

Part 3. Анализ логистики компании.

Определим **эффективность доставки**.

Эффективной доставка считается, **если** заказ был доставлен **вовремя**:

- ☐ **Standard Class**(стандартная доставка) ⇒ **доставка в течение 6-ти дней**
- ☐ **Second Class**(доставка вторым классом) ⇒ **доставка в течение 4-х дней**
- ☐ **First Class**(доставка первым классом) ⇒ **доставка в течение 3-х дней**
- ☐ **Same Day**(экспресс-доставка) ⇒ **экспресс-доставка в тот же день**

➤ Определим, какая **доля** заказов **выполняется в срок** по каждой категории доставки (запрос sql):

```

/*определяем статус доставки: успешно доставлена или же с опозданием*/
with category_delivery as
(select
    order_id,
    ship_mode,
    case
        when (ship_mode='Standard Class' and sd.ship_date-sd.order_date<=6)
            or (ship_mode='Second Class' and sd.ship_date-sd.order_date<=4)
            or (ship_mode='First Class' and sd.ship_date-sd.order_date<=3)
            or (ship_mode='Same Day' and sd.ship_date-sd.order_date=0)
        then 'success'
        else 'late' end status
    from sql.store_delivery sd)
select
    distinct ship_mode,
    count(order_id) over(partition by ship_mode) as orders_cnt,
    count(order_id) filter(where status='late')
        over(partition by ship_mode) as late_orders_cnt,
    /*-считаем количество доставок с опозданием в окне по категориям*/
    round((count(order_id) filter(where status='success')
        over(partition by ship_mode))::numeric
        *100/count(order_id) over(partition by ship_mode),2) as success
    /*-делим кол-во успешных доставок, умноженное на 100, на кол-во доставок в
    целом*/
    from
        category_delivery
    order by 4

```

Результат запроса:

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.

Глядя на таблицу выше, можно сделать вывод, что чаще всего **задерживались** доставки, отправленные **вторым классом**.

А доставки первым классом в **99.9%** были доставлены во время.

Рассмотрим **подробнее** категорию доставок вторым классом, чтобы понять: доставляются они с опозданием постоянно или только на протяжении ограниченного периода времени.

- Найдем **долю** заказов, отправленных **вторым классом**, которые были доставлены **с опозданием**, по кварталам, по данной категории (запрос sql):

select

date_trunc('quarter',sd.order_date)::date "date_quart", /*-определяем квартал от даты*/

round(count(case

when sd.ship_date-sd.order_date>4 then sd.order_id end)::numeric

***100/count(distinct sd.order_id),1) percent_later** /*-выставляем условия для доставок с опозданием*/

from sql.store_delivery sd

where sd.ship_mode='Second Class'

group by 1

Построим **график** на основе данных ([см Диаграмма 3 - Доля заказов, отправленных вторым классом с опозданием.](#))

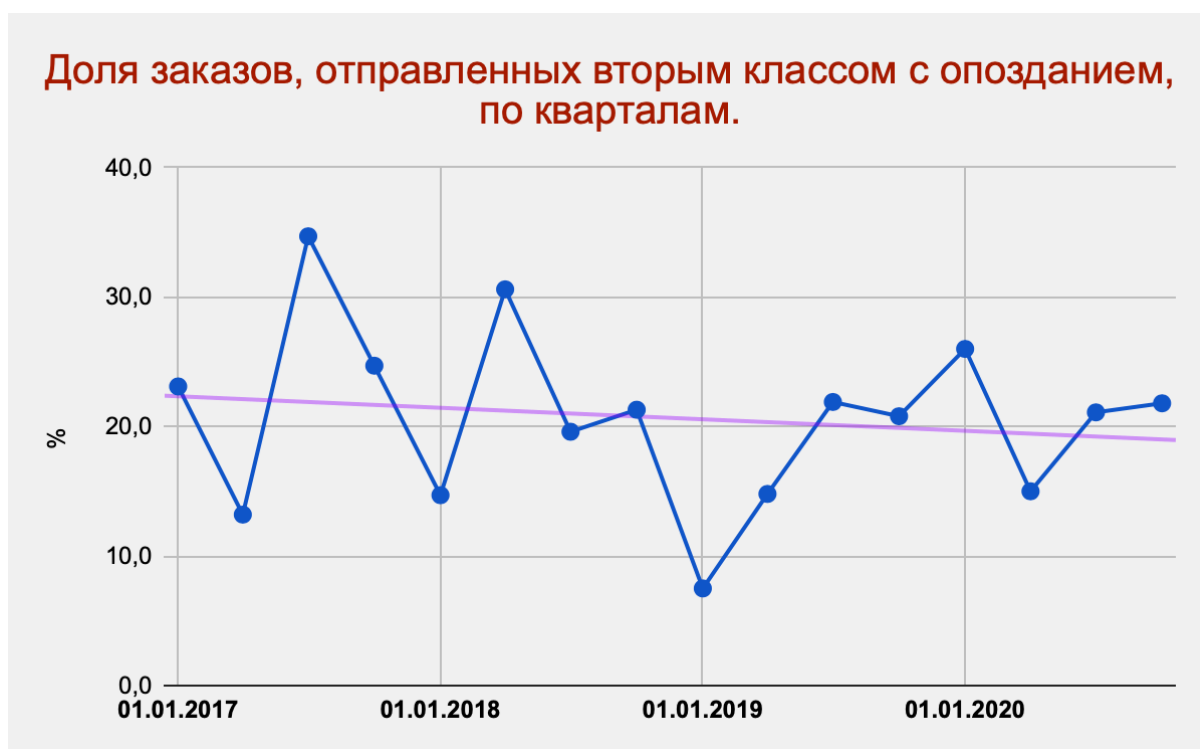
Как видно из графика, что самое **высокое** значение заказов с **опозданием** в году приходилось на **конец весны - начало лета**.

