



# E-commerce: анализ товарного ассортимента

интернет-магазин «Пока все ещё тут»

10.2018-10.2019

## Описание проекта и постановка задач

- Данные описывают транзакции интернет-магазина товаров для дома и быта.
- Задача: проанализировать ассортимент товаров на предмет разделения ассортимента на основной и дополнительный.

Определить, какие товары входят в основной и дополнительный ассортимент, чтобы грамотно предлагать покупателям дополнительные товары и оптимизировать закупки (нет смысла закупать много дополнительного товара, когда нет основного).

- Заказчик: менеджер проекта, который отвечает за актуальность товарного ассортимента.

# Описание предоставленных данных

3

- Получен датафрейм 6737 строк с шестью столбцами: дата, идентификатор покупателя, идентификатор заказа, наименование товара, количество товара в заказе, цена товара.
- Каждый товар из одного заказа идет отдельной строкой - как отдельная транзакция. Нет сведений, что это за транзакции - являются ли они покупками. **Сделали допущение: транзакция = покупка.**
- Заказов – 2784 уникальных номера.
- Период данных: с 01 октября 2018 по 31 октября 2019 - 13 месяцев.
- Обнаружили, что почти треть данных – это дубликаты логов в заказ в разное время, под разными идентификаторами покупателя, дубликаты состава заказа с разным количеством товара, и даже дубликаты транзакций по разной цене. **Сделали допущение: наиболее достоверные данные о транзакции – в последнем по дате-времени лог.**
- Обозначили выбросы по количеству товара в заказе и по цене как транзакции с числом штук более 100 или ценой более 6130 ден.ед.

# Отчет по предобработке

4

- Составили баг репорты по всем найденным дубликатам.
- Удалили все найденные дубликаты, оставив только последние по дате-времени логи, кроме дубликатов по цене. После удаления дубликатов в датафрейме осталось 4820 строк (71.5% строк от начального числа).
- Изучив ассортимент, выделили 12 категорий товаров и 1 дефолтную категорию `другое`.
- Присвоили всем транзакциям категории и выделили первое слово названия товара. Посчитали выручку с каждой транзакции.
- Разделили датафрейм на очищенный от выбросов и выбросы.

# Категории ассортимента

5

- ▶ `товары для санузла` - всё для ванной комнаты и туалета: коврики, сиденья и шторы для ванной, штанги для шторы, средства для прочистки труб, ершики и вантузы для унитаза, товары для стирки (емкости и порошки, отбеливатели), товары личной гигиены (зубная паста).
- ▶ `уборка` - тряпки, щетки, сметка, швабры и их составляющие, веники, совки, урны, мусорные контейнеры и ведра, средства для мытья и полировки, оконной мойки, перчатки и т.д.
- ▶ `операции с бельем` - все для сушки, глажки и развески белья: сушилки всех мастей, гладильные доски, чехлы/покрытия, рукава, подрукавники, вешалки, плечики и их наборы, прищепки и ролики для одежды, сменные блоки к ним.
- ▶ `текстиль` - постельное белье в комплектах и по отдельности, полотенца, халаты, салфетки из ткани, покрывала, пледы, одеяла.
- ▶ `посуда и скатерти` - всё для кухни.
- ▶ `сумки-тележки`.
- ▶ `тележки-стремянки`.

# Категории ассортимента

6

- ▶ `растения и рассада-семена` - сюда отнесли взрослые растения в горшках, а также рассаду, саженцы и семена, а также срезанные цветы (подарочные).
- ▶ `хранение и ковры` - ковры и придверные коврики (кроме ковриков для ванной комнаты), а также обувницы всех мастей, мебель и ящики-коробки-кофры для хранения, кроме чехлов для одежды.
- ▶ `стройтовары` - крепежи и инструмент, шнуры-шпагаты-фалы, чертежные инструменты, а также большие ёмкости типа котла-корыта.
- ▶ `электротовары` - весы, пылесос, электроштопор, соковарка, электрический чайник, миксер, утюг, лампа, фен. Мантоварка, овощеварка - тоже отнесли сюда, но надо уточнить у менеджера.
- ▶ `муляжи` - искусственные цветы и муляжи фруктов.
- ▶ `другое` - сюда попадали товары, для которых не нашлось категории, типа уличного термометра или почтового ящика, а также товары, названия которых нужно корректировать, чтобы они попали в правильную категорию.

## Варианты возможного разделения ассортимента на "основной" и "дополнительный":

### первый

- основной - те товары, которые непосредственно удовлетворяют потребность покупателя и приносят прибыль магазину,
- дополнительный – те товары, без которых не могут применяться основные, даже если для магазина они не очень выгодны.

### второй

- основной - те SKU, которые принесли магазину за период основную долю прибыли,
- дополнительный - те товары, которые принесли прибыли значительно меньше, чем основные.

Пропорции можно взять, например, из ABC анализа.

\* SKU, Stock Keeping Unit = стоковая единица хранения = единица складского учета. В нашем датафрейме – это товар с уникальным названием.

\*\* ABC анализ – классификация ресурсов фирмы по степени их важности, основанная на принципе Парето: 20% всех товаров дают 80% оборота.

