

E-commerce — Выявление профилей потребления

На базе датасета /datasets/ecom_dataset_upd.csv.

Выполнила: Кудрявцева Ксения Вячеславовна

Дата: 16 сентября 2023г.

Контактные данные: +7(982)7029663 тел,
ksiyshasv2013@gmail.com.

Описание проекта.

Интернет-магазину «Пока все еще тут» требуется расчёт основных метрик и создание гипотез на основе полученных данных для того чтобы стать лучше для своих клиентов.

Задача:

- Сегментировать покупателей по профилю потребления (на основе истории их покупок).

План исследования:

1. Загрузка данных и подготовка их к анализу
 - открытие файла с данными и изучение общей информации
 - проверка на пропуски и типы данных, их корректировка при необходимости
 - проверка на дубликаты
 - промежуточные выводы
2. Исследовательский анализ данных
 - установление общего количества пользователей в логе и количества дней анализа
 - изучение количества товаров в заказе
 - изучение распределения стоимости заказа
 - промежуточные выводы
3. Разделение товаров на группы
 - лемматизация
 - промежуточные выводы
4. RFM-анализ для сегментирования клиентов
 - RFM-анализ
 - промежуточные выводы
5. Формирование и проверка статистических гипотез.
6. Выводы и рекомендации.

Общие выводы.

В ходе работы мною были описаны полученные данные, изменен тип данных там, где это было необходимо, созданы новые столбцы для дальнейших расчетов, удалены явные и неявные дубликаты, выявлены и удалены аномалии в данных. Проведен исследовательский анализ данных при помощи методов визуализации; товары разделены на категории при помощи лемматизации; проведен RFM-анализ для разделения на категории самих клиентов с учетом частоты совершаемых ими покупок, их стоимости и давности совершенных покупок; профили клиентов рассмотрены в разрезе категорий товаров и сезонов; проверены 2 заявленные гипотезы с помощью метода Манна-Уитни.

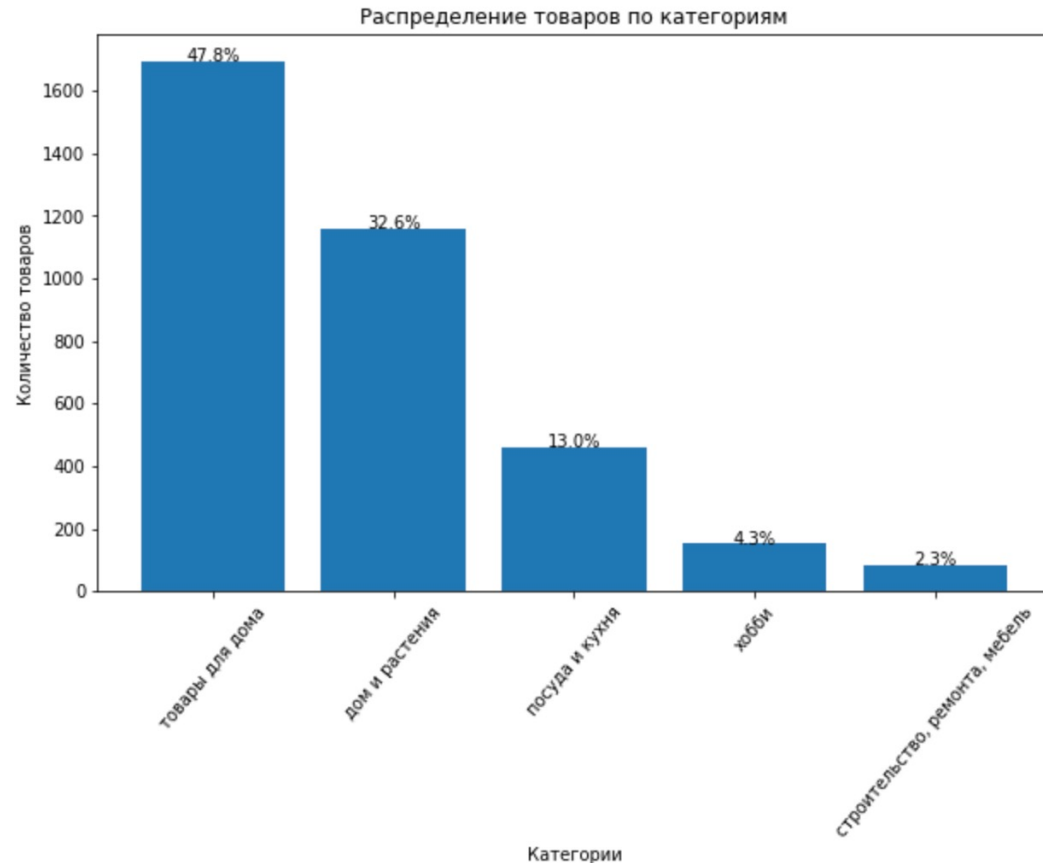
При помощи лемматизации товары были разделены мною на такие категории:

