**PROJECT-2. Решение бизнес-задач с помощью SQL**

**Выполнил: Рябов Д.В. DAPR-1006**

**Рассматривается компания *Pens and Pencils*. Компания базируется в США и там же осуществляет свою деятельность. С марта 2017 года её профиль — *B2B* и *B2C-*продажи офисных товаров (оргтехника, канцтовары и мебель) онлайн.**

**Задачи**

**Необходимо 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).**

### **Задание 27.4.1**

Заказы осуществлялись в период **с 2017-01-03 по 2020-12-30**

**Ответ в формате ГГГГ-ММ-ДД.**

### **Задание 27.4.2**

Всего **5009 заказов** было выполнено за это время.

### **Задание 27.4.3**

Объём общей выручки **1 446 157 usd**

**WITH RE AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT category, -- вывод нужных столбцов**

**subcategory,**

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

**FROM sql.store\_carts c -- объединение данных из таблиц**

**JOIN sql.store\_products p ON c.product\_id=p.product\_id**

**JOIN sql.store\_delivery d ON c.order\_id=d.order\_id**

**GROUP BY 1, -- группировка результатов**

**2**

**ORDER BY 3 DESC) -- сортировка выходных значений**

**SELECT SUM(revenue) -- вывод суммы по столбцу revenue**

**FROM RE**

## **27.5 PRACTICE. Шаг 1: Определяем эффективность продаж**

**Начнём с анализа данных о продажах и на основе полученных результатов сформулируем рекомендации.**

**Для этого ответим на следующие вопросы:**

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

### 

### 

### 

### **Задание 27.5.1 Считаем выручку.**

Напишите запрос, который выведет сумму выручки по месяцам:

* **date (месяц заказа) ― тип date;**
* **revenue (объём выручки) ― округлите значения до целых с помощью round. Отсортируйте запрос по дате заказа.**

**WITH tot\_rev AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT price\*quantity AS rev1, -- вывод нужных столбцов**

**(discount\*quantity\*price) disc,**

**order\_date**

**FROM sql.store\_carts c -- объединение данных из таблиц**

**JOIN sql.store\_products p ON c.product\_id=p.product\_id**

**JOIN sql.store\_delivery d ON c.order\_id=d.order\_id)**

**SELECT date\_trunc('month', order\_date)::date "date", -- выводит месяц заказа**

**round(sum(rev1-disc)) revenue --выручка, округленная до целого**

**FROM tot\_rev**

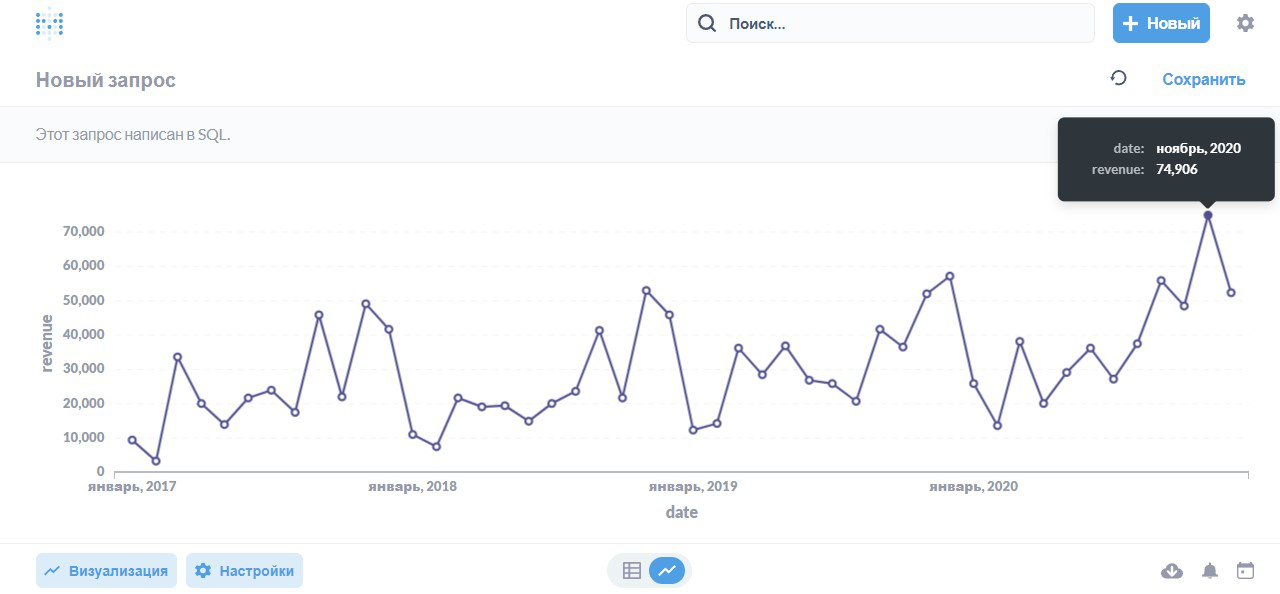
**WHERE order\_date IS NOT NULL -- условие вывода данных**

**GROUP BY 1**

**ORDER BY 1**

**Задание 27.5.2**

По полученным данным постройте график, который отразит динамику выручки.

