**Оптимизации**:

1. Исходя из поставленной задачи заметим, что сначала можно выяснить код страны United Kingdom, а затем по найденному коду фильтровать страны.

Таким образом исключаем формирование двух DF, DS или RDD, которые затем ещё и потребуется соединить.

2. Наверное можно ли считать автоматической оптимизацией то, что анализатор при фильтрации автоматически исключает пустые поля, заменяя условие

(Date#42 = 2018-12-01) AND (CountryCode#48 = 36)

на условие

(isnotnull(Date#42) AND isnotnull(CountryCode#48)) AND ((Date#42 = 2018-12-01) AND (CountryCode#48 = 36))

3. В ходе экспериментов выяснилось, что в использовании DataFrame и RDD подсчетах продуктов и транзакций нет смысла, потому что данных не много, и нехватки памяти не происходит.

Продукты и транзакции собраны в отдельные векторы.