



## АНАЛИЗ РАБОТЫ КОМПАНИИ PENS AND PENCILS

## ЦЕЛЬ

В рамках тестового задания просят 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).

## Шаг 1: Определяем эффективность продаж

Для этого ответим на следующие вопросы:

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

Напишем запрос, который выведет сумму выручки по месяцам:

- date (месяц заказа) — тип date;
  - revenue (объём выручки)
- Отсортируем запрос по дате заказа.

```
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
```



Как видно из графика: выручка растет. Также наблюдается сезонность роста и падения выручки. Отмечается падение выручки зимой и летом, наилучшие показатели осенью. Возможно из-за начала учебного года.

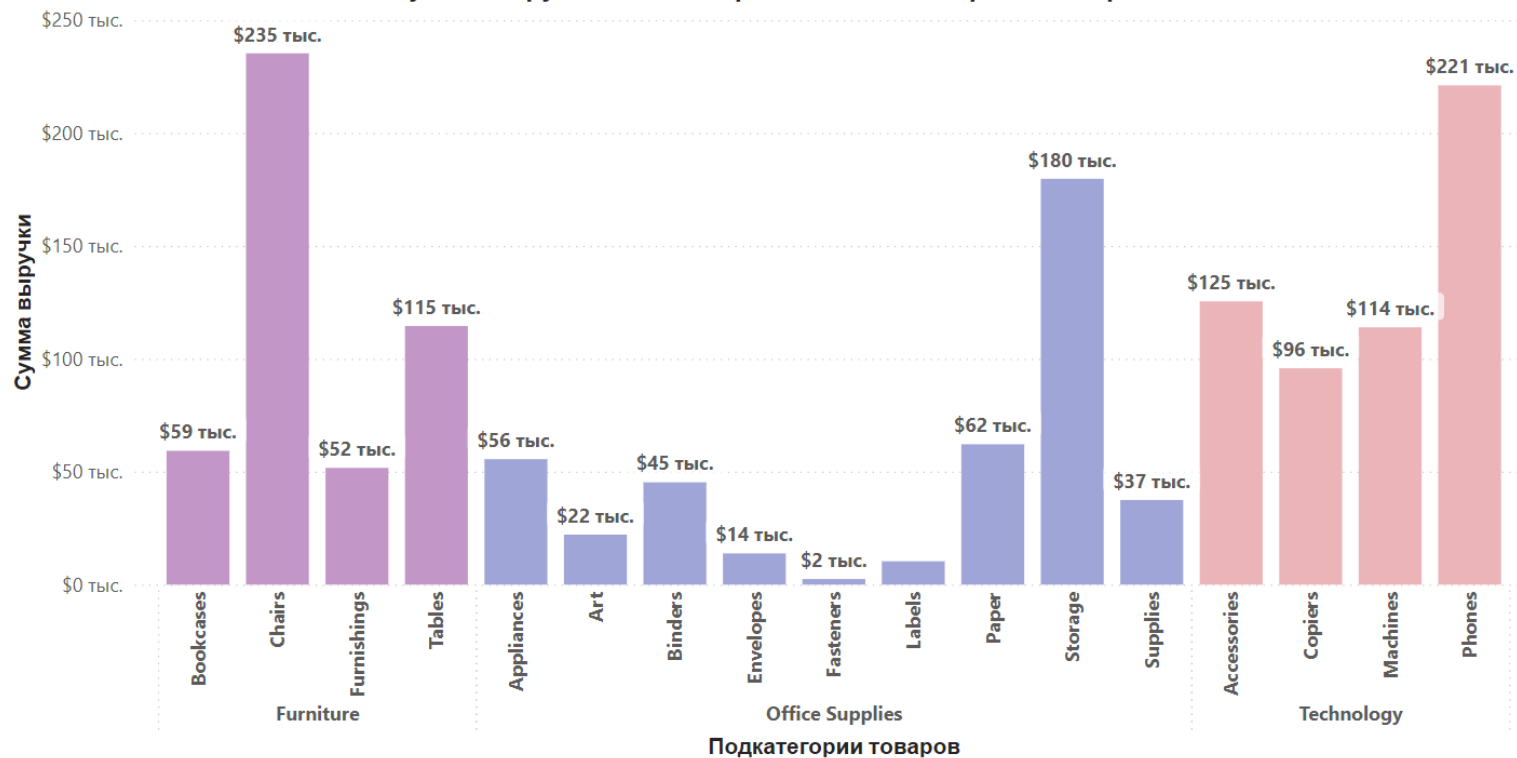
Напишем запрос, который выведет сумму выручки по различным категориям и подкатегориям:

