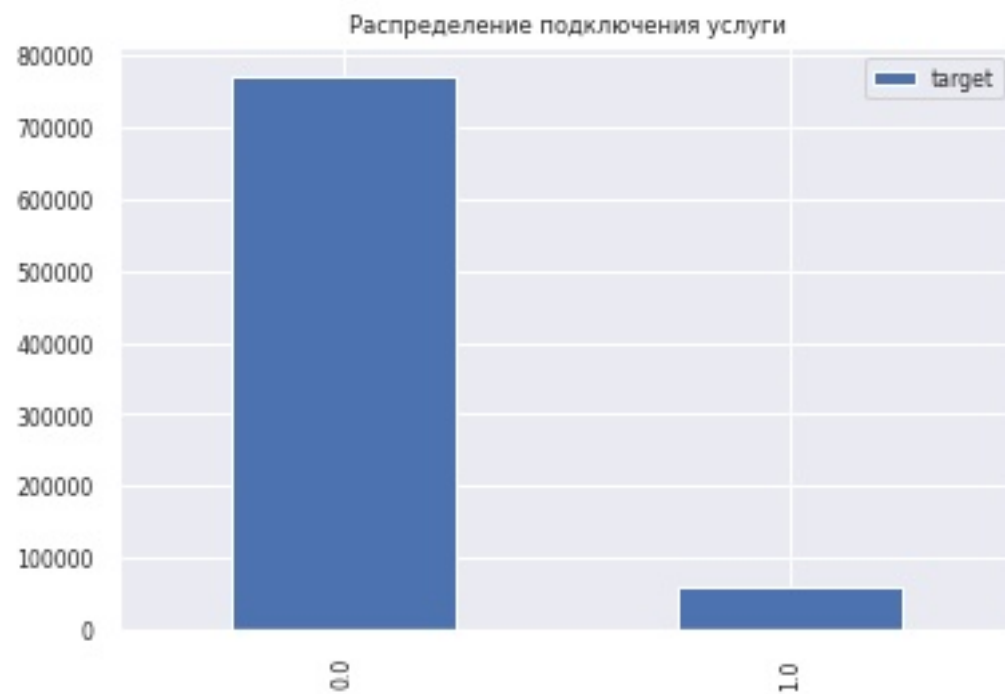
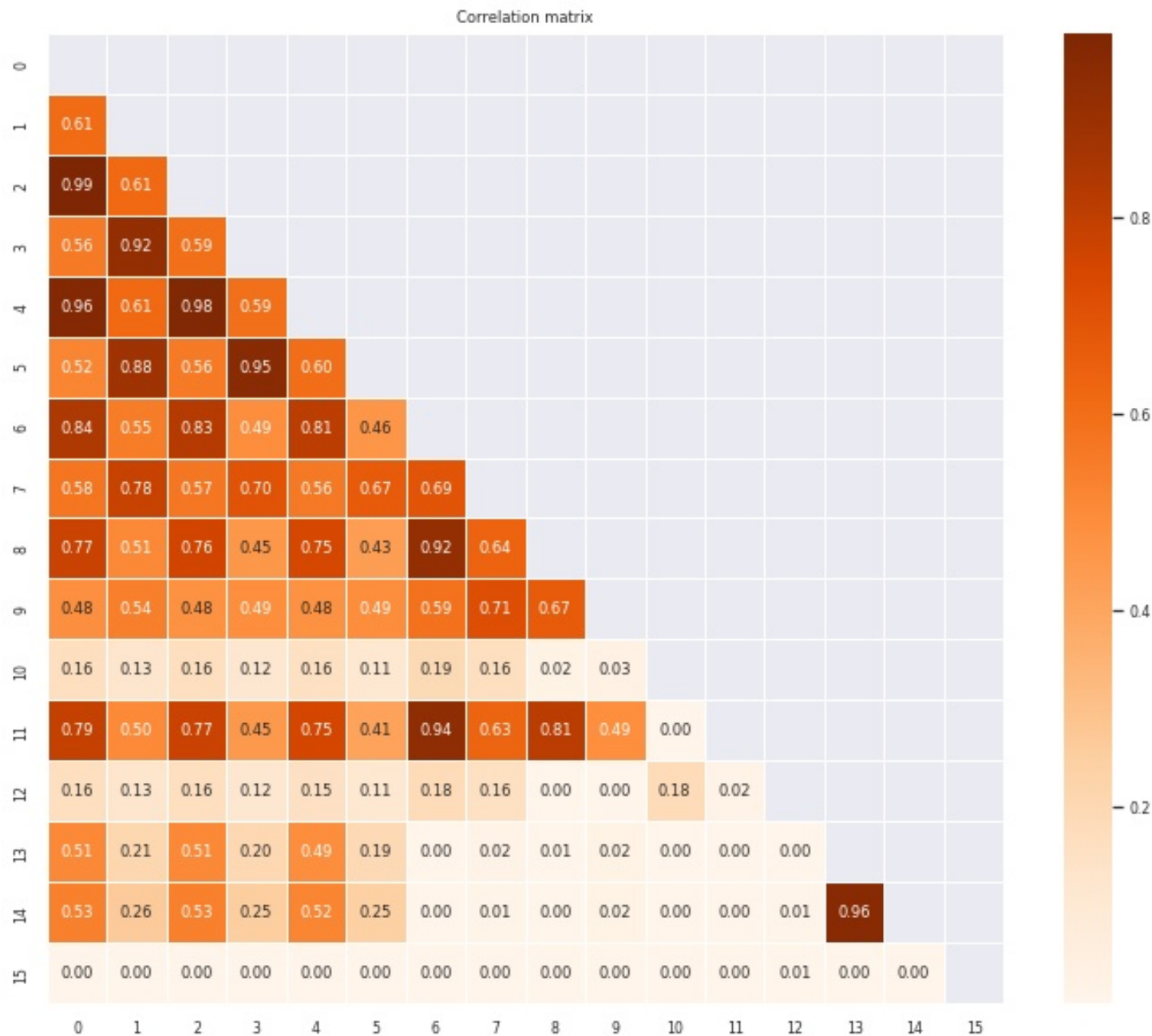


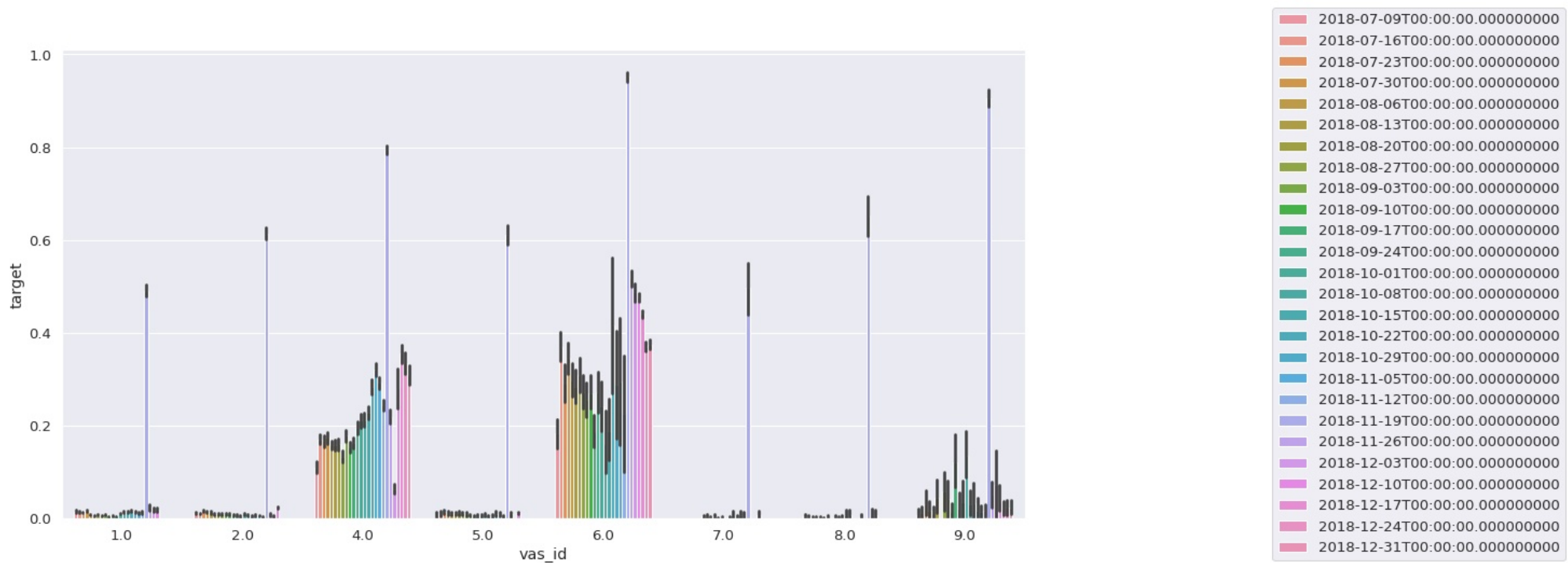
Анализ таймсерии показал, что имеется аномальный пик на последних числах ноября. Попытка исключить выброс из серии положительного эффекта не дало