# Исследовательский анализ

8

первый вариант деления на «основной» и «дополнительный»

Для поиска жестких взаимосвязей между товарами по первому типу разбиения ассортимента на «основной» и «дополнительный» разделили товары на категории, а в них на подкатегории=товарные группы по первому слову названия. Посмотрели многокомпонентные заказы (больше 1 SKU).

**Жестких корреляций между SKU не нашли.**

К взаимосвязанным товарам из имеющегося ассортимента можно отнести несколько сочетаний, типа:

- гладильные доски - рукава/подрукавники - чехлы/покрытия;
- шторы для ванной - кольца для штор;
- ролики для одежды - сменные блоки для них;
- растения - кашпо.

НО: при первичной покупке основной товар – гладильная доска, штора, ролик или растение - уже полностью готовы к применению, не обязательно покупать дополнительный товар сразу для использования основного, потребность в дополнительном возникает потом, когда основной поизносился/потратился/надоел.



# Исследовательский анализ

9

второй вариант деления на «основной» и «дополнительный»

Особенность наших данных: известна только выручка. Наценка/норма прибыли/маржинальность для разных категорий товаров и даже разных товарных групп внутри них может различаться, разделение SKU по выручке на **ABC не может служить однозначным критерием разделения ассортимента на основной и дополнительный.**

- сгруппировали транзакции очищенного от выбросов датафрейма по категориям и посчитали показатели: общую выручку с категории, варианты взвешенной выручки (доля в общей), число SKU в категории, число транзакций с этими SKU, и среднюю "стоимость" одной транзакции (сколько выручки приносит в среднем).
- проанализировали выбросы по категориям, числу транзакций и выручке.
- сгруппировали датафрейм по SKU и поделили на ABC - анализировали очищенные от выбросов данные. Сравнили все группы ABC с разделением на все категории по доле в выручке; по числу SKU и по числу транзакций. Попробовали посмотреть ABC внутри отдельной категории.
- **предложили условное разделение ассортимента по второму варианту и пути дальнейшей работы с ассортиментом.**

