

Тестовое задание

Задание 1. Ключевые метрики на каждый день

Для начала надо понять на какой стадии развития находится СберМаркет и какие главные цели.

По итогам июня 2021 СберМаркет уже есть во всех городах России.

Активная территориальная экспансия закончилась, начинается внутреннее усиление.

Если сначала СберМаркет ориентировался на одну схему «плановая закупка», то сейчас стоит задача увеличить ТО за счет:

- увеличение доли рынка в направлении экспресс-доставок
- проработка дарк-сторов, отслеживание эффективности и быстрое масштабирование при необходимости - залог повышения устойчивости бизнеса и возможность повысить маржу
- развитие самовывоза
- развитие быстрой доставки
- усиление работы с партнерами

По данным вводным я бы предложила следующие метрики:

- Количество пришедших заказов в разбивке по скорости доставки: плановая, экспресс-доставка, самовывоз
- Средний чек
- Количество новых уникальных клиентов
- Конверсия из захода на сайт/приложение в покупку. Каждый день эта конверсия должна быть примерно одинаковая, иначе можно судить по сбоям в системе.

Также не являются метриками, но очень важно отслеживать:

- Работоспособность приложения по маршрутизации водителей
- Работоспособность приложения для сборки и комплектации заказов
- Доступность количества курьеров / сборщиков в данный момент

Тестовое задание

Задание 2. Анализ среза транзакций

Разделы:

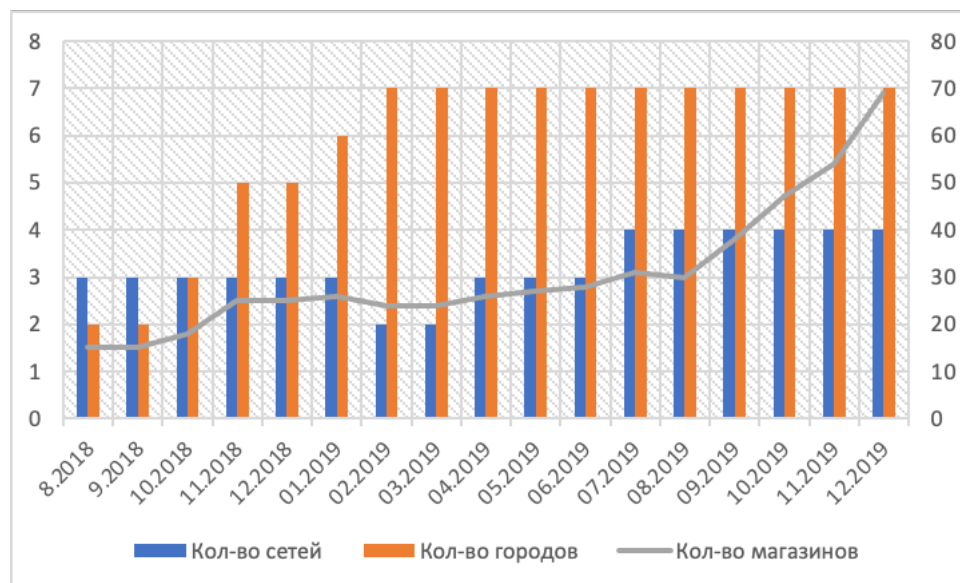
1. Активный рост. Общий срез
2. Анализ транзакций в разрезе по ритейлерам
3. Динамика по городам сети, показавшей падение
4. Гипотеза
5. Срез по магазинам
6. Уровень удержания клиентов
7. Найденные аномалии

1. Активный рост. Общий срез

Анализируемый период в срезе: 26.08.2018 - 18.12.2019.

На данный период в компании приходился активный рост.

