# **Анализ компании “Pens and Pencils”**

### **ЛЕГЕНДА**

В услугах аналитика нуждается молодая, но перспективная 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**

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

**ЦЕЛИ**

1. Дать оценку **динамике продаж** и **распределению 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**).

# 

# **Описание данных**

### **Таблица “store customers”**

* + Представляет собой справочник клиентов.

| **Название поля** | **Тип данных** | **Описание** |
| --- | --- | --- |
| **cust\_id** | varchar | id клиента, первичный ключ таблицы |
| **cust\_nm** | text | имя клиента |
| **category** | text | тип клиента; значения: *'Consumer'* (B2C-клиент, розничный покупатель), *'Corporate'* (B2B-клиент, компания) |

### 

### **Таблица “store products”**

* + Представляет собой справочник товаров.

| **Название поля** | **Тип данных** | **Описание** |
| --- | --- | --- |
| **id** | integer | id записи в таблице |
| **order\_id** | varchar | id заказа; внешний ключ к таблице **store\_delivery** |
| **product\_id** | varchar | id товара; внешний ключ к таблице **store\_products** |
| **quantity** | integer | количество единиц отдельного товара в заказе |
| **discount** | numeric | доля скидки на определенный товар; значение 0.1 означает скидку в 10% |

### **Таблица “store carts”**

* + Представляет собой таблицу со списком товаров в заказе, их количеством, а также скидкой для каждого товара.

| **Название поля** | **Тип данных** | **Описание** |
| --- | --- | --- |
| **product\_id** | varchar | id товара; первичный ключ таблицы |
| **category** | text | категория товара |
| **subcategory** | text | подкатегория товара |
| **product\_nm** | text | название, свойства: цвет, модель, бренд и т. д. |
| **price** | numeric | цена товара |

### 

### **Таблица “store delivery”**

* + Хранит информацию о заказах и их доставке.

| **Название поля** | **Тип данных** | **Описание** |
| --- | --- | --- |
| **order\_id** | varchar | id заказа, первичный ключ таблицы |
| **order\_date** | date | дата заказа |
| **ship\_date** | date | фактическая дата доставки |
| **ship\_mode** | text | тип доставки |
| **state** | text | штат доставки |
| **city** | text | город доставки |
| **zip\_code** | text | индекс доставки |
| **cust\_id** | varchar | id клиента; внешний ключ к таблице **store\_customers** |

Вся работа над проектом проводилась в **Metabase**.

Ссылки на все таблицы в формате ***csv*** для проверки:

