

Тестовое задание для кандидатов на бизнес-аналитика

Тестовое задание состоит из набора данных по которым нужно построить отчет и вопросов на которые нужно ответить.

Исходный датасет содержит в себе сессии (session_id - id сессии, time - время создания записи) сбора данных Озона за 1 мес. в категории “компьютеры и комплектующие”. Сессия представляет из себя сбор различных параметров продукта (id - id товара, title - наименование). Сырые данные из Ozon - Возможны выбросы, неполнота данных, некорректные значения*

<https://disk.yandex.ru/d/QlOTpva6Wa4eQw>

Для отчета:

1. Рассчитать продажи за день по каждому товару
2. Выявить топ 10 брендов по выручке, топ 10 продаваемых товаров
3. Выявить топ товаров, приносящих 80% выручки
4. Построить график продаж от времени (опционально)
5. Оценить качество данных
6. Объяснить полученную картину:
падения и пики на графике продаж
как можно улучшить расчет продаж на текущих данных?

Общие вопросы:

1. Как будете заполнять пропущенные значения в данных?
2. Что делать, если датасет несбалансированный?
3. Каковы потенциальные причины аномалий в наборе данных?
4. Что делать, если данных не хватает или они плохого качества?

*Возможны выбросы, неполнота данных, некорректные значения

*in_stock – это остатки, а продажи нужно рассчитывать через их разность.

Готовое тестовое отправлять на db@talentpowered.team