****

**Выручка растёт. Наблюдается сезонность роста и падения выручки.**

**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, -- вывод нужных столбцов**

**subcategory,**

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

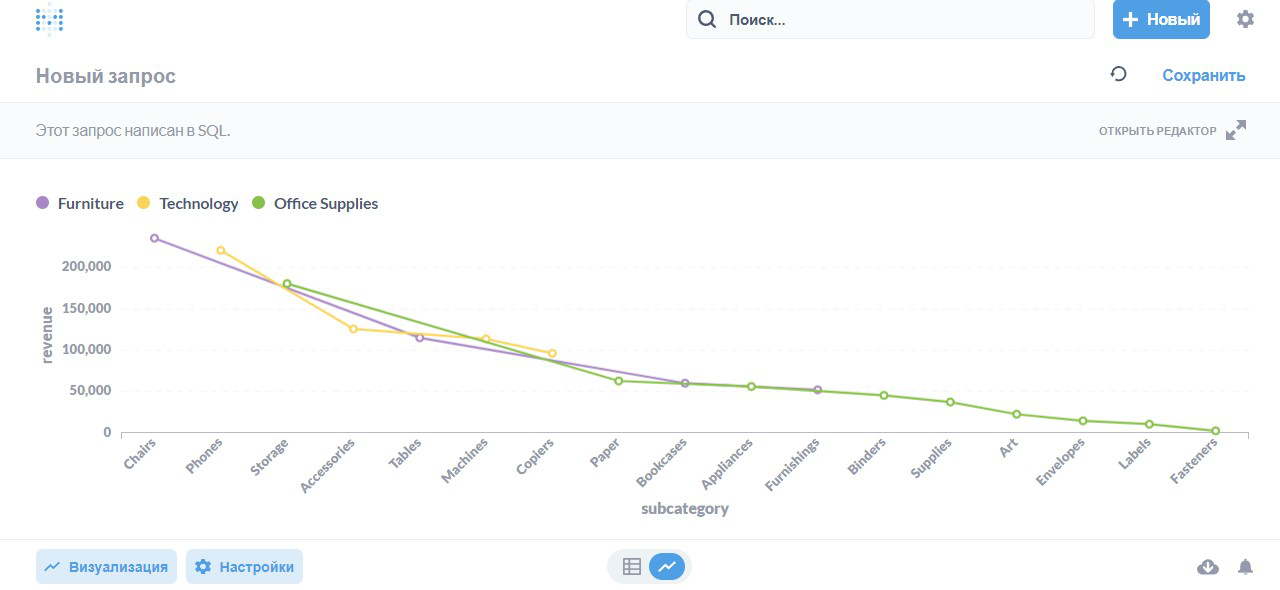
**FROM sql.store\_carts c -- объединение данных из таблиц**

**JOIN sql.store\_products p ON c.product\_id=p.product\_id**

**JOIN sql.store\_delivery d ON c.order\_id=d.order\_id**

**GROUP BY 1,2**

**ORDER BY 3 DESC**

****

**На основе результатов, полученных по выполнении запроса, делаем вывод: в Америке очень любят хорошие кресла — они в топе по продажам!**

### **Смотрим на товары**

**Важно не только понять, какие из товаров приносят больше всего выручки, но и определить, какую долю от общей выручки составляют их продажи.**

### 

### **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; Не забудьте просуммировать количество!**

**percent\_from\_total — доля от общей выручки в процентах. Не забудьте умножить на 100!**

**Примечание. Значения revenue и percent\_from\_total необходимо округлить до двух знаков после разделителя. Включите решение этого задания в итоговый документ. При необходимости снабдите дополнительными комментариями.**

**WITH top25 AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT sum((price\*quantity)-(price\*quantity\*discount)) AS total**

**FROM sql.store\_carts c -- объединение данных из таблиц**

**JOIN sql.store\_products p ON c.product\_id=p.product\_id)**

**SELECT product\_nm, -- вывод нужных столбцов**

**round(sum((price\*quantity)-(price\*quantity\*discount)), 2) AS revenue, -- выручка с учетом скидки,округленная до 2-го знака**

**sum(quantity) AS quantity,**

**round(sum((price\*quantity)-(price\*quantity\*discount))\*100/(total), 2) AS percent\_from\_total -- доля от общей выручки в процентах,округленная до 2-го знака**

**FROM sql.store\_carts c -- объединение данных из таблиц**

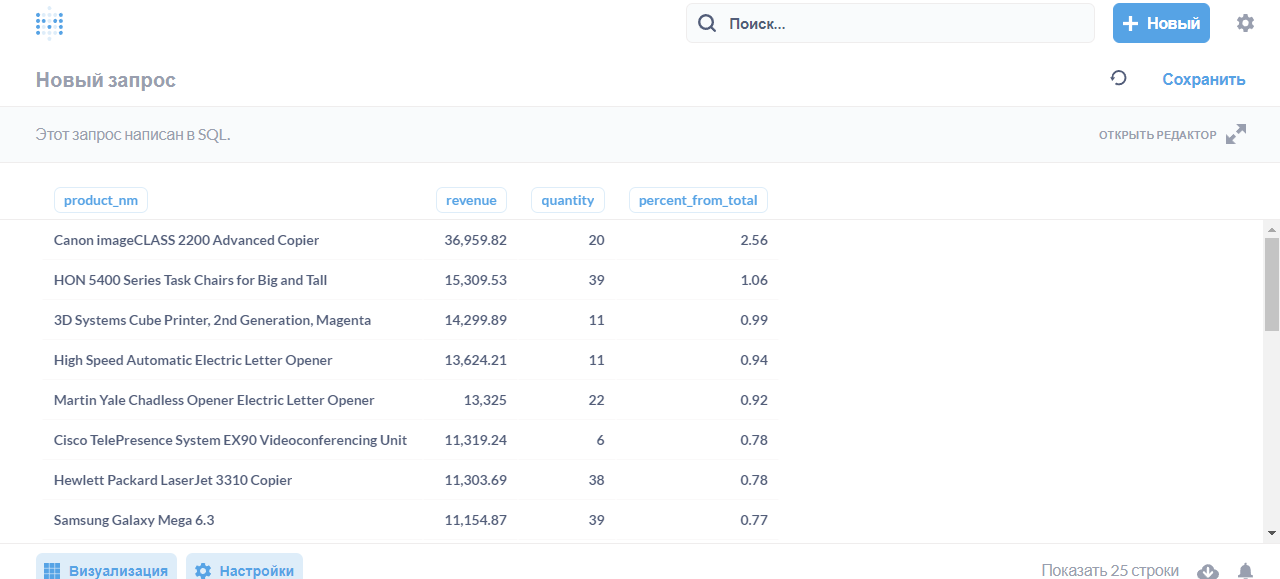
**JOIN sql.store\_products p ON c.product\_id=p.product\_id**

