

ОТЧЕТ О ПРОЕКТЕ: Обработка данных покупателей онлайн платформы

Выполнял студент: Малыга Дмитрий Алексеевич

Группа: Е4160

Цель проекта: Создание автоматизированной платформы для обработки транзакционных данных и сегментации клиентской базы с целью оптимизации маркетинговых затрат и удержания покупателей.

Бизнес-проблема: Рост клиентской базы без четкого понимания их лояльности приводит к неэффективным рассылкам и потере клиентов в сегменте «под угрозой». Необходим инструмент для быстрого выделения ключевых групп покупателей.

Ключевые задачи:

1. Разработка ETL-пайплайна для очистки данных и формирования БД.
2. Проведение RFM-сегментации (Recency, Frequency, Monetary).
3. Анализ удержания клиентов через когортный анализ.
4. Выявление географических паттернов продаж (ABC-анализ).

1. Методология и обработка данных (ETL)

Процесс реализован на языке Python с использованием библиотеки **Prefect** для оркестрации задач. Данные организованы по «медальонной» архитектуре:

- **Слой Silver (Очистка):** Обработка 500k+ строк. Удаление транзакций без ID клиента, фильтрация возвратов, коррекция типов данных. Данные сохранены в таблицу `cleaned_transactions`.
- **Слой Gold (Витрины):** Формирование аналитических таблиц для быстрого доступа из BI-систем или Jupyter Notebook (`mart_customer_segments`, `mart_countries_abc`).

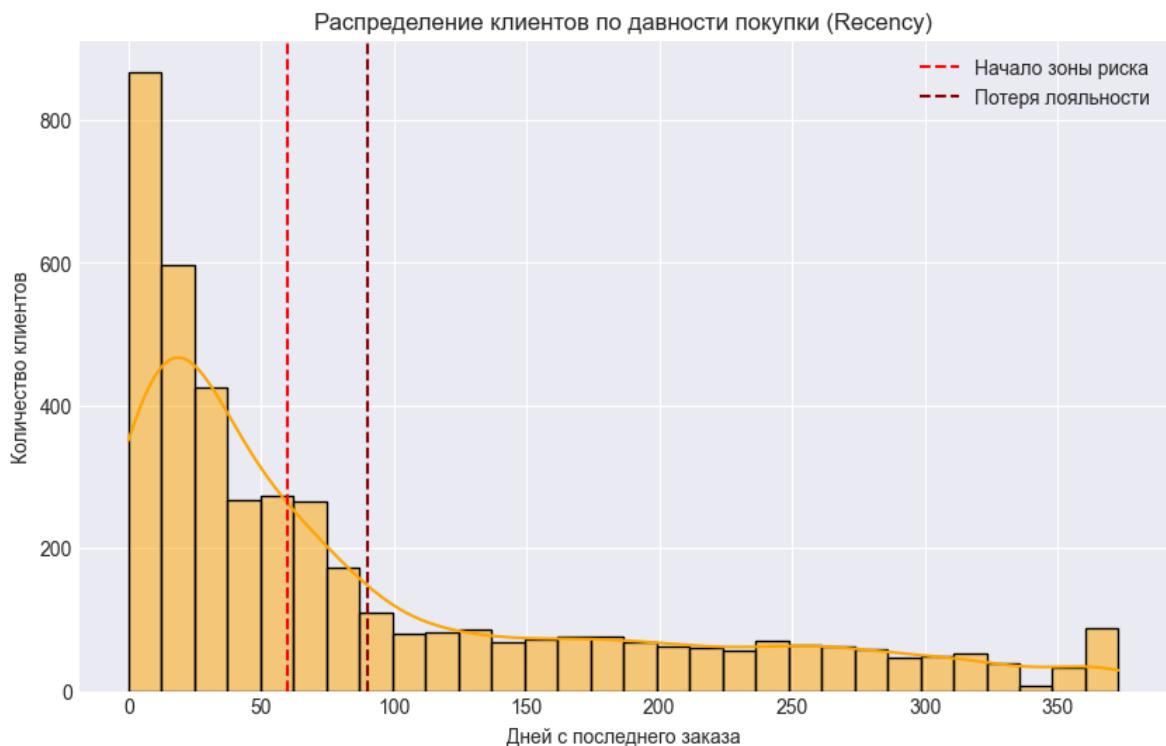
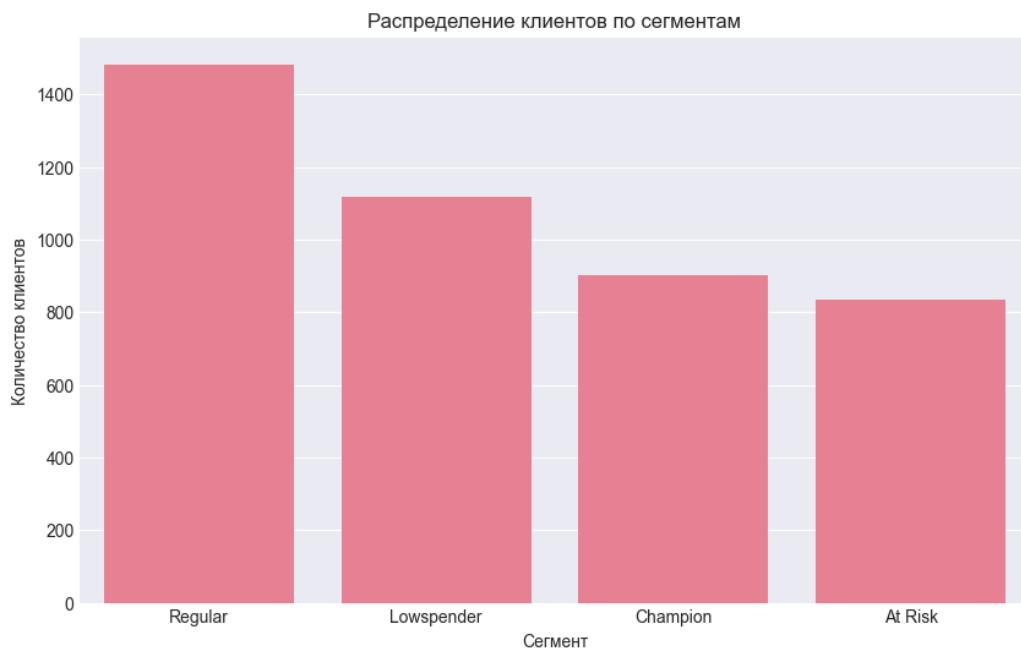
Tables	InvoiceNo	Total_Price
<code>audit_order_anomalies</code>	Filter	Filter...
<code>cleaned_transactions</code>	1 537657	9639.12
<code>mart_countries_abc</code>	2 537659	15885.49
<code>mart_customer_segments</code>	3 539101	6930
	4 539731	8520.92
	5 540267	7011.38
	6 540689	12797.52

