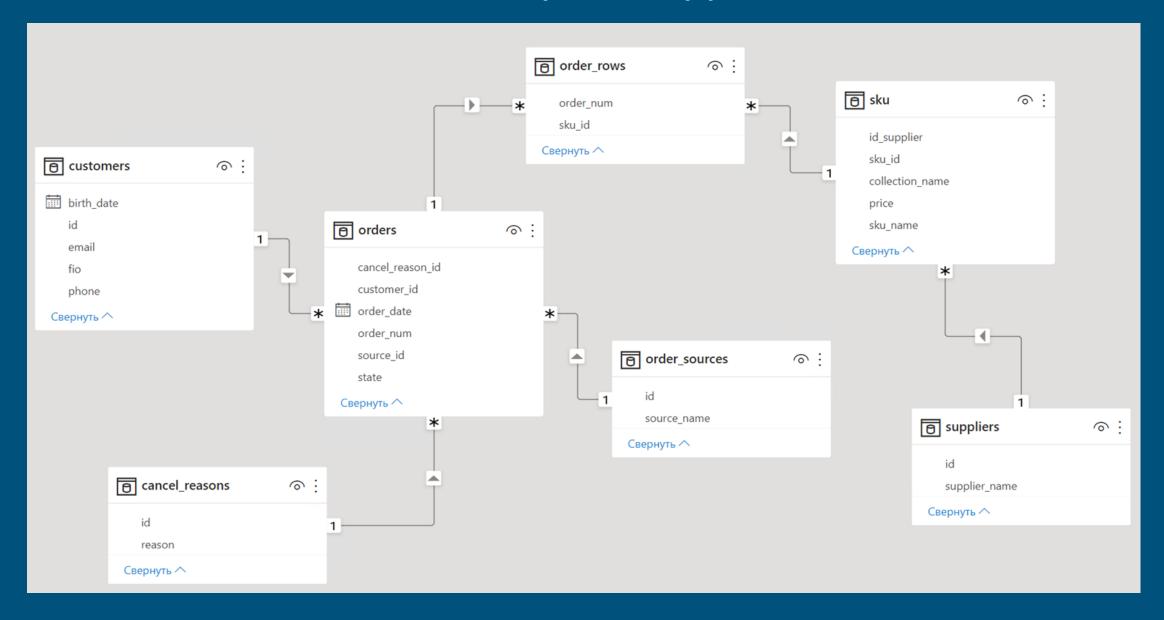
Отчёт компании **Бест Мебель»** за 2021 год

Исполнитель: Ярослав Филиппов

Описание задачи

- Поступил заказ проанализировать и визуализировать данные по продажам за 2021 год.
- Большая часть данных была предоставлена заказчиком в виде СУБД типа SQLite mebel.db.
- Кроме того, была предоставлена 1 таблица в виде csv-файла discounts.csv и 1
 Excel-таблица Ozon-Nov-Sale.xlsx.
- В ходе личного общения с представителем заказчика выяснилось, что качество данных в базе не на 100% идеально, в частности есть пропуски в справочнике номенклатуры.
- Представитель заказчика особо подчеркнул необходимость проанализировать причины отмены заказов.
- Также была предоставлена схема таблиц базы данных.
- Кроме того, представитель заказчика попросил, если это возможно, оценить изменение выручки, если бы клиентам было предоставлена дополнительная скидка за 6-й заказ в размере 250 рублей.

Схема таблиц базы данных



Подготовка календаря рабочих дней

Выполнение задания началось с календаря рабочих дней. Он был мной сделан еще до получения данного заказа, но я изложу процесс его создания в рамках данной презентации.

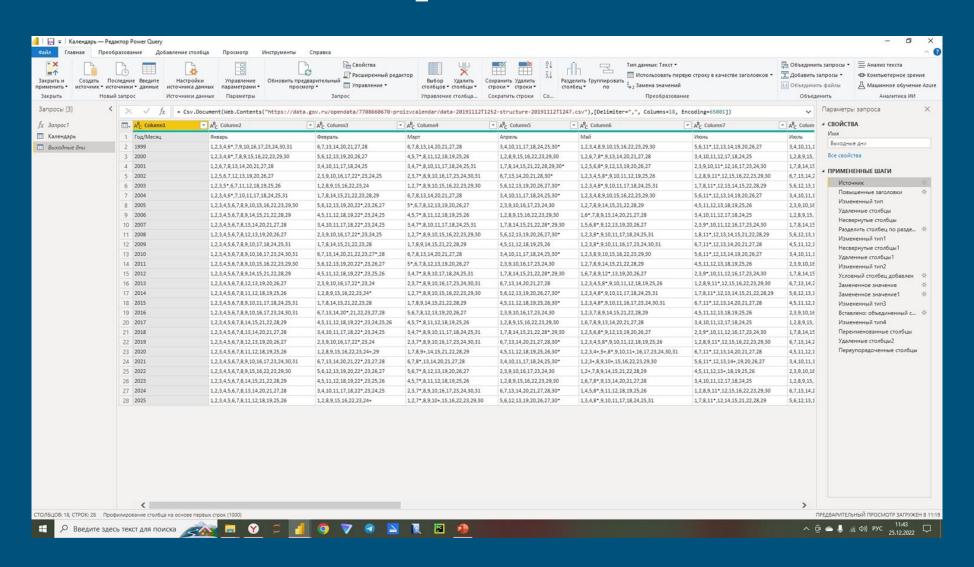
Исходный файл был скачан с сайта https://data.gov.ru/ и представлял собой сsv-файл.