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

Отсортируем запрос по убыванию выручки.

```
SELECT category,
          subcategory,
          round(sum(price * quantity * (1 - discount))) revenue
FROM sql.store_carts sc
JOIN sql.store_products sp ON sc.product_id = sp.product_id
GROUP BY 1,2
ORDER BY 3 DESC
```

Сумма выручки по категориям и подкатегориям товаров



Исходя из графика видно, что в США очень любят хорошие кресла — они в топе по продажам!

Напишем 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** t\_rev **AS**

(**SELECT** **sum**(price \* quantity \* (1 - discount)) total

**FROM** sql.store\_carts sc

**JOIN** sql.store\_products sp **ON** sc.product\_id = sp.product\_id) -- Подзапрос, который выводит объем общей выручки

**SELECT** product\_nm,

round(**sum**(price \* quantity \* (1 - discount)), 2) revenue,

**sum**(quantity) quantity,

round(**sum**(price \* quantity \* (1 - discount)) / **avg**(t\_rev.total) \* 100, 2)

percent\_from\_total

**FROM** sql.store\_carts sc

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

**JOIN** t\_rev **ON** TRUE

**GROUP BY** 1

**ORDER BY** 2 **DESC**

**LIMIT** 25

## Топ 25-товаров по продажам

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	13 325,00	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	9 187,50	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	7 798,00	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	7 019,50	28	0,49
Zebra ZM400 Thermal Label Printer	6 965,70	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

Как мы видим из таблицы лидерами продаж являются: копир Canon, кресло для руководителей серии HON 5400 и 3D принтер Magenta.

## Шаг 2: Составляем 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.

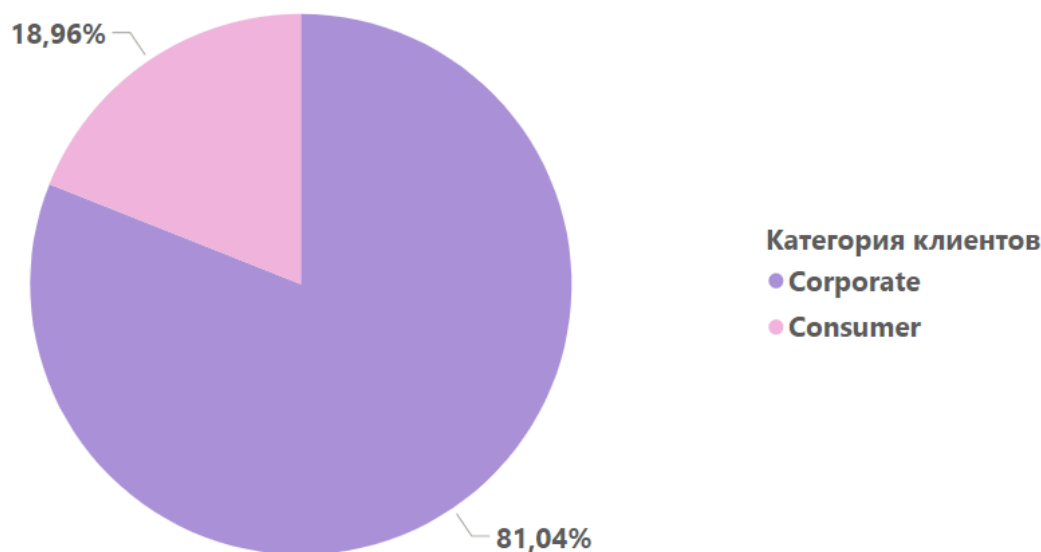
Напишем запрос, который выведет количество клиентов и выручку по категориям клиента:

- category (категория клиента);
- cust\_cnt (количество клиентов);
- revenue (объём выручки).

