

# Анализ компании и поиск инсайтов в данных.

Выполнил:  
Запсельский Виктор

## Содержание проекта:

**1.0**

**Анализ компании. Начало**

Условие. Работа с метриками. Оценка времени выполнения заказа.

**1.1**

**Анализ компании. Сегментация**

Сегментация пользователей: RFM анализ

**1.2**

**Анализ компании. Retention.**

Retention пользователей : когортный анализ

**1.3**

**Анализ компании.ТОП**

ТОП магазинов по кол-ву отмененных / замененных товаров

**1.4**

**Анализ компании. Выводы.**

Выводы: инсайты, закономерности.

# 1.0 Анализ компании. Начало.

Условие. Работа с  
метриками. Оценка  
времени  
выполнения заказа.

## 2.0 Решение второй части задания.

### Условие.

#### Условие:

Основная задача аналитиков в нашей компании - растить бизнес. Иногда мы сами не знаем где прячутся эти точки кратного роста, поэтому порой приходится кранчить данные и искать зависимости/аномалии, генерировать гипотезы в процессе и потом предлагать проекты. Мы дадим доступ к базе со срезом транзакций.

Твое задание: найти все интересные на твой взгляд инсайты и представь их в любом удобном виде (ноутбук желательно приложить)

Ссылка на ноутбук (collab):

[Link](#)

## 2.0 Решение второй части задания. Первичный анализ данных.

Рассмотрим следующие метрики и посмотрим их ежедневное изменение. Из такого анализа можно сделать выводы о "здоровье" компании, в какой стадии она находится, а также найти интересные закономерности.

### Метрики:

**GMV** : чем выше данный показатель, тем активнее клиенты пользуются нашим сайтом

**AOV**: данная метрика позволяет контролировать как изменяется средний чек пользователя

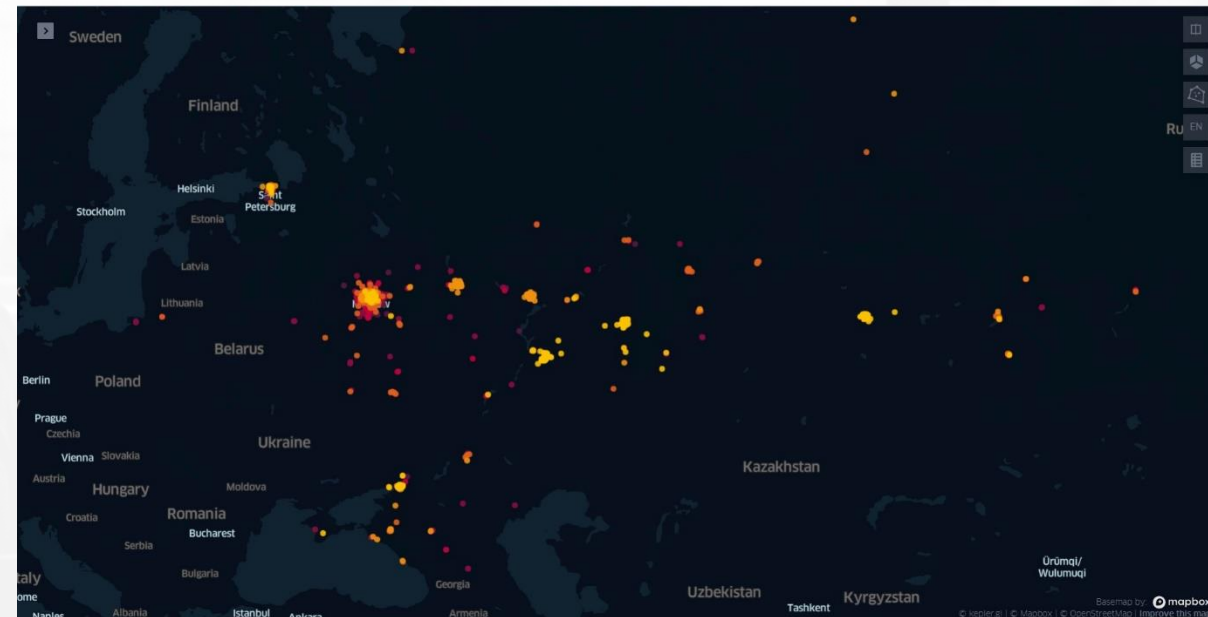
**Count\_orders** ( общее кол-во заказов за определенное время)  
Больше заказов -> больше доставок -> больше прибыль.

**Total\_cost** (суммарная сумма доставок за опред.период): так как доставки – основной инструмент заработка, то стоит отслеживать их изменения.

**Count\_of\_replaced** (кол-во замененных товаров за опред. период)

**Count\_of\_cancelled** (кол-во отмененных товаров за опред. период)

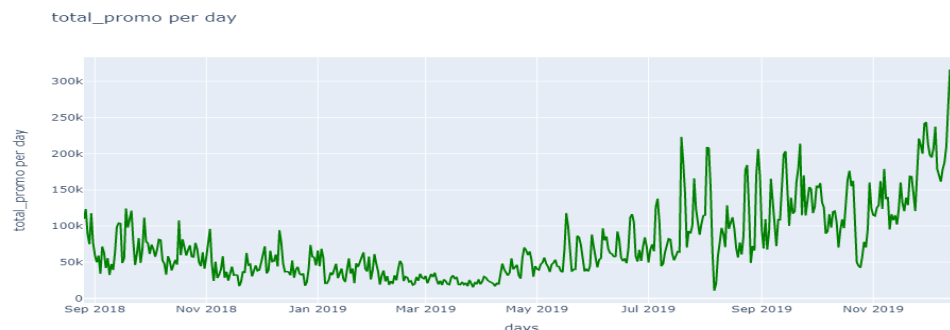
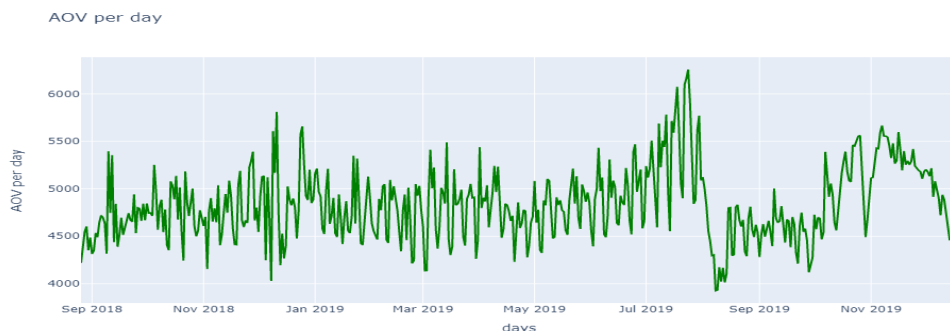
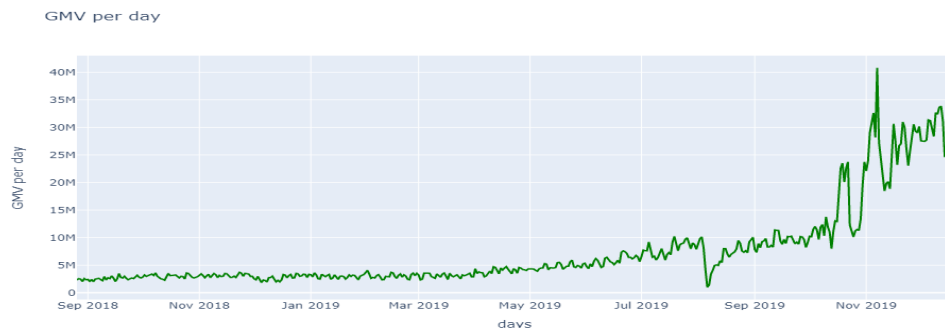
Сделаем визуализацию адресов доставки, посмотрим куда пользователи делают заказы:



Видно, что заказывают люди по всей стране и даже в Калининград, а вот магазины у нас только в 7 городах...