После загрузки этого файла стало понятно, что не получится его использовать по назначению без трансформации данных.

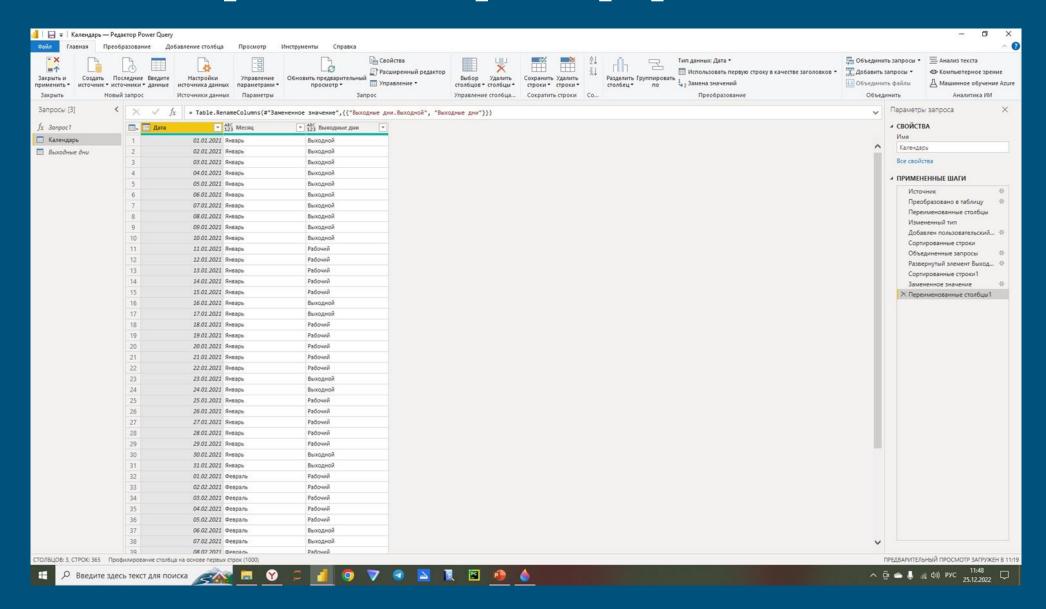
С помощью Power Query необходимые трансформации были сделаны и таблица сформирована. Более подробно с этим этапом работы можно ознакомиться в файле Календарь.pbix.

Для целей текущего заказа мне осталось только удалить данные за все годы кроме 2021-го и переименовать сокращенные дни в рабочие, так как в рамках данного заказа разница не имела значение.

Производственный календарь до обработки



Календарь после трансформации данных



Вариант №1

Обработка данных средствами Power BI, визуализация и анализ результатов.

Обработка данных и создание дашбордов

Следующим этапом была произведена загрузка данных в Power Bl и предварительная их обработка в Power Query:

- В таблице customers был добавлен столбец пол, заполненный в соответствии с последней буквой отчества клиента, предварительно были удалены все пробелы.
- В таблице sku был добавлен столбец Категория, заполненный на основании данных из столбца sku_id (использовал сочетание букв до дефиса). Столбец Коллекция имел пропуски поэтому он был полностью переписан на основании данных из столбца sku_id (использовал букву после дефиса).
- В таблице orders был добавлен столбец Город на основании данных из столбца order_num (использовал первую букву).

Обработка данных и создание дашбордов

Затем уже с помощью языка DAX были сформированы:

- Вспомогательная таблица Сумма заказа, в которой данные из таблицы order_rows были агрегированы по номеру заказа.
- Вспомогательная таблица Количество заказов в которой данные из таблицы order были агрегированы по номеру клиента.
- В таблице customers были добавлены столбцы Возраст, лет и Возрастная группа.
- Были прописаны несколько мер и построены 3 дашборда. Подробнее с этой работой можно ознакомиться в файле bestmebel.pbix.

