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

Елизавета Попова, darp-130

**Содержание**

1. [Описание отчета](#i1iv9v1kwu31)
2. [Эффективность продаж *Pens and Pencils*](#ib0r3n7bcja1)

2.1. [Динамика продаж](#99o9ewkfc63v)

2.2. [Выручка по категориям товаров](#2ogvrbfeldjr)

2.3. [Самые прибыльные товары](#5zqop5enh05w)

1. [Портрет клиента](#2zyi7wz7tsls)

3.1. [Выручка по категориям клиентов](#4suuq5nircu8)

3.2. [Характеристика B2B клиентов](#2c2dx6einatw)

1. [Логистика компании](#3ie30ax9417t)

4.1. [Эффективность доставки](#zf60b3roh5zt)

4.2. [Выбираем оффлайн-точку продаж](#7owvow4846oe)

1. [Выводы](#ujgboawti9lw)

**Описание отчета**

**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.

**Задача:**

Проанализировать работу компании с точки зрения ее 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).

**Эффективность продаж *Pens and Pencils***

Для оценки эффективности продаж Pens And Pencils ответим на следующие вопросы:

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

**Динамика продаж**

У нас есть данные о продажах компании за четыре года – с 2017-го по 2020-й. Общая сумма выручки составила - $1 446 154, а общее количество заказов – 5009 штук.

После небольшого спада в 2018-м году выручка компании каждый год увеличивалась.

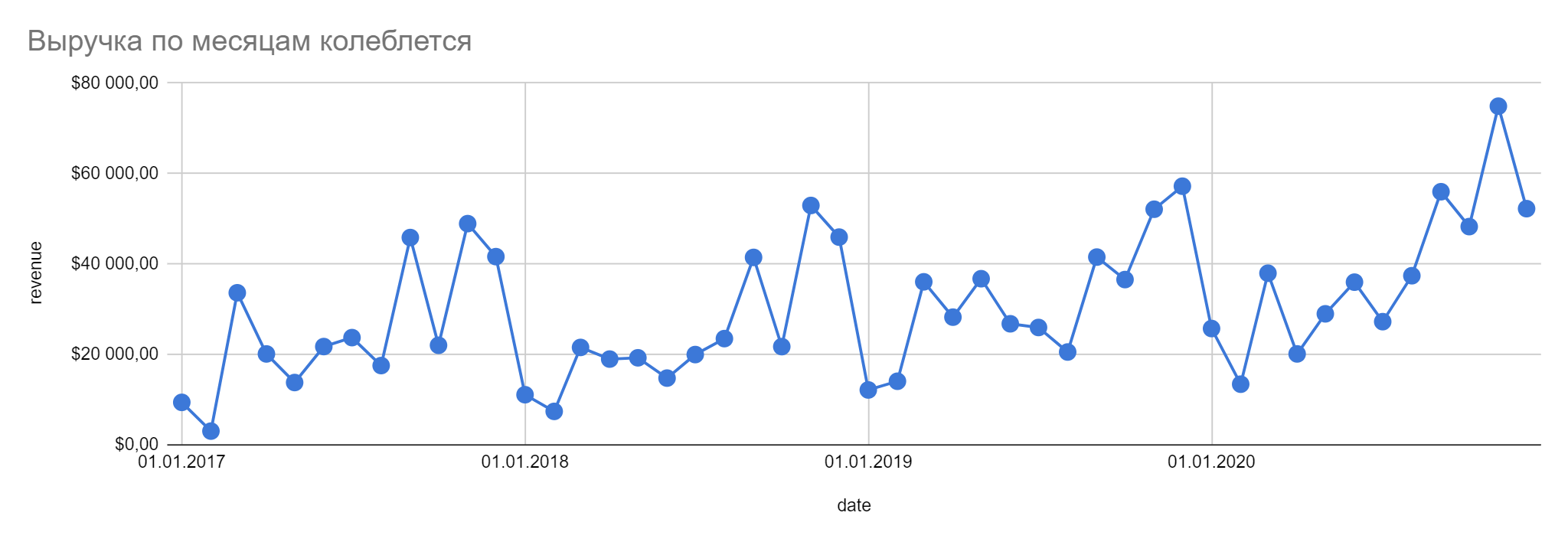
| **Год** | **Выручка** | **Изменения** |
| --- | --- | --- |
| 2017 | $301 395 |  |
| 2018 | $298 576 | -$2 819 |
| 2019 | $387 938 | $89 362 |
| 2020 | $458 248 | $70 310 |

