REAL TIME FIRST PRICE AUCTIONS

Dataset used: Ascendeum_Dataset2.csv

The dataset contains the revenue for each adunit for the month of June,2019

Assumptions:
1. Advertiser is not known prior. So the advertiser input has not be taken.
2. The *'total_revenue'* has values upto 4 decimal points and hence it is multiplied by 100. Currency is assumed as $.
3. It is assumed that the publisher is auctioning the different versions of the *ad_unit* in different *geo_id, device_category_id, os_id, ad_type_id,* and *monetization_channel_id* separately. So the reserve price for each is found out.
4. The negative 'total revenue' value has not been removed.
   As the data's are joined together, the effect of same might not affect much.

Conclusions after understanding the data:
1. All the ad unit id's are independent and are not depend on site_id. But site_id might have an influence on the number of impressions.
2. The characteristics for an adunit and their category count include

Number of site_id	=10
Number of ad_unit_id	=132
number of ad_type_id	=2
Number of device_category_id	=5
Number of os_id	=7
Number of geo_id	=219
Number of intergration_type_id	=1 *(not considered as it same for all)*
Number of monetization_channel_id	=5
Number of revenue_share_percent	=1

   (not considered as all advertiser give full amount to publisher)

3. The characteristics of advertiser and number of categories include
   Number of advertiser_id   =23

   (Not used as the publisher will not be knowing the advertiser during reserve price setting)

4. The reserve price for an adunit should depend on the adunit characteristics and past impression/reserve price which are known prior.

The system developed should be able to predict the reserve price from the adunit characteristics and past impressions

System modelling:

Input variables

Categorical:

site_id

ad_unit_id

ad_type_id

geo_id

device_category_id

os_id

monetization_channel_id

Numerical:

Total Impressions

Viewable Impressions

Measurable Impressions

Output variable

CPM or reserve price

Regression model used:

XGBoost

Explanation:

A system has been modelled taking the ad characteristics and the impressions obtained using XGBoost regressor. The system has been trained to predict the reserve price.

However, from the practical knowledge we know that the reserve price for an ad location in a website should not be set below the already generating revenue and the predicted price should not be too high as indicated in the information given.

So the predicted price if it is less than the current amount it is updated to the current amount and if the predicted amount has deviation of more than 20% from the current amount it is set to the 20% margin amount.

The predicted revenue range for July is found out using the model output obtained and the model output altered for the practical purpose.

QUESTIONS –

1. **What is the potential revenue range our publisher can make in July?**

    *The revenue generated during the month of June is $3956311.92/-. The potential revenue range our publisher can make in **July** is between **$3941315.10/-** to **$4213663.11/-** according the modelling performed.*

2. **What are the reserve prices that he/she can set?**

    There have been 40687 ad configurations including the different configurations of geo_id, adtype_id, device_category_id, os_id, monetisation_channel_id (given in assumptions). The reserve price for the same has been given in the csv file attached with email.