Отсортируем запрос по убыванию выручки.

```
SELECT sc2.category,  
       count(DISTINCT sc2.cust_id) cust_cnt,  
       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  
JOIN sql.store_customers sc2 ON sc2.cust_id = sd.cust_id  
GROUP BY 1  
ORDER BY 3 DESC
```

Соотношение выручки по категориям клиентов



Как мы видим: B2B-клиентов намного больше, и выручки они приносят тоже в разы больше!

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

- month (месяц) — тип date;
- new\_custs (количество новых клиентов).

Отсортируем запрос по первому столбцу в порядке возрастания.

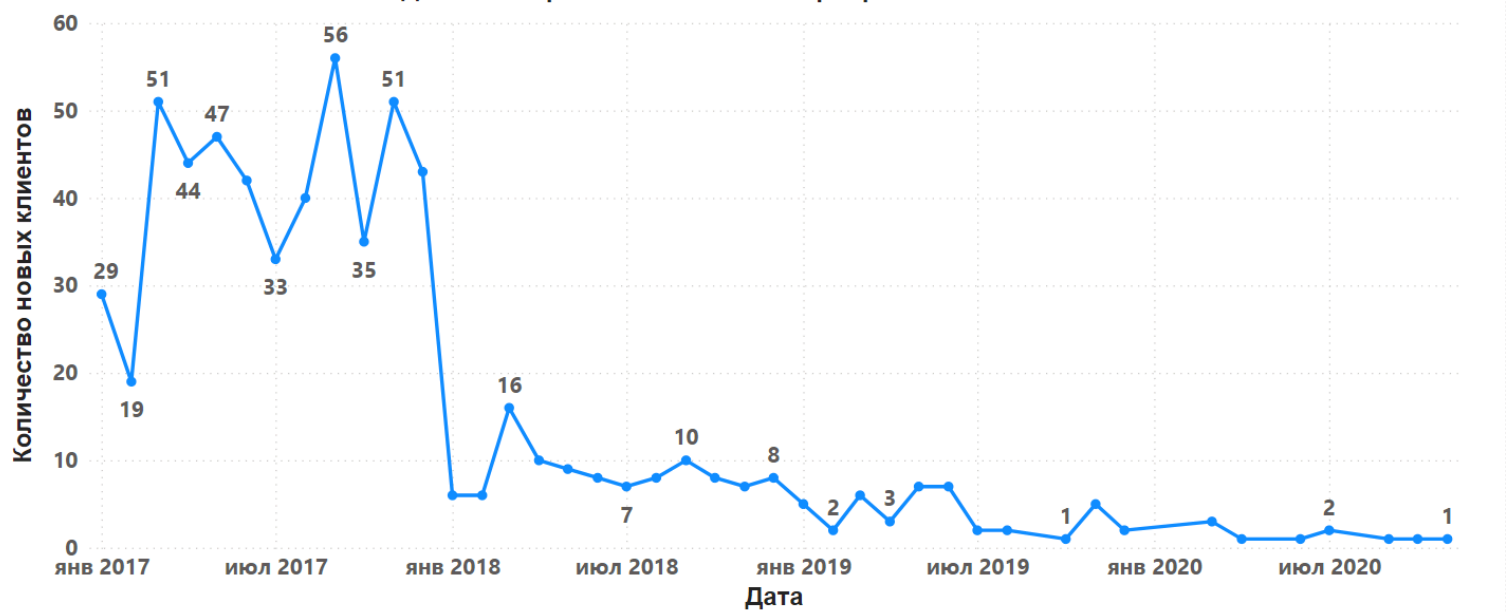
```
WITH first_order AS
(SELECT sd.cust_id,
  min(date_trunc('month', order_date))::date first_date
FROM sql.store_delivery sd
JOIN sql.store_customers sc ON sd.cust_id = sc.cust_id
WHERE category = 'Corporate'
GROUP BY 1)--Подзапрос, который выводит дату первого заказа
```

```
SELECT first_date,
  count(cust_id)
FROM first_order
GROUP BY 1
ORDER BY 1
```





Динамика привлечения новых корпоративных клиентов



Как видно из графика, в самом начале было привлечено много корпоративных клиентов, а после 2018 года новые клиенты почти не привлекались. Таким образом выручка растет за счет старых корпоративных клиентов. Следует обратить на это внимание, и выяснить почему в компанию перестали обращаться новые клиенты.