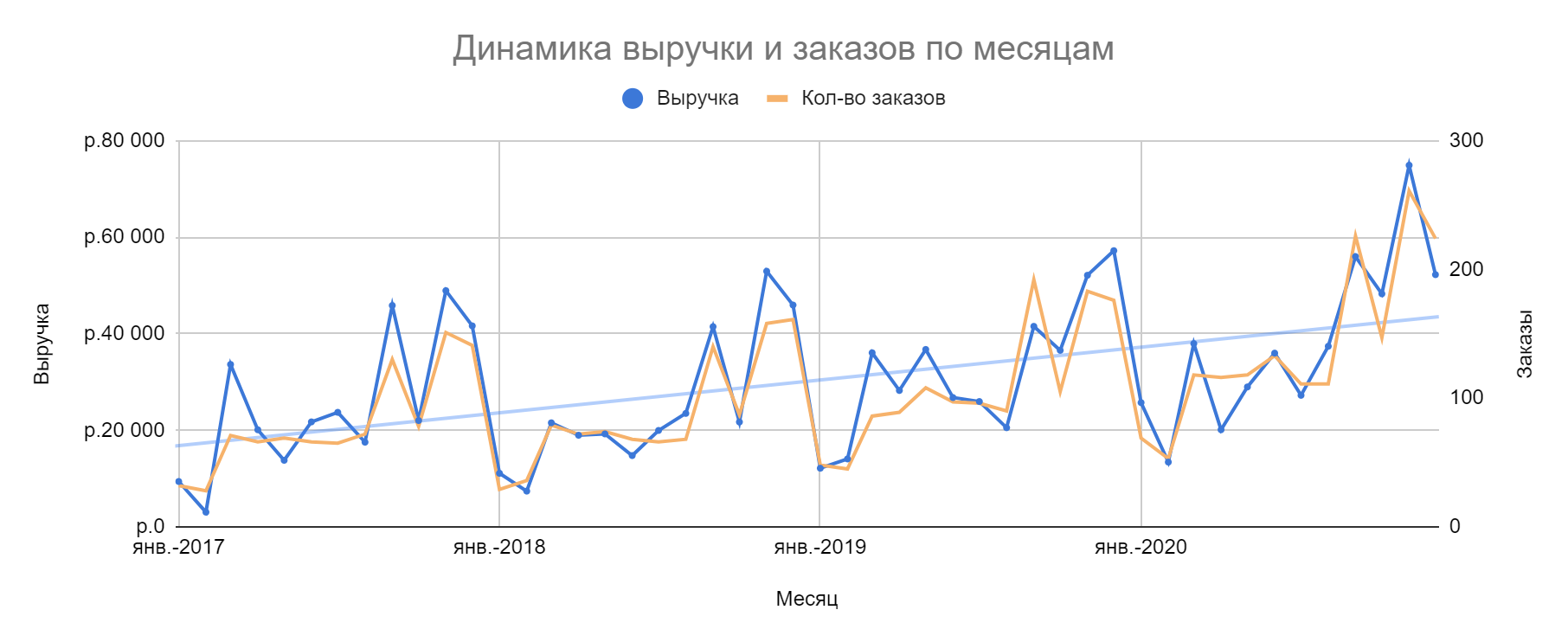
* [store customers;](https://drive.google.com/file/d/172u1sxz0T23fB88D1shwAkRzk1K1VgMY/view?usp=sharing)
* [store products;](https://drive.google.com/file/d/1Bx6y2V7OyBVz6YbI0Kxv-hXthTndmrBr/view?usp=sharing)
* [store carts;](https://drive.google.com/file/d/1ePZou1po8XU66lYNidF4DhkG-j6wR6x6/view?usp=sharing)
* [store delivery.](https://drive.google.com/file/d/1n3YiaHsastufNLJja0llrA2HdRIS365o/view?usp=sharing)

## **Часть 1. Определение эффективности продаж**

* Для начала посчитаем сумму выручки по месяцам и количество заказов:

| SELECT date\_trunc('month', order\_date)::date AS date,  count(distinct order\_id) as orders\_quantity,-- количество заказов в месяц  round(sum(price \* (1 - discount) \* quantity)) AS revenue -- округленный подсчет выручки FROM sql.store\_carts sc JOIN sql.store\_delivery sd USING(order\_id) JOIN sql.store\_products sp USING(product\_id) GROUP BY 1 ORDER BY 1 |
| --- |

Сортируем по месяцам для корректного отображения в графике.

По полученным данным построим график, который отразит динамику выручки.

###### График 1 - Динамика выручки и заказов по месяцам.

Исходя из графика выручки, можно прийти к выводу, что выручка имеет **сезонный характер** - в марте замечен скачок в выручке, затем небольшое снижение и стагнация на протяжении нескольких месяцев вплоть до августа, где опять наблюдается резкое повышение количества заказов в сентябре и ноябре с последующим спадом в декабре-январе.

В целом, за весь рассматриваемый период **выручка имеет тенденцию к росту за счет постепенного увеличения количества заказов**, что подтверждает линия тренда.

Рекомендуется рассмотреть возможность стимулирования продаж существующих клиентов и увеличения среднего чека заказа, ввести программу лояльности, если такой еще нет, проведение акций и сезонных предложений.

