Enhancing the Effectiveness of Online Experiments using CUPED: A Case Study of Booking.com

By Eric F Aragundi (efa2119), Spardha Sharma (ss6343) and Simran Padam (sdp2158)

Problem Description

Online experiments can be biased due to user characteristics and other factors, which can lead to inaccurate results and unreliable conclusions. In this case study, we will see how Booking.com is solving this problem of bias in online experiments. By using CUPED (Controlled-experiment Using Pre-Experiment Data), a technique that matches user profiles between the treatment and control groups, Booking.com ensures that both groups are similar in terms of relevant characteristics. This reduces the risk of bias and leads to more accurate results.

Practical Applications of CUPED

The practical application and significance of using CUPED is that it leads to more accurate and reliable results in online experiments. This has helped Booking.com to make better decisions and improve the user experience on their platform. By reducing variance, CUPED ensures that the results of online experiments are not skewed by user characteristics or other factors. This helps Booking.com to make data-driven decisions that improve the user experience and business outcomes.

Challenges of using CUPED

1. Complexity: Implementing CUPED requires a thorough understanding of statistical techniques such as covariate balancing and decoration. This can be difficult for non-experts to understand and implement.
2. Appropriateness for certain experiments: CUPED may not be appropriate for all types of online experiments. For example, experiments that rely on small sample sizes may not benefit from the additional complexity of CUPED.

Executive Summary

CUPED is a technique that helps to reduce bias in online experiments. Bias can occur when certain user characteristics or other factors influence the assignment of users to the treatment or control group in an experiment. This can lead to inaccurate results and unreliable conclusions. By using CUPED, a platform like Booking.com can ensure that both the treatment and control groups are similar in terms of relevant characteristics, reducing the risk of bias.

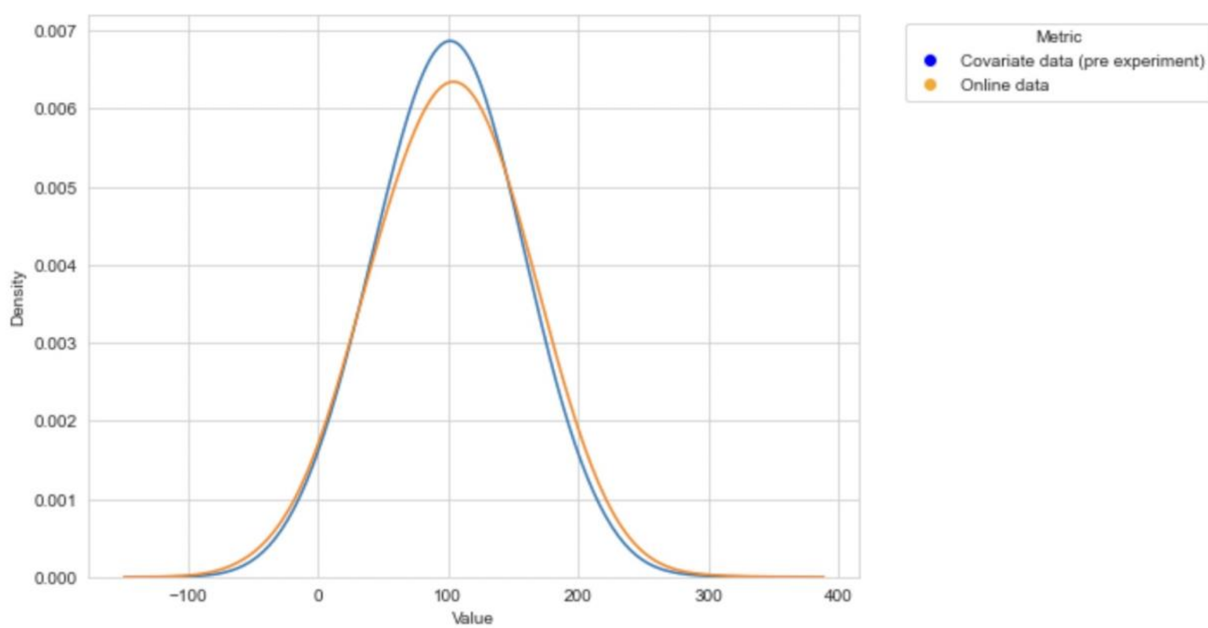
Methodology

We generate simulated data to demonstrate the use of covariate adjustment and the impact of the CUPED method. The data generation process involves generating pre-experiment covariate data for two groups, control and treatment, and then generating post-experiment data for both groups by adding random noise to the pre-experiment data. Finally, the CUPED method is applied to adjust the

