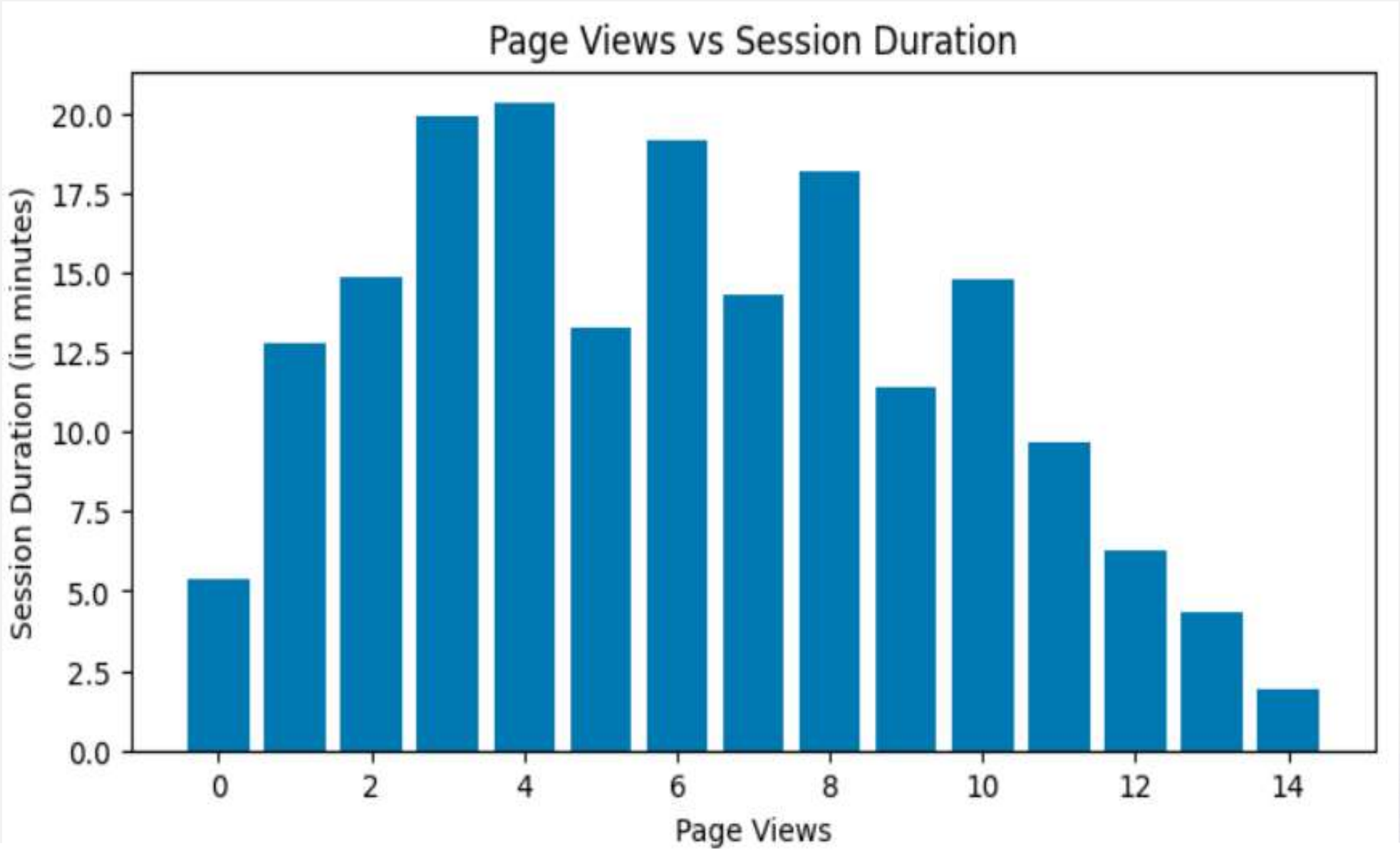

Page Views
1
0
2
4
6
8
10
12
14
Session Duration
1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Bounce Rate
1
0.0
0.2
0.4
0.6
0.8

