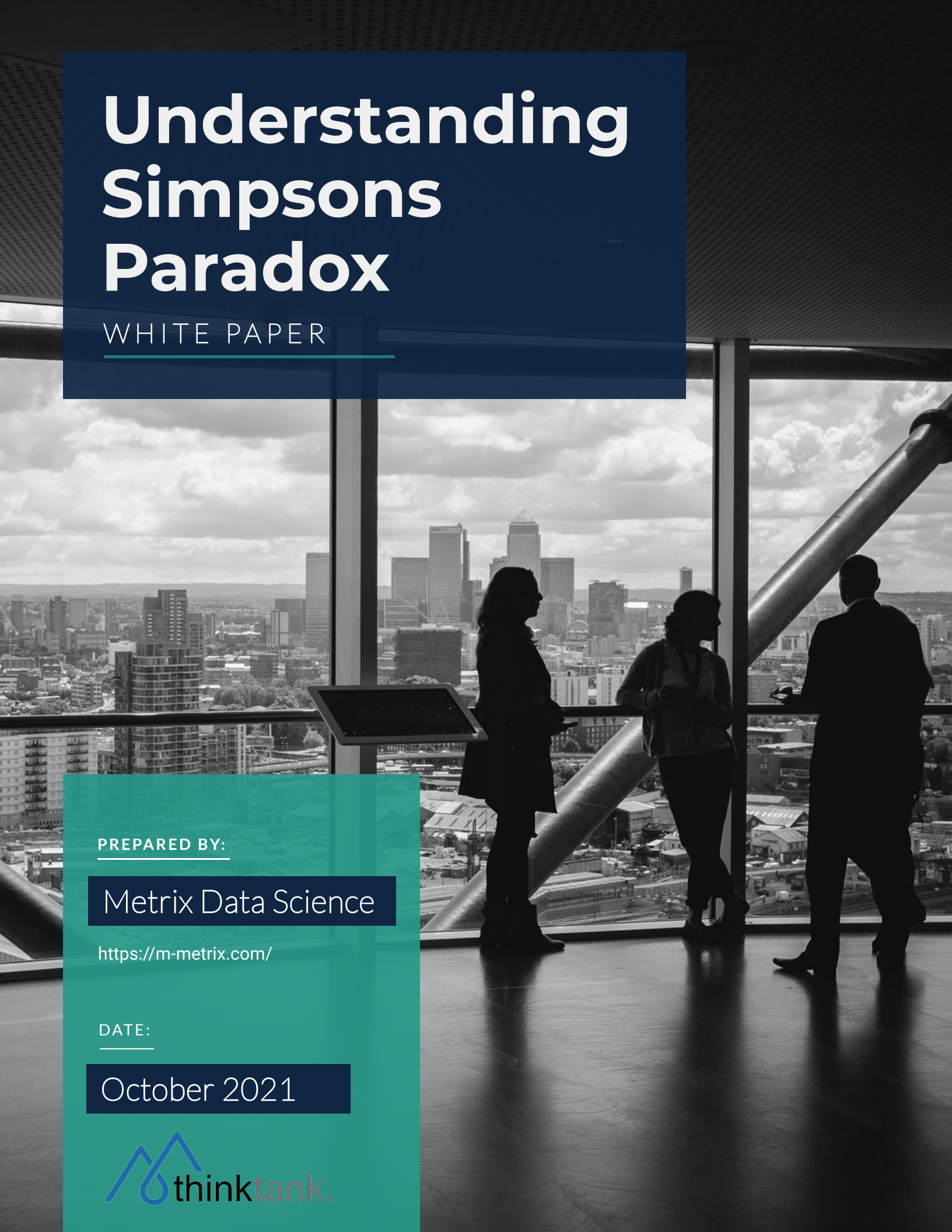
Understanding Simpsons Paradox

WHITE PAPER



PREPARED BY:

