**ЗАДАЧА ПРОЕКТА:**

проанализировать работу компании с точки зрения её эффективности и дать рекомендации по масштабированию бизнеса, а именно в каком штате лучше открыть офлайн-магазин.

**КОНКРЕТНЫЕ ШАГИ (ФОРМАЛИЗОВАННАЯ ЗАДАЧА):**

1.Оценить динамику продаж и распределение выручки по товарам.

2. Составить портрет клиента, а для этого — выяснить, какие клиенты приносят больше всего выручки.

3.Проконтролировать логистику компании (определить, все ли заказы доставляются в срок и в каком штате лучше открыть офлайн-магазин).

**Примечание: в целях проверки проекта проведена дополнительная структуризация работы по номеру заданий. Задания с маской 27.5… относятся к первому шагу, 27.6..- ко второму шагу, 27.7… - к третьему шагу.**

**ШАГ 1. Оценить динамику продаж и распределение выручки по товарам**

**Задание 27.5.1-27.5.4**

27.5.1. Напишите запрос, который выведет сумму выручки по месяцам:

select

date\_trunc('month', order\_date)::date as "date",

round(sum(price \* quantity \* (1 - 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

**Комментарии: Для выполнения запроса используем данные трех исходных таблиц: delivery(поле order\_date), carts(поля quantity,discounts), products (поле price), предварительно преобразовава поле order\_date в date и использовав функцию trunc для помесячного расчета. Осуществлена помесячная группировка. Для округления до целого использована функция round. Джойн проведен по соответствующим ключам -order\_id, product\_id( в дальнейшем этот, достаточно очевидный комментарий будет опускаться)**

27.5.2.По полученным данным постройте график, который отразит динамику выручки

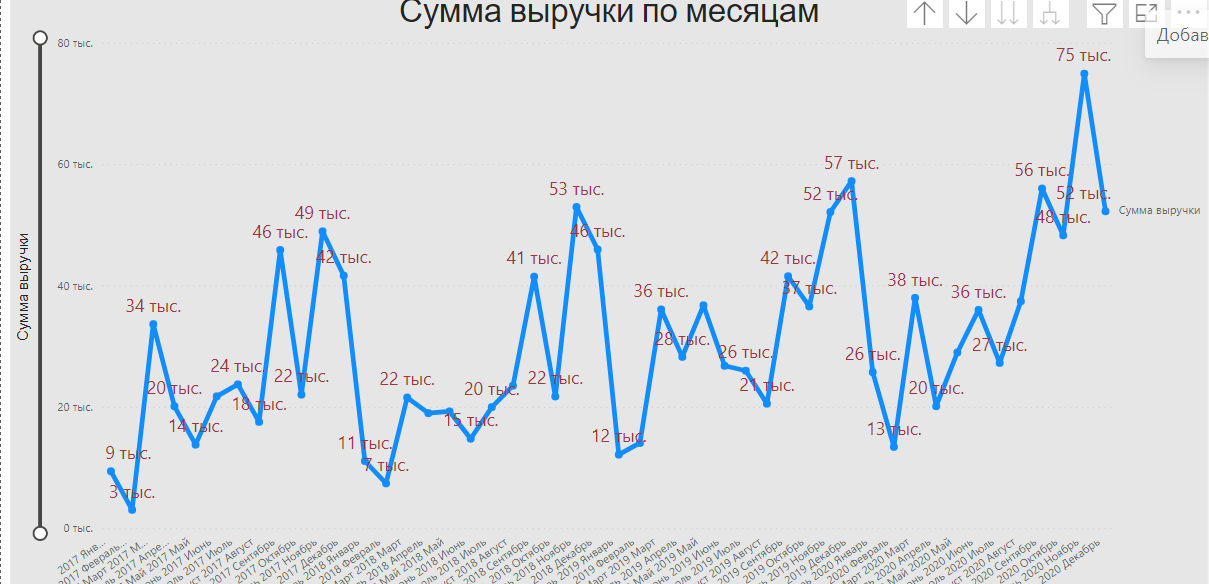


График с прогнозом

****

27.5.3.Какова динамика выручки?

Вывод: В рассматриваемом периоде объем выручки растет, Однако, при построении прогнозного графика в Power BI b с использованием аналитических функций предполагается падение выручки в дальнейшем.

27.5.4 Наблюдается ли сезонность роста и падения выручки?