Но начиная с **середины 2019** года ситуация более менее **стабилизируется** в лучшую сторону. Уже нет таких резких отрывов в показателях.

Можно предположить, что такое колебание связано с **сезонностью**. Но чтобы точнее ответить на этот вопрос, нужно **детальнее** подойти к анализу.

Принять все связывающие **факторы** во внимание. И уже тогда на основе этого анализа делать выводы и рекомендации.

Диаграмма 3 - Доля заказов, отправленных вторым классом с опозданием.



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*). С помощью офлайн-магазина можно привлечь **больше клиентов** и **снизить стоимость** доставки, нужно только выбрать, **где** его открыть.

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

➤ Найдем **штат** с самым **большим** количеством **доставок** (запрос sql):

```

select
    sd.state,
    count(sd.order_id) as count_delivery
from sql.store_delivery sd
group by 1
order by 2 desc

```

Таким штатом оказалась **Калифорния** с количеством доставок - **1 021**.

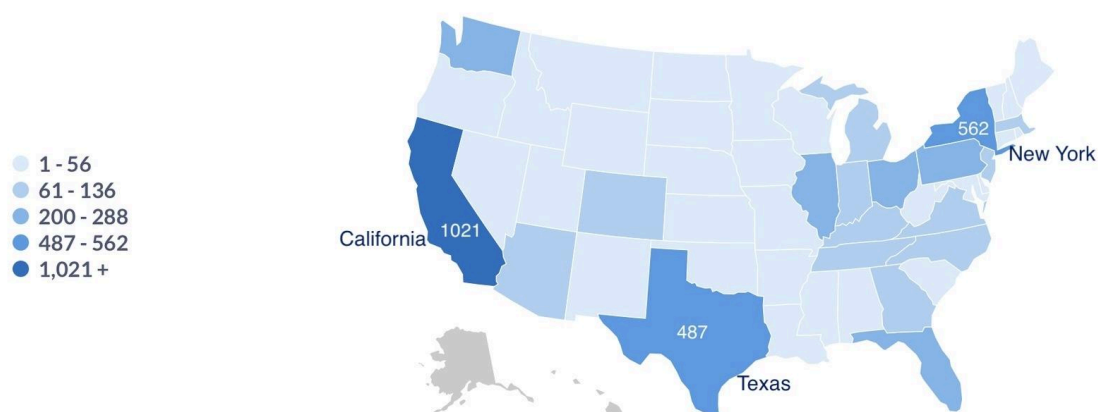
- Далее найдем **город** с самым **большим** количеством **доставок** (запрос sql):

```
select
    sd.city,
    count(sd.order_id) as count_delivery
from sql.store_delivery sd
group by 1
order by 2 desc
```

Самым **популярным** городом по количеству доставок оказался **Нью Йорк** с количеством доставок - **450**.

Ситуация **неоднозначная** и требует **визуализации** результата по доставкам по штатам с помощью **карты** в *Metabase*.

Диаграмма 4 - Карта с визуализацией доставок по штатам.



Как можно заметить по карте доставок с правой стороны у нас интенсивно окрашены штаты, а значит в ту область совершается **большое** количество доставок. Но не стоит забывать про **Калифорнию**, с лидирующим значением по количеству доставок.

В **топ 3** штата по доставкам входит **Нью Йорк, Калифорния и Техас**. В Техасе у нас уже есть склад. Рассмотрим оставшиеся **два** штата по прибыли.

- Рассчитаем какую **прибыль** нам приносят штаты **Нью Йорк** и **Калифорния** (запрос sql):

```
select
    sd.state,
    round(sum((sp.price*(1-sc.discount))*sc.quantity)) revenue
from sql.store_carts sc
    join sql.store_products sp on sc.product_id=sp.product_id
    join sql.store_delivery sd on sd.order_id=sc.order_id
where sd.state in ('California','New York')
group by 1
```

Штат **Нью Йорк** приносит выручку - **201 702\$**. **Калифорния** - **292 432\$**. Это на **30%** больше. Посмотрим как обстоят дела с доставками с опозданием.

