

# “Пока все ещё тут”

Интернет-магазин товаров для дома и быта.  
Профилирование покупателей.

# Содержание

1. Постановка задачи
2. Основные бизнес-показатели интернет-магазина
3. Исследовательский анализ данных
4. Профилирование покупателей
5. Рекомендации

# Постановка задачи

- Задача- сегментировать покупателей на основе их истории покупок.
- Сегментация необходима для более персонализированных предложений покупателем.

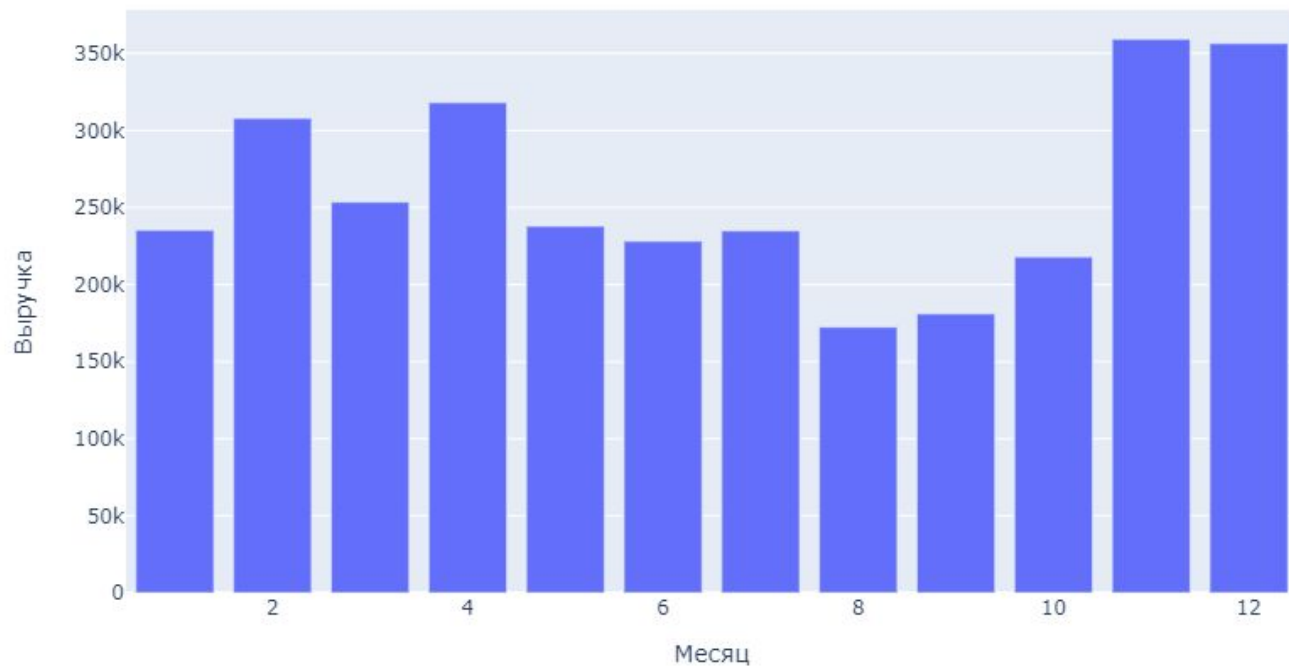
# Основные бизнес-показатели интернет-магазина

## Годовые показатели 01 Ноября 2018 - 31 Окт 2019

<b>Выручка</b>	3 104 000
<b>Покупателей уникальных</b>	2 320
<b>Покупок уникальных</b>	2 550
<b>Средний чек</b>	1 220
<b>Наименований товаров</b>	3 270

