Strata + Hadoop WORLD

PRESENTED BY

O'REILLY cloudera

Measuring benefit effect for customers with bayesian prediction modeling

Kwon, JeongMin (cojette@gmail.com)

Strata + Hadoop World 2015, London

strataconf.com

#StrataHadoop

Kwon,JeongMin

- Data Analyst @ SK Planet
- Interests: Data Problem Solving with R, Hive, SQL, Python and others

Offering

* LIMITED TIME OFFER *

$20 OFF

for you & your friends

Introduce your friends to the best in American made and give them 15% off their DODOcase purchase. For every friend that redeems your offer, you will receive a $20 coupon back!

CREATE YOUR INVITE LIST:

From: buyer@example.com IMPORT CONTACTS

Enter email addresses separted by commas

Congratulations! You Get 15% off at DODOcase.

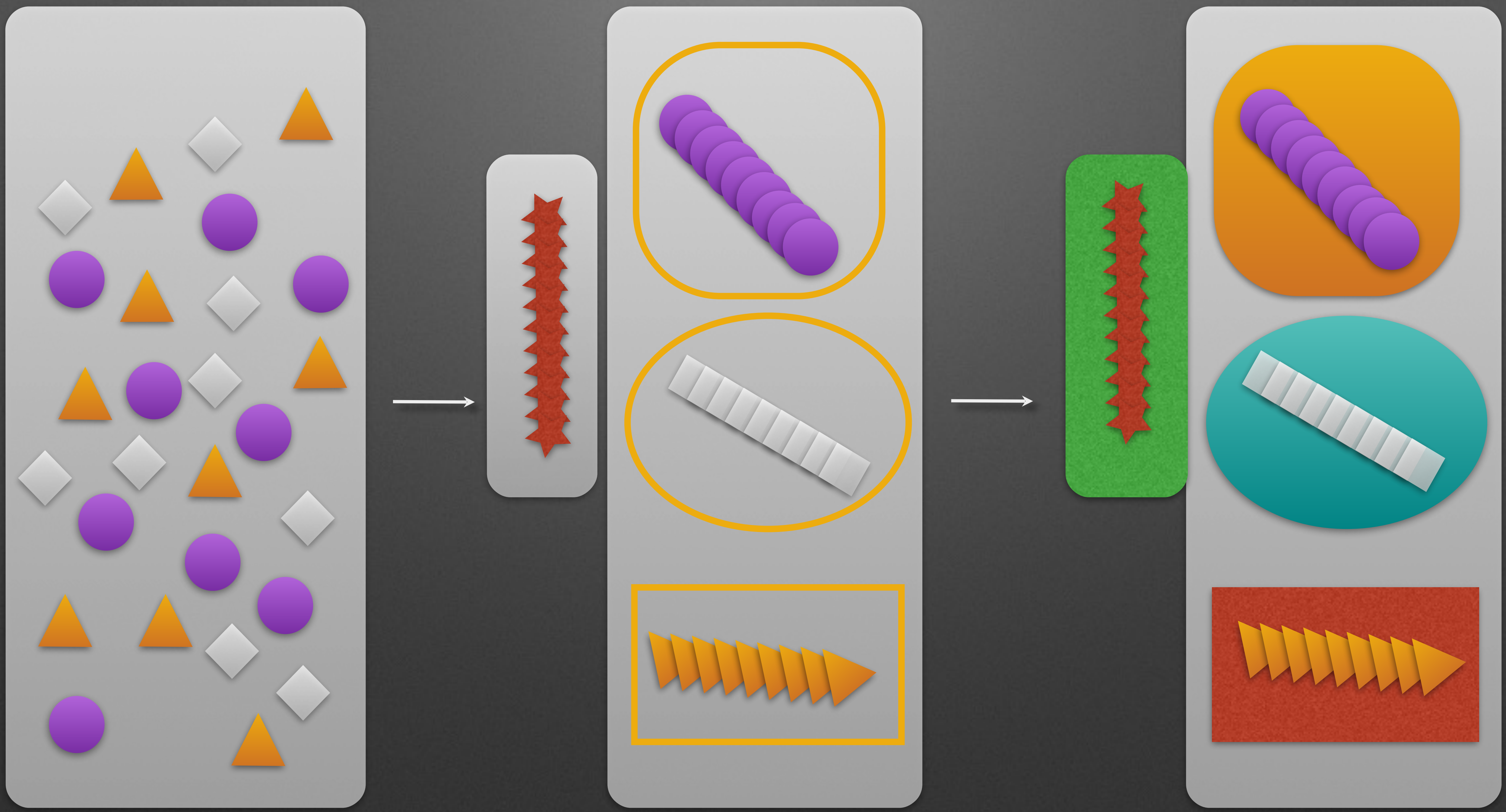
Add a personal message below to go in the email with the offer details:

I'm excited to share this great DODOcase deal with you. Enjoy shopping and supporting American craftsmanship!

