

Учебный проект № 2. Решение бизнес-задач с помощью SQL

Введение

Заказчиком исследования выступает молодая компания Pens and Pencils (США). С марта 2017 года её профиль — B2B и B2C-продажи офисных товаров (оргтехника, канцтовары и мебель) онлайн.

Целью является анализ работы компании с точки зрения её эффективности и выработка рекомендаций по масштабированию бизнеса, а именно выбор штата для открытия офлайн-магазина. На момент проведения анализа компания имеет один склад в г. Хьюстон (штат Техас).

Задачи:

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).

Анализ проведен на основании данных о заказах клиентов за период 2017–2020 гг.

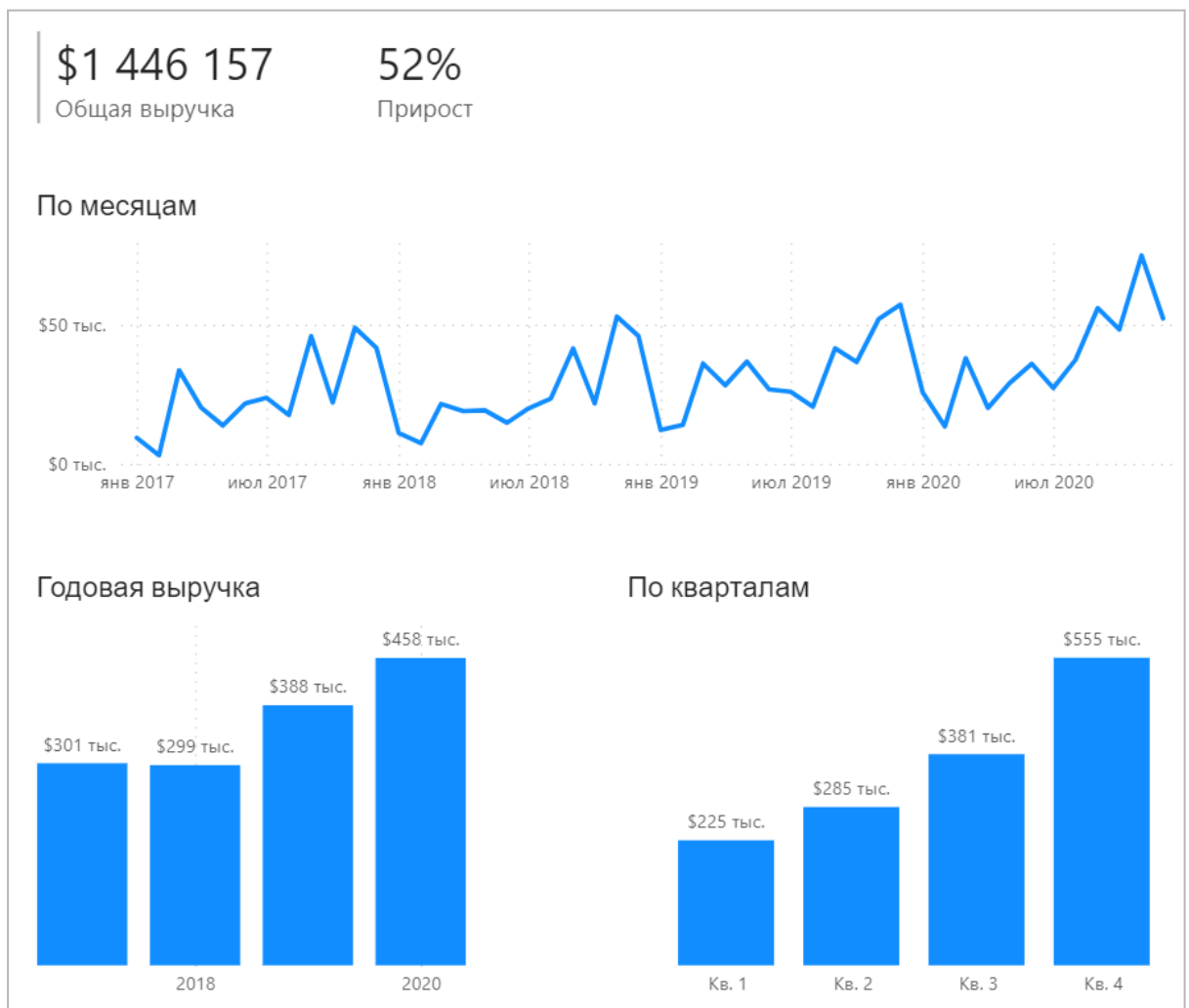
1. Динамика продаж и распределение выручки

