

Исследование данных супермаркета

(поиск инсайтов, составление рекомендаций стейкхолдерам, построение дашборда)

Дипломная работа по программе «Аналитик BI: с нуля до middle»

Мукаева Г. А.

Группа: ABU-77

2025 г.



Описание проекта

Задача проекта - комплексный анализ данных по заказам с целью увеличения дохода

- увидеть распределение объемов продаж
- выявить лидеров и наиболее низкие показатели
- провести анализ эффективности типов доставки
- сегментировать клиентов (RFM-анализ)

Стейкхолдеры:

- директор продаж и маркетинга
- отдел логистики
- финансовый департамент



Описание проекта

Гипотезы:

- более быстрый тип доставки увеличивает повторные покупки
- клиенты сегмента «Corporate» генерируют больший средний чек

Метрики для оценки:

- объем продаж
- темп роста продаж
- доля объема продаж
- среднее время доставки
- RFM



Описание данных и их предобработки

Датасет «Superstore Sales Dataset» на 9800 строк имеет высокое качество данных, без явных аномалий и зависимостей

Проблемы: отсутствует количество и стоимость единицы заказа

| Атрибут | Пример данных | Описание |
|---------------|-----------------------------------|------------------------|
| Row ID | 1 | Номер строки |
| Order ID | CA-2017-152156 | Номер заказа |
| Order Date | 08/11/2017 | Дата заказа |
| Ship Date | 11/11/2017 | Дата доставки |
| Ship Mode | Second Class | Тип доставки |
| Customer ID | CG-12520 | Идентификатор клиента |
| Customer Name | Claire Gute | Имя клиента |
| Segment | Consumer | Сегмент |
| Country | United States | Страна |
| City | Henderson | Город |
| State | Kentucky | Штат |
| Postal Code | 42420.0 | Почтовый индекс |
| Region | South | Регион |
| Product ID | FUR-BO-10001798 | Идентификатор продукта |
| Category | Furniture | Категория |
| Sub-Category | Bookcases | Подкатегория |
| Product Name | Bush Somerset Collection Bookcase | Название продукта |
| Sales | 261.9600 | Цена |

Шаги преобразования данных

- исключены лишние данные («Postal Code», «Country»)
- приведены к единому naming convention
- проверены на пустые значения и дубликаты
- изменены типы данных
- сохранена витрина данных



Этапы анализа

1. преобразование и очистка данных в Visual Studio Code
2. нормализация данных для хранения в базе данных
3. выгрузка данных
4. создание моделей
5. анализ витрины данных в Tableau Public
6. выводы и рекомендации на основе анализа