**JOIN top25 ON TRUE**

**GROUP BY 1, total**

**ORDER BY 2 DESC**

**LIMIT 25 -- вывод первых 25 строк**

****

## **27.6 PRACTICE. Шаг 2: Составляем портрет клиента**

**Чтобы бизнес рос и развивался, нужно знать своего клиента.**

**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.**

### **Считаем выручку клиентов по категориям**

**Начнём с простого — посмотрим, сколько у компании *B2B* и *B2C*-клиентов и какую долю от общей выручки они приносят.**

### **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 cu.category, -- вывод нужных столбцов**

**count(DISTINCT cust\_id) cust\_cnt,**

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

**FROM sql.store\_delivery d -- объединение данных из таблиц**

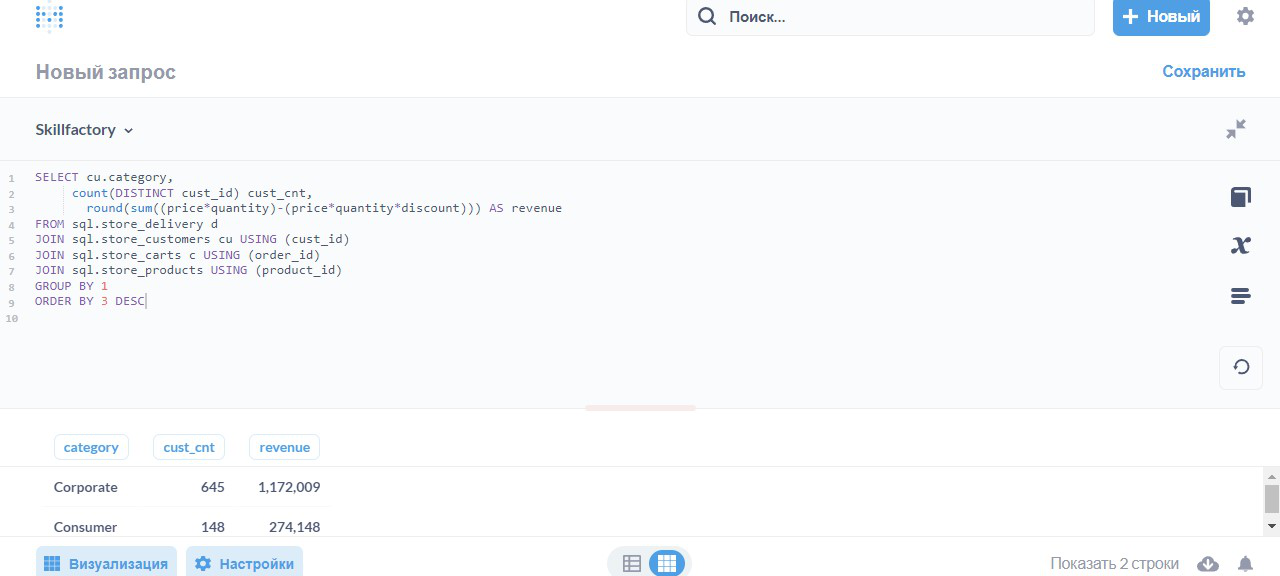
**JOIN sql.store\_customers cu USING (cust\_id)**

**JOIN sql.store\_carts c USING (order\_id)**

**JOIN sql.store\_products USING (product\_id)**

**GROUP BY 1**

**ORDER BY 3 DESC --сортировка по убыванию**

****

***B2B*-клиентов намного больше, и выручки они приносят тоже в разы больше!**

### **Составляем характеристику B2B-клиентов**

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.