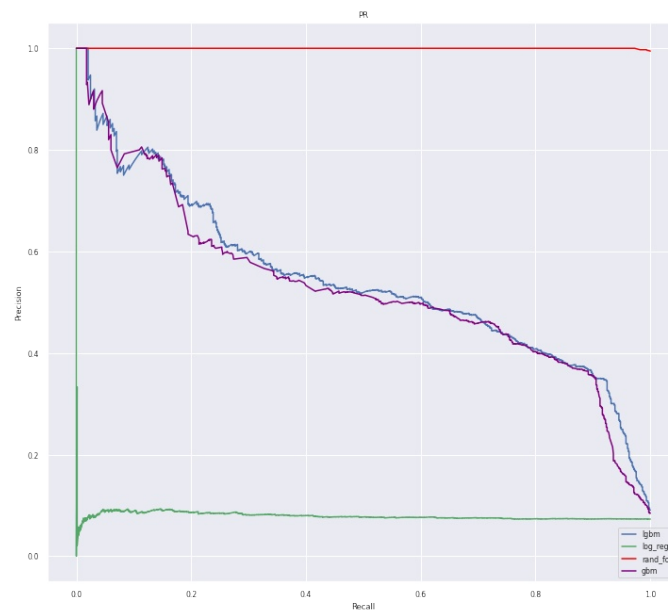
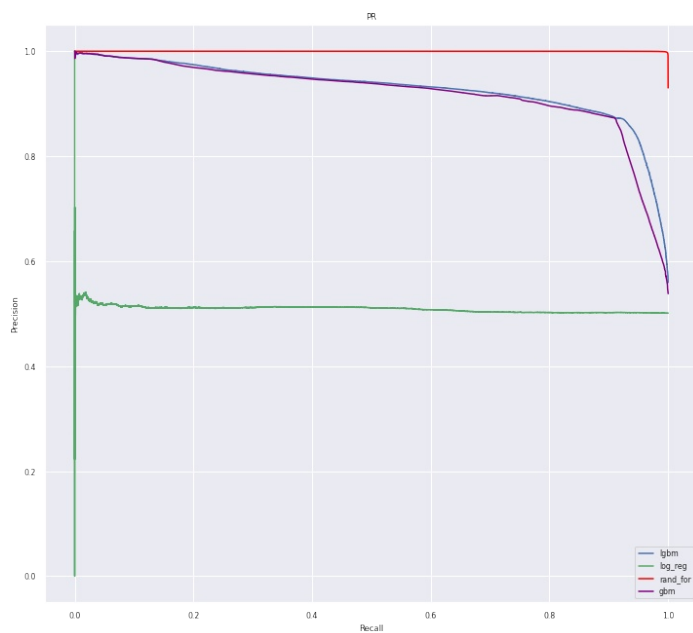
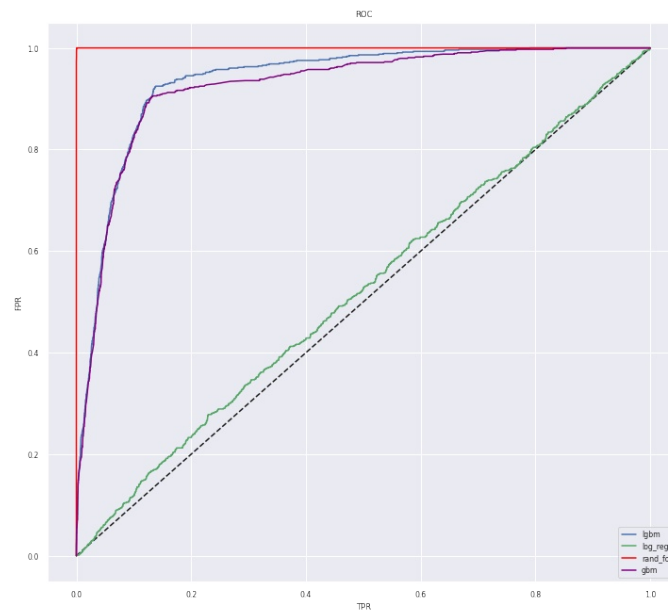
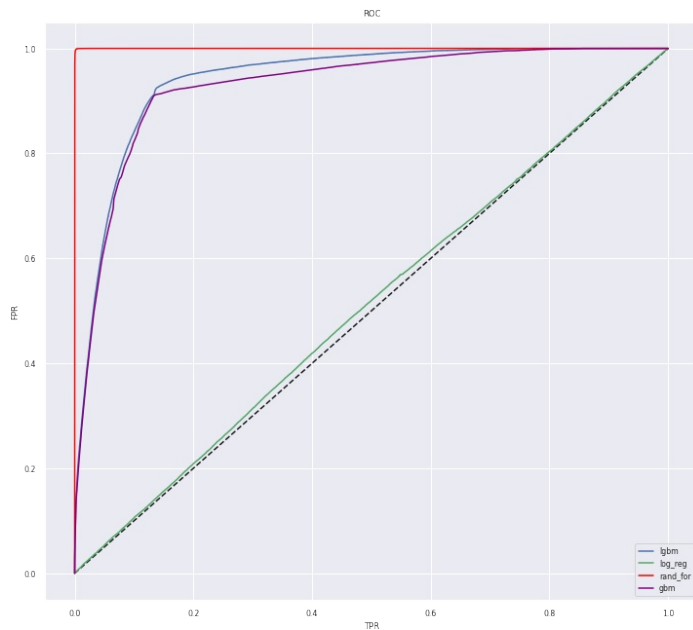
Видим наличие явного дисбаланса целевого класса. Имело смысл использовать оверсэмплинг



Многие фичи
имеют линейную
зависимость друг с
другом. Подготовка
датасета включала
в себя очистку от
столбцов с
корреляцией выше
пороговой



Наличие явных лидеров, среди услуг, предлагаемых пользователям позволило построить бейзлайн с достаточно высокой целевой метрикой



Из за
оверсэмплинга
модели получились
переобученными,
но даже в этом
случае показывали
метрики, лучше,
чем на базовом
датасете.

Из-за скорости
работы и одного из
лучших скоров была
выбрана модель
LightGBM с
небольшой
глубиной
решающих
деревьев для
снижения
переобучения