| | row_id | order_id | order_date | ship_date | ship_mode | customer_id | customer_name | segment | city | state | region | product_id | category | sub_category | product_name | sales |
|------|--------|----------------|------------|------------|----------------|-------------|------------------|-----------|-----------------|------------|---------|-----------------|-----------------|--------------|---|----------|
| 0 | 1 | CA-2017-152156 | 2017-11-08 | 2017-11-11 | Second Class | CG-12520 | Claire Gute | Consumer | Henderson | Kentucky | South | FUR-BO-10001798 | Furniture | Bookcases | Bush Somerset Collection Bookcase | 261.9600 |
| 1 | 2 | CA-2017-152156 | 2017-11-08 | 2017-11-11 | Second Class | CG-12520 | Claire Gute | Consumer | Henderson | Kentucky | South | FUR-CH-10000454 | Furniture | Chairs | Hon Deluxe Fabric Upholstered Stacking Chairs,... | 731.9400 |
| 2 | 3 | CA-2017-138688 | 2017-06-12 | 2017-06-16 | Second Class | DV-13045 | Darrin Van Huff | Corporate | Los Angeles | California | West | OFF-LA-10000240 | Office Supplies | Labels | Self-Adhesive Address Labels for Typewriters b... | 14.6200 |
| 3 | 4 | US-2016-108966 | 2016-10-11 | 2016-10-18 | Standard Class | SO-20335 | Sean O'Donnell | Consumer | Fort Lauderdale | Florida | South | FUR-TA-10000577 | Furniture | Tables | Bretford CR4500 Series Slim Rectangular Table | 957.5775 |
| 4 | 5 | US-2016-108966 | 2016-10-11 | 2016-10-18 | Standard Class | SO-20335 | Sean O'Donnell | Consumer | Fort Lauderdale | Florida | South | OFF-ST-10000760 | Office Supplies | Storage | Eldon Fold 'N Roll Cart System | 22.3680 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 9795 | 9796 | CA-2017-125920 | 2017-05-21 | 2017-05-28 | Standard Class | SH-19975 | Sally Hughsby | Corporate | Chicago | Illinois | Central | OFF-BI-10003429 | Office Supplies | Binders | Cardinal HOLDIt! Binder Insert Strips,Extra St... | 3.7980 |
| 9796 | 9797 | CA-2016-128608 | 2016-01-12 | 2016-01-17 | Standard Class | CS-12490 | Cindy Schnelling | Corporate | Toledo | Ohio | East | OFF-AR-10001374 | Office Supplies | Art | BIC Brite Liner Highlighters, Chisel Tip | 10.3680 |
| 9797 | 9798 | CA-2016-128608 | 2016-01-12 | 2016-01-17 | Standard Class | CS-12490 | Cindy Schnelling | Corporate | Toledo | Ohio | East | TEC-PH-10004977 | Technology | Phones | GE 30524EE4 | 235.1880 |
| 9798 | 9799 | CA-2016-128608 | 2016-01-12 | 2016-01-17 | Standard Class | CS-12490 | Cindy Schnelling | Corporate | Toledo | Ohio | East | TEC-PH-10000912 | Technology | Phones | Anker 24W Portable Micro USB Car Charger | 26.3760 |
| 9799 | 9800 | CA-2016-128608 | 2016-01-12 | 2016-01-17 | Standard Class | CS-12490 | Cindy Schnelling | Corporate | Toledo | Ohio | East | TEC-AC-10000487 | Technology | Accessories | SanDisk Cruzer 4 GB USB Flash Drive | 10.3840 |

9800 rows x 16 columns



Структура решения

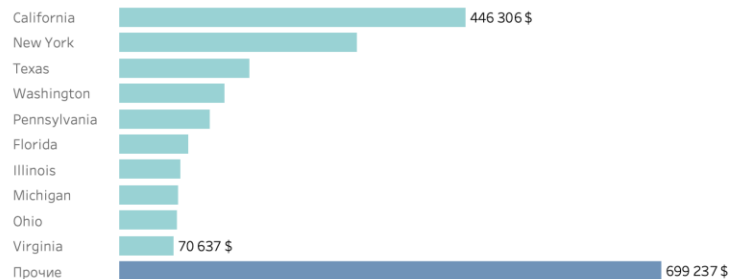


Анализ продаж супермаркета

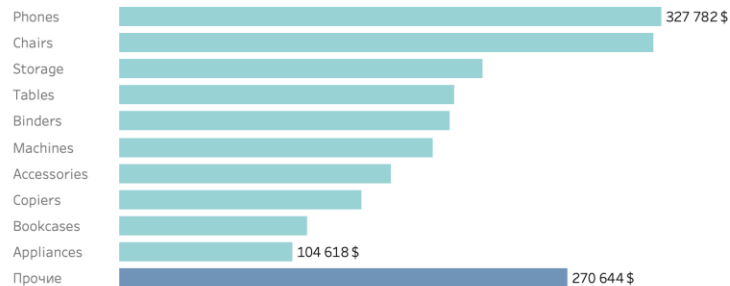
Выручка

2 261 537 \$

Штаты по объему продаж ▲



Подкатегории по объему продаж ▲



Направление сортировки Год Регион Тип доставки ➔

Максимум (All) (All) (All)

