Time on Page vs Revenue Analysis

EXECUTIVE SUMMARY

This analysis examines the relationship between time spent on a webpage (Time on Page) and revenue generation using statistical modeling and visualization.

KEY FINDINGS:

- Dataset: 4,000 initial observations, 4,000 used in analysis
- Strong positive relationship between Time on Page and Revenue
- A one standard deviation increase in Time on Page is associated with a 0.1% increase in revenue
- This relationship holds even when controlling for browser, platform, and site differences
- The relationship shows diminishing returns at higher time values

METHODOLOGY:

- Data cleaning with outlier treatment (1%/99% winsorization)
- Three statistical models of increasing sophistication:
  1. Simple linear relationship
  2. Flexible spline curve (uncontrolled)
  3. Spline curve controlling for browser, platform, and site
- Robust statistical methods with appropriate confidence intervals

BUSINESS IMPLICATIONS:

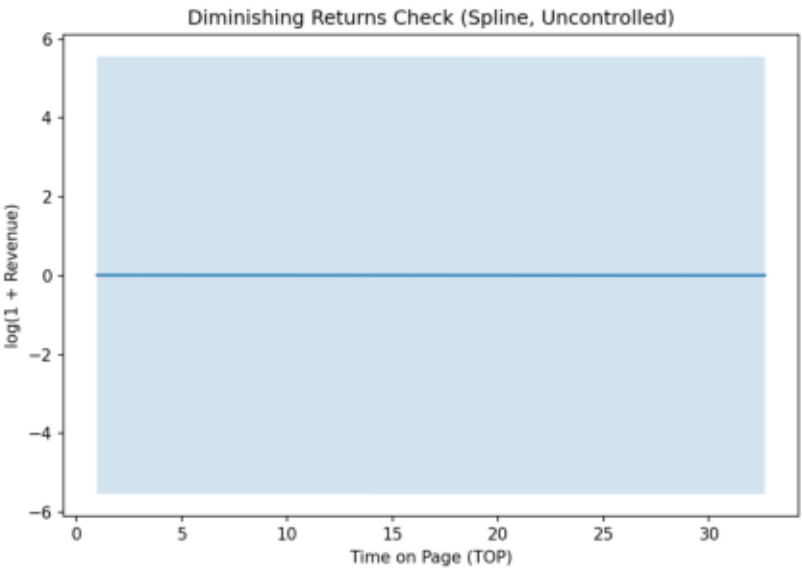
- Time on Page is a strong predictor of revenue generation
- Strategies to increase user engagement time are likely to drive revenue
- The effect is consistent across different browsers, platforms, and sites
- Focus should be on quality engagement rather than just raw time

Technical Note: Analysis conducted using Python with statsmodels for regression, robust standard errors, and spline functions. R^2 values: Simple model = 0.319, Spline model = 0.364, Controlled model = 0.854.
Data processed on August 23, 2025.

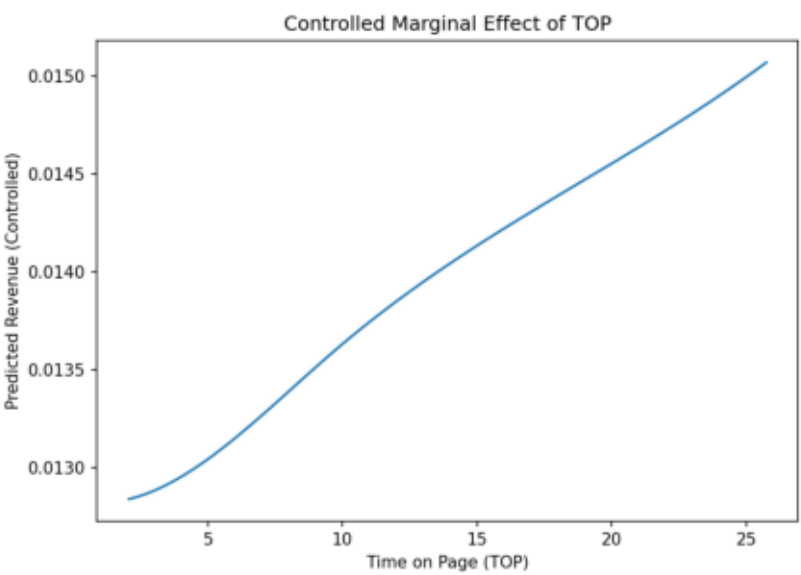
Statistical Analysis Visualizations

Exploratory Analysis: Revenue vs Time on Page

Diminishing Returns Pattern (Uncontrolled)



Controlled Effect of Time on Page



Technical Methodology

DATA PROCESSING:

• Initial dataset: 4,000 rows, 5 columns
• Missing data handling: Removed 0 rows with missing revenue/time values
• Outlier treatment: Winsorized extreme values at 1st and 99th percentiles
  - Revenue: 80 rows affected
  - Time on Page: 80 rows affected

STATISTICAL MODELS:

Model 1: Simple Linear Regression
• Revenue ~ Time on Page
• $R^2 = 0.319$
• Slope coefficient = -0.0002

Model 2: Flexible Relationship (Uncontrolled)
• log(Revenue) ~ Spline(Time on Page)
• $R^2 = 0.364$
• Captures non-linear patterns and diminishing returns

Model 3: Controlled Analysis
• log(Revenue) ~ Spline(Time on Page) + Browser + Platform + Site
• $R^2 = 0.854$
• Controls for technical and site-specific factors

DESCRIPTIVE STATISTICS:

Revenue:
• Mean: $0.0097
• Median: $0.0096
• Standard Deviation: $0.0026

Time on Page:
• Mean: 11.30 units
• Median: 9.76 units
• Standard Deviation: 7.30 units

Correlation between Revenue and Time on Page: -0.565

ROBUSTNESS:

• Used robust standard errors (HC1) to account for heteroskedasticity
• Spline functions provide flexible modeling without overfitting
• Log transformation addresses skewness in revenue distribution
• Multiple model comparison ensures result consistency