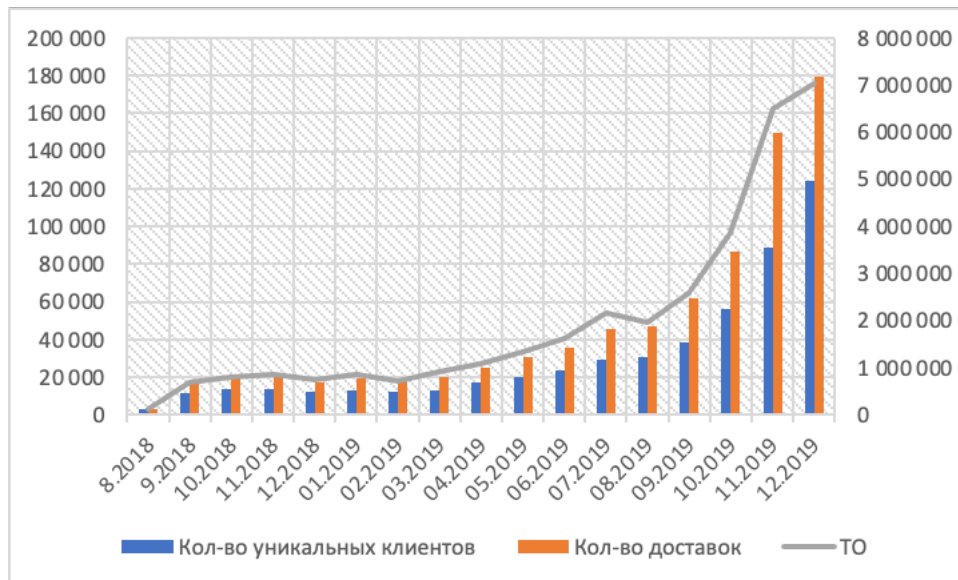
Задача данного периода: активно наращивать количество доставок за счет увеличения количества городов и количества ритейлеров.



Количество городов каждый месяц растет: от 2х городов в августе 2018 до 7 городов в декабре 2019.

Количество сетей растет незначительно с падением в период с февраля по март 2019. Более детальный срез показал, что в период с февраля 2019 по июнь 2019 перестали приходить заказы от одной из первых сетей.

Количество магазинов активно растет. Периоды падения: февраль 2019 - март 2019 - выбыло 2 магазина. Так же количество магазинов упало в июле 2019 (-1 магазин). Но в целом, по данному направлению динамика положительная.



На данном графике я вывела количество уникальных клиентов, количество доставок и ТО (сумма по полю orders.item_total).

Так как изначально данные за декабрь были только по 18.12.2019 у меня было 2 варианта: убрать декабрь 2019 из анализа (что не хотелось) или спрогнозировать показатели декабря на основании данных 18 дней. Я решила сделать прогноз.

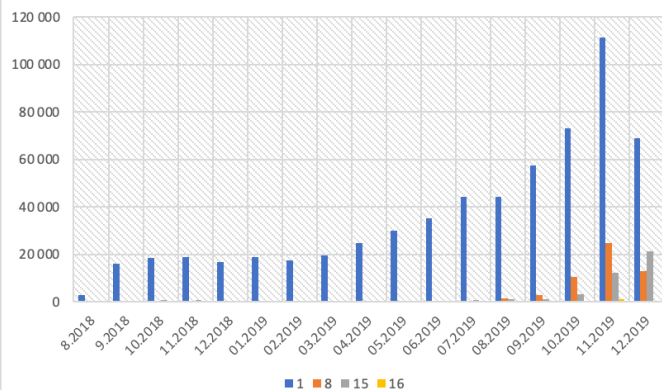
Выводы:

- Все показатели активно растут.
- Падение по всем показателям зафиксировано в **декабре 2018** и в **феврале 2019**.
- В августе 2019 падает ТО при положительной динамике по клиентам и доставкам.

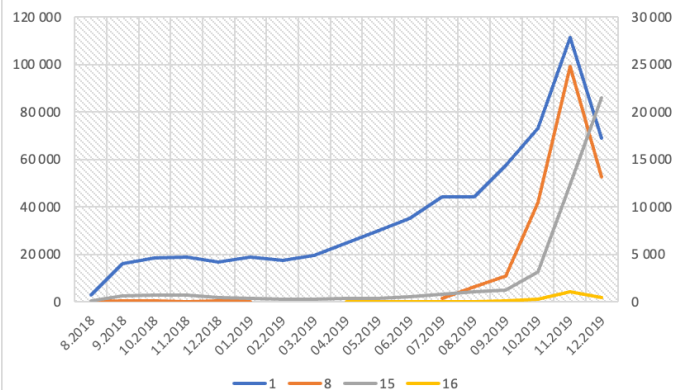
2. Анализ транзакций в разрезе по ритейлерам

Теперь я хочу проанализировать показатели роста между разными ритейлерами. На втором графике ритейлер № 1 относится к основной оси Y (слева), остальные ритейлеры с гораздо меньшим оборотом относятся к вспомогательной оси Y (слева).