Вывод: Отмечается выраженная сезонность объема выручки: ее падение в 1 квартале и рост к 4 кварталу

**Задание 27.5.5**

Напишите запрос, который выведет сумму выручки по различным категориям и подкатегориям:category (категория товара);subcategory (подкатегория товара);revenue (объём выручки) ― округлите до целых с помощью round.Отсортируйте запрос по убыванию выручки.

select

category,

subcategory,

round(sum(price \* quantity \* (1 - discount))) revenue

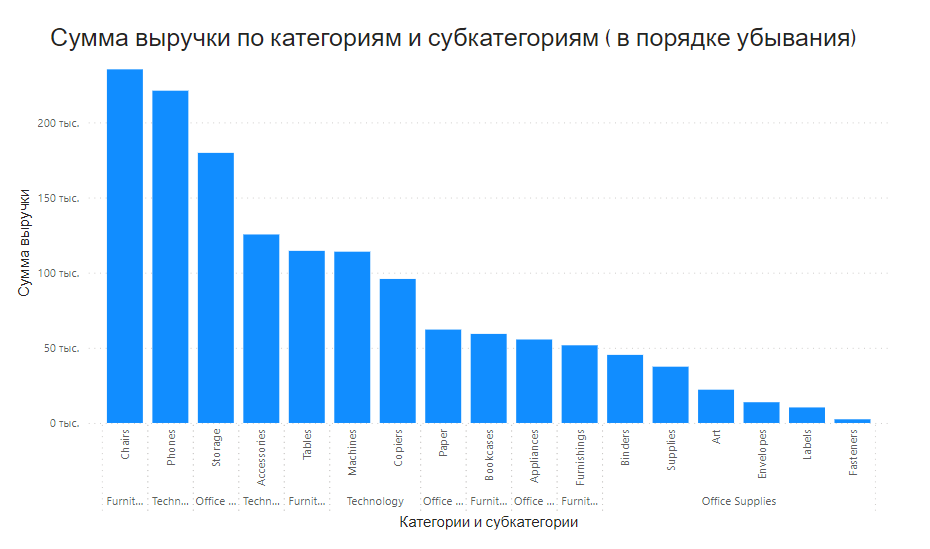
from

sql.store\_products sp

join sql.store\_carts sc on sc.product\_id = sp.product\_id

group by 1,2

order by 3 desc



**Комментарии: Для выполнения запроса используем данные исходных таблиц: carts(поля quantity,discounts), products (поля price,category ,subcategory. Группировка по категориям и субкатегориям, сортировка по убыванию прибыли.**

Выводы : Тройку лидеров по продажам на уровне подкатегорий составили стулья(кресла) из категории мебель, телефоны из категории технологии, места для хранения из категории офисные принадлежности

**Задание 27.5.6.**

Напишите запрос, который выведет данные топ-25 товаров по объёму выручки в следующем формате:product\_nm — наименование товара;revenue — объём выручки; quantity — количество проданных товаров; percent\_from\_total — доля от общей выручки в процентах.

WITH total\_revenue AS

(SELECT

sp.product\_nm,

ROUND(SUM(sc.quantity\*(1-sc.discount)\*sp.price),2) revenue,

SUM (sc.quantity) quantity

FROM sql.store\_carts sc

JOIN sql.store\_products sp ON sp.product\_id=sc.product\_id

GROUP BY 1)

SELECT

product\_nm,

revenue,

quantity,

ROUND(revenue/(SUM(revenue) OVER())\*100,2) percent\_from\_total

FROM total\_revenue

GROUP BY 1,2,3

ORDER BY 4 DESC, 2 DESC

LIMIT 25

**Комментарии: Для выполнения запроса используем данные исходных таблиц: carts(поля quantity,discounts), products (поля price,category ,subcategory. Однако для вычисления процентов используем CTE (с помощью которого предварительно рассчитываем выручку и количество проданных товаров, группируя их по наименованию), и. далее. в основном запросе выводим наименованию продукта, объем выручки, количество проданных товаров, а также долю от общей выручки в процентах с использованием оконной функции и округлением. затем сортируем и ограничиваем количество выводимых строк по условию задания.**

Выводы : Тройку лидеров по продажам на уровне подкатегорий составили стулья(кресла) из категории мебель, телефоны из категории технологии, места для хранения из категории офисные принадлежности

**ШАГ 2. Составить портрет клиента, а для этого — выяснить, какие клиенты приносят больше всего выручки.**

**Задание 27.6.1**

Напишите запрос, который выведет количество клиентов и выручку по категориям клиента:

category (категория клиента);

cust\_cnt (количество клиентов);

revenue (объём выручки) ― округлите до целых с помощью round.