# Исследовательский анализ

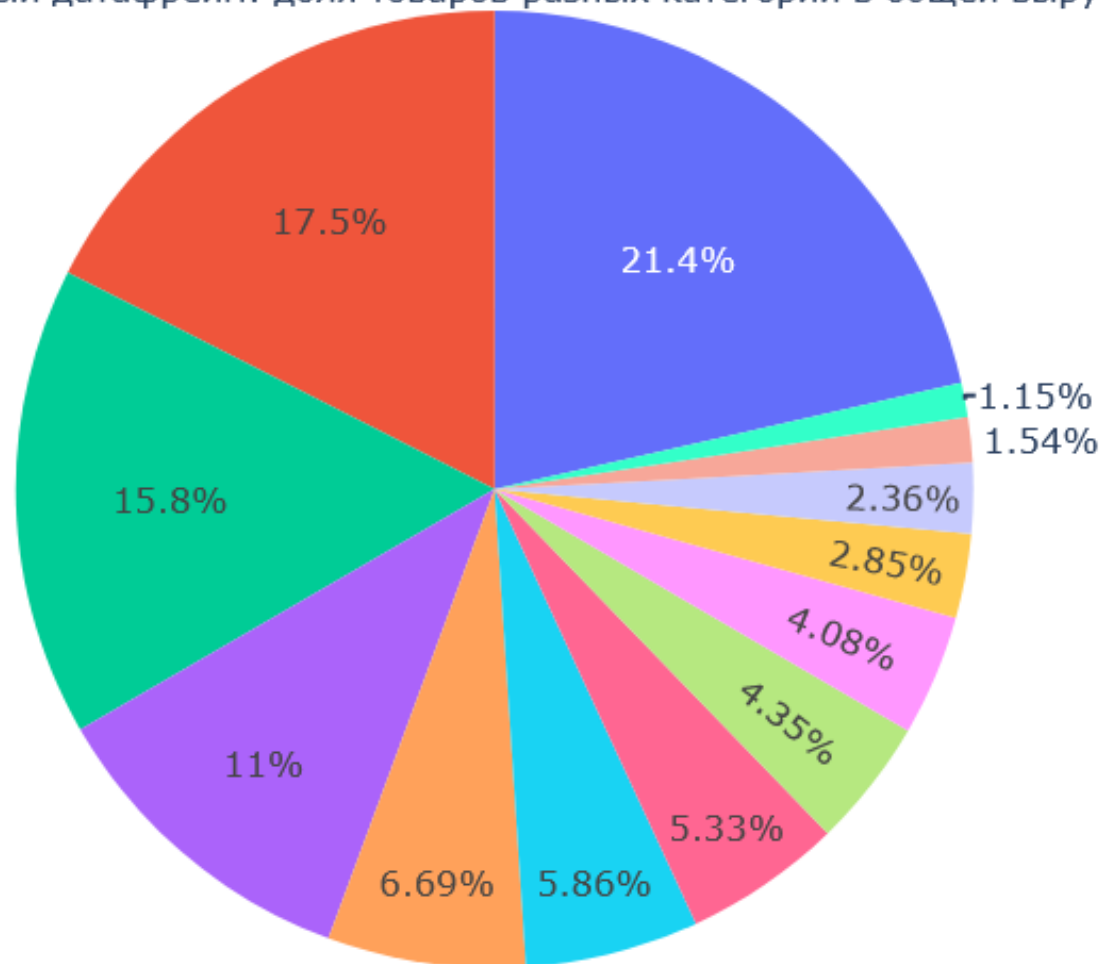
10

характеристики категорий ассортимента

	category	revenue	all_rev_share	clear_rev_share	product	transacts_count	quantity
2	операции с бельем	674842.0	0.1639	0.2138	331	615	1279
4	растения и рассада-семена	551689.0	0.1271	0.1696	917	2462	3803
6	сумки-тележки	499239.0	0.1217	0.1583	77	232	235
3	посуда и скатерти	346651.0	0.0828	0.1091	319	392	1282
8	тележки-стремянки	210664.0	0.0516	0.0668	53	142	201
9	товары для санузла	184736.0	0.0445	0.0583	122	186	444
11	хранение и ковры	168068.0	0.0413	0.0535	76	110	205
10	уборка	136960.0	0.0330	0.0433	64	85	574
7	текстиль	128384.0	0.0311	0.0407	35	36	120
0	другое	89875.0	0.0213	0.0284	127	169	468
1	муляжи	74499.0	0.0175	0.0232	129	270	1708
5	стройтовары	48577.0	0.0117	0.0151	39	53	585
12	электротовары	36260.0	0.0087	0.0114	30	34	37

# Исследовательский анализ

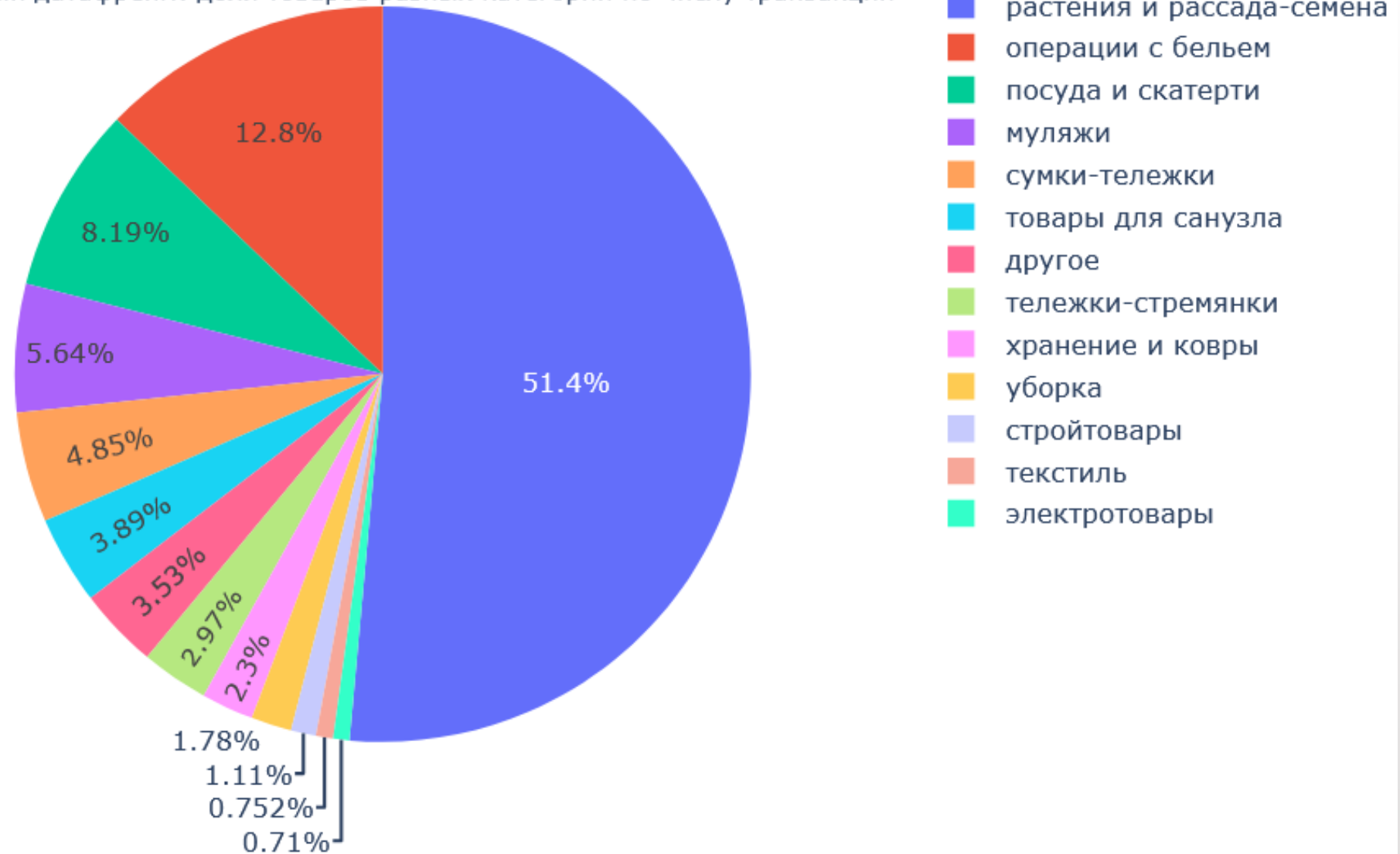
Очищенный датафрейм: доля товаров разных категорий в общей выручке



