# **I. Анализ данных о продажах**

Для того, чтобы сформировать выводы о продажах фирмы Pens and Pencils нужно ответить на следующие вопросы:

* Какова динамика продаж?
* Какие категории и подкатегории продаж приносят больше всего выручки?
* Какие товары продаются лучше всего?

Начнем с вопроса о динамике продаж. Для получения результатов по продажам нужно обратиться к базе данных. В нашем распоряжении есть 4 таблицы, которые представляют собой:

1. справочник клиентов;
2. справочник товаров;
3. таблицу со списком товаров в заказе, их количеством, а также скидкой для каждого товара;
4. таблицу с информацией о заказах и их доставке.

Определим сумму выручки по месяцам, для этого напишем следующий запрос:

SELECT

DATE\_TRUNC('MONTH',SD.ORDER\_DATE)::DATE DATE,

ROUND(SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT))) REVENUE

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CARTS SC ON SD.ORDER\_ID = SC.ORDER\_ID

JOIN SQL.STORE\_PRODUCTS SP ON SC.PRODUCT\_ID = SP.PRODUCT\_ID

GROUP BY 1

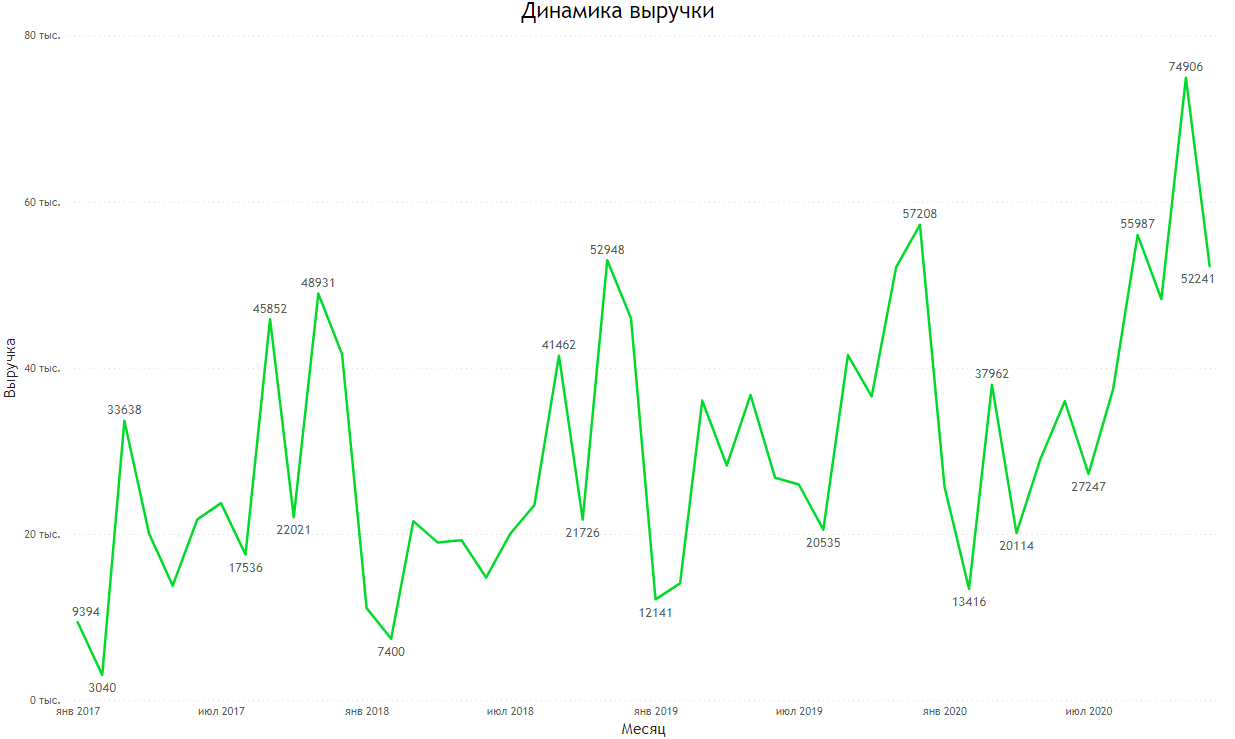
ORDER BY 1

Пояснения к запросу

В данном запросе нужно было объединить 3 таблицы по соответствующим ключам, в качестве столбцов выбрать дату заказа (преобразовав ее в месячный вид) и выручку, которую мы получаем с помощью агрегатных функций.