27.5.1

Напишем запрос, который выведет сумму выручки по месяцам:

| with revenune\_per\_position as (**SELECT**  order\_id, quantity, discount,  price, **case** **when** discount = 0 **then** price\*quantity **else** (price\*quantity)\*(1-discount) **end** total **from** sql.store\_carts sc **join** sql.store\_products sp **on** sc.product\_id = sp.product\_id **order** **by** 1),  --рассчитаем выручку по каждой позиции в заказе с учетом скидок   revenue\_per\_order\_with\_date **as** (**select**  rpp.order\_id, order\_date, **sum**(total) total **from** revenune\_per\_position rpp **join** sql.store\_delivery sd **on** rpp.order\_id = sd.order\_id **group** **by** 1, 2 **order** **by** 1)  --рассчитаем выручку по каждому заказу  **select** date\_trunc('month', order\_date)::date date, **round**(**sum**(total)) revenue **from** revenue\_per\_order\_with\_date rpo **group** **by** 1 **order** **by** 1  --рассчитаем выручку по каждому месяцу |
| --- |

27.5.2



27.5.3, 27.5.4

Выручка по месяцам колеблется в зависимости от сезона, но в целом наблюдается тенденция роста.

Самые прибыльные месяцы для компании – сентябрь, ноябрь и декабрь. Максимальную выручку за четыре года мы видим в ноябре 2020-го – $74 906, – это 16,3% от выручки за 2020-й год.

В октябре выручка падает, но стоит отметить, что с 2019 года спад стал менее резким.

Самая низкая прибыль у компании в январе и феврале. Самая низкая выручка за четыре года была в январе 2017-го – $3 040, – 1% от выручки за 2017-й год.

**Выручка по категориям товаров**

27.5.5

Напишем запрос, который выведет сумму выручки по различным категориям и подкатегориям и отсортируем его по убыванию выручки:

| with revenue\_per\_position\_with\_categories as (**SELECT** order\_id, **category**, subcategory, **case** **when** discount = 0 **then** price\*quantity **else** (price\*quantity)\*(1-discount) **end** total **from** sql.store\_carts sc **join** sql.store\_products sp **on** sc.product\_id = sp.product\_id **order** **by** 1)  --рассчитаем выручку за каждую позицию в заказе с указанием категории и подкатегории этой позиции  **SELECT** **category**, subcategory, **round**(**sum**(total)) **from** revenue\_per\_position\_with\_categories **group** **by** 1, 2 **order** **by** 3 **desc**  --просуммируем выручку для каждой категории и подкатегории |
| --- |

**Топ-5 самых ходовых подкатегорий:**

| **Category** | **Subcategory** | **Revenue** | **% from total revenue** |
| --- | --- | --- | --- |
| Furniture | Chairs | $235 318 | 16,3% |
| Technology | Phones | $221 110 | 15,3% |
| Office Supplies | Storage | $179 736 | 12,4% |
| Technology | Accessories | $125 442 | 8,7% |
| Furniture | Tables | $114 532 | 7,9% |

Наибольшую выручку приносят продажи кресел, телефонов, товаров для хранения, технических аксессуаров и столов. Вместе эти подкатегории приносят компании 60,6% выручки.

Наименее прибыльные подкатегории – мелкие недорогие товары из категории “Office Supplies”, – зажимы, этикетки, конверты и пр.

