

ЗАДАЧА

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

Дорожная карта

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).

1. Оценка динамики продаж и распределение выручки по товарам.

1.1. Определим период осуществления заказов

```
select
  min(order_date), --определим дату первого заказа
  max(order_date) --определим дату последнего заказа
From sql.store_delivery
```

Теперь мы знаем, что заказы производились в период с 2017-01-03 по 2020-12-30

1.2. Определим количество заказов за полученный период

```
select
  count(distinct order_id) --производим подсчет количества уникальных заказов
From sql.store_delivery
```

Количество заказов за исследуемый период составило 5009 шт.

1.3. Определим эффективность продаж

1.3.1. Определим объем выручки по всем заказам за указанный период

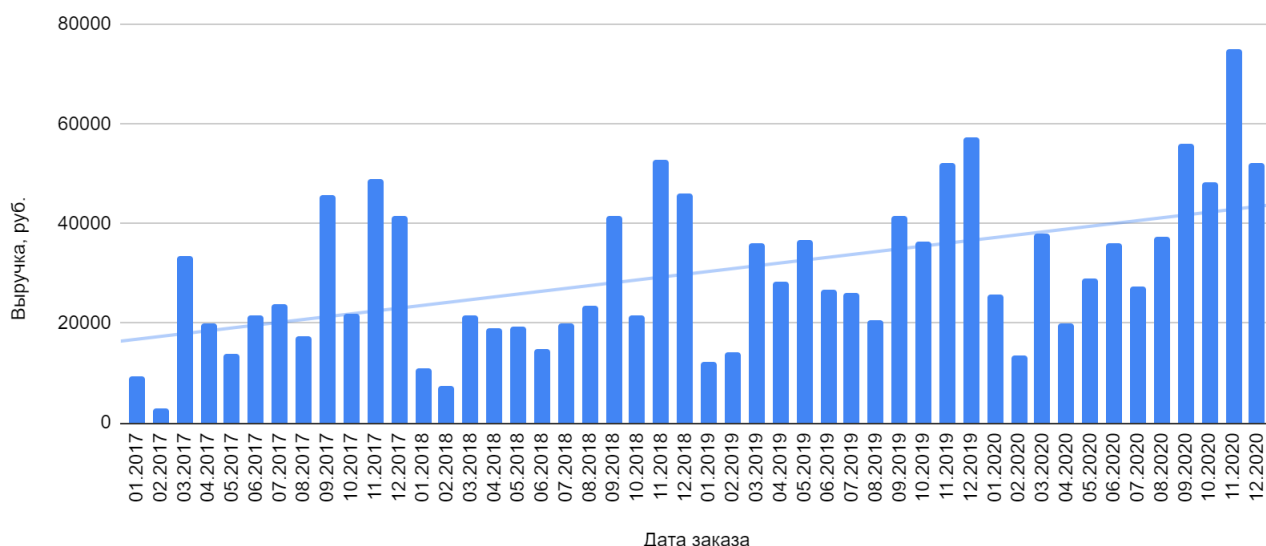
```
select
  sum(price*(1-discount)*quantity) -- рассчитываем объем выручки по всем заказам
From sql.store_carts sc
  join sql.store_products sp on sc.product_id=sp.product_id
```

Общая выручка по всем заказам за исследуемый период составляет 1 446 157 руб.

1.3.2. Определим объем выручки с разбивкой по месяцам

```
select
  date_trunc('month', order_date) date, -- определяем месяц полученной выручки
  ROUND(sum(price*(1-discount)*quantity)) revenue --определяем объем выручки с
округлением до целого
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
group by 1--группируем по дате
order by 1 --сортируем по дате
```

Для наглядности представим полученный результат в виде диаграммы



Оценив данные графика и обратив внимание на линию тренда, можно сделать вывод, что выручка в целом по компании растет. Однако стоит отметить, что выручка в компании подвержена сезонным изменениям. Так наблюдается снижение выручки в первые месяцы года и резкий подъем выручки к концу.