- товары для дома;
- дом и растения;
- посуда и кухня;
- строительство/ремонт/мебель;
- иное.

RFM-анализ выявил наличие клиентов в таких категориях как:

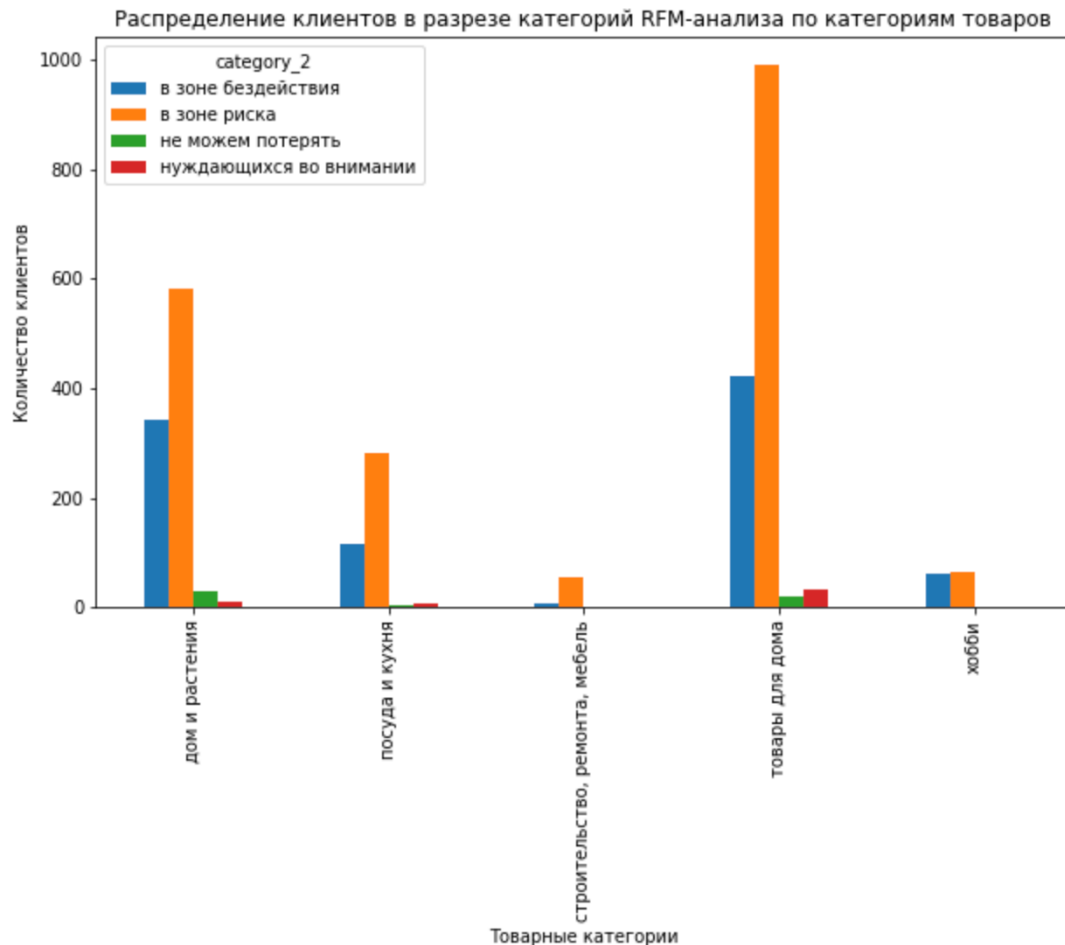
- в зоне риска - 1977 ед - совершили покупки в течении последних 250 дней с частотой примерно 1 раз в 6 дней, предпочитают такие категории товаров как "товары для дома" и "дом и растения", сезон активности - весь год, средний чек - 1180 уе;
- в зоне бездействия - 949 ед - совершили покупки в течении последних 250 дней с частотой примерно 1 раз в 6 дней, предпочитают такие категории товаров как "товары для дома" осенью, а весной товарам в категории "дом и растения", средний чек - 162 уе;
- нуждающиеся во внимании - 52 ед - совершили покупки в течении последних 250 дней с частотой примерно 1 раз в 6 дней, совершает покупки только осенью, в приоритете такие категории товаров "товары для дома" и "дом и растения", средний чек - 1383 уе;
- не можем потерять - 53 ед - совершили покупки в течении последних 250 дней с частотой примерно 1 раз в 6 дней, предпочитают покупать больше всего товаров осенью и весной и в категории "дом и растения", средний чек - 8086 уе.