- Рассчитаем **долю** доставок с опозданием по **штатам** (запрос sql):

```
with count_delivery as
(select
    sd.state,
    count(case
        when
            (ship_mode='Standard Class' and sd.ship_date-sd.order_date>6)
            or (ship_mode='Second Class' and sd.ship_date-sd.order_date>4)
            or (ship_mode='First Class' and sd.ship_date-sd.order_date>3)
            or (ship_mode='Same Day' and sd.ship_date-sd.order_date=0)
            then sd.order_id end)::numeric as late_delivery, /*-выставляем
условия для доставок с опозданием*/
    count(sd.order_id)::numeric as all_delivery
from sql.store_delivery sd
where sd.state in ('California', 'New York')
group by 1)
select
    state,
    late_delivery,
    all_delivery,
    round((late_delivery/all_delivery)*100,2) percent_late
from count_delivery
```

Результат запроса:

state	late_delivery	all_delivery	percent_late
New York	88	562	15.66
California	160	1021	15.67

Таблица 4 - Доля доставок с опозданием по штатам.

Доля доставок с опозданием у обоих штатов **совпала**. Получается **15,7%** заказов были доставлены **не во время**.

Пришло время **определить** штат, где выгоднее всего открыть **офлайн-магазин**.

Мы выяснили, что в **Калифорнии** количество **доставок** почти в 2 раза выше, чем в **Нью-Йорке**. Но сумма выручки **всего** на 30% больше. И тот факт, что штаты, находящиеся рядом с Нью Йорком, востребованы по количеству доставок, навевает мысль открыть магазин именно в Нью-Йорке. Но для начала я хочу убедиться, что они приносят **достаточно** выручки.

Далее я хочу взять в расчет штаты, которые **граничат** с Калифорнией и Нью-Йорком. И где количество доставок больше **61**. И посчитать выручку.

Воспользуюсь интерактивной картой в *metabase*, где я считала количество доставок по штатам, чтобы сделать выборку.

➤ Расчет **выручки** для штатов **Калифорния** и **Аризона** (запрос sql):

```
select
  round(sum((sp.price*(1-sc.discount))*sc.quantity)) as revenue
from sql.store_carts sc
  join sql.store_products sp on sc.product_id=sp.product_id
  join sql.store_delivery sd on sd.order_id=sc.order_id
where sd.state in ('California','Arizona')
```

Сумма **выручки** вышеупомянутых штатов - **315 334\$**.

Посмотрим, что с выручкой у **Нью-Йорка** и близрасположенных к нему штатов.

- Расчет **выручки** для штатов **Нью-Йорк, Пенсильвания, Огайо, Нью-Джерси и Массачусетс** (запрос sql):

select

```
round(sum((sp.price*(1-sc.discount))*sc.quantity)) as revenue
from sql.store_carts sc
join sql.store_products sp on sc.product_id=sp.product_id
join sql.store_delivery sd on sd.order_id=sc.order_id
where sd.state in ('New York','Ohio', 'Pennsylvania', 'Massachusetts', 'New Jersey')
```

Сумма **выручки** вышеупомянутых штатов **364 901\$**. Это на **15.7% больше** суммы выручки **Калифорнии и Аризона**.

Итак, несмотря на большую территорию Калифорнии и крупное количество доставок, территория правее Техаса выглядит **перспективнее**.

А значит, что **удачное** место для нового **офлайн-магазина** находится в **Нью-Йорке**.

- **Подводим итоги по третьей части анализа.**

Мы выяснили, что товары, которые были отправлены **вторым** классом доставлялись в **79.05%** во время. Остальные **20.9%** доставок дошли до клиентов с **опозданием**. Начиная с **2017** года до середины **2019** года процент доставок с опозданием колебался в пределах от **7.5%** до **34.7%**.

И начиная с середины **2019 г.** доля доставок с опозданием начала **стабилизироваться** и имела значение от **15%** до **26%**. (*см Диаграмма 3*)

Далее мы выяснили, что самым **популярным штатом** по доставкам оказалась **Калифорния** со значением **1 021** количества доставок. Самым популярным **городом** по доставкам оказался **Нью-Йорк**. Количество доставок туда составило **450**.

И на вопрос: **“В каком штате перспективнее всего открыть офлайн-магазин?”** свой выбор сделала в пользу штата **Нью-Йорк**.

Свое мнение подкрепила небольшим **анализом**, где решила выяснить какую **прибыль** нам приносит Нью-Йорк и близрасположенные штаты. Оказалось, что прибыль **пяти** небольших штатов **превысила** выручку большой по территории Калифорнии и Аризона. И свой выбор я обосную тем, что в **районах**, которые расположены справа от Техаса (как раз, где находится Нью-Йорк), штаты в **большей** мере делают заказы в нашей компании и приносят неплохую **выручку**.