

Project-2. Решение бизнес-задачи с помощью SQL

Цель

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

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).

1. Определение эффективности продаж

1.1. Определение динамики продаж

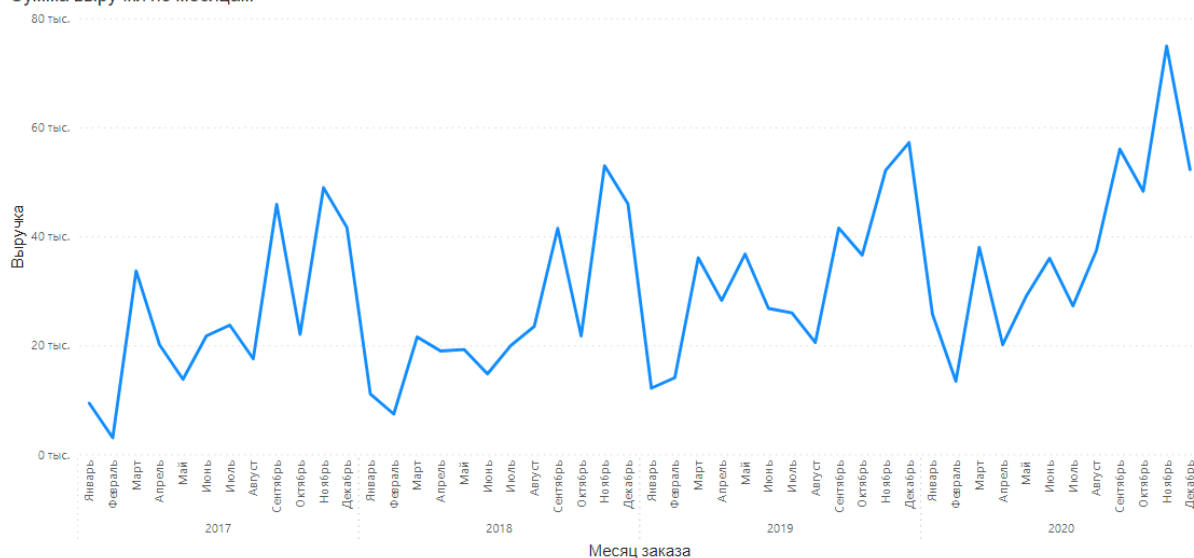
Для определения динамики продаж был составлен следующий запрос:

```
SELECT
    date_trunc ('month', order_date)::date as date,
    ROUND (SUM (sc.quantity*sp.price*(1-sc.discount))) as 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
```

Пояснения по запросу: рассчитываем необходимые для вывода столбцы с месяцем и суммой выручки, соединяя таблицы store_delivery и store_carts по столбцу order_id и таблицы store_carts и store_products по столбцу product_id.

На основе полученных данных был построен график за расчетный период:

Сумма выручки по месяцам



Можно сделать вывод о том, что выручка растет, но также при этом наблюдается сезонность роста и падения выручки.

1.2. Определение суммы выручки по категориям

Для определения суммы выручки по различным категориям и подкатегориям был составлен следующий запрос:

```

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

```

Из результатов, полученных при выполнении запроса, следует, что в топе продаж компании находятся кресла (Category: Furniture, Subcategory: Chairs, Revenue: 235 318).

1.3. Топ-25 товаров по объему выручки

Далее был составлен запрос для выведения данных топ-25 товаров по объему выручки.

```

WITH total_revenue AS
(SELECT
    SUM (sp.price*sc.quantity*(1-sc.discount)) as total_revenue
FROM
    sql.store_products sp
    JOIN sql.store_carts sc ON sp.product_id=sc.product_id)

SELECT
    sp.product_nm,
    ROUND (SUM (sp.price*sc.quantity*(1-sc.discount)),2) AS revenue,
    SUM (sc.quantity) AS quantity,
    ROUND (((SUM (sp.price*sc.quantity*(1-sc.discount)))/total_revenue)*100,2) AS
percent_from_total
FROM
    sql.store_products sp
    JOIN sql.store_carts sc ON sp.product_id=sc.product_id
    JOIN total_revenue ON TRUE
GROUP BY 1, total_revenue
ORDER BY 2 DESC
LIMIT 25

```

Пояснения по запросу: был составлен CTE total_revenue для расчета общей выручки от всех товаров, далее были выведены данные о наименовании товара, рассчитан объем выручки для каждого товара, количество проданных товаров, а также доля выручки рассматриваемого товара от общей выручки.