## 

## 

По полученным данным был построен следующий график:

### 

## 

## 

## 

## 

## 

## 

## 

По построенному графику отчетливо видно, что выручка растет с течением времени. Также наблюдается сезонность роста и падения выручки.

Следующим нашим вопросом стоит определение категорий и подкатегорий продаж приносящих больше всего выручки. Для ответа на него напишем запрос:

SELECT

SP.CATEGORY,

SP.SUBCATEGORY,

ROUND(SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT))) REVENUE

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CARTS SC ON SD.ORDER\_ID = SC.ORDER\_ID

JOIN SQL.STORE\_PRODUCTS SP ON SC.PRODUCT\_ID = SP.PRODUCT\_ID

GROUP BY 1, 2

ORDER BY 3 DESC

Пояснения к запросу

В данном запросе нужно было объединить такие же 3 таблицы по соответствующим ключам как и в предыдущем, но в качестве столбцов выбрать категории и подкатегории товаров, а также выручку, которую мы получали ранее.

На основе результатов запроса можно сформулировать вывод о том, что в фирме самыми ходовыми товарами являются кресла.

Следующее что нужно выяснить - лучшие по продажам наименования товаров. Для этого определим топ-25 товаров по объему выручки в компании посредством запроса:

WITH PRODUCT\_REVENUE AS

(SELECT

SP.PRODUCT\_NM,

SUM(SC.QUANTITY) QUANTITY,

SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT)) REVENUE

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CARTS SC ON SD.ORDER\_ID = SC.ORDER\_ID

JOIN SQL.STORE\_PRODUCTS SP ON SC.PRODUCT\_ID = SP.PRODUCT\_ID

GROUP BY 1),

TOTAL\_REVENUE AS

(SELECT

SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT)) REVENUE

FROM SQL.STORE\_CARTS SC

JOIN SQL.STORE\_PRODUCTS SP ON SC.PRODUCT\_ID = SP.PRODUCT\_ID)

SELECT

PR.PRODUCT\_NM,

PR.REVENUE,

PR.QUANTITY,

(PR.REVENUE / TR.REVENUE) \* 100 PERCENT\_FROM\_TOTAL

FROM PRODUCT\_REVENUE PR

CROSS JOIN TOTAL\_REVENUE TR

ORDER BY 4 DESC

LIMIT 25

Пояснения к запросу

Для выполнения данного запроса было написано 2 CTE. Одно PRODUCT\_REVENUE для определения объема выручки, наименования и количества товара. Второе TOTAL\_REVENUE для подсчета общей выручки за все время. После чего эти два CTE были объединены в запрос посредством CROSS JOIN и таким образом получилось посчитать долю от общей выручки для каждого товара.

В результате выполнения запроса мы выяснили, что самым продаваемым товаром является Canon imageCLASS 2200 Advanced Copier.

На все вопросы касательно эффективности продаж Pens and Pencils мы ответили, теперь перейдем к описанию клиентов компании.

# **II. Портрет клиента компании**

Сформулируем вопросы, которые помогут получить больше информации о клиентах компании:

* Как соотносится количество клиентов по категориям B2B и B2C?
* Как соотносится выручка по этим категориям?
* Каких клиентов больше, новых или старых?

Ответив на них, рассчитаем основные показатели по целевым клиентам: количество заказов, средняя сумма заказов и количество различных адресов доставки.