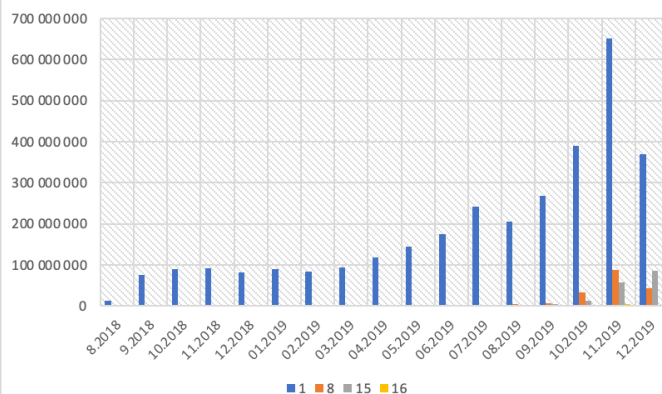
Динамика заказов по разным ритейлерам



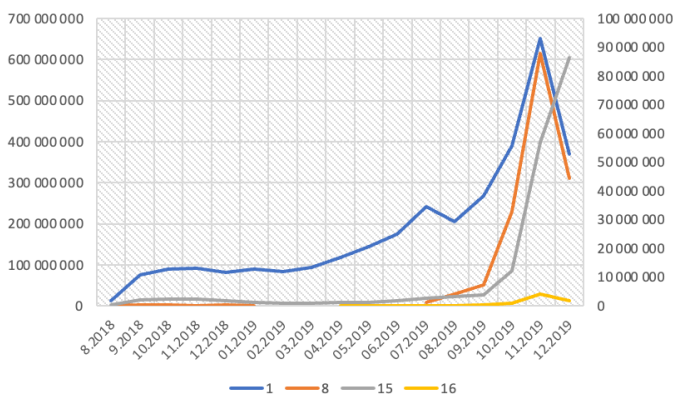
Динамика заказов по разным ритейлерам



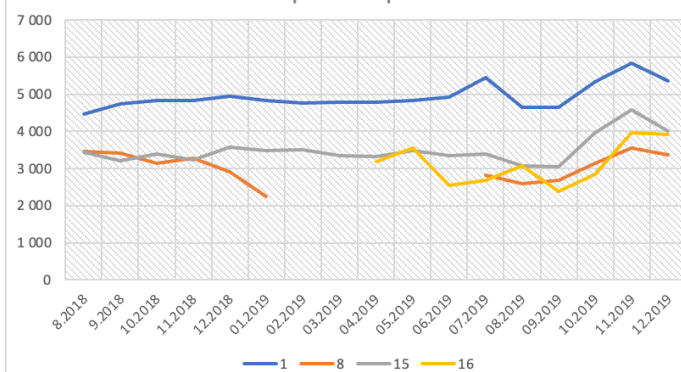
Динамика ТО по разным ритейлерам



Динамика ТО по разным ритейлерам



Динамика среднего чека по разным ритейлерам



Выводы:**Доставки:**

- Падение в декабре 2018 произошло в двух сетях (№1, №15), ритейлер №8 показал в декабре хороший рост.
- Сеть №15 продолжала падение с ноября 2018 по апрель 2019 (в то время как сеть №1 в марте 2019 стала расти)
- Сеть №8 выбыла из доставок на 5 месяцев, вернулась в июле 2019, показав хорошие показатели.
- Сеть №16 была подключена в апреле 2019, но вплоть до ноября не показывала хороших показателей.
- В декабре 2019 сеть №15 показала рост относительно ноября почти в 2 раза, в то время как другие сети показывали в декабре падение.

ТО: динамика по сетям примерно соответствует картине по динамике доставок.