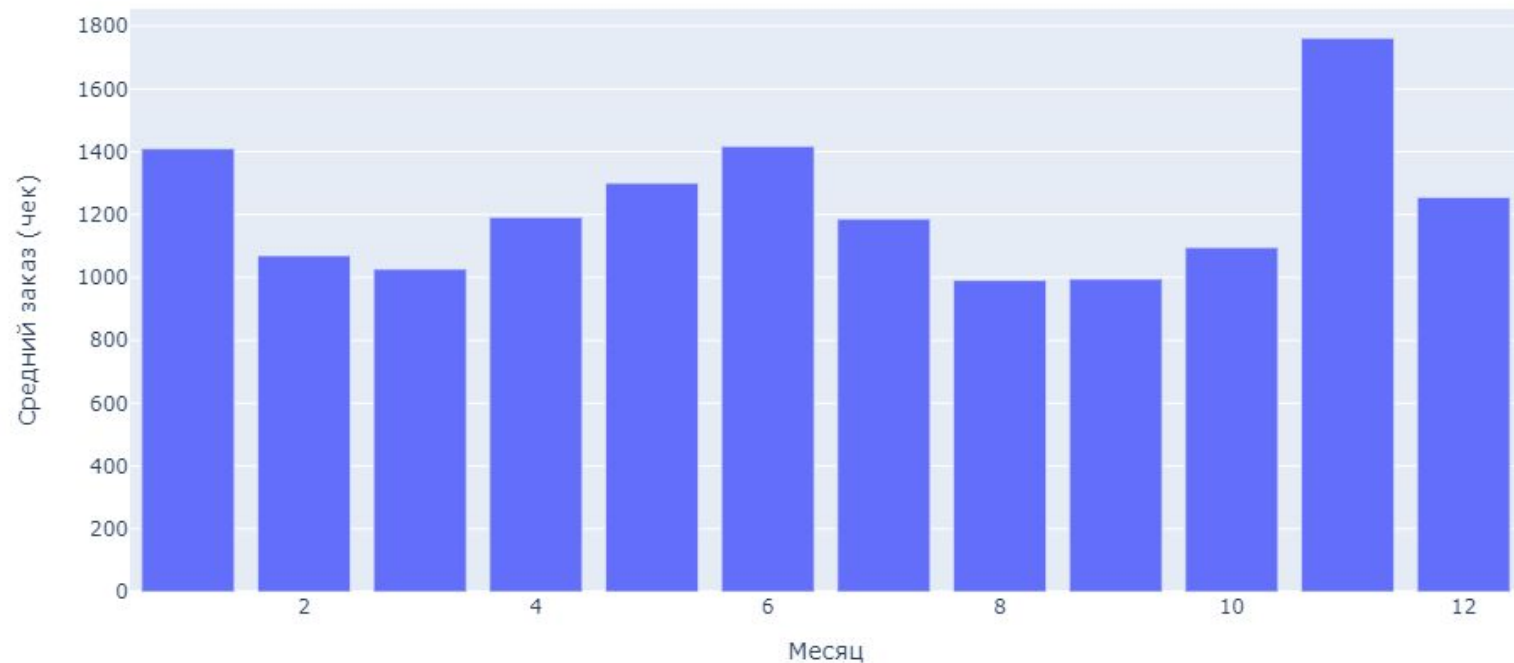
# Основные бизнес-показатели интернет-магазина

Выручка- месячная динамика



# Основные бизнес-показатели интернет-магазина

Средний заказ (чек)- месячная динамика



## Свойства исходных данных

- Датасет включает записи о транзакциях с 01 Окт 2018 по 31 Окт 2019.
- Одна строка в датасете представляет одну транзакцию - запись о дате продажи товара с информацией о наименовании, количестве единиц товара, покупателе и заказе.
- Заказ включает в себя от одной до нескольких транзакций для одного покупателя на одну дату.
- Наименования товаров не структурированы, категоризация товаров отсутствует
- Данные были очищены от тестовых и ошибочных записей (сбой формирования корзины)
- Данные предоставлены за год и 1 месяц со сдвигом от начала года. Для удобного годового анализа данные взяты за полный год с 01 Ноября 2018 по 31 Окт 2019