Для начала ответим на первый вопрос и посмотрим, сколько у компании B2B и B2C-клиентов и какую долю от общей выручки они приносят. Напишем следующий запрос:

SELECT

SCU.CATEGORY,

COUNT(DISTINCT SCU.CUST\_ID) CUST\_CNT,

ROUND(SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT))) REVENUE

FROM SQL.STORE\_CUSTOMERS SCU

JOIN SQL.STORE\_DELIVERY SD ON SCU.CUST\_ID = SD.CUST\_ID

JOIN SQL.STORE\_CARTS SC ON SD.ORDER\_ID = SC.ORDER\_ID

JOIN SQL.STORE\_PRODUCTS SP ON SC.PRODUCT\_ID = SP.PRODUCT\_ID

GROUP BY 1

ORDER BY 3 DESC

Пояснения к запросу

Для выполнения данного запроса мы объединили все 4 таблицы и вывели в столбцы категории клиентов, а также (с помощью агрегатных функций) их количество и прибыль, которую они принесли компании.

На основе результатов запроса видно, что B2B-клиентов намного больше, чем B2С-клиентов и выручки они приносят соответственно больше.

Составим характеристику B2B-клиентов. Для этого посмотрим на динамику новых B2B-клиентов по месяцам. Заодно выясним, выручка растёт за счёт увеличения продаж по старым клиентам или же за счёт привлечения новых. Напишем запрос, который выведет количество новых корпоративных клиентов по месяцам:

## 

WITH CUST\_DATE AS

(SELECT

SD.ORDER\_ID,

SCU.CUST\_ID,

SD.ORDER\_DATE,

ROW\_NUMBER()

OVER(PARTITION BY SCU.CUST\_ID

ORDER BY SD.ORDER\_DATE,

SD.ORDER\_ID) AS ROW

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CUSTOMERS SCU ON SCU.CUST\_ID = SD.CUST\_ID

WHERE SCU.CATEGORY = 'Corporate')

SELECT

DATE\_TRUNC('MONTH', CD.ORDER\_DATE)::DATE MONTH\_,

COUNT(CD.CUST\_ID)

FROM CUST\_DATE CD

WHERE CD.ROW='1'