Отсортируйте запрос по убыванию выручки.

SELECT

scust.category,

count(distinct scust. cust\_id) cust\_cnt,

ROUND(SUM(sc.quantity\*(1-sc.discount)\*sp.price)) revenue

FROM sql.store\_carts sc

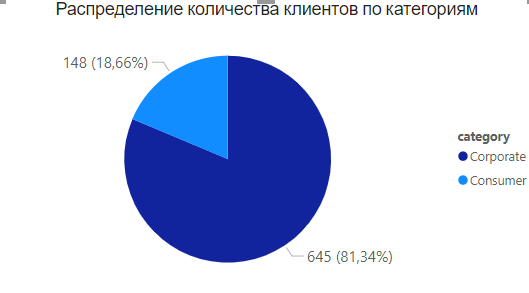
JOIN sql.store\_products sp on sp.product\_id=sc.product\_id

join sql.store\_delivery sd on sd.order\_id=sc.order\_id

join sql.store\_customers scust on scust.cust\_id=sd.cust\_id

group by 1

order by 3 desc





**Комментарии: Поскольку не удается напрямую связать таблицы customers c carts и products, используем для связи таблицу delivery по полю cust\_id . Для исключения повторов применяем оператор distinct в фрагменте count(distinct scust. cust\_id) cust\_cnt. Округляем и сортируем в соответствии с условиями задания.**

Выводы: Доля корпоративных клиентов и доля выручки от их приобретений составляют более 80 %

**Задания 27.6.2 -27.6.3**

Напишите запрос, который выведет количество новых корпоративных клиентов по месяцам:

month (месяц) ― тип date;

new\_custs (количество новых клиентов).

Отсортируйте запрос по первому столбцу в порядке возрастания.

with first\_date as(

select

cust\_id,

date\_trunc ('month',min(order\_date)) first\_date

from sql.store\_delivery

group by 1

)

select

first\_date::date as month,

count (distinct scu.cust\_id) new\_custs

from first\_date fd

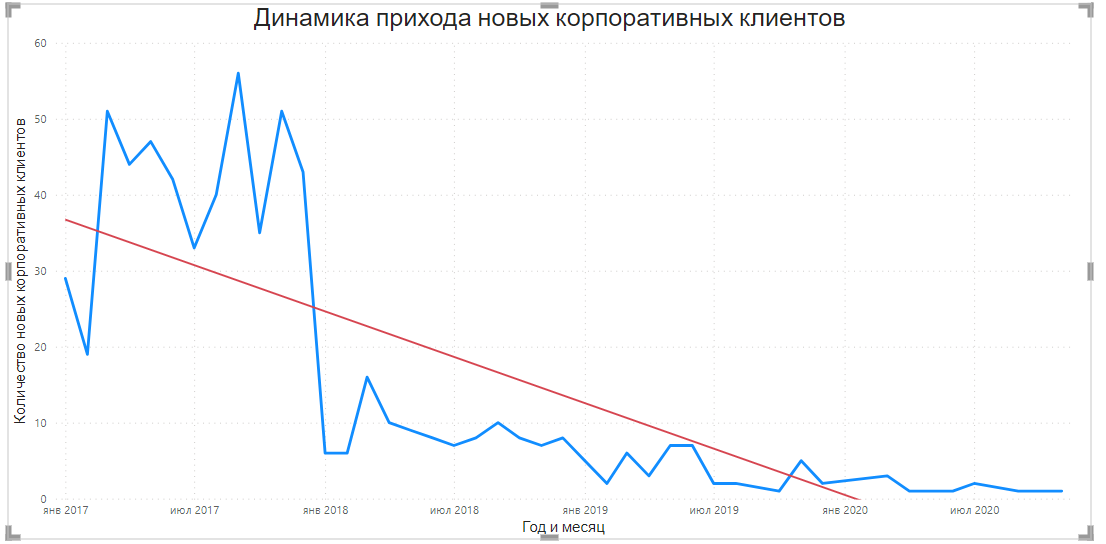
join sql.store\_customers scu on scu.cust\_id=fd.cust\_id

where scu.category = 'Corporate'

group by 1

order by 1

Динамика прихода новых корпоративных клиентов представлена на следующем графике:



**Комментарии: С помощью CTE первоначально осуществляем группировку клиентов по дате первого заказа. В основном запросе осуществляем подсчет этих клиентов, присоединяем таблицу customers для проведения фильтрации по категории “корпоративные”.**

Выводы : Активный прирост новых корпоративных клиентов отмечался в 2017 году, далее их количество резко снизилось.