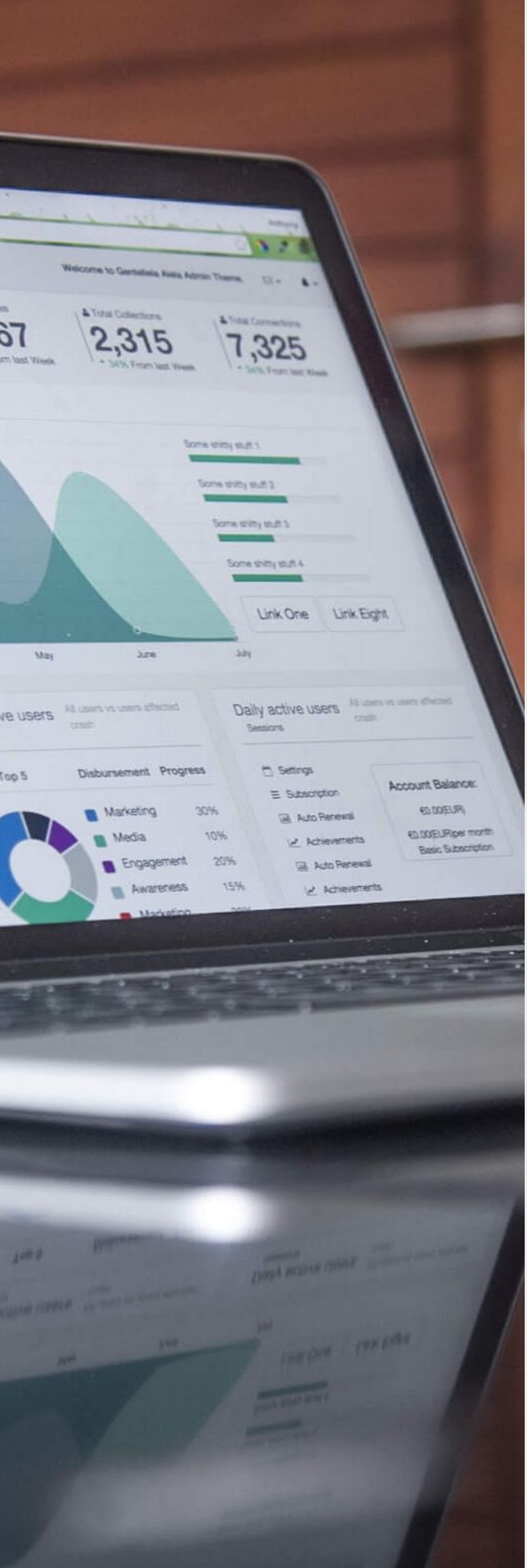
Metrix Data Science

https://m-metrix.com/

DATE:

October 2021

thinktank.



Summary

Even the most basic statistical reporting can cause controversy between internal departments in organisations, or between a brand and their advertising agency. A common problem that arises when reporting on marketing campaigns is that at first sight, a campaign may be performing well... however further analysis into the breakdown of the campaign reveals this is sometimes not the case.

This brings us to the tricky Simpson's Paradox, an effect that occurs when the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables. A mouthful, right? Let us help break it down for you...

A bit of context

Recently a client approached us to settle a debate with their media agency. They had lately teamed up for a mailing campaign showcasing their new product. In this campaign, the client wanted to stick to what they knew by sending bulk mail to their 'banker' list, using the tried and tested control. On the other hand, the agency pushed testing new creatives and several new lists of data with fresh prospects.

Unfortunately, neither had data scientists with statistical expertise within their teams to guide them through the test matrix.The results:On seeing the overall results, with the percentage of positive responses from the control being higher than the agencies test, the client concluded their control was the winner. See below:

The results

On seeing the overall results, with the percentage of positive responses from the control being higher than the agencies test, the client concluded their control was the winner. See below:

102542	Control List (no)	Test List (no)	Positive Responses (Control)	Positive Responses (Test)	% Success (Control)	% Success (Test)
Overall	100,000	20,000	4,500	800	279989	4%