Offering

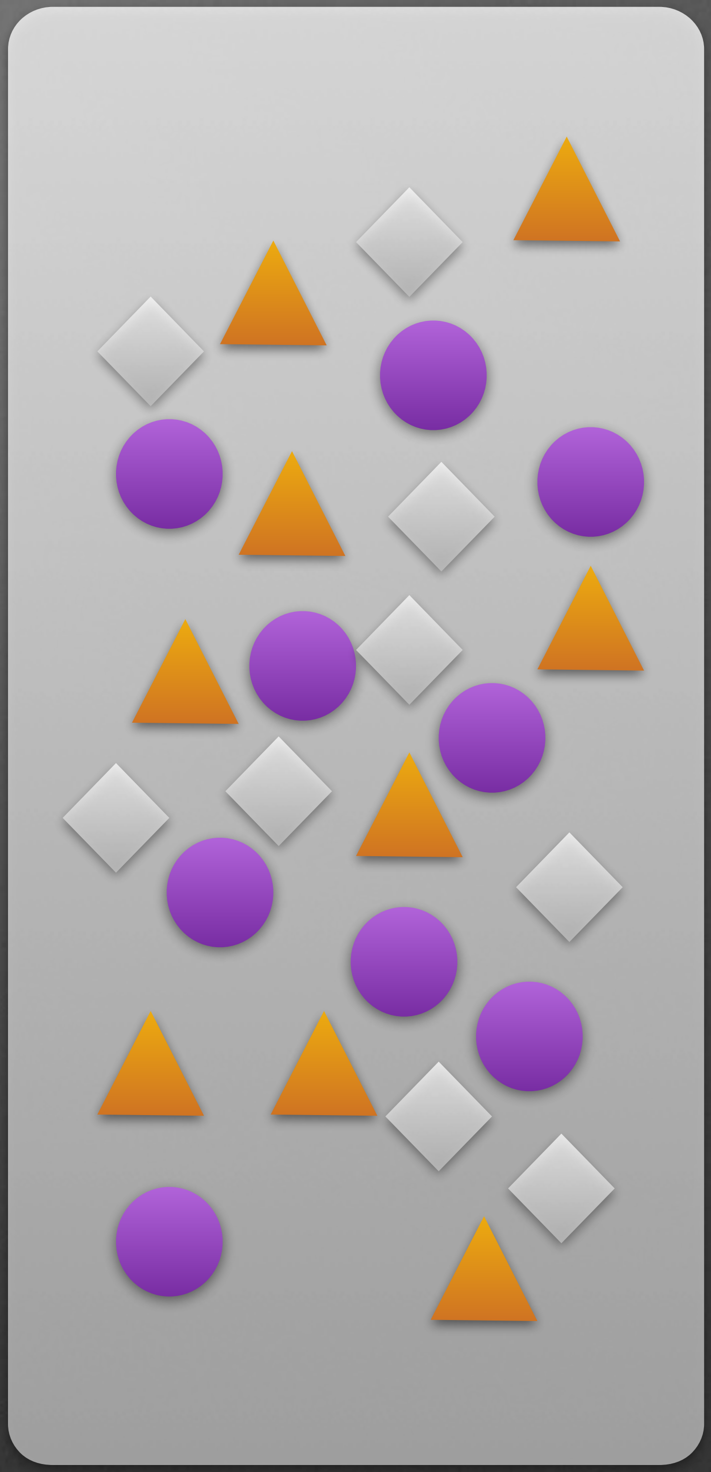
- Popular way to promote products

Offering Process



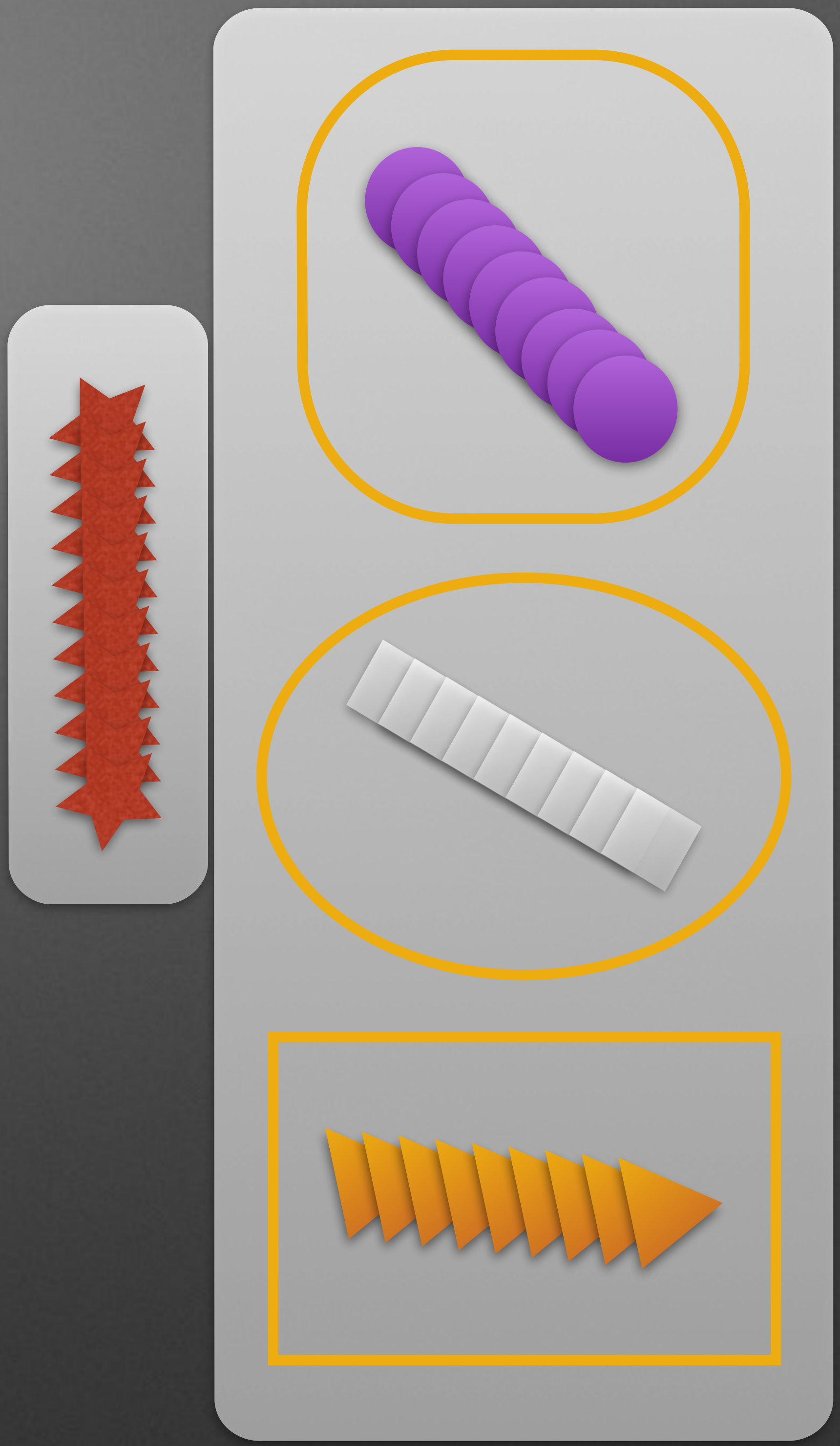
Step 1:
Customer Data Management

- Monitor customers' actions
- Keeping track of customer data and information
- Defining goals