Общая выручка компании за рассматриваемый период составила 1446 тыс. долларов.

В годовом выражении выручка стабильно растет с 2019 года (в 2018 фиксировалось небольшое ее снижение). В целом за четыре года выручка выросла на 52%

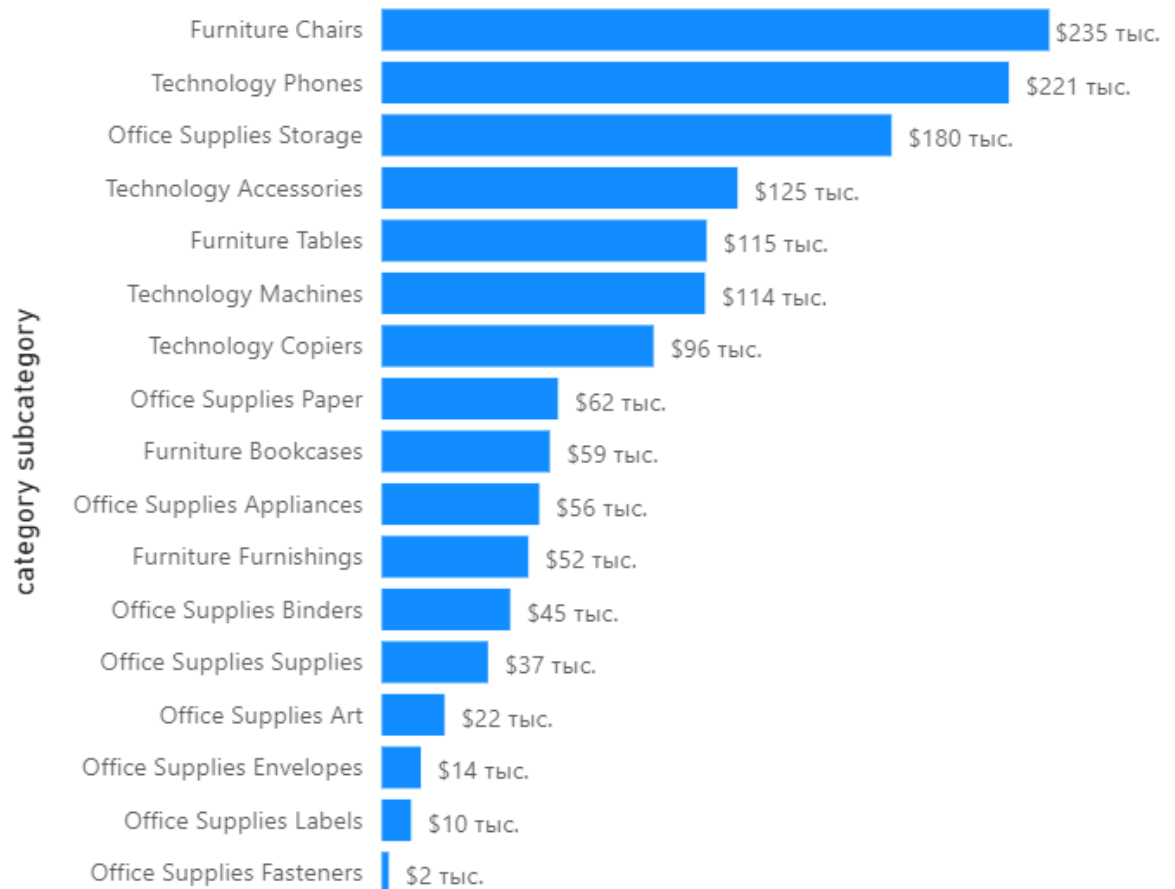
В течение года выручка распределена неравномерно: возрастая от начала к концу года и достигая максимума в четвертом квартале.