- операции с бельем
- растения и рассада-семена
- сумки-тележки
- посуда и скатерти
- тележки-стремянки
- товары для санузла
- хранение и ковры
- уборка
- текстиль
- другое
- муляжи
- стройтовары
- электротовары

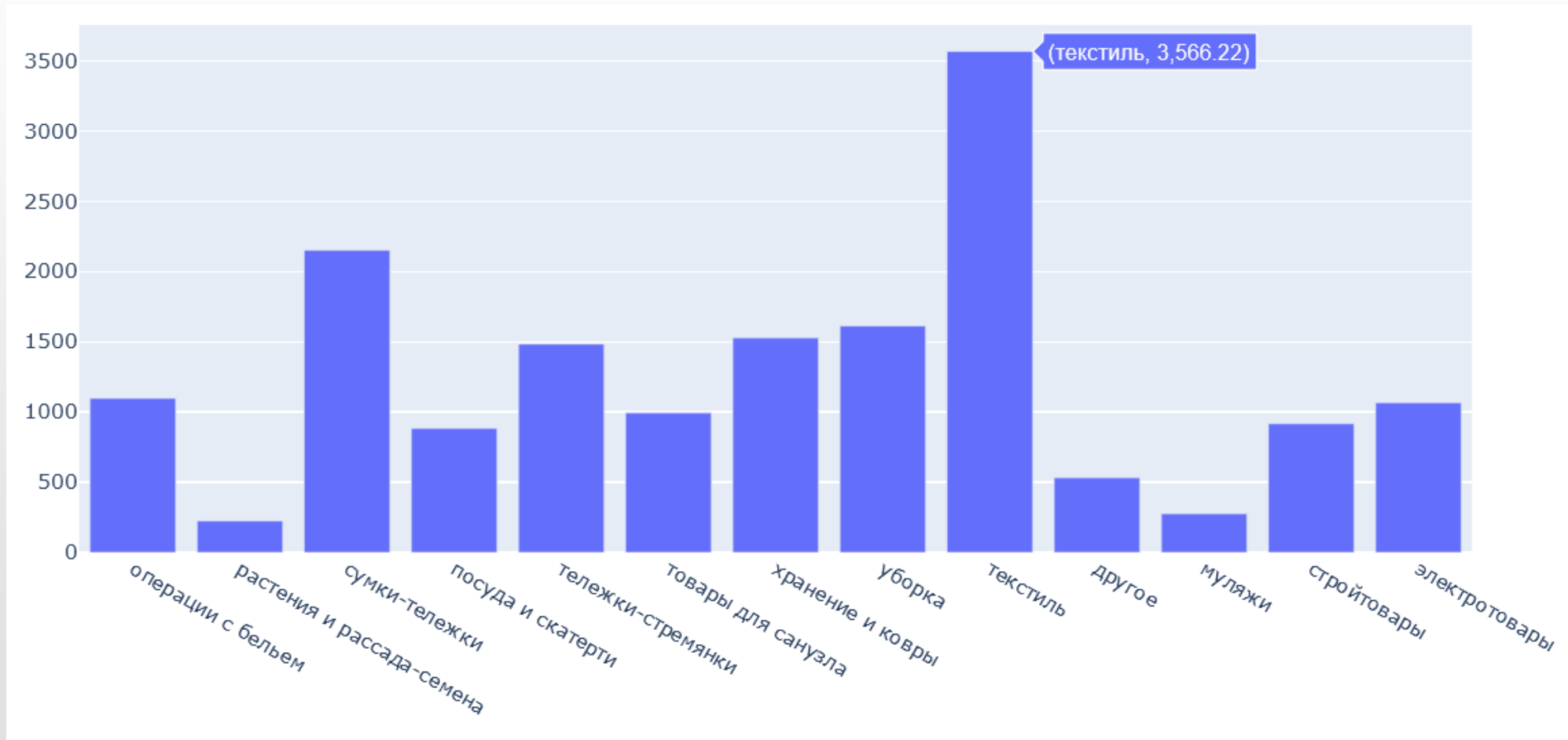
# Исследовательский анализ

Очищенный датафрейм: доля товаров разных категорий по числу транзакций