1.3.3. Определим выручку по категориям товаров

```
select
  category,
  subcategory,
  ROUND(sum(price*(1-discount)*quantity)) revenue --опред-ем объем выручки с
округлением до целого
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
```

group by 1,2 --группируем по категории и подкатегории
order by 3 desc --сортируем в порядке убывания

Для наглядности представим полученный результат в виде диаграммы

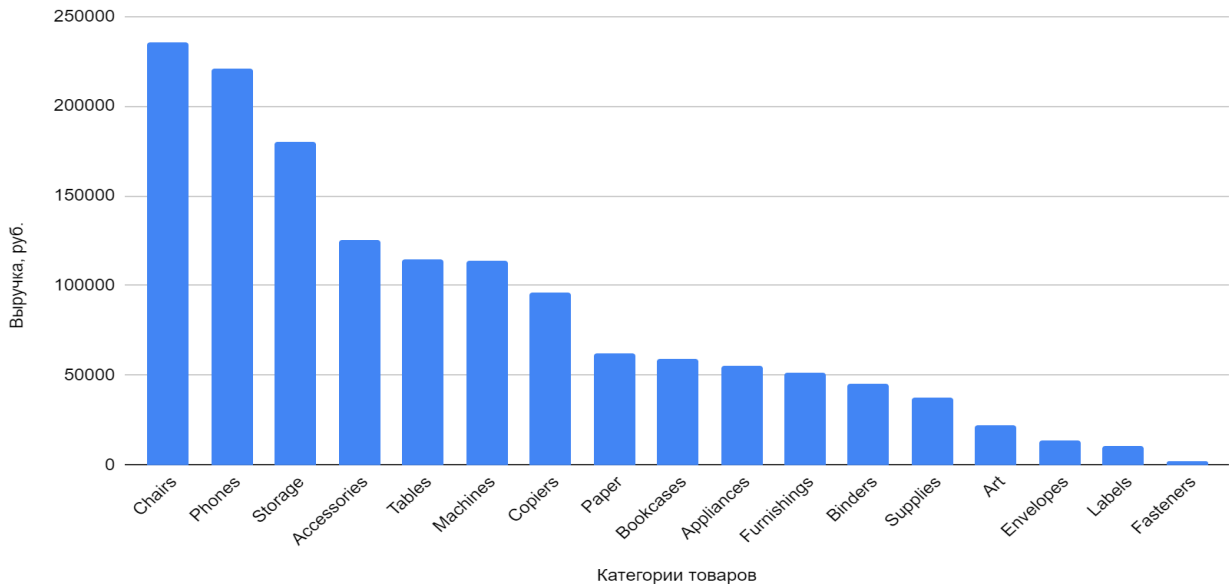


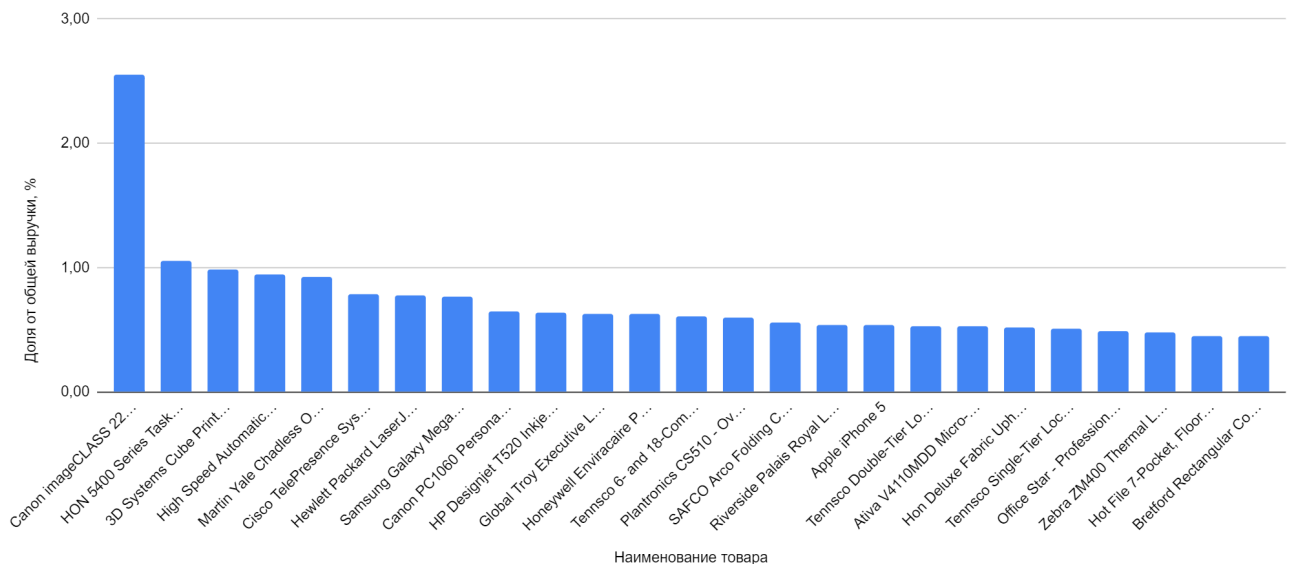
Диаграмма показывает, что наибольшую выручку приносят кресла, а наименьшую крепежные изделия. Это обусловлено не столько спросом сколько ценой приобретаемых товаров.

Определим топ-25 товаров по объему выручки

```
select
  product_nm,
  ROUND(sum(price*(1-discount)*quantity)) revenue, --опред-ем объем выручки с
округлением до целого
  count(quantity), --определяем количество товаров
  sum(price*(1-discount)*quantity)/
    (select --создадим подзапрос
      sum(price*(1-discount)*quantity) total --определим сумму всех
товаров
    from sql.store_carts sc
    join sql.store_products sp on sc.product_id=sp.product_id)
    *100 percent_from_total --опред-им долю выручки товаров
  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
group by 1 --сгруппируем по продукту
order by 2 desc --отсортируем по объему выручки
Limit 25 --определим ТОП-25
```

ТОП-25 товаров представлен в Таблице

Наименование товара	Выручка, \$.	Количество, шт.	Доля от общей выручки, %