2. Анализ и результаты

2.1. RFM-сегментация

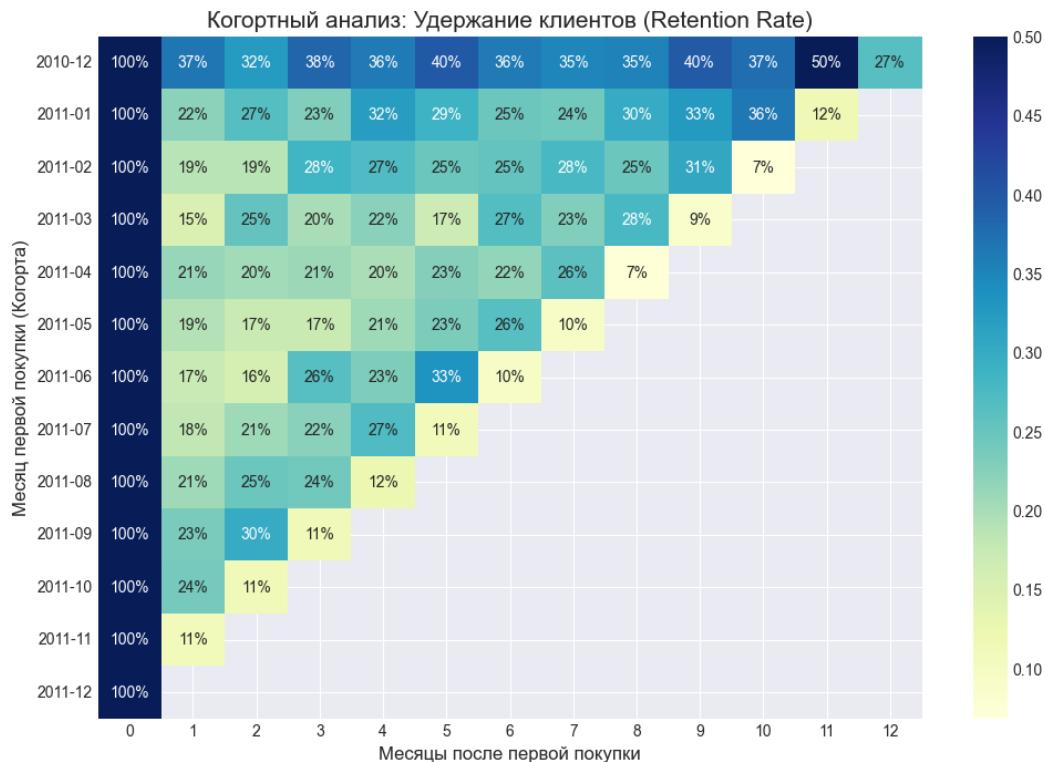
На основе давности, частоты и суммы покупок клиенты разделены на группы:

- **Champions:** Лояльные клиенты с высоким чеком.
- **At Risk:** Клиенты, не совершившие покупок более 60-90 дней.
- **Lowspender**
- **Regular**



2.2. Когортный анализ (Retention)

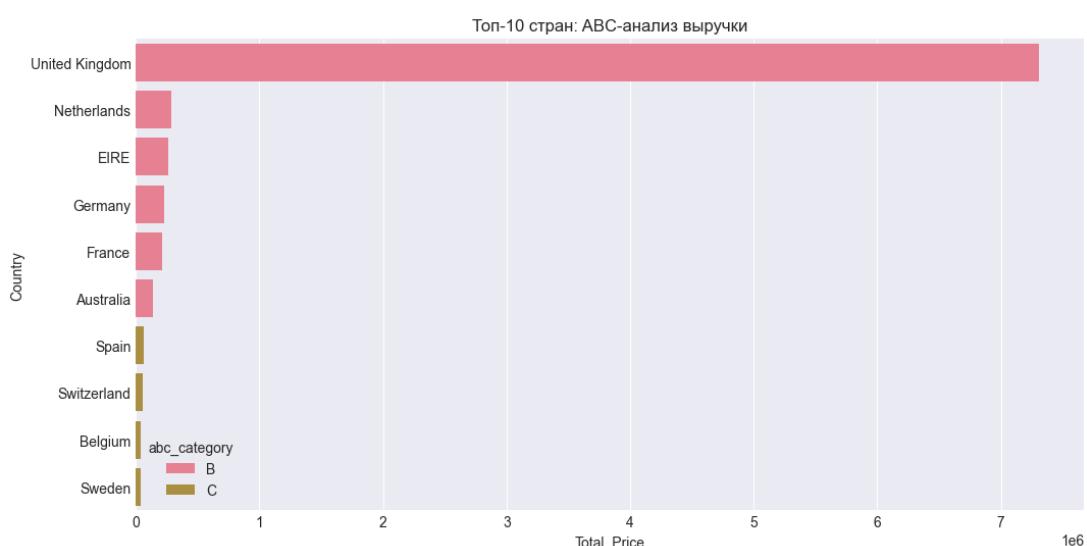
Построена матрица удержания, которая показала, что средний коэффициент удержания (Retention Rate) на второй месяц составляет ~20%. Выявлено, что январская когорта является самой стабильной.



2.3. Географический ABC-анализ

Выручка была проанализирована в разрезе стран:

- Категория А (80% выручки):** Великобритания.
- Категория В (15% выручки):** Германия, Франция, Ирландия.
- Категория С (5% выручки):** Остальные страны (30+ направлений).



3. Итоги и рекомендации

- Автоматизация:** Время подготовки отчетности сокращено с часов до минут благодаря переносу расчетов в ETL-пайплайн.
- Маркетинг:** Рекомендовано внедрить автоматическую рассылку для сегмента **At Risk** на 45-й день после последней покупки.
- Масштабирование:** Рынки категории **B** (Германия, Франция) имеют потенциал для роста и требуют локализованных рекламных кампаний.
- Безопасность:** Выявленные аномалии указывают на сегмент оптовых закупщиков, требующий отдельного бизнес-процесса обработки.