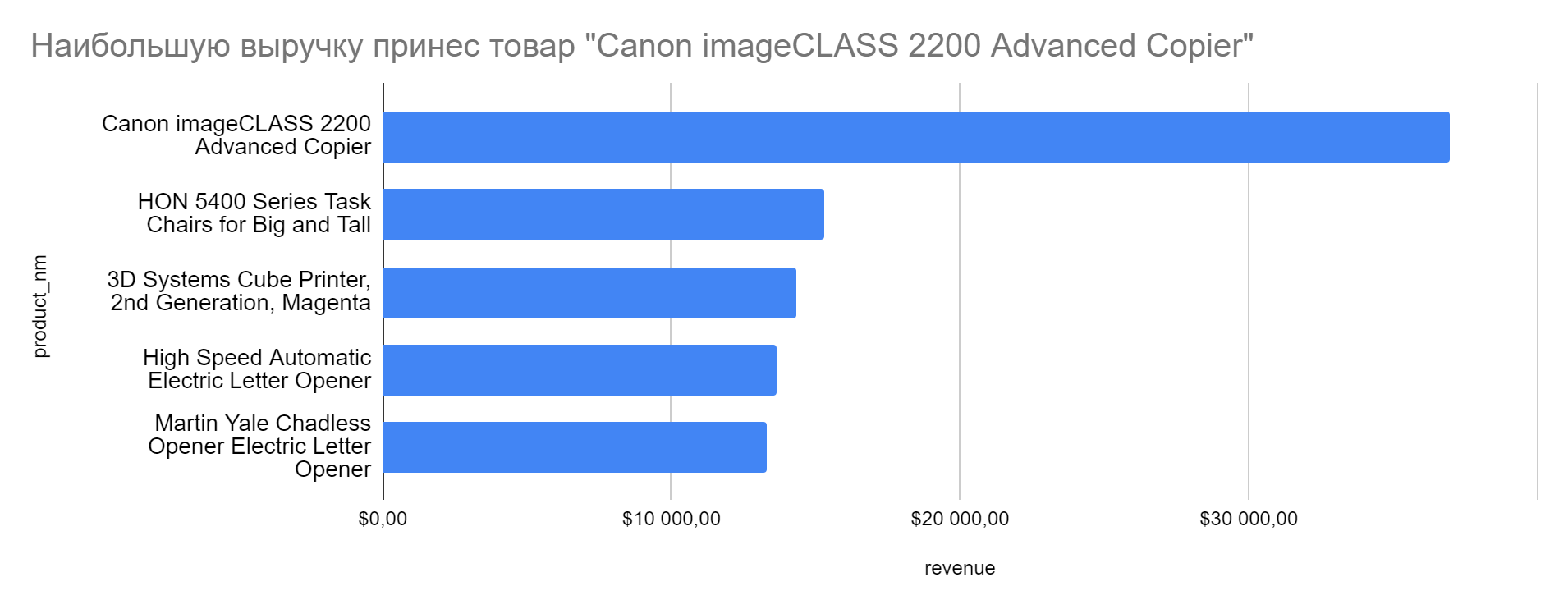
| **Category** | **Subcategory** | **Revenue** | **% from total revenue** |
| --- | --- | --- | --- |
| Office Supplies | Supplies | $37 457 | 2,59% |
| Office Supplies | Art | $22 148 | 1,53% |
| Office Supplies | Envelopes | $13 809 | 0,95% |
| Office Supplies | Labels | $10 345 | 0,72% |
| Office Supplies | Fasteners | $2 459 | 0,17% |