* Далее найдем сумму выручки по различным категориям и подкатегориям:

| SELECT category,  subcategory,  round(sum(price \* (1 - discount) \* quantity)) AS revenue FROM sql.store\_carts sc JOIN sql.store\_delivery sd USING(order\_id) JOIN sql.store\_products sp USING(product\_id) GROUP BY 1, 2 ORDER BY 3 DESC |
| --- |

### 

###### График 2 - Выручка по категориям товаров.

### Из Графика 2, представленного выше, видно, что **самая прибыльная категория - Technology (38,47%)**, где телефоны (Phones) занимают первое место с 15,29% от общего объема продаж. При этом, **самой прибыльной подкатегорией товаров являются кресла (Chairs)**, на долю которых приходится **16,27%** продаж. Среди офисных принадлежностей (Office Supplies) лидируют товары для организации хранения (Storage) с 12,43%.

* Определим топ-25 товаров по объему выручки, а также к каким категориям относятся эти товары.

| WITH total\_revenue AS (SELECT   sum(price \* (1 - discount) \* quantity) AS total\_rev  FROM sql.store\_carts sc  JOIN sql.store\_products sp USING(product\_id)) -- создаем CTE для подсчета общего объема продаж  SELECT category,  subcategory,  product\_nm,  round(sum(price \* (1 - discount) \* quantity), 2) AS revenue,  sum(quantity) AS quantity,  round(sum(price \* (1 - discount) \* quantity / total\_rev \* 100), 2) AS percent\_from\_total -- подсчитываем процент выручки от каждого товара  FROM sql.store\_carts sc JOIN sql.store\_delivery sd USING(order\_id) JOIN sql.store\_products sp USING(product\_id) CROSS JOIN total\_revenue GROUP BY 1,2,3, total\_rev ORDER BY 4 DESC LIMIT 25 |
| --- |

### 

###### График 3 - Категории топ-25 товаров по объему выручки.

Самыми прибыльными оказались копировальные машины, на втором месте - крупное оборудование для офисов, а на третьем - офисные кресла и стулья. В сумме на долю топ-25 товаров приходится 18,13% от общей выручки.

### **Итоги первой части анализа.**

* Динамика продаж у компании положительная. Более того, растет не только ежегодная выручка, но и количество заказов.
* Динамика продаж имеет **ярко выраженный сезонный характер**.
* Выявили самые прибыльные категории и подкатегории товаров:
  + **техника** является самой прибыльной **категорией**, где лидируют телефоны;
  + **кресла** являются самой прибыльной **подкатегорией** товаров;
  + среди офисных принадлежит наиболее прибыльными являются товары для хранения.
* Выявили топ-25 товаров по объему выручки (доля 18,13 от всей выручки).

## 

## **Часть 2. Составление портрета клиента**

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

| SELECT scust.category,  count(DISTINCT scust.cust\_id) AS cust\_cnt, -- нам нужно выбрать уникальных клиентов  round(sum(price \* (1 - discount) \* quantity)) AS revenue FROM sql.store\_carts sc JOIN sql.store\_delivery sd USING(order\_id) JOIN sql.store\_products sp USING(product\_id) JOIN sql.store\_customers scust USING(cust\_id) GROUP BY 1 ORDER BY 3 DESC |
| --- |

| **category** | **cust\_cnt** | **revenue** |
| --- | --- | --- |
| Corporate | 645 | 1 172 009 |
| Consumer | 148 | 274 148 |