Canon imageCLASS 2200 Advanced Copier	36 960,00	5	2,56
HON 5400 Series Task Chairs for Big and Tall	15 310,00	8	1,06
3D Systems Cube Printer, 2nd Generation, Magenta	14 300,00	2	0,99
High Speed Automatic Electric Letter Opener	13 624,00	3	0,94
Martin Yale Chadless Opener Electric Letter Opener	13 325,00	6	0,92
Cisco TelePresence System EX90 Videoconferencing Unit	11 319,00	1	0,78
Hewlett Packard LaserJet 3310 Copier	11 304,00	8	0,78
Samsung Galaxy Mega 6.3	11 155,00	6	0,77
Canon PC1060 Personal Laser Copier	9 296,00	4	0,64
HP Designjet T520 Inkjet Large Format Printer - 24" Color	9 188,00	3	0,64
Global Troy Executive Leather Low-Back Tilter	9 083,00	9	0,63
Honeywell Enviracaire Portable HEPA Air Cleaner for 17' x 22' Room	9 044,00	8	0,63
Tennsco 6- and 18-Compartment Lockers	8 740,00	11	0,60
Plantronics CS510 - Over-the-Head monaural Wireless Headset System	8 658,00	10	0,60
SAFCO Arco Folding Chair	8 101,00	13	0,56
Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	7 805,00	5	0,54
Apple iPhone 5	7 798,00	6	0,54
Tennsco Double-Tier Lockers	7 705,00	12	0,53
Ativa V4110MDD Micro-Cut Shredder	7 700,00	2	0,53
Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	7 446,00	12	0,51
Tennsco Single-Tier Lockers	7 327,00	8	0,51
Office Star - Professional Matrix Back Chair with 2-to-1 Synchro Tilt and Mesh Fabric Seat	7 020,00	8	0,49
Zebra ZM400 Thermal Label Printer	6 966,00	2	0,48
Hot File 7-Pocket, Floor Stand	6 511,00	13	0,45
Bretford Rectangular Conference Table Tops	6 498,00	12	0,45



