Adludio Data Science Challenge

We are expanding our team and are looking for driven individuals who can complement the skills of other team members, in a fast paced learning environment to deliver value to Adludio.

In the following, we ask you to demonstrate your data science knowledge, skill and attitude in a series of tasks. The tasks are open ended, and all that is expected from you is to do as much as you can. We also don't have a clear answer ourselves to some of the tasks, so use your time optimally.

We will evaluate all submissions based on the following criteria

- Understanding of the problem being asked (you can always ask by email if something is not clear in this description)
- Attempting as many of the tasks as possible in the time given, and answering the questions asked. This is our main indicator of hard work.
- Creative and Innovative analysis for informative insights

About Adludio business

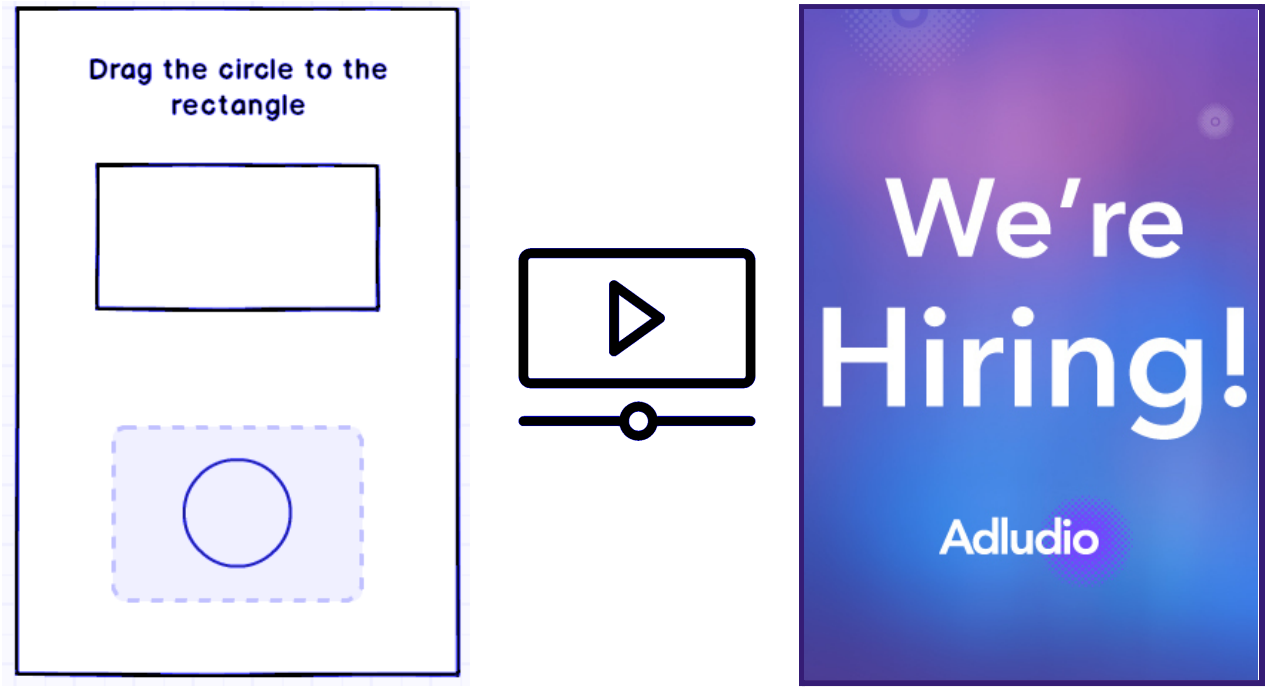
Adludio is an online mobile ad business. It provides the following service to its clients

- Design an interactive Ad - what is also called a "creative". A creative is a rich ad containing interaction elements through a mini game-engine, video, text, images etc..
- Serves these creatives to audiences on behalf of a client. In order to do that, adludio buys impressions from an open market through bidding.
- In order to make a good profit while also satisfying clients needs, adludio employs campaign managers and bidding and other algorithms to optimise the bidding strategy. One of the algorithms we use is called **inventory scoring** i.e, ranking inventories such as site, device type based on ads history performance. and it is used to optimise our targeting strategy.

