

Цель

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

Задачи

- 1) Оценить динамику продаж и распределение выручки по товарам.
- 2) Составить портрет клиента (выяснить, какие клиенты приносят больше всего выручки).
- 3) Проанализировать логистику компании (определить, все ли заказы доставляются в срок и в каком штате лучше открыть офлайн-магазин).

Характеристика компании

Название: Pens and Pencils

Страна: США

Вид деятельности: B2B и B2C-продажи офисных товаров (оргтехника, канцтовары и мебель) онлайн.

Задача 1. Оценить динамику продаж и распределение выручки по товарам

Определяем эффективность продаж.

Считаем сумму выручки по месяцам

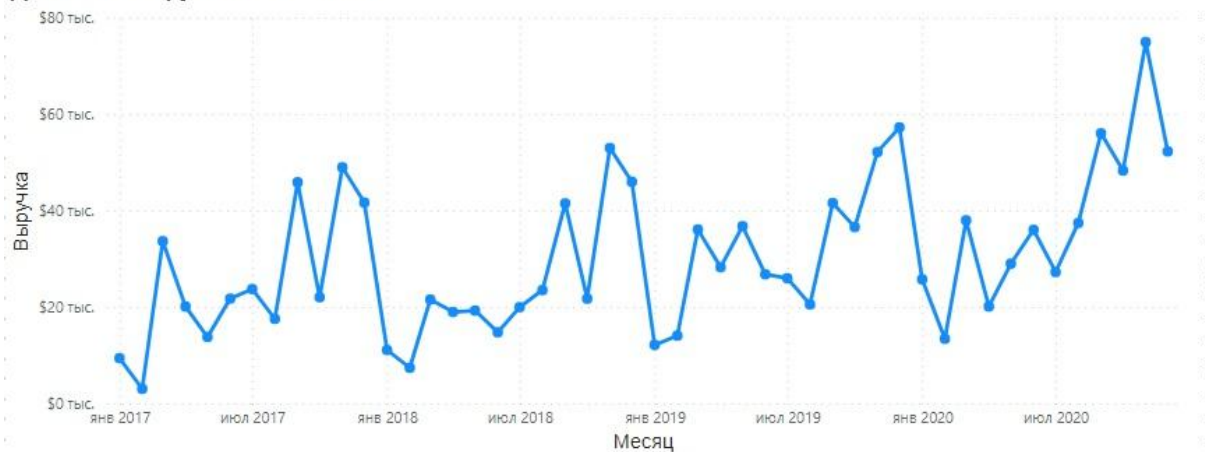
Для этого напишем запрос, который выведет столбцы:

- date (месяц заказа);
- revenue (объем выручки по месяцам).

```
-- сначала определим размер каждого заказа с помощью CTE
orders
WITH orders
    AS (SELECT order_id,
               SUM(quantity * price * ( 1 - discount ))
    amount
        FROM   SQL.store_carts sca
              join SQL.store_products sp
                ON sca.product_id = sp.product_id
        GROUP BY 1)
-- теперь можно посчитать выручку по месяцам
SELECT DATE_TRUNC('month', order_date) :: DATE date,
       ROUND(SUM(amount)) revenue
FROM   SQL.store_delivery sd
       join orders
        ON sd.order_id = orders.order_id
GROUP BY 1
ORDER BY date
```

Построим график на основе полученных данных.

Динамика выручки по месяцам



Как видно из графика, выручка компании растет. При этом в течение года характерна сильная сезонность. Так, минимальные значения выручки обычно наблюдаются в январе и феврале, а максимальные — в сентябре, ноябре и декабре (в октябре всегда спад).

Выявим самые популярные категории и подкатегории товаров

Для этого напишем запрос, который выведет столбцы:

- category (категория товара);
- subcategory (подкатегория товара);
- revenue (объем выручки по подкатегории).

```
SELECT category,  
       subcategory,  
       Round(Sum(quantity * price * ( 1 - discount ))) revenue  
FROM   sql.store_products sp  
       LEFT JOIN sql.store_carts sca  
           ON sp.product_id = sca.product_id  
GROUP BY 1,  
         2  
ORDER BY revenue DESC
```

Лидирует по продажам подкатегория “Кресла” из категории “Мебель” (\$235 318), на втором месте — подкатегория “Телефоны” из категории “Технологии” (\$221 110), на третьем — подкатегория “Системы хранения” из категории “Офисные принадлежности” (\$179 736).