№	product_nm	revenue	quantity	percent_from _total
1	Canon imageCLASS 2200 Advanced Copier	36,959.82	20	2.56
2	HON 5400 Series Task Chairs for Big and Tall	15,309.53	39	1.06
3	3D Systems Cube Printer, 2nd Generation, Magenta	14,299.89	11	0.99
4	High Speed Automatic Electric Letter Opener	13,624.21	11	0.94
5	Martin Yale Chadless Opener Electric Letter Opener	13,325	22	0.92
6	Cisco TelePresence System EX90 Videoconferencing Unit	11,319.24	6	0.78
7	Hewlett Packard LaserJet 3310 Copier	11,303.69	38	0.78
8	Samsung Galaxy Mega 6.3	11,154.87	39	0.77
9	Canon PC1060 Personal Laser Copier	9,295.83	19	0.64
10	HP Designjet T520 Inkjet Large Format Printer - 24" Color	9,187.5	12	0.64
11	Global Troy Executive Leather Low-Back Tilter	9,082.87	32	0.63
12	Honeywell Enviracaire Portable HEPA Air Cleaner for 17' x 22' Room	9,043.55	40	0.63
13	Tennsco 6- and 18-Compartment Lockers	8,740.17	43	0.6
14	Plantronics CS510 - Over-the-Head monaural Wireless Headset System	8,657.89	35	0.6
15	SAFCO Arco Folding Chair	8,100.95	53	0.56
16	Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	7,805.48	24	0.54
17	Apple iPhone 5	7,798	24	0.54
18	Tennsco Double-Tier Lockers	7,704.86	47	0.53
19	Ativa V4110MDD Micro-Cut Shredder	7,699.89	11	0.53
20	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	7,446.44	51	0.51
21	Tennsco Single-Tier Lockers	7,326.59	28	0.51
22	Office Star - Professional Matrix Back Chair with 2-to-1 Synchro Tilt and Mesh Fabric Seat	7,019.5	28	0.49
23	Zebra ZM400 Thermal Label Printer	6,965.7	6	0.48
24	Hot File 7-Pocket, Floor Stand	6,510.77	49	0.45
25	Bretford Rectangular Conference Table Tops	6,497.82	46	0.45

Таким образом, наибольшая выручка была получена от продажи Canon imageCLASS 2200 Advanced Copier в количестве 20 штук, выручка составила 2,56% от общей выручки компании за рассматриваемый период.

2. Портрет клиента

2.1. Определение количества клиентов по категориям

Для определения количества клиентов и выручки по категориям клиента был составлен следующий запрос:

```
SELECT
  sc.category,
  COUNT (DISTINCT sc.cust_id) AS cust_cnt,
  ROUND (SUM (sp.price*(1-c.discount)*c.quantity),0) AS revenue
FROM
  sql.store_customers sc
  JOIN sql.store_delivery sd ON sc.cust_id=sd.cust_id
  JOIN sql.store_carts c ON sd.order_id=c.order_id
  JOIN sql.store_products sp ON c.product_id=sp.product_id
GROUP BY 1
ORDER BY 3 DESC
```

Из результатов расчета следует, что количество *B2B*-клиентов (645) намного больше, чем количество *B2C*-клиентов (148), а также компании приносят в несколько раз большую выручку, чем розничные покупатели.

