**Приведение к нормальным формам:**

- БД в изначально данном формате (разделенном на таблицы «transaction» и «customer») требованиям 1НФ уже соответствовала – все атрибуты простые, все сохраняемые данные по ячейкам скалярны (есть столбец «address» в таблице «customer», формально состоящий из номера дома и улицы, но в контексте задач создаваемой БД возможно считать адрес в таком формате одной цельной единицей, не нуждающейся в дополнительном делении).

- 2 НФ – зададим ключи, по которым однозначно идентифицируются строки (и проконтролируем, что значения в строке зависят от всего ключа). Т.к. задавать составные ключи не планируется, то правило про зависимость от всего ключа выполняется автоматически, осталось определить сами ключи. Глобально будущая БД охватывает три сущности (3 таблицы): покупатель, продукт и связывающая их транзакция. Транзакции однозначно идентифицируются по transaction\_id т.к. поле уникально для каждой транзакции. Покупатели также однозначно идентифицируются по customer\_id т.к. поле уникально для каждого покупателя. Идентифицировать товар сложнее, т.к. можно заметить, что одному product\_id соответствуют явно разные товары (вплоть до того, что у них разные бренды), и ни один другой относящийся к товару столбец не идентифицирует продукт однозначно, поэтому создадим синтетический ключ под каждое уникальное описание товара (product\_unic\_id).

- 3 НФ – формально по итогам приведения ко 2НФ все 3 таблицы (customer, product, transaction) уже находятся и в 3НФ – отсутствуют транзитивные зависимости:

в таблице *transaction*: product\_unic\_id и customer\_id – это связи с другими таблицами базы, а transaction\_date, online\_order, order\_status, list\_price зависят только от transaction\_id;

в таблице *customer*: first\_name, last\_name, gender, DOB, job\_title, wealth\_segment, deceased\_indicator, owns\_car, address однозначно определяются только личностью конкретного покупателя и никак не зависят друг от друга; job\_industry\_category от job\_title зависит слабо (и в банках и в больницах работают люди очень разных профессий и, например, разработчика сейчас можно встретить в любой сфере) поэтому не выделяем в самостоятельную таблицу; теоретически можно вынести столбцы с составляющими адреса (почтовый код / штат / страна) в самостоятельную таблицу, но в данном случае это видится не очень полезным (мы не планируем, например, вносить массовые изменения в адреса, это просто часть анкетных сведений о клиентах) и допустимо оставить столбцы почтовый код / штат / страна в таблице т.к. они скорее все вместе формируют адрес (который является личной характеристикой клиента), чем как-то зависят друг от друга;

в таблице *product*: product\_id, brand, product\_line, product\_class, product\_size, standard\_cost являются частными характеристиками продукта, т.е. каждый из них зависит от одного product\_unic\_id, а не друг от друга. Но тем не менее выделим из этой таблицы в отдельные таблицы-справочники product\_line, product\_class, product\_size - эти данные явно стандартные по магазину и может возникнуть необходимость переименовать, например, продуктовую линейку (или ввести дополнительное деление), поменять таблицу размеров и т.д. (+будем хранить в большой таблице продуктов int-ы вместо занимающих больше места текстовых форматов).