Полностью результаты выглядят так:

category	subcategory	revenue
Furniture	Chairs	235 318 P
Technology	Phones	221 110 P
Office Supplies	Storage	179 736 P
Technology	Accessories	125 442 P
Furniture	Tables	114 532 P
Technology	Machines	113 978 P
Technology	Copiers	95 844 P
Office Supplies	Paper	62 148 P
Furniture	Bookcases	59 271 P
Office Supplies	Appliances	55 550 P
Furniture	Furnishings	51 695 P
Office Supplies	Binders	45 312 P
Office Supplies	Supplies	37 457 P
Office Supplies	Art	22 148 P
Office Supplies	Envelopes	13 809 P
Office Supplies	Labels	10 345 P
Office Supplies	Fasteners	2 459 P

Выявим, какие товары приносят самую большую выручку

Для более глубокого анализа важно понять не только, какие из товаров приносят больше всего выручки, но и определить, какую долю от общей выручки составляют их продажи.

Напишем запрос, который выведет топ-25 товаров по объему продаж. Столбцы:

- product_nm (наименование товара);
- revenue (выручка);
- quantity (количество проданных товаров);
- percent_from_total (доля от общей выручки, %).

--сначала рассчитаем выручку и количество проданных товаров по продуктам с помощью CTE products_revenue

WITH products_revenue AS

```
(
    SELECT      sp.product_id,
               product_nm,
               Sum(quantity) quantity,
               Sum(quantity * price * (1 - discount))
revenue
    FROM        sql.store_products sp
    LEFT JOIN   sql.store_carts sca
    ON          sp.product_id = sca.product_id
    GROUP BY    1,
               2
    ORDER BY    revenue DESC ),
```

--с помощью CTE `sum_revenue` посчитаем общую выручку, чтобы затем вычислить процент от общих продаж. Столбец `product_nm` необходим, чтобы потом можно было соединить результаты вычислений двух CTE

```
sum_revenue AS
(
    SELECT product_nm,
           Sum(revenue) OVER() total_revenue
    FROM   products_revenue )
SELECT   pr.product_nm,
         Round(revenue,2) revenue,
         quantity,
         Round(revenue * 100 / total_revenue,2)
percent_from_total
FROM     products_revenue pr
JOIN     sum_revenue sr
ON       pr.product_nm = sr.product_nm
ORDER BY revenue DESC limit 25
```

Результат удобнее сразу представить в виде таблицы:

product_nm	revenue	quantity	percent_from_total
Canon imageCLASS 2200 Advanced Copier	\$36 959,82	20	2,56%
HON 5400 Series Task Chairs for Big and Tall	\$15 309,53	39	1,06%
3D Systems Cube Printer, 2nd Generation, Magenta	\$14 299,89	11	0,99%
High Speed Automatic Electric Letter Opener	\$13 624,21	11	0,94%
Martin Yale Chadless Opener Electric Letter Opener	\$13 325	22	0,92%
Cisco TelePresence System EX90 Videoconferencing Unit	\$11 319,24	6	0,78%
Hewlett Packard LaserJet 3310 Copier	\$11 303,69	38	0,78%
Samsung Galaxy Mega 6.3	\$11 154,87	39	0,77%
Canon PC1060 Personal Laser Copier	\$9 295,83	19	0,64%
HP Designjet T520 Inkjet Large Format Printer - 24" Color	\$9 187,5	12	0,64%
Global Troy Executive Leather Low-Back Tilter	\$9 082,87	32	0,63%
Honeywell Enviracaire Portable HEPA Air Cleaner for 17' x 22' Room	\$9 043,55	40	0,63%
Tennsco 6- and 18-Compartment Lockers	\$8 740,17	43	0,60%
Plantronics CS510 - Over-the-Head monaural Wireless Headset System	\$8 657,89	35	0,60%
SAFCO Arco Folding Chair	\$8 100,95	53	0,56%
Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	\$7 805,48	24	0,54%
Apple iPhone 5	\$7 798	24	0,54%
Tennsco Double-Tier Lockers	\$7 704,86	47	0,53%
Ativa V4110MDD Micro-Cut Shredder	\$7 699,89	11	0,53%
Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	\$7 446,44	51	0,51%
Tennsco Single-Tier Lockers	\$7 326,59	28	0,51%
Office Star - Professional Matrix Back Chair with 2-to-1 Synchro Tilt and Mesh Fabric Seat	\$7 019,5	28	0,49%
Zebra ZM400 Thermal Label Printer	\$6 965,7	6	0,48%
Hot File 7-Pocket, Floor Stand	\$6 510,77	49	0,45%
Bretford Rectangular Conference Table Tops	\$6 497,82	46	0,45%

