

АНАЛИТИЧЕСКИЙ ОТЧЕТ

Магазин натуральной косметики -

Анализ продаж и покупателей

1. РЕЗЮМЕ ПРОЕКТА

Автор: Галия Исламова

Дата: 10.10.2025

1.1 Цель анализа:

Оценить продажи, поведение клиентов и продуктовую эффективность интернет-магазина натуральной косметики, а также выявить факторы роста выручки и удержания клиентов.

1.2 Данные

- 5 000 клиентов;
- 8 691 заказ;
- 15 929 позиций в заказах;
- 115 продуктов;
- Период: 01.01.2024 - 01.10.2025.

1.3 Ключевые выводы

- Значительная часть выручки формируется половиной покупателей (ABC-анализ);
- Сезонность продаж не наблюдается;
- Определенные категории продуктов дают наиболее высокий вклад в выручку;
- Уровень оттока клиентов составляет 51%, что указывает на потенциал программ удержания

1.4 Ключевые рекомендации

- Фокус на high-value клиентах
- Оптимизация скидочной политики.
- Развитие наиболее маржинальных категорий.
- Запуск retention-механик для новых клиентов.

2. БИЗНЕС-КОНТЕКСТ И ПОСТАНОВКА ЗАДАЧ

2.1 Описание бизнеса

Интернет-магазин натуральной косметики с доставкой, широким ассортиментом и системой скидок.

2.2 Основные аналитические вопросы:

- Кто приносит основную выручку?
- Какие продукты и категории наиболее эффективны?
- Как ведут себя клиенты во времени?
- Насколько скидки влияют на доход?
- Есть ли признаки оттока и сезонности?

3. ДАННЫЕ И СТРУКТУРА ХРАНИЛИЩА

3.1 Источники данных

Данные синтетически сгенерированы с использованием Python.

3.2 Таблицы:

- customers - 5 000 строк
- products - 115 строк
- orders - 8 691 строк
- order_items - 15 929 строк

3.3 Логическая модель

- Клиент → Заказы → Позиции заказа → Продукты
- Связи one-to-many



Рис 1. ERD

4. АЛТИЧЕСКИЙ ПАЙПЛАЙН И ПОДГОТОВКА ДАННЫХ

4.1 Этапы пайплайна:

- Генерация данных
- Валидация и базовая очистка
- Объединение таблиц
- Формирование аналитической витрины

4.2 Анализ и визуализация

- Ключевые шаги ETL (etl.ipynb):
 - Проверка типов и пропусков
 - Создание признака первой покупки
 - Подготовка дат для когорт
 - Формирование финальной таблицы cosmetics_analytics_ready.csv

Существенных проблем качества данных выявлено не было.

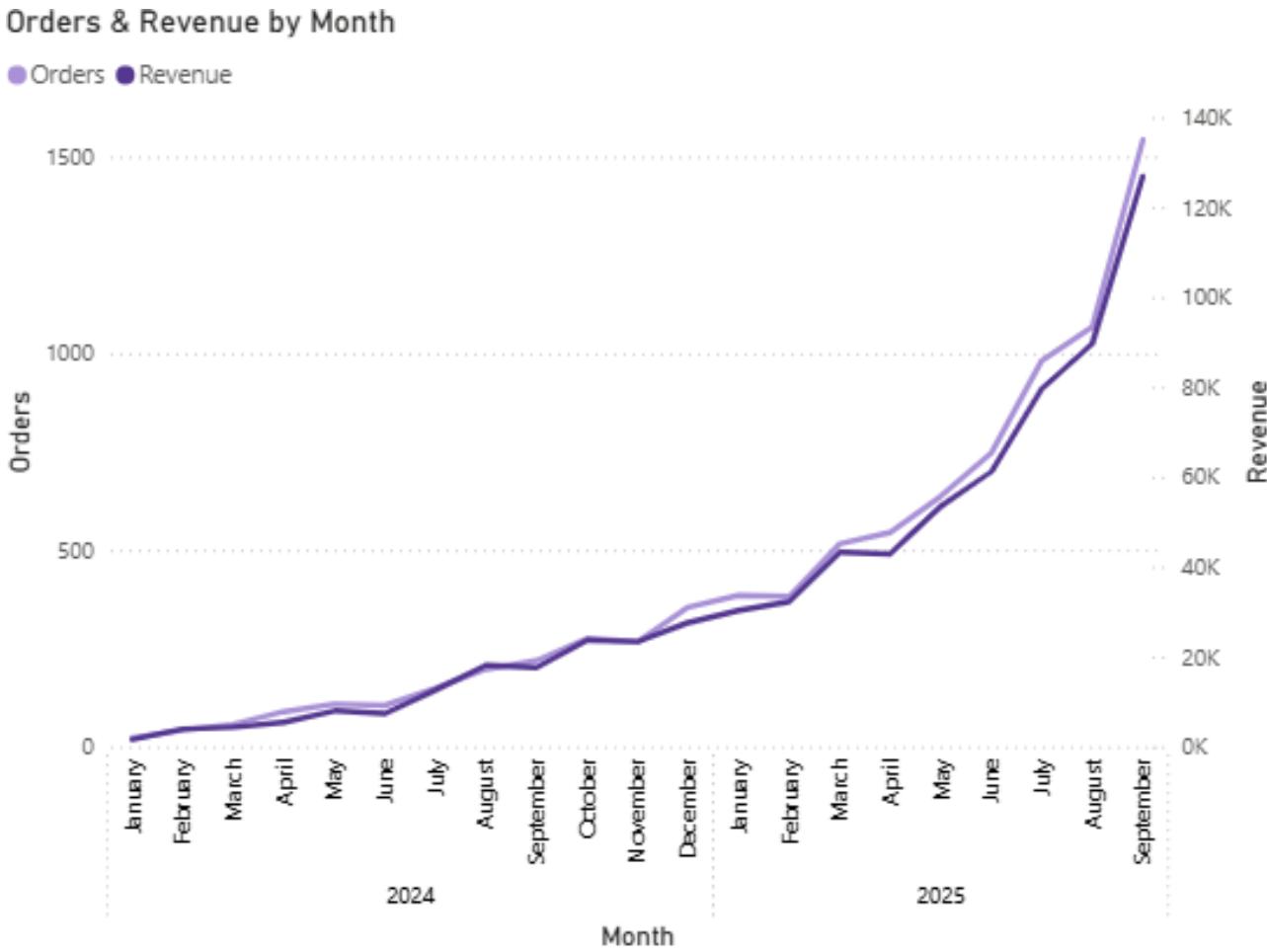
5. ОПИСАТЕЛЬНАЯ АНАЛИТИКА ПРОДАЖ

5.1 Общие показатели

- Total customers (5000) - общее количество клиентов
- Active customers (3369) - активные покупатели
- Total revenue (713,53 тыс. евро) - общая выручка
- Total orders (8691) - общее количество заказов
- Average order value (82,10 евро) - средний чек за заказ
- Average items per order (2,56) - среднее количество элементов в заказе