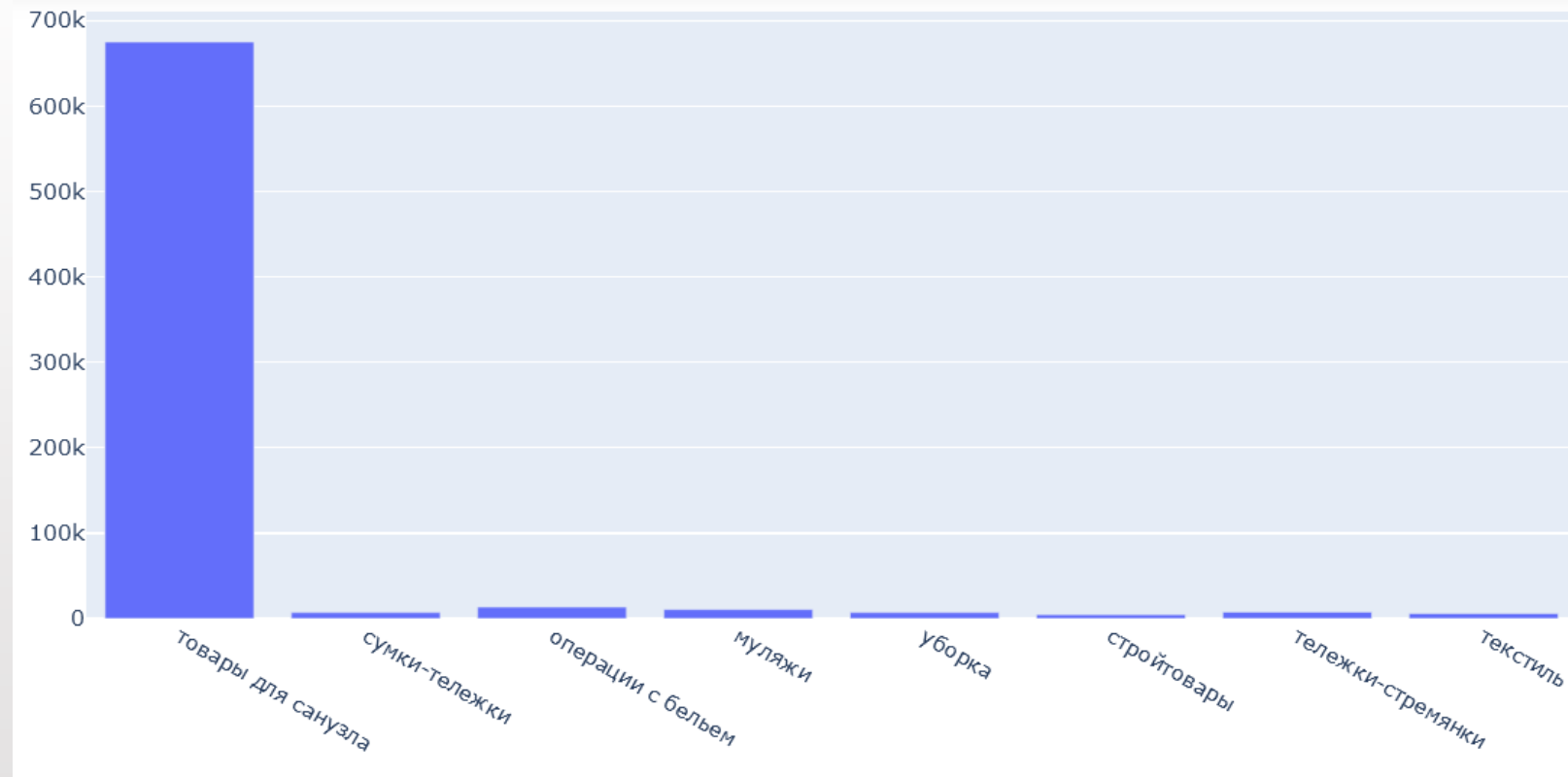
## Исследовательский анализ

очищенный датафрейм: средняя выручка с 1 транзакции в каждой категории, ден.ед/транзакция



## Исследовательский анализ

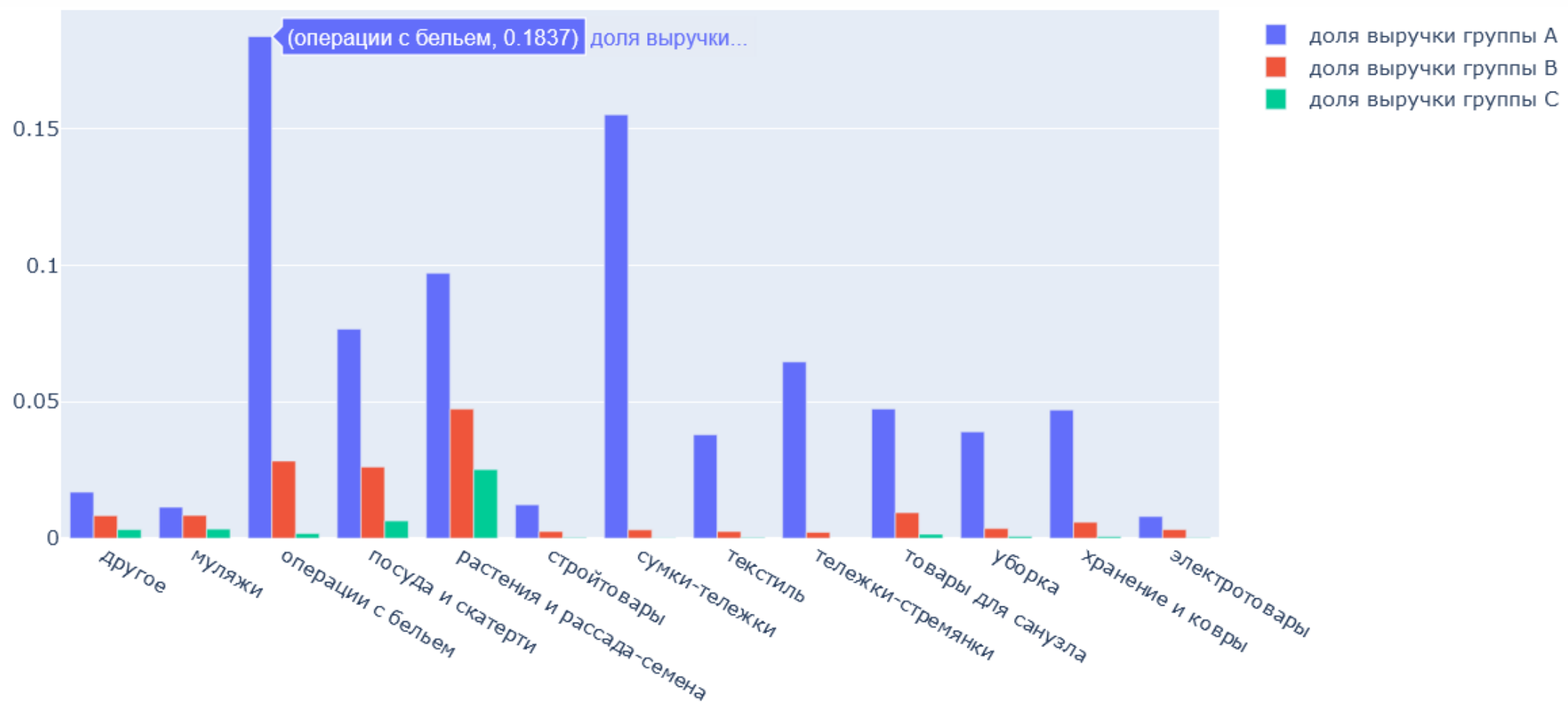
выбросы: средняя выручка с 1 транзакции в каждой категории, ден.ед/транзакция