Задача 2. Составить портрет клиента

Узнаем лучше клиентов компании.

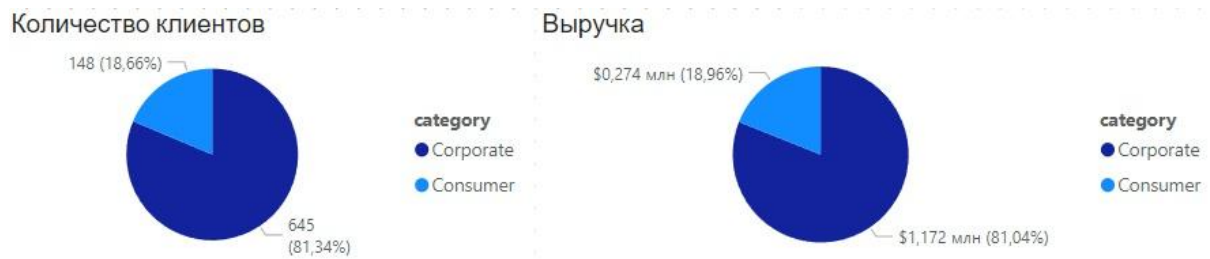
Рассчитаем выручку клиентов по категориям

Для этого напишем запрос, который выведет столбцы:

- category (категория клиента, Corporate или Consumer);
- cust_cnt (количество клиентов);
- revenue (сумма выручки по категории).

```
-- снова воспользуемся CTE orders, чтобы получить размер
каждого заказа
WITH orders
  AS (SELECT order_id,
             Sum(quantity * price * ( 1 - discount ))
amount
      FROM   sql.store_carts sca
      JOIN   sql.store_products sp
            ON sca.product_id = sp.product_id
      GROUP BY 1),
  --теперь рассчитаем выручку с каждого клиента с помощью
  CTE customers_revenue
  customers_revenue
  AS (SELECT cust_id,
             Sum(amount) cust_amount
      FROM   sql.store_delivery sd
      JOIN   orders
            ON sd.order_id = orders.order_id
      GROUP BY 1)
SELECT category,
       Count(DISTINCT scu.cust_id) cust_cnt,
       Round(Sum(cust_amount))      revenue
FROM   sql.store_customers scu
LEFT JOIN customers_revenue cr
      ON scu.cust_id = cr.cust_id
GROUP BY 1
ORDER BY revenue DESC
```

Как выяснилось, B2B-клиентов у компании 645, а B2C-клиентов — 148. Первые принесли ей \$1,172 млн, а вторые — \$0,274 млн.



Корпоративных клиентов у компании намного больше, и они приносят в разы больше денег, поэтому изучим их подробнее.