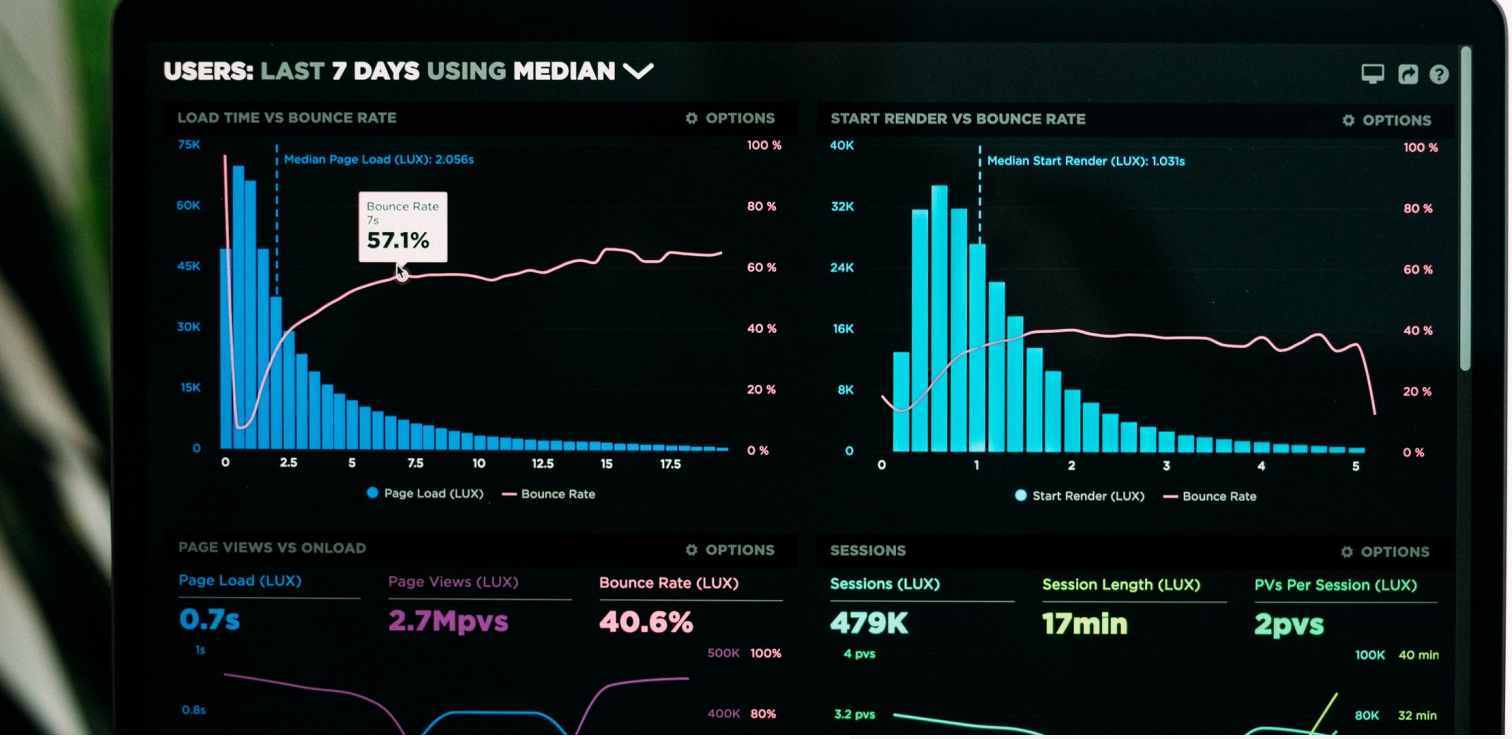
However, the agency analysed the results by list and concluded that the new creative was the winner. This is illustrated below:

Positive Responses (Control)	Control List Vol	Test List Vol	Responders (Control)	Responders (Test)	Control	Test
Aggregate	100,000	20,000	4,500	800	4.5%	4%
List 1	75,000	1,400	3,700	80	4.9%	5.7%
List 2	10,000	1,600	500	100	5%	6.3%
List 3	15,000	17,000	300	620	2%	3.6%

So, what's happening?

Do the Maths and you will see they are both right. Looking at the extended breakdown level of reporting, the test creative is the winner on every occasion, while looking at the aggregate results the control is the winner. This seemingly peculiar result is called Simpson's Paradox.As experienced Data Scientists, we often see this confusion when a scientific approach to sampling is not taken.

The Simpson's Paradox is caused by weighted averages. The list volume is, in this case, called a "lurking" variable or confounding variable. There is an uneven distribution across the three list volumes. In turn, the responders are also askew, hence causing the bizarre results we are seeing from the campaign reporting.In these situations – it's like comparing chalk and cheese. Only when we are sure the control and test groups are characteristically very similar, apart from their likelihood to respond to the marketing campaign, we will not be deceived by this paradox. Unfortunately this is not very often.



How to avoid the Paradox

It is essential when preparing a campaign to take a scientific approach to test design. Whether it be off-line, as in the above example (a very costly mistake) or on-line PPC digital testing. As marketing professionals begin to unleash the power of large sample online testing, it becomes paramount this trap is avoided.

Make sure that samples are spread evenly across lists. Or better yet employ stratified sampling techniques which recognises the different sub populations within the data and collects a simple random sample from each. Breaking lists down into equivalent sub-categories down into equivalent sub-categories enables organisations to understand what lists/creatives are working. This will also help avoid mistakes being made when testing what is the best option available. In the above example, both were correct when looking at the campaign performance reporting, but both were incorrect when preparing for the test matrix

When a dataset represents a large population covering characteristically diverse sub populations, ignoring this dynamic is at best sloppy. So not to be undone by this pernicious paradox, Metrix Data Science adopts a comprehensive approach to the collection of relevant data and rigorously analyses all relevant variables.

Conclusion

By nature, a paradox is difficult to understand and Simpsons paradox follows this trend. MDS recommend someone with data science and statistical experience is on campaign projects to ensure that these problems are known before diving into large scale sends.

An experienced data scientist would have been able to identify the problem before the send, and recommended that the best way to prevent the paradox is to keep the number of sends constant throughout.