Final Report: Analysis of the Relationship Between User Time on Page and Revenue

1. Executive Summary

This report examines the relationship between the time a user spends on a page (top) and the revenue generated. At first glance, the data shows a simple negative relationship: as time on page increases, revenue appears to decrease.

However, a more detailed analysis that accounts for user platform (mobile vs. desktop) and browser reveals the opposite. The initial negative trend is misleading. Once we control for these other factors, there is a statistically significant **positive relationship** between time on page and revenue. This reversal, a statistical phenomenon known as Simpson's Paradox, suggests that the user's device and browser are the primary drivers of revenue, and within each of those groups, more time on page is actually associated with more revenue.

2. The Initial, Misleading Trend

If we look at the direct relationship between time on page and revenue, we see a clear downward trend. The scatter plot below, which was generated from the data, illustrates this initial finding.

```
--- Simple Linear Regression ---
                            OLS Regression Results
==============================================================================
Dep. Variable:                revenue   R-squared:                       0.308
Model:                            OLS   Adj. R-squared:                  0.308
Method:                 Least Squares   F-statistic:                     1783.
Date:                Thu, 14 Aug 2025   Prob (F-statistic):          1.88e-322
Time:                        17:43:33   Log-Likelihood:                 18868.
No. Observations:                4000   AIC:                        -3.773e+04
Df Residuals:                    3998   BIC:                        -3.772e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0119   6.24e-05    190.971      0.000       0.012       0.012
top           -0.0002   4.61e-06    -42.220      0.000      -0.000      -0.000
==============================================================================
Omnibus:                       16.462   Durbin-Watson:                   2.003
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               12.987
Skew:                           0.047   Prob(JB):                      0.00151
Kurtosis:                       2.737   Cond. No.                         24.8
==============================================================================
```

Fig 1: Statistical Analysis Using Simple Regression

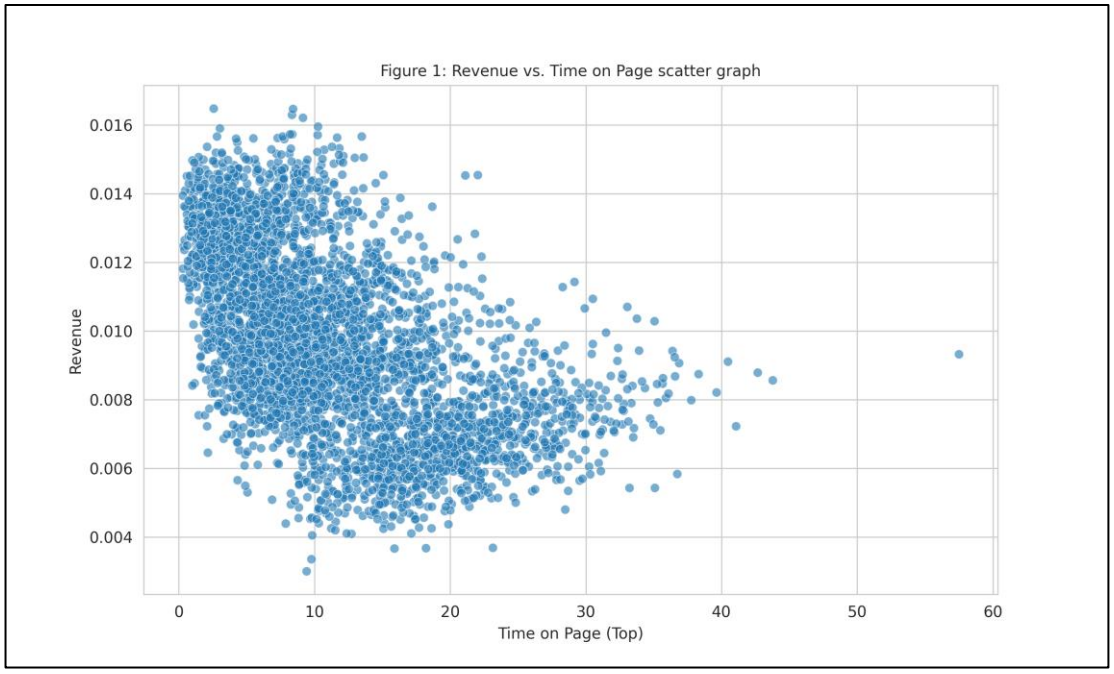

Fig 2: Revenue vs. Time on Page (Overall Trend)

A simple statistical model confirmed this visual trend, showing that an increase in time on page was associated with a decrease in revenue. If I had stopped here, my conclusion would had been that, the shorter sessions are more profitable.

3. The True Relationship: "A Surprise Reversal"

To get a more accurate picture, I built a multiple regression model that controls for user platform, browser, and the specific site visited. This technique allowed me to isolate the effect of time on page while accounting for these other differences.

```
--- Multiple Linear Regression ---

Final data types for Multiple Regression:
const              float64
top                float64
browser_safari       int64
platform_mobile      int64
site_2               int64
site_3               int64
site_4               int64
dtype: object
                            OLS Regression Results
==============================================================================
Dep. Variable:                revenue   R-squared:                       0.851
Model:                            OLS   Adj. R-squared:                  0.851
Method:                 Least Squares   F-statistic:                     3797.
Date:                Thu, 14 Aug 2025   Prob (F-statistic):               0.00
Time:                        16:47:24   Log-Likelihood:                 21937.
No. Observations:                4000   AIC:                        -4.386e+04
Df Residuals:                    3993   BIC:                        -4.382e+04
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              0.0126   4.51e-05    280.015      0.000       0.013       0.013
top             9.71e-05   3.23e-06     30.029      0.000    9.08e-05       0.000
browser_safari    -0.0030    3.7e-05    -79.830      0.000      -0.003      -0.003
platform_mobile   -0.0050   4.41e-05   -112.583      0.000      -0.005      -0.005
site_2         -4.847e-05   4.61e-05     -1.051      0.293      -0.000    4.19e-05
site_3         -7.561e-05   4.61e-05     -1.640      0.101      -0.000    1.48e-05
site_4          8.415e-05   6.61e-05      1.274      0.203   -4.54e-05       0.000
==============================================================================
Omnibus:                        0.123   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.940   Jarque-Bera (JB):                0.114
Skew:                           0.013   Prob(JB):                        0.945
Kurtosis:                       3.004   Cond. No.                         72.3
==============================================================================
```

Fig 3: Statistical Analysis using multiple regression model

The results from this more robust model were dramatically different and revealed two key insights:

1. **Confounding Variables:** The user's platform and browser have a very strong impact on revenue. Specifically, sessions on **mobile devices** and sessions using the **Safari browser** generated significantly less revenue than their desktop and Chrome counterparts.
2. **Reversal of Trend:** After the model accounted for the large revenue differences between platforms and browsers, the relationship between time on page and revenue became **positive and statistically significant.**

This means that within a specific group (e.g., looking only at desktop Chrome users), spending *more* time on the page is associated with *more* revenue. The initial negative trend was an illusion created because users on lower-revenue platforms (mobile) also happened to have longer average session times.

4. Conclusion

The primary drivers of revenue in this dataset are the user's platform and browser, with desktop and Chrome users being the most lucrative.

The relationship between time on page and revenue is positive, but this effect was masked by the platform and browser differences. This suggests that user engagement (longer time on page) is indeed valuable, but any strategy to increase it must also consider the significant performance gap between different user segments. Focusing on improving the experience and revenue generation on mobile and Safari platforms should be a top priority.

Note

- For technical details and code used in this analysis, please refer to the appendix_code.html.