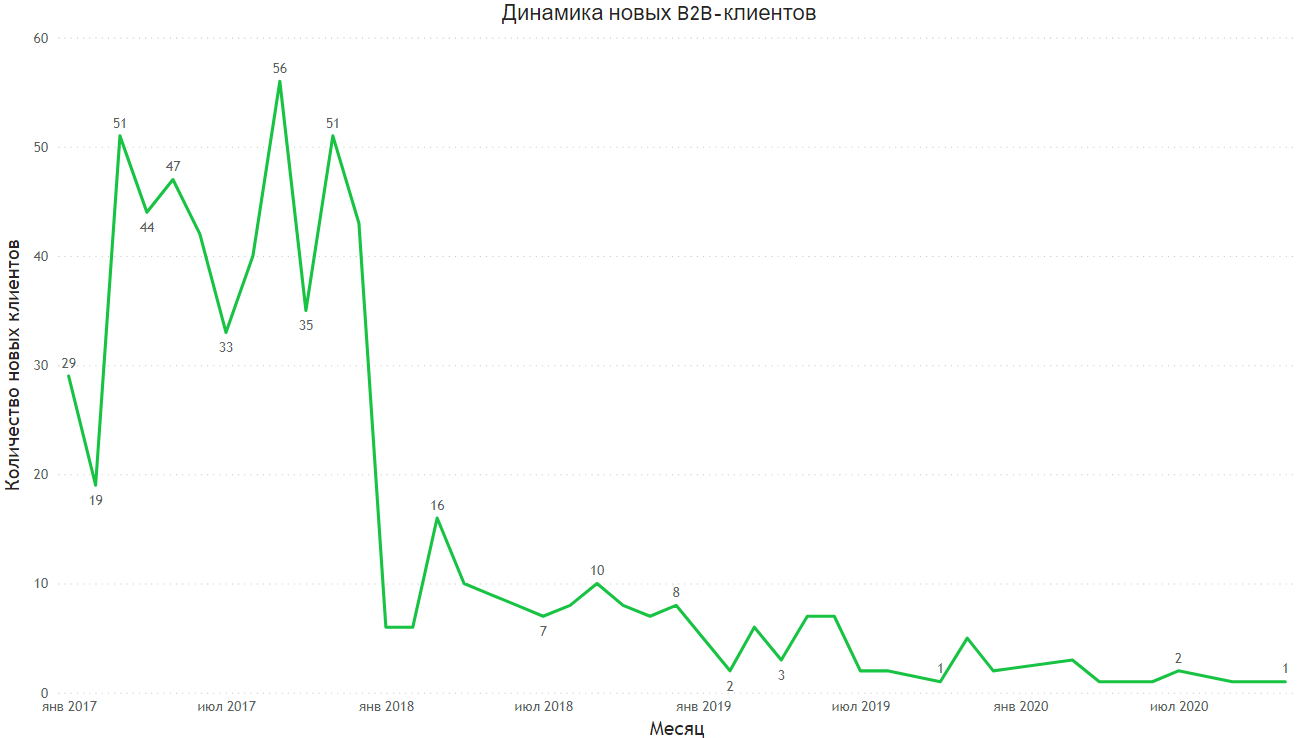
GROUP BY 1

ORDER BY 1

Пояснения к запросу

Для выполнения данного запроса было написано CTE для распределения очередности покупки посредством использования оконной функции. На основе уже созданного CTE CUST\_DATE были выбраны только первые заказы от каждого корпоративного клиента и сгруппированы по дате заказа (преобразованной в месячный вид).

Для наглядности полученных данных оформим их в виде графика.



## 

Как видно по графику количество новых клиентов падает. В самом начале деятельности фирмы было привлечено много корпоративных клиентов, но после 2018 года их количество резко снижается.

Теперь давайте изучим основные показатели по корпоративным клиентам: какова сумма заказа, сколько товаров в заказе и сколько различных офисов, куда они доставлялись. Оформим запрос:

WITH ORDERS AS

(SELECT

SD.ORDER\_ID,

COUNT(DISTINCT SC.PRODUCT\_ID) PRODUCTS,

SUM((SC.QUANTITY \* SP.PRICE) \* (1 - SC.DISCOUNT)) REVENUE

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CUSTOMERS SCU ON SCU.CUST\_ID = SD.CUST\_ID

JOIN SQL.STORE\_CARTS SC ON SD.ORDER\_ID = SC.ORDER\_ID

JOIN SQL.STORE\_PRODUCTS SP ON SP.PRODUCT\_ID = SC.PRODUCT\_ID

WHERE SCU.CATEGORY = 'Corporate'

GROUP BY 1),

OFFICE AS

(SELECT

SCU.CUST\_ID,

COUNT(DISTINCT SD.ZIP\_CODE) OFFICE

FROM SQL.STORE\_DELIVERY SD

JOIN SQL.STORE\_CUSTOMERS SCU ON SCU.CUST\_ID = SD.CUST\_ID

WHERE SCU.CATEGORY = 'Corporate'

GROUP BY 1)

SELECT

ROUND(AVG(ORD.PRODUCTS),1) PRODUCTS,

ROUND(AVG(ORD.REVENUE),1) REVENUE,

ROUND(AVG(OC.OFFICE),1) OFFICE

FROM ORDERS ORD

CROSS JOIN OFFICE OC

Пояснения к запросу

Для выполнения данного запроса было написано 2 CTE. Одно ORDERS подсчитывает уникальные товары и сумму заказов и корпоративных клиентов на основе всех 4 таблиц. Второе OFFICE считает посредством агрегатной функции количество офисов у корпоративных клиентов на основе двух таблиц.