## 2.0 Решение второй части задания. Работа с метриками. Анализ метрик.

### Графики:



### Выводы:

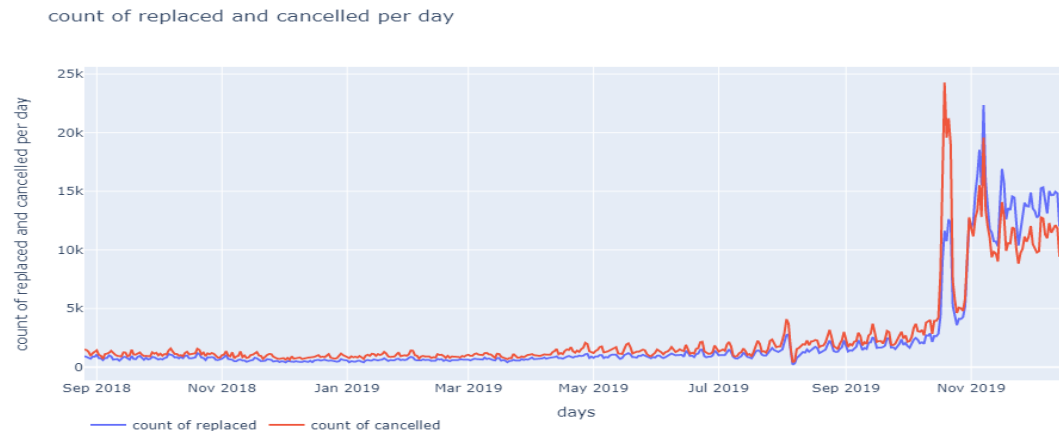
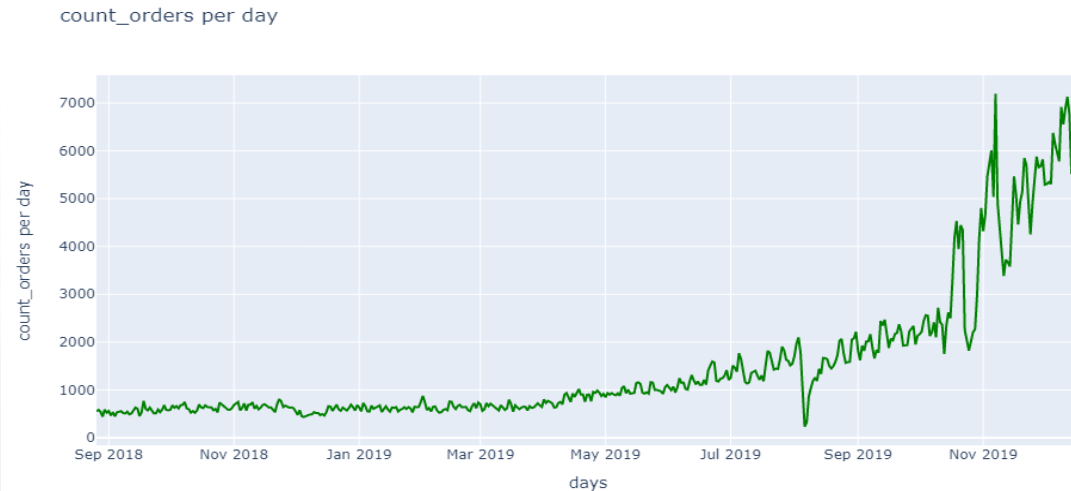
Долгое время показатель GVM увеличивался незначительно, но в последнее время **GVM имеет динамику роста**, однако есть временные промежутки, когда он проседает. Такой рост может быть связан с пандемией (после неё люди стали активнее пользоваться сервисом).

AOV имеет довольноную **высокую волатильность**. Интересно, что при высоком росте GMV, средний чек не испытывает большого роста, более того: после просадки средний чек не достиг своего максимального значения и видно, что за последние два месяца AOV **имеет динамику спада**.

Суммарная сумма промо-кодов постепенно падала, однако **с мая 2019 года заметна динамика роста**. Это может означать, что компания улучшила программу лояльности: ввела новые промо-коды пользователям для увеличения новой аудитории или удержания старой аудитории

## 2.0 Решение второй части задания. Работа с метриками. Анализ метрик.

### Графики:



### Выводы:

Заметно, что count\_orders имеет схожую динамику как с GVM. Кол-во заказов в день имеет **резкий скачок в октябре 2019 года**, после чего была просадка и далее последовала динамика роста. Как было сказано выше, такая просадка может зависеть от многих внешних факторов (в частности от пандемии)

Видно, что данные 2 графика повторяют динамику графика заказов за день. Интересно, что при первом скачке на графике count\_orders, кол-во отмененных товаров **сильно возросло**, но при дальнейшем росте и скачках count\_orders, кол-во отмененных товаров не имело значительного роста. Такой первый “скачек” отмены товаров мог быть связан с тем, что **магазины были не готовы** к резкому увеличению заказов, но в дальнейшем они оценили ситуацию и закупили необходимое товаров.