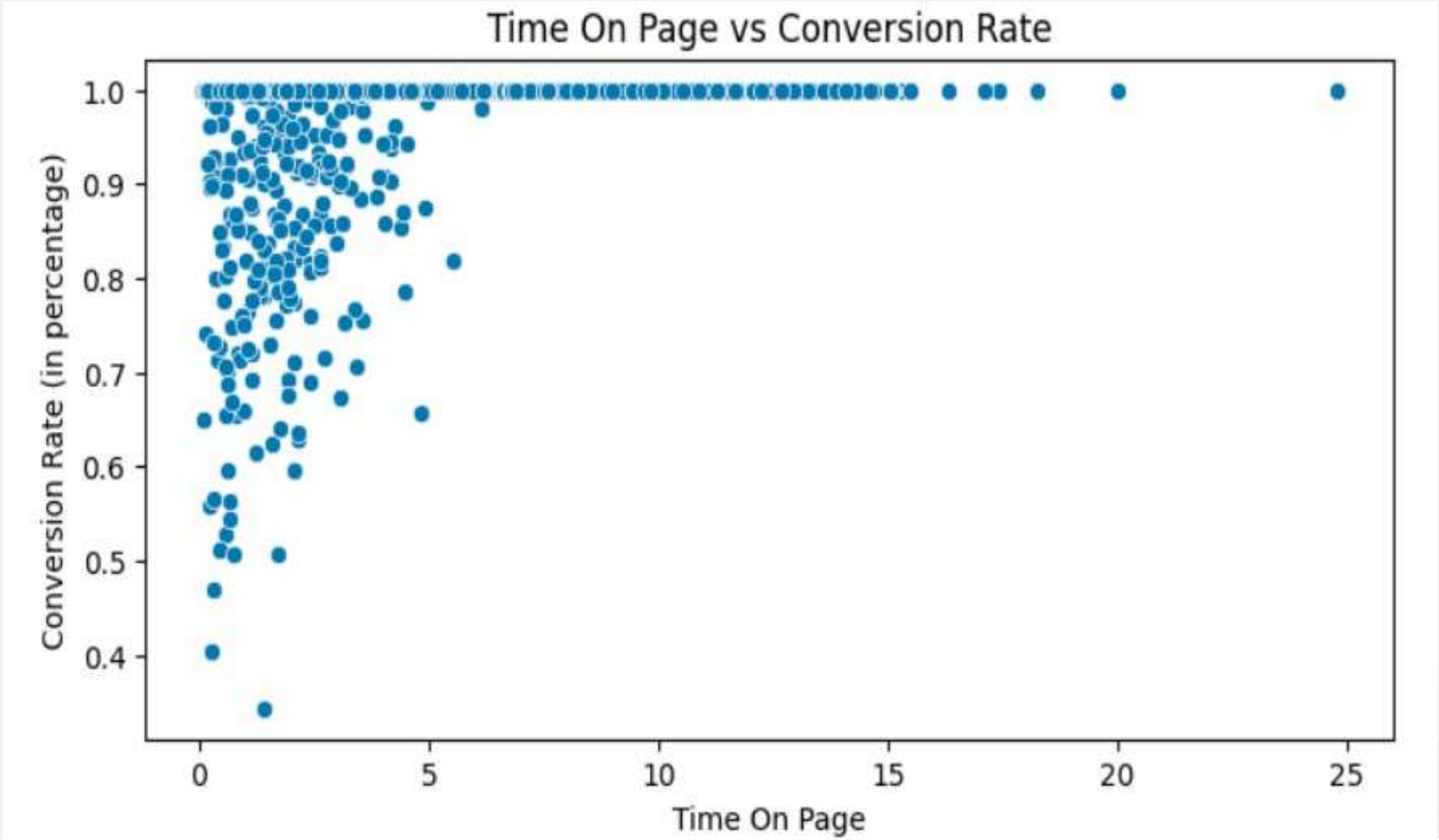

Time on Page
1
0
5
10
15
20
25
Previous Visits
1
0
2
4
6
8
Conversion Rate
1
0.4
0.5
0.6
0.7
0.8
0.9
1.0