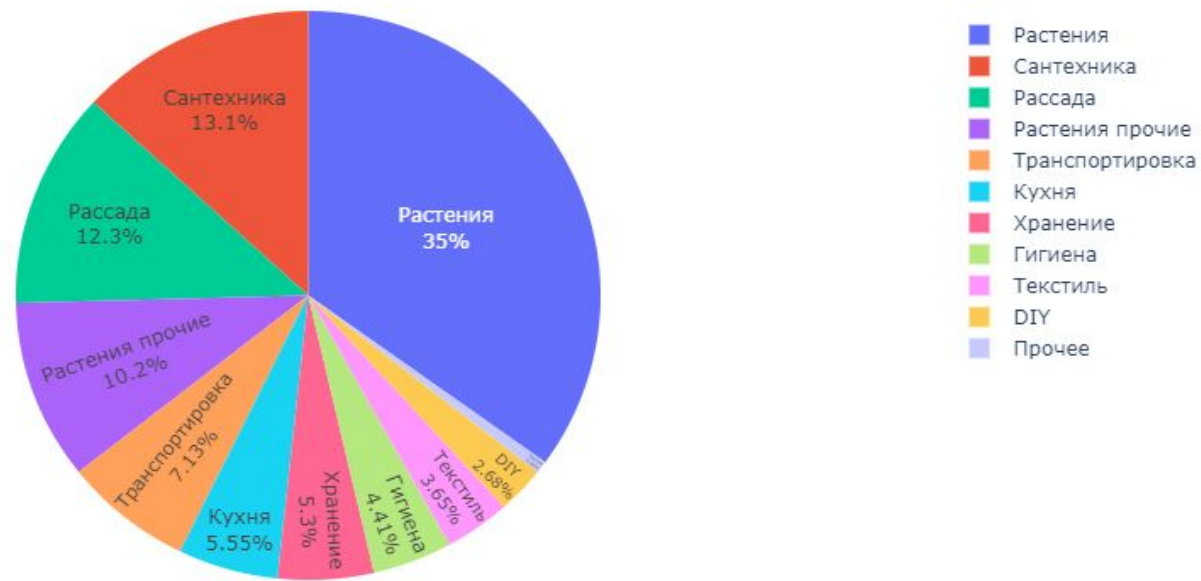
## Категоризация товаров

Для последующей сегментации покупателей по истории покупок выделим категории товаров:

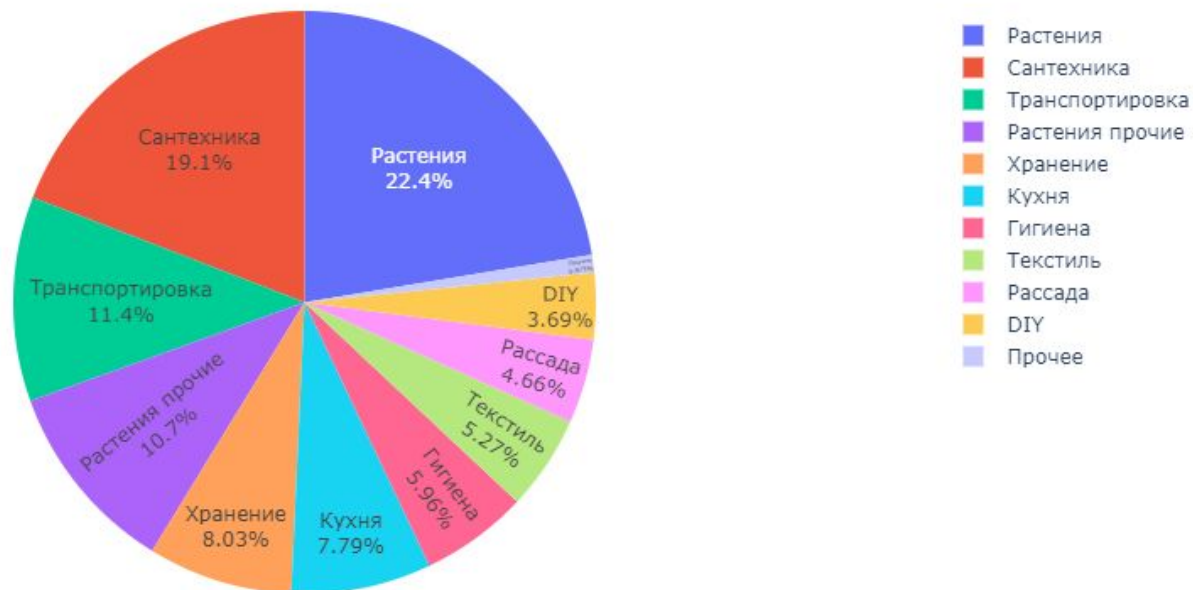
- Растения - товар идентифицируется только по наименованию растения
- Рассада - товары, в которых имеется обозначение рассады
- Растения прочие - растительные товары с дополнительными признаками
- Гигиена - товарная группа для поддержания личной гигиены и чистоты в доме
- Сантехника - товары, относящиеся к ванной комнате, санузлам и предметы для глажки белья.
- Хранение - товары для организации хранения в доме
- Текстиль - тканые ( и нетканые) материалы: подушки и одеяла, постельное белье, ковровые покрытия
- Транспортировка - товары для организации транспортировки (сумки, тележки и пр.)
- DIY (Do It Yourself) - инструмент, средства защиты, электротовары, скобяные изделия, такелаж
- Кухня -предметы кухонного интерьера, посуда, бытовая техника, кухонные приспособления.
- Прочее - товары, не вошедшие в другие категории



