

For our University of the Pacific Masters Capstone Project, we partnered with the Sacramento Kings and they wanted us to find out why customers who had already purchased tickets were not showing up to the games. Our team of four students decided to attempt to model ticket-holder attendance in order to determine what the most important factors were that influence attendance. All we were given was email, marketing, and transaction data for the previous two seasons, and the rest was up to us. After cleaning and understanding the data we were given, the most involved part of this project for us was the custom feature engineering. For that, we had to consider every potential reason why someone who had purchased a ticket would not show up to the game, and then find reliable data that would accurately capture those potential reasons. We came up with many features, ranging from weather on game day to player injuries to how good your seat was. Once we had all of our features, we had to decide on a modeling approach. We decided to each explore a different machine learning method, apply that method to our data, and see which one (if any) gave us the best results. After many iterations, our best performing model was able to predict ticket-holder attendance with ~92% accuracy, which the Kings were very pleased with. Please note that I have left out the specifics about what features were found to be the most important, as well as details about our final model due to the NDA agreement that we had with the Kings.

#### My Role in Capstone Project:

- Involved with initial data cleaning and identifying and removing outlier data points
- Engineering of all basketball-related features (6 in total)
  - o Used jersey sales data to compile a list of most popular players in the league
  - o Used basketballreference.com data to determine how many popular players were on each team, and whether or not most popular players from each team played in the game or not
  - o Used basketballreference.com data to determine power rankings and current win/loss streaks for both teams
- Learned about random forest modeling approach and applied it to our data in many different variations. Then ran a grid search to determine the best performing random forest model
- Used Tableau to create visualizations that we used in our final presentation
- Presented our findings to Sacramento Kings Business Operations team