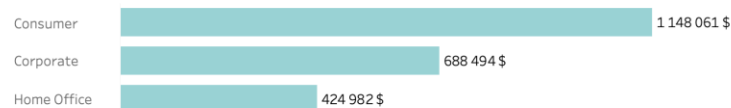
Количество заказов

4 922

Города по объему продаж ▲



Продажи по сегментам



Продажи по категориям



Структура решения

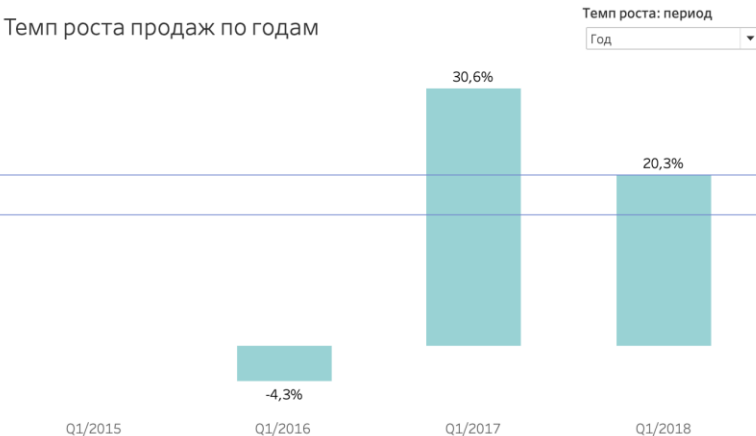


Анализ доставки и продаж

Динамика продаж и прогноз



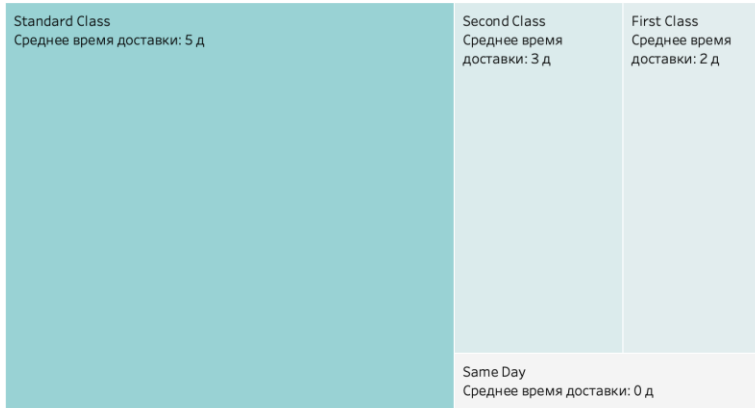
Темп роста продаж по годам



Сегмент: (All) Подкатегория: (All) Категория: (All) ← →

Доля повторных покупок в каждом типе доставки

| | |
|----------------|--------|
| Same Day | 100,0% |
| Second Class | 99,8% |
| Standard Class | 99,6% |
| First Class | 99,6% |