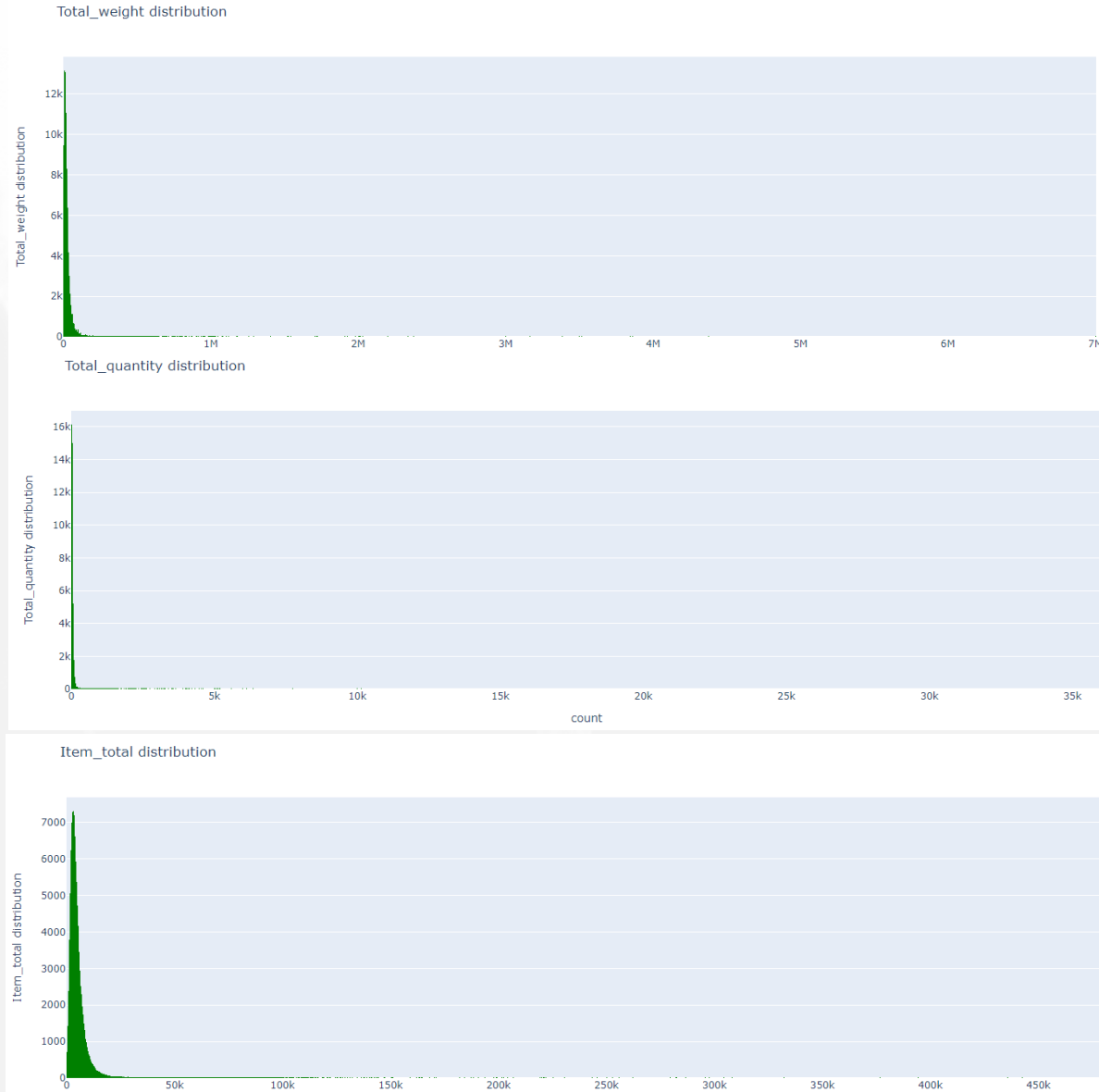


## 2.0 Решение второй части задания.

### Работа с метриками. Распределение данных.

Графики:

Выводы:



Из гистограммы распределения веса заказа видно, что большинство заказов в нашей компании **не являются слишком тяжелыми** (до 100кг), однако есть заказы, с суммарным весом **7 тонн!!**

Распределение количества товаром также имеет большое смещение. Есть заказы с количеством продуктов **36к!** Однако мода данного распределения равна **22 товарам**, то есть большинство пользователей заказывают по 22 товара за один заказ.

Наибольшая стоимость заказа равна **480к!** Но в основном Пользователи заказывают на **3к**.



## 2.0 Решение второй части задания. Первичный анализ данных.

### Информация из БД:

### Выводы:



В нашем срезе БД имеется **720к** заказов и **257к** пользователей



В среднем пользователь делает **2,8** заказа



Всего мы **доставляем из 162 магазинов**, расположенных в 7 городах, однако пользователи делают заказы по всей стране



Данный факт может сильно влиять на **время доставки** и привести к тому, что оно **будет большим** для многих заказов.



Большое кол-во продуктов было заменено (1.250к) и отменено (1.444к)



Это **негативно** сказывается на отношении пользователей к нашему сервису (далее поработаем с такими пользователями отдельно)



Несмотря на большое количество замен/отмен продуктов в заказах, малое количество заказов было отменено. Всего **15 штук**



Пользователи **могут понимать** проблему сервиса и поэтому не делают отмену заказа в случае отсутствия Товара в магазине.



Кол-во пользователей, которые имели бесплатную доставку до применения промо-кода **~33к**, а после применения промо-кода **~143к**



В компании довольно хорошо развита **программа лояльности**, позволяющая пользователям экономить на доставке продуктов



Максимальная стоимость доставки равнялась **30к!**

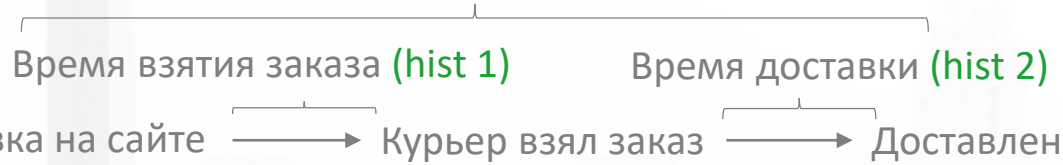


Это может быть связано с тем, что нам приходится доставлять в далекие города.

## 2.0 Решение второй части задания.

### Оценка времени выполнения заказа.

**Полное время заказа:** клиентам важно, чтобы заказы доходили до них быстро и вовремя, то целесообразно оценить время, за которое они получают свой заказ. Оно будет формироваться из **2ух составляющих** (см. диаграмму). Каждая из компонент важна, так как позволяет оценить работу курьеров и класстеризовать пользователей на группы по времени планирования заказа для дальнейших коммуникаций с ними

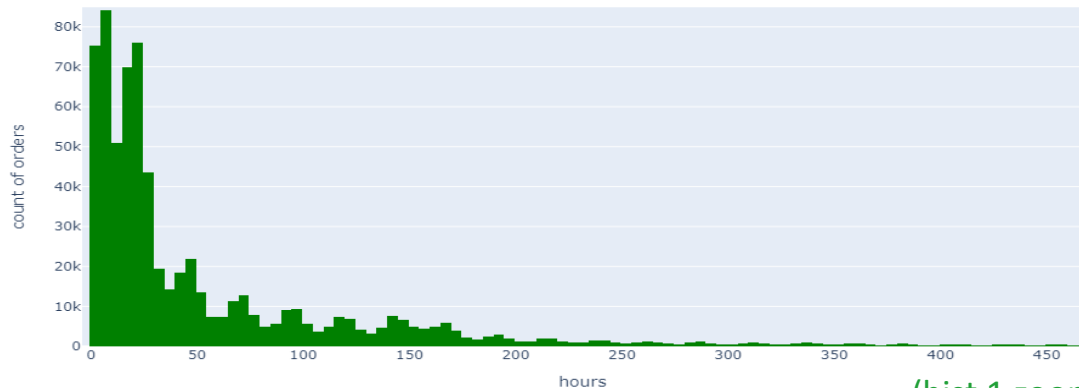


#### Выводы:

Рассм. кластеры клиентов, посмотрим среднее время заказа по ним:

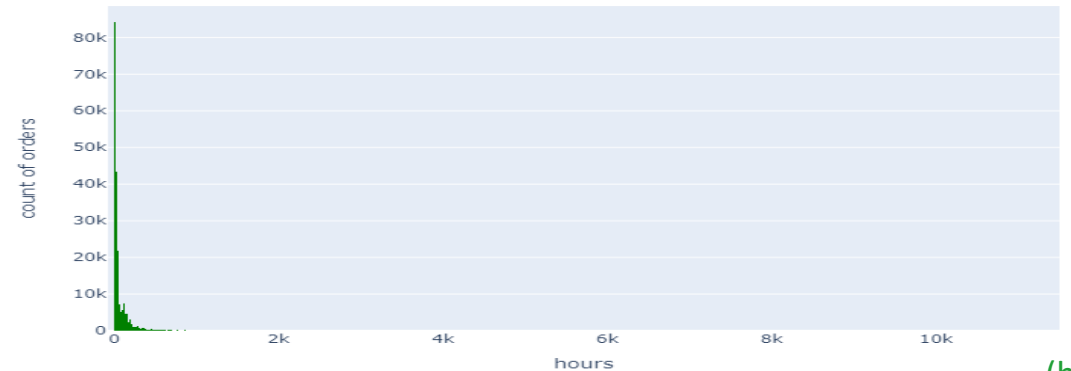
- 1) **Заказы до 48 часов:** большинство заказов планирует свои заказы на небольшой (до 2 дней) промежуток времени. Среднее время: 17 часов
- 2) **Заказы от 48 часов до недели:** меньше пользователей планирует заказы на неделю вперед. Среднее время: 124 часа
- 3) **Аномальные заказы (до 10к часов):** интересно, что есть и аномальные клиенты с огромным временем доставки (до 10к часов!)