5.2 Динамика продаж

График показывает продажи и выручку по месяцам



Описание графика и выводы:

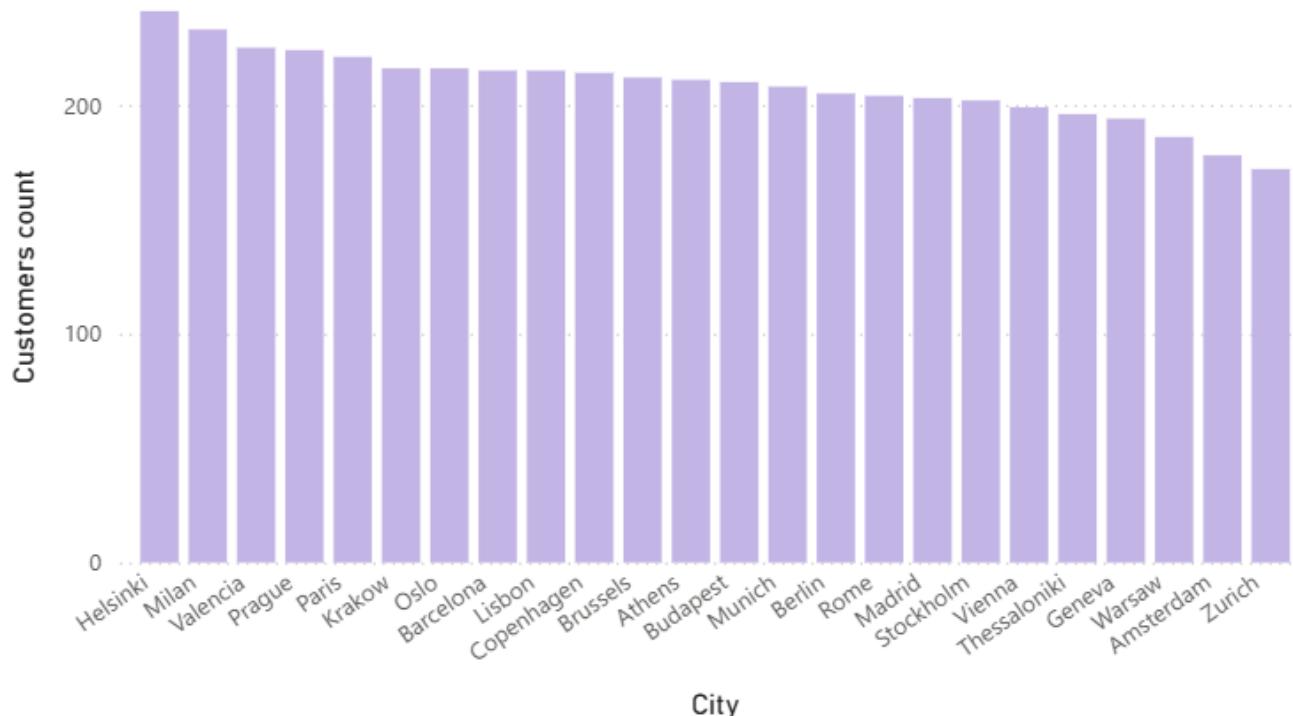
- Устойчивый и экспоненциальный рост:** Оба показателя демонстрируют сильный и устойчивый восходящий тренд на протяжении всего наблюдаемого периода. Рост начинается как линейный в 2024 году, но ускоряется и становится более крутым, приближаясь к экспоненциальному в 2025 году.
- Параллельная динамика:** Линии заказов и выручки идут практически параллельно, что говорит о сохранении среднего чека или средней стоимости заказа на относительно стабильном уровне на протяжении всего периода. Рост выручки напрямую обусловлен ростом количества заказов.
- Значительное масштабирование:** за 21 месяц количество заказов выросло с нескольких сотен до более чем 1500, а выручка с нескольких десятков тысяч до 140 тысяч евро.
- Отсутствие выраженной сезонности:** ярко выраженных сезонных спадов и пиков, не наблюдается. Общий тренд роста доминирует над любыми потенциальными сезонными колебаниями.

6. ГЕОГРАФИЧЕСКИЙ АНАЛИЗ КЛИЕНТОВ И ВЫРУЧКИ

6.1 Клиенты по городам:

График отображает количество клиентов по городам варьирующийся от 172 до 241.

