

E-commerce

Анализ товарного
ассортимента

Петрусеv И.А.

Техническое задание

- Для анализа представлен набор данных с историей транзакций интернет-магазина товаров для дома и быта "Пока все ещё тут".
- Необходимо:
 - предобработать данные,
 - провести исследовательский анализ данных,
 - проанализировать товарный ассортимент,
 - выделить основные категории товаров,
 - провести кластеризацию товаров с применением *machine learning*,
 - проверить статистическую гипотезу о равенстве средних чек в основных и дополнительных категориях.

Описание данных

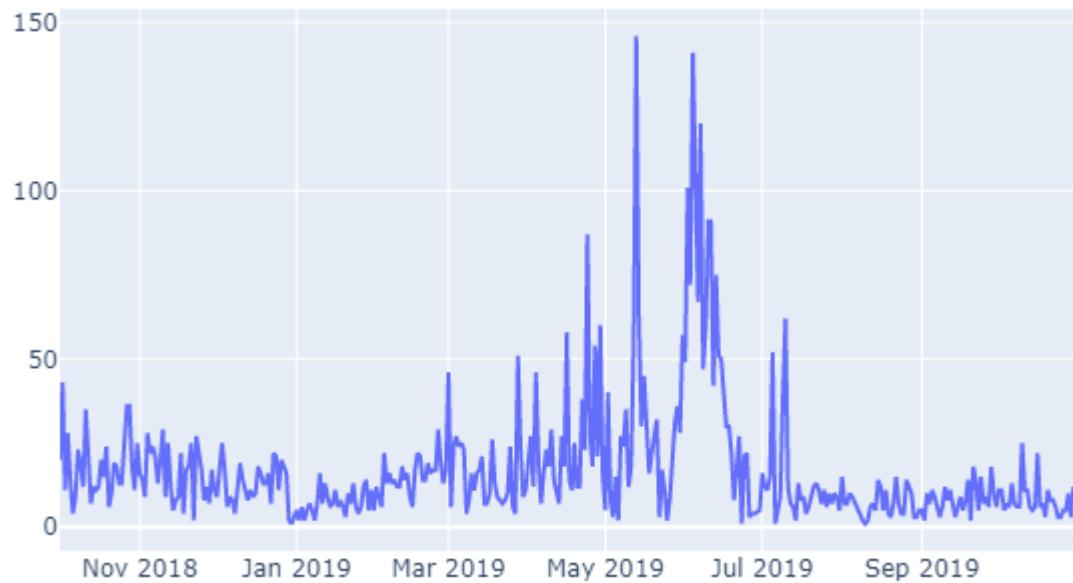
Для анализа представлены сведения о 6737 наблюдениях.

Файл **ecommerce_dataset.csv** содержит колонки:

- date – Дата заказа
- customer_id – Идентификатор покупателя
- order_id – Идентификатор заказа
- product – Наименование товара
- quantity – Количество товара в заказе
- price – Цена товара

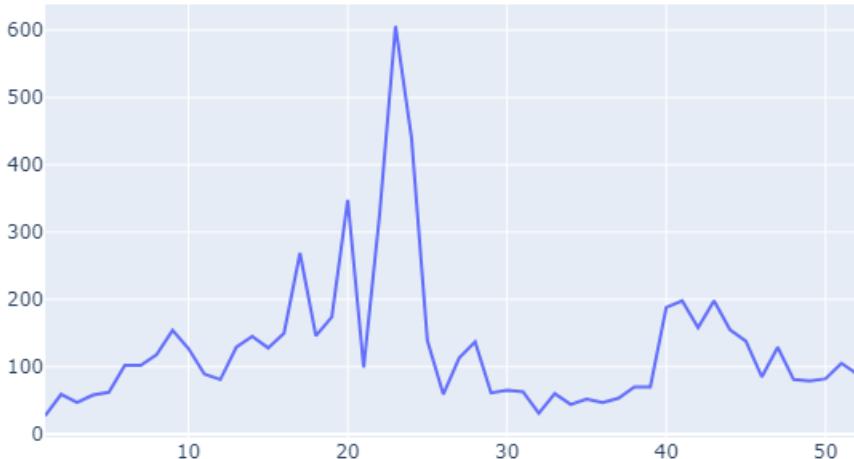
График, показывающий распределение покупок по дням в период с 01.10.2018 - 31.10.2019

- Пик продаж приходится на 13 мая 2019 и июнь 2019 года.
- Поскольку наш магазин позиционируется на узком ассортименте товаров, такое поведение можно объяснить тем, что в этот период начинается летний сезон, все активнее закупают товары для отдыха и дома.