Histogram of all\_time by orders



(hist 1 zoomed)

Histogram of all\_time by orders



(hist 1)

Также, ранее мы указали, что магазины есть всего лишь в **7 городах**, а доставлять нам приходится в **самые далёкие города России**, из-за этого могут быть получены такие результаты, то есть можно выдвинуть гипотезу о том, что в нашей компании имеется нехватка курьеров, но не стоит делать быстрых выводов, так как мы не проанализировали само время доставки (hist 2).

## 2.0 Решение второй части задания.

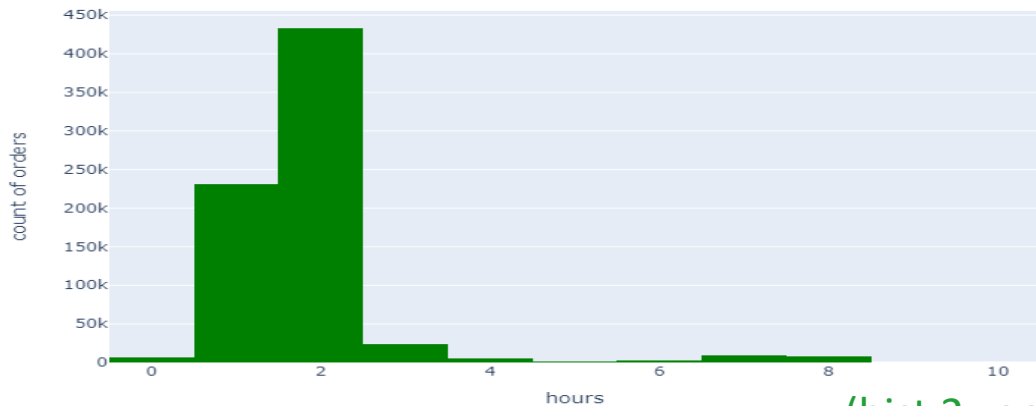
### Оценка времени выполнения заказа.

#### Выводы:

Из второй диаграммы видно, что само **время доставки заказа не является большим**. Наибольшее кол-во заказов доставляется **за 2 часа** (после того, как его взял курьер) . То есть курьеры справляются со своей работой довольно быстро, но, возможно, их количество невелико и заказы долго находятся в ожидании из-за того, что их просто некому брать. Таким образом, стоит обратить внимание на штат курьеров и может быть нанять.

Получим, что распределение пользователей по планированию заказа **очень разное** и каждому кластеру необходим свой подход. Стоит отметить, что служба доставки справляется с своей задачей **отлично** и доставляет заказы за довольно короткое время!

Histogram of dilivery\_time by orders



(hist 2 zoomed)

# 1.1 Анализ компании. Сегментация.

Сегментация  
пользователей: RFM  
анализ.

## 2.1 Решение второй части задания.

### Сегментация пользователей: RFM анализ.

Выполним **RFM анализ** с целью сегментирования пользователей на группы и оценки "здоровья" бизнеса. Такой метод анализа позволит помочь определить клиентов, приносящих большой оборот продукции, GMV, и своевременно выявить клиентов, которые потенциально могут уйти в отток. Реализуем квантильный RFM.

#### Основные показатели:

**Recency** — давность покупки (промежуток времени с момента последней покупки)

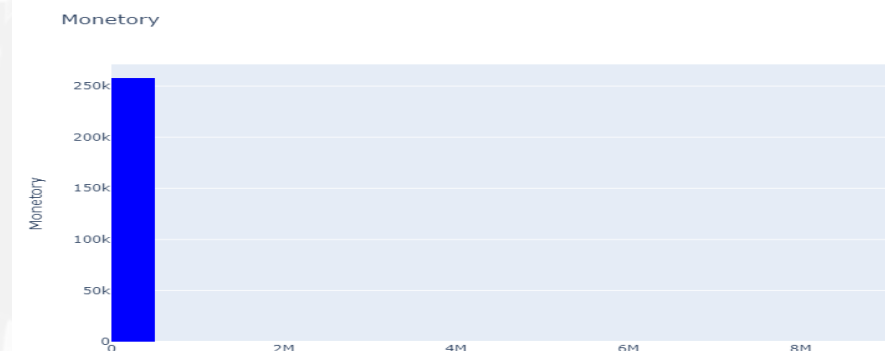
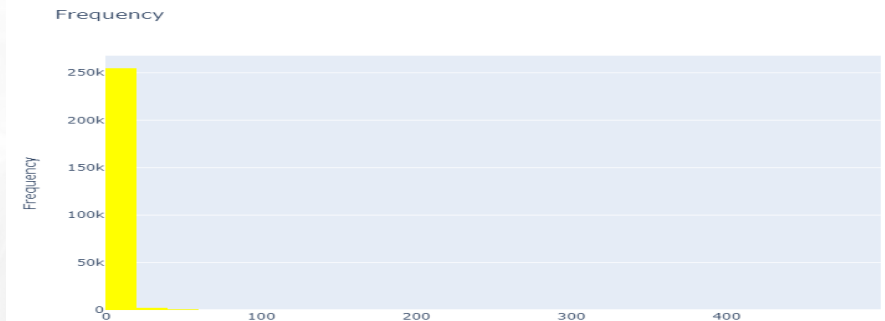
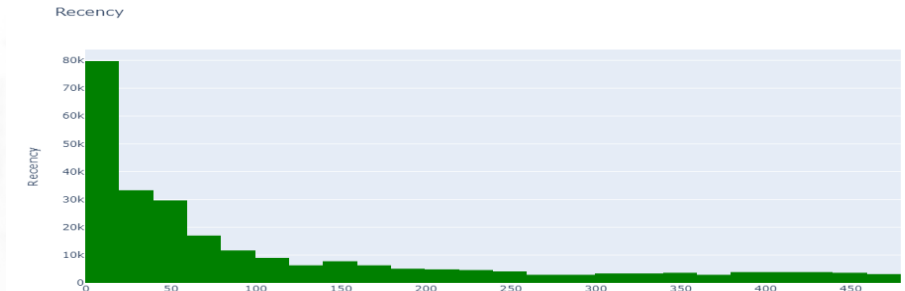
**Frequency** — частота покупок (количество покупок за данный период)

**Monetary** — сумма покупок (сумма всех покупок за данный период)

Проводим сегментацию пользователей на основе следующей таблицы:

