

3. В задаче сквозной аналитики вам предоставили данные по местоположению пользователей. Т. е. для каждого `user_id` известна последовательность координат (широта/долгота), когда они требовались приложению для полноценной работы. Как бы вы добавили эти сведения в таблицу визитов и покупок? Для составления ответа можно использовать вопросы:

- У каждого пользователя известен набор координат. А для связывания с визитом или фактом покупки скорее всего потребуется одно-два числа. Как их получить?
- Наборы координат одного и того же пользователя могут быть значительно удалены друг от друга. Как это отразится на вопросе расчета пункта 1?
- Какие дополнительные признаки можно получить из координат? Ведь это просто числа, которые сами по себе мало что дают.

Обработка экономических показателей скорее всего будет храниться отдельно от координат, поэтому хранить два набора данных - (визит/покупка) и координаты – необходимо в отдельных таблицах (файлах, базах) – для скорости вычисления экономических метрик и экономии оперативной памяти.

Минимальный набор полей:

Набор координат

user_id	coord1	coord2

Данные о визитах и покупках

user_id	product	cost
		<i>0 = визит</i>
		<i>!= 0 стоимость покупки</i>

Для идентификации (потенциального) клиента обе таблицы содержат код клиента (`user_id`). подробные данные о пользователе – в другой таблице.

Множественность посещений, покупок клиентом и сложность обработки дублирующихся строк требуют дополнительных полей, например, дата-время. Возможно, дополнительно и удобней иметь идентификатор посещения/покупки (индекс) - event_id.

Набор координат

user_id	coord1	coord2	date	event_id

Данные о визитах и покупках

user_id	product	cost	date	event_id

Фактически наличие одного поля event_id в обеих таблицах достаточно для однозначного соотнесения события (визит, покупка) и места расположения клиента (координат) и, соответственно, по этому полю объединять таблицы.

Дополнительные признаки из координат:

- близость клиента к местам продаж, складов (среднее расстояние – может быть использовано в качестве показателя вместо координат)
- мобильность клиента (количество мест, из которых клиент обращается в он-лайн магазин)
- плотность клиентов по регионам – выбор мест для наружной рекламы