### **Задание 27.6.2**

**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\_month AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT min(MIN (order\_date)) OVER(PARTITION BY cust\_id) AS min\_date, -- оконная функция ищет дату привлечения клиента**

**cust\_id**

**FROM sql.store\_delivery -- объединение данных из таблиц**

**JOIN sql.store\_customers USING (cust\_id)**

**WHERE category = 'Corporate' -- только корпоративные клиенты**

**GROUP BY 2**

**ORDER BY 2)**

**SELECT date\_trunc('month', min\_date)::date AS "month",**

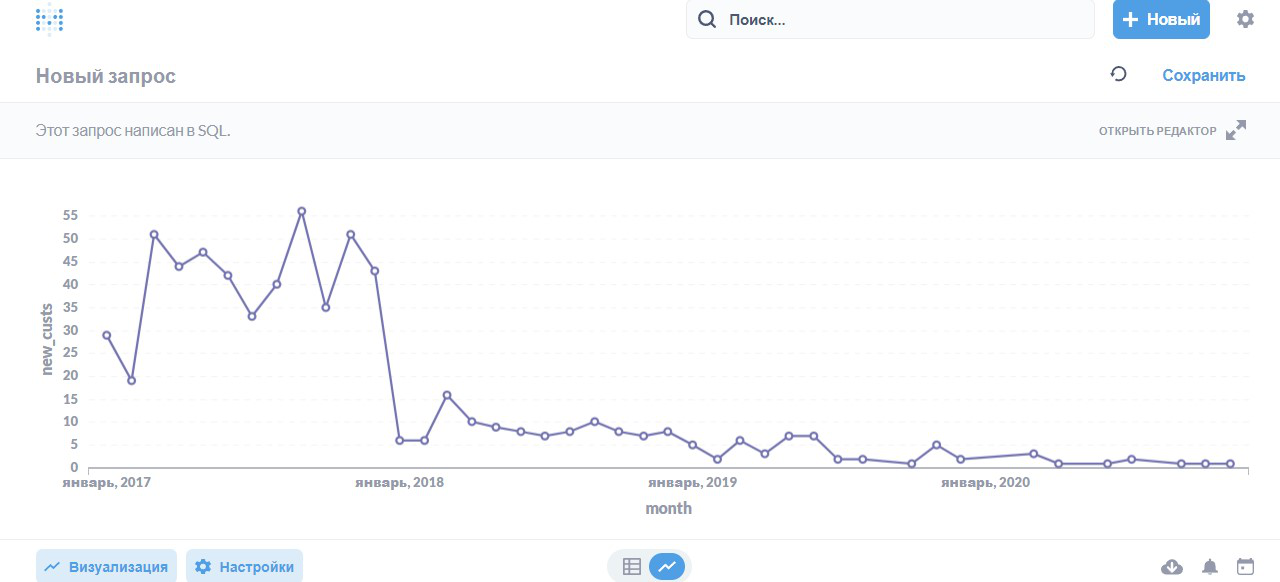
**count(fm.cust\_id) AS new\_custs -- количество новых клиентов**

**FROM first\_month fm**

**GROUP BY 1**

**ORDER BY 1**

**По полученным данным строим график динамики новых клиентов.**

****

**Как видно из графика, в самом начале было привлечено много клиентов, а после 2018 года новые клиенты почти не привлекались. Кажется, мы нашли точку роста.**

### **Составляем характеристику B2B-клиентов**

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

**Задание 27.6.4**

**Найдите ответы на следующие вопросы (*все ответы округлите до одного знака после разделителя*):**

### **Задание 27.6.4.1**

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

**WITH cnt\_distinct\_products AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT count(DISTINCT c.product\_id) AS cdp, -- вывод нужных столбцов**

**c.order\_id**

**FROM sql.store\_delivery d -- объединение данных из таблиц**

**JOIN sql.store\_customers cu USING (cust\_id)**

**JOIN sql.store\_carts c USING (order\_id)**

**WHERE category = 'Corporate'**

**AND order\_id IS NOT NULL**

**GROUP BY 2)**

**SELECT round(avg(cdp), 1) AS avg\_distinct\_products -- среднее количество различных товаров у корпоративных клиентов**

**FROM cnt\_distinct\_products**

**В среднем 2 различных товара в заказах у корпоративных клиентов.**

### **Задание 27.6.4.2**

1. **Какая в среднем сумма заказов у корпоративных клиентов?**

**WITH sum\_corporate\_orders AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT DISTINCT cust\_id, -- вывод нужных столбцов**

**sum((price\*quantity)-(discount\*quantity\*price)) sum\_order,**

**order\_id**

**FROM sql.store\_carts c -- объединение данных из таблиц**

**JOIN sql.store\_products p USING(product\_id)**

**JOIN sql.store\_delivery d USING(order\_id)**

**JOIN sql.store\_customers cu USING(cust\_id)**

**WHERE order\_id IS NOT NULL -- order\_id не равно нулю**

**AND cu.category = 'Corporate' -- только корпоративные клиенты**

**GROUP BY 1,**

**3)**

**SELECT round(avg(sum\_order), 1) avg\_corporate\_order**

**FROM sum\_corporate\_orders**

**Средняя сумма заказов корпоративных клиентов 285.9 usd**

### **Задание 27.6.4.3**

**3.Сколько в среднем различных офисов у корпоративных клиентов?**

**WITH cnt\_distinct\_zip\_corporate AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT count(DISTINCT zip\_code) cnt\_zip, -- вывод нужных столбцов**

**cust\_id**

**FROM sql.store\_delivery d -- объединение данных из таблиц**

**JOIN sql.store\_customers cu USING(cust\_id)**

**WHERE cu.category = 'Corporate'**

**GROUP BY 2)**

**SELECT round(avg(cnt\_zip), 1) avg\_zip\_corporate**

**FROM cnt\_distinct\_zip\_corporate**

**Среднее кол-во офисов для корпоративных клиентов 6.2 штуки.**

**Выводы на основании этих показателей: Для увеличения количества продаж и выручки нужно увеличить количество офисов.**

## **27.7 PRACTICE. Шаг 3: Анализируем логистику компании**

**Нам 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? (Ответ на этот вопрос поможет определить наиболее удачное место для офлайн-магазина.)**