Найдем ответы на следующие вопросы:

- 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?
- Поскольку данных о полном адресе у нас нет, разные адреса будем считать по различным индексам (zip\_code).

**WITH** orders **AS**

```
(SELECT d.cust_id,  
      d.order_id,  
      count(DISTINCT c.product_id) product_qty,  
      sum(c.quantity * p.price * (1-c.discount)) order_amt
```

**FROM** sql.store\_carts **c**

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

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

**JOIN** sql.store\_customers sc **ON** d.cust\_id=sc.cust\_id

**WHERE** sc.category = 'Corporate'

**GROUP BY** 1,

2),-- Подзапрос, который выводит сколько товаров в заказах и сумму заказов у

корпоративных клиентов

cust\_zip **AS**

```
(SELECT sc.cust_id,  
      count(DISTINCT d.zip_code) address_num
```

**FROM** sql.store\_customers sc

**JOIN** sql.store\_delivery d **ON** d.cust\_id=sc.cust\_id

**WHERE** sc.category = 'Corporate'

**GROUP BY** 1)-- Подзапрос, который выводит сколько различных почтовых индексов

у корпоративных клиентов

```
SELECT round(avg(o.product_qty), 1) avg_product_qty,  
      round(avg(o.order_amt), 1) avg_order_amount,  
      round(avg(z.address_num), 1) avg_address_number
```

**FROM** orders o

**CROSS JOIN** cust\_zip z --CROSS JOIN использовала, так как все со всем нужно соединить, это же фактически справочник

Таким образом корпоративные клиенты, которые представляют наибольшую ценность для компании, выглядят следующим образом:

- средняя стоимость заказа  $\geq 285,9$
- количество уникальных товаров в заказе  $\geq 2$
- количество офисов  $\geq 6,2$

## 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.)

Напишем 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).

Отсортируем запрос по четвёртому столбцу в порядке возрастания.

```
WITH schedule AS
  (SELECT 'Standard Class' ship_mode,
          6 days_plan
   UNION ALL SELECT 'Second Class',
                    4
   UNION ALL SELECT 'First Class',
                    3
   UNION ALL SELECT 'Same Day',
                    0), --Подзапрос, который выводит классы доставки
total_orders AS
  (SELECT ship_mode,
          count(*) orders_cnt
   FROM sql.store_delivery sd
   GROUP BY 1), --Подзапрос, который выводит общее количество заказов
late_orders AS
  (SELECT sd.ship_mode,
          count(*) late_orders_cnt
   FROM sql.store_delivery sd
   JOIN schedule s ON sd.ship_mode = s.ship_mode
   AND sd.ship_date - sd.order_date > s.days_plan
   GROUP BY 1) --Подзапрос, который выводит количество заказов, которые пришли с
опозданием
SELECT t.ship_mode,
       t.orders_cnt,
       l.late_orders_cnt,
       round((t.orders_cnt - l.late_orders_cnt) * 100.0 / t.orders_cnt, 2) "% success"
FROM total_orders t
LEFT JOIN late_orders l ON t.ship_mode = l.ship_mode
ORDER BY 4
```

ship_mode	orders_cnt	late_orders_cnt	% success
Second Class	964	202	<b>79.05</b>
Standard Class	2994	309	89.68
Same Day	264	12	95.45
First Class	787	1	99.87

Как видим, чаще всего с опозданием доходят заказы, отправленные вторым классом. Давайте посмотрим, насколько систематично это происходит. Возможно, у службы доставки были проблемы только на протяжении ограниченного периода.

