

Relationship Between Time on Page (top) and Revenue

Reading CSV from: testdata.csv

1) Data cleaning + searching (profiling)

Checks missing values, data types, duplicates, and ensures `top` and `revenue` are numeric.

Preview

	revenue	top	browser	platform	site
0	0.0078	30.9603	safari	mobile	3
1	0.0142	9.4006	chrome	desktop	3
2	0.0123	14.1199	safari	desktop	3
3	0.0107	29.9071	chrome	mobile	3
4	0.0095	19.3632	chrome	mobile	3
5	0.0062	20.6915	safari	mobile	4
6	0.0083	9.601	chrome	mobile	4
7	0.0063	9.765	chrome	mobile	2
8	0.0088	10.7436	chrome	mobile	2
9	0.0123	9.1016	safari	desktop	3

Column profile

	column	dtype	missing_r
0	revenue	float64	
1	top	float64	
2	browser	object	
3	platform	object	
4	site	int64	

Duplicate rows: 0

Rows before cleaning: 4,000

Rows after cleaning: 4,000

Detected control variables (all other columns besides `top` and `revenue`):

```
browser, platform, site
```

2) EDA (visual + statistical)

We visualize the raw relationship and compute correlation as a quick signal check.

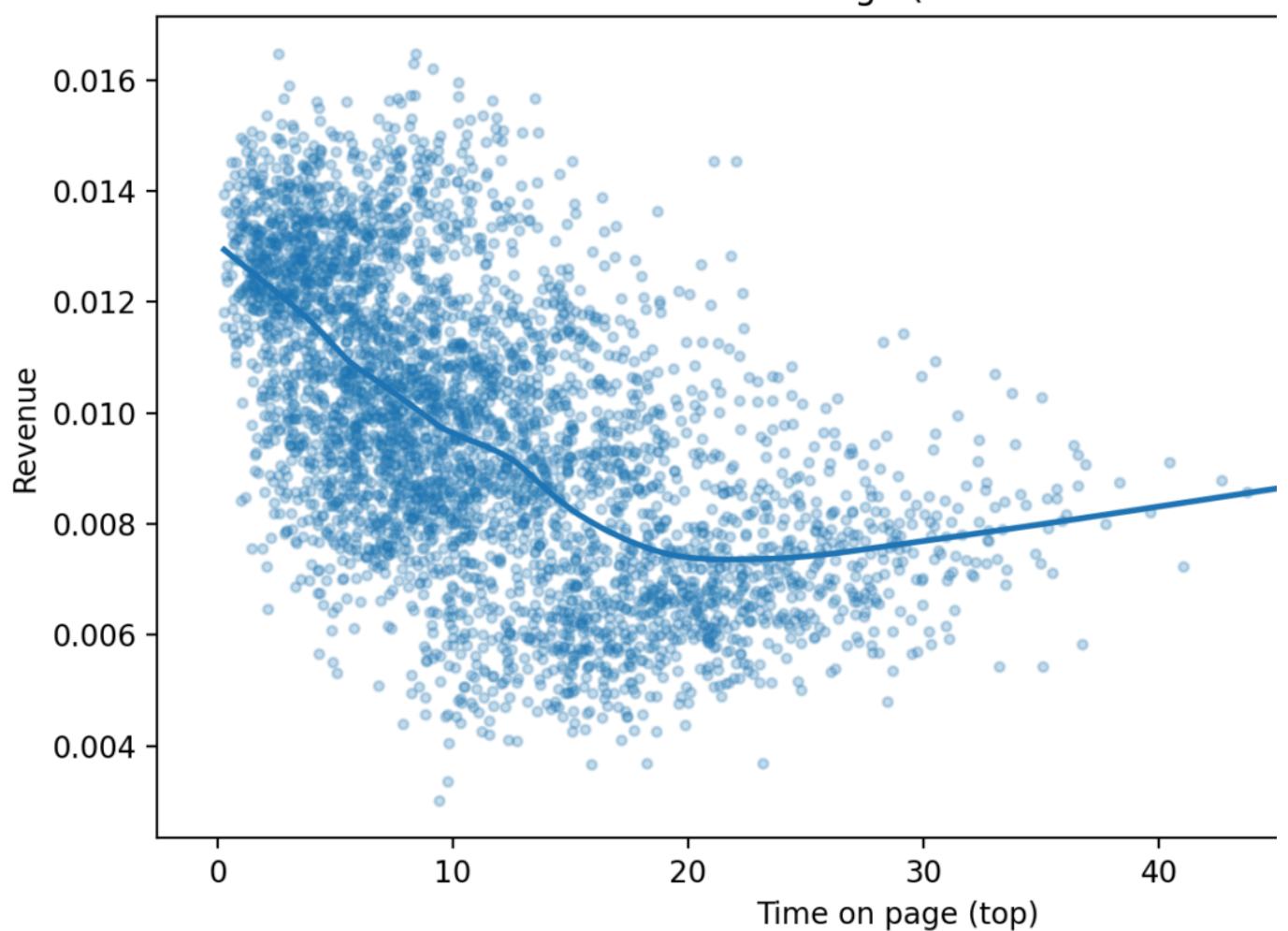
Pearson correlation (linear)