Наибольшую долю выручки получаем от реализации копировальных аппаратов Canon, в то время как стулья по доле выручки располагаются на втором месте в рейтинге. Данное расположение в рейтинге связано с более высокой стоимостью единицы копировальной техники без учета количества проданных товаров.

2. Определим портрет клиента

2.1. Найдем у компании количество клиентов в разных сегментах и какую долю от общей выручки они приносят

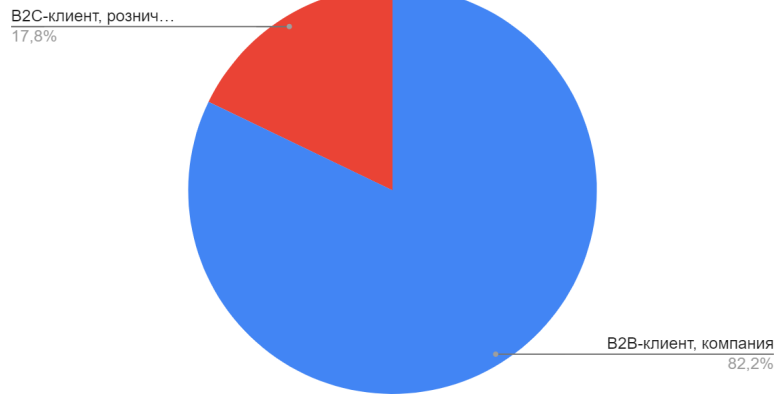
```

select
  scs.category,
  count(scs.cust_id) cust_cnt, --определяем количество заказов по категориям
  ROUND(sum(price*(1-discount)*quantity)) revenue --определим объем выручки
From sql.store_customers scs
  join sql.store_delivery sd on sd.cust_id=scs.cust_id
  join sql.store_carts sc on sc.order_id=sd.order_id
  join sql.store_products sp on sc.product_id=sp.product_id
group by 1 --группируем по категории клиентов
order by 3 desc --сортируем по объему выручки

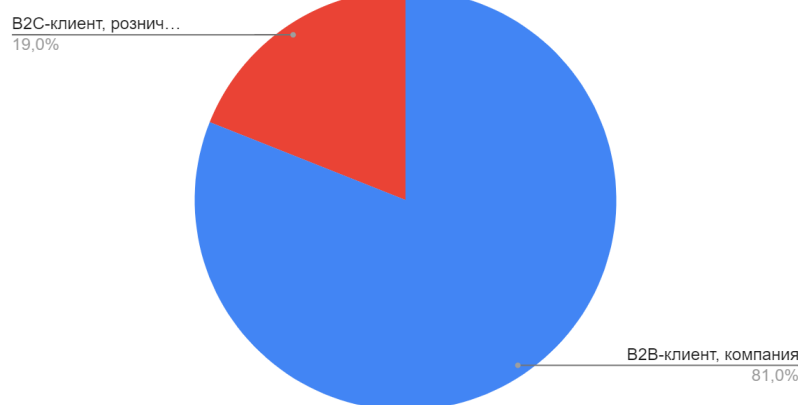
```

Категория клиента	Кол-во, шт.	Выручка, \$
B2B-клиент, компания	8 211,00	1 172 009,00
B2C-клиент, розничный покупатель	1 783,00	274 148,00

Доля клиентов



Доля выручки по сегментам клиентов



Из диаграмм видно, что доля клиентов по каждой категории сопоставимы с объемами выручки, которые они приносят. Стоит отметить, что преобладающий сегмент клиентов - B2B-клиент, Компании. Сопоставив данный факт с неравномерностью сезонной выручки, можно сделать вывод, что она связана с проведением компаниями торгово-закупочных процедур и освоению денежных средств в конце производственного периода.

