

## Результаты EDA, процесс обучения, оценки модели

- сформирован датасет, который содержит только те строки, где от регистрации пользователя до начала игры прошло не более 24 часов.
  - проверка на типы признаков = категориальных признаков не обнаружено => заменять на численные не нужно
  - проверка на пропущенные значения = таковых не оказалось => никаких удалений и замен в строках не нужно производить
  - построение тепловой карты = избыточных признаков не обнаружено => удалять никакие признаки не нужно пока
  - построение паирплота для наглядного отображения зависимостей/распределений признаков
  - так как признаки типа `datetime64[ns]` могут мешать при обучении моделей, они были удалены, а их информативность сохранилась в новом признаке, который равен отношению времени регистрации пользователя к времени начала игры
  - проверка на соотношение классов целевой переменной = обнаружен дисбаланс классов => было принято решение использовать в качестве оценки модели ROC AUC
  - проверка на корреляцию признаков с целевой переменной = признаки "round", "winner" имеют очень маленький коэфф корреляции – слабую линейную зависимость = удалила их, чтобы модель уделила больше внимания и ресурсов первостепенным признакам¶
- 
- датасет разбила на обучающую и тестовую выборки
  - для обучения была выбрана модель `LGBMClassifier`
  - при помощи сетки нашла наилучшие значения некоторых гиперпараметров
  - при помощи кросс-валидации проверила качество модели метрикой ROC AUC
  - обучила модель `LGBMClassifier`
  - предсказала значения