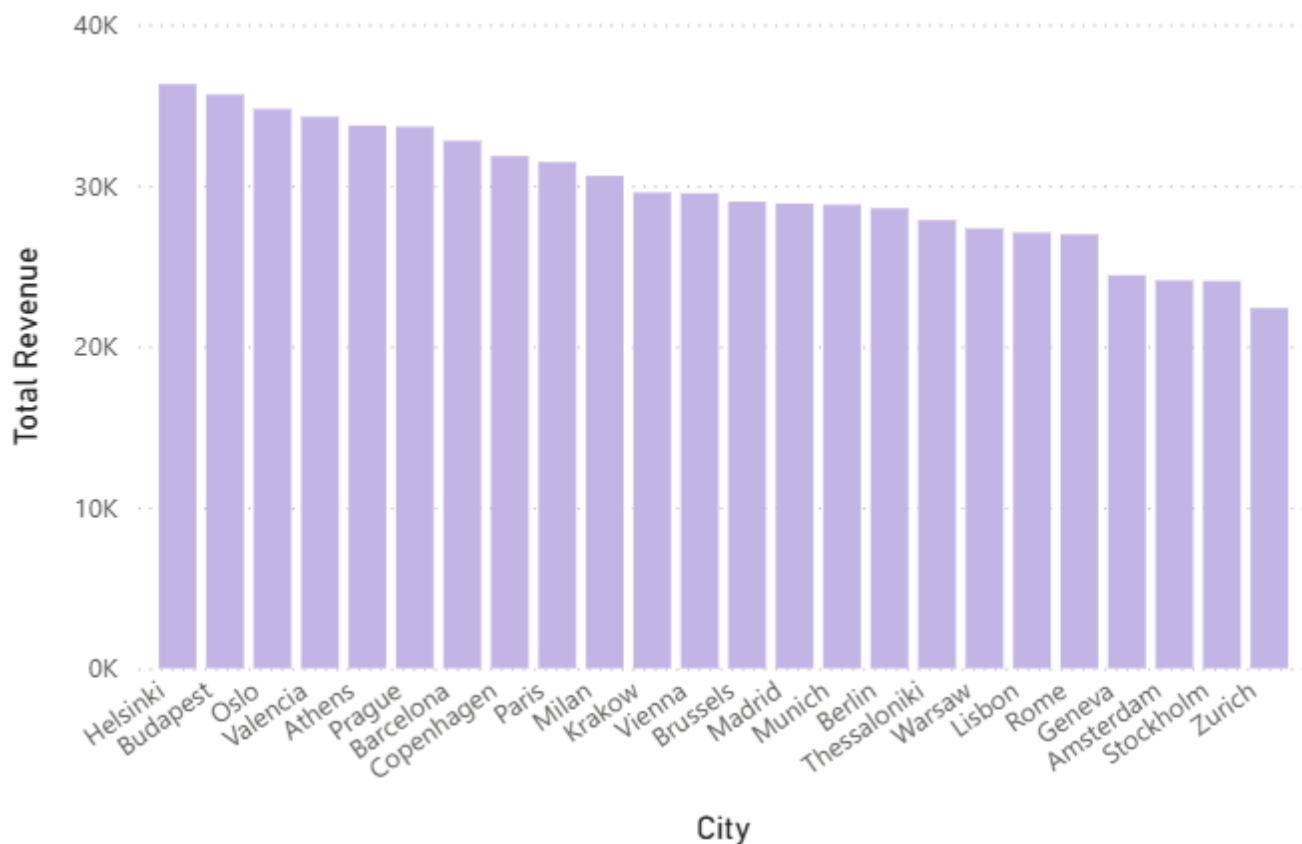
Customers by City



6.2 Выручка по городам:

График отображает общую выручку масштабированную от 22382 до 36295

Total Revenue by City



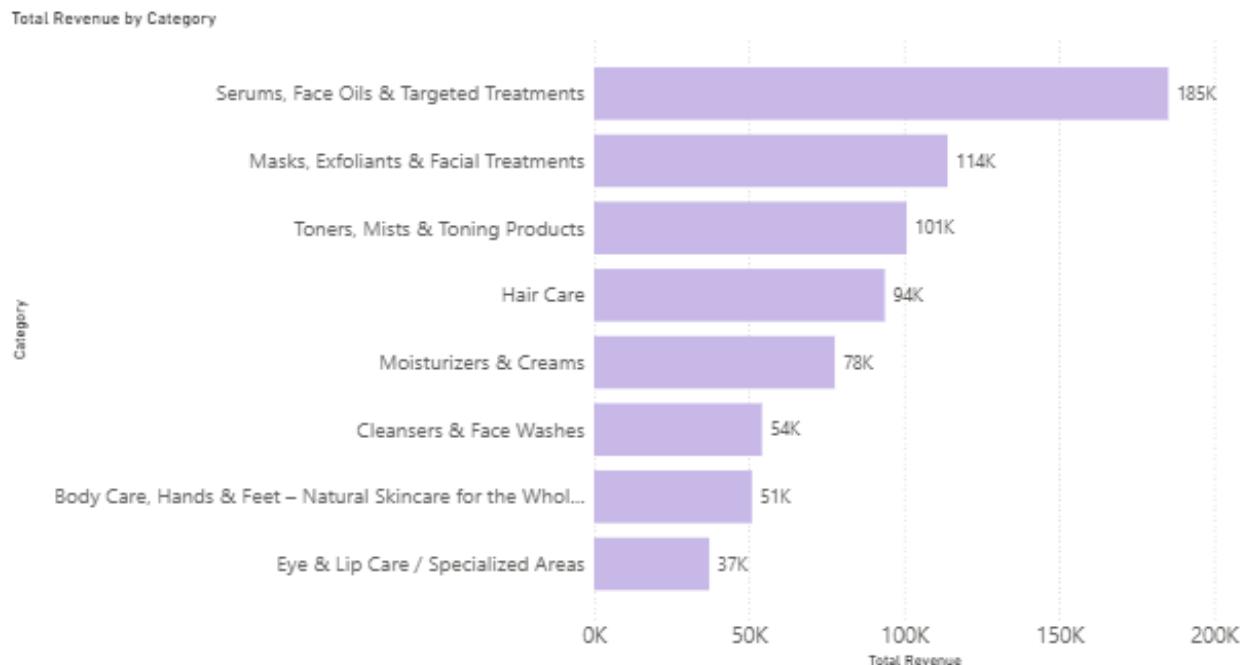
Описание графиков и общие выводы:

- **Критически важный показатель ARPU (средняя выручка на покупателя):** несовпадение позиций городов в рейтингах прямо указывает на необходимость расчета ARPU для точной оценки эффективности. Например:
 1. Милан: занимает второе место по количеству клиентов, но лишь десятое место в общей выручке. Говорит о том, что покупатели в Милане тратят значительно меньше в среднем, чем в других городах.
 2. Будапешт: занимает второе место по общей выручке, но находится на двенадцатом месте по количеству клиентов. Это указывает на очень высокий чек и высокую покупательскую способность покупателей в Будапеште.
- Выявление типов рынков:
 3. Высокоэффективные (Высокий ARPU, много клиентов): Хельсинки – лидер в обоих рейтингах, является наиболее ценным и приоритетным рынком.
 2. Потенциальные (Высокий ARPU, мало клиентов): Будапешт, Осло – рынки с «богатыми» клиентами, где есть потенциал для привлечения большего числа похожих клиентов.
 3. Проблемные (Низкий ARPU, много клиентов): Милан, Париж – здесь много клиентов, но они приносят меньше денег, что требует изучения их покупательского поведения.
 4. Стабильные (Средние показатели): большинство остальных городов демонстрируют эффективность.
- Географические тренды: наблюдается высокая концентрация выручки и клиентов в Северной Европе (Хельсинки, Осло, Копенгаген), тогда как некоторые города в Южной Европе (Милан, Рим, Мадрид) имеют больше клиентов, но относительно меньшую выручку из каждого из них.

7. ПРОДУКТОВАЯ АНАЛИТИКА

7.1 Выручка по категориям продуктов

Диаграмма иллюстрирует общую выручку по категориям продуктов, показывая какие из них приносят наибольший доход.



