# Information about data used in "Setting reserve prices in second-price auctions with unobserved bids"

Jason Rhuggenaath, Alp Akcay, Yingqian Zhang
Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology,
{j.s.rhuggenaath, YQZhang, a.e.akcay}@tue.nl

Uzay Kaymak
Jheronimus Academy of Data Science, Eindhoven University of Technology, u.kaymak@ieee.org

---

## Data used in Experiments

This document contains a description of the datasets used in the paper "Setting reserve prices in second-price auctions with unobserved bids" by J. Rhuggenaath, A. Akcay, Y. Zhang and U. Kaymak which has been submitted to INFORMS Journal on Computing.

### Ebay Data

The experiments in Section 6 of the main text and Section EC.3 of the Supplementary Materials make use of an eBay dataset that was used in Mohri and Medina (2016). The original data set can be found at the following website: `http://cims.nyu.edu/~munoz/data`. In the paper we performed a number of pre-processing steps on the original data and the pre-processed data is subsequently used in our experiments. The file `ebay_data_clusters_10.csv` contains the data after performing the clustering as described in Section 6. The experiments in Section EC.3 are based on a different clustering of the bids. The file `ebay_data_clusters_6.csv` contains the data after performing the clustering as described in Section EC.3.

Each file has three variables: `SB`, `TB`, `cluster_id`. The variable `TB` indicates the highest bid, the variable `SB` indicates the second highest bid and the variable `cluster_id` indicates the cluster in which the bids are placed.

## Header bidding Data

The experiments in Section 7 of the main text and Section EC.4 of the Supplementary Materials make use of a dataset that was provided by our industry partner. In the paper we performed a number of pre-processing steps (which are detailed in Section 7.1) on the raw data and the pre-processed data is subsequently used in our experiments. The files `data_website_A.csv` and `data_website_B.csv` contain the data after performing the pre-processing steps.

Each file has three variables: `max_bid_scaled`, `hour`, `second_bid_scaled`. The variable `max_bid_scaled` indicates the highest bid, the variable `second_bid_scaled` indicates the second highest bid and the variable `hour` indicates the hour at which the bids are placed.

## References

Mohri M, Medina AM (2016) Learning algorithms for second-price auctions with reserve. *J. Mach. Learn. Res.* 17(1):2632–2656.