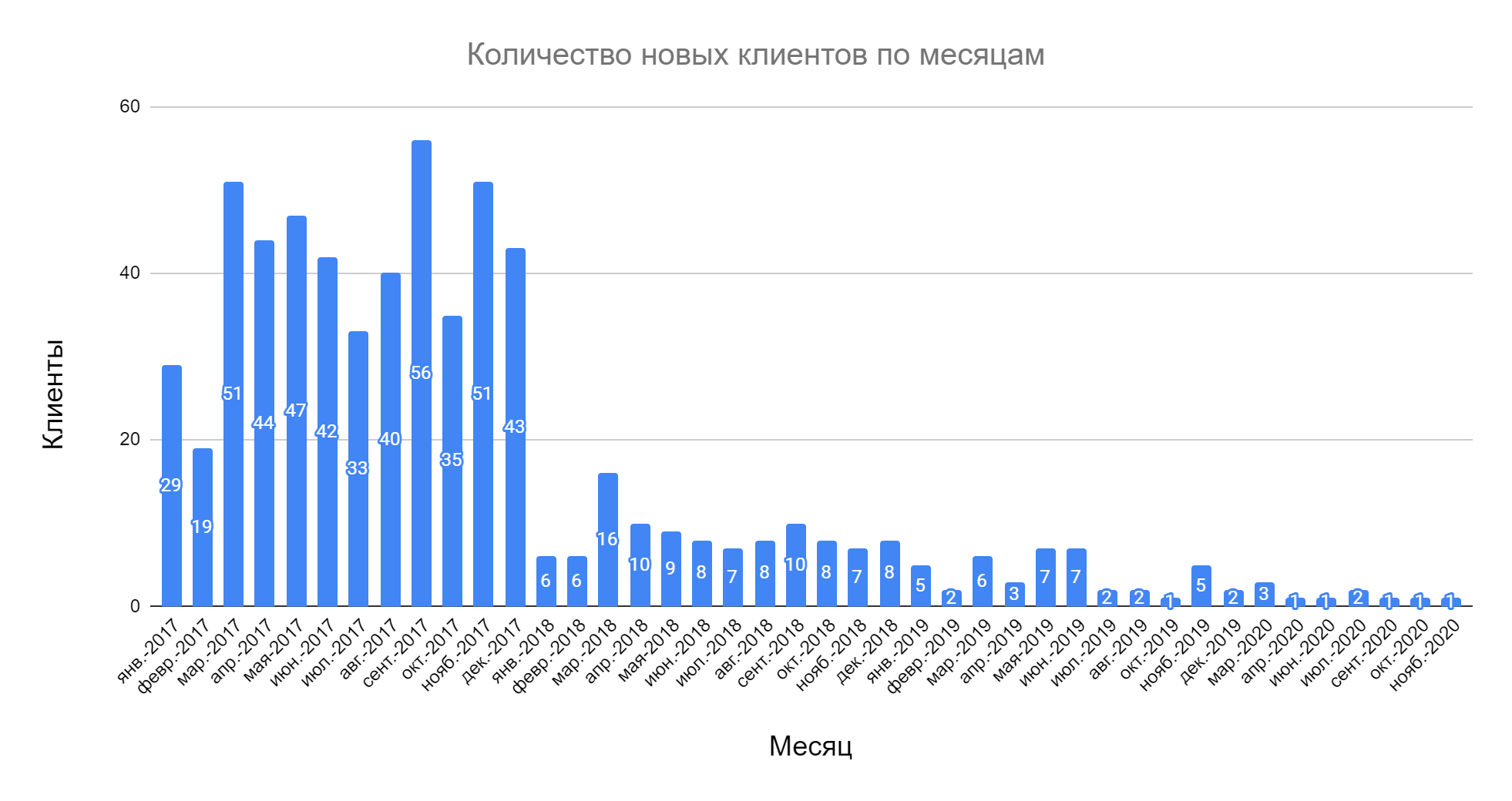
###### Таблица 1 - Выручка клиентов по категориям.

Исходя из Таблицы 1, мы видим, что **ключевыми для компании являются B2B-клиенты**. **Количество корпоративных клиентов** превышает количество физических лиц **в 4 раза**, а **выручка от них составляет 81%** (общий объем выручки составляет 1 446 157 $).

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 first\_order\_month AS -- CTE для определения месяца первого заказа корпоративных клиентов  (SELECT sd.cust\_id,  min(date\_trunc('month', order\_date))::date AS month  FROM sql.store\_delivery sd  LEFT JOIN sql.store\_customers scust USING(cust\_id)  WHERE category like 'Corporate'  GROUP BY 1) SELECT month::date,  count(DISTINCT sd.cust\_id) AS new\_custs FROM sql.store\_delivery sd JOIN first\_order\_month fom USING(cust\_id) GROUP BY 1 ORDER BY 1 |
| --- |



### *График 4 - Количество новых клиентов по месяцам.*

**Количество новых клиентов падает**. Как видно из графика, в самом начале было привлечено много клиентов, но **начиная с 2018 года, мер по привлечению новых клиентов не проводилось**. Принимая во внимание этот факт и то, что за весь рассматриваемый период выручка продолжала расти, можно сделать вывод, что у компании очень высокая лояльность клиентов. В данном случае у компании есть большой потенциал к росту, поэтому отделу маркетинга рекомендуется провести мероприятия по привлечению новых клиентов.

