DS home assignment

Prediction task

(a)

Attached are two CSV files: "train" and "test". The target variable that we want to predict is *org price usd following 30 days*. Your task is to send predictions on the test file and minimize the root mean squared error (RMSE). **Do not** use the *treatment* or *org price usd following 30 days after impact* columns as features.

Note: Please send us the results in a .py file that contains only code that contributes to the RMSE loss.(Unfortunately we won't have time to review all the things you've tried). Also, please be careful not to overfit as the test set contains the target variable for convenience, and we will test the loss in a similar test set with different dates.

(b)

What are the three most important features that contributed to the prediction?

Recommendation task

We are interested in increasing the target variable. For that, we ran a randomized experiment where we assigned to 50 percent of the population 10 dollar offer, and 2 dollar to the rest (in the "treatment" column). The experiment yielded a new target variable, *org price usd following 30 days after impact* that showed the result of the experiment in terms of income.

(a)

Using the test data, set the treatment that would maximize the target variable per user. Add a column for each user that you think would be the best treatment for that user (either 10 or 2), and please include the code that created that column.

(b)

What are the three most important features that contributed to the decision to give users a specific treatment?