post-experiment data by regressing it on the pre-experiment covariate data.The data is visualized using density plots to compare the pre-experiment covariate data to the post-experiment data and the unadjusted post-experiment data to the CUPED adjusted post-experiment data. The CUPED adjusted post-experiment data is also compared to the unadjusted post-experiment data in a separate plot to demonstrate the effectiveness of the CUPED method in removing bias. A Pandas dataframe is created to store the generated data and the adjusted post-experiment data. Finally, a t-test is performed on the unadjusted post-experiment data and the CUPED adjusted post-experiment data for the control and treatment groups separately to demonstrate the impact of the CUPED method on statistical inference.

Data Analysis

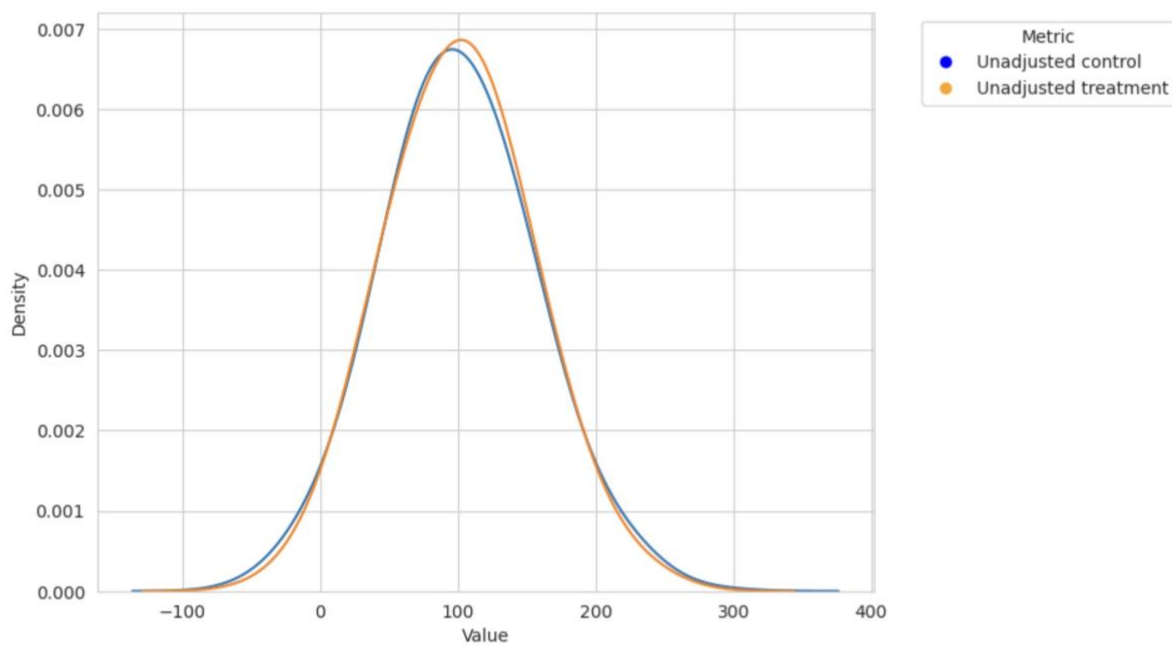
The code simulates a randomized controlled trial (RCT) where we have a control and a treatment group with the same number of individuals, and we want to measure the treatment's effect on an outcome variable. We start by generating a sample of 1000 individuals for each group. The outcome variable is generated by adding random noise to a baseline value. The control group's baseline value is sampled from a normal distribution with a mean of 100 and a standard deviation of 50, and the treatment group's baseline value is sampled from the same distribution. The noise is added to each group separately, with a standard deviation of 20 for both groups. Additionally, a treatment effect of 5 units is added to the treatment group. We plot pre-experiment and experimental data using kernel density estimation. We can see that both metrics: pre-experiment and online data follow a normal distribution.