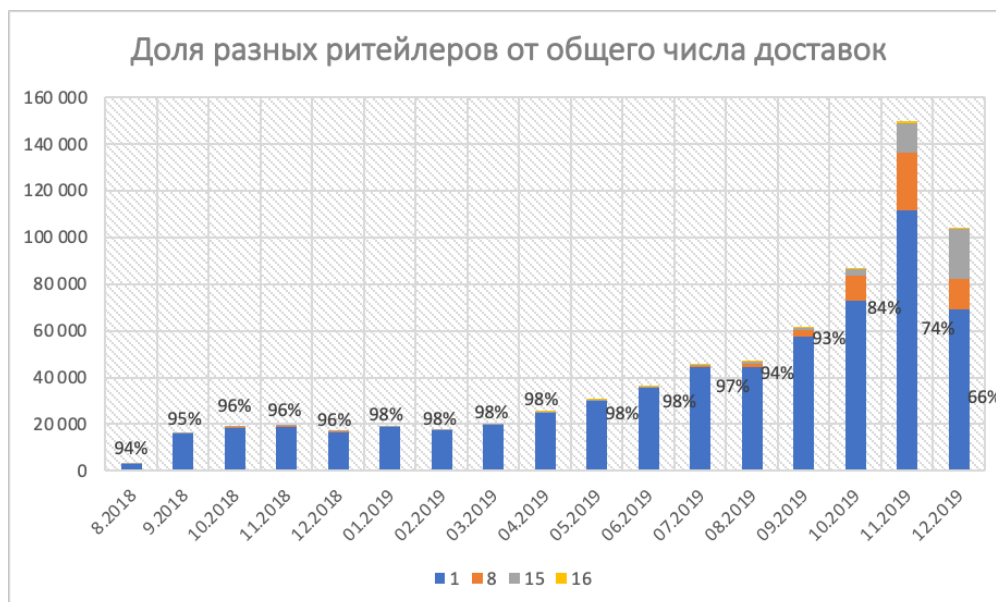
Средний чек:

- самый большой средний чек у сети №1
- самый маленький средний чек у сети №8
- самый большой средний чек зафиксирован у всех сетей в ноябре 2019
- в июле 2019 в сети №1 был резкий скачок по среднему чеку

Доля каждого из ритейлеров в общем количестве доставок

Ритейлер №1 вплоть до сентября 2019 занимал долю от всех доставок 92% - что является для бизнеса очень неустойчивой позицией.

И только в ноябре доля ритейлера №1 стала снижаться за счет роста других сетей.



3. Динамика по городам ритейлеров, показавшим падение

Ритейлер № 1

Как видно из таблицы, проблема в ноябре и декабре оказалась в городе № 1. Это самый крупный город с точки зрения продаж.

Город	8.2018	9.2018	10.2018	11.2018	12.2018
Город 1	2 881	15 121	17 383	17 167	14 280
Город 2	177	835	906	914	972
Город 5			283	475	582
Город 7				449	580
Город 8				19	235

Ритейлер № 15

В этом ритейлере продажи упали тоже в городе 1.

Город	8.2018	9.2018	10.2018	11.2018	12.2018
Город 1	150	661	761	621	397
Город 2				104	93

4. Гипотеза

Из анализа 2018 имеем следующую проблему: в городе №1, сеть №1 в декабре 2018 на 12% упали продажи относительно ноября.

Это был самый крупный город с точки зрения продаж.

И самый крупный ритейл по сравнению с другими.

Падение на 12% сильно повлияло на всю компанию.

Что могло привести к этому падению:

- увеличение стоимости доставки
- уменьшение доли скидки от общей покупательской корзины
- ухудшение сервиса доставки
- отсутствие курьеров (отсутствие свободных слотов времени для заказа доставки)

В своей работе я решила проверить такую гипотезу:

Увеличение времени ожидания в октябре- ноябре привело к оттоку постоянных клиентов в декабре.

Проверка гипотезы

Какие поля я использовала для анализа:

orders.id

orders.shipped_at

orders.delivery_window_id

delivery_windows.id

delivery_windows.ends_at

Мои шаги:

1. К каждому заказу в таблице orders я соотнесла время окончания доставки delivery_windows.ends_at по имеющемуся orders.delivery_window_id.
2. Сравнила фактическое время доставки с временем окончания слота доставки.
3. Получила количество минут, за которое заказ был доставлен до окончания выбранного слота времени (это значения со знаком +), а так же получила количество минут, через сколько после окончания выбранного слота времени был доставлен заказ (это значения идут со знаком минус).
4. Сгруппировала полученные значения по 60 минут.