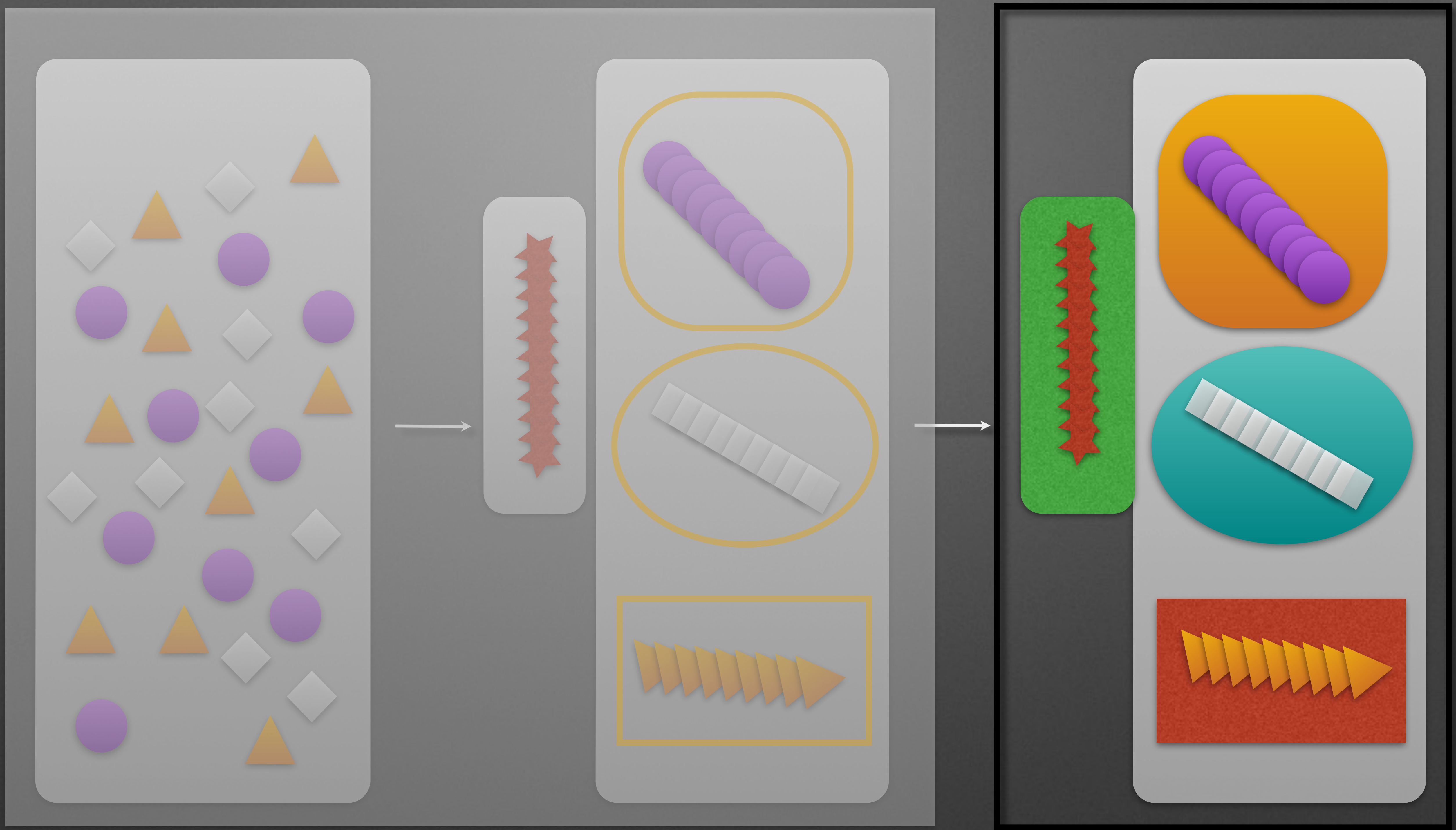
Step 2: Targeting

- Customer segmentation and selection with goals at step 1
- Based on demographic informations and log collections
- Statistical methods and data mining algorithms

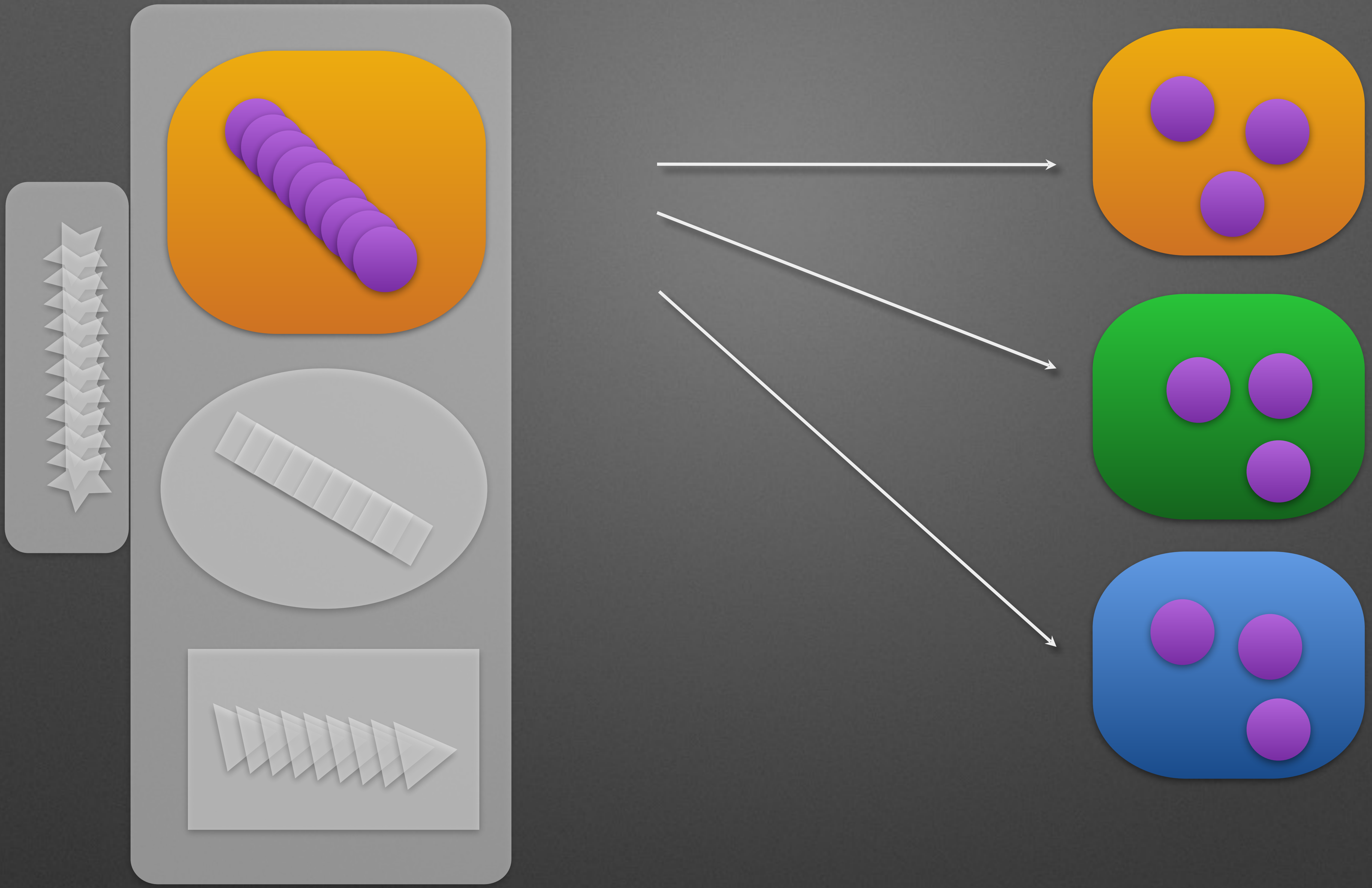


Step 3: Offering Benefit (Campaign)