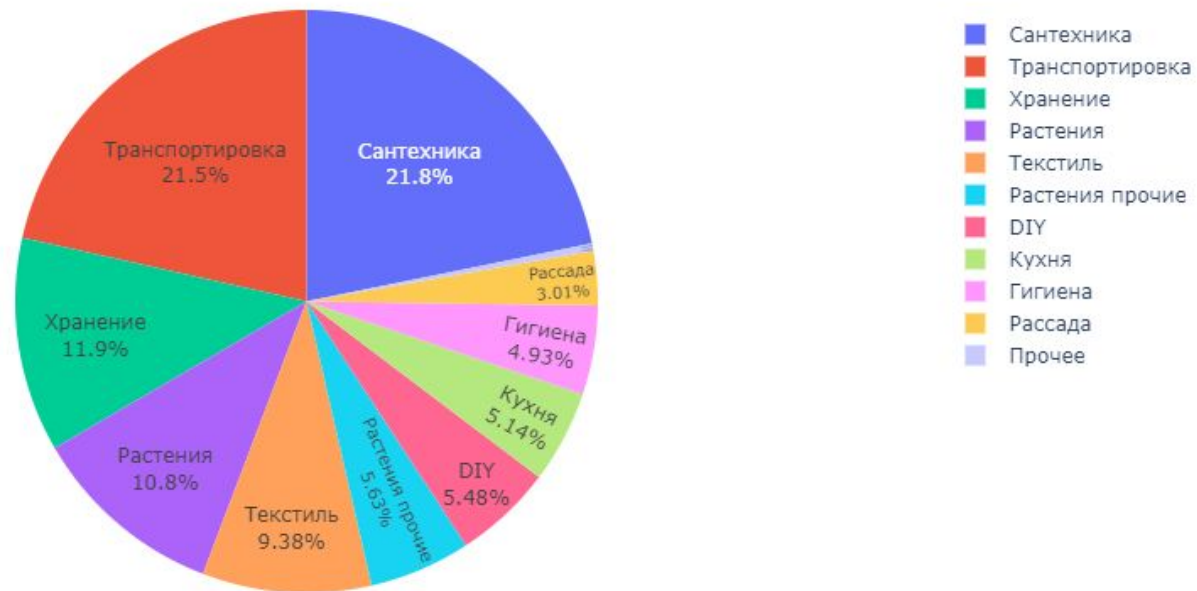
Распределение транзакций по категориям



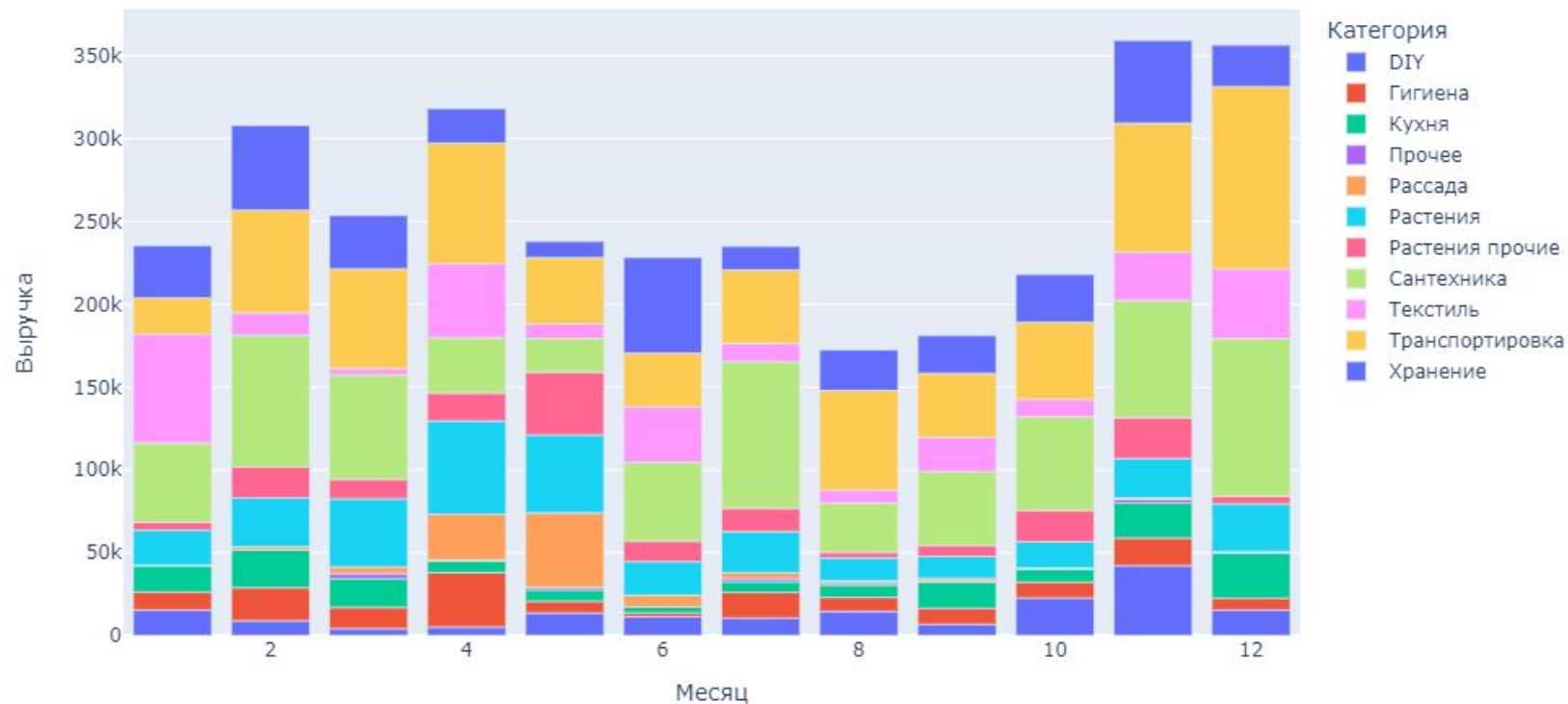
## Уникальные покупатели и категории



Суммарная выручка по категориям



Выручка- месячная динамика по категории



## Основные показатели

- **Lifetime.** Для 96% покупателей лайфтайм составляет 1 день. Максимальный лайфтайм - 137 дней.
- **Цена товаров.** 91% покупателей предпочитает товары не дороже 2000 р.
- **Категории товаров и ассортимент.** Покупатели также не сильно предпочитают разнообразие: 93-94 % покупателей покупают товары не больше 1 категории и 1 наименования в заказе. Максимально - 10 категорий и 105 наименований товаров.
- **LTV (выручка за весь лайфтайм).** Диапазон LTV для покупателей: 14 - 131 269 р. Медианное значение LTV - 686. Для 92 % покупателей LTV не более 3300.
- **Средний чек (заказ).** Диапазон средних чеков для покупателей: 14- 53 232. Медианное значение - 524. Для 91 % покупателей средний чек не более 2700.