► [SQL-запрос](#)



Основной вклад в формирование выручки компании вносит продажа мебели, оргтехники и систем хранения. Офисные принадлежности (за исключением систем хранения) дают около 25% дохода.

Выручка по категориям товаров



ТОП-25 товаров по выручке в совокупности приносят 18% дохода.

Наименование товара	Выручка	Продано , шт.	Доля в общей выручке, %
Canon imageCLASS 2200 Advanced Copier	36959,82	20	2,56
Hon 5400 Series Task Chairs for Big and Tall	15309,53	39	1,06
3D Systems Cube Printer, 2nd Generation, Magenta	14299,89	11	0,99
High Speed Automatic Electric Letter Opener	13624,21	11	0,94
Martin Yale Chadless Opener Electric Letter Opener	13325,00	22	0,92
Cisco TelePresence System EX90 Videoconferencing Unit	11319,24	6	0,78
Hewlett Packard LaserJet 3310 Copier	11303,69	38	0,78
Samsung Galaxy Mega 6.3	11154,87	39	0,77
Canon PC1060 Personal Laser Copier	9295,83	19	0,64
HP Designjet T520 Inkjet Large Format Printer - 24" Color	9187,50	12	0,64
Global Troy Executive Leather Low-Back Tilter	9082,87	32	0,63

Наименование товара	Выручка	Продано , шт.	Доля в общей выручке, %
Honeywell Enviracaire Portable HEPA Air Cleaner for 17' x 22' Room	9043,55	40	0,63
Tennsco 6- and 18-Compartment Lockers	8740,17	43	0,60
Plantronics CS510 - Over-the-Head monaural Wireless Headset System	8657,89	35	0,60
SAFCO Arco Folding Chair	8100,95	53	0,56
Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	7805,48	24	0,54
Apple iPhone 5	7798,00	24	0,54
Tennsco Double-Tier Lockers	7704,86	47	0,53
Ativa V4110MDD Micro-Cut Shredder	7699,89	11	0,53
HON Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	7446,44	51	0,51
Tennsco Single-Tier Lockers	7326,59	28	0,51
Office Star - Professional Matrix Back Chair with 2-to-1 Synchro Tilt and Mesh Fabric Seat	7019,50	28	0,49
Zebra ZM400 Thermal Label Printer	6965,70	6	0,48
Hot File 7-Pocket, Floor Stand	6510,77	49	0,45
Bretford Rectangular Conference Table Tops	6497,82	46	0,45
Итог	262180,06	734	18,13

2. Характеристика клиентов

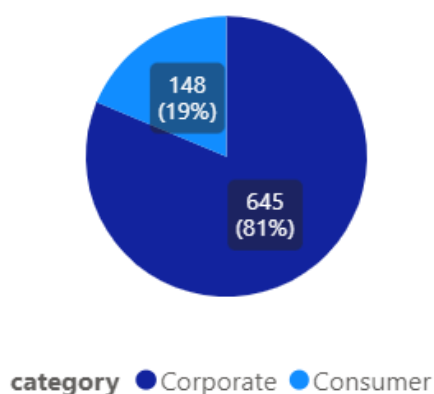
Основную массу клиентов (81%) составляют корпоративные клиенты. они же приносят 81% выручки.

► [SQL-Запрос](#)

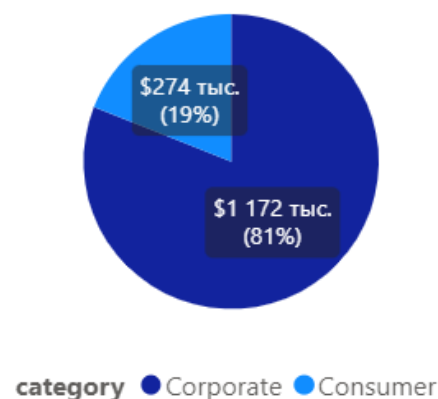
Компания активно привлекала новых клиентов в 2017 году, однако с 2018 данный показатель значительно упал и продолжает уменьшаться.