**Характеристика B2B-клиентов**

Теперь более подробно изучим показатели по корпоративным клиентам. Нам следует выяснить:

1. Сколько в среднем различных товаров в заказах у корпоративных клиентов?
2. Какая в среднем сумма заказов у корпоративных клиентов?
3. Сколько в среднем различных офисов у корпоративных клиентов?

* Для ответа на вышестоящие вопросы пишем следующий код:

| WITH corporate\_info AS -- CTE для вывода сумм заказов и количества уникальных товаров в нем для каждого корпоративного клиента  (SELECT cust\_id,  order\_id,  count(DISTINCT product\_id) AS number\_of\_products, -- количество уникальных товаров в каждом заказе  sum(price \* (1 - discount) \* quantity) AS order\_sum -- подсчет суммы заказа  FROM sql.store\_delivery sd  LEFT JOIN sql.store\_carts sc USING(order\_id)  JOIN sql.store\_products sp USING(product\_id)  JOIN sql.store\_customers scust USING(cust\_id)  WHERE scust.category like 'Corporate'  GROUP BY 1, 2  ORDER BY 1),  unique\_corporate\_zip\_codes AS -- CTE для вывода количества уникальных индексов для определения количества офисов для для каждого корпоративного клиента  (SELECT cust\_id,  count(DISTINCT zip\_code) zip  FROM sql.store\_delivery sd  JOIN sql.store\_customers scust USING(cust\_id)  WHERE scust.category like 'Corporate'  GROUP BY 1)  SELECT round(avg(number\_of\_products), 1) average\_products, -- среднее количество различных товаров у корпоративных клиентов  round(avg(order\_sum), 1) average\_order\_amount, -- средняя сумма заказов корпоративных клиентов  round(avg(zip), 1) average\_unique\_zipcodes -- среднее количество офисов у корпоративных клиентов FROM corporate\_info CROSS JOIN unique\_corporate\_zip\_codes |
| --- |

1. Среднее количество товаров в заказах корпоративных клиентов: 2.
2. Средняя сумма заказов корпоративных клиентов: 285,9.
3. Среднее количество офисов у корпоративных клиентов: 6,2.

### **Итоги второй части анализа.**

### Количество **B2B**-клиентов **в 4 раза больше**, чем физических.

* На долю **корпоративных клиентов** приходится **81% выручки**.
* Количество новых корпоративных клиентов сокращается.
* Доход растет за счет продаж старым лояльным клиентам.

Опираясь на уже проведенный анализ можно предложить некоторые рекомендации:

* Провести мероприятия по привлечению новых клиентов. Компания обладает огромным потенциалом к росту.
* Так как среди топ-25 товаров в основном присутствуют товары с долгим сроком эксплуатации (мебель, копировальные машины и т.д.) следует обратить внимание на возможность проведения акций, связанных с продажей сопутствующих товаров для данных категорий (сжатый воздух, бумага, масло для подъемных механизмов в креслах и т.д.).
* Выявить самые непопулярные товары компании и провести сравнительный анализ цен на эти товары или их аналоги в других магазинах.
* Сравнить товарный ассортимент с ассортиментом конкурентов. При необходимости пополнить его.