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

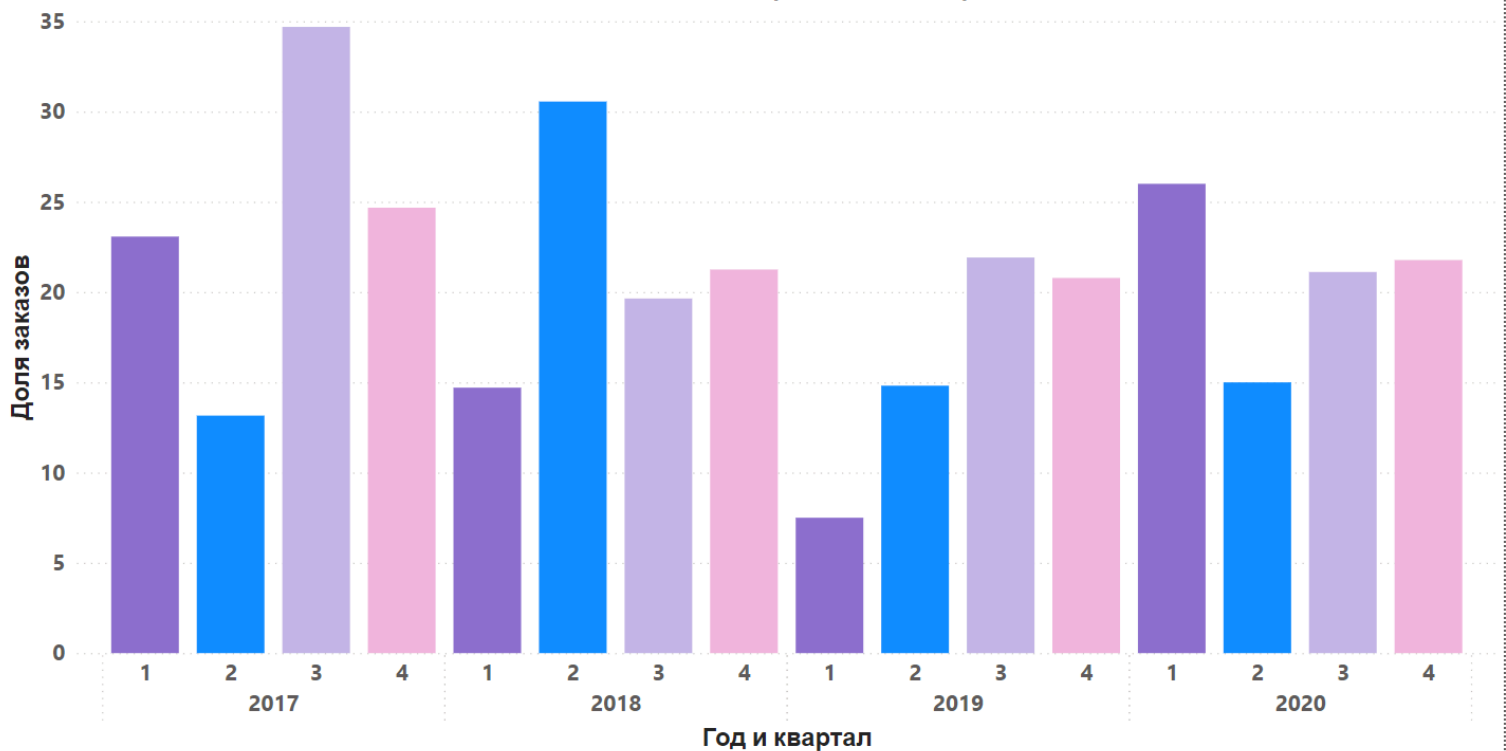
```

WITH Not_in_time_t AS
  (SELECT DISTINCT to_char(order_date, 'Q') QUARTER,
    to_char(order_date, 'YYYY') "year",
    COUNT(order_id)::float orders_cnt_nit
  FROM sql.store_delivery
  WHERE ship_mode='Second Class'
  AND (ship_date-order_date)>4
  GROUP BY 1,
    2
  ORDER BY 2,
    1),--Подзапрос, который выводит количество заказов, отправленных вторым
классом с опозданием
In_time_t AS
  (SELECT DISTINCT to_char(order_date, 'Q') QUARTER,
    to_char(order_date, 'YYYY') "year",
    COUNT(order_id)::float orders_cnt
  FROM sql.store_delivery sd
  WHERE ship_mode='Second Class'
  GROUP BY 1,
    2
  ORDER BY 2,
    1)--Подзапрос, который выводит количество заказов, отправленных вторым
классом вовремя

SELECT it.Quarter,
  it.year,
  ((nit.orders_cnt_nit) / (it.orders_cnt)) * 100 fraction
FROM In_time_t it
LEFT JOIN Not_in_time_t nit ON it.Quarter = nit.Quarter
AND it.year = nit.year
ORDER BY 2,
  1

```

Доля заказов с опозданием, отправленных вторым классом



Какой вывод можно сделать: наименьшее количество опозданий было в 2019 году. Четвертый квартал "заколдованный" - во все года доля заказов, доставленных с опозданием примерно одинаковая и достаточно высокая - около 20%. А вот первый, второй и третий кварталы отличаются. Наихудшим был третий квартал 2017 года - 34,69% опозданий. Необходимо проанализировать более детально и подробно работу службы доставки в эти периоды, чтобы сделать более конкретные выводы, чтобы можно было скорректировать действующие алгоритмы и протоколы работы службы доставки, и таким образом уменьшить количество доставок с опозданием.