- Delivering proper benefits to targeted customer groups
- Methods: Promotion, Event, Advertisement and others
- Measurement and prediction of campaign effects



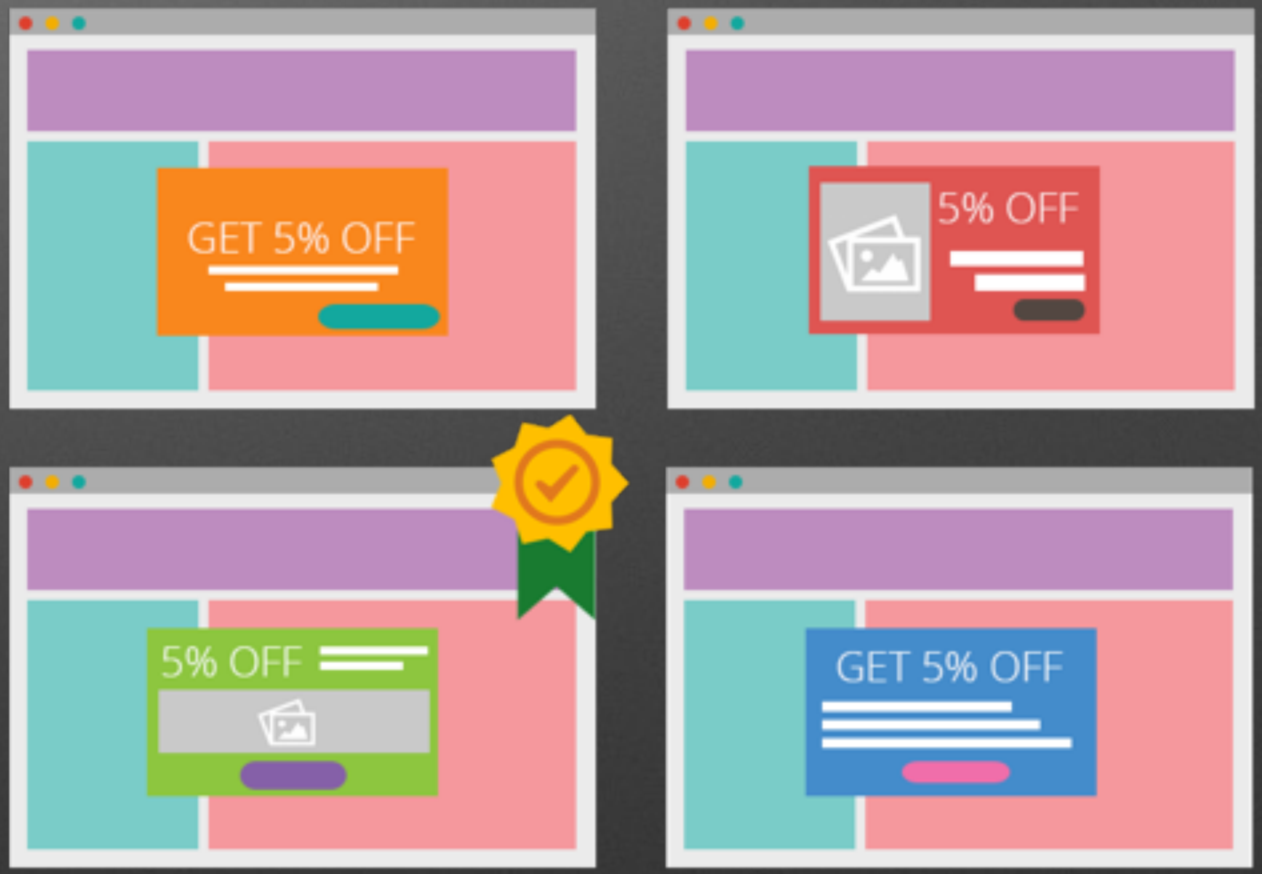
How to measure and predict



How to <u>compare</u> offering effects

Multivariate Testing

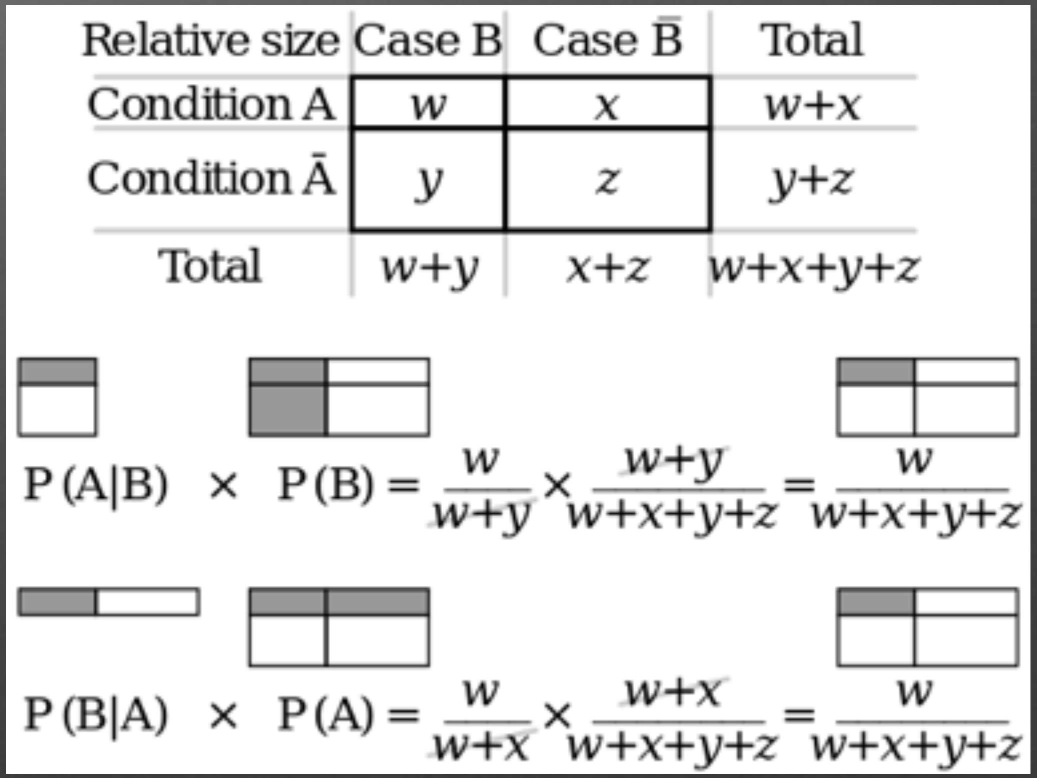
- Technique for testing a hypothesis with multiple variables
- Issues for offering
  - Lack of long-term prediction
  - data, benefit limitations