## 

## **Часть 3. Анализ логистики компании**

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

Поле ship\_mode таблицы store\_delivery содержит четыре различных значения.

| **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 |

###### Таблица 2 - Категории доставок

* Выясним, какая доля заказов доставляется в сроки для каждого типа доставки.

| WITH delivery\_time AS -- CTE для вывода заказов, доставленных с опозданием, и метода доставки  (SELECT order\_id,  ship\_mode,  order\_date,  ship\_date,  ship\_date - order\_date days\_delivery,  CASE  WHEN ship\_mode = 'Standard Class'  AND ship\_date - order\_date > 6 THEN 'late'  WHEN ship\_mode = 'Second Class'  AND ship\_date - order\_date > 4 THEN 'late'  WHEN ship\_mode = 'First Class'  AND ship\_date - order\_date > 3 THEN 'late'  WHEN ship\_mode = 'Same Day'  AND ship\_date - order\_date > 0 THEN 'late'  ELSE 'on\_time'  END AS is\_late -- пишем условие для определения сроков доставки  FROM sql.store\_delivery),  delivery\_status AS -- CTE для подсчета общего количества заказов, с опозданием и без  (SELECT DISTINCT  ship\_mode,  count(\*) OVER(PARTITION BY ship\_mode)::numeric orders\_cnt,  count(\*) filter (WHERE is\_late = 'late') OVER(PARTITION BY ship\_mode)::numeric late\_orders\_cnt,  count(\*) filter (WHERE is\_late = 'on\_time') OVER(PARTITION BY ship\_mode)::numeric successful\_orders\_cnt  FROM delivery\_time)  SELECT ship\_mode,  orders\_cnt,  late\_orders\_cnt,  round((successful\_orders\_cnt / orders\_cnt) \* 100, 2) AS "%success" FROM delivery\_status ORDER BY 4 |
| --- |

| **ship\_mode** | **orders\_cnt** | **late\_orders\_cnt** | **%success** |
| --- | --- | --- | --- |
| Second Class | 964 | 202 | 79,05 |
| Standard Class | 2994 | 309 | 89,68 |
| Same Day | 264 | 12 | 95,45 |
| First Class | 787 | 1 | 99,87 |

###### Таблица 3 - Эффективность доставки

Заказы, доставляемые **вторым классом**, чаще всего **приходят с опозданием**.

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

| SELECT date\_trunc('quarter', sd.order\_date)::date AS quarter,  count(sd.order\_id) AS orders FROM sql.store\_delivery sd WHERE sd.ship\_mode = 'Second Class'  AND ship\_date - order\_date > 4 GROUP BY 1 |
| --- |

###### ДиаграммаГрафик *5 - Динамика количества заказов, доставленных с опозданием, по кварталам.*

**Задержки** в доставке **имеют сезонный характер** по аналогии с выручкой и количеством заказов.

Можно предположить, что это происходит в связи с высокой загруженностью склада в связи с возросшим количеством заказов, либо задержка происходит на стороне службы доставки, так как наибольшее количество задержек случается в 3-ем и 4-ом кварталах, что соответствует началу учебного года (конец августа - начало сентября) и сезону праздников (День благодарения и Рождество) или концу финансового года, сопровождающимся увеличенным количеством отчетности и необходимости в большем количестве канцелярских товаров.

**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).

Находим самые популярные штаты по количеству доставок для визуализации:

| SELECT state,  count(order\_id) FROM sql.store\_delivery GROUP BY 1 |
| --- |

###### ДиаграммаГрафик *6 - Распределение количества доставок по штатам.*

За рассматриваемый период самым “популярным” штатом по количеству доставок был штат Калифорния. Также на этот штат приходится самое большое количество адресов доставок (если рассматривать именно индексы). При этом Техас - штат, в котором находится склад компании - находится на втором месте по количеству доставок.

| -- код для проверки количества адресов доставок по штатам --  SELECT state,  count(distinct zip\_code) FROM sql.store\_delivery GROUP BY 1 |
| --- |

Для того, чтобы облегчить работу склада и сосредоточиться на доставках по Техасу и на восточном и юго-восточном направлениях, **рекомендуется открыть магазин в Калифорнии**. Это поможет снять часть нагрузки на склад компании, так как большая часть заказов с западного побережья приходит именно из Калифорнии. Местные клиенты смогут быстрее получать свои заказы (при самовывозе) или доставке по штату.