Glossary of Useful terms

Creative - an advertisement (ad) that a user sees and interacts when browsing a website or using an ad powered mobile app.
Inventory - a digital space advertised by a publisher. These spaces are bought and sold in an ad exchange service, like thetradedesk (TTD) .
Impression: A single impression is registered when one ad is rendered on a user's device. Different impressions are indexed by an auction_id - a unique string.
Advertiser - a client who is advertising for a single or multiple products. Different advertiser companies are distinguished by **advertiser_ids**. An advertiser may advertise multiple products

through different campaigns. The advertiser can be the company with the product or a publisher acting as a middleman between a company and Adludio.
Campaign - an ad placement session for a single client and single product. A **campaign_id** is used to distinguish different campaigns. Within a single campaign, there could be multiple strategies who are served under what is called an adgroup.
Adgroup - an ad channel in a campaign with a unique serving stratergy. Different adgroups are distinguished by an **adgroup_id**.
Campaign Manager - professionals who manage campaigns and interact with the client.
Engagement - an event triggered when a user interacts the first time with an ad
Click Through - the last click a user make before leaving the interaction with the ad
Engagement Rate (ER) - number of engagements divided by the the number of impressions
Click Through Rate (CTR) - the number of clicks divided by the number of engagements
KPI - Key performance indicators. The two KPIs used commonly are **ER** and **CTR**.



Pre-Engagement **Post-Engagement** **End-screen**

Fig 1 : Example of a user interacting with an ad (Engagement Event)

Challenge Description

Data Definition

Please download the data here.

Id: a unique identifier for each row. Each row is recorded whenever an ad is rendered on a user's device.
CampaignId: A campaign is an ad placement session for a single client and single product. The campaignId is used to distinguish different campaigns. Within a single campaign, there could be multiple strategies who are served under what is called an adgroup, identified by the next column.
Adformat: the size of the creative expressed as width and height.
FoldPosition:the declared position of the ad in the view space of the user's device. Possible values are 1 (any), 2 (above), 3 (below), 4 (unknown).
DeviceType: the numeric ID for the type of device on which the impression was served. For supported values, see Device Types.
OS: The numeric ID for the operating system (OS) associated with the impression event. For supported values, see Operating Systems.
Site: The address of the website where the impression was purchased and the creative was displayed. This is the column you are supposed to give a score to and rank in task one.

engagement: a binary value that indicates if the user interacted with the ad or not.

You can calculate engagement rate for a site by dividing the number of impressions (total number of rows having that site) by the number of engagements (number of rows having the value "1" in the engagement column) .

Task 1

What we would like is for you to come up with an algorithm that assigns a score to each site in each campaign and rank them accordingly. The scoring should be done in such a way that a higher score promises having high engagement rates for ads rendered on that site in the future.

Write a function that

- Accepts a campaignId as an input
- Filters the data by the campaignId provided
- Assign a score to each site
- Sort the data by the assigned score

Hint: Bear in mind that, directly taking the engagement rate to assign a score is problematic. Assume we have 3 engagements out of 4 impressions for site A and 70 engagements out of 100 for site B. Even if site A has a better engagement rate (75%) than site B (70%), the value might be a result of random chance due to its low number of impressions and thus we have less certainty.

Task 2

Build a dashboard that allows the user to select a specific campaign from a list of available ones and view the top ranked sites based on the algorithm you worked on (Task 1).

- Prepare a Flask API end-point that accepts a campaignId and returns the top-sites data in JSON format
- Build a front-end using React and consume the API

Task 3

This task is explorative. There is no right/wrong answer, just different approaches. We want to know what you were thinking while trying to complete this task, and what you tried and what worked well. You will not be graded on the results, but rather on the methods you adopted.

We now want to know the level of decisiveness of the four categorical variables (AdFormat, FoldPosition, OS, DeviceType) on whether a user will engage with an ad or not.

Hint: Mutual information is one of many quantities that measures how much one random variable tells us about another.

Submission

You are expected to submit your results, codes (including jupyter notebooks), by the defined date. Submit your notebook using one of the following methods:

- Google colab link
- Github link

- Direct email attachment

Deadline

Friday End of the day, 1st October 2021 before 6 pm EAT time.