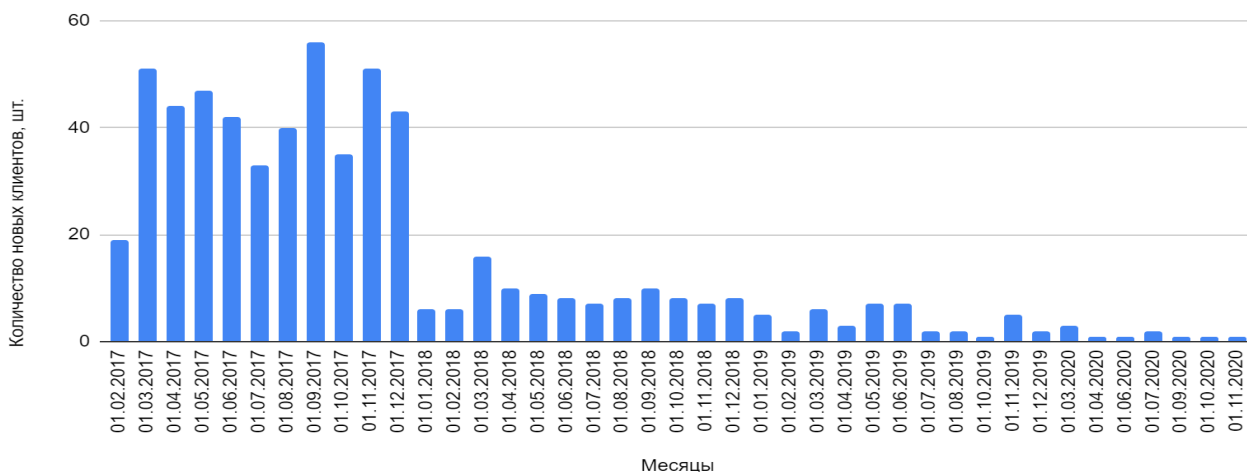
2.2. Определим количество новых корпоративных клиентов по месяцам

--создадим вспомогательную таблицу

```
with a as(
  select
    distinct cust_id,
    min(order_date) month_ --определим дату первой покупки
  From sql.store_delivery
  group by 1) --сгруппируем по коду покупателя
select
  date_trunc('month', a.month_) month_, --определим месяц первой покупки
```

```
count(a.cust_id) new_custs --определяем количество покупателей совершивших
покупку
```

```
From sql.store_customers scs
  join a on a.cust_id=scs.cust_id
where scs.category= 'Corporate'
group by 1 -- группируем по месяцу
order by 1 desc --сортируем по месяцу
```



Как видно из графика, в самом начале работы было привлечено много клиентов, а после 2018 года новые клиенты почти не привлекались. Несмотря на то, что количество новых клиентов падает, выручка по годам растем, данный факт может говорить о высокой возвращаемости клиентов и удовлетворенности клиентов предлагаемыми услугами. Тем не менее очень важно не останавливаться на достигнутом, необходимо возобновить кампанию по привлечению клиентов.

2.3. Определим среднее количество различных товаров в заказах у корпоративных клиентов

```
--создадим вспомогательную таблицу
with a as(
  select
    order_id,
    count(distinct product_id) quantity --определим количество различных продуктов в
заказе
```

```
From sql.store_carts
group by 1)
```

```
select
  round(avg(quantity),1) --определим среднее кол-во различных продуктов в
заказе
```

```
From a
  join sql.store_delivery sd on sd.order_id=a.order_id
  join sql.store_customers sc on sd.cust_id=sc.cust_id
where sc.category= 'Corporate' --только для корпоративных клиентов
```

Среднее количество различных товаров в заказах корпоративных клиентов составляет 2 шт.