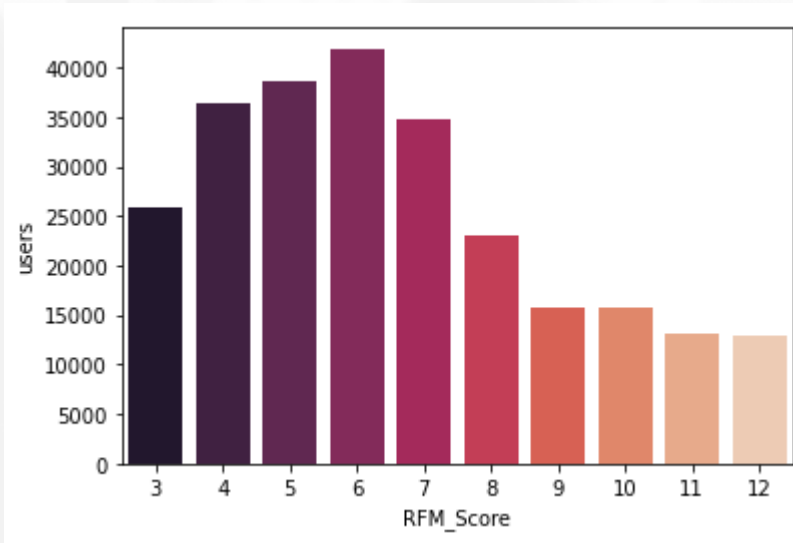
| R | 4 | Отток    | Спящие        |   |   |
|---|---|----------|---------------|---|---|
|   | 3 | Риск     | Перспективные |   |   |
|   | 2 | Лояльные |               |   |   |
|   | 1 |          |               |   |   |
|   |   | 1        | 2             | 3 | 4 |
|   |   | F        |               |   |   |

#### Распределение R, F, M.



## 2.1 Решение второй части задания.

### Сегментация пользователей: RFM анализ. Диаграмма RFM score.



#### Анализ RFM score. Рекомендации по работе с пользователями:

Кол-во пользователей с  $R=F=M=1$  достаточно **велико**, что **не очень хорошо** для нас, то есть в нашем магазине много пользователей, которые имеют низкую частоту заказов, высокий показатель давности заказа и низкую метрику Monetary

Наибольшее количество пользователей имеют  $RFM\_Score=6$ , что является **средним значением** по данной метрике. Однако стоит обратить внимание на этих пользователей и создать им условия для перехода в 7 и последующие группы.

По мере увеличения метрики  $RFM\_Score$  количество пользователей уменьшается, причём большее количество пользователей находится до среднего значения  $RFM\_Score$ . Это является **негативным фактором** для нашего бизнеса.



## 2.1 Решение второй части задания.

### Сегментация пользователей: RFM анализ. Диаграмма RF .

#### Анализ RF диаграммы. Рекомендации по работе с пользователями:

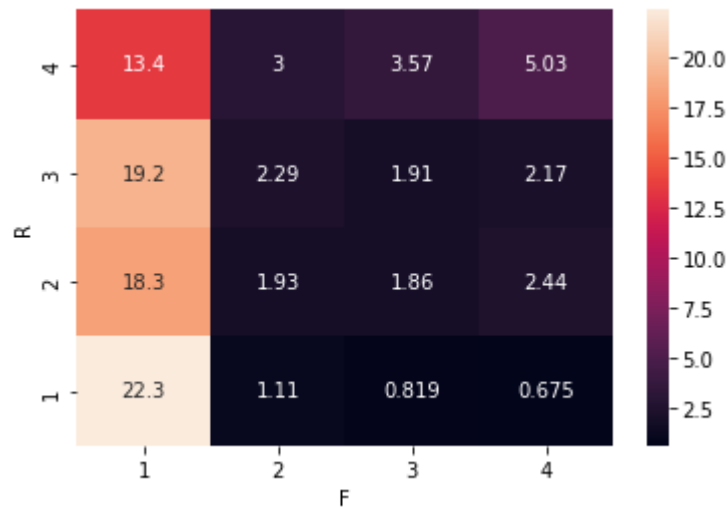
**Отток** - это пользователи, имеющие R=4 и F=1, т.е. они давно заходили в наш магазин и имели мало заказов. ~13%. Это довольно высокий показатель, поэтому отток пользователей является большим. Менеджеру проекта стоит поработать над этими людьми для уменьшения кол-ва данных пользователей.

**Риск**- пользователи имеющие R=3 и F=1. Видно, что их мало ~1%. Для уменьшения доли этих пользователей следует рассказать их о наших новых акциях, новых продуктах, т.е. постараться вернуть их обратно и не дать им перейти в группу "оттока".

**Лояльные**-пользователи имеющие R=2 и 1 , F=1. Видим, что таких пользователей ~40%.

**Спящие**-пользователи, которые довольно часто пользовались нашим сайтом в прошлом. ~10%. Следует разослать им пуш уведомления о предстоящих акциях, так как есть вероятность, что они вернуться на наш сайт, так как раньше часто заказывали товары у нас.

**Перспективные**-пользователи имеющие R=1 F=2;R=1 F=3;R=2 F=2,R=2 F=3;R=3,F=3;R=3 F=2;R=3 F=4. ~12% от всех пользователей. В дальнейшем мы надеемся, что эти люди станут нашими постоянными клиентами, поэтому стоит уделить им особое внимание. Так как они уже покупали заказывали у нас продукты, мы можем связаться с ними и предложить какой-либо подарок или предоставить бесплатную подписку на наши сервисы в течение определенного времени.





## 1.2 Анализ компании. Retention.

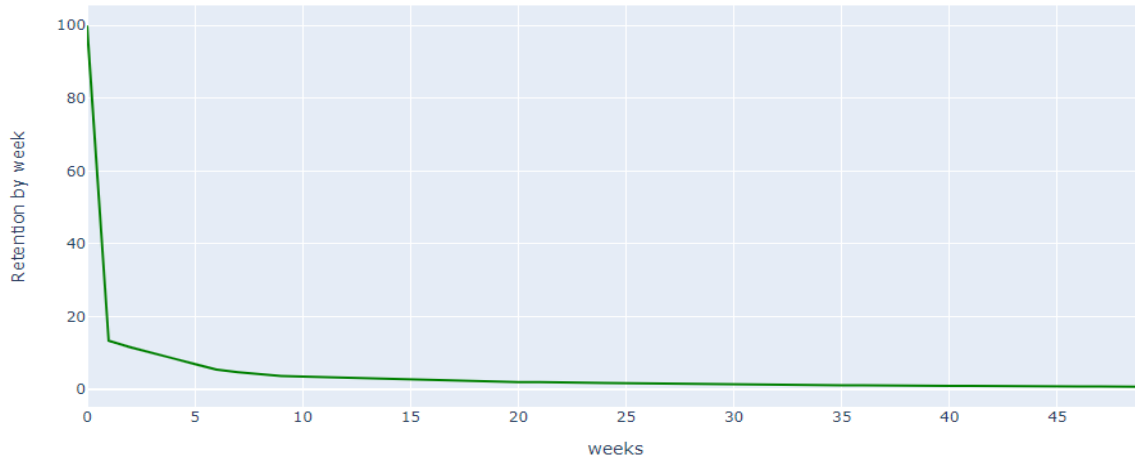
Retention пользователей.  
Графики retention.  
Когортный анализ.