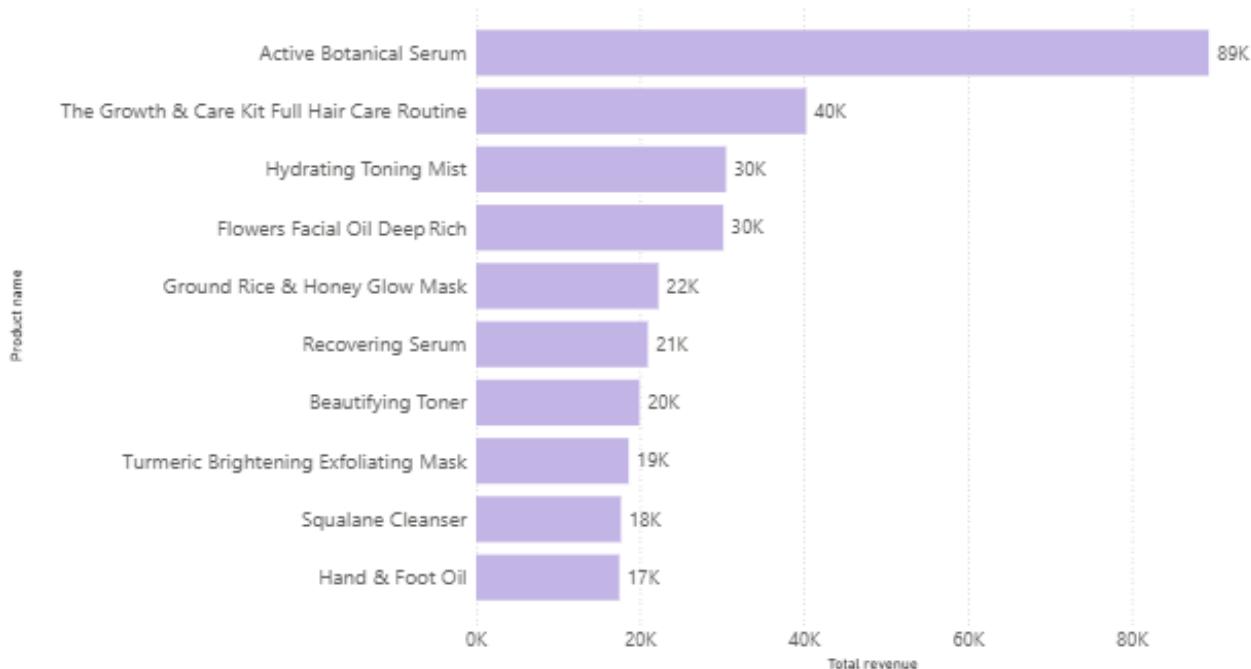
Описание графика и выводы:

- Лидеры продаж: «Serums, Face Oils & Targeted Treatments» () является абсолютным лидером по выручке, значительно опережая все остальные и достигая отметки ... тыс. евро
- Второй эшелон: «Masks, Exfoliants & Facial Treatments» () и «Toners, Mists & Toning Products» () – формируют второй уровень выручки, каждая приносит около 100 тыс. евро
- Средняя выручка: Категории «Hair Care» () и «Moisturisers & Creams» () имеют средние показатели в диапазоне 60-80 тыс. евро.
- Отстающие категории: Категории «Cleansers & Face Washes», «Body Care, Hands & Feet - Natural Skincare for Whole Body» и «Eye & Lip Care/Specialized Areas » приносят наименьшую выручку, около 30-50 тыс. евро.
- Неравномерное распределение: выручка распределена крайне неравномерно: три верхние категории генерируют большую часть общего дохода, в то время как нижние категории значительно отстают.

7.1 Топ продукты по выручке

График отображает распределение выручки среди 10 самых продаваемых товаров компании с сортировкой по убыванию дохода.

Top 10 products by Revenue



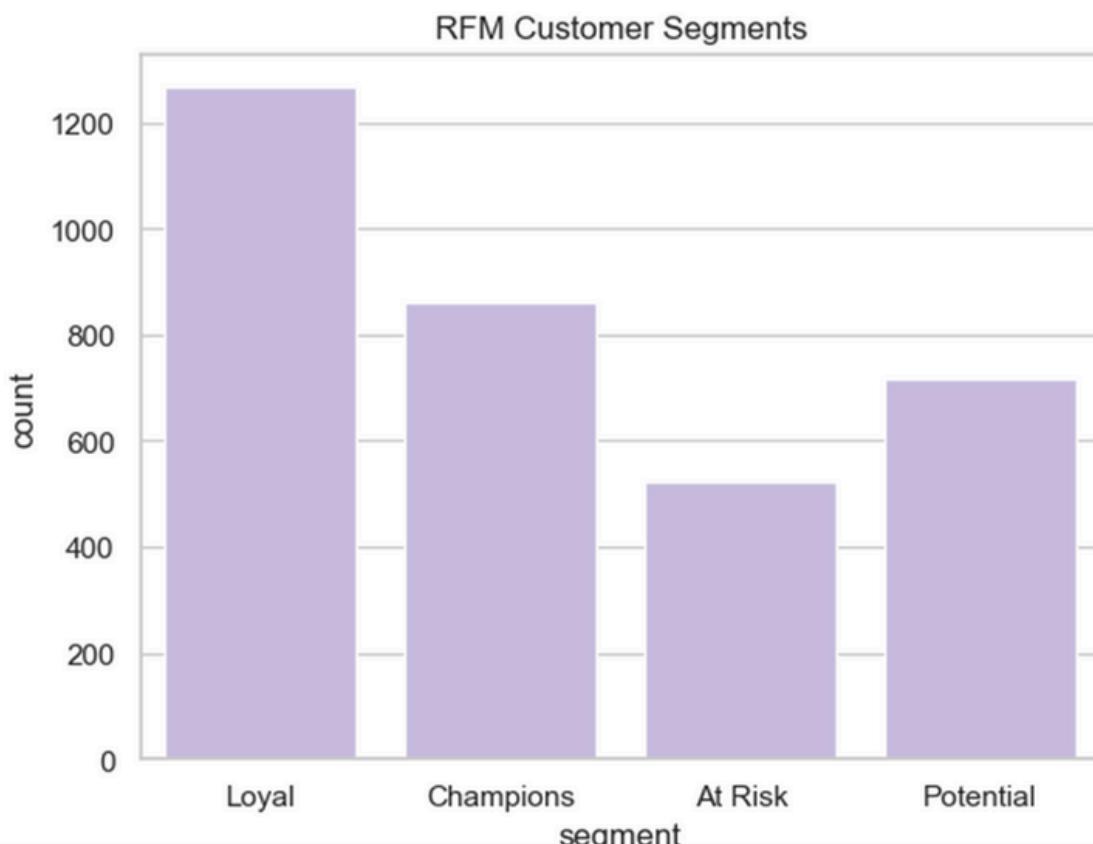
Выводы:

- Лидер продаж «Active Botanical Serum» значительно опережает все остальные товары по объему выручки. Его продажи составляют 89 тыс. евро, в то время как его ближайший конкурент, «The Growth & Care Kit Full hair Care Routine», приносит примерно в два раза меньше дохода (40 тыс. евро.)
- Значительный разрыв между позициями: наблюдается резкое падение выручки от первого ко второму продукту, и далее тренд снижения продолжается. Это указывает на высокую концентрацию дохода вокруг нескольких ключевых позиций.
- Широкий ассортимент с разным вкладом: несмотря на то, что это топ-10, выручка последних продуктов («Squalane Cleanser», «Hand & Foot Oil») значительно ниже, чем у лидера (18 тыс. евро и 17 тыс. евро соответственно).

