Conversion Benchmark Report Analysis Methodology

This methodology was prepared by the data team at Unbounce and lays out the research process in well-considered detail. If you’re looking for a simpler guide to understanding how to apply our findings to your marketing campaigns, read the section called “How We Created This Report.”

1. Industry Classification

Using the machine-generated topic mixture, we began the manual (i.e., not machine) process of assigning topics to industries. We started with the set of machine-generated topics, the Global Industry Classification Standard, and the North American Industry Classification System, and began to map a relationship between topics and industries.

Our goal was to assign each topic to a single industry and to generate a final list that described the landing pages in our dataset. The priority was to define industries that fit the data, rather than imposing an existing classification system as is. We found we needed more granular classification to capture the conversion rate variation within the industry, and so we created a second layer of hierarchy using subcategories. Since each page in our dataset received a topic mixture, it was also assigned a normalized industry and subcategory mixture; thus each page could be grouped under multiple topics, industries, and subcategories.

2. Industry Inclusion

Our analysis showed that the majority of pages in the dataset were assigned a mixture of 1 to 4 industries. When we included pages whose industry mixture was at least 16%, the majority of pages were assigned 2 industries. Given our set of industries and pages, we decided it was an appropriate description of the data to allow a page to be split between 2 industries. We also decided to only include pages with industries or subcategories that had at least 16% inclusion in the given mixture element.
3. Topic Exclusion

We found that some topics had a much higher conversion rate than others, and resulted in a conversion rate distribution that wasn’t representative of the whole. Some of these topics were manually removed from the analysis to ensure an accurate representation of the dataset. The excluded topics were:

- Politics
- Religion
- Rewards and Affiliate Programs
- Sales Vernacular
- Terms and Conditions
- Testimonials
4. Sample Size

We haven’t included analysis on any data subsets with fewer than 400 pages. We explored a number of options to decide how large a subset should be to ensure statistically significant results.

The exploration included an examination of the standard error on the distribution of conversion rate, a Chi-Square test to check the statistical significance of categorical factors on conversion rate, and a two-sample Kolgomorov-Smirnov test to compare distributions of subsets with varying sample sizes. We chose a conservative value of 400 to ensure we had enough data in each analysis to confidently report on the results.

5. Conversion Type

For our analysis of conversion types, we defined CTA types based on conversion goals set in the Unbounce builder. There were five possible types:

- **Click**: all conversion goals on the page are set up as click conversions
- **Form**: all conversion goals on the page are set up as form conversions
6. Mean vs Median

Conversion rates, by nature, have a wide range of values depending on the amount of traffic the page naturally receives. We found that the **mean** conversion rate was much higher than the **median**, due to a skew from a handful of very high converting pages.

To account for both outliers and skew, we used median across the majority of the report. We believe that median better represents conversion rate benchmarks because it describes where a typical page will fall within subsets of the data.
7. Regression Analysis

We used a linear regression model to determine the extent to which each factor in the analysis (conversion type, number of words, and reading ease) affected conversion rate. For these factors, we used a nonparametric lowess model (locally weighted linear regression) to approximate the trend line, as we felt it described the data better than a linear model.

We can view the relationship between conversion rate and these factors using the overall trend produced by the regression. For readability, we further bucketed the data points into 10 groups, visualizing the median of each group, to show the concentration of data within the subset.

![Nonparametric regression model for conversion rates by reading ease (real estate)](image)

8. Sentiment Analysis

We expressed sentiment on a page by computing the number of words that the EmoLex algorithm identified by the total number of stemmed words on the page. In this way, each page was assigned a percentage for each sentiment.

We investigated the relationship between sentiment and conversion rate using regression analysis, as with the other factors on a page. While it may have seemed natural and intuitive that sentiment could be a good predictor of conversion rate, we didn’t find this to be generally true in our research and analysis.
Using several rigorous methods, we were unable to show that sentiment causes conversion rates to rise or fall. However, we know that all options for analysis haven’t been explored, so we can’t say for certain that a relationship doesn’t exist in some capacity.

The data we saw in our analysis is generally noisy and not clearly patterned with a clear correlation. While some pages may convert better, and show an increased percentage of sentiment, this potentially is random, and shouldn’t be relied upon to alter conversion rates in a predictable way.

The visualization of sentiment is based on a linear regression with a small number of averaged data points. We choose this for visual clarity, and to illustrate the limited predictability of sentiment on conversion rates.

**Additional visualizations of sentiment analysis**

1. SaaS anticipation
2. Ecomm anticipation

3. Ecomm joy

4. Ecomm negative
5. Business Services anticipation

6. Legal anger

7. Legal fear
8. Real Estate joy

9. Real Estate positive

10. Real Estate trust