Распределение товаров по категориям.



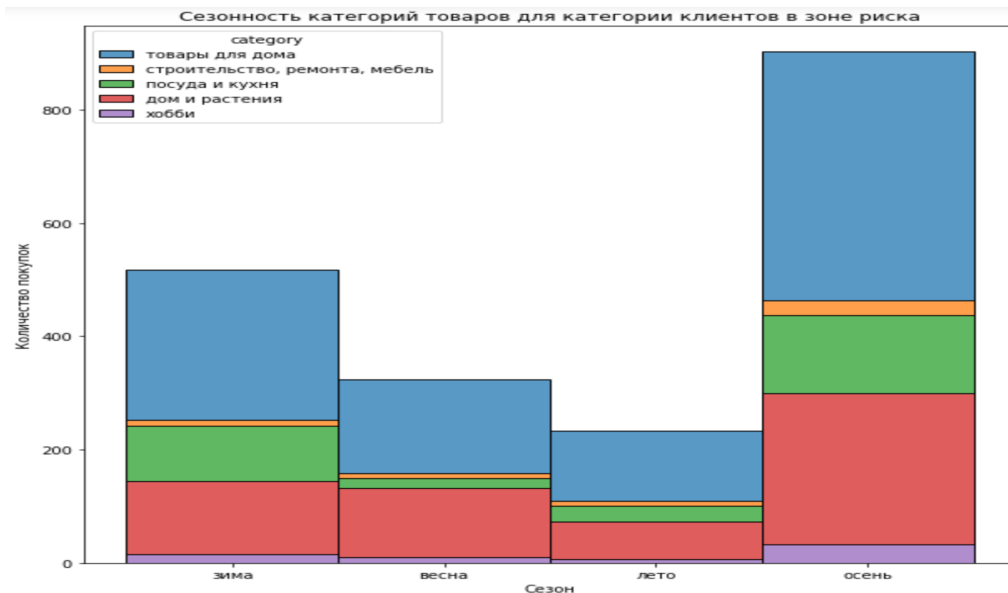
Вывод. С помощью библиотек `pymystem3` и `collections` мне удалось разделить товары на такие категории как: дом и растения, товары для дома, посуда и кухня, строительство/ремонт/мебель и иное. Больше всего клиенты покупают товары в категории Дом и растения (47% от всех товаров) и меньше всего в категории Строительство/ремонт/мебель (2,3% от всех товаров).

Распределение клиентов в разрезе категорий RFM-анализа по категориям товаров.