На основании дашбордов были сделаны следующие выводы:

- Самая распространенная причина отмены заказов отсутствие товаров на складе. По этой причине произошло почти 2/3 отказов. Почти четверть отказов произошло по причине низкого качества товара.
- Кресло Берг имеет самый большой процент возврата. Кресло Спейс напротив имеет минимальный процент возврата.
- Динамика продажи показывает почти линейный рост, если учесть сезонное снижение летом – период отпусков.
- В Москве продажи ожидаемо выше, чем в Санкт-Петербурге.

На основании дашбордов были сделаны следующие выводы:

- Доля выручки в выходные в среднем примерно соответствует доле этих дней в году. Но в онлайн магазинах эта доля на несколько процентов выше. А в оффлайн ниже.
- Среди клиентов преобладают женщины над мужчинами. Среди возрастных групп лидирует группа от 30 до 40 лет.
- Медианные выручки мужчин и женщин отличаются незначительно.

Поскольку этот вариант выполнения задания я реализовал исключительно в рамках возможностей Power BI, я не смог выполнить последний пункт задания (вычислить каждую 6-ю продажу клиента). Тем интереснее было попробовать использовать дополнительные средства.

Вариант №2

Предварительная обработка данных с помощью Python

Подробный скрипт расписан в файле bestmebel.ipynb.

Файл discounts.csv уже не загружался в Power BI, так как нам известна логика его формирования: начиная с седьмого доставленного заказа полагался купон на плюс 500 рублей к каждому следующему доставленному заказу.

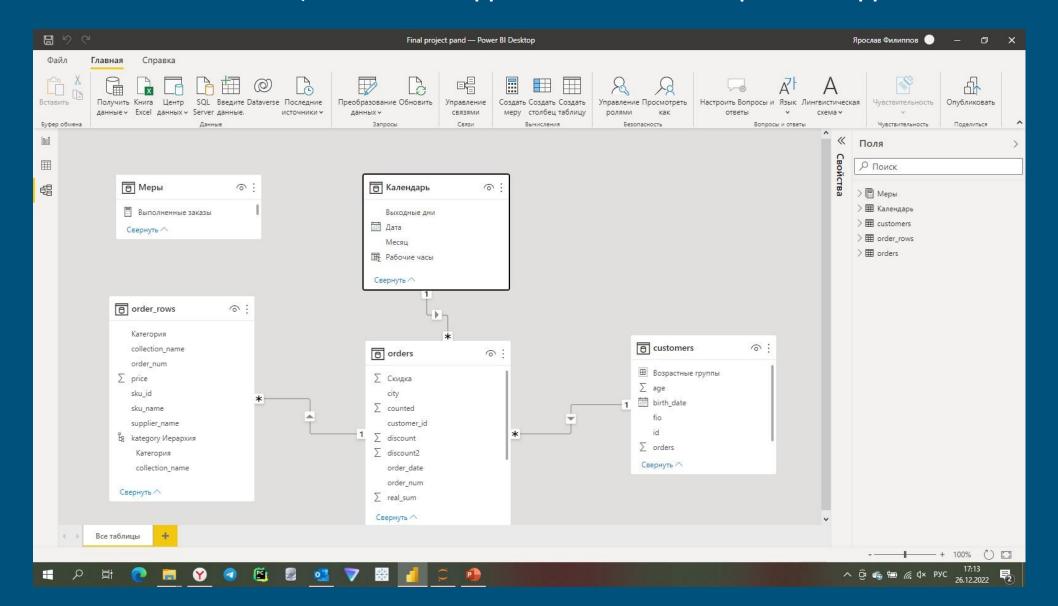
Кроме того, счетчик заказов каждого клиента позволяет реализовать дополнительное задание заказчика – посчитать дополнительный дисконт за шестые заказы.

Загруженные данные осталось только немного поправить в Power Query. Это главным образом касалось изменения типов данных.

Непосредственно в Power BI осталось только добавить столбец Возрастные группы в таблице customers, написать меры и сформировать дашборды аналогично Варианту №1.

Подробнее можно посмотреть в файле bestmebel_pand.pbix

Так как вспомогательные таблицы были объединены с основными, схема соединения таблиц выглядит так:



Вариант №3

Предварительная обработка данных с помощью SQL

Подробный скрипт расписан в файле Script_sql.txt.

В этом скрипте данные трансформируются аналогично Варианту №2 и мы получаем точно такие же таблицы.

Отличие только в том, что данные из Excel-таблицы Ozon-Nov-Sale.xlsx пришлось загружать в СУБД вручную.

Схема соединения таблиц такая же как в Варианте №2.

После этого в Power BI выполняются все те же действия, что и в Варианте №2.

Подробнее можно посмотреть в файле bestmebel_sql.pbix

Заключение

Сравнивая все 3 варианта лично мне больше понравился Вариант №2:

- Он позволил выполнить задание на все 100%, в отличии от Варианта №1.
- Доступна загрузка данных из файлов всех форматов и не пришлось загружать данные вручную, как в Варианте №3.