► [SQL-запрос](#)

Количество клиентов



Сумма выручки



Новые клиенты



В среднем корпоративные клиенты единоразово заказывают 2 наименования общим числом 7,6 единиц товара на сумму 285,9 долл. Среднее число офисов (адресов доставки) у корпоративных клиентов – 6,2.

► [SQL-запрос](#)

Категория клиента	Corporate	CONsumer
Заказов на 1 клиента, шт.	6,4	6,1
Среднее число позиции в заказе	2,0	2,0
Среднее число товаров в заказе, шт.	7,6	7,4

Средний чек, долл.	285,9	301,6
Среднее число адресов доставки на 1 клиента	6,2	6,0

Чтобы интерпретировать результаты целесообразно сравнить их со средними по данному сегменту рынка. К сожалению, в доступных источниках информации о торговле канцтоварами в США найти не удалось. По данным портала IRP Commerce ([Ecommerce Market Data and Ecommerce Benchmarks for the Electrical & Commercial Equipment Market for December 2020](#)) средний чек на декабрь 2020 в сегменте “Electrical & Commercial Equipment” составил около 59£ или с учетом обменного курса \$73. Можно сказать, что средний чек в анализируемой компании довольно высок. В то же время, за период 2017–2020 выручка в данном сегменте рынка выросла на 220%, т.е. в нашей компании выручка растет недостаточно быстро, что, возможно, связано с недостаточным притоком новых клиентов.

Дополнительно: по рассмотренным выше показателям корпоративные клиенты мало отличаются от частных, предпочтения клиентов по категориям товаров также практически совпадают. Из чего можно сделать вывод, что сегментация по категориям клиентов мало информативна.

► [SQL-запрос](#)

Категория товара	Подкатегория товара	Продано клиентам, шт.		Популярность (место среди других категорий товара)	
		corporate	consumer	corporate	consumer
Office Supplies	Binders	1 079	237	1	1
Office Supplies	Paper	967	224	2	2
Furniture	Furnishings	713	164	3	3
Technology	Phones	663	151	4	4
Office Supplies	Storage	648	129	5	7
Office Supplies	Art	593	138	6	5
Technology	Accessories	582	136	7	6
Furniture	Chairs	478	98	8	8
Office Supplies	Appliances	384	67	9	9

Office Supplies	Labels	280	66	10	10
Furniture	Tables	256	51	11	11
Office Supplies	Envelopes	209	40	12	13
Furniture	Bookcases	192	32	13	14
Office Supplies	Fasteners	171	44	14	12
Office Supplies	Supplies	158	29	15	15
Technology	Machines	90	22	16	16
Technology	Copiers	55	13	17	17

3. Логистика

Более половины заказов оформляются со стандартной доставкой (в течение 6 дней). В абсолютном выражении наибольшее число заказов, доставленных с опозданием, также приходится на стандартную доставку. Вместе с тем, в относительном выражении наибольший процент опозданий приходится на доставку вторым классом.