**Самые прибыльные товары**

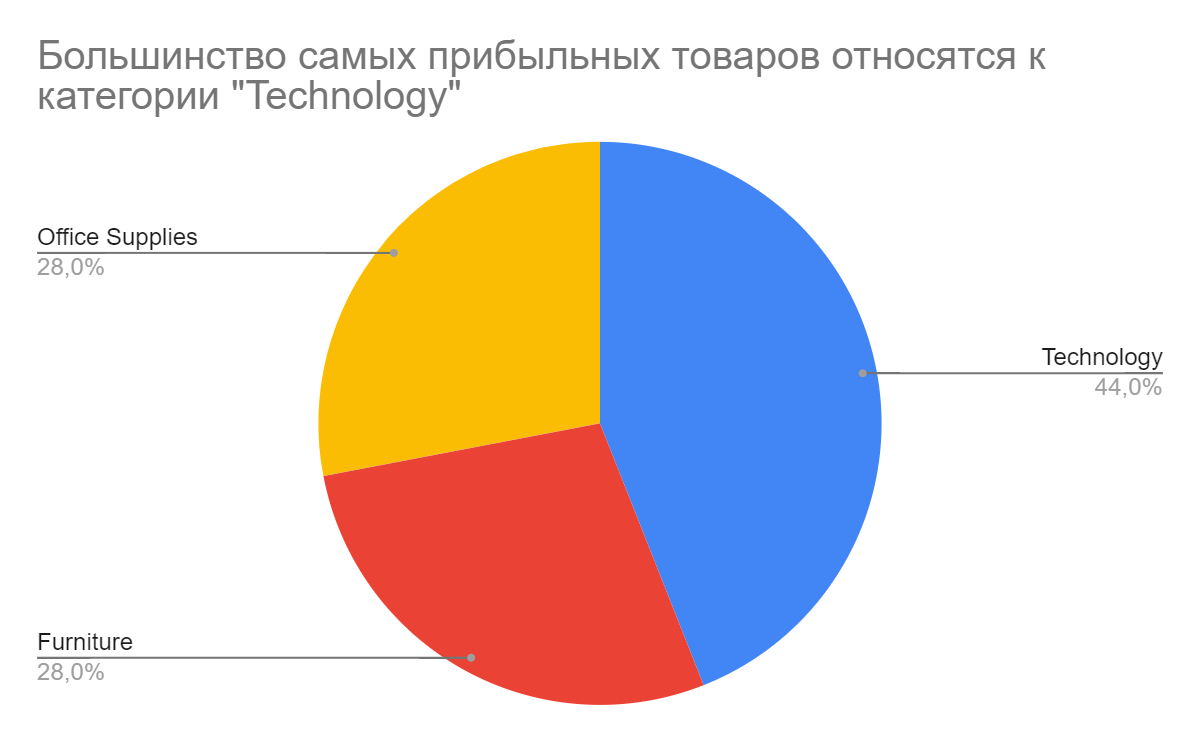
27.5.6

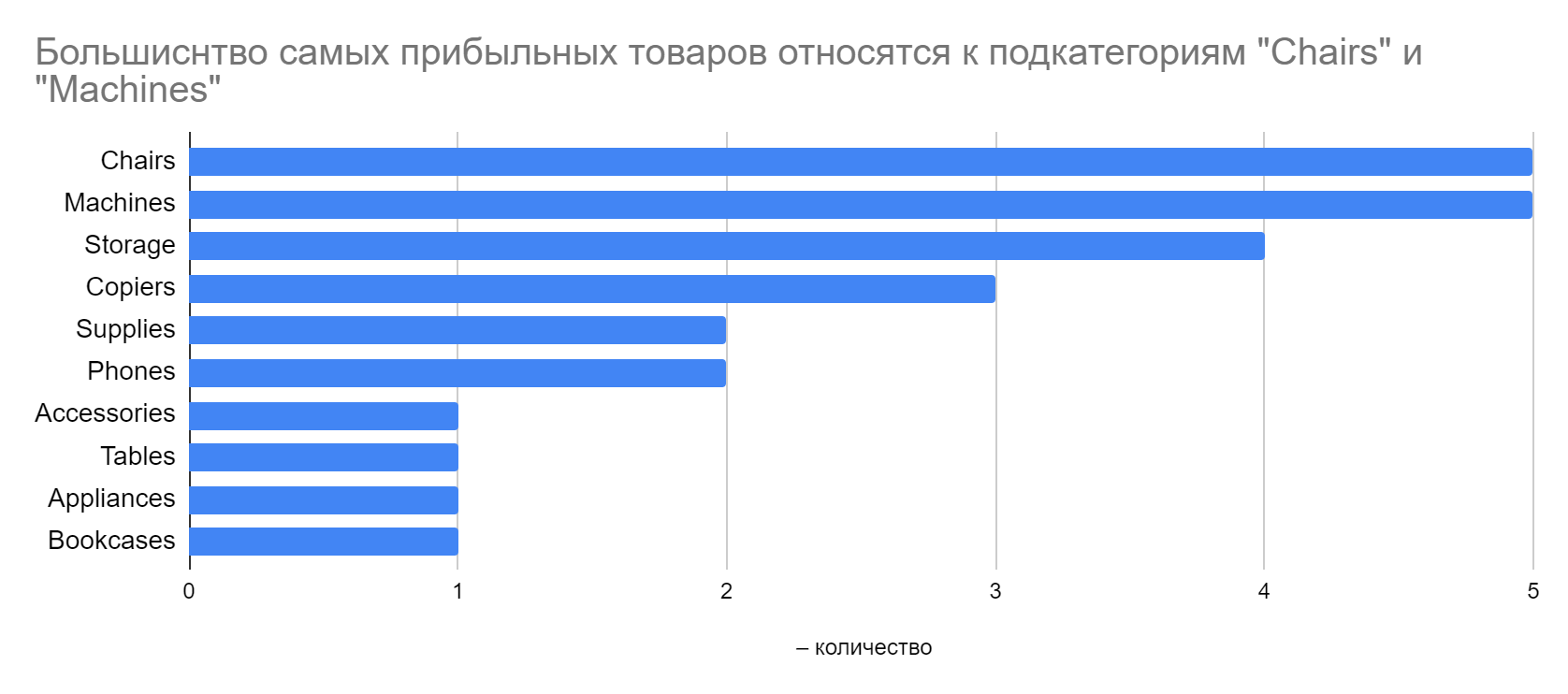
Напишем запрос, который покажет нам топ-25 товаров, приносящих больше всего выручки. Выведем выручку, количество проданных товаров и долю от общей выручки в процентах:

| with revenue\_per\_product\_name as (**SELECT** product\_nm, quantity, **case** **when** discount = 0 **then** price\*quantity **else** (price\*quantity)\*(1-discount) **end** total **from** sql.store\_carts sc **join** sql.store\_products sp **on** sc.product\_id = sp.product\_id **order** **by** 1),  --рассчитаем выручку с учетом скидок для каждого товара и выведем количество проданных единиц  total\_revenue **as** (**select**  **sum**(total) total\_rev **from** revenue\_per\_product\_name)  --просуммируем выручку за все товары  **SELECT** product\_nm, **round**(**sum**(total), 2) revenue, **sum**(quantity) quantity, **round**((**sum**(total) / total\_rev) \* 100, 2) percent\_from\_total **from** revenue\_per\_product\_name **cross** **join** total\_revenue **group** **by** product\_nm, total\_rev **order** **by** 2 **desc** **limit** 25  --выведем количество проданных единиц товара и выручку с него, рассчитаем процент от общей выручки для 25 самых прибыльных товаров |
| --- |



Суммарно выручка от этих товаров составляет 18% выручки ($262 180) за четыре года. Самый продаваемый товар – *Canon imageCLASS 2200 Advanced Copier*. Его продажи составляют 2,56% от общей выручки, продано 20 штук. Затем идет кресло *HON 5400 Series Task Chairs for Big and Tall* – 1,06% от общей выручки, продано 39 штук. И на третем месте *3D Systems Cube Printer, 2nd Generation, Magenta –* 0,99% от общей выручки, продано 11 штук.





**Портрет 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.

**Выручка по категориям клиентов**

27.6.1

Напишем запрос, который выведет количество клиентов и выручку по категориям клиентов, отсортируем вывод по убыванию выручки:

