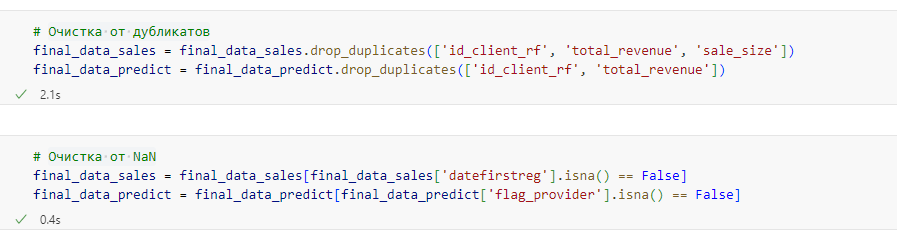
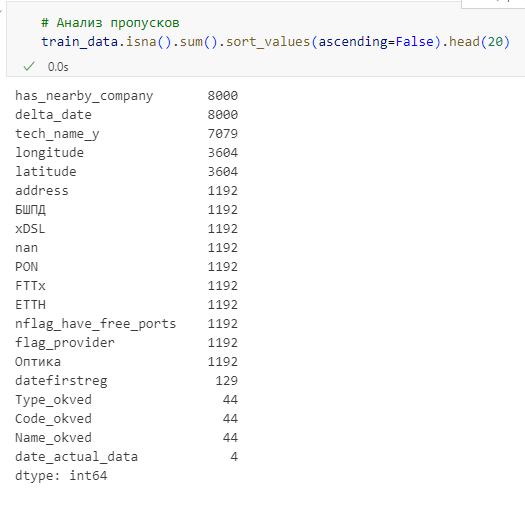
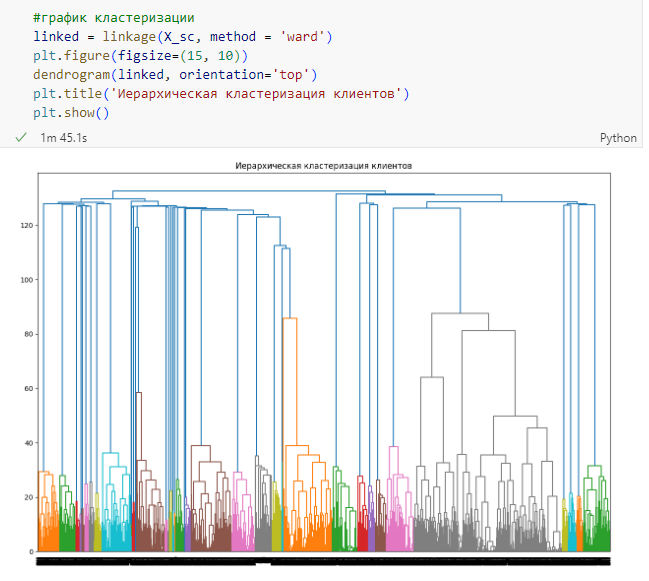
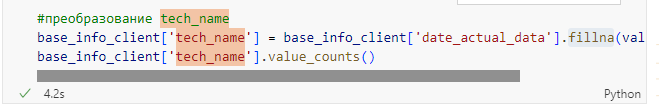
Вопросы к скрипту:

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.
5. Кластеризация по дендраграммам по методу Варда вроде бы используется, когда небольшое кол-во кластеров:  
   
6. Для чего в base\_info\_client создается поле 'tech\_name' записывается date\_actual\_data? Или это опечатка?



1. Касательно Силуэта не был импортирован метод:  
   from sklearn.metrics import silhouette\_score

Я вставил еще другие методы определения числа кластеров. Но я пока не силен в этом анализе, может тебе поможет.