**Задание 27.6.4**

Найдите ответы на следующие вопросы (все ответы округлите до одного знакапосле разделителя):

Сколько в среднем **различных** товаров в заказах у корпоративных клиентов?

Какая в среднем сумма заказов у корпоративных клиентов?

Сколько в среднем различных офисов у корпоративных клиентов?

Какие выводы можно сделать на основании этих показателей? Какие рекомендации можно дать?

with offices as( select category, d.cust\_id, count(distinct d.zip\_code) as dist\_offices

FROM sql.store\_delivery d

LEFT JOIN sql.store\_customers c ON c.cust\_id = d.cust\_id

WHERE c.category = 'Corporate'

GROUP BY 1,2),

m as

(select c.category, crt.order\_id, count(distinct crt.product\_id) as unique\_products, sum(crt.quantity\*price\*(1-discount)) as sum\_orders from sql.store\_carts crt

left join sql.store\_delivery d on crt.order\_id = d.order\_Id

left join sql.store\_customers c on c.cust\_id = d.cust\_id

left join sql.store\_products p on crt.product\_id = p.product\_id

where c.category ='Corporate'

group by 1, 2

)

select

round ( avg(unique\_products),1)corp\_avg\_prod,

round (avg(sum\_orders),1)corp\_avg\_sum,

round(avg(dist\_offices),1)corp\_avg\_office

from m

left join offices o on o.category = m.category

**Комментарии: В запросе используем 2 CTE Первым считаем количество офисов у корпоративных клиентов, вторым -количество уникальных товаров и сумму заказов у этой же категории. Основным запросом определяем среднее вышеуказанных параметров с округлением до одного знака.**

Выводы :У корпоративных клиентов среднее число **различных** купленных товаров -2, сумма заказов - 285,9, различных офисов-6,2. Для сравнения выполнен запрос аналогичных показателей для категории ‘consumer’. Показатели следующие: среднее число **различных** купленных товаров -2, сумма заказов- 301,6, различных офисов-6. Таким образом Как корпоративные так и ‘consumer’ клиенты в пересчете на одного приносят примерно одинаковую прибыль. Необходимо сосредоточить усилия для привлечения корпоративных клиентов.

**ШАГ 3. Проконтролировать логистику компании (определить, все ли заказы доставляются в срок и в каком штате лучше открыть офлайн-магазин)**

**Задание 27.7.1**

Напишите запрос, который выведет:

тип доставки;

общее количество заказов (orders\_cnt);

количество заказов, которые не были доставлены вовремя (late\_orders\_cnt);

долю выполненных вовремя заказов, в процентах (% success), округлённую до двух знаков после запятой.

Отсортируйте запрос по четвёртому столбцу в порядке возрастания.

WITH delivery\_time AS

(SELECT

order\_id,

ship\_mode,

ship\_date-order\_date days\_fact,

CASE WHEN ship\_mode='Standard Class' THEN 6

WHEN ship\_mode='Second Class' THEN 4

WHEN ship\_mode='First Class' THEN 3

WHEN ship\_mode='Same Day' THEN 0

END norm\_delivery\_time

FROM sql.store\_delivery

GROUP BY 1, 2, 3)

