This contains all bets placed for a certain slot game for 3 days in May 2020. Each row denotes a single, unique bet.

**Data Dictionary**

gameutctime: Timestamp of when the bet occurred

betWager: Amount of money that the player puts up for the bet

betPayout: Amount of money that the player receives from the bet (this will be 0 in most cases because most bets do not pay out) (if betPayout < betWager he has effectively lost money on that bet and vice versa)

sessionBalance: Amount of money that the player sees in his account after the betWager and betPayout amounts have been logged in the system (the session balances from bet to bet don’t always reconcile on wagers and payouts alone because sometimes the player can debit/credit funds from/into his account) -> debit/credit as a feature

userid\_masked: unique identifier of a player

sessionid\_masked: unique identifier of a session (the last bet in each session is the ‘session-ending’ bet)

**Task**

We want you to tell us:

**What kind of experiences/triggers leads to an increased likelihood of a session ending?**

You’ll be evaluated based on the following:

* **How you structure your analysis**
* **Critical thinking/innovation/breadth and depth of thought when spotting causes/factors/predictors**
* **How you decide which causes/factors/predictors are important/significant/predictive**

Your submission can be in the form of a combination of the following:

* Write-up/report
* Slide deck
* Charts
* Anything else that helps you communicate your insights

**Approach**

* Data cleaning
* Feature engineering based on themes
* Exploratory Data Analysis
  + Visualization
  + Feature engineering
  + Hypothesis testing
* Explanatory Modeling (Why is this the last bet?)
  + Logistic regression
    - No train test split
    - P-value
    - Log odds
* Predictive Modeling (Will this be the last bet?)
  + Model per user OR Global model for all users
    - We will go with building models tuned to an average population as most users only have 1 session of data
  + Random forest
    - Feature importance
    - Limitation: model-dependent
  + Propensity to end bet model
    - Logistic regression
  + Performance metrics
    - Precision (False positive is more costly)
    - Recall (False negative is more costly)
    - F1 (Both false positive and false negative are costly)
    - Dummy predictor = Predict all to be positive (session-ending bets)