2.4. Определим среднюю сумму заказов у корпоративных клиентов

```
--создадим вспомогательную таблицу
with a as(
select
  order_id,
  sum(price*(1-discount)*quantity) revenue --определим выручку по каждому заказу
From sql.store_carts sc
  join sql.store_products sp on sc.product_id=sp.product_id
group by 1) --сгруппируем по заказам

select
  round(avg(revenue),1) --определим среднюю выручку по заказам
From a
  join sql.store_delivery sd on sd.order_id=a.order_id
  join sql.store_customers sc on sd.cust_id=sc.cust_id
where sc.category= 'Corporate' --только для корпоративных клиентов
```

Средняя сумма заказов у корпоративных клиентов составляет 285.9 руб.

2.5. Определим среднее количество офисов у корпоративных клиентов

```
--создадим вспомогательную таблицу
with a as(
select
  cust_id,
  count(distinct zip_code) adres --определим количество адресов доставки у каждого заказчика
From sql.store_delivery
group by 1)

select
  round(avg(adres),1) --определим среднее количество адресов у заказчиков
From a
  join sql.store_customers sc on a.cust_id=sc.cust_id
where sc.category= 'Corporate' --только для корпоративных клиентов
```

Среднее количество офисов у корпоративных клиентов равно 6,2

Невысокое среднее количество товаров в заказе и низкая средняя цена заказа при достаточно большом количестве офисов и высокой выручке по сегменту говорит о большом количестве низкостоймостных заказов и большой площади расположения компаний по территории страны.

3. Проанализируем логистику компании

3.1. Определим долю заказов выполняется в срок по каждой категории доставки

--создадим вспомогательную таблицу

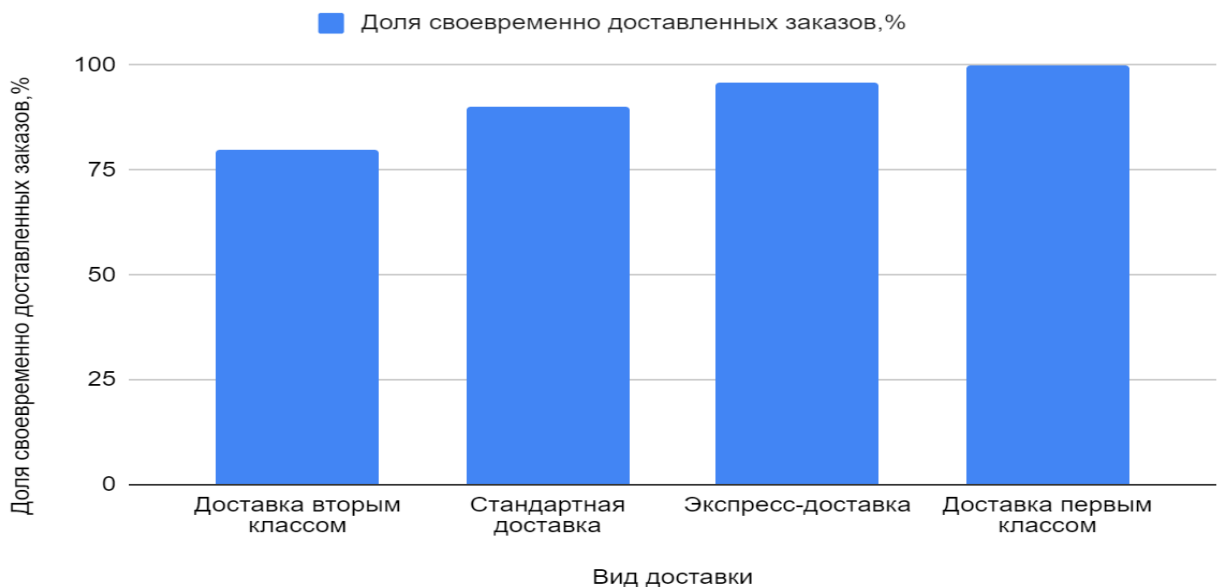
```
with a as(
select
  order_id,
  ship_mode,
  CASE WHEN ship_mode='Standard Class' and (ship_date-order_date)>6 THEN 'Не вовремя'
        WHEN ship_mode='Second Class' and (ship_date-order_date)>4 THEN 'Не вовремя'
        WHEN ship_mode='First Class' and (ship_date-order_date)>3 THEN 'Не вовремя'
        WHEN ship_mode='Same Day' and ship_date<>order_date THEN 'Не вовремя'
        ELSE 'Вовремя'
      END AS late_orders --определяем статус заказа по своевременной доставке в
                          каждой категории доставки
From sql.store_delivery),
```