SELECT

ship\_mode,

COUNT(order\_id)order\_cnt,

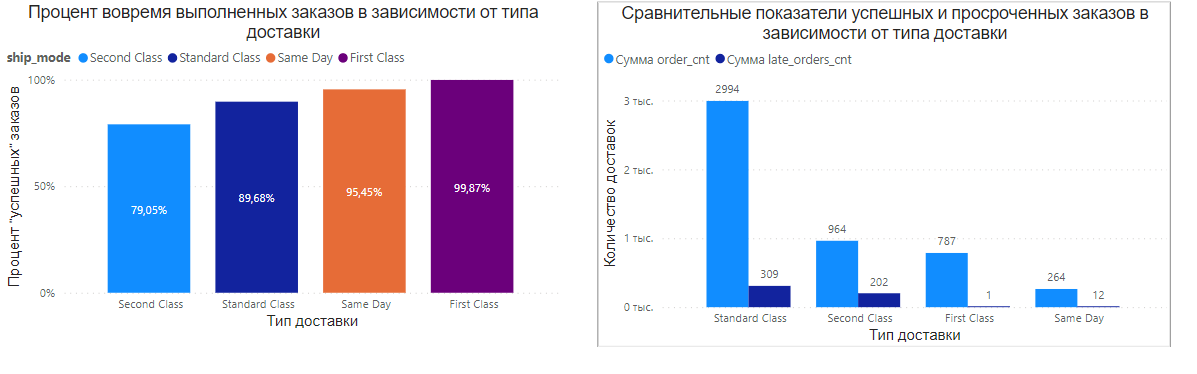
COUNT(CASE WHEN days\_fact>norm\_delivery\_time THEN order\_id END) late\_orders\_cnt,

ROUND(COUNT(CASE WHEN days\_fact<=norm\_delivery\_time THEN order\_id END)\*100.00/COUNT (order\_id),2 ) "% success"

FROM delivery\_time

GROUP BY 1

ORDER BY 4



**Комментарии: Первым шагом с помощью CTE формируем таблицу, включающую номер заказа, класс доставки, гарантированное и фактическое число дней доставки. Далее в основном запросе подсчитываем общее число заказов, заказов прибывших с опозданием и процент вовремя доставленных заказов. В данных запросах используем выражение case… when…then**

Наименьший процент вовремя доставленных заказов у отправленных вторым классом

**Задание 27.7.2**

Напишите запрос, чтобы вывести долю заказов, отправленных вторым классом, которые были доставлены с опозданием, по кварталам. По получившимся данным постройте график.

with delivery\_dates as

(select

date\_trunc('QUARTER', order\_date) quarters,

order\_id,

ship\_date,

order\_date,

ship\_date-order\_date difference\_date,

case when ship\_mode='Second Class' then 4

end planned\_time\_delivery

from sql.store\_delivery

where ship\_mode='Second Class'),

by\_quarters as (

select

date\_trunc('QUARTER', order\_date) quarters,

count (difference\_date) all\_late,

count (case when planned\_time\_delivery<difference\_date then order\_id end) late\_orders\_cnt

from delivery\_dates

group by 1

order by 1)

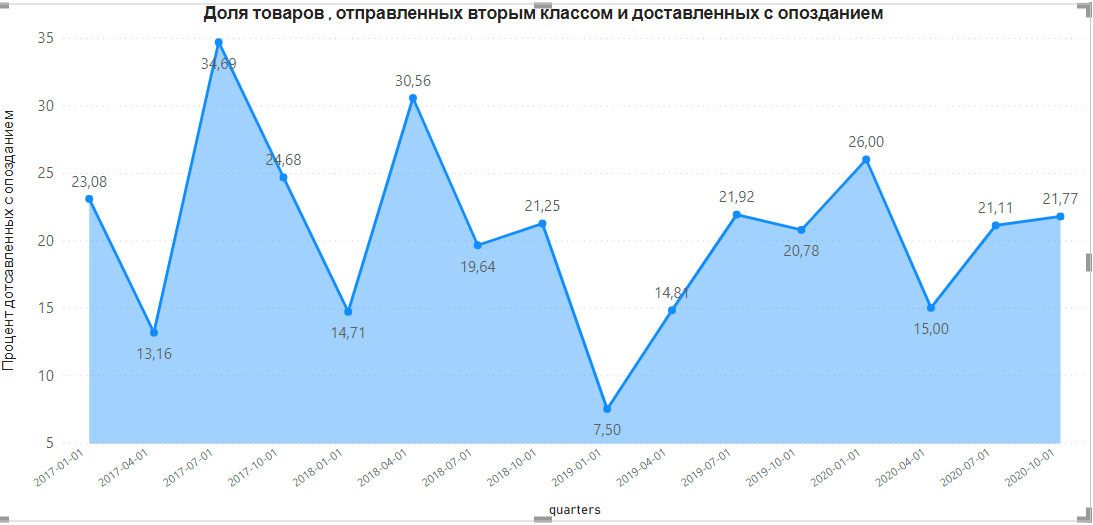
select

quarters::date,

(late\_orders\_cnt::numeric/all\_late::numeric)\*100 "%"

from by\_quarters

order by 1



**Комментарии:В данном задании используем 2 CTE. Первым выделяем доставку второго класса с распределением заказов по кварталам, подсчетом дней между днем заказа и днем доставки. Вторым подзапросом проводим подсчет заказов, доставленных с опозданием.И в основном запросе выводим процент заказов, доставленных с опозданием с поквартальным распределением. В запросе использован оператор \_trunc ('QUARTER')**

**Задание 27.7.3**

Какой вывод можно сделать: заказы доставляются с опозданием постоянно или только на протяжении ограниченного периода времени?