| with revenune\_per\_position as (**SELECT** order\_id, quantity, discount,  price, **case** **when** discount = 0 **then** price\*quantity **else** (price\*quantity)\*(1-discount) **end** total **from** sql.store\_carts sc **join** sql.store\_products sp **on** sc.product\_id = sp.product\_id),  --рассчитаем выручку за каждую позицию в заказе с учетом скидок  customer\_categories\_and\_orders **as** (**select** order\_id, d.cust\_id, **category** **from** sql.store\_delivery d **join** sql.store\_customers c **on** d.cust\_id = c.cust\_id)  --выведем информацию о том, какие клиенты совершили заказ  **select** **category**, **count**(**distinct** cust\_id) cust\_cnt, **round**(**sum**(total)) revenue **from** customer\_categories\_and\_orders ccao **join** revenune\_per\_position rpp **on** ccao.order\_id = rpp.order\_id **group** **by** **category** **order** **by** 3 **desc**  --рассчитаем выручку по категориям клиентов |
| --- |

| **Category** | **Cust\_cnt** | **Revenue** |
| --- | --- | --- |
| Corporate | 645 | $1 172 009 |
| Consumer | 148 | $274 148 |

Наибольшую выручку приносят корпоративные клиенты - 81% от всей выручки. Поэтому присмотримся к ним повнимательнее.

**Характеристика 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.

27.6.2, 27.6.3

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



| with corporate\_clients\_start\_date as (**select** d.cust\_id, **min**(order\_date) start\_date **from** sql.store\_delivery d **join** sql.store\_customers c **on** d.cust\_id = c.cust\_id **where** **category** = 'Corporate' **group** **by** d.cust\_id **order** **by** 1)  --выведем информацию только о корпоративных клиентах и дате их первого заказа  **select** date\_trunc('month', start\_date)::date, **count**(cust\_id) **from** corporate\_clients\_start\_date **group** **by** 1 **order** **by** 1  --рассчитаем количество корпоративных клиентов за каждый квартал |
| --- |



В 2017-м у *Pens and Pencils* появилось больше всего корпоративных клиентов, но к началу 2018-го оно резко упало, а к концу 2020-го новых клиентов практически не появляется. Привлечение новых B2B клиентов – точка роста для компании.

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

27.6.4

Узнаем, сколько в среднем различных товаров в заказах у корпоративных клиентов, а также среднюю сумму заказа:

| with corporate\_clients as (**select** c.cust\_id, order\_id, zip\_code **from** sql.store\_customers c **join** sql.store\_delivery d **on** c.cust\_id = d.cust\_id **where** **category** = 'Corporate'),  --выведем id клиентов, id заказов и индексы только корпоратиных клиентов  revenune\_per\_position **as** (**SELECT**  order\_id, quantity, discount,  price, sc.product\_id, **case** **when** discount = 0 **then** price\*quantity **else** (price\*quantity)\*(1-discount) **end** total **from** sql.store\_carts sc **join** sql.store\_products sp **on** sc.product\_id = sp.product\_id **order** **by** 1),  --рассчитаем выручку и количество единиц каждой позиции в заказе  revenue\_per\_order **as** (**select**  rpp.order\_id, **sum**(total) total, **count**(**distinct** product\_id) positions **from** revenune\_per\_position rpp **join** sql.store\_delivery sd **on** rpp.order\_id = sd.order\_id **group** **by** 1 **order** **by** 1)  --рассчитаем выручку и количество единиц в каждом заказе  **select** **round**(**avg**(total), 1) avg\_revenue, **round**(**avg**(positions), 1) avg\_positions **from** revenue\_per\_order rpo **join** corporate\_clients cc **on** rpo.order\_id = cc.order\_id  --рассчитаем среднее количество единиц в заказе и среднюю выручку только у корпоративных клиентов |
| --- |

Помимо этого, выведем среднее количество офисов у клиентов:

| with corporate\_clients as (**select** c.cust\_id, order\_id, zip\_code **from** sql.store\_customers c **join** sql.store\_delivery d **on** c.cust\_id = d.cust\_id **where** **category** = 'Corporate'),  --отфильтруем корпоративных клиентов  distinct\_zip\_codes **as** (**select** cust\_id, **count**(**distinct** zip\_code) zip **from** corporate\_clients **group** **by** 1)  --рассчитаем количество офисов у клиентов  **select** **round**(**avg**(zip), 1) **from** distinct\_zip\_codes  --рассчитаем среднее количество офисов |
| --- |

В среднем корпоративные клиенты покупают 2 единицы товара в заказе, а средняя сумма заказа составляет $285,9. Также у наших клиентов в среднем 6,2 различных офисов.