b as(--создадим вспомогательную таблицу

```
select
  ship_mode,
  count(late_orders) late_orders_cnt --определяем кол-во заказов в каждой категории
From a
where late_orders='Не вовремя' --только для не вовремя доставленных
group by 1)
```

```
select
  sd.ship_mode,
  count(order_id) orders_cnt, --определяем количество заказов в каждой категории
  late_orders_cnt,
  round(100-(late_orders_cnt*100/count(order_id)),2) "% success" --определяем долю
заказов доставленных своевременно
From b
  join sql.store_delivery sd on b.ship_mode=sd.ship_mode
group by 1,3 --группируем
order by 4 --сортируем
```

Вид доставки	Количество заказов, шт.	Количество заказов не доставленных в срок, шт.	Доля своевременно доставленных заказов, %
Доставка вторым классом	964	202	80,00
Стандартная доставка	2994	309	90,00
Экспресс-доставка	264	12	96,00
Доставка первым классом	787	1	100,00



Чаще всего с опозданием доставляются заказы, отправленные вторым классом.

3.2. Определим долю заказов, отправленных вторым классом, которые были доставлены с опозданием, по кварталам.

--создадим вспомогательную таблицу

```
with a as(
select
    order_id,
    sd.ship_mode,
    EXTRACT(QUARTER FROM order_date) order_QUARTER, --определяем квартал
    CASE WHEN ship_mode='Second Class' and (ship_date-order_date)>4 THEN 'Не вовремя'
    ELSE 'Вовремя'
    END AS late_orders --выделяем заказы направленные вторым классом и
    доставленные не вовремя
From sql.store_delivery sd),
```

--создадим вспомогательную таблицу

```
b as(
select
    a.ship_mode,
    order_QUARTER,
    count(late_orders) late_orders_cnt --определяем кол-во заказов доставленных вторым
    классом не вовремя
From a
where late_orders='Не вовремя' --только для статуса не вовремя
group by 1,2)
```

```
select
    sd.ship_mode,
    order_QUARTER,
    count(order_id) orders_cnt, --определяем количество заказов по кварталам
```

```

late_orders_cnt,
round(late_orders_cnt*100/count(order_id),2) "% success" --определяем долю заказов
вторым классом доставленных не вовремя по кварталам
From b
join sql.store_delivery sd on b.ship_mode=sd.ship_mode
group by 1,2,4 --группируем
order by 2 --сортируем

```

Вид доставки	Квартал	Количество заказов, шт.	Количество заказов не доставленных в срок, шт.	Доля не своевременно доставленных заказов, %
Доставка вторым классом	1	964	27	2
Доставка вторым классом	2	964	33	3
Доставка вторым классом	3	964	63	6
Доставка вторым классом	4	964	79	8

