Website Traffic Analysis

1. Project Overview

This report summarizes the analysis of a website traffic dataset to identify key user behaviour patterns and measure website performance. The primary goal was to extract actionable insights that can inform marketing strategies and guide website optimization efforts.

The project involved a complete data lifecycle: generating a synthetic dataset with Python, cleaning and preparing the data using MySQL, performing in-depth analysis through SQL queries, and visualizing the findings in a comprehensive dashboard.

2. Objectives

- To analyse website traffic to understand where users are coming from (traffic sources) and what devices they use.
- To evaluate user engagement by measuring key performance indicators (KPIs) such as session duration, bounce rate, and pageviews.
- To measure website effectiveness through conversion and user retention rates.
- To develop a visual dashboard to communicate these findings effectively to stakeholders.

3. Methodology and Tools

The project was executed in four distinct phases:

1. Data Generation: A synthetic dataset was created using Python. It's important to note that as a simulated dataset, it contained certain imbalances which were accounted for during the analysis.
2. Data Cleaning and Preparation (MySQL): The raw data was loaded into a MySQL database. Initial preparation involved handling missing values and removing extreme outliers to ensure data quality. A dedicated SQL view cleaned_traffic was created to streamline the analysis process.
3. Data Analysis (MySQL): All key metrics were calculated using MySQL queries on the cleaned_traffic view. This included aggregations for total sessions, bounce rates by traffic source, conversion rates, and retention rate.
4. Dashboard Development: The results were visualized in a dashboard to provide an at-a-glance understanding of website performance, featuring charts for hourly sessions, conversions by device, and traffic source analysis.

Tools Used: Python (for data generation), MySQL (for data storage, cleaning, and analysis), and Power BI for dashboarding.

4. Key Performance Indicators (KPIs) & Analysis

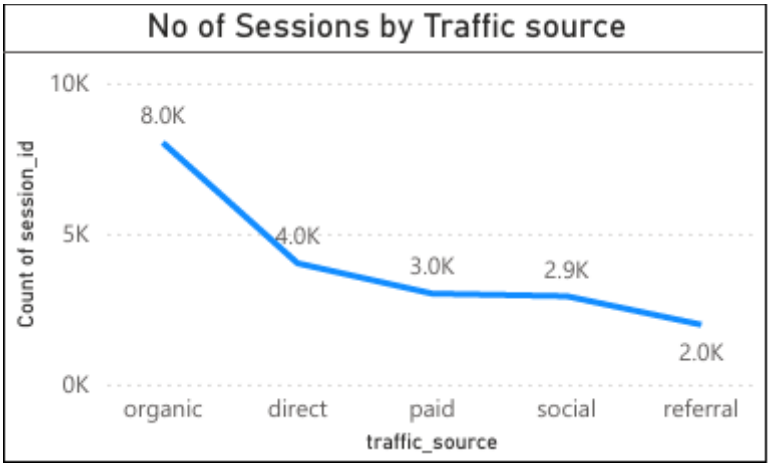
The analysis yielded the following top-level website performance metrics:

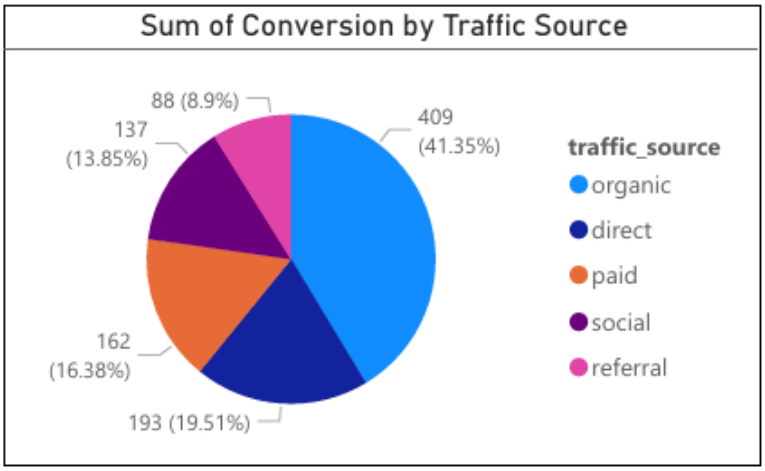
- Overall Sessions: The website registered a total of 20,000 user sessions.
- Average Session Duration: The average user engagement per session was 299.93 seconds (approx. 5 minutes).
- Overall Bounce Rate: The site-wide bounce rate stands at 40.40%, indicating the percentage of single-page sessions.
- User Retention Rate: The analysis shows a retention rate of 92.47%, measuring the percentage of unique users who returned for a second session on a different day.

5. Detailed Insights and Recommendations

Insight 1: Organic Search is the Most Valuable Traffic Channel

The 'organic' traffic source is the top performer, driving the highest number of sessions (8,000) and also contributing the largest share of conversions (41.35%).