### **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 is\_it\_late AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT ship\_mode, -- вывод нужных столбцов**

**count(order\_id) AS cnt\_orders,**

**count(CASE –- CASE – условный оператор языка SQL**

**Данный оператор позволяет осуществить проверку условий и возвратить в зависимости от выполнения того или иного условия тот или иной результат.**

**WHEN ship\_mode = 'Standard Class'**

**AND (ship\_date-order\_date)>6 THEN order\_id**

**WHEN ship\_mode = 'Second Class'**

**AND (ship\_date-order\_date)>4 THEN order\_id**

**WHEN ship\_mode = 'First Class'**

**AND (ship\_date-order\_date)>3 THEN order\_id**

**WHEN ship\_mode = 'Same Day'**

**AND (ship\_date-order\_date)>0 THEN order\_id**

**END) AS cnt\_late\_orders**

**FROM sql.store\_delivery**

**GROUP BY 1)**

**SELECT ship\_mode,**

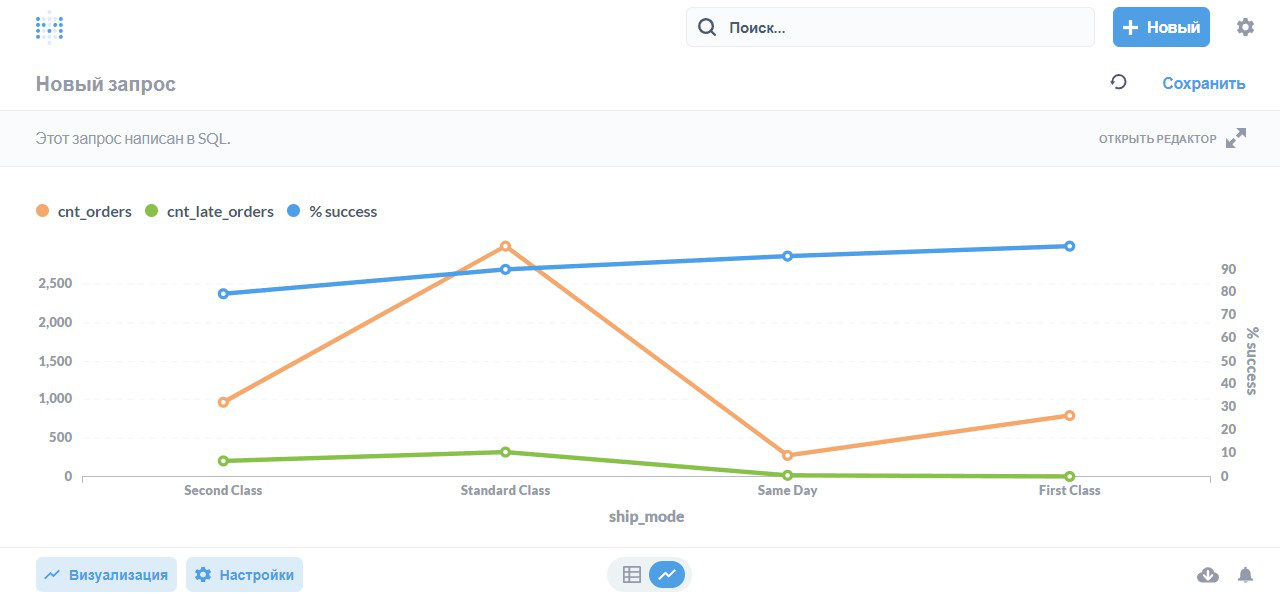
**cnt\_orders,**

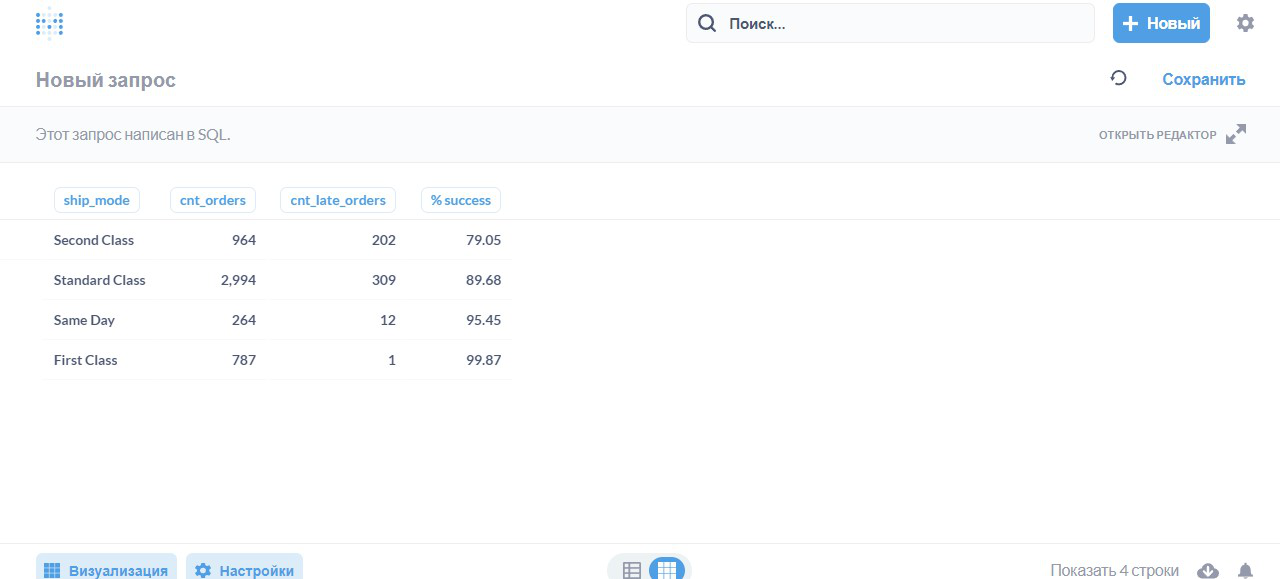
**cnt\_late\_orders,**

**round(((cnt\_orders-cnt\_late\_orders::decimal) / cnt\_orders::decimal\* 100),2) AS "% success" - - доля выполненных вовремя заказов, в процентах (% success), округлённая до двух знаков после запятой.**

**FROM is\_it\_late**

**ORDER BY 4**

****

****

**Заметили, что чаще всего с опозданием доходят заказы, отправленные вторым классом.**

**Из 964 заказов, отправленных вторым классом 202 пришло с опозданием, вовремя доставили 79,05%.**

**Из 2994 заказов, отправленных стандартным классом 309 пришло с опозданием, вовремя доставили 89,68%.**

**Из 787 заказов, отправленных первым классом 1 пришло с опозданием, вовремя доставили 99,87%.**

Посмотрим, насколько систематично это происходит. Возможно, у службы доставки были проблемы только на протяжении ограниченного периода.

**Задание 27.7.2**

Напишите запрос, чтобы вывести долю заказов, отправленных вторым классом, которые были доставлены с опозданием, по кварталам.