\* Удивительные красные вантузы с деревянной ручкой в количестве 1000 штук за 1 транзакцию сделали больше выручки, чем средние два с половиной месяца работы магазина: "товары для санузла" = 675000 ден.ед с 1 реальной транзакции! Повторится ли такое вновь?..

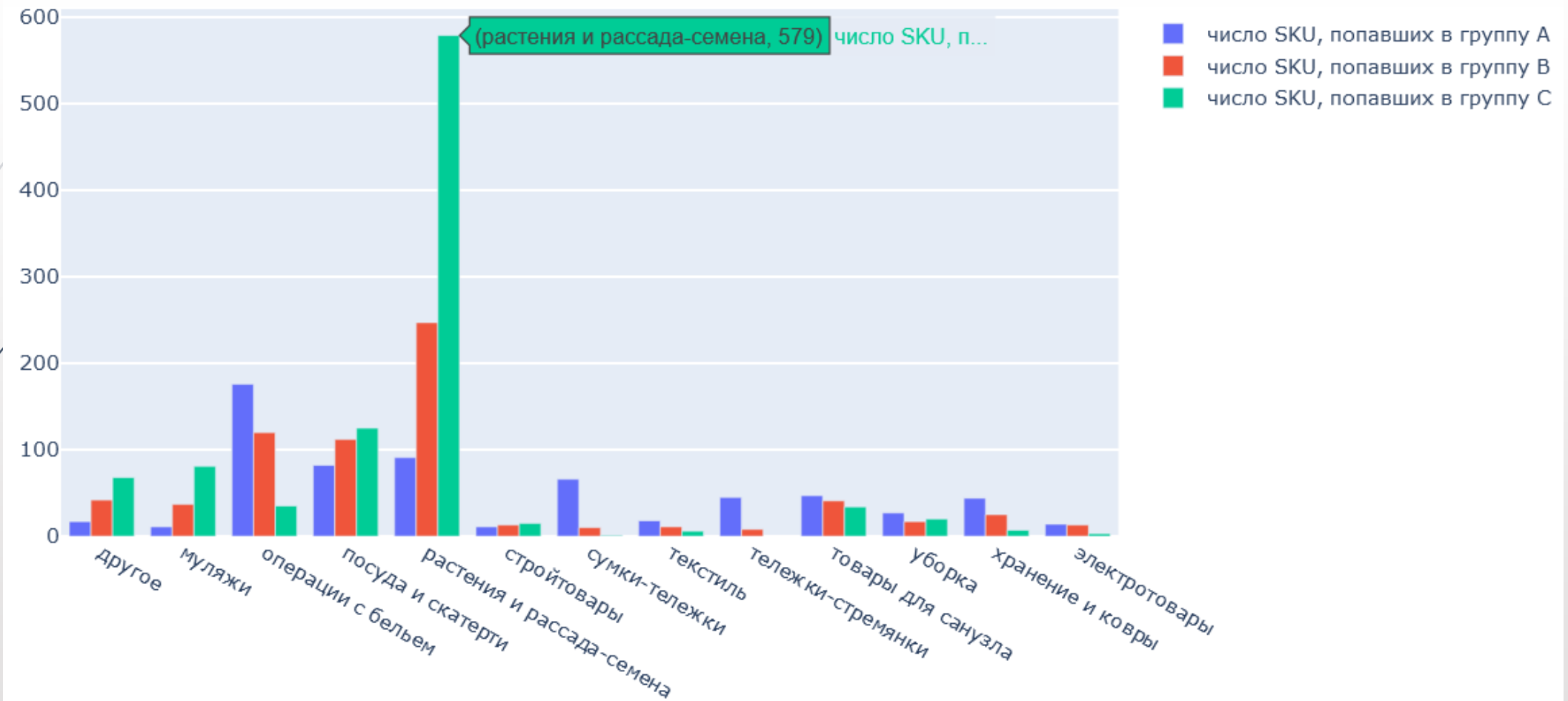
# Исследовательский анализ

ABC-анализ разделили очищенный датафрейм на группы  
A: 28% товаров (649 SKU) дали 79.70% выручки,  
B: 30% товаров (696 SKU) дали 15.08% выручки,  
C: оставшиеся 42% товаров (974 SKU) дали 4.37% выручки.



# Исследовательский анализ

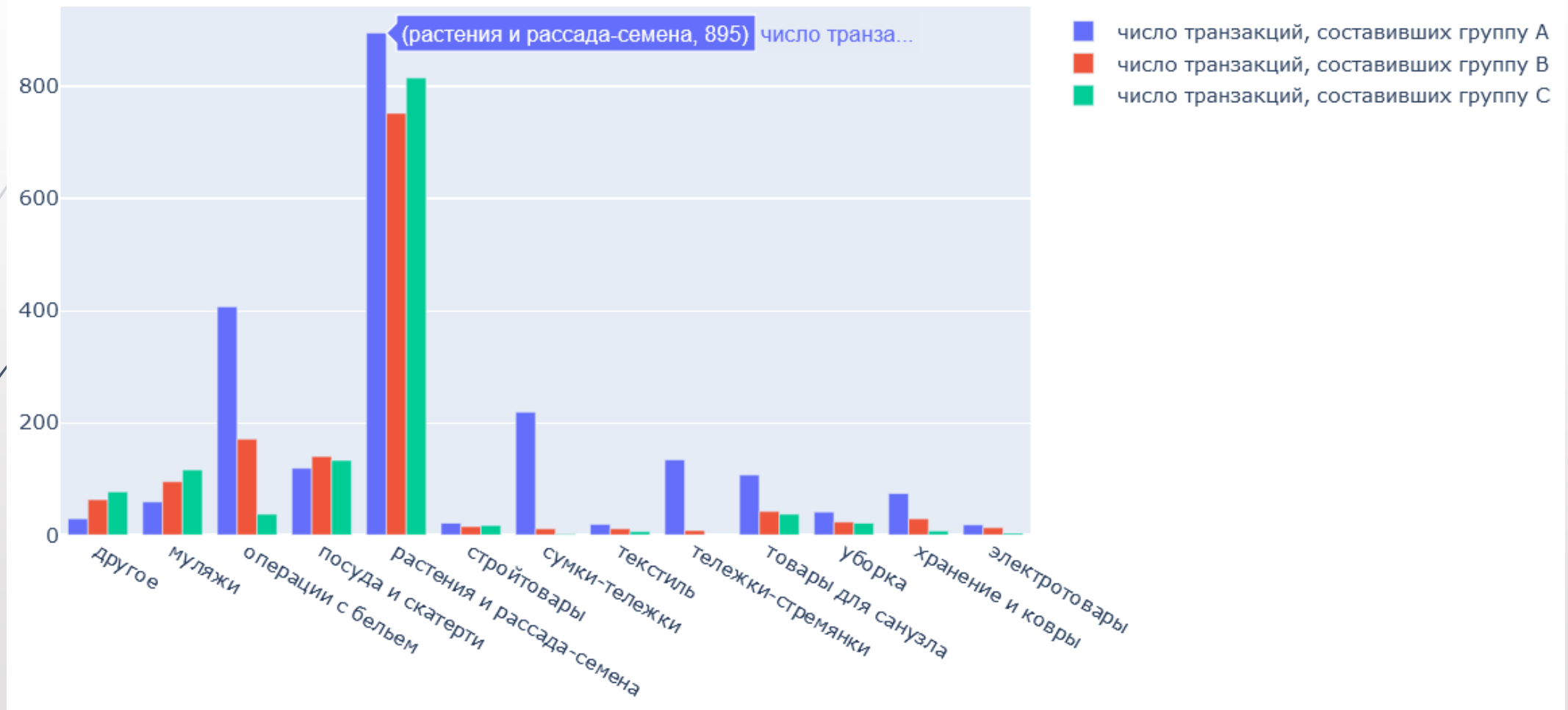
ABC-анализ число SKU





# Исследовательский анализ

## ABC-анализ число транзакций



# Исследовательский анализ

18

результаты анализа категорий в порядке убывания шансов категории считаться «основной»:

- "операции с бельем": отличная категория, но предполагаются некие логические ограничения на продажу одному и тому же покупателю, то есть нужно наращивать число новых покупателей.
- "сумки-тележки": прекрасная категория, отлично продается, наша "звезда". Предлагать эти товары всем "неотележкиным" еще покупателям.
- очень привлекательно выглядит категория "текстиль". Нет каких-то жестких ограничений по числу покупок товаров этой категории на покупателя (как, скажем, со стремянками). Рекомендация: развивать категорию - наращивать число транзакций - предлагать товары "текстиль" при выборе товаров других категорий, а также мэтчить товары "текстиля" между собой.
- "тележки-стремянки": хорошая категория, "дойная корова". Думать как удерживать долю.
- "посуда и скатерти": хороша тем, что также нет каких-то жестких ограничений по числу покупок на покупателя. Рекомендация: делить на подкатегории и товарные группы, анализировать их состав, убирать каннибализирующие и маловостребованные SKU, но внимательно с дополняющими товарами, промоутировать более дорогие товары этой категории.