## Основные корреляции

- **Цена- количество категорий.** Заметно снижение диапазона цен товаров для покупателей, приобретающих товары разных категорий. Есть отдельные исключения.
- **Цена- заказы, цена- кол-во товаров.** Есть транзакции с разнообразным количеством товаров (и количеством заказов) на покупателя и небольшой ценой, а также ряд транзакций с высокими ценами и небольшим количеством товара в транзакции. При этом практически нет транзакций с большим количеством дорогих товаров.
- **Цена- выручка, цена- средний чек.** Наблюдается положительная корреляция между признаками из-за в основном единичного количества товаров в транзакции и корзине.
- **Количество категорий - кол-во заказов.** Для покупателей с 1-2 категориями товаров наблюдается наибольший диапазон количества заказов, выручки и среднего чека, количества товарных единиц.

# Сегментация (профилирование) покупателей

Учитывая структуру продаж магазина и характерные особенности основной массы клиентов, профилировать покупателей можно различными способами. Наиболее интересные показатели обнаружили следующие сегменты:

1. **Типичный разовый** - типичный покупатель одного дня ( $\text{life Time} = 1$ ) приобретающий товары только одной категории в количестве 1 единицы.
2. **Не типичный** - профиль, обратный типичному- все покупатели датасета за вычетом типичных.
3. **Снабженец** - покупатель возвращается в магазин за повторными покупками, в его корзине товары различных категорий более, чем в единичном экземпляре. Поведенчески профиль противоположен типичному, но не перекрывает все варианты сочетаний признаков покупателей. "Снабженец" является частным случаем "не типичного" профиля.

