Arpeely Take-Home Assignment

Hello and thank you for taking the time to complete the below.

The questions are compiled to both pique your interest and help us evaluate your strengths in different areas.

Please, complete these questions entirely by yourself incl. no outside review of your final results.

Complete as many questions as possible within your time constraints.

The entire assignment should take about two hours to complete.

We'd like to see your code. **Please provide it as a single Jupyter notebook that includes both the final answers and the code**.

Assignment

Please read sections **2.1, 2.2** (pages 8-11) , **3.1** (page 17-18 excluding 3.1.1) in this RTB manual as a concise introduction to the RTB world. Other parts of this manual may be helpful as well but not mandatory(You can ignore the math stuff).

Please, download this dataset https://arpeely.com/data_is_the_new_oil/data_assignment.zip
This dataset is a small subset of user impressions. Each row describes an request/impression of an ad to a user, along with user properties and bidding participation info.

Dataset scheme:

column	description
time	time of the impression
user_id	unique identifier of a user (obfuscated)
domain	domain (obfuscated)
url	url (obfuscated)
ad_slot	ad slot identifier (specific advertisement placement in a webpage) (obfuscated)
country	country code (ISO)
postal_code	postal code
state_code	state code

city	city
user_agent	user agent
cookie_age_seconds	the time in seconds since the user_id was created
detected_language	detected user language
device_type	the type of device on which the ad will be shown
device_brand	device brand
device_model	device model
screen_orientation	the screen orientation of the device when the ad request is sent
publisher_properties	properties of the publisher (obfuscated)
is_wifi	wifi indicator
viewability	an estimate of the likelihood that this slot will be viewable by the user. -1 indicates that viewability could not be estimated.
session_depth	total number of impressions served to this user in browsing session, plus 1. -1 indicates that session_depth could not be estimated.
auction_type	type of the auction
bidded	indicates whether we participated in the auction or not
bid	the bid we placed (in CPM)
won_bid	indicator whether or not we won the auction
feedback_bid	**in a first price auction**: when won, it's the 2nd highest bid in the auction. when lost, it's the winning bid. **in a second price auction**, it's the winning bid (in CPM).
conversion	whether the user clicked or not

1. Please summarize the main characteristics of the dataset. Feel free to add visualizations as well

2. Let's say we're interested in predicting the click-through rate. What are the most promising features for this task at hand?
   The expected outcome for this question is to code the selection process and provide a final set of features

3. You've been asked to provide a way where we can guess (pre bid) the bid that will allow us to win X% of the time for every given url & ad slot . Given this dataset, how would you solve it?

BONUS (not mandatory, it's just a glimpse to some of our challenges):

Assume there's an entity named Bidder which all it does is bidding, that is, for each request (impression) it returns a bid, i.e. float number greater than zero, or ignores the request. The number of Bidders can be one or more and each Bidder can have **its own pricing and bidding strategy**. For example, Bidder A can have a fixed pricing method, i.e. returns the same bid for each request, whereas Bidder B is more selective - it has a differential pricing method and will bid only when certain conditions are satisfied, for example, on previously known urls only.

You've been asked to design a bidding ecosystem for a completely new activity. It's up to you to decide what are the strategies for each Bidder, and (possible) interactions between them. Some of the core issues you may want to address are exploration, exploitation and/or pricing - but it's not mandatory.

How would you design such a system that will allow a synergy between different Bidders?
Can you think of components* that are crucial for such an ecosystem?

* you don't need to elaborate on the implementation of these components, but rather their functionality.

Good luck!