(Image from https://www.ownedit.com/features)

Bayesian Interpretation

- Diachronic Interpretation
  - Probability of the hypotheses changes over time
  - Prior and posterior based on background information
  - Good for simulation, decision and prediction



(Image from http://en.wikipedia.org/wiki/Bayes%27_theorem)

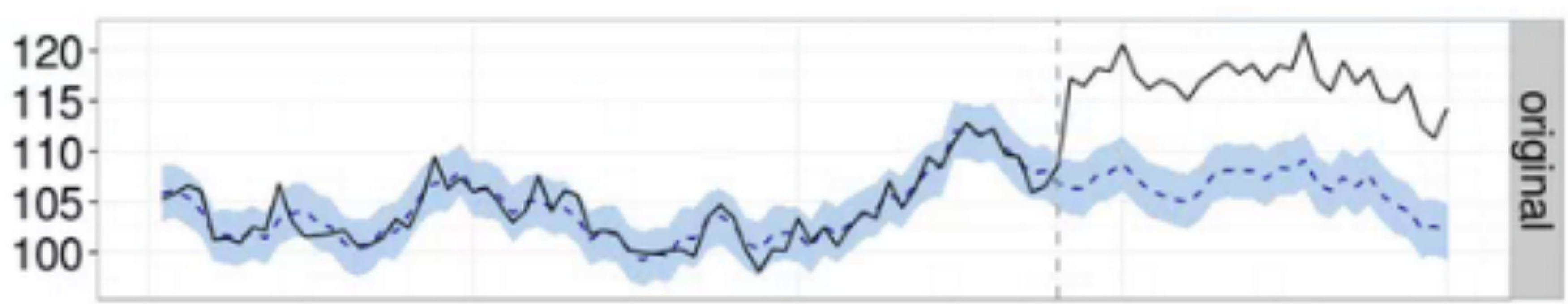
September 12, 2014

Google uses R to calculate ROI on advertising campaigns

Google has just released a new package for R: CausalImpact. Amongst many other things, this package allows Google to resolve the classical conundrum: how can we asses the impact of an intervention (for example, the effect of an advertising campaign on website clicks) when we can't know what would have happened if we *hadn't* run the campaign? For a marketer, the worry is that the spike in clicks was partially or wholly the result of something unrelated (say, a general increase in web traffic) rather than your campaign.

The CausalImpact package uses Bayesian structural time-series models to resolve this question. All you need is a *second* time series to act a a "virtual" control, which is unaffected by your actions but which is still subject to the extraneous effects you're worried about. (For the marketing example, you might choose web clicks from a region where the campaign didn't run.) Then, you can model the extraneous effects and subtract them from your actual results, so see how your things would have played out had the intervention *not* occurred.

In the chart below (from the Google Open Source blog post) you can see the results of the campaign in black, with the campaign launch at the dotted line. The blue line shows the estimated results had the campaign *not* run, clearly showing that it was effective.



Google uses R and the CausalImpact package to measure the return-on-investment on advertising campaigns its customers run:

We've been testing and applying structural time-series models for some

About this blog
Comments Policy
About Categories
About the Authors
R Community Calendar
Local R User Group Directory

R for the Enterprise

Got comments or suggestions f
Email David Smith.

Follow David on Twitter: @
+David Smith

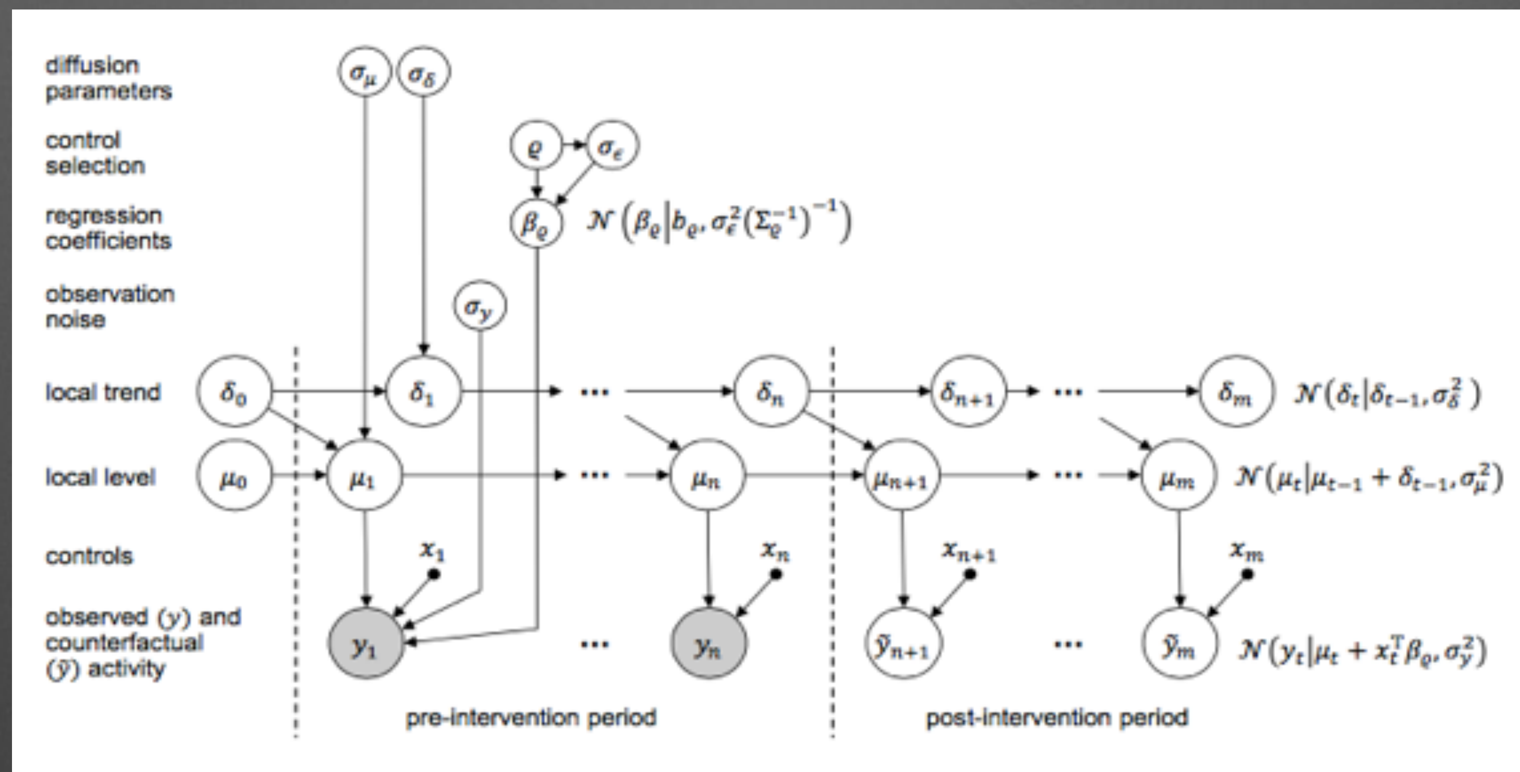
Subscribe to this blog's feed

Categories

#StrataHadoop

Strata+Hadoop WORLD

CausalImpact



(Image from the paper)