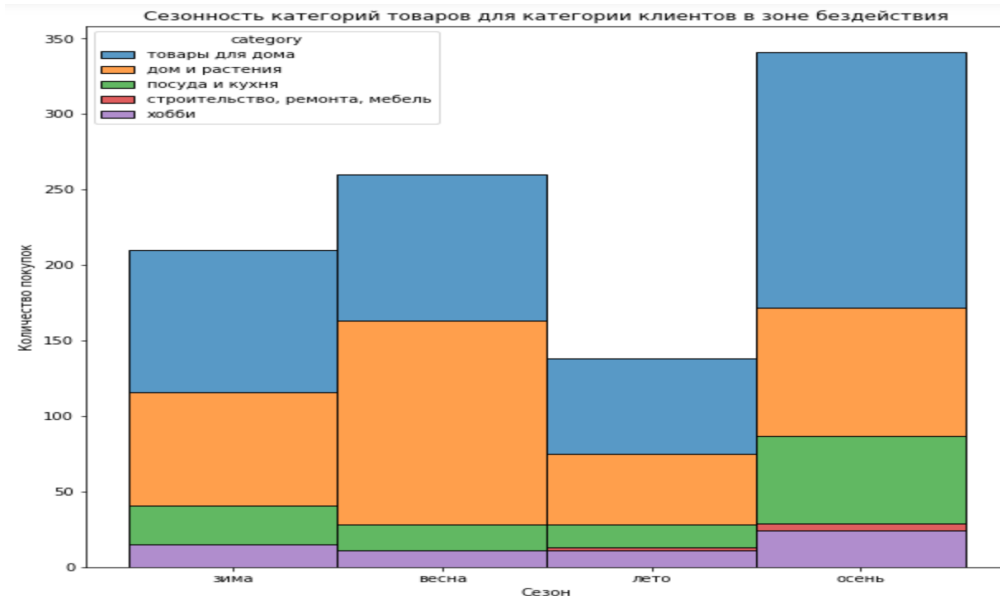


Вывод. По значениям, которые у нас есть по нашим клиентам мы получаем 4 категории: "в зоне риска" - 1977 человек, "в зоне бездействия" - 949 человек, "нуждающихся во внимании" - 52 человек, "не можем потерять" - 53 человек. В зоне риска находятся те пользователи, которые совершали покупки давно и в среднем количестве. В зоне бездействия находятся малоактивные пользователи, заходили давно и покупали мало. Не можем потерять - это люди, которые делали много покупок, но много дней назад. В зоне бездействия-находятся покупатели на грани оттока, которых важно не потерять и нужно активировать. Как мы видим из графиков выше и ниже излюбленной категорией товаров для всех сегментов покупателей является "дом и растения" и "товары для дома", по ним разделены покупатели с самым распространенным чеком - со средним - это сегмент "в зоне риска» и "нуждающиеся во внимании". Меньше всего внимания сегменты покупают товаров для строительства и ремонта, а так же есть единичные покупки в категории "иное".

Сезонность по категориям клиентов



Вывод. На графике по данным категории клиентов "в зоне риска" мы видим, в течение всего года они отдают предпочтение товарам в Категории "товары для дома", товары в категории "дом и растения" чуть чаще покупают осенью, на остальные категории товаров для данной категории клиентов сезонность не оказывает значительного влияния.



Клиенты же в категории "в зоне бездействия" осенью отдают предпочтение покупкам в категории "товары для дома", а весной товарам в категории "дом и растения".