Составим характеристику B2B-клиентов

Для начала изучим динамику появления новых B2B-клиентов по месяцам. Для этого напишем запрос, который выводит столбцы:

- date (месяц первого заказа);
- new_custs (количество новых клиентов).

```
--определим дату первого заказа для каждого клиента с помощью
CTE first_order_date
WITH first_order_date
    AS (SELECT DISTINCT cust_id,
                        MIN(order_date)
                        over(
                            PARTITION BY cust_id) first_order
        FROM    SQL.store_delivery)
SELECT DATE_TRUNC('month', first_order) :: DATE date,
       COUNT(fod.cust_id) new_custs
FROM    first_order_date fod
       join SQL.store_customers scu
         ON fod.cust_id= scu.cust_id
WHERE   category = 'Corporate'
GROUP BY 1
ORDER BY date
```

Отообразим результаты на графике.



Как видно из графика, основная часть корпоративных клиентов была привлечена еще в 2017 г. Затем количество новых клиентов постепенно снижалось, а в некоторые месяцы 2019-2020 гг. их вовсе не было.

Теперь изучим основные показатели у B2B-клиентов.

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

Для ответа на этот вопрос напомним запрос, который выведет нужное значение:

--с помощью CTE `dist_products_in_order` узнаем, сколько разных товаров в заказах B2B-клиентов

```
WITH dist_products_in_order
  AS (SELECT sd.order_id,
            sd.cust_id,
            Count(DISTINCT product_id) dist_products
  FROM   sql.store_delivery sd
  JOIN   sql.store_customers scu
        ON sd.cust_id = scu.cust_id
  JOIN   sql.store_carts sca
        ON sd.order_id = sca.order_id
  WHERE  category = 'Corporate'
  GROUP BY 1,
           2)

SELECT Round(Avg(dist_products), 1) avg_dist_products
FROM   dist_products_in_order
```

2. Какая в среднем сумма заказов у корпоративных клиентов?

Нужное значение выведет запрос:


```

-- снова воспользуемся CTE orders, чтобы получить размер
каждого заказа
WITH orders
  AS (SELECT order_id,
             Sum(quantity * price * ( 1 - discount ))
amount
      FROM   sql.store_carts sca
            JOIN sql.store_products sp
              ON sca.product_id = sp.product_id
      GROUP BY 1),
--выясним размер заказов B2B-клиентов
amount_per_order
  AS (SELECT sd.order_id,
             sd.cust_id,
             amount
      FROM   sql.store_delivery sd
            JOIN orders
              ON sd.order_id = orders.order_id
            JOIN sql.store_customers scu
              ON sd.cust_id = scu.cust_id
      WHERE  category = 'Corporate')
SELECT Round(Avg(amount), 1) avg_amount
FROM   amount_per_order

```

3. Сколько в среднем различных офисов у корпоративных клиентов?

Нужное значение выведет запрос:

```

--с помощью CTE b2b_offices узнаем, сколько разных офисов у
каждого B2B-клиента
WITH b2b_offices
  AS (SELECT scu.cust_id,
             Count(DISTINCT zip_code) cnt_offices
      FROM   sql.store_customers scu
            JOIN sql.store_delivery sd
              ON scu.cust_id = sd.cust_id
      WHERE  category = 'Corporate'
      GROUP BY 1)
SELECT Round(Avg(cnt_offices), 1) avg_offices
FROM   b2b_offices

```

Получился следующий портрет B2B-клиентов. В среднем за один раз они заказывают 2,0 различных товаров, средний чек равен \$285,9. В среднем у них 6,2 различных офиса.

Задача 3. Проанализировать логистику компании

Оценим эффективность логистики и определим штат, где лучше открыть новый офлайн-магазин.

Определяем эффективность доставки

В этом поможет запрос, который выводит столбцы:

- ship_mode (тип доставки);
- orders_cnt (количество заказов);
- late_orders_cnt (количество заказов, доставленных с опозданием);
- % success (доля доставленных вовремя заказов, %).

