SEIS 763 Team 7 – Assign4 – Project Proposal

**Dataset Being Used:**

* 10k instances
* League of Legends EUW Diamond Solo-Queue Ranked game status at 10 minutes

<https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min>

**Class Imbalance:**

* Win/Loss variable is nearly 50/50 between Red and Blue win

**Attributes:**

* Dataset contains 36 numerical variables and 4 categorical variables (including the target)
* More variables can be obtained from Riot’s Developer API (<https://developer.riotgames.com/apis>)

**Exploration:**

* Outliers: more than 1% of the data in the columns blueWardsPlaced, blueWardsDestroyed, blueTowersDestroyed, redWardsPlaced, redWardsDestroyed, and redTowersDestroyed are outliers with a z-score > 3.
  + All the outliers in these columns are positive, so they are all right-skewed.
  + All these columns have more than 25 outliers with a z-score>5 (>.25% of the column)
* No missing values

**Hypothesis:**

* A team that prioritizes map/global objectives (herald & dragon) early tends to win.

**ML Techniques:**

* We plan on using a variety of classification techniques in an ensemble such as k-NN, SVM, Naïve Bayes, and Decision Trees & Random Forest.