8. КЛИЕНТСКАЯ АНАЛИТИКА

8.1 RFM-анализ

Диаграмма иллюстрирует распределение общего количества покупателей по четырем ключевым сегментам, определенным на основе их покупательского поведения (recency (давность), frequency (частота), monetary (денежной ценности) покупок): Loyal (лояльные) – которые покупают и, вероятно, тратят значительные суммы; Champions (чемпионы) – которые недавно совершали покупки, тратят больше всего, и покупают чаще других; At risk (в зоне риска) – которые тратили большие деньги и покупали часто, но их последняя покупка была давно. Их можно потерять, если не принять меры; Potential (потенциальные) – недавние покупатели, которые уже потратили приличную сумму и совершили более одной покупки, но еще не перешли в категорию лояльных.

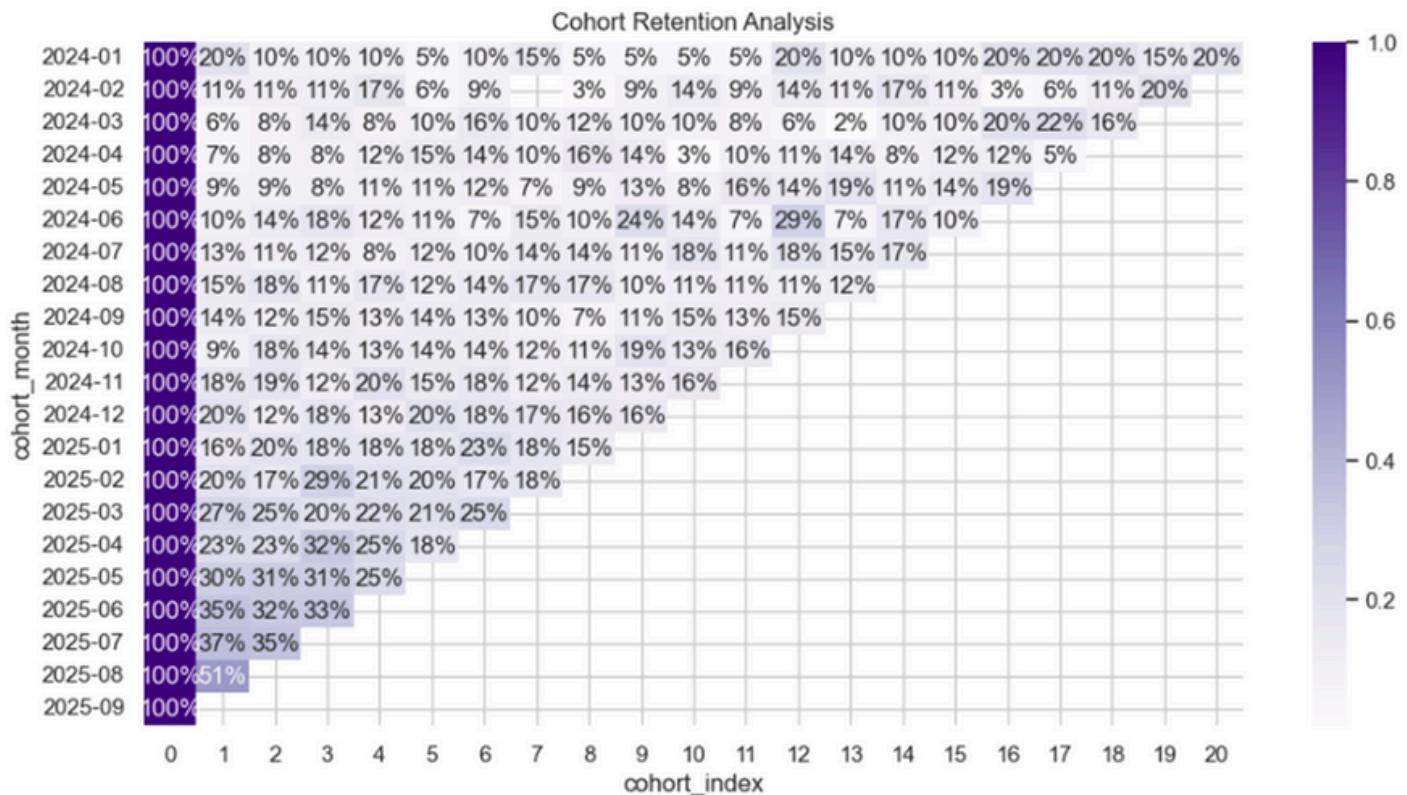


Выводы:

- Лояльные и чемпионы составляют наибольшую долю клиентской базы (более 1200 и около 850 покупателей соответственно), что свидетельствует о сильном ядре постоянных покупателей. Это самые ценные группы для бизнеса.
- Значительное количество покупателей находится в зоне риска (около 520). Этой группе требуется немедленное внимание, чтобы предотвратить их отток.
- Сегмент потенциальных клиентов достаточно велик, что указывает на хорошие возможности для их развития и перевода в более лояльные категории.

8.2 Cohort analysis

Тепловая карта показывает процент пользователей, вернувшихся к покупкам после первого заказа. Каждая строка представляет собой отдельную когорту – группу пользователей, совершивших покупку в определенный месяц, а столбцы – это прошедшие месяцы (индексы когорт).



Выводы:

- Ранний отток покупателей (в большинстве когорт наблюдается резкое снижение процента удержания уже во втором месяце, например, когорта января 2024 года теряет 80% покупателей. Проблема указывает на потенциальные проблемы с онбордингом (первым опытом использования), или первоначальным соответствием продукта ожиданиям покупателей).
- Стабилизация удержания (после первых 2-3 месяцев показатели удержания для большинства когорт стабилизируются и снижаются медленнее, оставаясь в среднем в диапазоне 10-20% для когорт 2024 года)
- Сезонность и эффективность привлечения (когорты 2025 года показывают значительно лучшие результаты удержания по сравнению с когортами 2024 года. Например, когорта мая 2025 года удерживает 31% покупателей даже на 4 месяц, это говорит о том, что изменения в продуктах, маркетинговых кампаниях или процессах привлечения в 2025 году оказались более эффективными).

