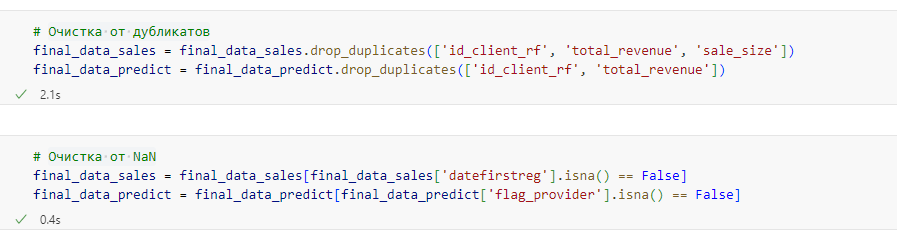
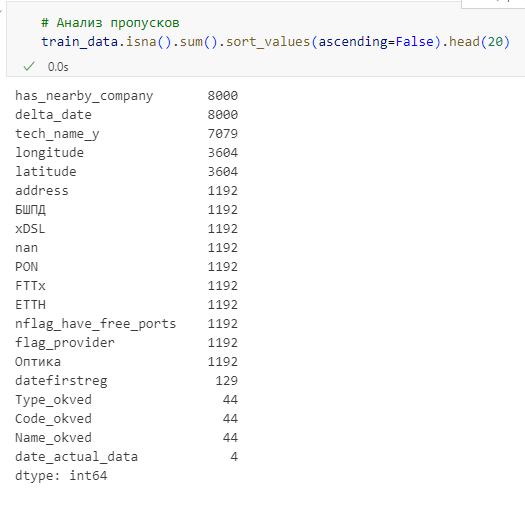
Вопросы к скрипту:

1. Зачем удалять дубли. Решили вроде что филиалы по одному клиенту не будем удалять:  
   
2. В train\_data остались NaN. Насколько это критично?  
   
3. Оставляем только первые 2 цифры Code\_okved – done
4. У нас в данных по клиентам нет ИНН. Поэтому подтянуть внешние источники не понял, по какому ключу.