Основной запрос же объединяет обе CTE с помощью CROSS JOIN и считает средние значения по всем трем пунктам.

В результате выполнения запроса видно, что в среднем корпоративный клиент берет 2 продукта, приносит 285,9 единицы выручки и имеет 6,2 офисов, в которые доставляют товары. В купе с предыдущими данными о корпоративных клиентах можно сформулировать выводы о том, что привлекательность продуктов компании упала среди данной категории клиентов. Стоит применить стимулирующие действия по привлечению новых покупателей, например запустить рекламную компанию.

Теперь мы знаем больше о корпоративных клиентах, которые представляют наибольшую ценность для компании, а значит пора перейти к последнему вопросу.

# **III. Анализ логистики компании**

Осталось оценить текущую картину по логистике доставок и найти штат, в котором лучше всего открыть офлайн-магазин.

Вновь сформулируем вопросы, на которые нам предстоит найти ответы в рамках задачи:

* Насколько эффективно выполняются текущие доставки?
* Как распределяются доставки и выручка по штатам и городам?

Для начала определим, какая доля заказов выполняется в срок по каждой категории посредством запроса:

WITH LATE AS

(SELECT

SD.SHIP\_MODE,

COUNT(SD.ORDER\_ID) ORDERS\_CNT,

COUNT(CASE

WHEN SD.SHIP\_MODE = 'Standard Class'

AND SD.SHIP\_DATE > SD.ORDER\_DATE + 6 THEN 1

WHEN SD.SHIP\_MODE = 'Second Class'

AND SD.SHIP\_DATE > SD.ORDER\_DATE + 4 THEN 1

WHEN SD.SHIP\_MODE = 'First Class'

AND SD.SHIP\_DATE > SD.ORDER\_DATE + 3 THEN 1

WHEN SD.SHIP\_MODE = 'Same Day'

AND SD.SHIP\_DATE > SD.ORDER\_DATE THEN 1

END) LATE\_ORDERS\_CNT

FROM SQL.STORE\_DELIVERY SD

GROUP BY 1)

SELECT

L.SHIP\_MODE,

L.ORDERS\_CNT,

L.LATE\_ORDERS\_CNT, ROUND(((L.ORDERS\_CNT-L.LATE\_ORDERS\_CNT)/L.ORDERS\_CNT::NUMERIC)\*100, 2) "%\_success"

FROM LATE L

ORDER BY 4

Пояснения к запросу

Для выполнения данного запроса было написано CTE с функцией CASE для распределения количества заказов доставленных не в срок по типам доставки с учетом времени доставки, а также подсчета общего количества доставок по типа доставки. После чего в общий запрос были взяты данные из CTE, а также добавлен столбец с подсчетом доли успешно доставленных заказов.

В результате выполнения запроса мы выяснили, что чаще всего с опозданием доходят заказы, отправленные вторым классом.

Для получения более детальных данных по задержкам напишем запрос который распределит все заказы отправленных вторым классом, которые были доставлены с опозданием, по кварталам.

WITH LATE AS

(SELECT

SD.ORDER\_DATE,

SD.SHIP\_MODE,

COUNT(CASE

WHEN SD.SHIP\_MODE = 'Second Class'

AND SD.SHIP\_DATE > SD.ORDER\_DATE + 4 THEN 1

END) LATE\_ORDERS\_CNT

FROM SQL.STORE\_DELIVERY SD

GROUP BY 1,2)