## 2.1 Решение второй части задания. Retention пользователей.

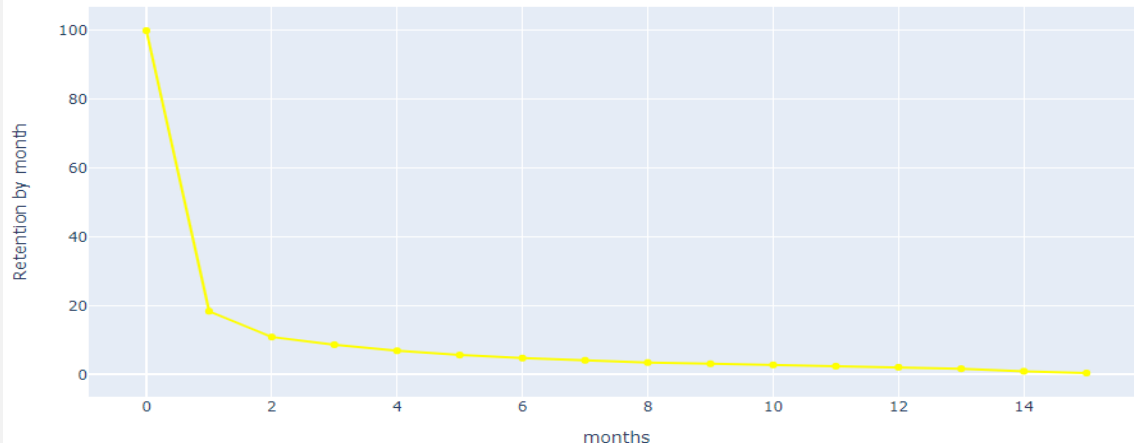
Важно понимать какой % пользователей возвращается к нам **за повторным заказом**, поэтому построим график **retention'a** пользователей, далее построим **когорты** и посмотрим какое кол-во пользователей возвращается сделать повторные заказы в нашем приложении.

Графики недельного и месячного retention'a:

Retention by week, %



Retention by month, %



### Выводы:

Из графиков retention'a, видно, что у нас **низкое** кол-во пользователей возвращается за повторной покупкой.

Так, в первый месяц возвращается 18%, во второй 10%, в третий – 8% пользователей. Поэтому стоит поработать над **удержанием** пользователей и придумать стратегию коммуникаций с ними.

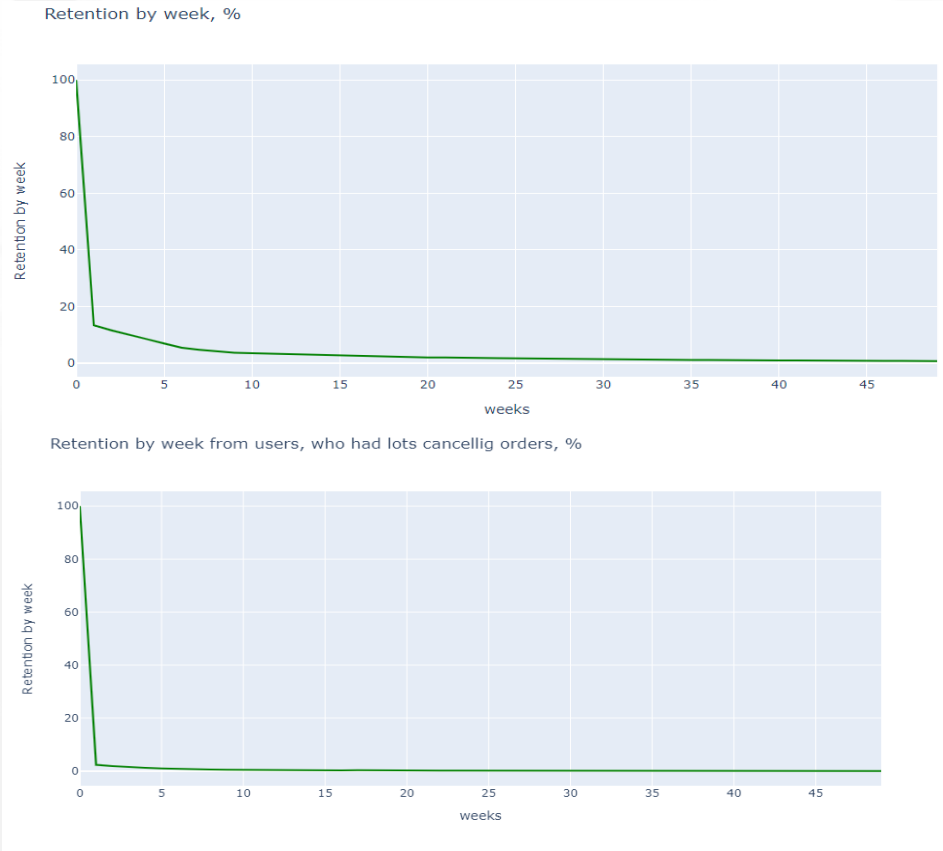
Если рассматривать недельный retention, то его значения будут следующими: первая неделя – 13%, вторая – 11%, третья – 9%. Что тоже **не является хорошими** показателями.

Но так как мы имеем высокое кол-во отмененных товаров, то целесообразно посмотреть на retention у пользователей, которые имеют **процент отмены продуктов в заказе** > 20%.

Предположим, что такие люди должны с меньше вероятностью возвращаться к нам. Сравним этот график с графиком полной аудитории.

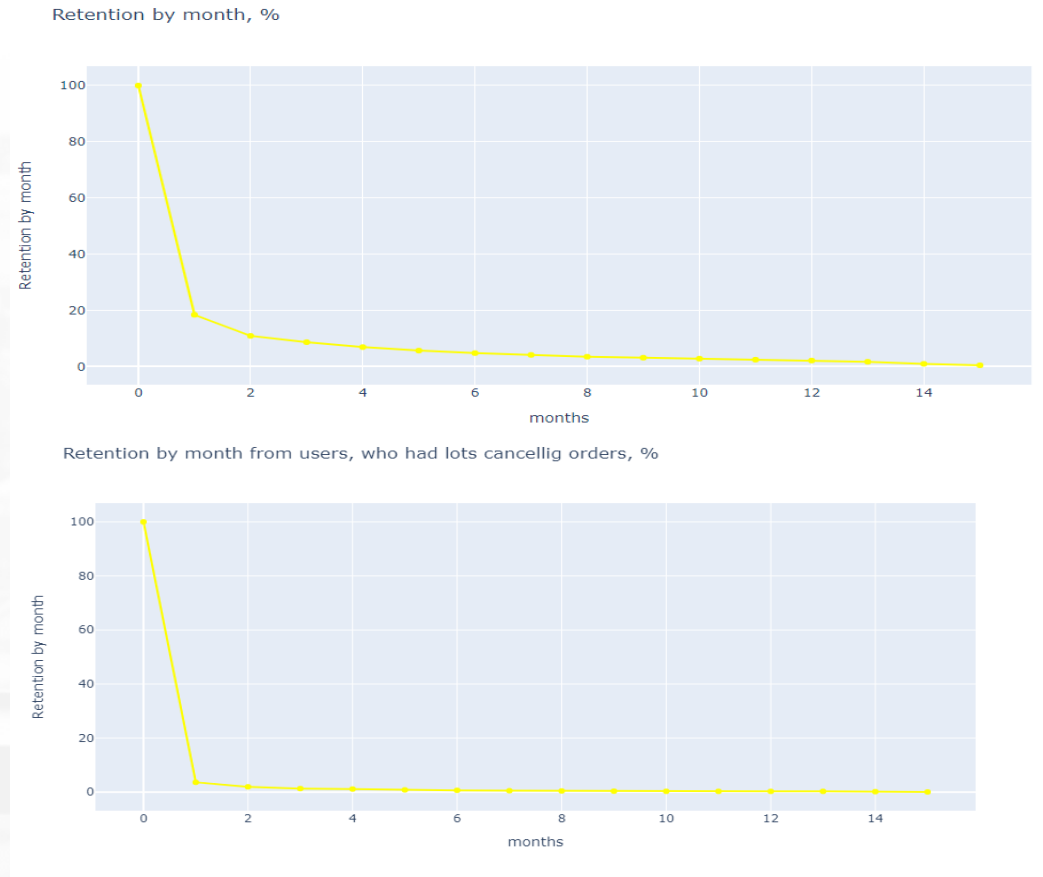
## 2.1 Решение второй части задания. Retention пользователей.

