**Web Analytics Report for**[www.datadb.com](http://www.datadb.com/)

**1. Executive Summary**

This analysis explores web activity data from [www.datadb.com](http://www.datadb.com/) to understand user engagement and identify factors influencing key website metrics: unique page views, exits, time on page, and bounces. Using R for exploratory data analysis and regression modelling, we find that visits are the primary driver of unique page views and exits, while bounce rate strongly reduces time spent on site. However, most variation in time on page remains unexplained by the available variables. Recommendations are provided to improve user retention and engagement.

**2. Data Overview**

* **Observations:** 32,109 sessions
* **Variables:** 8 (Bounces, Exits, Continent, Source, Timeinpage, Uniquepageviews, Visits, BouncesNew)
* **No missing values**
* **Distribution:** Most sessions have low engagement (e.g., 0–1 visits/pageviews), with a small number of high outliers.
* **Geography:** Majority of users are from North America, by Europe and Asia.
* **Traffic sources:** Main sources are Google, direct, Others, and tableausoftware.com.

**3. Exploratory Data Analysis**

**Univariate Findings**

* **Time on Page:** Highly right-skewed; most sessions are short, but some last hours.
* **Bounces/Exits:** Most sessions have 0 or 1, but rare cases go much higher.
* **BouncesNew:** About 36% of sessions have zero bounces; non-zero values are mostly small.

**Categorical Insights**

* **Continent:** North America dominates traffic.
* **Sourcegroup:** Google and direct traffic are largest.

**Bivariate Insights**

* **Visits vs Unique Page Views:** Strong positive linear relationship.
* **Boxplot:** Time on Page varies little by source group, with all groups showing many low-time outliers and a few extreme highs.

**4 Modelling & Results**

**A. Unique Page Views Dependence on Visits**

* **Model:** Uniquepageviews ~ Visits
* **R² = 0.66** (strong fit)
* **Interpretation:** Each additional visit increases unique page views by ~0.69. This is highly significant (p < 0.001).
* \*\*Plot Shows a clear upward trend.

**B. Factors Affecting Exits**

* **Model:** Exits ~ . ( Bounces, Uniquepageviews)
* **R² = 0.73** (very strong fit)
* **Key predictors:**
* **Visits**): More visits, more exits.
* **BouncesNew** (+): Higher bounce rate, much higher exits.
* **Timeinpage** (+): Longer time, slightly more exits.
* **Sourcegroup:** Sessions from reddit.com, t.co, Others, and visualisingdata.com have fewer exits than the reference group.
* **Continent:** Not significant after controlling for other variables.

**C. Factors Affecting Time on Page**

* **Model:** Timeinpage ~ . (excluding Bounces, Exits)
* **R² = 0.10** (weak fit)
* **Key predictors:**
* **Uniquepageviews** (+): More unique page views, more time.
* **Visits** (+): More visits, more time.
* **BouncesNew** (–): Higher bounce rate, much less time.
* **South America:** Sessions from SA spend more time on page.
* **Other continents/sourcegroups:** Not significant.
* \*\*Interpretation Most variation in time on page is not explained by these variables.

**D. Factors Impacting Bounces**

* **Model:** `Bounces ~ . (excluding Exits, Uniquepageviews)
* **R² ≈ 1.00** (perfect fit, likely due to mathematical redundancy withouncesNew)
* **Note:** Model is overfit; suggests need for alternative modelling (e.g., logistic regression or removing redundant variables).

**5. Diagnostics & Limitations**

* **Residuals:** Diagnostic plots show heavy clustering at zero, with outliers and non-normality. Some heteroscedasticity is present.
* **Perfect fit in Bounces model:** Indicates multicollinearity/redundancy; interpret with caution.
* **Skewed distributions:** Many variables are zero-inflated or right-skewed, limiting the effectiveness of standard linear models.
* **Unexplained variance:** Especially for time on page, suggesting important predictors (like content type, device, or session duration) are missing.

**6. Recommendations**

* **Focus on increasing visits:** Since visits drive both unique page views and exits, strategies to encourage repeat visits may boost engagement.
* **Reduce bounce rate:** High bounce rates dramatically reduce time on page and increase exits. landing page relevance and usability could help.
* **Targeted interventions:** Consider further segmentation by source group and continent for marketing or UX improvements.
* **Data collection:** Capture additional features (content type, device, entry/exit pages) to better explain time on page and user behaviour.
* **Model refinement:** For future work, consider robust regression, GLMs, or machine learning methods to handle skewed and zero-inflated data.