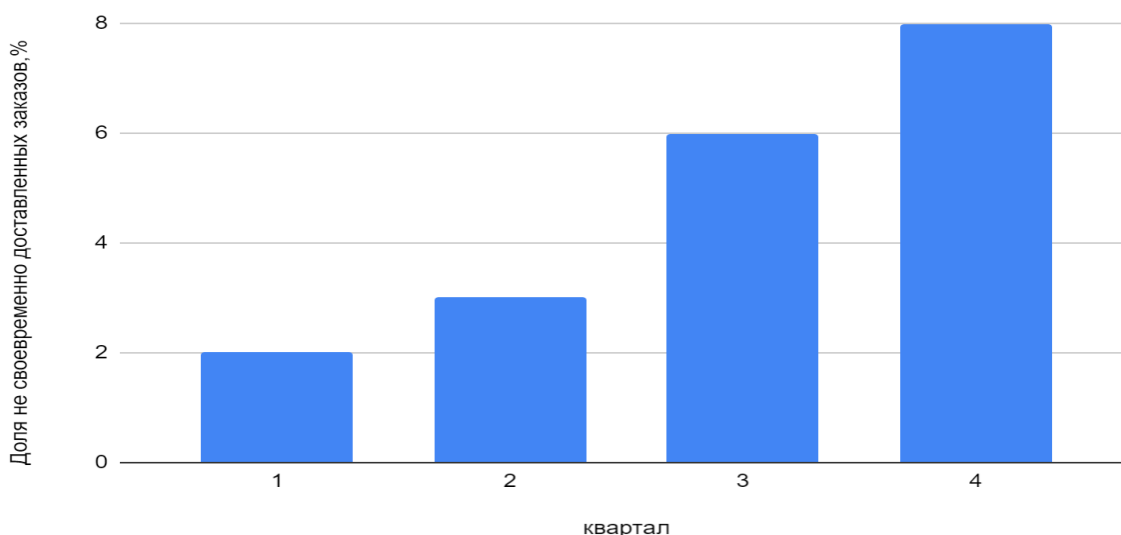


Диаграмма показывает, что количество заказов доставленных с опозданием возрастает во второй половине года. Это связано с увеличением количества заказов к концу года. Необходимо снизить этот показатель для повышения удовлетворенности клиентов качеством обслуживания и избежания их оттока.

3.3. Определим наиболее популярный по количеству доставок штат

```

select
state,
count(order_id) --определяем количество заказов по штатам
From sql.store_delivery
group by 1 --группируем по штату
order by 2 desc --сортируем по количеству в порядке убывания
limit 1 --ограничиваем лимит одним значением

```

Наибольшее количество заказов доставляется в штат Калифорния (1021 шт.)

3.4. Определим наиболее популярный по количеству доставок город

```
select
  city,
  count(order_id) --определяем количество заказов по городам
From sql.store_delivery
group by 1 --группируем по городу
order by 2 desc --сортируем по количеству в порядке убывания
limit 1 --ограничиваем лимит одним значением
```

Наибольшее количество заказов доставляется в город Нью-Йорк (450 шт.).

3.5. Определим распределение заказов по штатам

```
select
  state,
  count(order_id) --определяем количество заказов по штатам
From sql.store_delivery
group by 1
```



Таким образом мы получаем два самых востребованных направления доставки, которые располагаются в противоположных концах страны.

Несмотря на то, что в Калифорнии 40 млн потенциальных покупателей против 8,4 млн в Нью-Йорке и количество доставляемых заказов в этот штат больше, чем в Нью-Йорк в 2 раза я советую открыть офлайн магазин именно в г. Нью-Йорк.

На это имеется ряд причин:

- плотность населения в городе выше, а значит и количество потенциальных покупателей будет выше, которые обратятся в магазин по совету постоянных клиентов;
- по карте видно, что большинство заказов доставляются на восточное побережье США. Оно является наиболее населенным, в отличие от центральной и северо-западной части, а значит более перспективно в части привлечения новых клиентов. Тем более население этих штатов имеет наибольшую покупательскую способность. Также там располагается большое количество офисов нуждающихся в наших товарах.

Из вышесказанного предлагаю в Нью-Йорке открыть не только магазин, но и склад, который позволит снизить транспортные расходы (заказы в среднем состоят из двух наименований и небольшие по стоимости) и время доставки заказов.

Таким образом открывая в Нью-Йорке магазин и склад мы получаем огромный потенциал для развития в части привлечения новых клиентов и увеличения выручки, снижаем транспортные затраты, уменьшаем количество заказов доставленных с опозданием.