SELECT

DATE\_TRUNC('QUARTER', L.ORDER\_DATE)::DATE QUAR,

SUM(L.LATE\_ORDERS\_CNT) LATE\_ORDERS

FROM LATE L

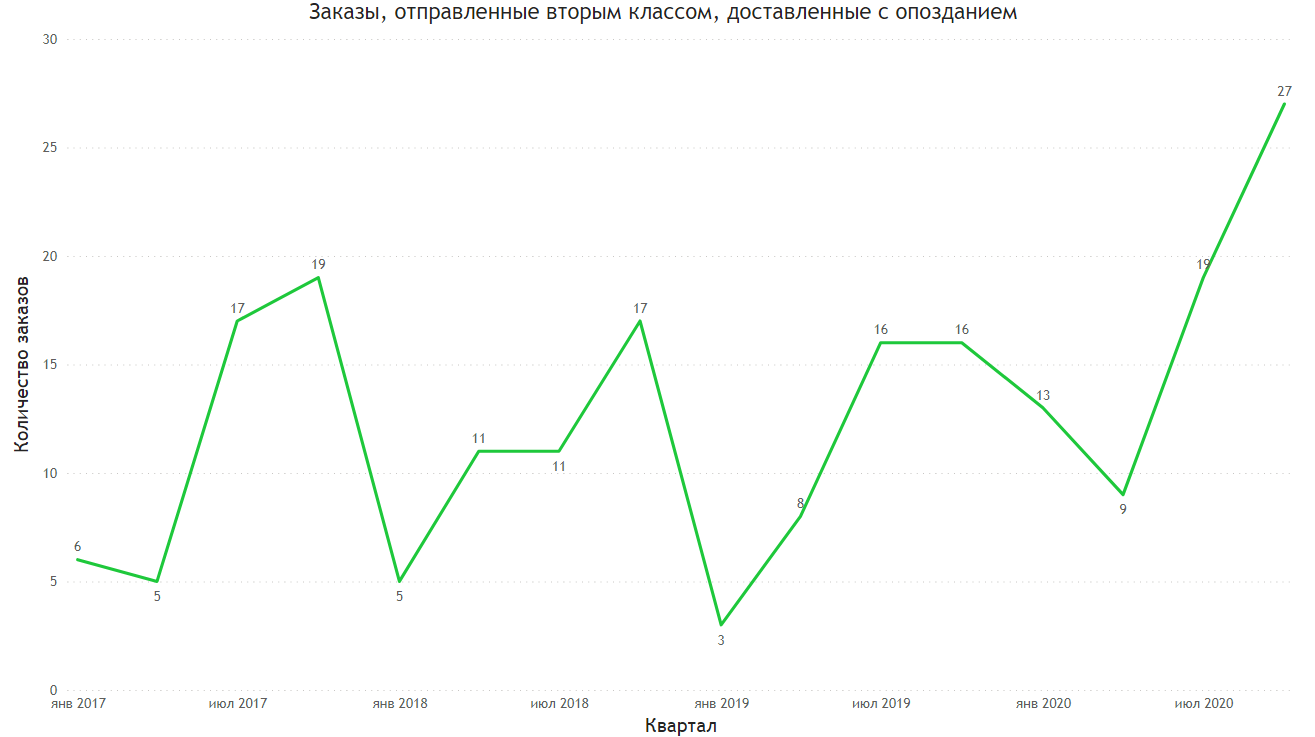
WHERE L.LATE\_ORDERS\_CNT>0

GROUP BY 1

ORDER BY 1

Пояснения к запросу

Для выполнения данного запроса было взято CTE с функцией CASE из предыдущего запроса и сокращено по типу доставки только из второго класса, а также добавлен столбец с датой заказа (было решено взять дату заказа, а не дату доставки, так как это более менее фиксированная дата в отличие от даты доставки). После чего в общем запросе были использованы данные из CTE, а именно дата заказа (сокращенная до квартала) и сумма просроченных заказов по кварталам.



## 

После получения данных и построения графика можно заметить, что опоздание доставки происходит сезонно в летний период.

Перейдем к исследованию места для открытия офлайн-магазина.

Сейчас есть только склад, откуда отправляются все товары, — находится он в городе Хьюстон, штат Техас (Houston, Texas).