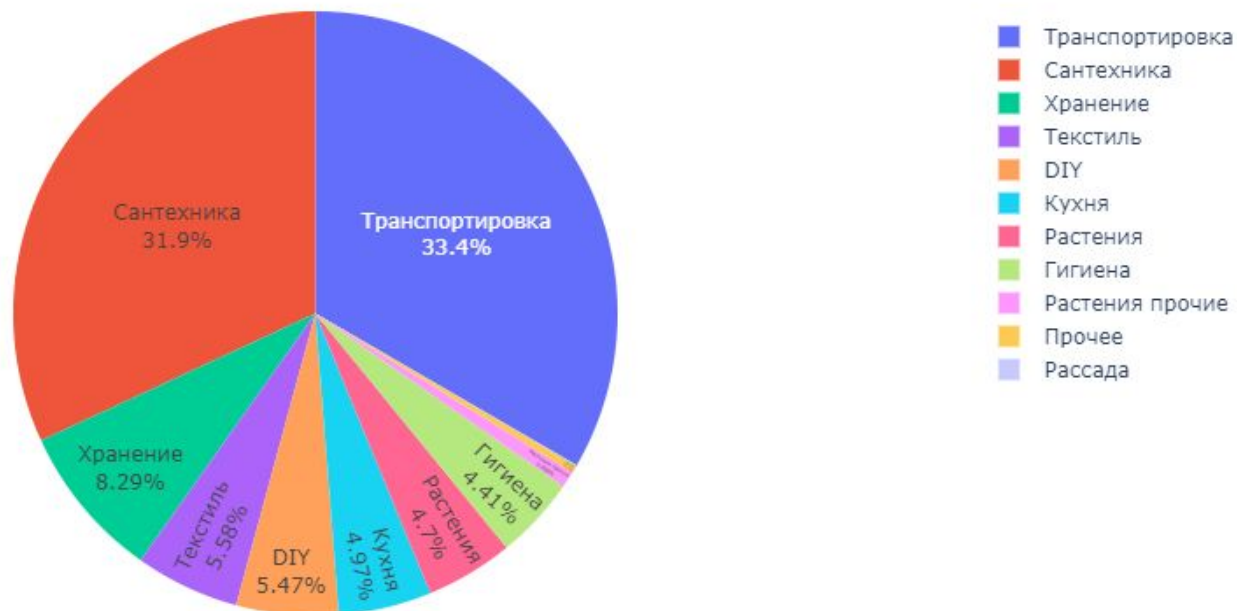
## Типичный разовый

- Типичный покупатель одного дня (Lifetime = 1), приобретающий товары только одной категории в количестве 1 единицы.
- Количество уникальных покупателей в сегменте 1399 ( 61.6 % от всех)
- Количество транзакций в сегменте 1399 ( 32 % от всех)
- Суммарная выручка сегмента: 1 473 879 (47 % от всей)
- ТОП-3 категории по выручке - Транспортировка, Сантехника и Хранение - 74 % от всей выручки сегмента
- Средний чек - 622.
- Больше 60% покупателей приобретают сумки, тележки, что-то из сантехники или для организации хранения и не возвращаются. И такие покупатели приносят немногим меньше половины выручки.



# Профиль покупателя - Типичный разовый

Суммарная выручка по категориям

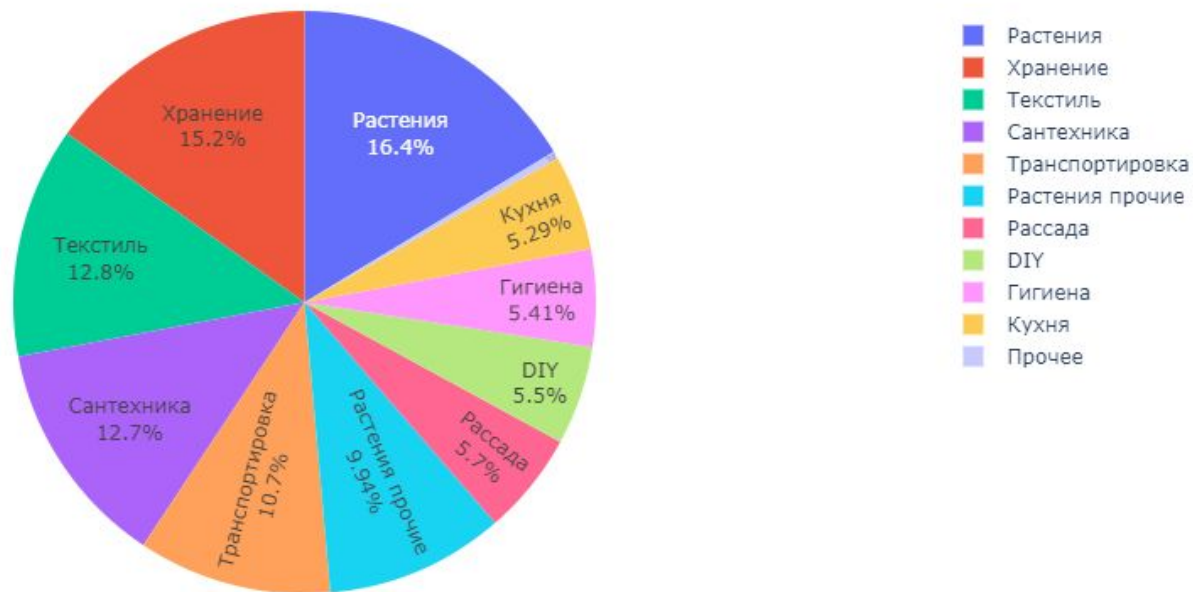


## Не типичный

- Профиль, обратный типичному- все покупатели датасета за вычетом типичных.
- Количество уникальных покупателей в сегменте 873 ( 38.4 % от всех)
- Количество транзакций в сегменте 2967 ( 68 % от всех)
- Лайфтайм медианный 1 день, максимальный - 137 дней.
- Суммарная выручка сегмента: 1 629 822 (53 % от всей)
- ТОП-4 категории по транзакциям - Растения, Рассада, Растения прочие и Сантехника- 82,26 % от всех транзакций сегмента
- ТОП-4 категории по выручке - Растения, Хранение, Текстиль, Сантехника- 57 % от всей выручки сегмента
- Средний чек - 378
- “Не типичный” сегмент очень похож на истинного садовода, в корзине которого часто и много бывают растения, но который приобретает товары и других категорий и иногда возвращается за повторными покупками.

