

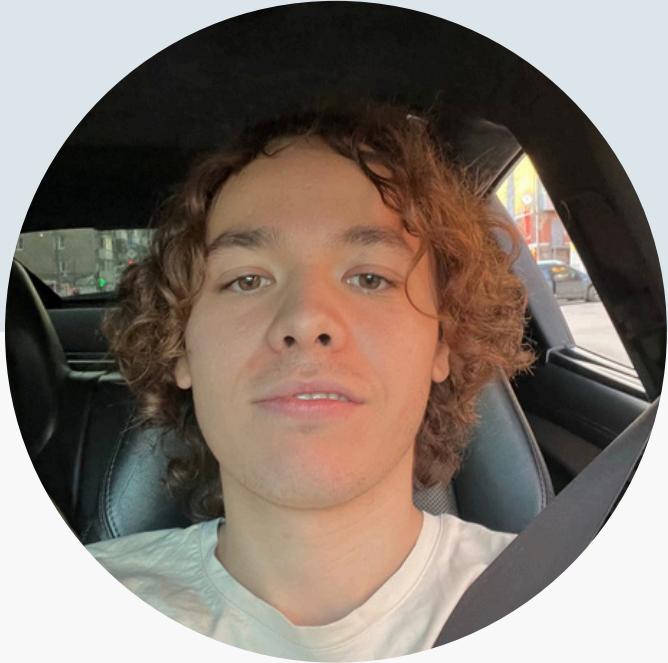
Прикладная наука о данных

Командный проект

Снижение стоимости возвратов в e-commerce через аналитику и
ML



Участники команды



Камбетов Давид

Product Manager



Динь Тхи Хонг Ань

Project Manager



Панасенкова Полина

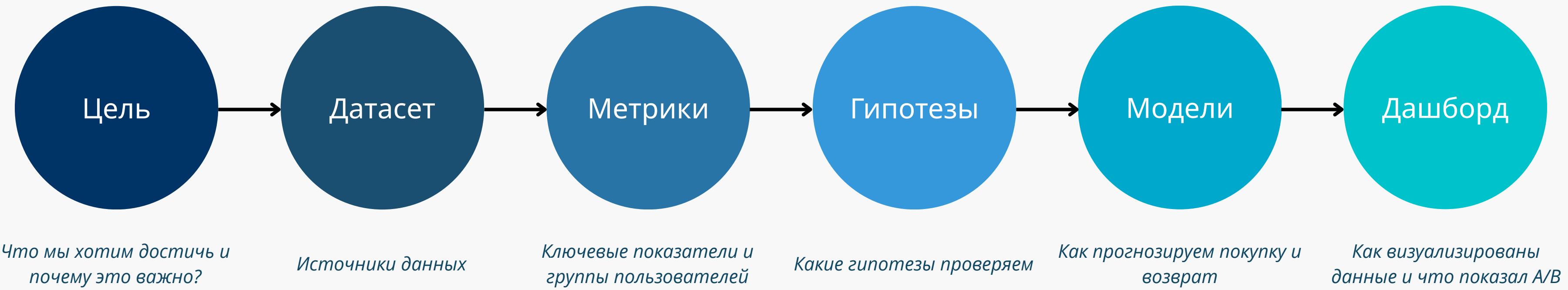
Data-Analyst



Мануйлов Павел

Data-scientist

Структура презентации



План реализации проекта

Этап 1 → Этап 2 → Этап 3 → Этап 4

Неделя 1 Подготовка

Созвоны команды,
распределение ролей,
аудит и очистка
данных

Неделя 2 Аналитика

Базовые метрики,
сегментация,
проверка гипотез Н1-
Н3

Неделя 3 Модели

Подготовка
признаков, обучение
Р2В/R2R, дашборд

Неделя 4 Итоги

Финальные расчёты,
визуализация,
презентация и отчёт

Каждая неделя - шаг от сырых данных к продуктовым инсайтам

Цель и идея проекта

Основная цель:

Снизить долю и стоимость возвратов, сохранив рост продаж и среднего чека с помощью аналитики, скринговых моделей и визуальных дашбордов

Задачи:

- Построить метрики продаж и возвратов
- Сегментировать пользователей и проверить гипотезы
- Визуализировать результаты в интерактивный дашборд



Факты из датасета

Основные цифры, на которых строится аналитика проекта

Сессии

172 838

Общее кол-во
визитов на сайт

Заказы

131 500

Сформированные
покупки

Выручка

≈ 71,1 млн ₽

Сумма продаж по
всем заказам

Возвраты

≈ 12,79 млн ₽

Почти 18% от общей
выручки

Чистая выручка

≈ 58,3 млн ₽

После вычета
возвратов

Средний чек

≈ 540,7 ₽

Средний доход с
заказа

Гипотеза H_1 - Возвраты и дорогие покупки

Проверяем, как стоимость заказа влияет на вероятность возврата



Вывод:

Гипотеза подтверждена
частично: возвраты чаще
среди дорогих покупок

- Возвраты чаще встречаются в корзинах 250-500₽ и 1000-1500₽
- Дорогие заказы возвращают немного чаще, но незначительно
- Средняя сумма возврата = 566,9₽, что выше средней покупки = 540,7₽

Гипотеза H2. Высокий объём продаж ≠ высокий средний чек

Анализируем различия между странами по количеству заказов и среднему чеку



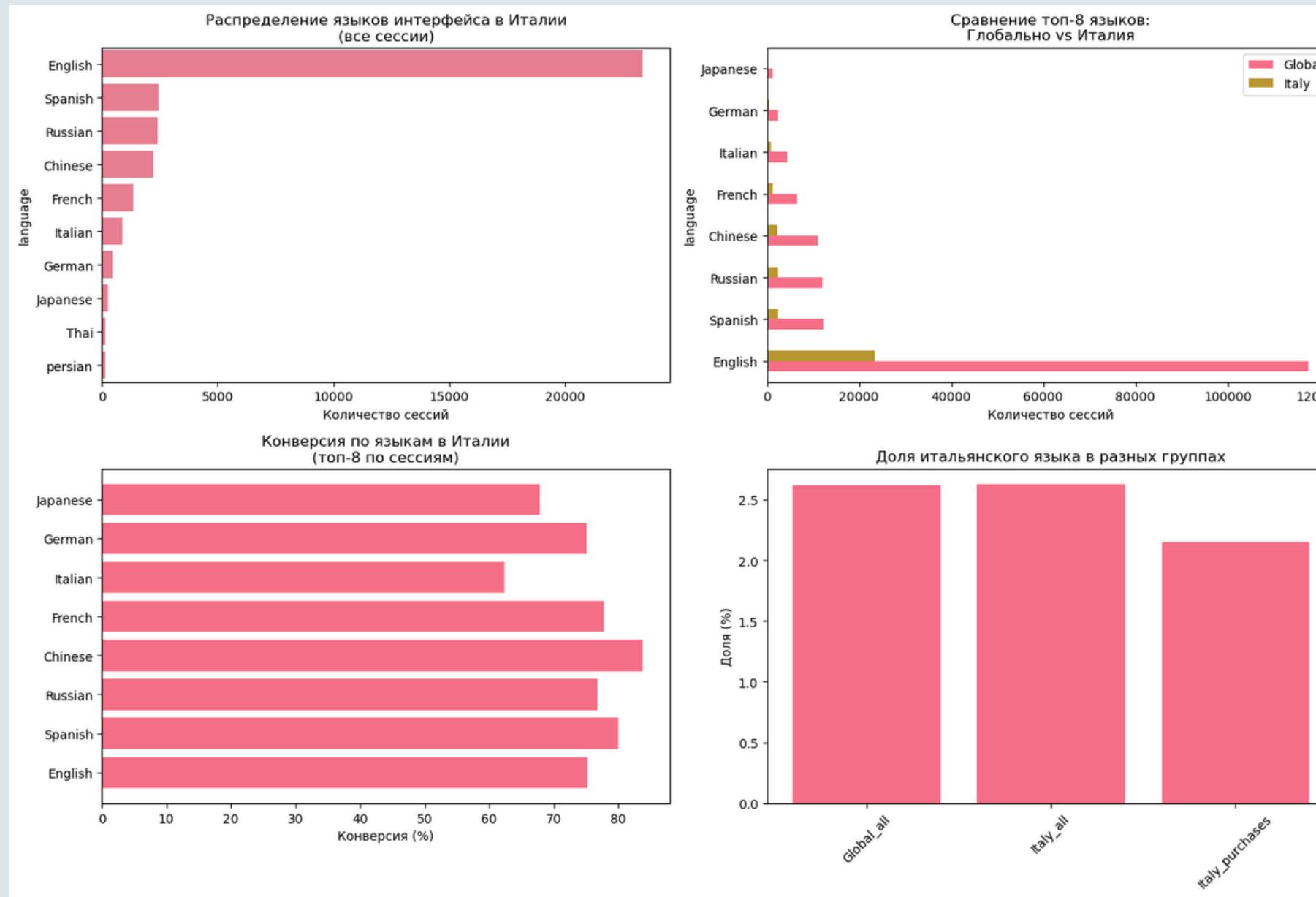
Вывод:

Стратегии роста должны быть разными: в одних странах повышаем чек, в других - конверсию