Вывод. На основании запроса из предыдущего задания доля товаров, доставленных с опозданием колеблется от 7 до 34 процентов Тенденции не определяются. Таким образом, заказы отправленные вторым классом систематически оставляются с опозданием.

**Задание 27.7.4**

Какой штат наиболее популярный по количеству доставок?

select

state,

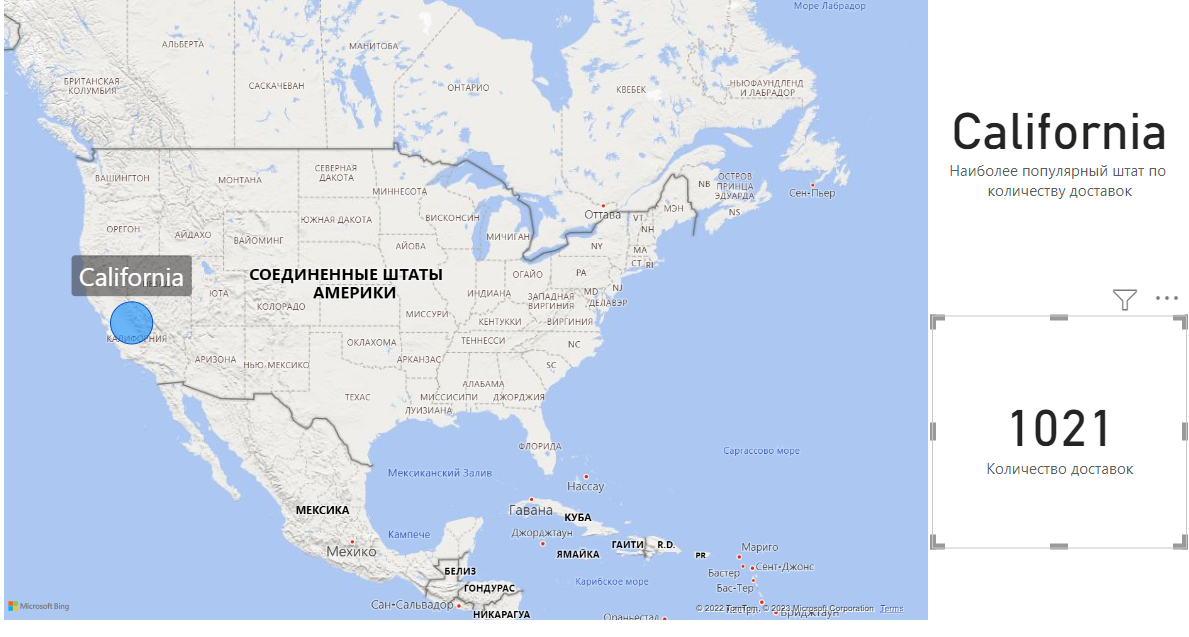
count(order\_id)

from sql.store\_delivery

group by 1

order by 2 desc

Limit 1



**Комментарии: Для написания запроса используем таблицу store\_delivery, группируя по штатам, сортируя в нисходящем порядке по количеству доставок и ограничивая по количеству выводимых строк (1)**

Вывод: Наиболее популярным по количеству доставок является штат Калифорния (1021 доставка)

**Задание 27.7.5**

Какой город наиболее популярный по количеству доставок?

select

city,

count(order\_id)

from sql.store\_delivery

group by 1

order by 2 desc

Limit 1



**Комментарии: Запрос выполняется аналогично предыдущему, только изменяем поле ‘state’ на ‘city’**

Вывод: Наиболее популярным городом по количеству доставок является Нью-Йорк (450 доставок)

**Задание 27.7.6**

Напишите запрос, который выведет количество доставок по штатам и визуализируйте результат с помощью карты в Metabase. Какой штат является наиболее перспективным для открытия офлайн-магазина? Почему?