- Based on the paper [Inferring causal impact using Bayesian structural time-series models], Google, 2014
- CausalImpact Package in R
    - https://github.com/google/CausalImpact

CompareImpacts

- Integration of bayesian time series prediction model with multivariate tests
- For simple comparison of causal effects

```r
CompareImpacts <- function(data = NULL,
                           pre.period = NULL,
                           post.period = NULL,
                           model.args = NULL,
                           bsts.model = NULL,
                           post.period.response = NULL,
                           alpha = 0.05) {
  # by JeongMin Kwon (cojette@gmail.com)
  # CompareImpact() performs causal inference comparison based on local functions
  # in CausalImpact package.
  #

  data.length <- dim(data)[2]
  impact.list <- list()

  # Check input
  if (data.length <3)
  {
    assert("Nothing to compare! You should use CausalImpact. ")
  }
  else
  {
    data.list <- list()
    for (i in 1:length-1)
    {
      data.list[i]<- cbind(data[,i], data[,length])
    }

    for (i in 1:length-1)
    {
      checked <- FormatInputForCausalImpact(data, pre.period, post.period,
                                            model.args, bsts.model,
                                            post.period.response, alpha)
      data <- checked$data
      pre.period <- checked$pre.period
      post.period <- checked$post.period
      model.args <- checked$model.args
      bsts.model <- checked$bsts.model
      post.period.response <- checked$post.period.response
      alpha <- checked$alpha

      # Depending on input, dispatch to the appropriate Run* method()
      if (!is.null(data)) {
        impact.list[i] <- RunWithData(data, pre.period, post.period, model.args, alpha)
      } else {
```