-0.555

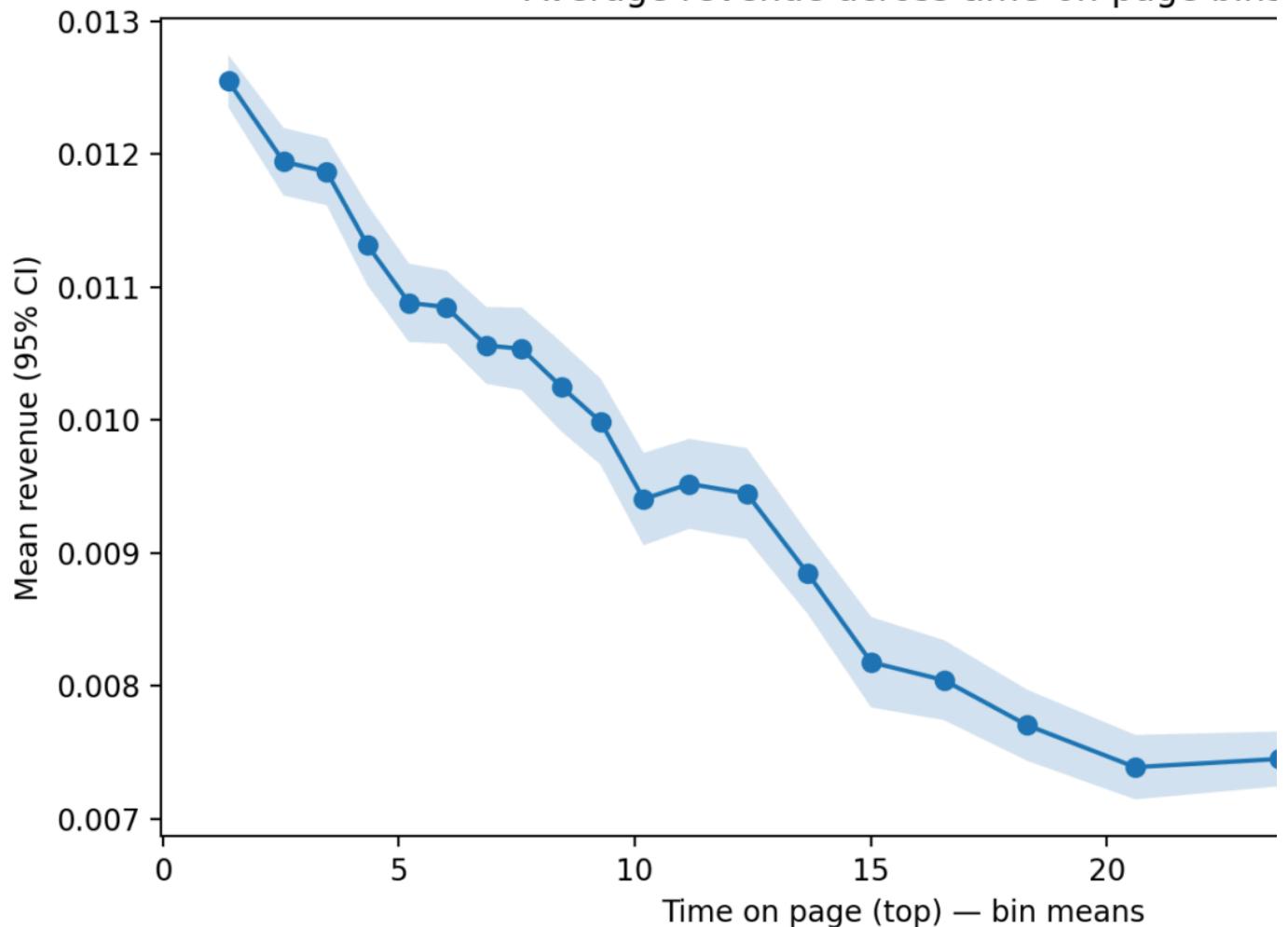
Spearman rho (rank-based)

-0.608

Revenue vs Time on Page (scatter + smooth trend)



Average revenue across time-on-page bins



3) Modeling (simple vs controlled)

We fit two models:

- **Simple:** `revenue ~ top`
- **Controlled:** `revenue ~ top + (all other variables)`

Then we compare how the `top` effect changes.

