**ADENIO Data Science Project**

ADENIO Digital is a marketing technology company that develops solutions for the digital advertising space (AdTech).

**Problem Description**

ADENIO Digital has run a digital advertising campaign for one of our clients, targeting mobile users in the Southeast. The goal of the campaign was to drive conversions -- users clicking through the ad and accepting our client's offer. To execute this campaign, we built an audience of target users based on historical information about the likelihood to convert.

Our campaigns are executed on real-time bidding (RTB) exchanges. Web pages or mobile apps submit bid requests to potential advertisers on these exchanges. If ADENIO Digital sees a bid request for a user we want to target on a web page or mobile app we want to serve on, we will submit a bid for that advertising placement. If we submit the highest (or only) bid, we win the placement and serve an ad. Given that we have won and served the ad, a user then has the opportunity to convert.

We'd like you to take a look at the campaign's results to answer two primary questions:

1. We ran two creatives (distinct ad designs) during the campaign: a baseline creative in line with our client's previous campaigns, and a new creative designed to increase conversions by more prominently displaying the offer to the user.  To do this, we assigned some users to a "test" group that received the new ad during the campaign.  *Did the new creative increase conversions?*
2. We are equally likely to bid on any user at any time during a campaign.  Other advertisers have found success by "retargeting" users -- increasing the likelihood of a bid for users that have previously landed on a page or actually converted in the past (e.g. redeeming a digital coupon). *Would retargeting have improved performance for this campaign?*

**Data**

There are two files with this project.  These files are as follows:

* *user\_attributes.csv* is a CSV file of target users' attributes
  + *user\_id* is a unique identifier for each user
  + *attributes* is a text blob of the user's attributes
    - *age* is the user's age in years
    - *gender* is the user's inferred gender
    - location is the user's state of residence
    - *test* is a binary variable indicating that the user belongs to the test group
* *bid\_requests.csv* is a CSV file of bid requests that we have seen for the targeted users. Each row represents one (bid) request.
  + *timestamp* is the time the bid request was received. This is unique for each request and can be treated as a unique key.
  + *user*\_id is the unique identifier for user the request is for
  + *bid* is a binary variable indicating whether we made a bid for this request
  + *win* is a binary variable indicating whether we submitted the winning bid and served an ad
  + *conversion* is a binary variable indicating whether the user converted after seeing the ad

**Exercises / Questions**

1. What do the different distributions of requests by user tell us?

Please describe the distributions related to requests and the users we targeted.

1. Based on insights from #1, was the test effective?

Were test users more likely to convert than control users?

1. Should we consider retargeting users that have previously converted?

Would retargeting users in this campaign have improved performance?