Use Cases

- Same offerings in various groups

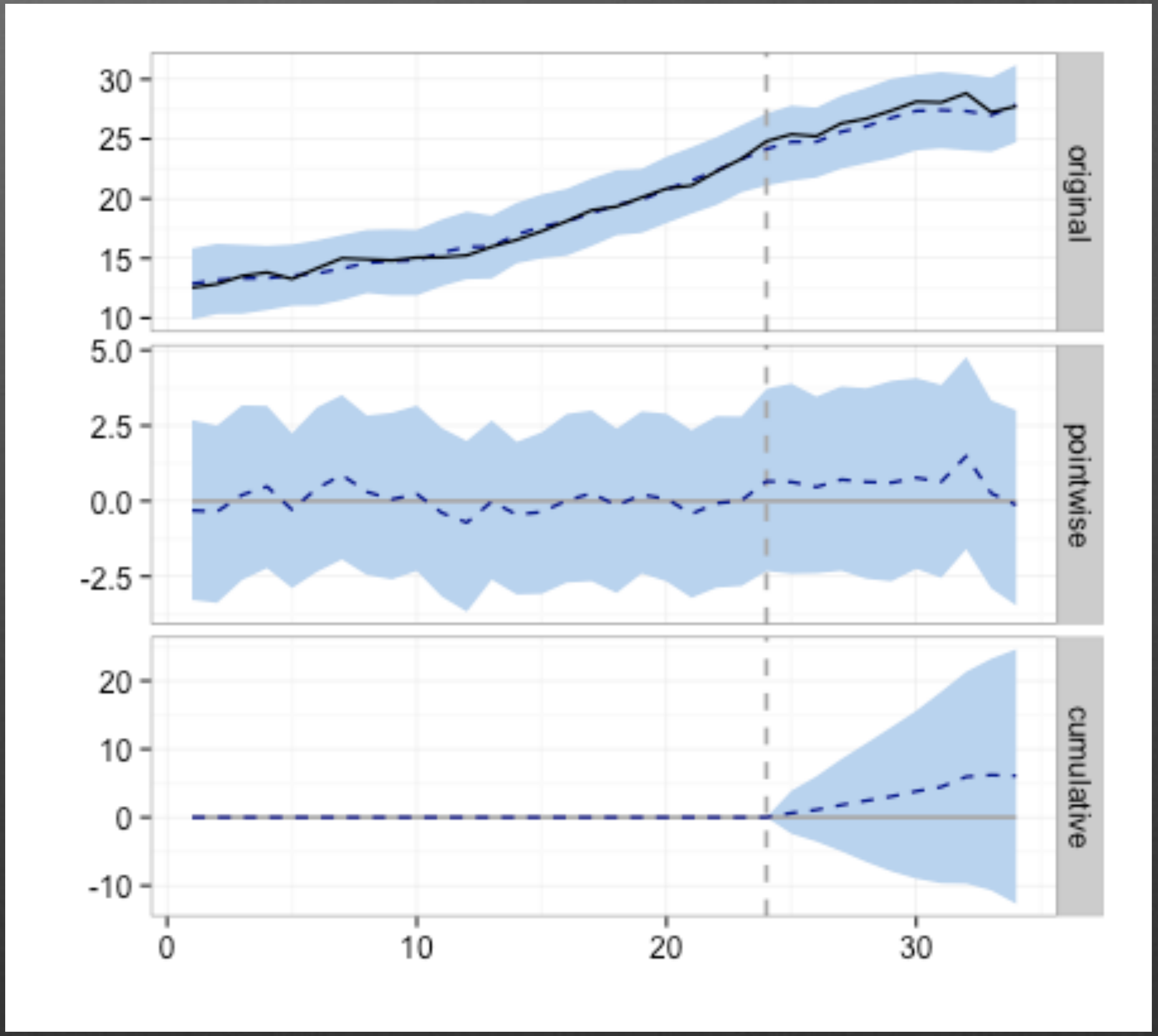
Use Cases

- Various offerings in a group

Basic Model

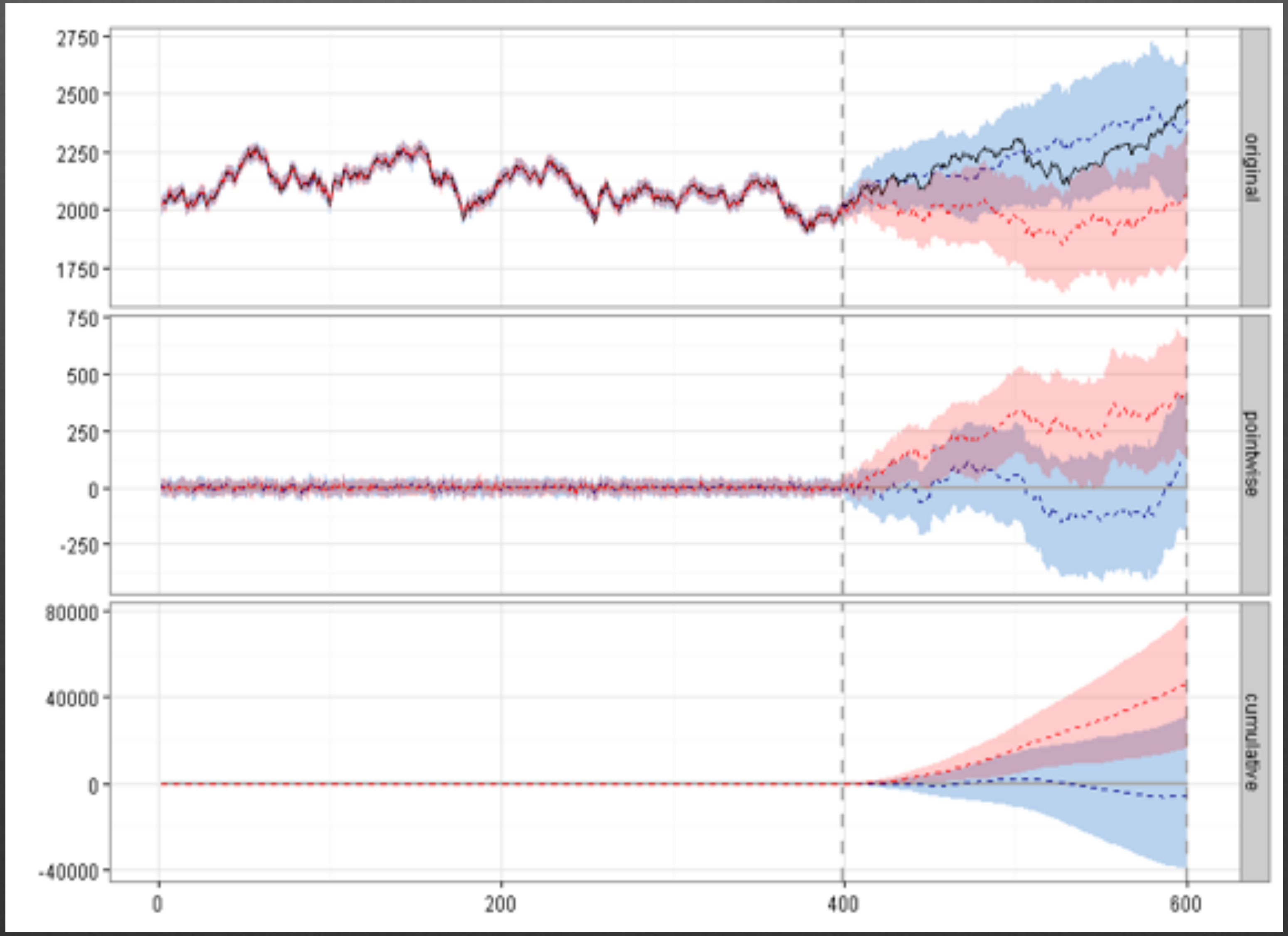
- CausalImpact results

```
> summary(imp1)
Posterior inference {CausalImpact}

                         Average        Cumulative
Actual                   27             271
Prediction (s.d.)        26 (1)         265 (10)
95% CI                   [25, 28]       [246, 284]

Absolute effect (s.d.)   0.61 (1)       6.07 (10)
95% CI                   [-1.3, 2.5]    [-12.7, 24.6]

Relative effect (s.d.)   2.3% (3.8%)    2.3% (3.8%)
95% CI                   [-4.8%, 9.3%]  [-4.8%, 9.3%]

Posterior tail-area probability p:   0.25527
Posterior prob. of a causal effect:  74%
```