Напишем запрос, который выведет количество доставок по штатам и визуализируем результат с помощью карты в Metabase.

Какой штат является наиболее перспективным для открытия офлайн-магазина? Почему?

```
WITH delivery_with_status AS
```

```
(SELECT *,
```

```
  CASE
```

```
    WHEN ship_mode ~* 'same'
```

```
      AND ship_date > order_date
```

```
      OR ship_mode ~* 'first'
```

```
      AND ship_date > order_date + 3
```

```
      OR ship_mode ~* 'second'
```

```
      AND ship_date > order_date + 4
```

```
      OR ship_mode ~* 'stand'
```

```
      AND ship_date > order_date + 6 THEN 'late'
```

```
    ELSE 'on time'
```

```
  END AS status
```

```
FROM sql.store_delivery)--Подзапрос, который выводит все характеристики  
доставок: дата заказа, дата доставки, класс доставки, город, почтовый код, id клиента,  
статус доставки вовремя-невовремя
```

```
SELECT state,
```

```
  count(DISTINCT dlv.order_id) AS cnt,
```

```
  100.0*count(DISTINCT dlv.order_id) filter(
```

```
    WHERE status='late')/count(DISTINCT dlv.order_id) AS lates,
```

```
  sum(price*quantity*(1-discount)) AS rev
```

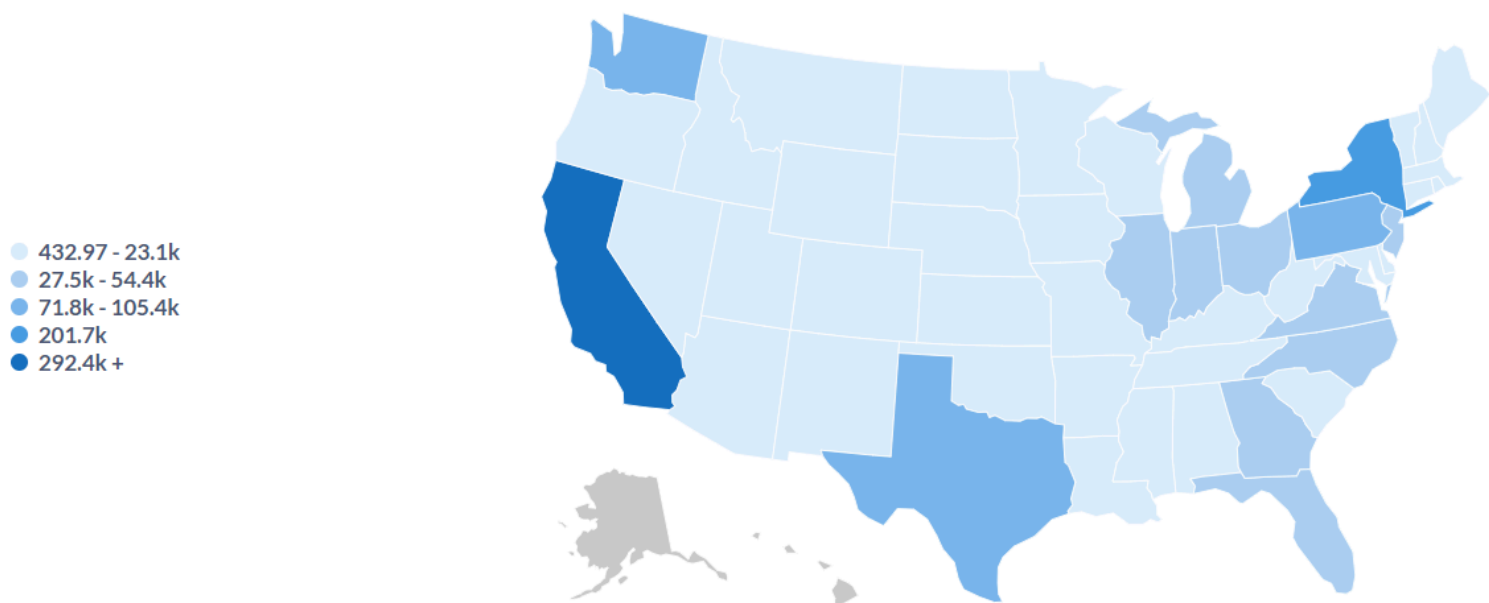
```
FROM delivery_with_status AS dlv
```

```
JOIN sql.store_carts AS crt ON dlv.order_id = crt.order_id
```

```
JOIN sql.store_products AS prd ON prd.product_id = crt.product_id
```

```
GROUP BY state
```

```
ORDER BY cnt DESC
```



