

INTRODUCTION

60% of the digital ad inventory is sold by publishers in **Real Time first price Auctions**.

Once a user lands on a webpage, *bidders (advertisers)* bid for different ad slots on the page and the one with the highest winning bid displays their ad in the ad space and pays the amount he bid. This process encourages bid shading – bidding lesser than the perceived value of the ad space to maximize utilization for self while maintaining a particular win rate at lowest prices.

Hence, for publishers, it becomes important to **value their inventory** (all the users that visit their website * all the ad slots they have on their websites) correctly so that a **reserve price, or a minimum price can be set up for the auctions**. The minimum price

PROBLEM STATEMENT

In a first price auction, the highest bidder wins and pays the price they bid if it exceeds the reserve price. The optimal strategy of a bidder is to shade their bids (bid less than their true value of the inventory). However, bidder needs to win a certain amount to achieve their goals. This suggests they need to shade as much possible while maintaining a certain win rate.

A bidder perceives a certain value out of every impression they win. Each bidder would like to maintain the value they derived out of this set of websites (given in the dataset) in June with a maximum deviation of 20%.

Setting a reserve price induces this by causing bidders to lose at lower bids which encourages higher bidding and more publisher revenue. However, since most of these takes place through automated systems, there might be an unknown delay in setting reserve prices & reducing win rate of bidder & bidder changing their bid shading algorithm & increased publisher revenue.

IMPORTANT TERMS:

- Publisher – person who owns and publishes content on the website
- Inventory – all the users that visit the website * all the ad slots present in the website for the observation period
- Impressions - showing an ad to a user constitutes one impression. If the ad slot is present but an ad is not shown, it falls as “unfilled impression”. Inventory is the sum of impressions + unfilled impressions.
- CPM – cost per Mille. This is one of the most important ways to measure performance. It is. Calculated as $\text{revenue/impressions} * 1000$. 'bids' and 'price' are measured in terms of CPM

DATASET :

The dataset provided to you has data for several websites owned by the same company and they are asking for your help for what should be their approach to set reserve prices and what is the range for reserve prices they should be setting for July. The data is only of the actual revenue generation and not at bid level. The dataset has the following columns:

1. Date
2. site_id : each id denotes a different website
3. ad_type_id : each id denotes a different ad_type. These can be display ads , video ads, text ads etc
4. geo_id : each id denotes a different country. our maximum traffic is from english speaking countries
5. device_category_id : each id denoted a different device_category like desktop , mobile, tablet
6. advertiser_id: each id denotes a different bidder in the auction
7. order_id : can be ignored
8. line_item_type_id : can be ignored
9. os_id : each id denotes a different operating system for mobile device category only (android , ios etc) . for all other device categories, os_id will correspond to not_mobile
10. integration_type_id : it describes how the demand partner is setup within a publisher's ecosystem - can be adserver (running through the publisher adserver) or hardcoded
11. monetization_channel_id : it describes the mode through which demand partner integrates with a particular publisher - it can be header bidding (running via prebid.js), dynamic allocation, exchange bidding, direct etc
12. ad_unit_id - each id denotes a different ad unit (one page can have more than one ad units)
13. total_impressions - measurement column measuring the impressions for the particular set of dimensions
14. total_revenue - measurement column measuring the revenue for the particular set of dimensions
15. viewable_impressions - Number of impressions on the site that were viewable out of all measurable impressions. A display ad is counted as viewable if at least 50% of its area was displayed on screen for at least one second
16. measurable_impressions - Impressions that were measurable by Active View out of the total number of eligible impressions. This value should generally be close to 100%. For example, an impression that is rendering in a cross-domain iframe may not be measurable.
17. Revenue_share_percent - not every advertiser gives all the revenue to the publisher. They charge a certain share for the services they provide. This captures the fraction of revenue that will actually reach the publishers pocket.

QUESTIONS –

1. What is the potential revenue range our publisher can make in July?
2. What is the reserve prices that he/she can set ?

REQUIREMENTS :

1. Answer file with potential revenue range, suggested reserve prices and logic behind choosing the same
2. Python/R code used to reach the above answer