select

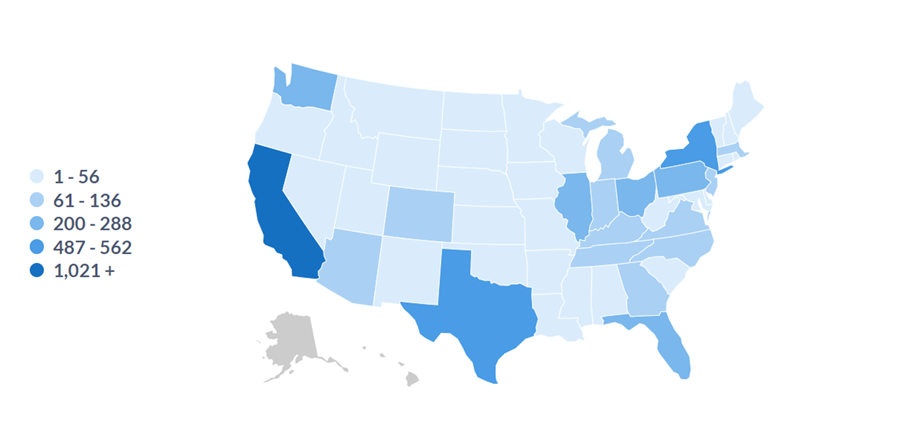
state,

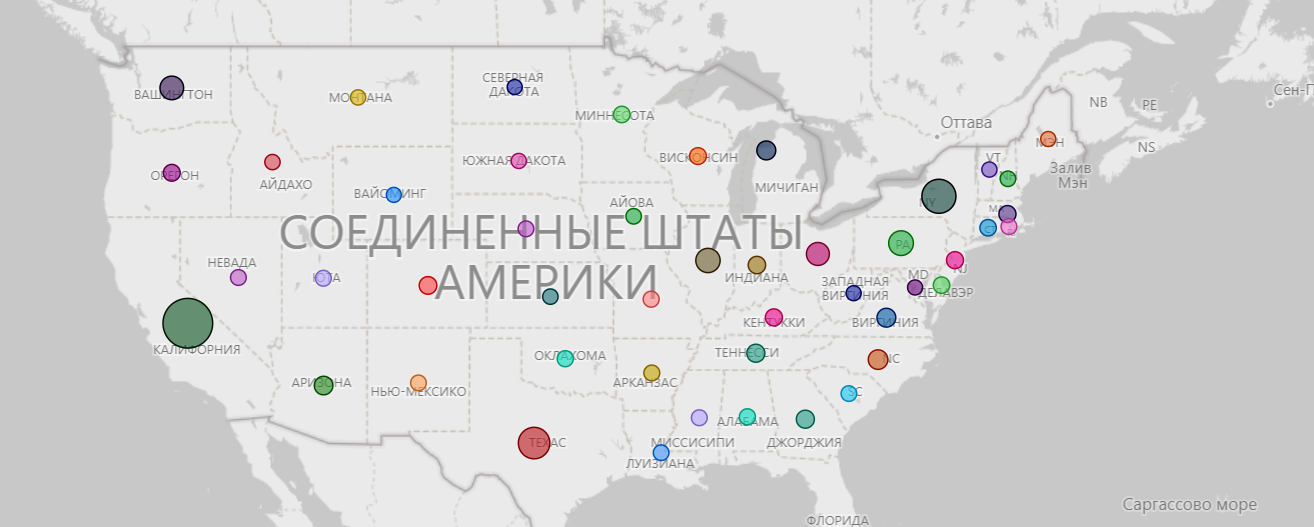
count(order\_id)