*Доля выручки по штатам*



*Количество доставок по штатам*

Лидером по количеству доставок и выручки является штат Калифорния, второе место занимает Нью-Йорк, а третье - Техас.

Логичным становится вывод об открытии офлайн магазина в одном из штатов лидеров. Я бы дополнительно уточнила некоторые данные, может в одном из этих штатов есть уже склад компании например и там будет проще, быстрее и дешевле открыть точку офлайн продаж. А если склад находится в Калифорнии, то выбор очевиден.

## Шаг 4: Выводы

Резюмируя выше проанализированные данные, подведем итоги:

- Общий объем выручки растет. Также наблюдается сезонность роста и падения выручки. Отмечается падение выручки зимой и летом, наилучшие показатели осенью. Возможно из-за начала учебного года (ведь университеты и исследовательские центры корпоративные клиенты). Однако, имеет смысл дополнительно проанализировать почему так происходит.
- Как мы видим из таблицы топ -25 товаров: лидерами продаж являются: копир Canon, кресло для руководителей серии HON 5400 и 3D принтер Magenta. Тоже причина задуматься, над третьей позицией оборудование для 3D печати может быть задействовано во многих областях (от строительства до медицины), я бы порекомендовала изучить структуру корпоративных клиентов, которые покупают это оборудование, возможно будет иметь смысл расширить ассортимент для них. За счет этого мы будем удерживать лояльность старых клиентов, так как им еще удобнее и проще будет с нами взаимодействовать, плюс сможем начать привлекать новых клиентов.
- Наши основные клиенты - это корпоративные клиенты, их намного больше, и выручки они приносят тоже в разы больше!
- При анализе по привлечению клиентов, мы выяснили, что в самом начале было привлечено много корпоративных клиентов, а после апреля 2018 года новые клиенты почти не привлекались. Таким образом выручка растет за счет старых корпоративных клиентов. Следует обратить на это внимание, и выяснить почему в компанию перестали обращаться новые клиенты, дать команду активизироваться команде маркетинга.
- При анализе работы службы доставки выявили следующую проблему: у нас достаточно большая доля заказов приходит с опозданием, особенно если это доставка вторым классом. Необходимо проанализировать более детально и подробно работу службы доставки, чтобы сделать более конкретные выводы. Возможно, задержки в доставках являются причиной того, что к нам приходит меньше новых клиентов.
- Повторим, что лидером по количеству доставок и выручки является штат Калифорния, второе место занимает Нью-Йорк, а третье - Техас.
- Логичным становится вывод об открытии офлайн магазина в одном из штатов лидеров. Как я уже отмечала, я бы дополнительно уточнила некоторые данные, может в одном из этих штатов есть уже склад компании, например, и там будет проще, быстрее и дешевле открыть точку офлайн продаж. А если склад находится в Калифорнии, то выбор очевиден. Может будет даже лучше открыть несколько офлайн магазинов, это может снизить нагрузку на службу доставки, и уменьшить количество заказов, доставленных с опозданием. Это одна из возможных причин, но необходимо провести более детальный анализ.
- При открытии офлайн магазина, не лишним будет подумать и о привлечении дополнительных розничных клиентов. Как правило, большая часть обычных клиентов любит покупать канцелярские принадлежности вживую, чтобы их можно было потрогать, понюхать (ароматизированные ручки одно время были очень популярны)).