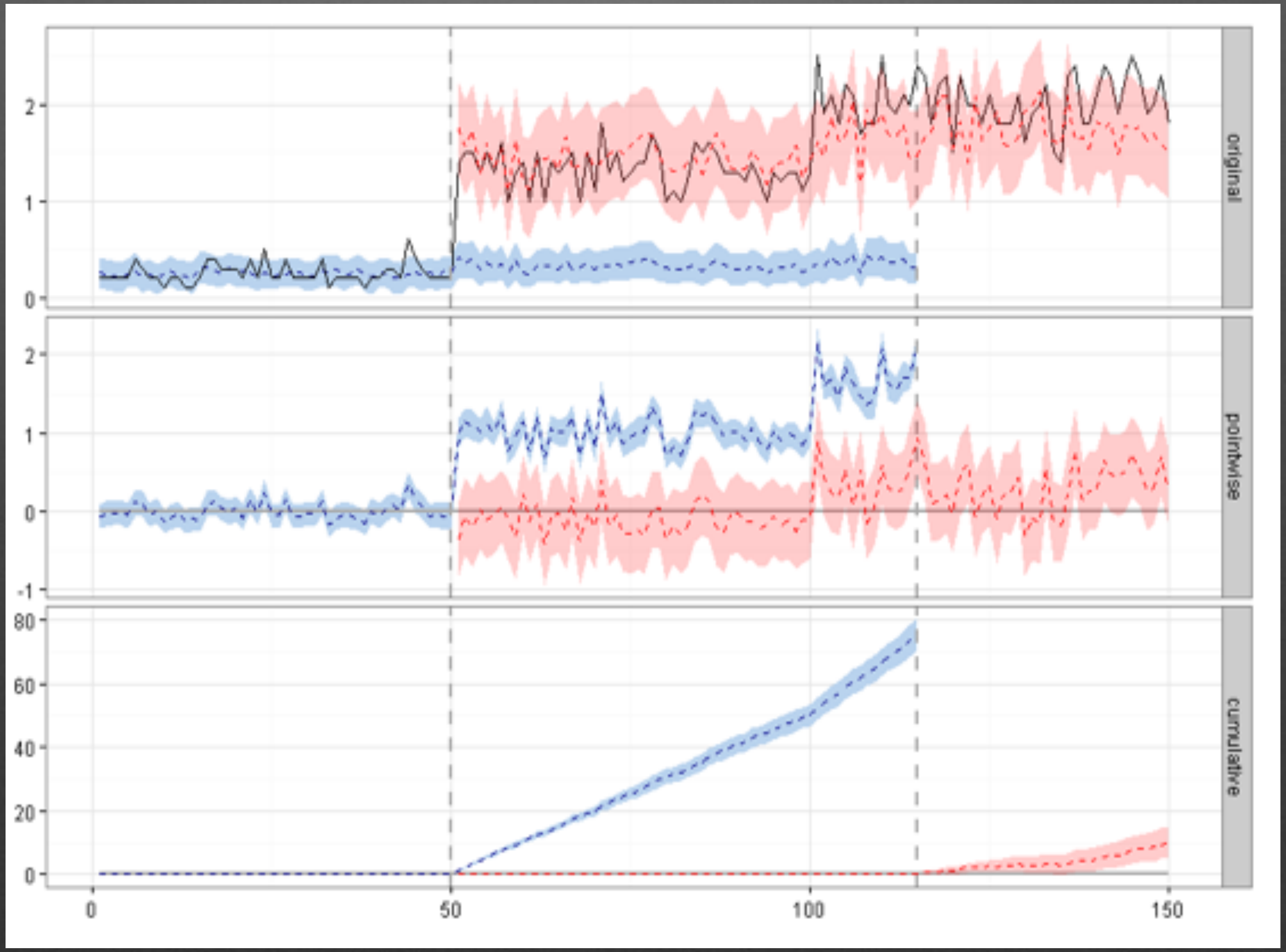
Use Case Results

- Same offerings in three groups

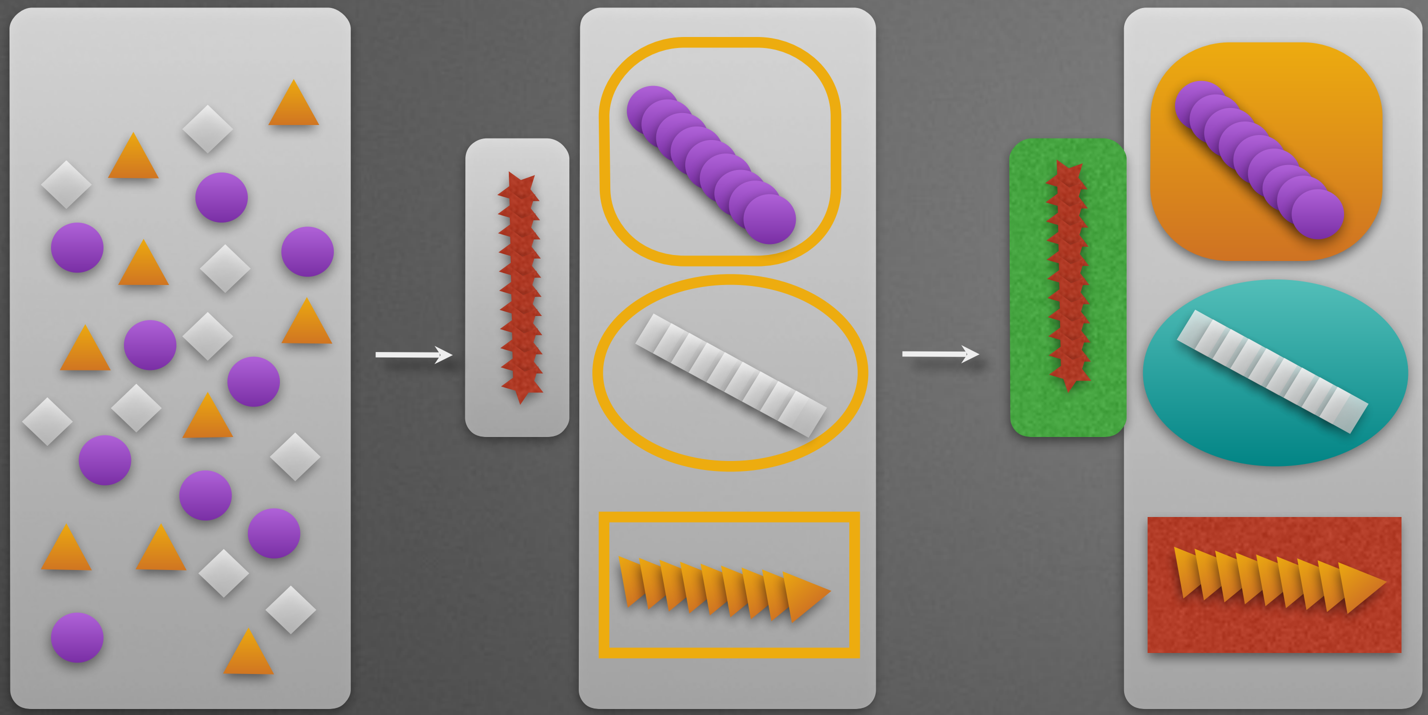


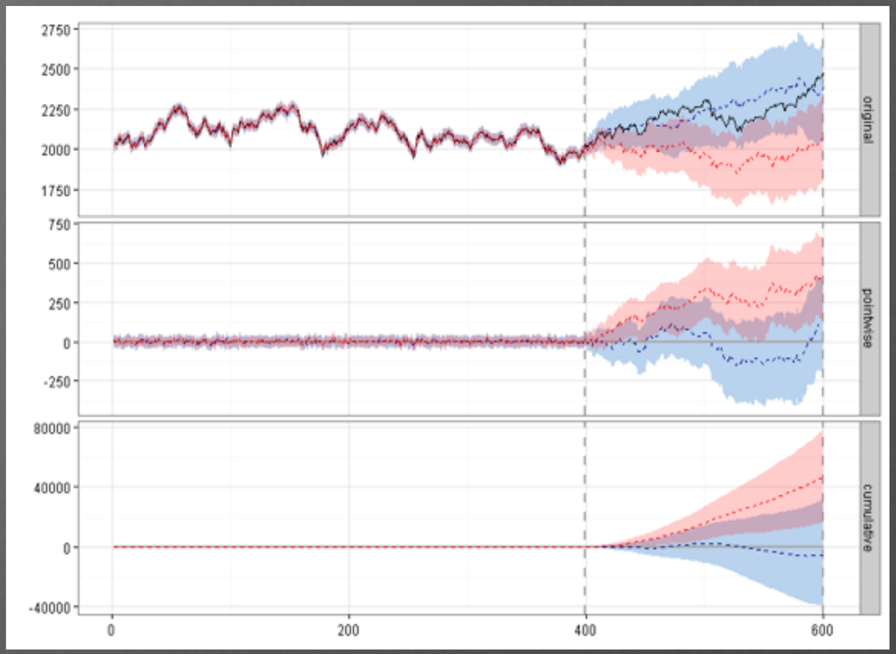
Use Case Results

- Offerings in a group with time differences




```
> CreateCompImpPlot(qq)
[1] "series 2 would generally not be considered statistically significant and plot is omitted."
```





New Tool for Offering Comparison: Multivariate Test + Bayesian Time-Series Analysis

Office Hour: 15:25~16:05, Table A

E-mail : cojette@gmail.com