8.2 CLV

Выводы

- Методология: Код группирует данные по клиентам для расчета общего дохода и количества заказов, затем вычисляет средний чек (avg_order_value) и общую CLV для каждого клиента. Далее данные объединяются с информацией о городе, и рассчитывается средняя CLV для каждого города.
- Лидеры по CLV: Города Будапешт (Budapest), Осло (Oslo) и Хельсинки (Helsinki) демонстрируют самую высокую среднюю жизненную ценность клиента.
- Приоритеты: Клиенты из этих городов приносят компании наибольший доход в долгосрочной перспективе, что делает их наиболее ценным сегментом.

8.2 Churn = 0.51

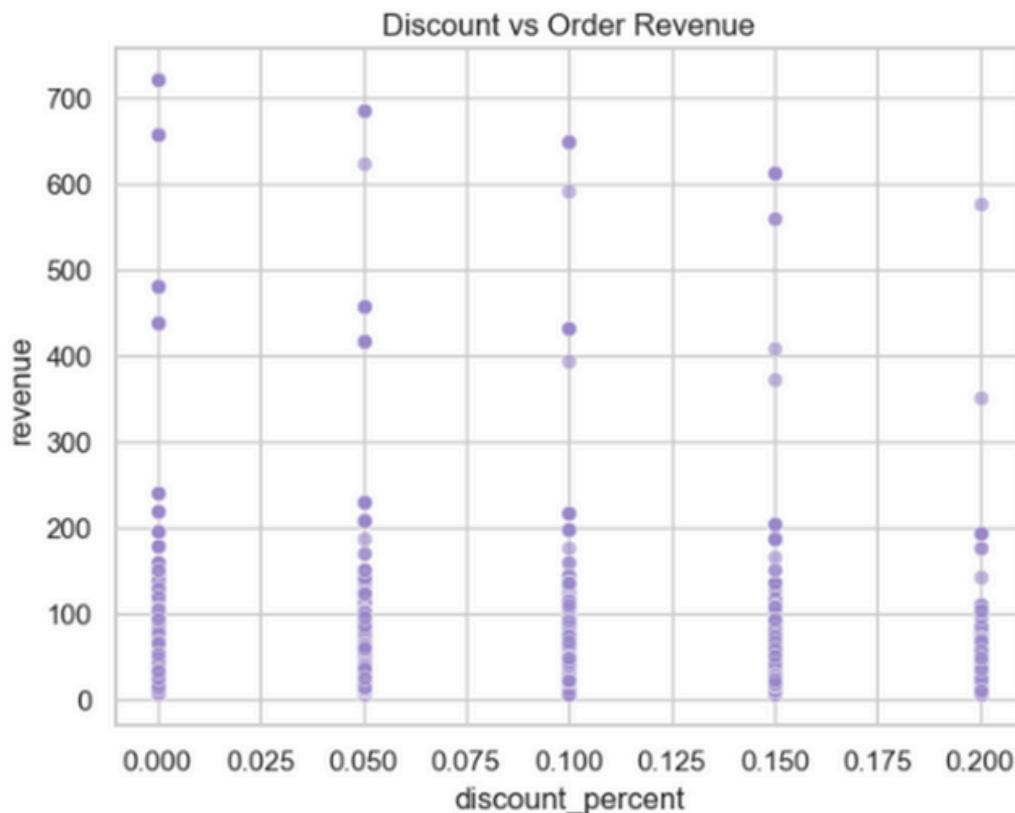
Выводы

- Методология: Код вычисляет дату отсечения (churn_cutoff) как 90 дней до даты снимка. Клиенты, чей последний заказ был до этой даты, помечаются как "оттекшие" (churned).
- Показатель оттока: Рассчитанный уровень оттока составляет 0.5105 или приблизительно 51%.
- Интерпретация: Такой высокий показатель оттока означает, что более половины клиентской базы, существовавшей на начало анализируемого периода, прекратили взаимодействие с компанией (не делали заказы более 90 дней). Это очень высокий показатель, который обычно указывает на серьезные проблемы в бизнесе, требующие немедленного внимания.

8. ЦЕНООБРАЗОВАНИЕ И СКИДКИ

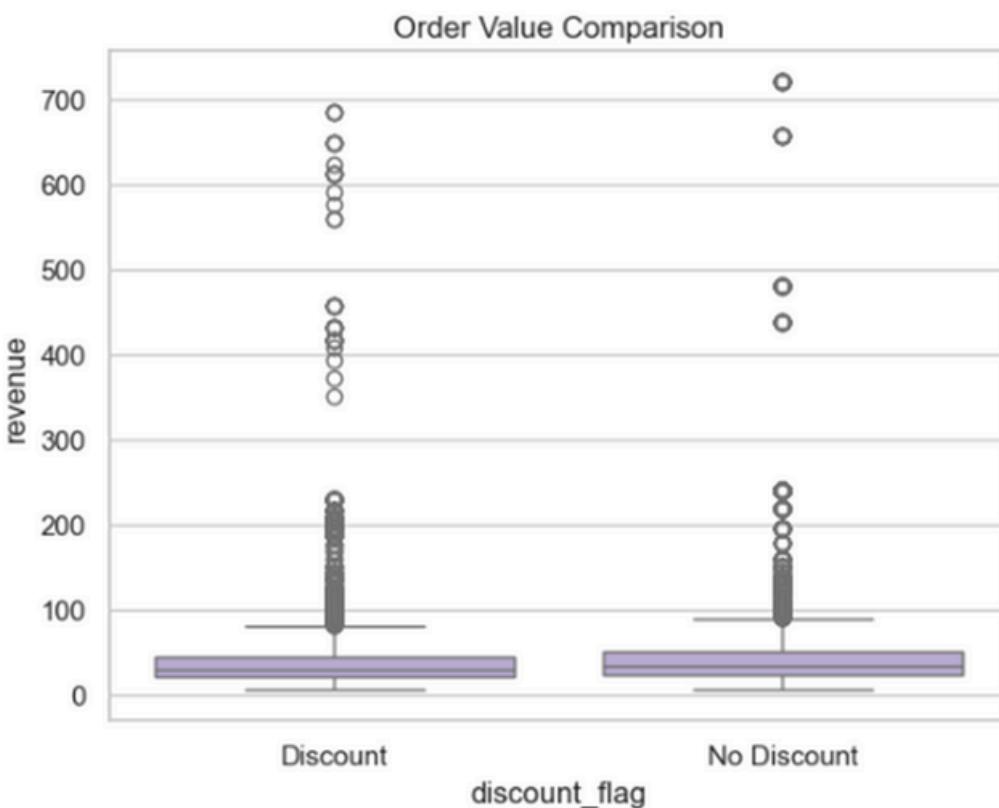
8.1 Скидки и выручка

График (диаграмма рассеяния) показывает взаимосвязь между точным процентом скидки и выручкой каждого отдельного заказа.