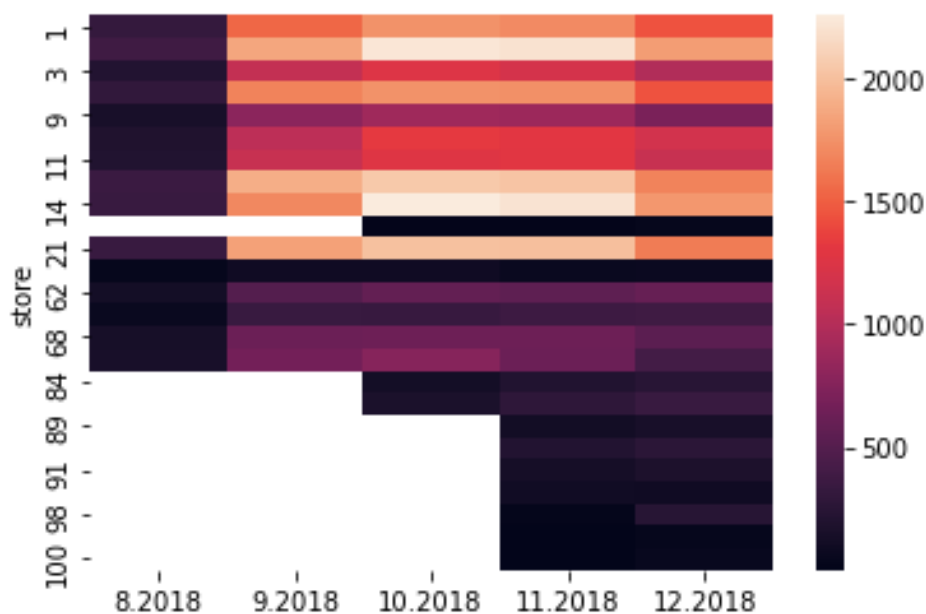
Количество минут	8.2018	9.2018	10.2018	11.2018	12.2018
<-240	43	412	224	347	400
-240--180	7	20	22	21	16
-180--120	17	60	51	34	80
-120--60	78	248	215	212	352
-60-0	262	1 346	1 314	1 296	1 590
0-60	1 116	5 350	5 812	5 526	5 096
60-120	980	5 410	6 678	5 967	4 753
120-180	227	1 541	2 157	2 642	1 435
>180	151	734	910	1 122	558

Вывод:

в таблице я не увидела каких-то запредельных отклонений по доставке.
Поэтому моя гипотеза не подтвердилась.

5. Срез по магазинам

Падение в декабре идет почти по всем старым магазинам, в среднем на -17%. Но надо отметить, что тренд к падению начался в ноябре.



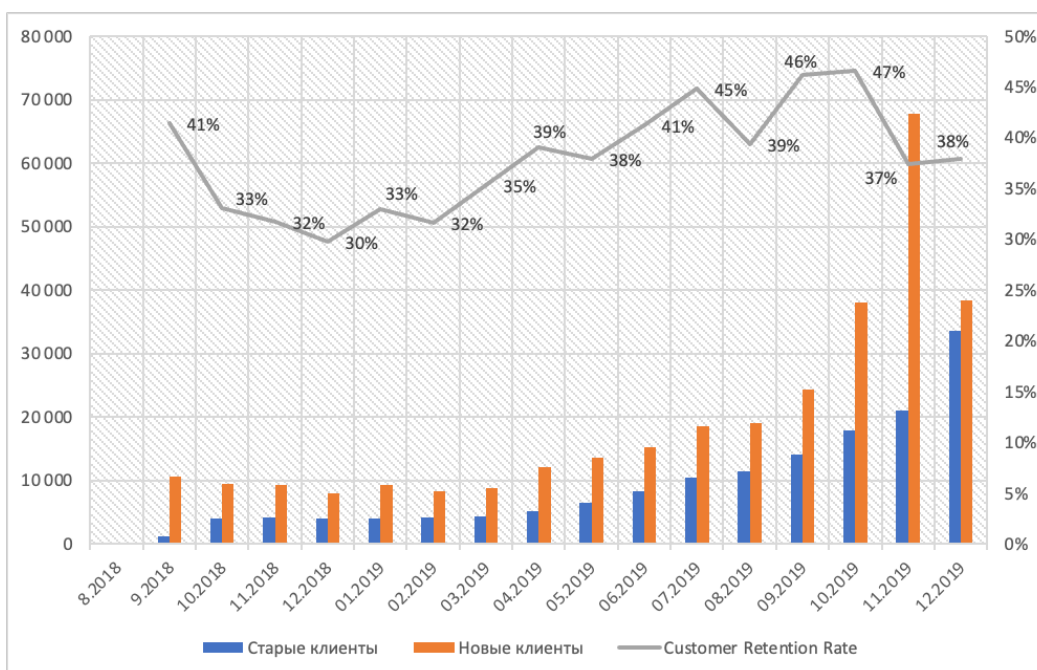
store	10.2018	11.2018	12.2018	Δ 11/10	Δ 12/11
1	1743	1 712	1 440	-2%	-16%
2	2221	2 193	1 806	-1%	-18%
3	1253	1 186	984	-5%	-17%
8	1745	1 741	1 440	0%	-17%
9	894	873	701	-2%	-20%
10	1322	1 302	1 169	-2%	-10%
11	1266	1 291	1 101	2%	-15%
12	2046	2 022	1 666	-1%	-18%
14	2253	2 208	1 774	-2%	-20%
18	12	13	34	8%	162%
21	1999	1 993	1 631	0%	-18%
57	93	58	67	-38%	16%
62	573	552	597	-4%	8%
63	333	362	375	9%	4%
68	629	633	534	1%	-16%
70	761	621	397	-18%	-36%
84	123	197	246	60%	25%
85	160	278	336	74%	21%
89		113	143		27%
90		211	262		24%
91		125	175		40%
92		104	93		-11%
98		19	235		1137%
99		2	36		1700%
100		1	45		4400%

