Результаты EDA, процесс обучения, оценки модели

- сформирован датасет, который содержит только те строки, где от регистрации пользователя до начала игры прошло не более 24часов.

- 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.

- затем сделана группировка по юзеру и агрегация, в результате чего каждый юзер получил новые характеристики, а именно:

```
Index(['user_id_', 'length_min', 'length_max', 'magic_used_min',
       'magic_used_max', 'magic_used_mean', 'player_cards_min',
       'player_cards_max', 'round_count', 'round_max', 'datetime_max',
       'datetime_mean'],
      dtype='object')
```

-проверка на корреляцию признаков с целевой переменной = признаки, которые имели низкий коэфф.корреляции были удалены (на скриншоте выше уже с удаленными признаками – из названий понятно значение и способ получения каждого)

- 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

- проверка на соотношение классов целевой переменной = обнаружен дисбаланс классов => было принято решение использовать в качестве оценки модели ROC AUC

– 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, избежав переобучения

- обучила модель LGBMClassifier

- предсказала значения