- Италия - лидирует по продажам, но средний чек = 539 ₽
- Польша и германия - средний чек высокий, но заказов мало
- Видна сильная разница между странами по модели "объём-чек"

Гипотеза Н3. Почему в Италии доминирует английский язык интерфейса

Проверяем, как языковые настройки влияют на поведение пользователей



- В Италии английский язык используют чаще, чем итальянский
- Итальянский не входит в топ-5 по количеству сессий
- Возможные причины: стандартные лендинги на английском, экспаты и туристы

Вывод:

Отсутствие строгой локализации снижает доверие и повышает долю возвратов - интерфейс нужно адаптировать

Применение скоринговых моделей: P2B и R2R

Модель	Задача	Метрика качества	Вывод
P2B	Прогноз вероятности покупки	ROC AUC = 0,8	Хорошо отделяет покупателей от неактивных пользователей
R2R	Прогноз вероятности возврата	ROC AUC = 0,8	Уверенно определяет риск возврата

**На основе очищенных данных
обучили две скоринговые модели:**

- P2B - вероятность покупки
- R2R - риск возврата товара

Это позволило сегментировать пользователей и предложить персональные UX-интервенции

Ключевые инсайты:

- Модели помогают определить группы с высоким потенциалом покупки и с высоким риском возврата
- Эти данные легли в основу дашборда и сегментации

Сегмент → Модель → Интервенция → Выручка / Возвраты ↓

Применение скоринга позволило перейти от общей аналитики к персонализированным бизнес-решениям