6. Уровень удержания клиентов



Одним из важных показателей бизнеса является - **уровень удержания клиентов**. По итогам 2019 г CRR (Customer Retention Rate) = 39%.

Если смотреть в динамике по месяцам, то уровень удержания клиентов увеличился во втором полугодии 2019 и достиг максимального значения в сентябре и октябре 2019.



Факторы, влияющие на удовлетворенность клиента:

- свободные слоты времени соответствуют желанию клиента
- курьер привозит товар вовремя, не позднее указанного при заказе времени
- выбранный курьером товар качественный (не просрочен, нет повреждения упаковки)
- курьер при необходимости правильно подбирает замену товаров (угадывает/ согласовывает предпочтения клиентов)

7. Найденные аномалии

По условиям СберМаркета, можно выбрать доставку максимум через 7 дней от даты заказа. Я решила сравнить даты заказа и доставки и выявила довольно много несоответствий.

1. 3 записи подряд были доставлены раньше даты заказа (id заказа 980548, 980564, 980590). По этим записям я проверила окно доставки. Окончание окна доставки пришлось на день оформления заказа. Я сделала вывод, что в системе неправильно отразилось время доставки (shipped_at) и сделала замену в этих ячейках на конечное окно доставки.
2. Сравнила store_ID из таблицы orders и store_id из таблицы delivery_windows (по ключу orders.delivery_window_id) . Обнаружила, что в delivery_windows отсутствуют такие окна доставки как 973326, 1013194, 1242161, 1242161. Вопрос: куда они могли деться, и где найти для них соответствие?
3. Сравнила **дату доставки и конечное время в окне доставки**. Допустила, что оптимальное время доставки: 120 мин до окончания выбранного слота доставки и 120 мин после окончания выбранного слота доставки. Из всех доставок (720 434 штуки) всего 630 288 шт подошло под такую выборку. То есть **90 тыс заказов было доставлено или позже, чем на 2 часа** после окончания выбранного слота, или более чем на 2 часа раньше окончания выбранного слота. На мой взгляд, это очень большое расхождение.
4. Проанализировала **total_cost** (финальная стоимость доставки) и **cost** (начальная стоимость доставки):
 - 3 доставки: стоимость доставки стояла со знаком минус.
 - По ряду заказов были в значении стоимости доставки дробные числа. Поменяла знак точки на запятую, чтобы можно было их сравнить друг с другом.
 - 27 заказов: в стоимости была указана дата. Скорее всего ошибка

Стоимость доставки	2018	2019
0-50	31 053	122 218
50-100	483	104 325
100-150	1 519	28 637
150-200	2 430	206 461
200-250	7 476	30 782
250-300	23 714	88 568
>300	9 795	62 968

Стоимость доставки и стоимость заказа:

916 заказов было отгружено с стоимостью доставки больше чем стоимость всех товаров в заказе.

1211 заказов было с стоимостью доставки в 2 р меньше чем стоимость заказа.

1311 заказов со стоимостью доставки больше 1000 рублей.

70 заказов было с весом больше 1 т кг (если вес корзины указан в граммах)

Статусы заказов

720 404 заказа было доставлено (статус shipped), это очень хороший показатель.