**WITH is\_secclass\_late\_by\_quarters AS - - (*CTE*) или обобщенное табличное выражение**

**(SELECT ship\_mode, -- вывод нужных столбцов**

**date\_trunc('quarter', order\_date)::date AS QUARTER,*--вычисляю кварталы, перевожу в тип даты***

**count(order\_id) AS cnt\_orders, -- количество заказов**

**count(CASE**

**WHEN ship\_mode = 'Second Class'**

**AND (ship\_date-order\_date)>4 THEN order\_id**

**ELSE NULL**

**END) AS cnt\_late\_orders**

**FROM sql.store\_delivery**

**WHERE ship\_mode = 'Second Class' -- только второй класс**

**GROUP BY 1,**

**QUARTER**

**ORDER BY 2)**

**SELECT QUARTER,**

**cnt\_orders, *--cnt\_late\_orders,***

**sum(((cnt\_late\_orders)/(cnt\_orders::decimal))\*100) AS**

**"%late\_sec\_class"--процент опоздавших доставок второго класса по отношению ко всем доставкам второго класса**

**FROM is\_secclass\_late\_by\_quarters**

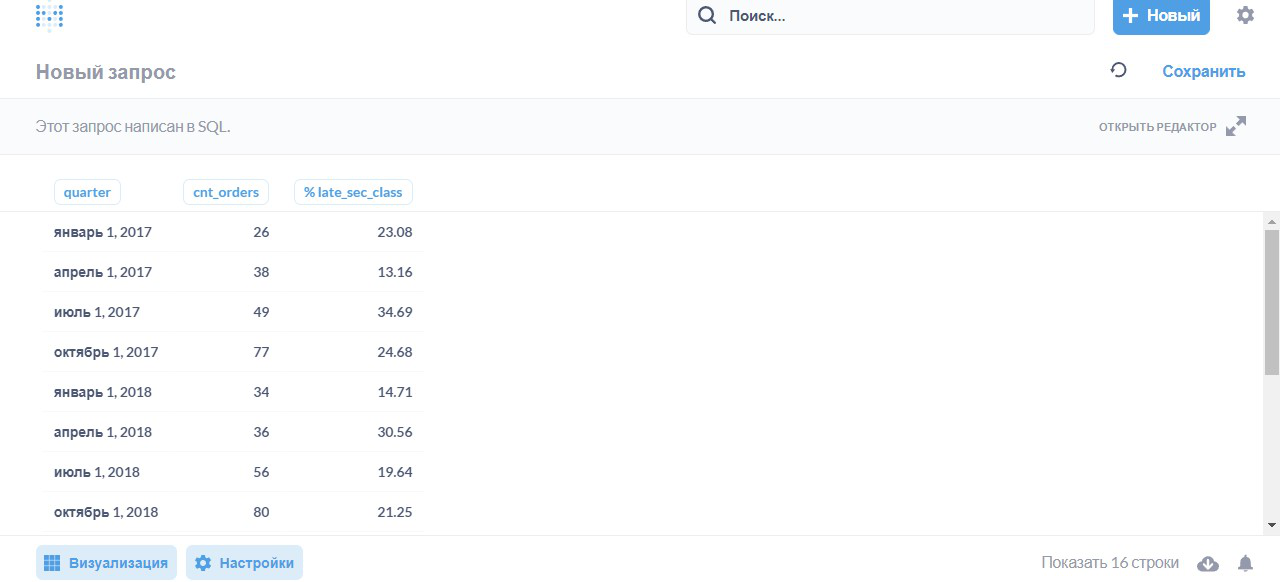