Next, we create a pandas DataFrame to store the simulated data. We add a user ID and an assignment column that indicates whether each user is in the control or treatment group. We also add two columns to store the outcome variable: one for the unadjusted outcome and one for the outcome adjusted using covariate adjustment with the CUPED method. The following is a glance of the random five rows in our simulated dataset.

	User id	Assigment	Online synthetic metric	Covariate (pre syntheic data)
1095	1096	treatment	79.454720	72.539291
1868	1869	treatment	59.833271	69.932246
1778	1779	treatment	76.400231	55.863242
374	375	control	53.952427	41.911214
338	339	control	94.844383	84.387060

Next, we plot the unadjusted data for control and treatment groups using kernel density estimation.

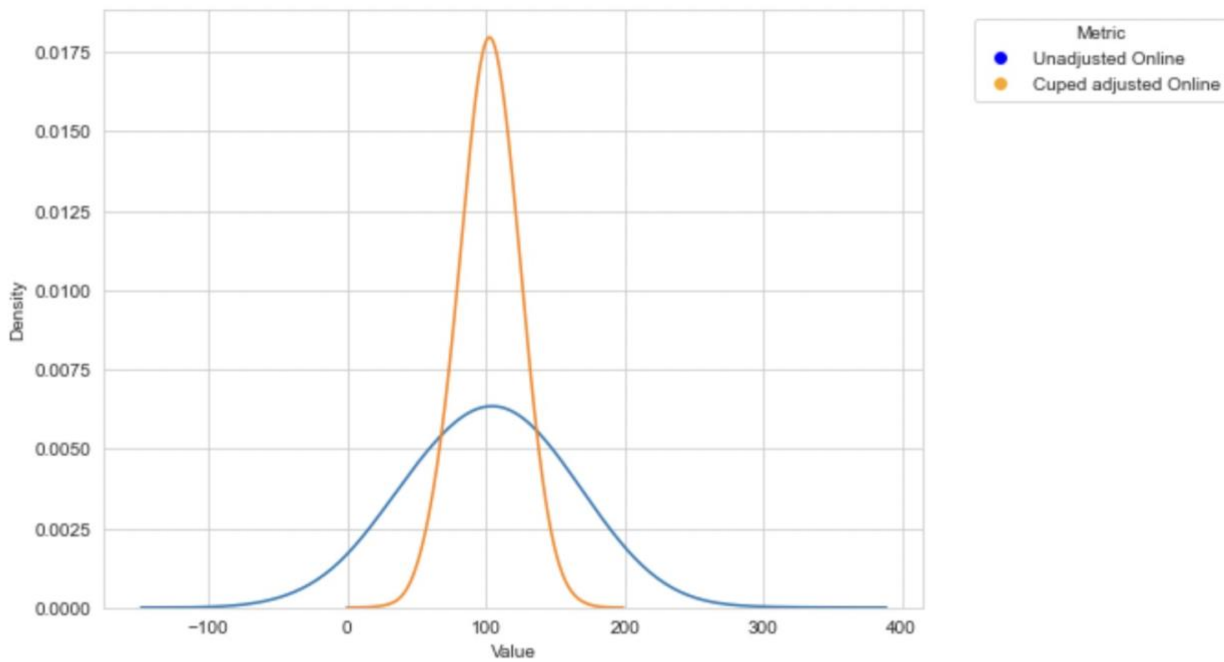


Two independent sample t-test is run on unadjusted control and treatment to check for detectable effect. P-value being high, 0.83, greater than alpha = 0.05, we have sufficient evidence to not reject the null hypothesis and conclude that there is no significant detectable difference.