Недельный retention для 2-ух групп:



Видно, что недельный retention у таких пользователей **хуже**: первая неделя – 2.1%, вторая неделя – 1.9%, третья неделя – 1.5%, поэтому необходимо искать способы для уменьшения % отмен товаров в заказах

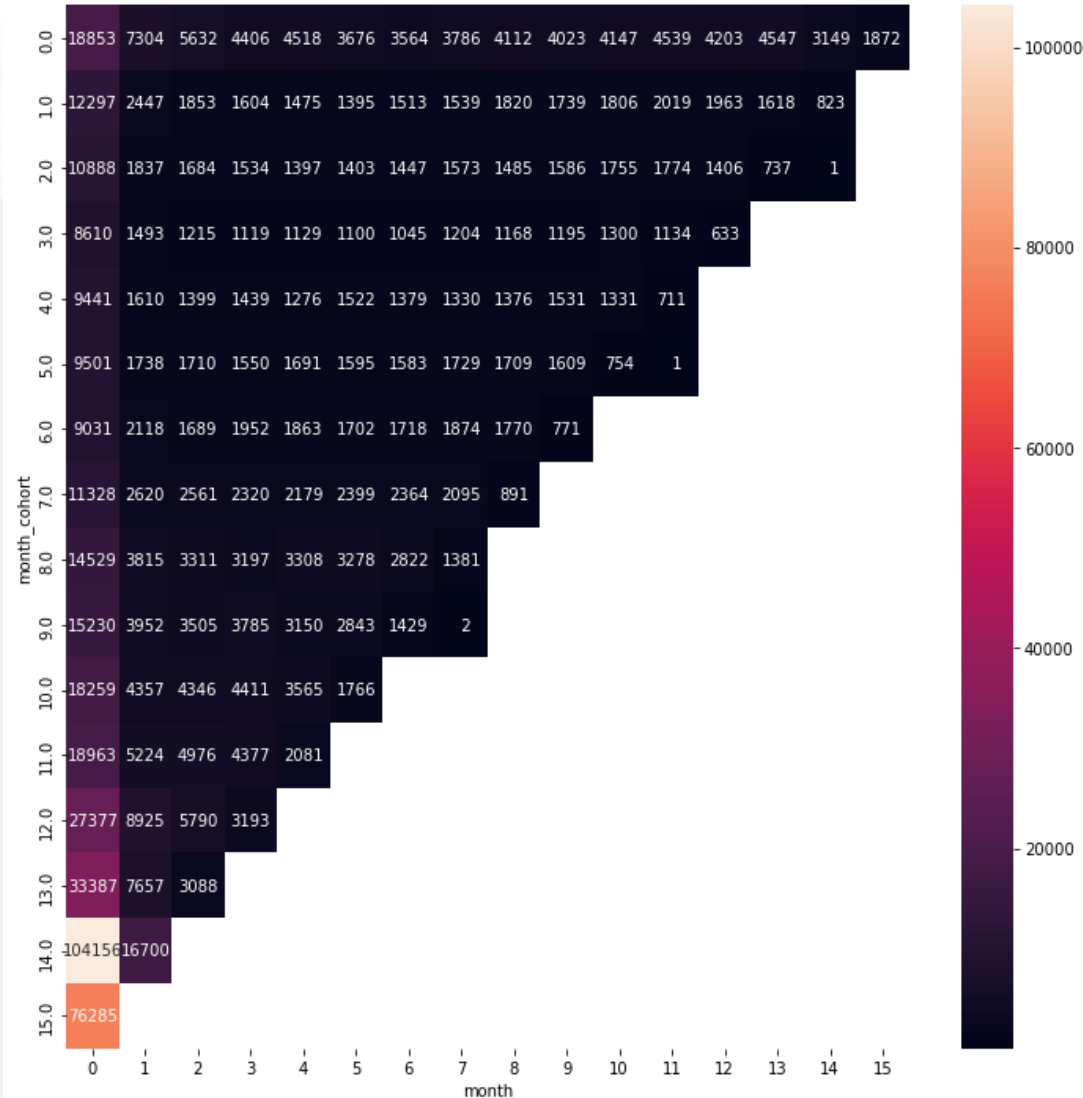
Месячный retention для 2-ух групп:



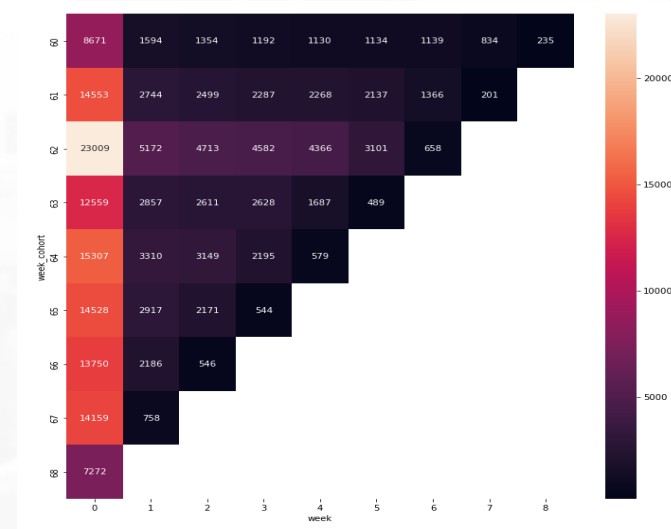
Месячный retention также **хуже**: первый месяц – 3%, второй месяц – 2%, третий месяц – 1.3%. То есть наша гипотеза о том, что данная группа людей имеет retention ниже - **подтвердилась**

## 2.2 Решение второй части задания. Retention пользователей. Когортный анализ.

Месячные когорты:



Недельные когорты (последние 8 недель):



### Выводы:

Из когортного анализа видно, что в последние месяцы увеличилась динамика прироста новых пользователей: может быть была запущена рекламная компания, поэтому на наш сайт/приложение пришло столько много новых пользователей.

Однако % их возврата остаётся всё такой же невысокий, стоит обратить на это внимание.

## 2.3 Анализ компании. ТОП магазинов.

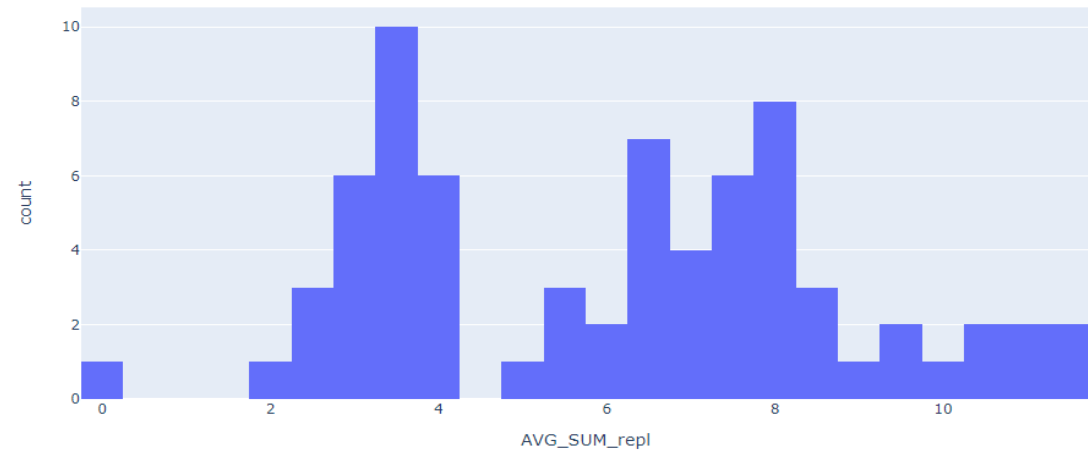
Анализ магазинов. ТОП  
магазинов по кол-ву  
отмененных /  
замененных товаров.