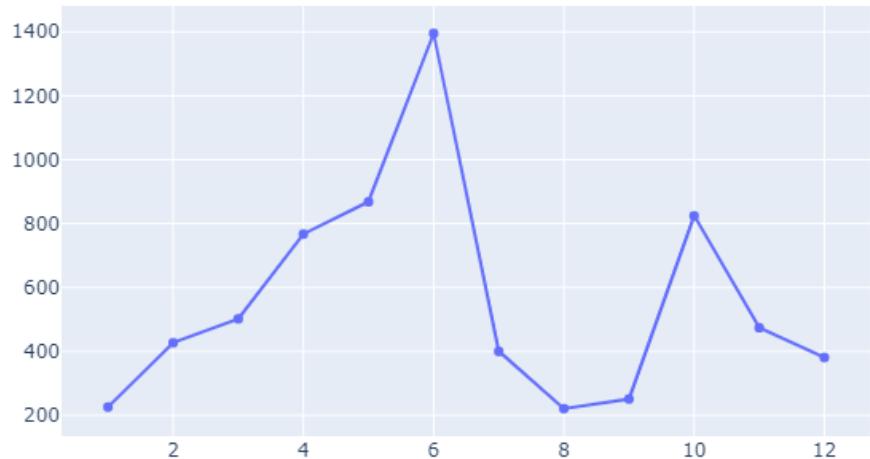


По информации с сайта "Пока все ещё тут"

Распределение покупок по неделям



Распределение покупок по месяцам



По информации с сайта "Пока все ещё тут"

Анализ товарных предпочтений

- По результатам проверки покупательской активности выявлено, что многие товары покупаются активнее других.
- После преобразований товаров и лемматизации получилось выделить 10 категорий товаров:
 - рассада
 - растение / цветы
 - фрукты/овощи
 - багаж
 - муляж
 - посуда
 - электрические приборы
 - хозтовары
 - система хранения
 - постельное белье/трикотаж.

Соотношение категорий товаров по количеству

- Чаще всего покупают товары в категории «растение/цветы»
- Товары в категории хозтовары покупают чуть реже, чем рассаду (15,2% против 20,4%).
- Товары категории посуда, фрукты/овощи, муляж, постельное белье/трикотаж и система хранения составляют примерно равный процент от общих продаж в магазине (менее 5%).



По информации с сайта "Пока все ещё тут"

Соотношение категорий товаров по доходности

- Самой доходной категорией является хозтовары и багаж, хотя товары в категории багаж покупают в два раза реже.
- Доходы от электротоваров незначительно превышают доходы от продажи растений и цветов, тогда как процентном соотношении растения покупают в 7 раз чаще.
- 6 крайних категорий товаров приносят менее 20% общих доходов.



По информации с сайта "Пока все ещё тут"

Выводы по категориям

- Несмотря на то, что по количеству товаров лидируют одни категории, основную прибыль приносят другие.
- Так, 15,2% проданных товаров в категории **хозтовары** приносят бо'льшую часть прибыли – 28,7%.
- Товары из категории **багаж**, составляющие от проданного объема товаров всего 6,87%, приносят 27,2% от общего дохода.
- **Электрические приборы**, составляющие 5,86% от числа продаж, приносят 12,6% прибыли.
- **Растения и цветы**, которые покупают относительное большинство клиентов, приносят всего 11,7% прибыли.
- Предлагается эти четыре категории выделить в качестве основных, поскольку на них приходится бо'льшая часть приносимой компании прибыли. На них, при проектировании маркетинговой компании и следует обратить первостепенное внимание.
- К дополнительным категориям можно отнести: **постельное белье/трикотаж, система хранения, посуда, рассада, муляж, фрукты / овощи.**

Проверка статистической гипотезы и вывод

- ❑ По результатам проверки гипотезе о равенстве средней выручке для покупателей в основных и дополнительных категориях, нулевая гипотеза была отвергнута (р-значение: 0.0165) - средняя выручка клиентов в основных категориях отличается от дополнительных категорий.
- ❑ Средний чек среди покупателей в основных категориях выше, чем в дополнительных.
- ❑ Очевидно, что при планировании маркетинговой компании следует сделать акцент на такие категории:
 - хозтовары,
 - багаж,
 - электрические приборы,
 - растения и цветы.

Работа с документацией

- Лекции и вебинары в Яндекс.Практикум
- Использование регулярных выражений
- Работа со словарями
- Лемматизация и выделение категорий
- Документация Plotly для построения графиков
- Кластерный анализ - кластеризация товаров