Simple model

Association between time on page and revenue without controlling for other variables.

Effect of +10 top units on revenue

-0.00195

95% CI: [-0.00204, -0.00185]

R²: 0.308

Does the relationship change after controlling?

- Change in +10-unit effect: 0.00292 (controlled minus simple). If this shifts a lot, it suggests the raw relationship was partly driven by differences across other variables.

Below is the **partial relationship** after removing the influence of control variables (residualized view).

Controlled model

Association after controlling for other variables (category)

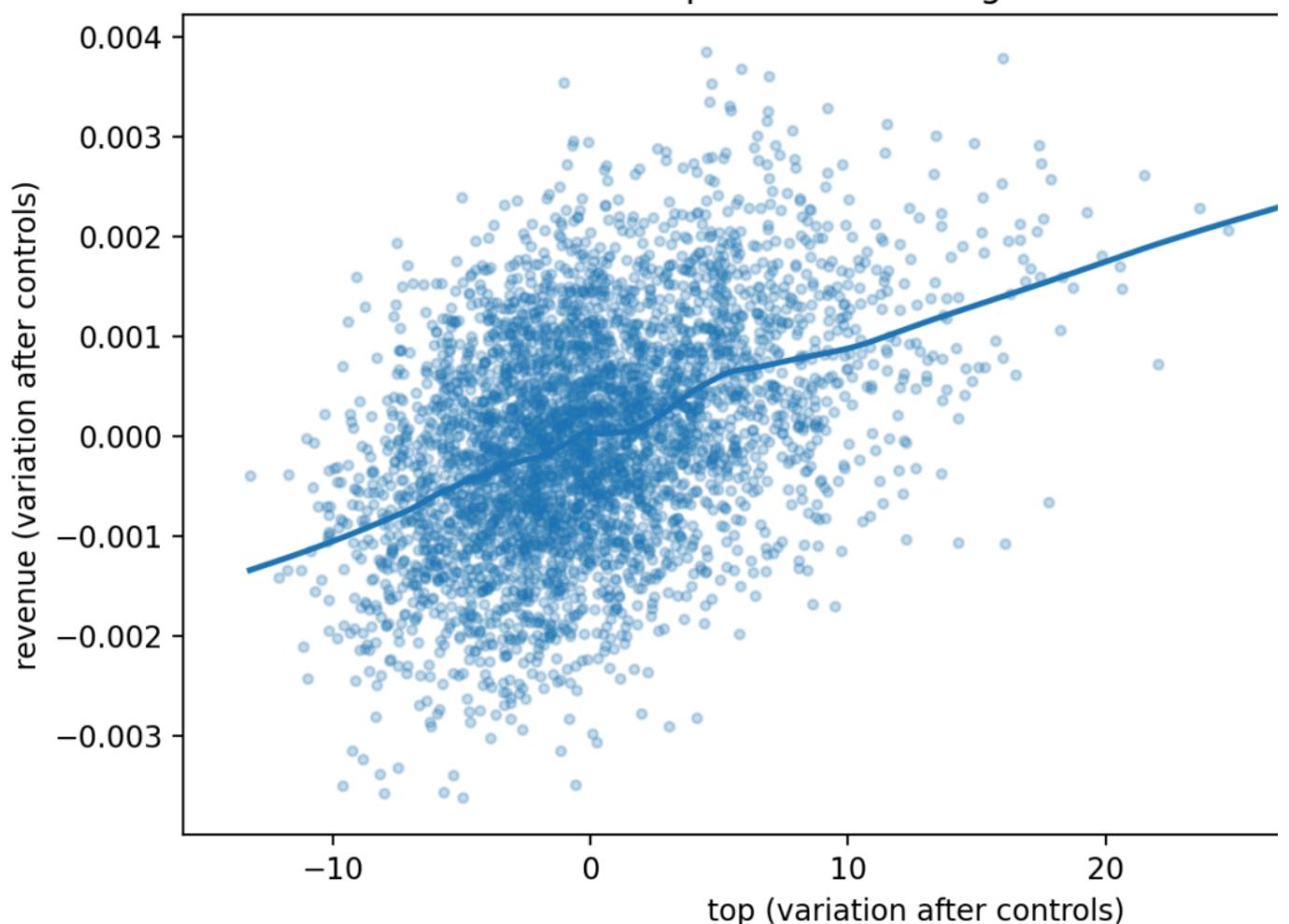
Effect of +10 top units on revenue

0.00097

95% CI: [0.00091, 0.00103]

R²: 0.851

Relationship after controlling for other variables



Optional: show model formulas and detailed stats output (technical)

4) Download a brief PDF report

This PDF is meant to be readable to mixed audiences and avoids dumping raw statistical tables.

[Download PDF report](#)