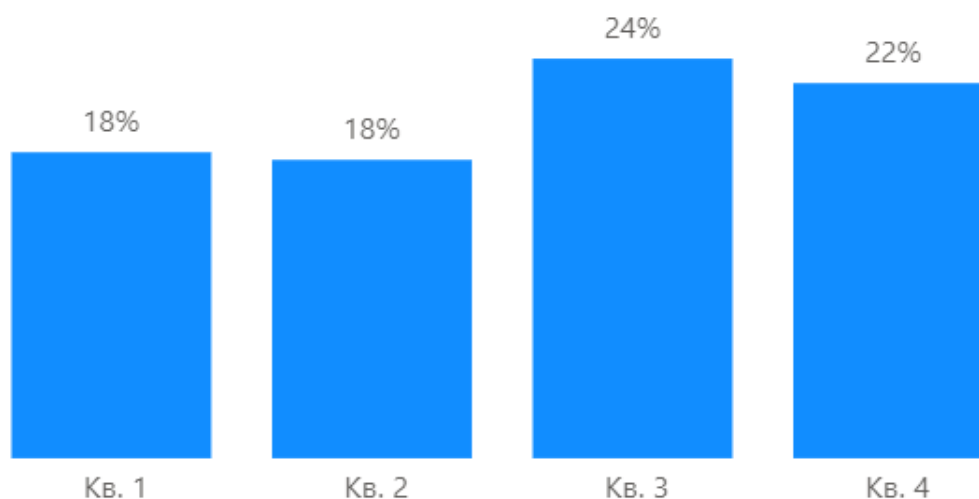
► [SQL-запрос](#)

Класс доставки	Заказов, шт.	Доставлено с опозданием, шт.	Доставлено вовремя, %
Second Class	964	202	79,05%
Standard Class	2994	309	89,68%
Same Day	264	12	95,45%
First Class	787	1	99,87%
Всего	5009	524	89,54%

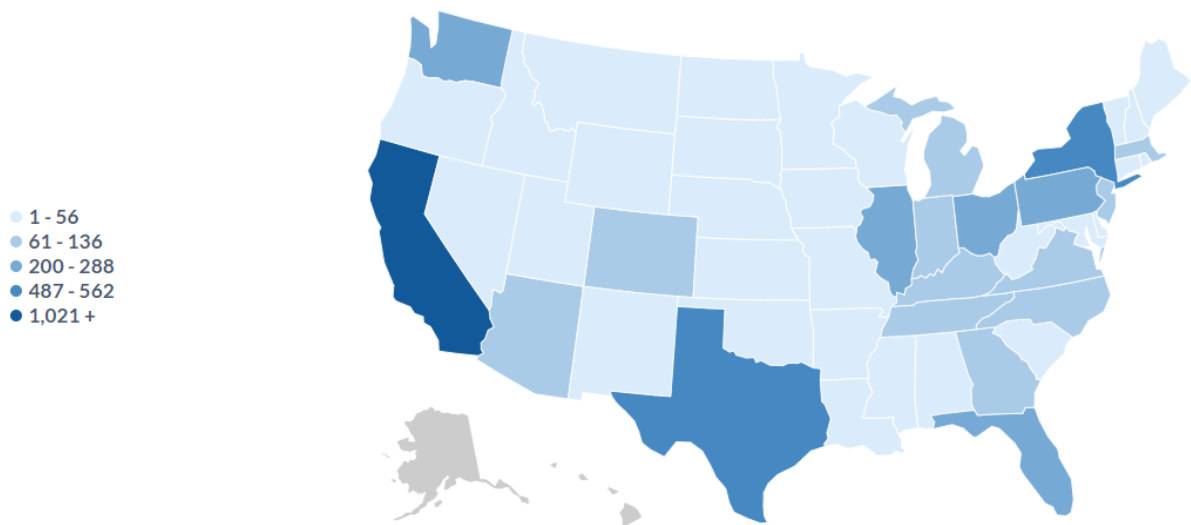
Среди доставок вторым классом опоздания чаще происходят в третьем и четвертом кварталах.

► [SQL-запрос](#)

Доля опозданий по кварталам



Наибольшее число заказов (20%) доставляется в Калифорнию. Следующий по числу доставок штат – Нью-Йорк – отстает по данному показателю почти в 2 раза (11%). Из городов на первом месте по числу заказов находится г. Нью-Йорк.



При выборе места для оффлайн-магазина следовало бы оценить условия организации доставки (оптимальный с точки зрения цены объем груза), затраты на доставку в различные штаты, стоимость аренды помещения, причины задержек при доставке и другие факторы. При отсутствии этих данных можно предложить открыть магазин в Калифорнии, т. к. большинство заказов приходится на этот штат.