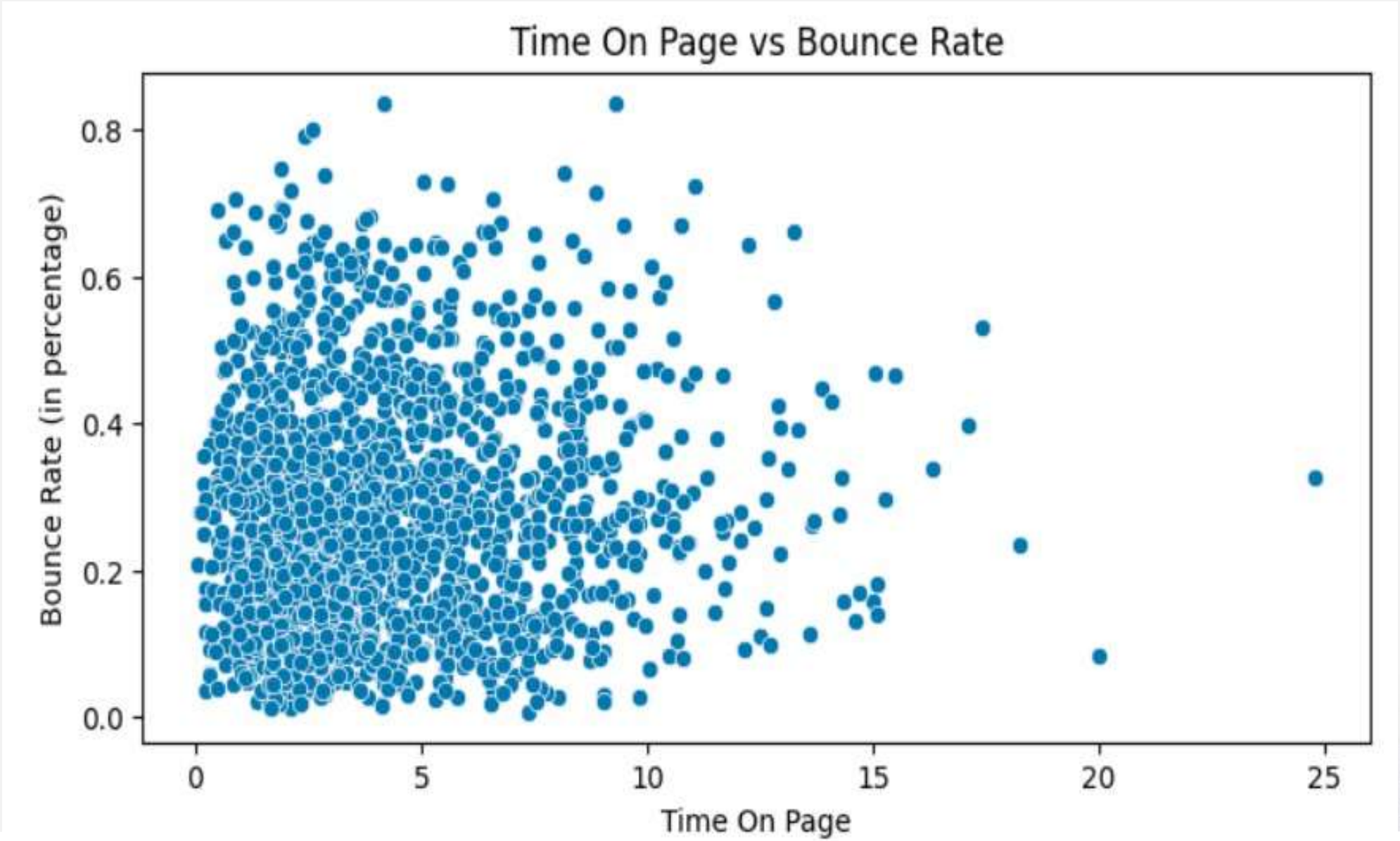

Page Views
Session Duration
Bounce Rate
Time on Page
Previous Visits
1.00
-0.04
0.03
-0.01
0.01
-0.04
1.00
0.00
-0.08
-0.06
0.03
0.00
1.00
0.06
-0.03
-0.01
-0.08
0.06
1.00
-0.07
0.01
-0.06
-0.03
-0.07
1.00
Page Views
Session Duration
Bounce Rate
Time on Page
Previous Visits
1.0
0.8
0.6
0.4
0.2
0.0


Traffic Source Distribution
Count
700
600
500
400
300
200
100
0
Organic
Paid
Referral
Social
Direct
Traffic Source


Session Duration vs Bounce Rate
Bounce Rate (in percentage)
0.0
0.2
0.4
0.6
0.8
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Session Duration (in minutes)


Page Views vs Session Duration
Session Duration (in minutes)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
2
4
6
8
10
12
14
Page Views


Time On Page vs Conversion Rate
Conversion Rate (in percentage)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
5
10
15
20
25
Time On Page


Time On Page vs Bounce Rate
Bounce Rate (in percentage)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Time On Page

Time-series analysis of traffic

Time-series analysis involves studying the website traffic data points collected or recorded over time. This analysis helps reveal trends, cycles, and seasonal variations, assisting in forecasting future traffic patterns.

Key engagement metrics (bounce rate, session duration)

Key engagement metrics such as bounce rate and session duration provide insights into user behavior. A high bounce rate might indicate issues with content quality, while longer session durations typically suggest better engagement and interest in the site.

Conclusions

In conclusion, the exploratory data analysis on website traffic reveals significant insights into user behavior, helping identify key patterns and trends that can drive decision-making. Future strategies should focus on leveraging these insights for website optimization.

Thank you!

Do you have any questions?