**WHERE ship\_mode = 'Second Class'**

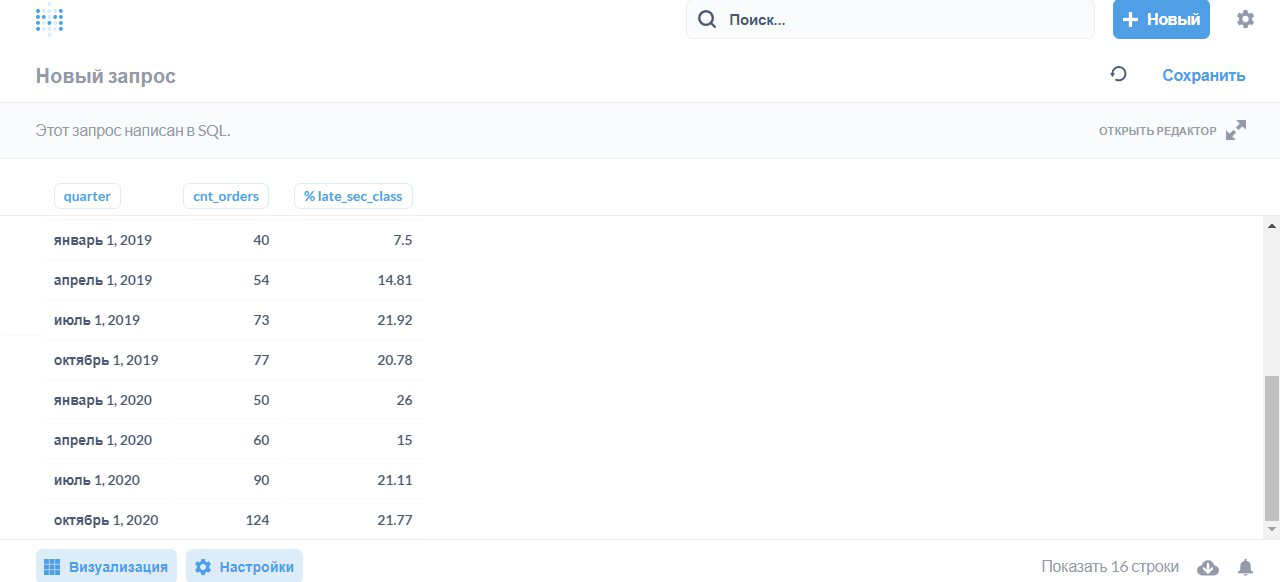
**GROUP BY 1,**

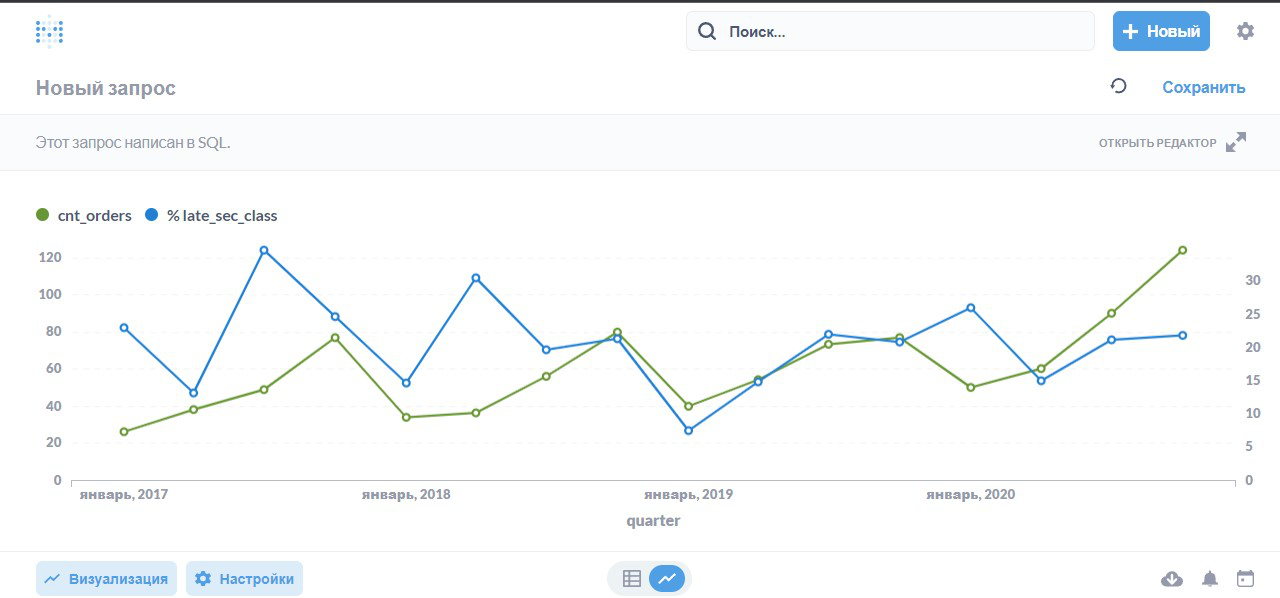
**2**

**ORDER BY 1 ASC**

**По получившимся данным строим график.**

****

****

****

**Задание 27.7.3**

**Заказы доставляются с опозданием в 12 из 16 кварталов от 15% до 34,69% от общего количества заказов. Можно сказать, что заказы доставляются с опозданием на протяжении наблюдаемого периода времени.**

### **Выбираем офлайн точку продаж**

Перейдём 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 найдём город и штат, куда совершается больше всего доставок.

### **Задание 27.7.4**

Какой штат наиболее популярный по количеству доставок?