from sql.store\_delivery

group by 1

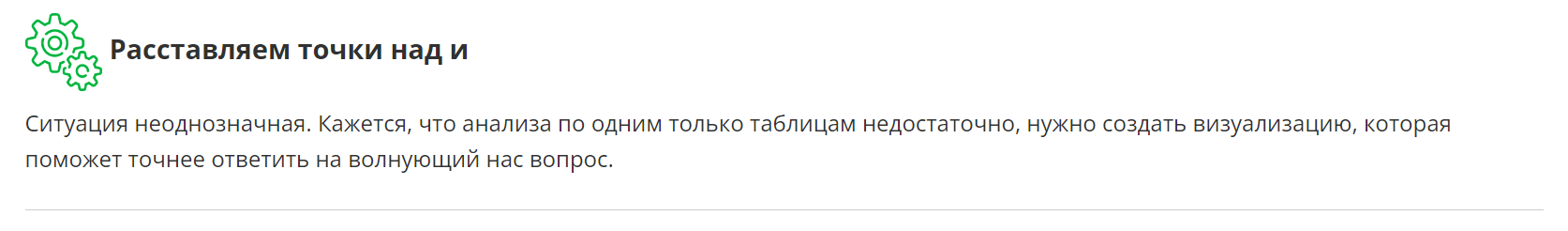
order by 2 desc





**Комментарии: Для написания запроса используем таблицу store\_delivery, группируя по штатам, сортируя в нисходящем порядке по количеству доставок**

Выводы: Ситуация действительно неоднозначная, счел нужным привести скриншот из условия задания:



По результатам визуализации видно, что наибольшее число покупок совершают жители восточных штатов. Но трудно представить себе ситуацию, что житель, например, Вирджинии, Флориды или Кентукки поедет в Нью-Йорк за покупками в оффлайн магазин.

Вспомнив, что согласно выводу из задания 27.6.1. доля корпоративных клиентов и доля выручки от их приобретений составляют более 80 %, Напишем дополнительный запрос, определив количество корпоративных клиентов в **штатах Калифорния и Нью-Йорк**

select

state,

category cat,

count(order\_id)orders

from sql.store\_delivery d

join sql.store\_customers cs on d.cust\_id=cs.cust\_id

group by 1,2

order by 3 desc

По результатам запроса определяем, что количество корпоративных клиентов в штате Калифорния составляет 836 , в штате Нью-Йорк - 469.

Это численное соотношение сохраняется и при определении количества клиентов без распределения по категориям:

select

state,

count(order\_id)orders

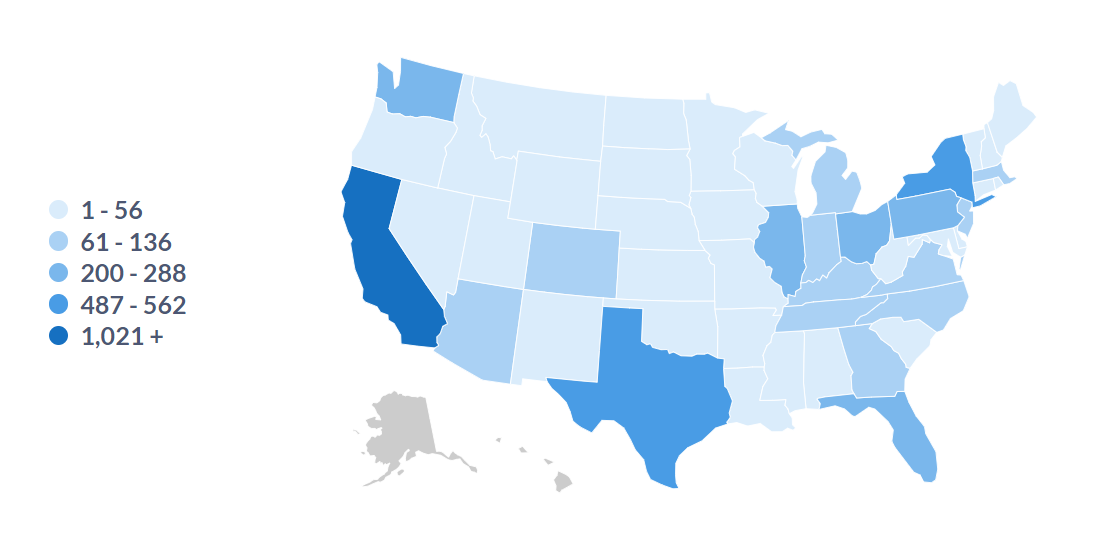
from sql.store\_delivery d

join sql.store\_customers cs on d.cust\_id=cs.cust\_id

group by 1

order by 2 desc

Калифорния- 1021, Нью-Йорк -562



Таким образом, приходим к выводу, что приоритетным является открытие офлайн магазина в штате Калифорния. Если же необходимо открыть **дополнительный склад,**  то предпочтительными штатами могут быть Кентукки, Вирджиния или Западная Вирджиния (географический центр покупок по восточным штатам).