Мы можем повысить среднюю сумму заказа с помощью дополнительных продаж сопутствующих товаров. Также можно предлагать клиенту совершить покупку сразу для нескольких офисов.

**Логистика 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.)

**Эффективность доставки**

В *Pens and Pencils* есть несколько типов доставки:

| **Тип доставки** | **Описание** | **Планируемое время доставки** |
| --- | --- | --- |
| Standard Class | Стандартная доставка | Доставка в течение шести дней |
| Second Class | Доставка вторым классом | Доставка в течение четырёх дней |
| First Class | Доставка первым классом | Доставка в течение трёх дней |
| Same Day | Экспресс-доставка | Экспресс-доставка в тот же день |

Определим, какая доля заказов выполняется в срок по каждой категории.

27.7.1

Напишем запрос, который выведет общее количество доставок по категориям, количество заказов, доставленных с опозданием и долю заказов, доставленных в срок:

| with delivered\_for as (**select** order\_id, ship\_date, order\_date, ship\_date - order\_date days\_in\_delivery, ship\_mode **from** sql.store\_delivery),  --рассчитаем, сколько дней заказ был в доставке  late\_orders **as** (**select** ship\_mode, **count**(**distinct** order\_id) orders\_cnt, **count**(**case** **when** ship\_mode **in** ('Same Day') **and** days\_in\_delivery > 0 **then** order\_id **when** ship\_mode **in** ('First Class') **and** days\_in\_delivery > 3 **then** order\_id **when** ship\_mode **in** ('Second Class') **and** days\_in\_delivery > 4 **then** order\_id **when** ship\_mode **in** ('Standard Class') **and** days\_in\_delivery > 6 **then** order\_id **end**) late\_orders\_cnt **from** delivered\_for **group** **by** 1)  --рассчитаем количество заказов, пришедших с опозданием  **select** ship\_mode, orders\_cnt, late\_orders\_cnt, **round**((orders\_cnt::decimal - late\_orders\_cnt::decimal) / orders\_cnt::decimal \* 100, 2) "%\_success" **from** late\_orders **order** **by** 4  --рассчитаем долю заказов, пришедших вовремя |
| --- |

Как мы видим, меньше всего опозданий у типа доставки “First Class” – всего одно за четыре года. В категории “Same Day” опозданий было больше – 12. Почти 90% заказов были доставлены вовремя в типе “Standard Class”.

Чаще всего с опозданием доходят заказы, отправленные вторым классом – лишь 79% заказов доставлены вовремя.

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

27.7.2

Выведем долю заказов, отправленных вторым классом, которые были доставлены с опозданием, по кварталам:

| with second\_class\_orders as (**select** order\_id, ship\_date, order\_date, ship\_date - order\_date days\_in\_delivery, ship\_mode **from** sql.store\_delivery **where** ship\_mode = 'Second Class'),  --рассчитаем, сколько дней шли заказы типы "Second Class"  late\_orders\_second\_class **as** (**select** date\_trunc('quarter', order\_date)::date **quarter**, **count**(order\_id) orders, **count**(order\_id) filter(**where** days\_in\_delivery > 4) late\_orders **from** second\_class\_orders **group** **by** 1 **order** **by** 1)  --рассчитаем общее количество заказов и количество заказов, пришедших с опозданием  **select** **quarter**, orders, late\_orders, late\_orders::decimal / orders::decimal \* 100 late\_orders\_percent **from** late\_orders\_second\_class **order** **by** 1  --рассчитаем долю заказов, пришедших с опозданием |
| --- |



27.7.3

Заказы доставляются с опозданием постоянно, самый высокий процент был в третьем квартале 2017 года – 34,69%. Компании стоит поработать над улучшением сервиса в данном типе доставки.