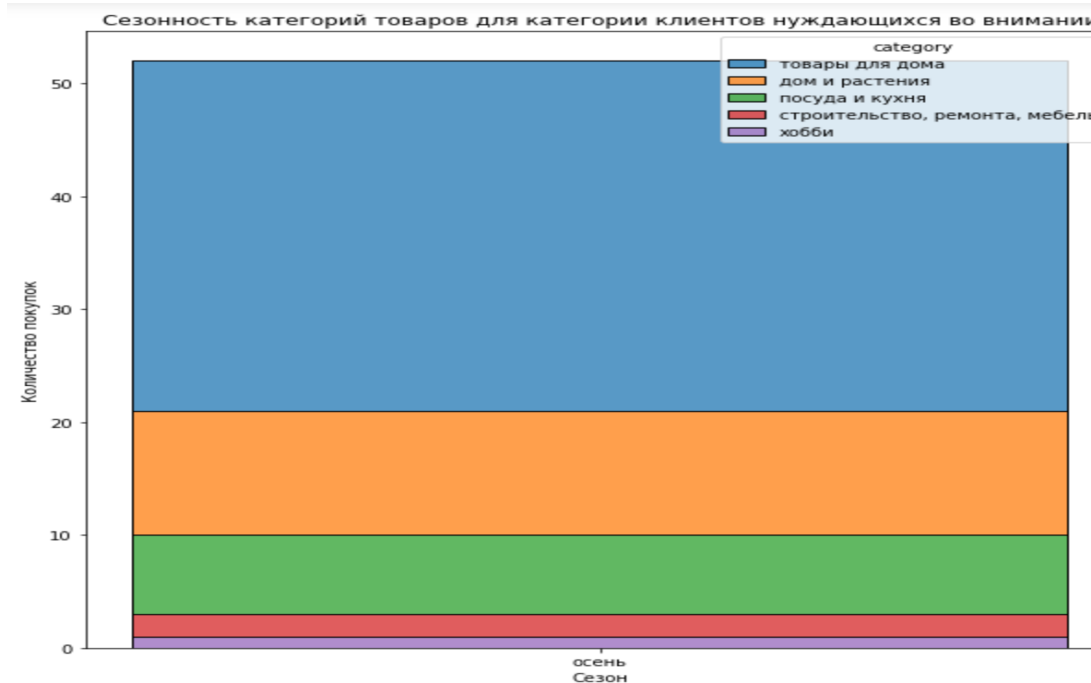
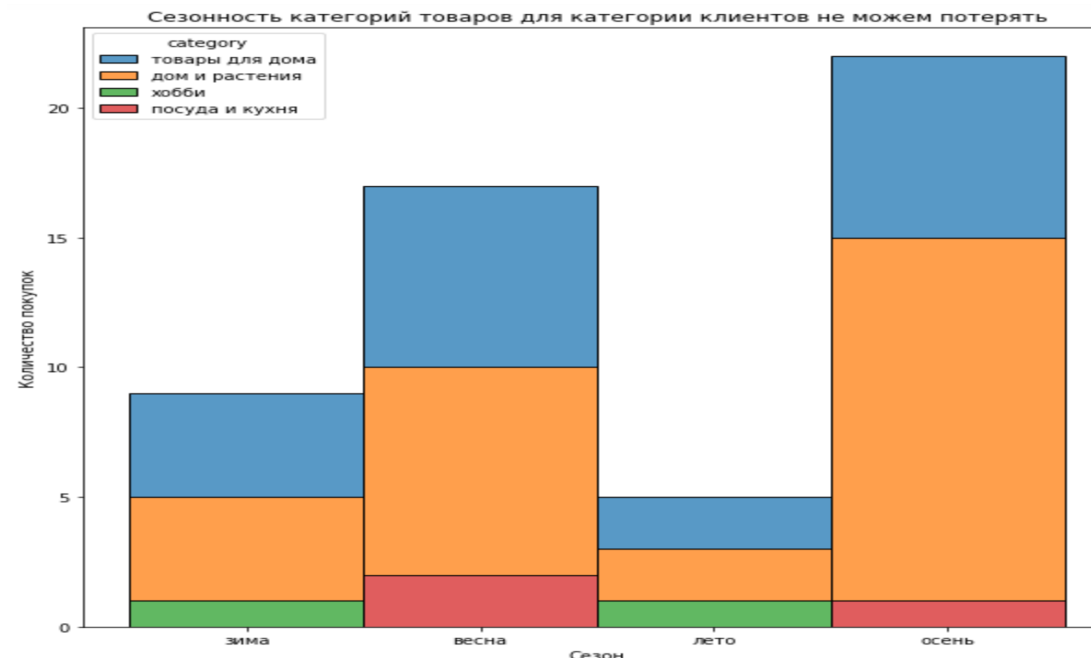
Источник: /datasets/ecom_dataset_upd.csv

Сезонность по категориям клиентов

Вывод. Клиенты из категории "не можем потерять" предпочитают покупать больше всего товаров осенью и весной и в категории "дом и растения".

Категория клиентов "нуждающихся во внимании" вообще совершает покупки только осенью, в приоритете у них так же категории "товары для дома" и "дом и растения".

Источник: /datasets/ecom_dataset_upd.csv



Рекомендации по рассылке:

- кластер "в зоне риска" включить в круглогодичную рекламную рассылку на категории "товары для дома" и "дом и растения";
- кластер "в зоне бездействия" - рассылка осень на категорию "товары для дома", а весной в категории "дом и растения";
- кластер "нуждающиеся во внимании" рассылка осень на категории "товары для дома" и "дом и растения";
- кластер "не можем потерять" рассылка осенью и весной и в категории "дом и растения".

Возможно увеличить количество продаваемых товаров в таких категориях как "товары для дома" и "дом и растения", это поможет повысить возвращаемость покупателей.