**California - штат наиболее популярный по количеству доставок.**

### **Задание 27.7.5**

Какой город наиболее популярный по количеству доставок?

**New York City - город наиболее популярный по количеству 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?

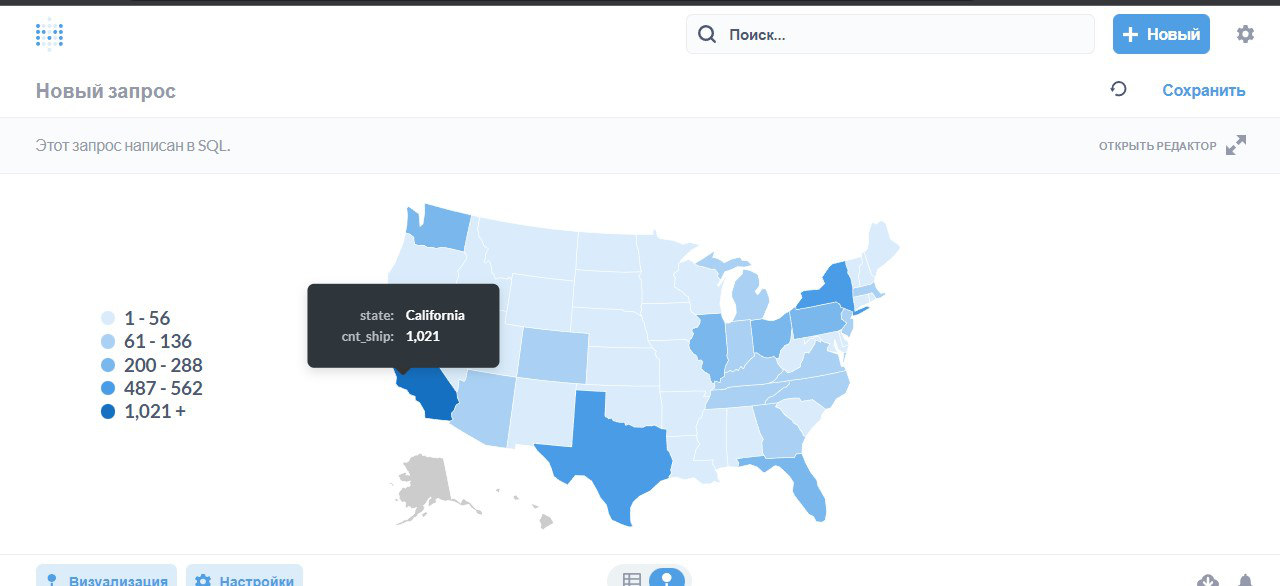
**SELECT state,**

**count(\*) AS cnt\_ship**

**FROM sql.store\_delivery d**

**GROUP BY 1**

**ORDER BY 2 DESC**

****

**Штат California является наиболее перспективным для открытия офлайн-магазина, т. к. в этом штате максимальное количество заказов по штатам.**

**Выводы по задачам**

1. **Оценить динамику продаж и распределение выручки по товарам.  
   Заказы осуществлялись в период с 2017-01-03 по 2020-12-30**

**Ответ в формате ГГГГ-ММ-ДД.**

**Всего 5009 заказов было выполнено за это время.**

**Объём общей выручки 1 446 157 usd**

**Выручка растёт. Наблюдается сезонность роста и падения выручки.**

**В Америке очень любят хорошие кресла — они в топе по продажам!**

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

***B2B*-клиентов намного больше, и выручки они приносят тоже в разы**

**больше!**

***B2B*-клиентов 545 и они принесли выручки 1 172 009 usd**

***B2B*-клиентов 148 и они принесли выручки 274 148 usd**

**выручка растёт за счёт увеличения продаж по старым клиентам,** **в**

**самом начале было привлечено много клиентов, а после 2018 года**

**новые клиенты почти не привлекались.**

**Нужно привлекать новых клиентов.**

**В среднем 2 различных товара в заказах у корпоративных клиентов.**

**Средняя сумма заказов корпоративных клиентов 285.9 usd**

**Среднее кол-во офисов для корпоративных клиентов 6.2 штуки.**

**Выводы на основании этих показателей:**

**Для увеличения количества продаж и выручки нужно**

**увеличить количество офисов и средней суммы заказов**

**корпоративных клиентов.**

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

**Заметили, что чаще всего с опозданием доходят заказы, отправленные**

**вторым классом.**

**Из 964 заказов, отправленных вторым классом 202 пришло с**

**опозданием, вовремя доставили 79,05%.**

**Из 2994 заказов, отправленных стандартным классом 309 пришло с**

**опозданием, вовремя доставили 89,68%.**

**Из 787 заказов, отправленных первым классом 1 пришло с**

**опозданием, вовремя доставили 99,87%.**

**Заказы доставляются с опозданием в 12 из 16 кварталов от 15% до**

**34,69% от общего количества заказов. Можно сказать, что заказы**

**доставляются с опозданием на протяжении наблюдаемого периода**

**времени.**

**Нужно подумать о заключении договора на поставку с**

**другими компаниями.**

**California - штат наиболее популярный по количеству**

**доставок.**

**New York City - город наиболее популярный по количеству**

**доставок.**

**Штат California является наиболее перспективным для**

**открытия офлайн-магазина, т. к. в этом штате максимальное**

**количество заказов по штатам.**

**New York City - город является наиболее перспективным для**

**открытия офлайн-магазина, т. к. в этом городе максимальное**

**количество заказов по городам.**