**Выбираем оффлайн-точку продаж**

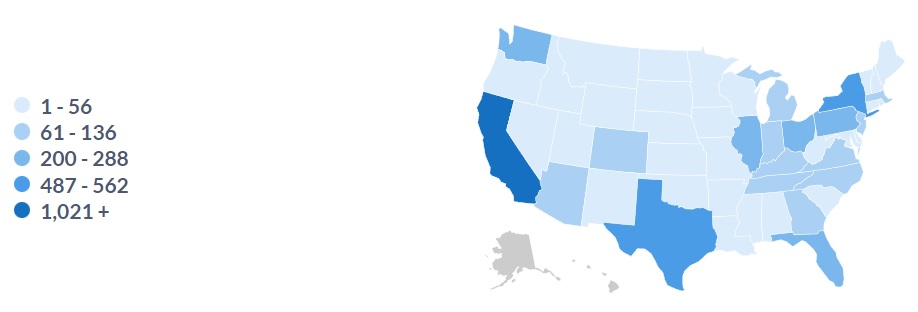
Перейдем к исследованию места для открытия оффлайн-магазина.

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

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

Напишем запрос, который выведет количество доставок по штатам.

| **select** state, **count**(**distinct** order\_id) **from** sql.store\_delivery **group** **by** 1 **order** **by** 2 **desc** |
| --- |



Самые популярные штаты: Калифорния, Нью-Йорк, Техас

Также выясним самые популярные города для доставки:

| **select** city, **count**(**distinct** order\_id) **from** sql.store\_delivery **group** **by** 1 **order** **by** 2 **desc** |
| --- |

Это Нью-Йорк, Лос Анджелес, Сан Франциско

Открытие магазина в штате Калифорния поможет снизить стоимость доставки, так как туда отправляется больше всего заказов. Лучше всего подойдет Лос Анджелес – это второй по популярности город после Нью-Йорка. Это место кажется нам наилучшим для открытия магазина, но можно рассмотреть и другие варианты.

Также возможно открытие нового магазина и в городе Нью-Йорк. Доставка товаров на данный момент туда обходится дороже, чем в Лос-Анджелес. Однако, хоть и доставок в Нью-Йорк больше (450), чем в Лос-Анджелес (384), доставок в Калифорнию (1021) больше практически в два раза, чем в штат Нью-Йорк (562).

Есть плюсы и у открытия магазина при складе в Хьюстоне: оффлайн продажи привлекут больше клиентов и само открытие магазина обойдется дешевле. Но сэкономить на доставке в другие штаты не удастся.

**Выводы**

Итак, по задачам отчета можно сделать следующие выводы:

1. **Эффективность продаж.** Выручка *Pens and Pencils* в целом растет, но колеблется в зависимости от сезона. Сезонные месяцы – это сентябрь, ноябрь и декабрь. “Не сезон” наступает в январе и феврале. Наибольшую прибыль компании приносит продажа кресел, телефонов, товаров для хранения, аксессуаров к технике и столов. Товар, принесший больше всего прибыли за все время ($36 960) – *Canon imageCLASS 2200 Advanced Copier.*
2. **Портрет клиента.** 81% выручки приносят корпоративные клиенты. В основном это старые клиенты, новых компания привлекает очень мало, – это точка роста для *Pens and Pencils*. Корпоративные клиенты компании в среднем добавляют 2 разные позиции в заказ и средняя сумма заказа составляет $285,9. В среднем у клиентов 6,2 различных офисов. Помимо привлечения новых клиентов, увеличить выручку со старых можно с помощью дополнительных продаж сопутствующих товаров, а также продаж сразу в несколько офисов клиента.
3. **Логистика компании и оффлайн-магазин.** Все типы доставок в компании работают достаточно эффективно, кроме типа “Second Class”, в котором лишь 79% заказов были доставлены вовремя. Заказы в данном типе доставки приходят с опозданием систематически – здесь компании стоит поработать над улучшением сервиса.

Для открытия оффлайн-магазина лучше всего подойдет Лос Анджелес, штат Калифорния. Калифорния – самый популярный штат для доставки, а Лос Анджелес – второй по популярности город. Открытие магазина позволит сэкономить на доставке со склада в Хьюстоне и увеличить выручку.