8.2 Распределение выручки

График (ящик с усами) сравнивает распределение выручки для двух дискретных групп: с примененной скидкой и без нее.



Выводы:

- Низкая медианная выручка: подавляющее большинство заказов независимо от наличия скидки, имеют очень низкую выручку, что подтверждается сжатыми «ящиками» на втором графике и плотным скоплением точек внизу на первом графике.
- Отсутствие сильной корреляции: ни один из графиков не демонстрирует четкой положительной связи между представлением скидки и ростом выручки отдельного заказа. Скидки не гарантируют более высокий средний чек.
- Крупные заказы (выбросы): редкие заказы с очень высокой выручкой встречаются в обеих группах. При этом заказы без скидки иногда достигают даже немного более высоких максимальных значений.

10. MARKET BASKET ANALYSIS

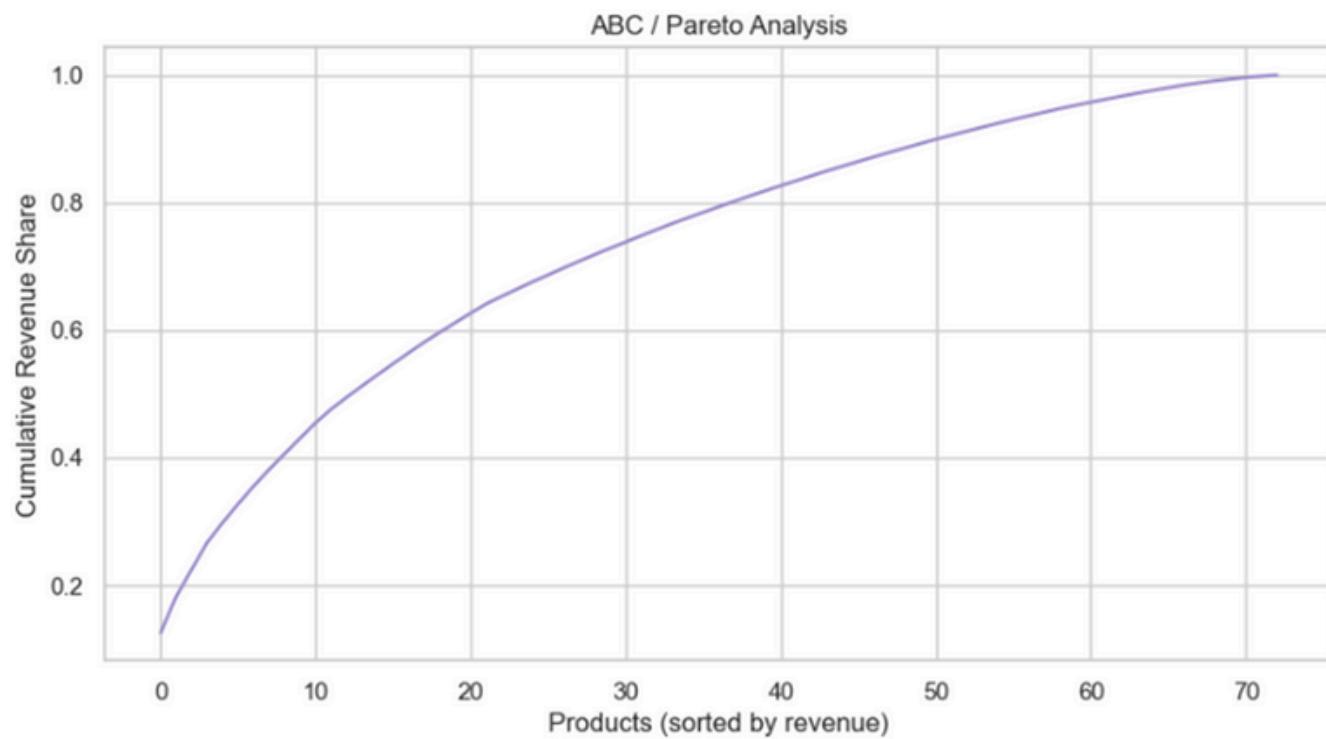
		products count
8	(Daily Moisturising Cream, Ground Rice & Honey...)	30
161	(Rosemary Mint Scalp & Hair Strengthening Oil,...)	25
547	(Brightening AHA Peel Mask, Ground Rice & Hone...	25
125	(Ground Rice & Honey Glow Mask, Hydrating Toni...	24
131	(Ground Rice & Honey Glow Mask, Turmeric Bright...	23
415	(Ground Rice & Honey Glow Mask, Squalane Clean...	23
279	(Flowers Facial Oil Deep Rich, Ground Rice & H...	22
36	(Brightening AHA Peel Mask, Squalane Cleanser)	22
321	(Beautifying Toner, Ground Rice & Honey Glow M...	22
316	(Organic Flowers Water Cream, Squalane Cleanser)	20

Выводы

- Наиболее популярная пара: Крем Daily Moisturising Cream и Ground Rice & Honey Glow Mask покупают вместе чаще всего - 30 раз.
- Частые комбинации: Продукты Ground Rice & Honey Glow Mask и Squalane Cleanser; Organic Flowers Water Cream и Squalane Cleanser; Brightening AHA Peel Mask и Squalane Cleanser также часто встречаются в одних и тех же заказах (по 20-23 раза).
- Популярность отдельных продуктов: Судя по частым комбинациям, продукты Ground Rice & Honey Glow Mask и Squalane Cleanser являются очень популярными и входят в большинство лучших пар.

11. ABC / PARETO-АНАЛИЗ

Был проведен ABC -анализ ассортимента магазина (73 SKU) для определения ключевых продуктов, приносящих основную выручку, и оптимизации управления запасами на основе правила Парето.



Продукты были отсортированы по убыванию вклада в общую выручку. Границы категорий определены следующим образом, на основе кумулятивной (накопленной) доли выручки (на графике):

- A: 0% - 79% накопленной выручки (доля в выручке 80%)
- B: 80% - 95% накопленной выручки (доля в выручке 15%)
- C: 96% - 100% накопленной выручки (доля в выручке 5%)

Выводы:

51% (37 наименований) продуктов дает 80% дохода. Это говорит о том, что бизнес имеет сильное ядро локомотивных продуктов, но также обладает высокой устойчивостью т.к. выручка не концентрируется на 2-3 позициях.

12. КЛЮЧЕВЫЕ ИНСАЙТЫ

- Выручка магазина формируется половиной продуктов. Продукты категории А (по ABC-анализу) обеспечивают основную долю выручки, что делает их критически важными для финансовой устойчивости бизнеса.
- Существует высокая зависимость от одного продуктового лидера. Active Botanical Serum является ключевым драйвером выручки, что увеличивает риски при перебоях в поставках или изменении спроса.