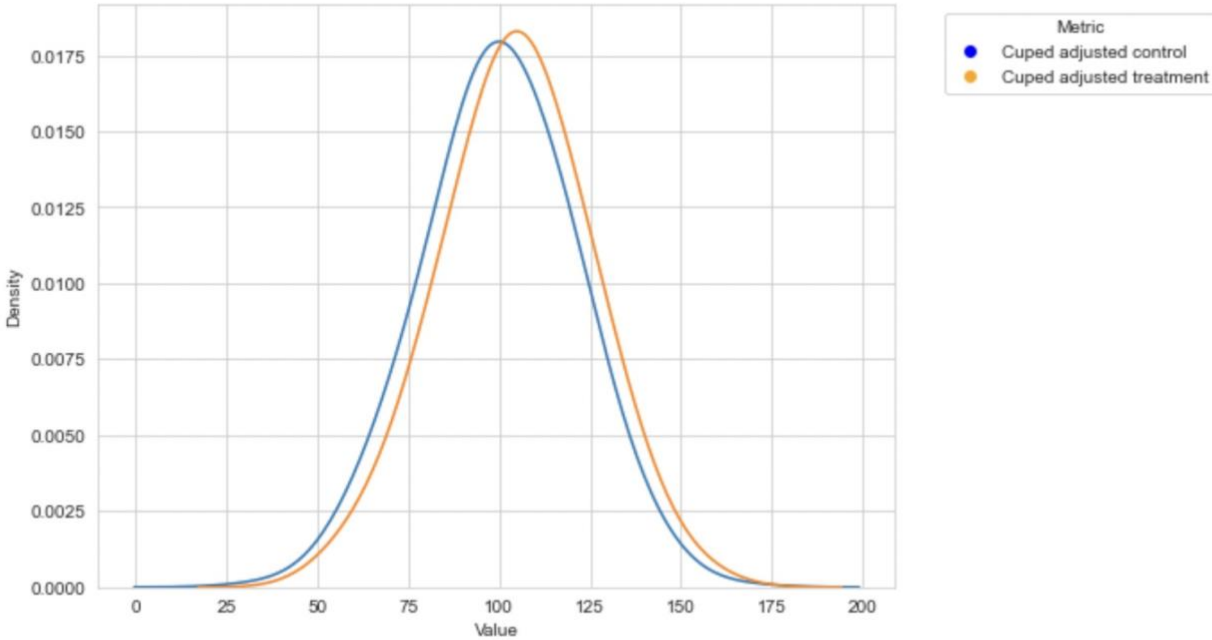
After calculating the theta value for CUPED adjustment, we create a new column in our dataset and call it CUPED adjusted metric.

	User id	Assigment	Online synthetic metric	Covariate (pre syntheic data)	Cuped adjusted metric
0	1	control	21.517699	46.212591	74.180595
1	2	control	142.234067	132.842587	108.443111
2	3	control	38.011016	38.876750	97.994837
3	4	control	72.890766	76.664851	95.163321
4	5	control	38.912655	44.935965	92.849581

Notice that in this experiment we are not considering the scenario of not having data about a user on pre-experiment for covariate. This means that the user will have NULL for the covariate column. If that were the case then the CUPED adjusted metric for that user will be just the online metric itself without adjustment. Next, we plot the density of the unadjusted and adjusted outcome variables for each group using seaborn kde plot. We can see that the adjusted outcome variables are closer for both groups, indicating that the covariate adjustment has reduced the effect of confounding variables.



Let's' see what happens when we apply CUPED and we used our adjusted data:



H_0 = no detectable effect vs H_1 = there is some effect

Finally, we perform a t-test to compare the unadjusted and adjusted outcomes for the treatment group. The p-value is much lower for the adjusted outcome, indicating a stronger treatment effect.

Conclusion

Calculated p-value is 0.00016 < 0.05. Thus, we reject the null hypothesis and conclude that there is an effect between unadjusted and adjusted outcomes for the treatment group.

Citation

[1] How Booking.com increases the power of online experiments with CUPED: https://booking.ai/how-booking-com-increases-the-power-of-online-experiments-with-cuped-995 d186fff1d

Source Code

https://drive.google.com/file/d/1YX2jAHrVs1OxlnOX5hBq6j2GFrgOhFq8/view?usp=sharing