# Исследовательский анализ

19

результаты анализа категорий в порядке убывания шансов категории считаться «основной»:

- "товары для санузла": потенциал хороший. Развивать, предлагать с другими категориями, мэтчить товары между собой. По "удивительным" вантузам - отдельно разбираться, чем они удивительны, и как такие продажи повторить.
- "хранение и ковры": анализировать наиболее успешные SKU, наращивать число транзакций.
- "растения и рассада-семена": с помощью специалиста по растениеводству и/или категорийного менеджера категории "растения и рассада-семена" выделить отдельно подкатегорию "рассада-семена-саженцы", также выделить отдельно подкатегорию "цветы подарочные и нг дерево" для срезанных цветов и новогодних елей - анализировать их по отдельности. Группировать и анализировать товарные группы, убирать нерентабельные товары, каннибализирующие позиции.
- "уборка": промоутировать товары подороже - наращивать число транзакций.
- "муляжи": анализировать товары категории, убирать нерентабельные, мэтчить между собой и с товарами других категорий, изучать покупателей, продумать как стимулировать покупать больше штук за раз.

# Исследовательский анализ

20

результаты анализа категорий в порядке убывания шансов категории считаться «основной»:

- "стройтовары": мало транзакций - разбираться, кто покупатели, и искать похожих и предлагать им. Предлагать с другими категориями ("растения", "санузел").
- "электротовары": мало транзакций - анализировать, кто и с чем покупал товары этой категории, предлагать похожим покупателям. Разбираться, имеет ли вообще эта категория будущее в нашем магазине, если да - промоутировать категорию.
- "другое": разбирать, что попало в эту категорию, особенно 17 SKU в 29 транзакциях из группы А , переименовывать названия товаров четко и строго без повторений слов, анализировать категорию.

# Исследовательский анализ

21

выводы и пути дальнейшего анализа ассортимента:

1. Разделение на «основной» и «дополнительный» ассортимент.

- ▶ К основному ассортименту магазина условно можно отнести:
  - большую часть товаров категорий: "операции с бельем", "сумки-тележки", "текстиль", "тележки-стремянки";
  - часть товаров категорий "посуда и скатерти", "товары для санузла", "хранение и ковры", "уборка" и часть растений из категории "растения и рассада-семена".
- ▶ К дополнительному ассортименту условно можно отнести:
  - семена и часть растений из категории "растения и рассада-семена";
  - категории "электротовары", "стройтовары", "муляжи", "другое";
  - и часть товаров категорий "посуда и скатерти", "товары для санузла", "хранение и ковры", "уборка", "операции с бельем", "сумки-тележки", "текстиль", "тележки-стремянки".
- ▶ Разделение - очень условное - с каждой категорией нужно работать отдельно.

# Исследовательский анализ

22

выводы и пути дальнейшего анализа ассортимента:

2. Для более бизнес-значимого анализа и выводов нужна дополнительная информация про данные: характер логов, уникальная идентификация SKU, информация о себестоимости и затратах либо норма прибыли/наценка для каждой категории/подкатегории.
3. Следующими шагами после уникальной идентификации SKU видятся:
  - формирование товарных групп внутри категорий и их анализ, включая ABC,
  - более подробный анализ многокомпонентных заказов,
  - формирование словаря со всеми идентификаторами покупателя (девайса) каждого реального пользователя – анализ покупательского поведения,
  - создание (придумывание) возможных вариантов сочетаний между товарами одной группы, категории и разных категорий в корзине,
  - анализ каждой категории, подкатегории по сезонности, месяцу, дню недели, времени продажи.

## Проверка статистических гипотез

сравнение выручки категории "хранение и ковры" и "уборка"

- ▶ Проверили равенство дисперсий выручки с транзакций выборок критерием Левене: не получилось отвергнуть нулевую гипотезу о равенстве дисперсий выручки со SKU для категорий "хранение и ковры" и "уборка": для равенства средних решили применять тест Стьюдента.
- ▶ Проверили гипотезу о равенстве средних двух генеральных совокупностей по взятым из них выборкам  $t$ -тестом: не получилось отвергнуть нулевую гипотезу о том, что в очищенном от выбросов датафрейме средняя выручка на SKU от товаров категорий "хранение и ковры" и "уборка" не различается - полученное значение  $p$ -value свидетельствовало о том, что различие в 3.3% средней выручки на SKU у товаров категории "хранение и ковры" и товаров категории "уборка" с вероятностью 90% было получено случайно.



# Спасибо за внимание

[emira.nunuka@yandex.ru](mailto:emira.nunuka@yandex.ru)

2021