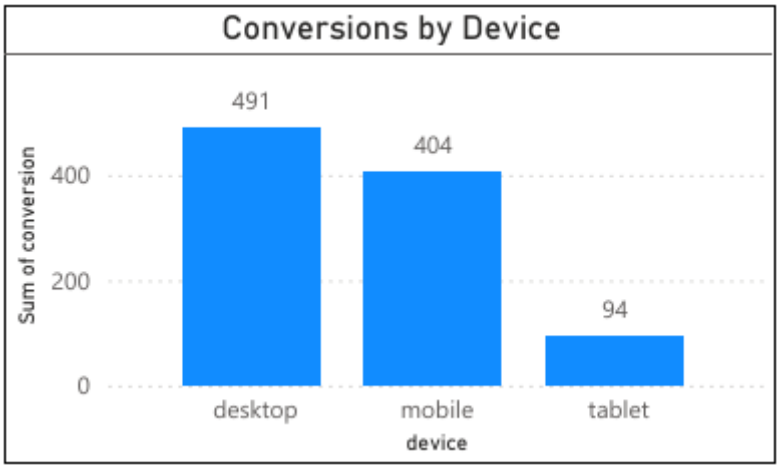




- Recommendation: Marketing efforts and budget should be prioritized towards Search Engine Optimization (SEO) to maintain and grow this high-performing channel.

Insight 2: Desktop Users Drive the Majority of Conversions

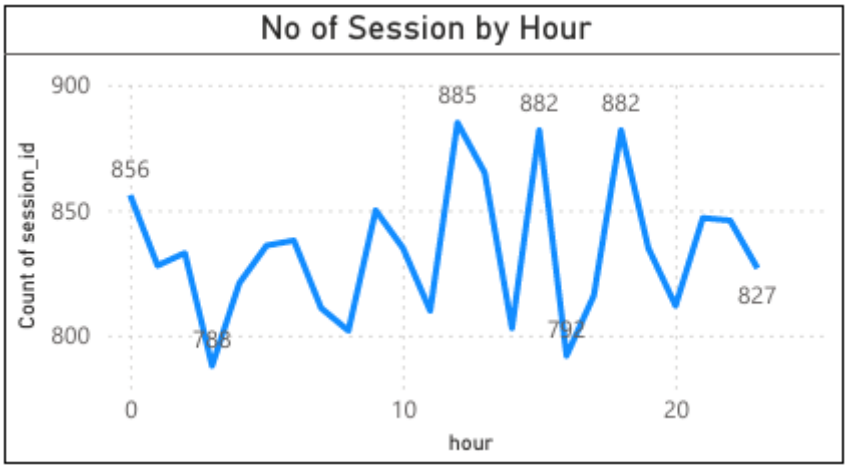
The dashboard clearly shows that desktop users convert at a much higher rate than users on other devices. Desktops accounted for 491 conversions, compared to 404 for mobile and just 94 for tablets.



- Recommendation: While desktop performance is strong, the mobile user experience should be critically reviewed. An audit of the mobile site's user interface, page speed, and checkout process could uncover friction points and help close the conversion gap.

Insight 3: User Activity Peaks at Specific Times of Day

Session data reveals distinct traffic peaks during the day, notably around 12:00, 15:00, and 18:00.

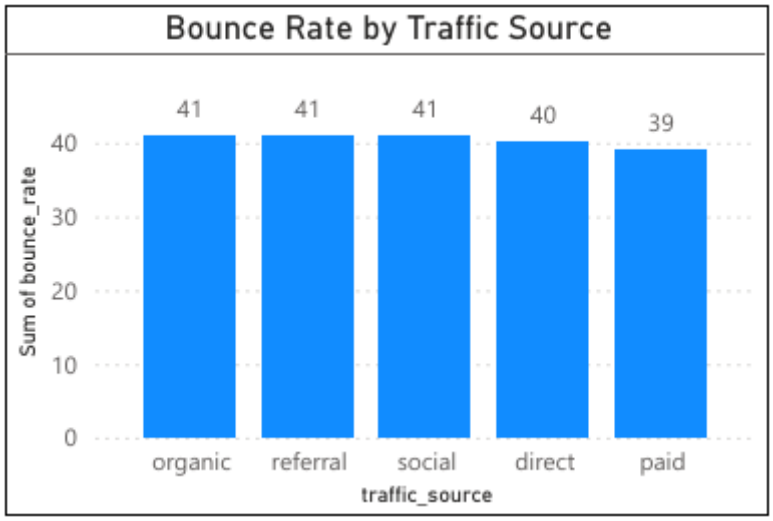


- Recommendation: Schedule marketing campaigns, social media posts, and email newsletters to coincide with these peak traffic hours to maximize reach and engagement.

Insight 4: Bounce Rates Are Consistent Across Channels

The bounce rate is relatively uniform across all major traffic sources, hovering between

39% and 41%. This suggests that no single channel is bringing in particularly low-quality traffic.



- Recommendation: Efforts to reduce the bounce rate should focus on universal improvements, such as optimizing landing page design, improving site speed, and ensuring clear calls-to-action for all incoming users.

6. Limitations of the Analysis

- Synthetic Dataset: The primary limitation is that this analysis is based on a synthetically generated dataset. The trends and figures, while useful for this demonstration, may not perfectly mirror real-world user behaviour.
- Data Imbalance: The dataset was created with known imbalances. For example, the calculated retention rate of 92.47% is exceptionally high and is likely an artifact of the data simulation process rather than a reflection of typical website performance.

7. Conclusion

This project successfully demonstrates the power of a data-driven approach to understanding user behaviour. By leveraging MySQL for robust analysis and a visual dashboard for clear communication, we have identified critical insights into traffic sources, device performance, and user engagement patterns. The recommendations provided offer a clear path toward optimizing marketing strategies and improving overall website effectiveness.