

Sojern Data Challenge

A large, multinational bank with branches in the US and UK would like to use digital advertising to drive more traffic to their website to improve brand recognition. This bank has reached out to Sojern to manage their advertising campaigns, and your goal is to help the account manager (AM) who is working with the bank.

One campaign that the AM is managing is a *Search Campaign*. A search campaign is:

- an ad,
- a set of keywords that represent different search intents that match to that ad,
- One bid per keyword which represents how much we are willing to pay for a click on that keyword.

Whenever a user types a query into the search box that matches the keyword, an auction takes place. The marketers with the winning bids then get their ad placed on the search results page in order of descending bids where the highest bid is put in position one, the next highest at position two and so on. Regardless of what we bid, we only pay for the ad if the user clicks on the ad, and the amount we pay is equal to the bid price for the next highest bid because SEM is second price auction. For more context, read sections 1,3,5 in this [tutorial](#).

We have been running a nineteen keyword search campaign for the bank for the past three months. Your goal is to help the AM make better decisions for the next three months. The AM needs to know:

- What metric(s) should we use to gauge performance?
- Are there differences in how the keywords perform?
- Is there a difference between US and UK performance? Can you prove it?
- Should any keywords be removed from the campaign? If so, why?
- Which keywords should have their bid increased? If so, why?

Your deliverable is a jupyter notebook that has answers to these questions. Make sure to:

- Document assumptions and any recommendations you'd make to the AM
- Include useful plots, analysis or commentary to support your conclusions
- If you do any data clean up, please document that as well

Along with this document, you should have received a dataset called `data_challenge.csv` which has the following schema:

- `date` : date the data was recorded
- `country` : country the data was recorded for
- `keyword` : keyword associated with the campaign
- `global_monthly_searches` : total monthly searches for that keyword on google globally
- `ad_position` : how close to the top of the search result page the ad appeared (1 is the top)
- `daily_clicks` : number of clicks that occurred that day for the keyword
- `daily_cost` : amount of money spent for the clicks on that keyword on that day
- `local_monthly_searches` : total monthly searches for that keyword on google for the country
- `impressions` : number of times the ad was shown to a user that day for that keyword

While this problem is extremely open-ended, we don't expect you to spend more than five hours of focused work after starting.