https://bit.ly/PerfABPattern

Trustworthy A/B Patterns – Page Load Time (Performance) Pattern

Ronny Kohavi, Jakub Linowski, and Lukas Vermeer
Last update: November 19, 2025

This is a summary of the Page Load Time (Performance) pattern, part of the Trustworthy A/B Patterns project (https://bit.ly/trustworthyABPatterns).

We have an experiment on performance reported by Talabat, a food delivery service owned by Delivery Hero, which uses Eppo to run experiments. We expect to have more replications in the future and add them to this document.

Website performance matters. In Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing (Kohavi, Tang, & Xu, 2020), the Speed Matters Chapter provides this motivation: An engineer that improves server performance by 10 msec (that's 1/30 of the speed that our eyes blink) more than pays for his (or her) fully-loaded annual costs. Every millisecond counts. At Bing, a 100 msec speedup improved revenue by 0.6%. At Amazon, 100msec improved sales by 1%. But these were speedups for every page, and we want to get more data in other settings where a single page, like the home page, improves.

As part of the Trustworthy A/B Testing project (https://bit.ly/trustworthyABPatterns), we suggested replicating this pattern.

Talabat Experiment – Speeding the Home Page

TL; DR: an experiment was run on Talabat's home page using Eppo as the experimentation platform with CUPED enabled, measuring the revenue impact of speedup to the home page achieved via smart caching. The speed up for time-to-interactive metric improved by about 58%, and the estimated impact to revenue was about 0.36%.

As with many first experiments, the first run had an SRM (Sample Ratio Mismatch). The cause was identified, so the results below are from the second run, which was balanced (no SRM). The second run used Eppo's sequential inference, which tends to exaggerate the estimate, but otherwise is trustworthy and an interesting result to share.

https://bit.ly/PerfABPattern

About Talabat

Talabat is the Middle East & North Africa's largest food-ordering and quick-commerce platform, operating across eight markets with around 450 employees in product and tech. IPO in November 2024 valued the company at $24B.

Talabat Experiment

The experiment had four variants with a control, the treatment of interest, and two treatments with partial performance changes to understand what is contributing to the speedup. For replication purposes, we are not interested in these other treatments and focus just on the treatment of interest. The speedup in Time-To-Interactive was from ~2.1 seconds to ~1.0 seconds.

Metric	Treatment effect	95% confidence	P-value (CUPED)
Time-to-Interactive (avg)	-58.05%	[-58.39%, -57.72%]	<0.001
Time-to-Interactive (p95)	- 28.95%	[-29.19%, -28.71%]	<0.001
Revenue/user (GMV)	+0.36%	[+0.02%, 0.70%]	0.037
Orders/user	+0.34%	[+0.03%, 0.66%]	0.031

Commentary:

1. The allocation was balanced with a split of around 1.1M users per variant, corresponding to a chi-squared p-value of 0.97.
2. The experiment was using sequential inference and stopped upon reaching significance. These techniques control type-I error, but do not correct the expected exaggeration of the treatment effect.
3. The total number of users participating in the experiment was around 2.2M, which provides 80% power for an MDE of 0.5 and about 20% power for an MDE of 0.2 (our recommended MDE for a single page performance improvement).

Summary

This conceptual replication experiment is interesting to share with the community. The dramatic improvement to performance (Time-To-Interactive) for the home page is

impressive, and the evidence shows material improvement to revenue and orders for the business.