# Профиль покупателя - Не типичный

Суммарная выручка по категориям

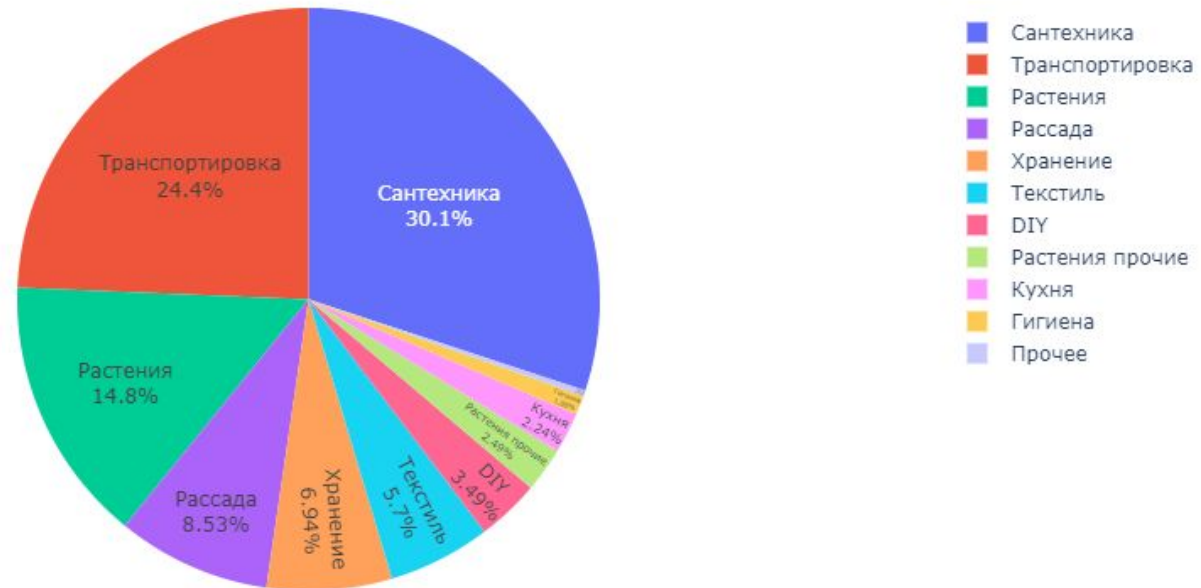


## Снабженец

- Покупатель возвращается в магазин за повторными покупками, в его корзине товары различных категорий более, чем в единичном экземпляре.
- Количество уникальных покупателей в сегменте 29 ( 1.3 % от всех)
- Количество транзакций в сегменте 405 ( 9,3 % от всех)
- Лайфтайм медианный 11 дней, максимальный - 126 дней.
- Суммарная выручка сегмента: 207 037 (6.7 % от всей)
- ТОП-4 категории по выручке - Сантехника, Транспортировка, Растения, Рассада- 78 % от всей выручки сегмента
- Средний чек - 284.
- Снабженцы наиболее активны в октябре-декабре, феврале- начале марта.
- Следует обратить внимание на эту группу лояльных клиентов: 1.3% клиентов приносит 6.7% выручки.

# Профиль покупателя - Снабженец

Суммарная выручка по категориям



# Рекомендации

Для более глубокого анализа и широкого поиска зависимостей, количественной оценки и качественных выводов и рекомендаций ( в том числе по разработке предложений покупателям) к датасету с транзакциями следует добавить

- исходный бизнес-план при открытии магазина и настоящие цели развития, без этого сложно оценить текущее состояние экономических показателей.
- информацию по транзакциям за предыдущие годы для оценки динамики,
- информацию по затратам и себестоимости для оценки прибыли в различных срезах данных,
- качественные и количественные ориентиры маркетинговой политики в целом и ближайшей маркетинговой кампании в частности.