

Задание

Вам даны таблицы с исходными данными о движении лидов (от поступления до продажи). Необходимо агрегировать эти данные и представить их в единой трансформированной таблице, где содержится:

1. Количество полученных лидов
2. Количество лидов, отправленных на модерацию
3. Количество лидов, отбракованных модераторами
4. Количество лидов, переданных юристам
5. Количество отбракованных юристами лидов (с компенсацией денег за брак)
6. Количество фактически проданных лидов*
7. Gross Revenue (выручка за проданные* лиды)
8. Конверсия в проданный лид (CR)
9. Выручка за проданный лид (APL)
10. Выручка за полученный лид (EPL)
11. Показатели 1-9 в разрезе по регионам (Москва и МО/СПб и ЛО/Остальные регионы).
12. Показатели 1-9 в разрезе по поставщикам лидов (Pravoved/CPA/Satellites).

Проданные лиды находятся как разность между лидами, переданными в продажу и лидами, отбракованными юристами (с компенсацией денег за брак)

Описание таблиц

Статусы лидов **'test_data1.csv'**. В ней могут встречаться дублирующие события по одному и тому же лиду (это не нормально). Также нужно исключать из выборки все события по лидам, с которыми произошло событие **'val_lead_absorbed'**.

- 'id' - номер события (всегда уникален, даже у дублей)
- 'event' - статус/событие по лиду
 - 'val_new' - получение лида
 - 'val_lead_absorbed' - у лида создан дубликат с другим номером
 - 'val_lead_moderation' - лид был отправлен на модерацию
 - 'val_lead_valid' - лид одобрен отделом модерации
 - 'val_lead_invalid' - лид отбракован отделом модерации
 - 'val_lead_bought' - лид передан юристу
 - 'val_lead_defected' - переданный лид отбракован юристом
- 'event_date' - дата и время события
- 'entity_id' - уникальный номер лида
- 'region_id' - код региона
 - 55 и 77 - Москва и МО

- 47 и 78 - СПб и ЛО
- 'user_registered' - наличие пользователя в числе зарегистрированных (1 - да, 0 - нет)
- 'user_id' - уникальный номер пользователя
- 'price' - цена лида
- 'is_refunded' - были ли возвращены (компенсированы) деньги за лида при его отбраковке юристами

Таблицы **'test_data2(reg_users).csv'**, **'test_data3(unreg_users).csv'** содержат данные по поставщикам лидов в зависимости от того, принадлежит ли пользователь к группе зарегистрированных пользователей.

- 'user_id' - номер пользователя
- 'invitor_id' - номер поставщика лида

Таблица **'test_data4(sources).csv'** содержит информацию о поставщиках лидов.

- 'invitor_id' - уникальный номер поставщика лида
- 'role' - роль поставщика лида
 - 144115188075855896 - CPA
 - 144115188075855947 - Satellites
 - все остальное - Pravoved