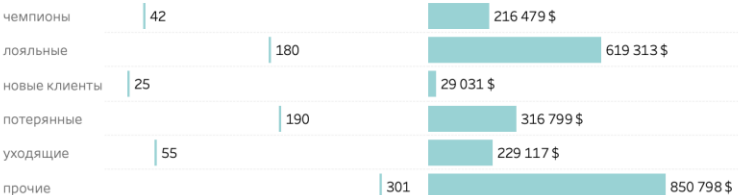


Структура решения



RFM-анализ показатели считаются на 2019-01-01

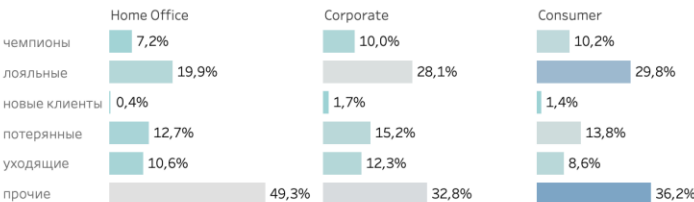
Количество клиентов и объем продаж



Ценность клиентского сегмента

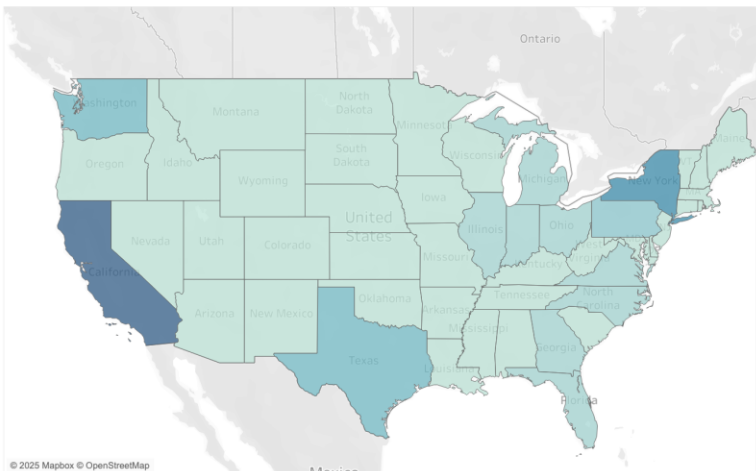


Объем продаж по сегментам



Сегмент (All) Год (All) Тип доставки (All)

Продажи по штатам



По клиентам

| RFM-group | Customer Name | Recency | Frequency | Monetary | |
|-----------|-----------------|---------|-----------|----------|-----|
| чемпионы | Alan Hwang | 8 | 9 | 4 805 | 555 |
| | Annie Thurman | 15 | 10 | 3 832 | 554 |
| | Arianne Irving | 15 | 10 | 4 376 | 555 |
| | Ben Ferrer | 18 | 11 | 5 908 | 555 |
| | Ben Peterman | 3 | 9 | 3 676 | 554 |
| | Brosina Hoffman | 25 | 8 | 6 255 | 545 |
| | Carlos Soltero | 27 | 11 | 3 037 | 554 |
| | Chris Cortes | 22 | 11 | 3 670 | 554 |



Выводы и рекомендации

Выводы:

- общая выручка более 2 млн. долларов
- общее количество заказов около 5 тысяч
- лидеры по продажам и заказам: штат Калифорния, город Нью-Йорк, сегмент – «Consumer», категория – технологии, подкатегория – телефоны
- имеют наиболее низкие показатели: штат Северная Дакота, город Миссури, сегмент – домашний офис, категория – офисные принадлежности, подкатегория – крепежи
- средний чек сегмента «Corporate» - 467,77 \$, гипотеза неверная
- заметна положительная динамика роста продаж
- чаще всего пользуются стандартной доставкой
- доля повторных покупок в каждом типе доставок практически одинаковая – гипотеза неверная
- резкий скачок продаж в 2017 году
- заметно снижение объема продаж в первом квартале каждого года
- RFM-анализ: по объему продаж лидирует группа лояльных, по количеству клиентов лидирует группа потерянных, количество новых клиентов год от года падает



Выводы и рекомендации

Рекомендации:

- запустить программы лояльности и персональные предложения для лидеров
- оптимизировать логистику в лидирующих штатах
- развивать ассортимент
- добавить эксклюзивные предложения
- развивать кросс-продажи сопутствующих товаров
- проанализировать причины спада продаж и заказов
- оптимизировать ассортимент
- изучить причины скачка продаж в 2017 году и повторить эффективные акции и промо
- проанализировать предложения конкурентов по срокам доставки
- запустить спецакции по ускоренной доставке в пиковые периоды
- проанализировать триггеры повторных продаж
- проанализировать причины спада продаж по кварталам
- запустить спецакции в «низкий» сезон
- разработать методы возвращения потерянных клиентов
- усилить рекламу для новых клиентов



**Спасибо за
внимание!**