## 2.2 Решение второй части задания.

### Анализ магазинов. ТОП магазинов по кол-ву отмененных / замененных товаров.

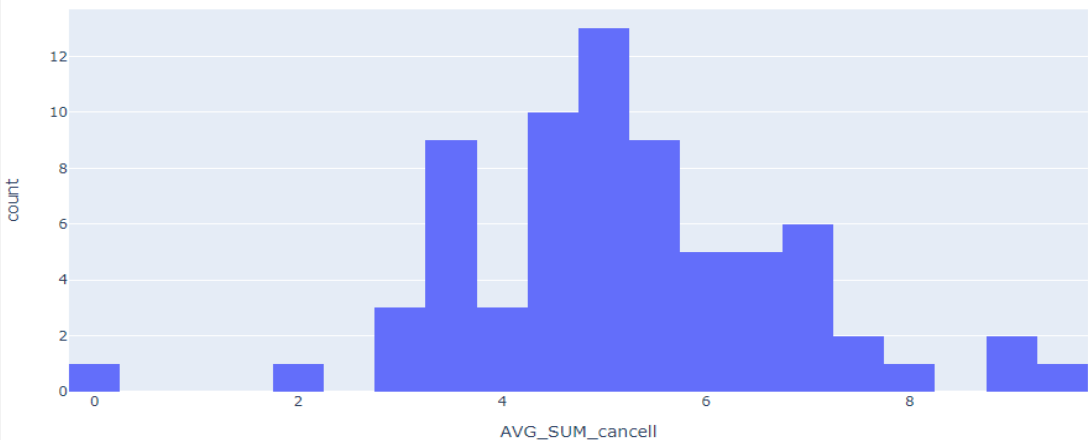
Так как мы имеем большое количество заказов с отмененными/замененными товарами, то целесообразно будет выяснить в каких магазинах было **наибольшее** количество отмены и замены товаров. Найдем % количества отмен/замен товаров в заказе по магазинам, составим **ТОП 10** магазинов с наибольшим показателем отмен/замен товаров.

SUM\_replaced\_by\_store



| store_id |     |
|----------|-----|
| 61       | 252 |
| 53       | 238 |
| 48       | 183 |
| 54       | 239 |
| 57       | 246 |
| 46       | 177 |
| 50       | 185 |
| 44       | 171 |
| 62       | 253 |
| 45       | 172 |

SUM\_cancelled\_by\_store



| store_id |     |
|----------|-----|
| 11       | 57  |
| 31       | 128 |
| 53       | 238 |
| 57       | 246 |
| 44       | 171 |
| 45       | 172 |
| 54       | 239 |
| 64       | 257 |
| 30       | 120 |
| 46       | 177 |

Будем смотреть на следующую метрику:

Кол-во отмененных заказов в магазине / Кол-во заказов в магазине \* 100

#### Выводы:

Видно, что **мода** на графике AVG\_sum\_repl равна **~3%**, то есть можно утверждать, что большинство магазинов имеют % отмен в заказе 3%. Максимальное значение отмен по магазинам составляет 12%.

**Мода** на графике AVG\_sum\_repl равна **~5%**, то есть большинство магазинов имеют % замен в заказе 5%. Максимальное значение отмен по магазинам составляет 10%.

Из таблицы с магазинами видно, что почти в одних и тех же магазинах наблюдается высокое кол-во отмен и замен продуктов. Стоит обратить внимание на данные магазины и попытаться **устранить** данную проблему

## 2.4 Анализ компании.

### Выводы.

Выводы: инсайты,  
закономерности.



## 2.4 Решение второй части задания.

### Выводы: инсайты, закономерности.

#### ■ Инсайты по компании:

- Максимальный вес заказа был равен **7,000кг**, при кол-ве товаров в заказе 6370 штук
- Максимальное кол-во единиц товаров в заказе было равно **~36к**, при суммарном весе заказа 200кг
- Наибольшая стоимость заказа равна **480к!** Но в основном пользователи заказывают на 3к.
- Один город гораздо сильно преобладает над другими по количеству магазинов –партнеров (возможно, это Москва)
- Несмотря на большое количество замен/отмен продуктов в заказах, малое количество заказов было отменено. Всего **15 штук**.
- Кол-во пользователей, которые имели бесплатную доставку до применения промо-кода **~33к**, а после применения промо-кода **~143к**. В компании довольно хорошо развита программа лояльности, позволяющая пользователям экономить на доставке продуктов.
- Максимальная стоимость доставки равнялась **30к!** Это может быть связано с тем, что нам приходится доставлять в далекие города.
- Имеются заказы с суммарным временем доставки **до 10к часов!!**

## 2.4 Решение второй части задания.

### Выводы: инсайты, закономерности.

#### ■ Аудиторный анализ: интересные закономерности

- Распределение пользователей по планированию заказа **очень разное** и каждому кластеру необходим свой подход. Имеются заказы с суммарным временем доставки до 10к часов!!
- Среди клиентов имеется много лояльных пользователей **~40%**. Стоит обратить на это внимание и проработать стратегию коммуникаций с данным кластером.
- Спящих пользователей **~10%**. Следует разослать им пуш уведомления о предстоящих акциях, так как есть вероятность, что они вернуться на наш сайт.
- Имеется довольно высокий процент оттока клиентов: **~13%**. Нужно обратить на это внимание!
- Также есть довольно большой кластер перспективных пользователей: **~12%**
- **Малое** кол-во пользователей возвращается за повторной покупкой (низкий retention пользователей). Ещё меньше он у тех пользователей, кто имеет кол-во отмены продуктов в заказе > 20%.
- В последние месяцы **увеличилась** динамика прироста новых пользователей: может быть была запущена рекламная компания. Однако % возврата остаётся всё такой же невысокий.

#### ■ Анализ компании: интересные закономерности

- С мая 2019 года заметна **динамика** роста показателя суммарных промо акций. Это может означать, что компания улучшила программу лояльности.
- Всего мы **доставляем из 162 магазинов**, расположенных в 7 городах, однако пользователи делают заказы по всей стране. Это усложняет задачу логистики.
- В последнее время **GVM имеет заметную динамику роста**, однако есть временные промежутки, когда он проседает.
- Интересно, что при высоком росте GMV, **средний чек** не испытывает большого роста, более того: после просадки средний чек не достиг своего максимального значения и видно, что за последние два месяца AOV **имеет динамику спада**.
- Время доставки заказа **не является большим**. Наибольшее кол-во заказов доставляется **за 2 часа** (после того, как его взял курьер) . Курьеры справляются со своей работой довольно быстро.
- Максимальный % Кол-Во отмененных заказов в магазине/Кол-во заказов в магазине составляет **~9%**, а замен **~11%**, что является довольно высокой цифрой, которую необходимо стараться уменьшить.