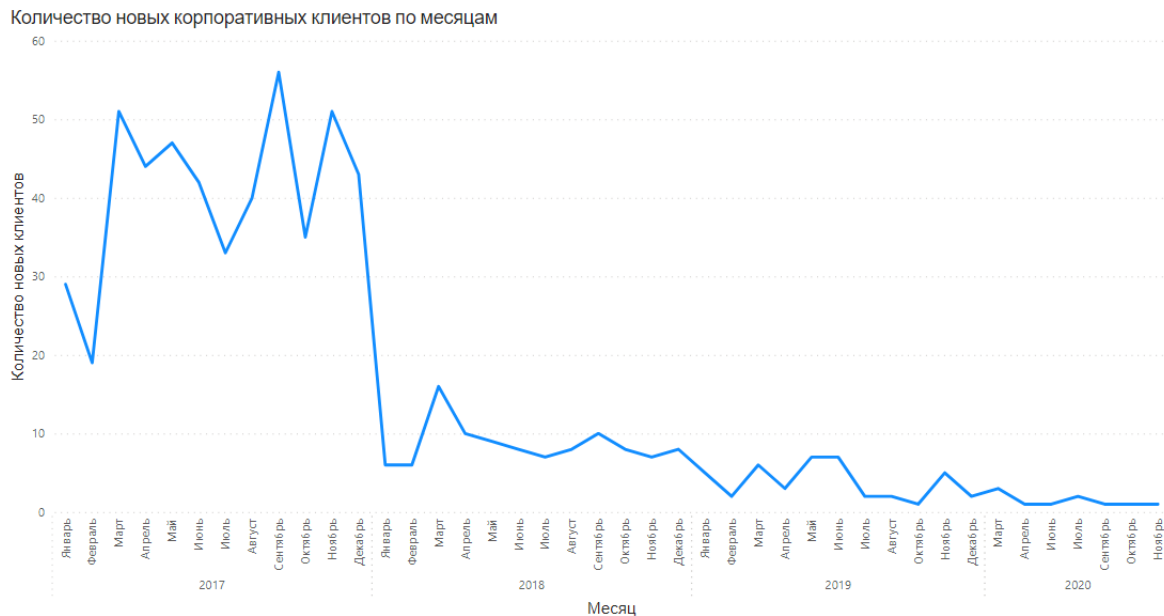
2.2. Динамика новых B2B-клиентов

Для поиска количества новых корпоративных клиентов по месяцам был составлен следующий запрос:

```
WITH first_order AS
(SELECT
  cust_id,
  MIN (DATE_TRUNC ('month', order_date))::date AS month
FROM
  sql.store_delivery
  GROUP BY 1)
SELECT
  first_order.month,
  COUNT (first_order.cust_id) AS new_custs
FROM
  first_order
  JOIN sql.store_customers sc ON first_order.cust_id=sc.cust_id
WHERE sc.category='Corporate'
GROUP BY 1
ORDER BY 1
```

Пояснения по запросу: был составлен СТЕ для определения месяца первого заказа для каждого клиента, далее был проведен расчет количества новых клиентов по каждому месяцу на основании соединения СТЕ с таблицей store_customers по cust_id.

По полученным данным был построен график динамики новых B2B-клиентов.



Как видно из графика, количество новых клиентов растет до 2018 года, а затем идет значительный спад. Новые клиенты с 2018 года почти не привлекались. Так как доля корпоративных клиентов в общей выручке компании высока, рекомендуется обратить внимание на привлечение новых B2B-клиентов.

2.3. Характеристика B2B-клиентов

Для получения характеристики B2B-клиентов был составлен следующий запрос:

```
WITH B2B_products AS
(SELECT
  sd.cust_id,
  sd.order_id,
  COUNT (sc.product_id) AS products,
  SUM (sp.price*(1-sc.discount)*sc.quantity) AS revenue
FROM
  sql.store_delivery sd
  JOIN sql.store_carts sc ON sd.order_id=sc.order_id
  JOIN sql.store_customers c ON sd.cust_id=c.cust_id
  JOIN sql.store_products sp ON sc.product_id=sp.product_id)
```

```
WHERE c.category='Corporate'  
GROUP BY 1,2),
```

```
B2B_offices AS  
(SELECT  
    cust_id,  
    COUNT (DISTINCT zip_code) as offices_qty  
FROM  
    sql.store_delivery  
GROUP BY 1)
```

```
SELECT  
    ROUND (AVG (products),1) AS avg_products_qty,  
    ROUND (AVG (revenue),1) AS avg_revenue_per_order,  
    ROUND (AVG (offices_qty),1) AS avg_offices_qty  
FROM B2B_products  
JOIN B2B_offices ON TRUE
```

Пояснения по запросу: были составлены два CTE: B2B_products - для подсчета количества товаров и выручки по каждому заказу клиента на основании данных всех 4-х таблиц (store_delivery, store_carts, store_customers, store_products) с фильтрацией для корпоративных клиентов, B2B_offices - для подсчета количества офисов для каждого клиента. Далее в основном запросе были взяты средние значения приведенных параметров, округленные до 1-го знака после запятой.

В результате расчета получено, что среднее количество различных товаров в заказах у корпоративных клиентов равно 2, средняя сумма заказа - 285,9, а среднее количество различных офисов у корпоративных клиентов составляет 6,2. На основании этих данных можно сделать вывод о том, что B2B-клиенты, показатели заказов которых превышают приведенные средние значения, представляют наибольшую ценность для компании.

