

# Attribution and allocation for D2C clothing

**Motivation:** Our new direct to consumer (D2C) clothing business has been running experiments to understand the best advertising channels for our product. To do this, we've collected a week of data from our website traffic monitoring platform to observe how spend relates to customer conversion (conversion = customer bought the product). For our paid advertising channels we're worried about market saturation: is spending more money on marketing valuable or do we start to see diminishing returns? To test this we ran identical ad campaigns on three different (identical) population samples that only differed by total level of spend: tier 1 (\$50), tier 2 (\$100), and tier 3 (\$150). We're interested in understanding our marketing effectiveness by channel and developing a spending allocation plan for next week's advertising budget.

## Instructions

**Part 0, optional:** Listen to interview with Kevin Frisch, Uber CMO on detecting attribution fraud on behalf of programmatic agency partners: <https://www.marketingtodaypodcast.com/194-historic-ad-fraud-at-uber-with-kevin-frisch/>

**Part 1, attribution:** Allocate conversions by channel (*social, organic\_search, referral, email, paid\_search, display, direct*) and evaluate effectiveness

- Test 3 methods for allocation
- Calculate CAC for each of the channels
- Discuss observations and potential conclusions from CAC calculations

**Part 2, allocation:** For one of the allocation methods, calculate the marginal CAC by spending tier by channel

- Discuss how your observations/conclusions from the previous section may change
- This week we spent \$1,500 total on advertising across all platforms. Next week we want to allocate the same budget in \$50 increments, i.e. \$700 spend on display adds, \$50 spend on social ads, etc. Determine how you would most effectively allocate this budget.

## Hints

- The data set is large, so looping over a DataFrame to apply attribution functions can be time consuming. For attribution, we only care about allocating credit for visitors who converted...not for observations where the potential customer didn't convert. This may help pare down the number of calculations you have to perform
- The data set is large, so as is best practice for data science, I recommend **randomly** selecting observations to test code on while in the development stage, then scaling up the sample until running the code on the full set once you're reasonably assured the code is working as intended

## Submission

Submit 5-7 slides as a group to Assignment 3 responding to the analysis request above

On your title slide be sure to include the names of group members and a link to one GitHub repository containing a complete version of the code. Only one version will be reviewed. All are welcome to have the code in their repos – and it's encouraged – but it will not be reviewed. You can work in groups of up to 5 people. Fewer (or individual work) is ok; groups are for your convenience and learning.

3 groups will be selected on Wednesday (4/22) to give ~5-10 minute presentation. Understanding and preprocessing the data may take a deceptively substantial portion of your time depending on your comfort with Python/pandas, so make sure you start early. Part 1 can be done in its entirety before Part 2, so you do not need to wait for the discussion on allocation to begin attribution modeling.

Submissions of presentation and code for all groups are due by Wednesday 4/22/2010 9am Shanghai time.