--с помощью CTE delivery_control, узнаем какие заказы были доставлены вовремя, а какие - нет

```
WITH delivery_control
  AS (SELECT order_id,
             ship_mode,
             CASE
               WHEN ship_mode = 'Standard Class'
                 AND ship_date - order_date <= 6 THEN
                 'yes'
               WHEN ship_mode = 'Second Class'
                 AND ship_date - order_date <= 4 THEN
                 'yes'
               WHEN ship_mode = 'First Class'
                 AND ship_date - order_date <= 3 THEN
                 'yes'
               WHEN ship_mode = 'Same Day'
                 AND ship_date - order_date = 0 THEN
                 'yes'
               ELSE 'no'
             END in_time_delivery
  FROM SQL.store_delivery)
```

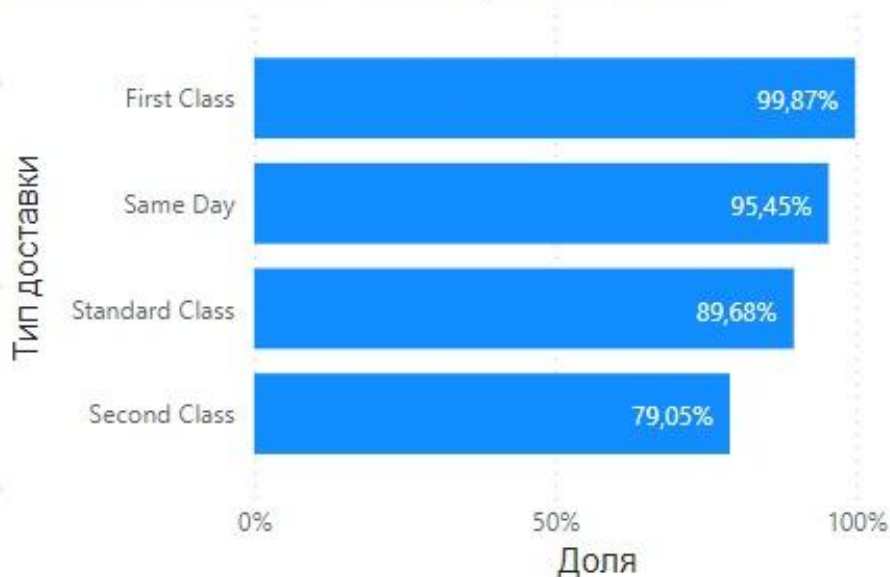
```

SELECT ship_mode,
       Count(order_id)
       orders_cnt,
       Count(CASE
            WHEN in_time_delivery = 'no' THEN order_id
            END)
       late_orders_cnt,
       Round(Count(CASE
            WHEN in_time_delivery = 'yes' THEN
order_id
            END) :: NUMERIC * 100 / Count(order_id) ::
NUMERIC, 2)
       "% success"
FROM   delivery_control
GROUP BY 1
ORDER BY "% success"

```

Реже всего — только в 79,05% случаев — вовремя приходят заказы, отправленные вторым классом. Также много опозданий при отправке заказов стандартным классом (вовремя приходят 89,68%). При отправке в тот же день (95,45% вовремя) или первым классом (99,87% вовремя) опоздания случаются гораздо реже.

Доля доставленных вовремя заказов



Подробнее взглянем на проблему частых опозданий заказов, отправленных вторым классом. Они доставлялись с опозданием постоянно или только в некоторых кварталах?

Для этого напомним запрос, который выводит столбцы:

- date (квартал);
- late_rate (доля доставленных с опозданием заказов, %).

```
--снова воспользуемся CTE delivery_control, узнаем какие
заказы были доставлены вовремя, а какие - нет
WITH delivery_control
    AS (SELECT order_id,
              ship_mode,
              CASE
                WHEN ship_mode = 'Standard Class'
                  AND ship_date - order_date <= 6 THEN
                    'yes'
                WHEN ship_mode = 'Second Class'
                  AND ship_date - order_date <= 4 THEN
                    'yes'
                WHEN ship_mode = 'First Class'
                  AND ship_date - order_date <= 3 THEN
                    'yes'
                WHEN ship_mode = 'Same Day'
                  AND ship_date - order_date = 0 THEN
                    'yes'
                ELSE 'no'
              END in_time_delivery
    FROM   SQL.store_delivery)
SELECT Date_trunc('quarter', order_date) :: DATE date,
       Round(Count(CASE WHEN dc.ship_mode = 'Second Class'
                        AND in_time_delivery = 'no' THEN
sd.order_id END) :: NUMERIC * 100 /
           Count(CASE WHEN dc.ship_mode = 'Second Class'
                        THEN sd.order_id END) :: NUMERIC, 2)
late_rate
FROM   SQL.store_delivery sd
       join delivery_control dc
       ON sd.order_id = dc.order_id
GROUP BY 1
ORDER BY date
```

Результат отображен на графике.