С помощью офлайн-магазина можно привлечь больше клиентов и снизить стоимость доставки, нужно только выбрать, где его открыть. Для этого найдем город и штат, куда совершается больше всего доставок.

## 

SELECT

SD.STATE,

COUNT(SD.ORDER\_ID)

FROM SQL.STORE\_DELIVERY SD

GROUP BY 1

ORDER BY 2 DESC

Пояснения к запросу

Простой запрос на основе одной таблицы по доставкам.

В результате мы выяснили, что наиболее популярный штат по количеству доставок - California.

SELECT

SD.CITY,

COUNT(SD.ORDER\_ID)

FROM SQL.STORE\_DELIVERY SD

GROUP BY 1

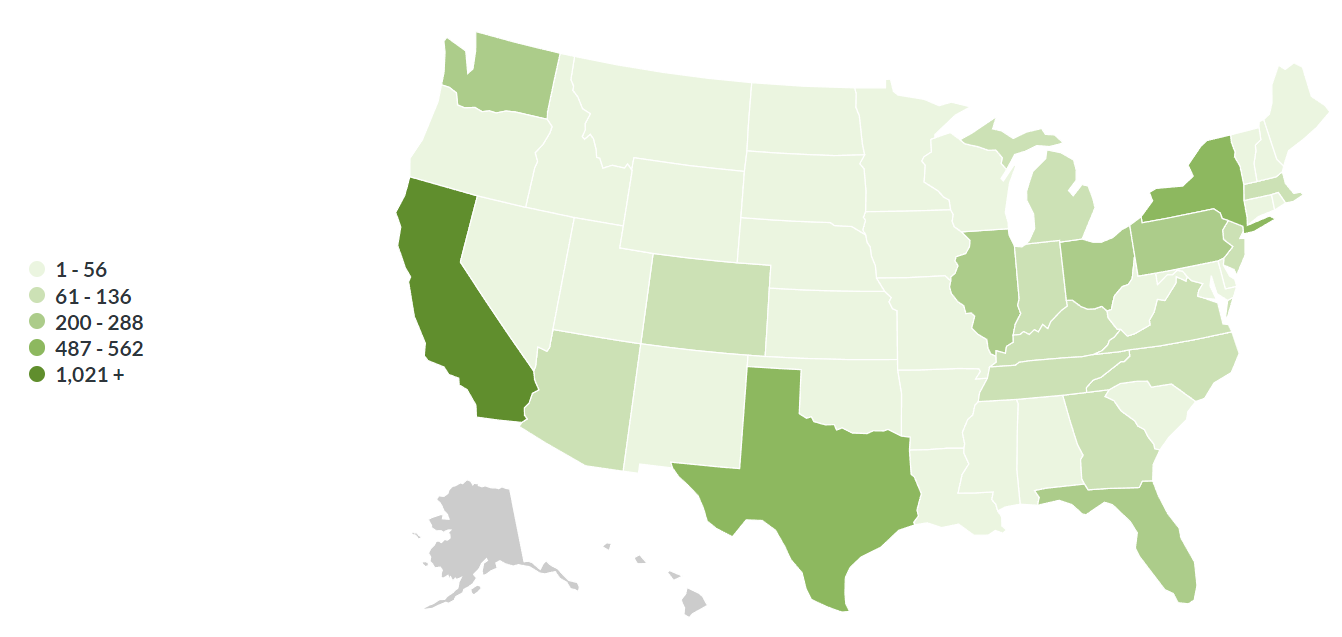
ORDER BY 2 DESC

Пояснения к запросу

Простой запрос на основе одной таблицы по доставкам.

В результате мы выяснили, что наиболее популярный город по количеству доставок - New York City.

Построим график доставок по штатам.



Судя по графику, самый перспективный штат для открытия магазина - это Калифорния, но не все так просто, на основе всего исследования можно сделать совсем другие выводы. Перейдем же к ним.