3. Анализ логистики компании

3.1. Эффективность выполнения текущих доставок

Для определения эффективности выполнения текущих доставок был составлен следующий запрос:

```
WITH delivery_time AS  
(SELECT  
    ship_mode,  
    COUNT (order_id) AS orders_cnt,  
    COUNT (CASE WHEN ship_mode='Standard Class' and ship_date-order_date<=6 THEN  
'order_on_time'  
    WHEN ship_mode='Second Class' and ship_date-order_date<=4 THEN 'order_on_time'
```

```

    WHEN ship_mode='First Class' and ship_date-order_date<=3 THEN 'order_on_time'
    WHEN ship_mode='Same Day' and ship_date-order_date=0 THEN 'order_on_time'
    ELSE NULL
  END) AS on_time_orders_cnt
FROM
  sql.store_delivery
GROUP BY 1)

SELECT
  ship_mode,
  orders_cnt,
  orders_cnt-on_time_orders_cnt AS late_orders_cnt,
  ROUND (on_time_orders_cnt*100.00/orders_cnt,2) AS "%success"
FROM delivery_time
GROUP BY 1,2,on_time_orders_cnt
ORDER BY 4

```

Пояснения по запросу: в CTE delivery_time было подсчитано общее количество заказов и количество вовремя доставленных заказов для каждого типа доставки. В основном запросе было выведено общее количество заказов, количество заказов, которые не были доставлены вовремя, а также доля вовремя выполненных заказов по каждому типу доставки.

Результаты запроса представлены в таблице:

ship_mode	orders_cnt	late_orders_cnt	%success
Second Class	964	202	79.05
Standard Class	2,994	309	89.68
Same Day	264	12	95.45
First Class	787	1	99.87

На основании результатов можно сделать вывод о том, что чаще всего с опозданием приходят заказы, отправленные вторым классом.

3.2. Анализ заказов, отправленных вторым классом

Для анализа доли заказов, отправленных вторым классом, которые были доставлены с опозданием (по кварталам) был составлен следующий запрос:


```

WITH Delivery_status AS
(SELECT
    order_id,
    order_date,
    CASE WHEN ship_mode='Second Class' and ship_date-order_date>4 THEN
'late_order'
    ELSE NULL END Order_status
FROM sql.store_delivery
WHERE ship_mode='Second Class')

```

```

SELECT
    DATE_TRUNC ('quarter', order_date)::date AS quarter,
    ROUND (COUNT (CASE WHEN Order_status='late_order' THEN order_id
END)*100.00/COUNT (order_id),2) AS "%_late_orders"
FROM Delivery_status
GROUP BY 1
ORDER BY 1

```

Пояснения по запросу: в CTE Delivery_status были выведены id и даты заказа для заказов второго класса, а также был выведен столбец, в котором указывается значение “late_order” в случае, если заказ не доставлен вовремя. В основном запросе выводится квартал, в котором был сделан заказ, а также подсчитана доля заказов, которые были доставлены с опозданием.

На основании полученных данных был построен график:



Таким образом, можно сделать вывод о том, что заказы второго класса доставляются с опозданием постоянно. В среднем около 20% заказов второго класса были доставлены с опозданием за рассмотренный период.

3.3. Выбор точки офлайн продаж

На данный момент в компании 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.

Запрос для поиска наиболее популярного по количеству доставок штата:

```
SELECT
    state,
    COUNT (order_id)
FROM
    sql.store_delivery
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1
```

Таким образом, наиболее популярный штат по количеству доставок - Калифорния (1021 заказ).

Запрос для поиска наиболее популярного по количеству доставок города:

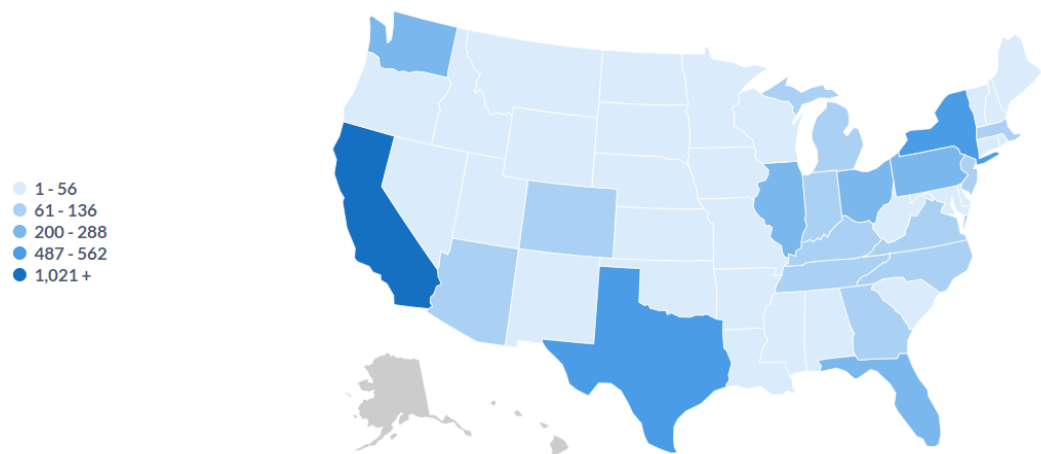
```
SELECT
    city,
    COUNT (order_id)
FROM
    sql.store_delivery
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1
```

Таким образом, наиболее популярный город по количеству доставок - Нью-Йорк (450 заказов).