Приложение. SQL-запросы

27.5.1 - сумма выручки по месяцам

Запрос выдает суммарную стоимость заказов с учетом скидки с группировкой по месяцу оформления заказа.

```
SELECT
    date_trunc('month',order_date)::date AS date,
    round(sum(quantity*price*(1-discount))) AS revenue
FROM
    sql.store_delivery AS dlv
    JOIN sql.store_carts AS crt ON dlv.order_id = crt.order_id
    JOIN sql.store_products AS prd
        ON crt.product_id = prd.product_id
GROUP BY 1
ORDER BY 1
```

27.5.5 - сумма выручки по категориям и подкатегориям товаров

Запрос выдает суммарную стоимость заказов с учетом скидки с группировкой по категории и подкатегории товара.

```
SELECT
    category,
    subcategory,
    round(sum(quantity*price*(1-discount))) AS revenue
FROM
    sql.store_products AS prd
    LEFT JOIN sql.store_carts AS crt
        ON crt.product_id = prd.product_id
GROUP BY
    category,
    subcategory
ORDER BY
    revenue DESC
```

27.5.6 - ТОП-25 товаров по объёму выручки

Запрос выдает наименования 25 товаров, принесших наибольшую выручку за период 2017-2020, количество проданных единиц товара (поле quantity) и их вклад в процентах в суммарную выручку (поле percent_FROM_total).

```
WITH products_revenue AS (
    SELECT
        product_nm,
        sum(quantity*price*(1-discount)) AS revenue,
        sum(quantity) AS quantity
    FROM
```

```

    sql.store_products AS prd
    LEFT JOIN sql.store_carts AS crt
      ON crt.product_id = prd.product_id
GROUP BY
  product_nm
ORDER BY
  revenue DESC
)

SELECT
  product_nm,
  round(revenue,2) AS revenue,
  quantity,
  round(100*revenue/sum(revenue) OVER (),2) AS percent_from_total
FROM
  products_revenue
GROUP BY
  product_nm,
  revenue,
  quantity
ORDER BY
  revenue DESC
LIMIT 25

```

27.6.1 - количество клиентов и выручка по категориям клиента

Запрос выдает количество клиентов (поле cust_cnt) и суммарную стоимость их заказов с учетом скидки для двух категорий клиентов (поле revenue).

```

SELECT
  cst.category,
  count(DISTINCT cst.cust_id) AS cust_cnt,
  round(sum(quantity*price*(1-discount))) AS revenue
FROM
  sql.store_customers AS cst
  LEFT JOIN sql.store_delivery AS dlv ON cst.cust_id = dlv.cust_id
  JOIN sql.store_carts AS crt ON dlv.order_id = crt.order_id
  JOIN sql.store_products AS prd ON crt.product_id =
prd.product_id
GROUP BY
  cst.category
ORDER BY
  revenue DESC

```

27.6.2 - количество новых корпоративных клиентов по месяцам

Запрос выдает число корпоративных клиентов, совершивших первую покупку, в каждом месяце.

```
with first_orders AS (
/* месяц совершения первого заказа для каждого клиента */

    SELECT
        cust_id,
        min(date_trunc('month',order_date))
        OVER(PARTITION BY cust_id)::date AS start
    FROM
        sql.store_delivery
    GROUP BY
        cust_id,
        order_date
)

/*
число корпоративных клиентов, совершивших первую покупку,
в каждом месяце
*/
SELECT
    start AS month,
    count(DISTINCT ord.cust_id) AS new_custs
FROM
    first_orders AS ord
    JOIN sql.store_customers AS cyst ON cust.cust_id = ord.cust_id
WHERE
    category ~* 'corp'
GROUP BY
    month
ORDER BY
    month
```

27.6.4 – поведение клиентов

Запрос выдает усредненные данные о заказах для двух категорий клиентов:

- среднее число заказов на одного клиента;
- среднее число наименований товаров в заказе;
- среднее количество единиц товара в заказе;
- среднюю стоимость заказа;
- среднее количество различных адресов доставки на одного клиента.

```
with orders AS (
/*
Обобщенная информация о заказе:
```

```

- количество позиций (различных наименований товаров) в заказе,
- количество заказанных единиц товара в заказе,
- сумма заказа с учетом скидки,
- клиент
- место доставки
*/
SELECT
    dlv.order_id,
    dlv.cust_id,
    dlv.zip_code,
    count(DISTINCT product_id) AS positions,
    sum(quantity) AS quantity,
    sum(price*quantity*(1-discount)) AS cost
FROM
    sql.store_delivery AS dlv
    NATURAL JOIN sql.store_carts AS ctr
    NATURAL JOIN sql.store_products AS prod
GROUP BY
    dlv.order_id,
    dlv.cust_id,
    dlv.zip_code
)

```

```

SELECT
    category,
    round(count(DISTINCT order_id)::numeric/
        count(DISTINCT cust_id),1) AS ord_user,
    round(sum(positions)::numeric/count(DISTINCT order_id),1)
        AS avg_pos,
    round(sum(quantity)::numeric/count(DISTINCT order_id),1)
        AS avg_q,
    round(sum(cost)::numeric/count(DISTINCT order_id),1) AS aov,
    round(1.0*count(DISTINCT cust_id || zip_code)/
        count(DISTINCT cust_id),1) AS offices_cnt
FROM
    sql.store_customers AS cust
    NATURAL JOIN orders
GROUP BY
    category

```

27.6.4 (дополнительно) – предпочтения клиентов по категориям товаров

Запрос выдает перечень категорий и подкатегорий товаров, число единиц товаров данной подкатегории, проданных частным и корпоративным клиентам, а также место данной подкатегории товаров по числу продаж.

```
WITH favs AS (
    SELECT
        prod.category,
        prod.subcategory,
        count(DISTINCT dlvs.order_id) FILTER (WHERE cust.category ~*
'corp') AS corp,
        count(DISTINCT dlvs.order_id) FILTER (WHERE cust.category ~*
'cons') AS cons

    FROM
        sql.store_products AS prod
        LEFT JOIN sql.store_carts AS cart
            ON prod.product_id = cart.product_id
        JOIN sql.store_delivery AS dlvs
            ON dlvs.order_id = cart.order_id
        JOIN sql.store_customers AS cust
            ON cust.cust_id = dlvs.cust_id
    GROUP BY
        prod.category,
        prod.subcategory
)

SELECT
    category,
    subcategory,
    corp,
    cons,
    dense_rank() OVER(ORDER BY corp DESC) rank_corp,
    dense_rank() OVER(ORDER BY cons DESC) rank_cons
FROM favs
ORDER BY
    3 DESC, 4 DESC
```

27.7.1 - опоздания

Запрос выдает общее количество заказов (orders_cnt), количество заказов, доставленных с опозданием (late_orders_cnt) и процент заказов, доставленных вовремя (% success), с группировкой по категориям доставки.

```
with delivery_with_status AS (
    /*
```

```

сведения о доставках с указанием статуса:
- вовремя ('on time') или
- с опозданием ('late')
*/
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
)

SELECT
    ship_mode,
    count(order_id) AS orders_cnt,
    count(order_id) FILTER(WHERE status = 'late')
    AS late_orders_cnt,
    round(100.0*count(order_id) FILTER(WHERE status <>
        'late')/count(order_id),2) AS "% success"
FROM
    delivery_with_status
GROUP BY
    ship_mode
ORDER BY
    4

```

27.7.2 – доля заказов, отправленных вторым классом, которые были доставлены с опозданием, по кварталам

```

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
)

SELECT
    extract(quarter from order_date) AS quarter,
    100.0*count(order_id) FILTER(WHERE status = 'late')/
        count(order_id) AS late_percent
FROM
    delivery_with_status
WHERE
    ship_mode ~* 'second'
GROUP BY
    quarter
```

27.7.6 - количество доставок по штатам

```
SELECT
    state,
    count(DISTINCT order_id) AS orders_count
FROM
    sql.store_delivery
GROUP BY
    state
ORDER BY
    orders_count DESC
```