# BSTS Synthetic Control Generated Revenue Data

Jonathan Hershaff

2025-06-29

```
# Define the file path
file_path <- "/Users/jonathan/Desktop/projects/udacity_course/lessons/lesson4_synthetic_
control/data/synthetic_control_revenue_data.csv"

# Read the CSV file
revenue_data <- read_csv(file_path)
```

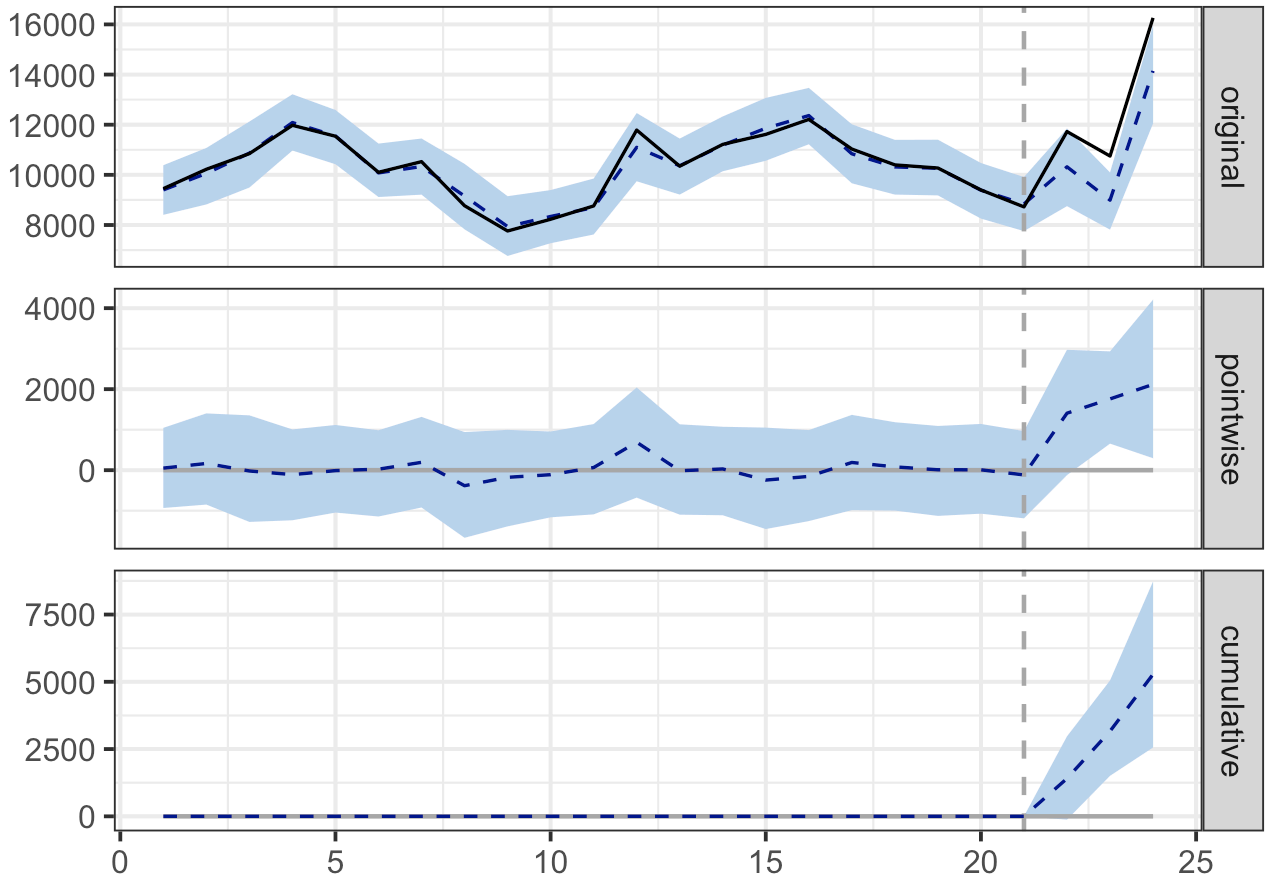
```
## Rows: 24 Columns: 11
## ── Column specification ────────────────────────────────────────────────────────
────────────────────────────────────────────────────────
## Delimiter: ","
## dbl  (10): Region 1, Region 2, Region 3, Region 4, Region 5, Region 6, Region 7, Regi
on 8, Region 9, Region 10
## date  (1): Month
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Replace spaces with underscores in column names
colnames(revenue_data) <- gsub(" ", "_", colnames(revenue_data))

# View the first few rows of the data
print(head(revenue_data))
```

```
# Prepare the data for CausalImpact
# Extract the treated region (Region 5) and the control regions (Region 1 to Region 4, R
egion 6 to Region 10)
treated_region <- revenue_data$Region_5
control_regions <- revenue_data[, c("Region_1", "Region_2", "Region_3", "Region_4",
                                    "Region_6", "Region_7", "Region_8", "Regio
n_9", "Region_10")]

# Combine treated and control regions into a matrix
impact_data <- cbind(treated_region, control_regions)

# Define the pre-treatment and post-treatment periods
pre_period <- c(1, which(revenue_data$Month == "2024-09-01"))
post_period <- c(which(revenue_data$Month == "2024-10-01"), nrow(revenue_data))

# Run the CausalImpact analysis
impact <- CausalImpact(impact_data, pre_period, post_period)

# Print the summary of the impact analysis
summary(impact)
```

```
## Posterior inference {CausalImpact}
##
##                         Average          Cumulative
## Actual                  12913            38738
## Prediction (s.d.)       11152 (533)      33455 (1599)
## 95% CI                  [10003, 12059]   [30009, 36178]
##
## Absolute effect (s.d.)  1761 (533)       5282 (1599)
## 95% CI                  [853, 2910]      [2560, 8729]
##
## Relative effect (s.d.)  16% (5.7%)       16% (5.7%)
## 95% CI                  [7.1%, 29%]      [7.1%, 29%]
##
## Posterior tail-area probability p:   0.00102
## Posterior prob. of a causal effect:  99.89848%
##
## For more details, type: summary(impact, "report")
```

```
# Plot the results
plot(impact)
```




```
summary(impact, "report")
```

```
## Analysis report {CausalImpact}
##
##
## During the post-intervention period, the response variable had an average value of ap
prox. 12.91K. By contrast, in the absence of an intervention, we would have expected an
average response of 11.15K. The 95% interval of this counterfactual prediction is [10.00
K, 12.06K]. Subtracting this prediction from the observed response yields an estimate of
the causal effect the intervention had on the response variable. This effect is 1.76K wi
th a 95% interval of [0.85K, 2.91K]. For a discussion of the significance of this effec
t, see below.
##
## Summing up the individual data points during the post-intervention period (which can
only sometimes be meaningfully interpreted), the response variable had an overall value
of 38.74K. By contrast, had the intervention not taken place, we would have expected a s
um of 33.46K. The 95% interval of this prediction is [30.01K, 36.18K].
##
## The above results are given in terms of absolute numbers. In relative terms, the resp
onse variable showed an increase of +16%. The 95% interval of this percentage is [+7%, +
29%].
##
## This means that the positive effect observed during the intervention period is statis
tically significant and unlikely to be due to random fluctuations. It should be noted, h
owever, that the question of whether this increase also bears substantive significance c
an only be answered by comparing the absolute effect (1.76K) to the original goal of the
underlying intervention.
##
## The probability of obtaining this effect by chance is very small (Bayesian one-sided
tail-area probability p = 0.001). This means the causal effect can be considered statist
ically significant.
```