- Распределение клиентов по городам не коррелирует напрямую с их ценностью. Города с большим количеством клиентов не всегда демонстрируют высокий ARPU, что указывает на необходимость переоценки географических приоритетов.
- Не наблюдается выраженная сезонность продаж и ускорение роста в конце 2024 - начале 2025 года.
- Текущая стратегия скидок не демонстрирует устойчивого положительного влияния на выручку заказа. Заказы со скидками, включая максимальные скидки, не показывают значимого роста среднего чека по сравнению с заказами без скидок.
- Основной отток клиентов происходит в первые месяцы после первой покупки.
- Когортный анализ показал резкое снижение удержания на ранних этапах жизненного цикла клиента.
- Уровень оттока клиентов остаётся высоким (~51%), несмотря на рост выручки.
- Рост бизнеса в большей степени обеспечивается привлечением новых клиентов, а не удержанием существующих.
- Небольшая доля клиентов генерирует основную часть выручки.
- RFM- и Pareto-анализы подтверждают наличие высокоценных сегментов клиентов с существенно большей жизненной ценностью.
- Существует потенциал роста за счёт увеличения повторных покупок.
- Значительная часть клиентов совершает ограниченное число заказов, что указывает на неиспользованный потенциал retention-стратегий.
- Некоторые категории ассортимента демонстрируют системно низкий вклад в выручку.
- Это может свидетельствовать как о проблемах позиционирования, так и о неэффективном управлении ассортиментом в данных категориях.

13. РЕКОМЕНДАЦИИ

1. Управление ассортиментом и запасами

- Сфокусировать управление запасами на продуктах категории А (ABC-анализ), внедрив строгий контроль неснижаемого остатка для предотвращения out-of-stock, так как их отсутствие оказывает критическое влияние на выручку.
- Провести дополнительный анализ оборачиваемости и рентабельности продуктов категории С. Рассмотреть вывод неликвидных позиций из ассортимента или проведение распродаж для высвобождения оборотного капитала.
- Обеспечить постоянное наличие ключевых продуктов-лидеров (в частности, Active Botanical Serum), одновременно снижая зависимость бизнеса от одного драйвера за счёт развития продуктов среднего сегмента.

2. Оптимизация продуктовой структуры и категорий

- Усилить маркетинговую поддержку и управление запасами для категорий-лидеров по выручке с целью максимизации их потенциала. Провести детальный анализ причин низкой выручки в отстающих категориях. Развитие стратегии кросс-продаж.

3. Клиентская стратегия и удержание

- Использовать результаты RFM-анализа для дифференциированного подхода к клиентам:
 - поддерживать высокий уровень вовлечённости для «чемпионов» и лояльных клиентов через эксклюзивные предложения и персонализацию;
 - запустить реактивационные кампании для клиентов в зоне риска;
 - стимулировать повторные покупки у потенциально лояльных клиентов.
- Сосредоточить усилия на улучшении онбординга новых клиентов, так как наибольший отток наблюдается в первые месяцы после первой покупки.
- Для стабилизации долгосрочного удержания (4–6 месяцев) разработать и внедрить программы лояльности, стимулирующие повторные покупки.

4. Географическая оптимизация и рост ARPU

- Перейти от оценки эффективности регионов по количеству клиентов или общей выручке к приоритезации по ARPU.
- Усилить маркетинговые активности в городах с высоким средним чеком для масштабирования наиболее прибыльных рынков.
- Для регионов с низким ARPU:
 - провести дополнительный анализ причин (ценовая чувствительность, конкурентная среда, структура покупок);
 - внедрить локальные программы лояльности и стимулирования среднего чека;
 - адаптировать ассортимент и ценовую политику с учётом региональной специфики.

5. Пересмотр скидочной политики

- Пересмотреть текущую стратегию скидок, так как предоставление скидок не демонстрирует однозначного положительного эффекта на рост выручки заказа.
- Сместить фокус анализа со скидок на прибыльность, включая оценку маржинальности заказов со скидкой, особенно в сегменте скидок 20%.
- Провести сегментацию покупателей и более глубокий анализ факторов, приводящих к редким крупным заказам, с целью масштабирования успешных сценариев без избыточного снижения цен.

6. Масштабирование и стратегическое планирование

- Поддерживать текущий темп роста бизнеса, параллельно масштабируя операционные процессы для предотвращения ухудшения качества обслуживания.
- Проанализировать изменения в конце 2024 – начале 2025 года, приведшие к ускорению роста, и использовать полученные инсайты для дальнейшего масштабирования.
- В условиях роста выручки усилить контроль операционных расходов и себестоимости для сохранения устойчивой прибыльности.
- Использовать выявленные тренды для прогнозирования продаж и стратегического планирования запасов, персонала и инвестиций.

14. ОГРАНИЧЕНИЯ АНАЛИЗА

- Синтетические данные. Все данные сгенерированы с помощью Python для демонстрации аналитических методик. Они не отражают реального поведения клиентов или точных финансовых показателей компании.
- Отсутствие дополнительных бизнес-данных. В анализе не учитывались каналы привлечения, маркетинговые кампании, сезонные промо-акции и фактическая себестоимость товаров, что может влиять на интерпретацию выручки и прибыльности.
- Ограниченные временные рамки. Данные охватывают период до октября 2025 года. Долгосрочные тенденции и циклы поведения клиентов не выявлены.
- Упрощённые расчёты CLV и RFM. Метрики пожизненной ценности и сегментации клиентов рассчитаны упрощёнными методами, без прогнозирования будущих покупок или учёта влияния маркетинговых активностей.
- Не учтены внешние факторы. Конкуренция, изменения цен на рынке и макроэкономические условия не были включены в анализ, что ограничивает точность стратегических выводов.
- Ограничения визуализации и представления данных. Некоторые графики, например по ежедневным продажам, не были включены в основной отчёт из-за ограничений читаемости и качества визуализации.

15. ЗАКЛЮЧЕНИЕ

В ходе анализа продаж и клиентского поведения интернет-магазина натуральной косметики выявлены ключевые закономерности и возможности для роста.

- Основная часть выручки обеспечивается половиной и небольшой долей клиентов, что подчёркивает важность фокусировки на высокодоходных категориях и сегментах клиентов.
- Рост выручки наблюдается с конца 2024 года, что подтверждает положительную динамику бизнеса и потенциал для масштабирования операций.
- Уровень оттока клиентов остаётся значительным (~51%), при этом наибольший риск связан с новыми покупателями в первые месяцы после первой покупки.
- Текущая политика скидок не даёт очевидного прироста среднего чека, а некоторые продукты и категории требуют дополнительного внимания с точки зрения ассортимента и прибыльности.