

Dataset

Spring 2018 Sports Marketing Analytics Contest: Season ticket churn

Description: The data will be presented in the form of an Excel spreadsheet. The spreadsheet will consist of four tabs.

One tab will contain the data dictionary, providing a description of the variables in the dataset.

The calibration tab contains Year 1 customer data for all 4608 customers that defected in Year 2, a random sample of 4608 customers that renewed in Year 2, and a vector indicating whether the customer defected or renewed in Year 2. This balanced sample of defections and renewals allows us to mask the team's actual renewal rate, while taking away nothing from the analytical task and the resulting insights.

The validation tab is similar to the calibration tab, with the following differences. The customer data now corresponds to Year 2 data, both for the 4771 customers that defected in Year 3, and a random sample of 4771 customers that renewed in Year 3. The defection/renewal data for Year 3 is not included – predicting those values is the analytical task of the contest.

The submission tab will consist of three vectors. Column A will contain the IDs for the validation sample. Teams will insert their scores for each customer in the validation sample into Column B. After applying a cut-off value to the scores, teams will populate Column C with their predictions, 1=defect, 0=renew.

Availability: The dataset will be available for download from Canvas once the team is registered and the invitation to join the Canvas course has been accepted. The dataset will remain available on Canvas through Wednesday, April 18.