Несмотря на то что в некоторые кварталы ситуация была менее критичной, очевидно, что такие заказы доставляются с опозданием постоянно. Эту проблему необходимо решать.

Определяем наиболее удачное место для нового офлайн-магазина

Для этого найдем город и штат, куда совершается больше всего доставок.

Для штатов запрос выводит столбцы:

- state (штат);
- orders (количество заказов).

```
SELECT state,  
        Count(order_id) orders  
FROM    sql.store_delivery  
GROUP BY 1  
ORDER BY orders DESC
```

Для городов запрос выводит столбцы:

- city (город);
- orders (количество заказов).

```
SELECT city,  
        Count(order_id) orders  
FROM    sql.store_delivery  
GROUP BY 1  
ORDER BY orders DESC
```

Результат неоднозначный. Среди штатов на первом месте оказалась Калифорния, а среди городов — Нью-Йорк.

Поэтому взглянем на карту. Рядом с Нью-Йорком находится гораздо больше штатов, где также много заказов. В западной части страны, кроме Калифорнии, много клиентов только в штатах Вашингтон, Аризона и Колорадо. Но последние два находятся недалеко от Техаса, где и так есть склад. Учитывая все эти факторы, целесообразнее открыть новый магазин в штате Нью-Йорк.



Выводы

Выручка компании растет. При этом в течение года характерна сильная сезонность. Минимальные значения выручки обычно наблюдаются в январе и феврале, а максимальные — в сентябре, ноябре и декабре.

Лидирует по продажам подкатегория “Кресла” из категории “Мебель” (\$235 318), на втором месте — подкатегория “Телефоны” из категории “Технологии” (\$221 110), на третьем — подкатегория “Системы хранения” из категории “Офисные принадлежности” (\$179 736).

B2B-клиентов у компании 645, а B2C-клиентов — 148. Первые принесли ей \$1,172 млн, а вторые — \$0,274 млн.

Портрет B2B-клиентов:

- в среднем за один раз они заказывают 2,0 различных товаров;
- средний чек равен \$285,9;
- в среднем у них 6,2 различных офиса.

Основная часть корпоративных клиентов была привлечена еще в 2017 г. Затем количество новых клиентов постепенно снижалось, а в некоторые месяцы 2019-2020 гг. их вовсе не было.

Реже всего — только в 79,05% случаев — вовремя приходят заказы, отправленные вторым классом. Также много опозданий при отправке заказов стандартным классом (вовремя приходят 89,68%). При отправке в тот же день (95,45% вовремя) или первым классом (99,87% вовремя) опоздания случаются гораздо реже.

В некоторые кварталы ситуация с опозданием заказов, отправленных вторым классом, была менее критичной. Однако такие заказы доставляются с опозданием постоянно.

Среди штатов на первом месте по количеству заказов находится Калифорния, а среди городов — Нью-Йорк. Но рядом с Нью-Йорком находится гораздо больше штатов, где также много заказов. В западной части страны, кроме Калифорнии, много клиентов только в штатах Вашингтон, Аризона и Колорадо. Но последние два находятся недалеко от Техаса, где и так есть склад. Учитывая все эти факторы, целесообразнее открыть новый магазин в штате Нью-Йорк.

Рекомендации

1. Компания должна быть готова к сезонному всплеску спроса в конце года. В частности, в сентябре, ноябре и декабре.
2. Необходимо активнее привлекать новых B2B-клиентов, чтобы ускорить рост выручки.
3. Необходимо решить систематическую проблему с опозданием заказов, отправленных вторым классом.
4. Новый офлайн-магазин лучше открыть в штате Нью-Йорк.