Полученных данных недостаточно для выбора места для открытия магазина, поэтому был создан запрос для определения количества заказов по штатам.

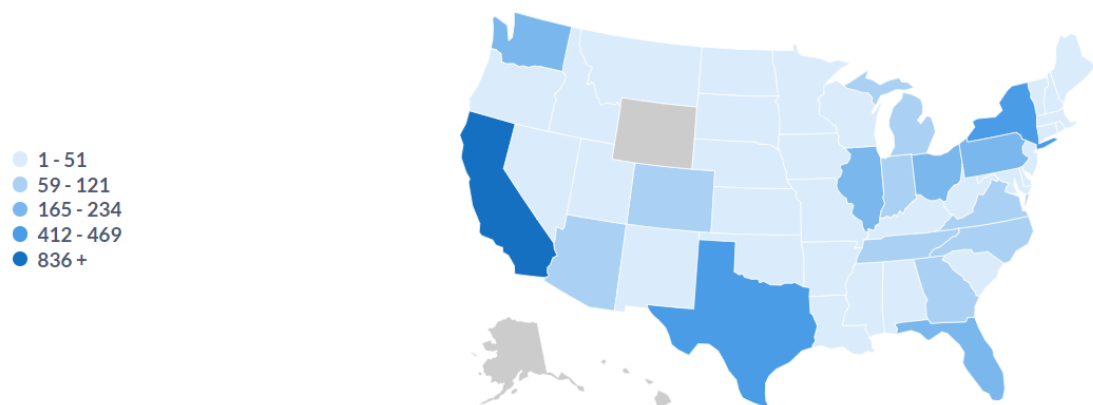
```
SELECT
    state,
    COUNT (order_id)
FROM
    sql.store_delivery
GROUP BY 1
```

На основе результатов запроса была создана визуализация для количества заказов по штатам.



Для корпоративных клиентов получаем похожие результаты:

```
SELECT
  state,
  COUNT (order_id)
FROM
  sql.store_delivery sd
  JOIN sql.store_customers sc ON sd.cust_id=sc.cust_id
WHERE sc.category='Corporate'
GROUP BY 1
```



Несмотря на то, что наибольшее количество доставок было выполнено в Калифорнии, на карте отображено, что количество заказов на Восточном побережье США больше, чем на Западном побережье, особенно у штата Нью-Йорк. Поэтому на основании проанализированных данных по количеству заказов целесообразно открывать магазин именно в штате Нью-Йорк.

Выводы

1. По данным в рассматриваемом периоде (январь 2017 года - декабрь 2020 года) выручка компании растет, при этом наблюдаются сезонность роста и падения выручки.
2. При оценке выручки по категориям было выявлено, что в топе продаж находятся кресла, а при анализе товаров было получено, что в лидерах продаж по объему выручки Canon imageCLASS 2200 Advanced Copier.
3. Наибольшее количество клиентов составляют корпоративные клиенты, а также объем выручки от заказов B2B-клиентов в разы превышает объем выручки от продаж розничным покупателям.
4. Из оценки динамики появления новых корпоративных клиентов было получено, что в компании был рост числа корпоративных клиентов до 2018 года, а затем пошел спад. Было рассчитано, что среднее количество различных товаров в заказах у корпоративных клиентов равно 2, средняя сумма заказа - 285,9, а среднее количество различных офисов у корпоративных клиентов составляет 6,2. Привлечение новых B2B-клиентов могло бы повысить выручку компании.
5. Была проанализирована эффективность выполнения текущих доставок, где было получено, что чаще всего с опозданием приходят заказы, отправленные вторым классом.
6. Наиболее перспективным для открытия офлайн магазина был выбран штат Нью-Йорк, так как город Нью-Йорк - лидер из городов по количеству доставок, а также близлежащие штаты на Восточном побережье США показывают большее суммарное количество заказов, чем штаты на Западном побережье. Так что с точки зрения логистики наиболее оптимально открытие магазина в штате Нью-Йорк.