Более того, в Калифорнии находятся города-миллионники Лос-Анджелес и Сан-Франциско, которые находятся на втором и третьем месте соответственно по количеству доставок (Таблица 4).

| -- код для проверки количества доставок по городам и штатам с -- SELECT DISTINCT  state,  city,  count(order\_id) OVER(PARTITION BY city, state) AS city\_deliveries\_cnt, -- разделяем по городам и штатам, так как названия городов встречаются в разных штатах  count(order\_id) OVER(PARTITION BY state) AS state\_deliveries\_cnt FROM sql.store\_delivery sd GROUP BY 1,2, sd.order\_id  ORDER BY 3 DESC  LIMIT 10 |
| --- |

| **state** | **city** | **city\_deliveries\_cnt** | **state\_deliveries\_cnt** |
| --- | --- | --- | --- |
| New York | New York City | 450 | 562 |
| California | Los Angeles | 384 | 1 021 |
| California | San Francisco | 265 | 1 021 |
| Pennsylvania | Philadelphia | 265 | 288 |
| Washington | Seattle | 212 | 256 |
| Texas | Houston | 188 | 487 |
| Illinois | Chicago | 171 | 276 |
| California | San Diego | 88 | 1 021 |
| Texas | Dallas | 80 | 487 |
| Ohio | Columbus | 55 | 236 |

###### Таблица 4 - Топ-10 городов по количеству доставок.

# 

# ВЫВОДЫ

Целью данного анализа была оценка эффективности работы компании и рекомендация места для открытия офлайн-магазина. Компания занимается продажами мебели, орг.техники, канцтоваров и различных принадлежностей для офиса.

Было проведено исследование динамики выручки и количества заказов компании ([см. График 1](#_qb0nexihc341)). **Продажи** компании показывают ярко выраженный **сезонный характер**, но при этом **выручка и количество заказов** в компании **растет год от года**, что говорит о хорошей операционной деятельности компании. Во второй половине года наблюдается рост выручки, который достигает своего пика в ноябре-декабре.

Мы выяснили, что ключевыми для компании являются корпоративные клиенты ([см. Таблица 1](#_tiymdvg5a1lp)). **На долю B2B-клиентов приходится 81% от всего объема продаж**. При этом, удалось найти важную точку роста: исходя из [Графика 4](#_h1bkoh4ev83w) видно, что начиная с 2018 года мероприятий по привлечению клиентов не проводилось. В связи с этим стоит обратить внимание на работу отдела маркетинга и провести мероприятия по привлечению новых клиентов.

Большую часть выручки компании приносят товары с долгим сроком службы ([см. График 2](#_foehphp54lx8)) и, в связи с этим, стоит уделить внимание акциям, направленным на продажу сопутствующих товаров для данных категорий. На данный момент среднее количество уникальных позиций в корзине у корпоративных клиентов равно двум, что крайне мало. Следует провести сравнительный анализ цен на непопулярные подкатегории товаров - это поможет улучшить ценовую политику компании, что приведет в увеличению среднего чека, который на данный момент составляет 285,9 долларов.

**Клиентская база** компании *Pens and Pencils* **отличается своей лояльностью** и поэтому рекомендуется сравнить ассортимент компании с ассортиментом конкурентов и пополнить его при необходимости. Таким образом, постоянные клиенты при составлении заказа смогут найти ранее недостающие товары среди ассортимента *Pens and Pencils* и с меньшей вероятностью обратятся к конкурентам.

На последнем этапе анализа была проведена оценка логистики компании. В результате исследования было выявлено, что товары доставляемые вторым классом чаще других доходят до клиента с опозданием ([см. Таблица 3](#_lsrbkhvcmgfr)). **Процент опозданий составляет 21%.** При этом, **задержки** в доставке **имеют сезонный характер** по аналогии с выручкой и количеством заказов ([см. График 5](#_3ygl5jaw2ul8)). Это свидетельствует сильной нагрузке на склад и службы доставки.

В целях снижения нагрузки на склады рекомендуется открыть **офлайн-магазин в штате Калифорния** в Лос-Анджелесе, так он находится на втором месте по количеству доставок среди городов, а территориальное расположение города также удобно по отношение к Сан-Франциско и Сан-Диего, занимающими 3 и 8 места соответственно по количеству доставляемых заказов ([см. Таблица 4](#_8pz3cqvg32fe))