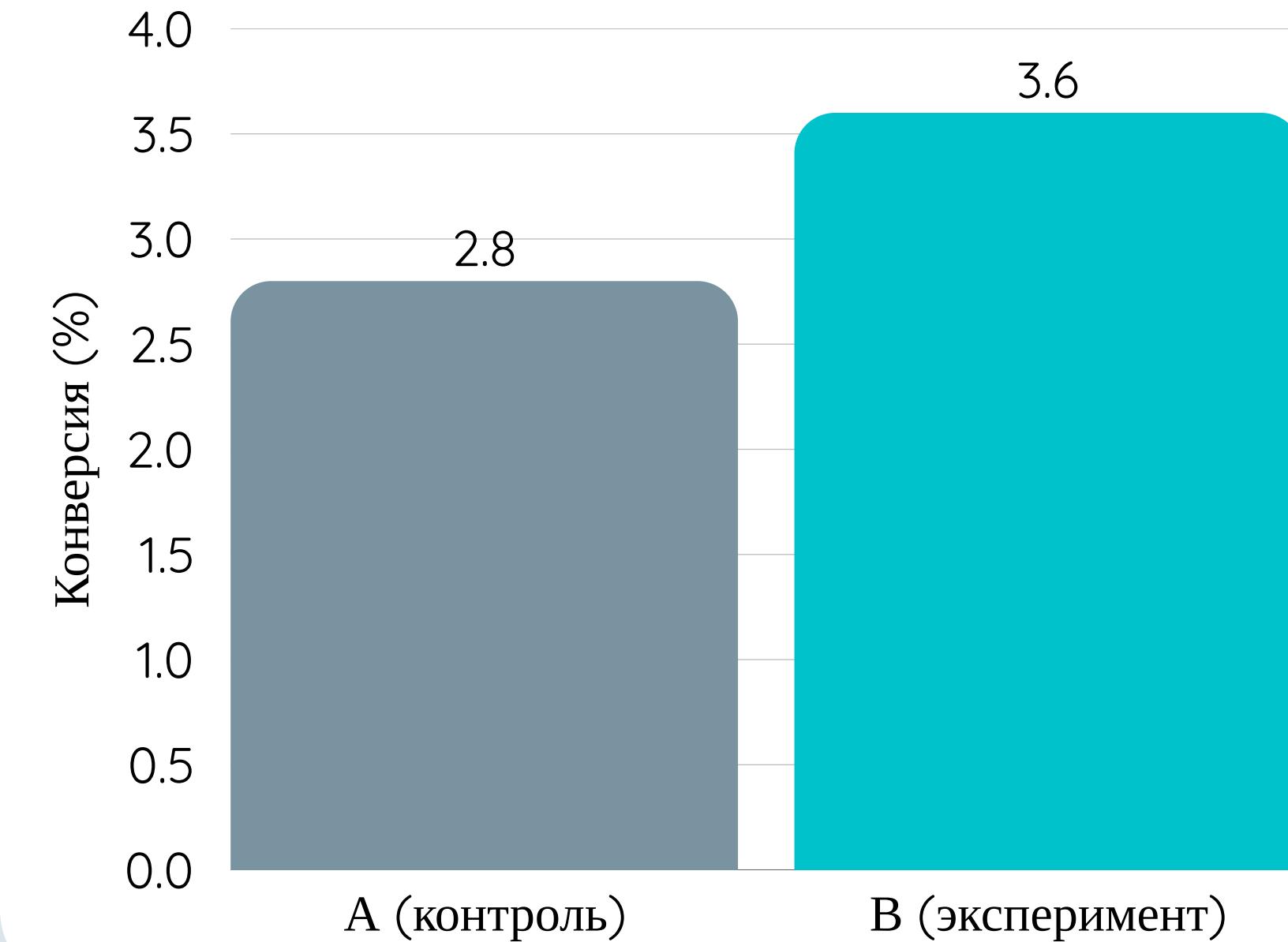
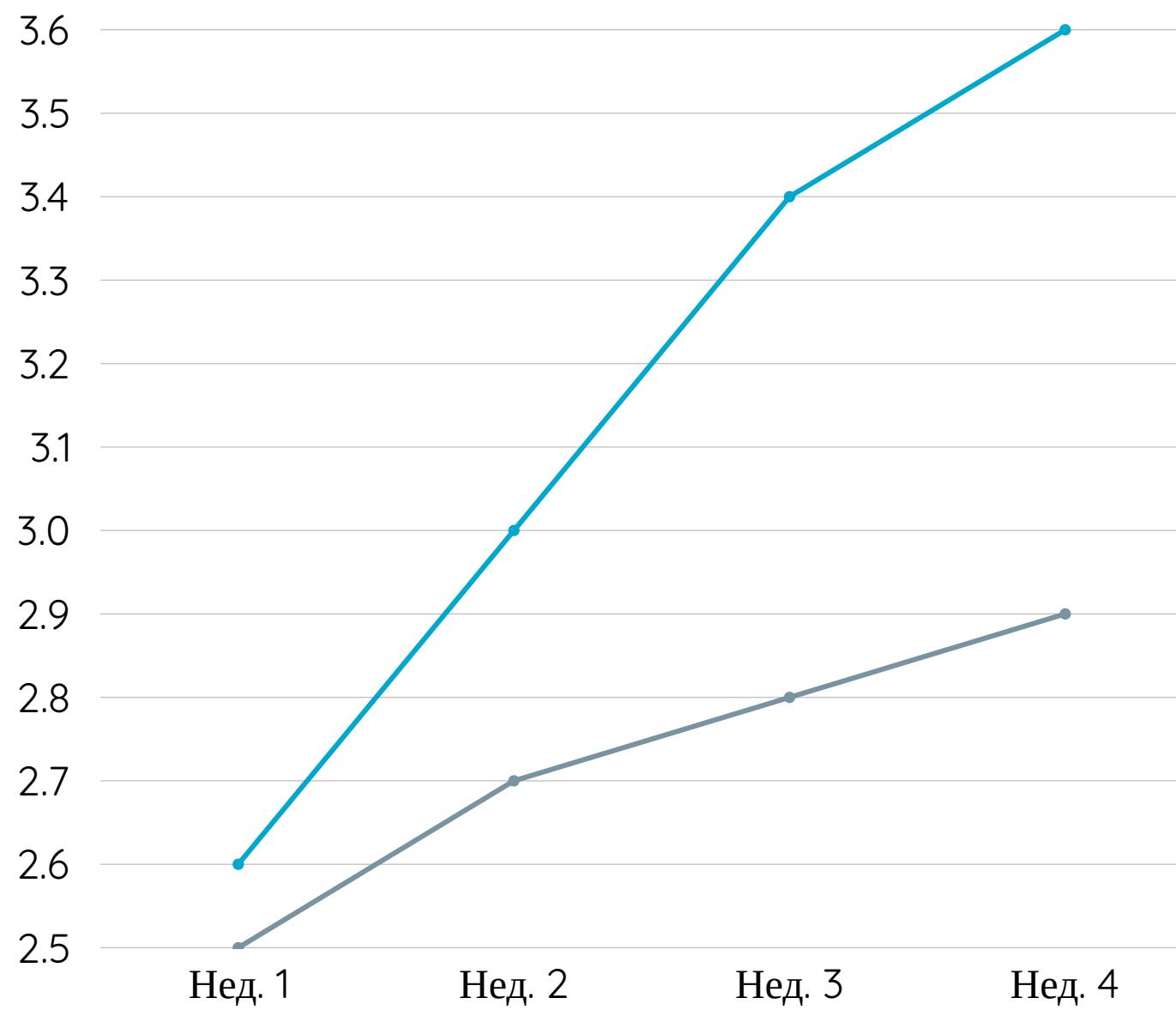
A/B-тест: влияние персонализированных интераций

После обучения моделей пользователи были случайно разделены на две группы:

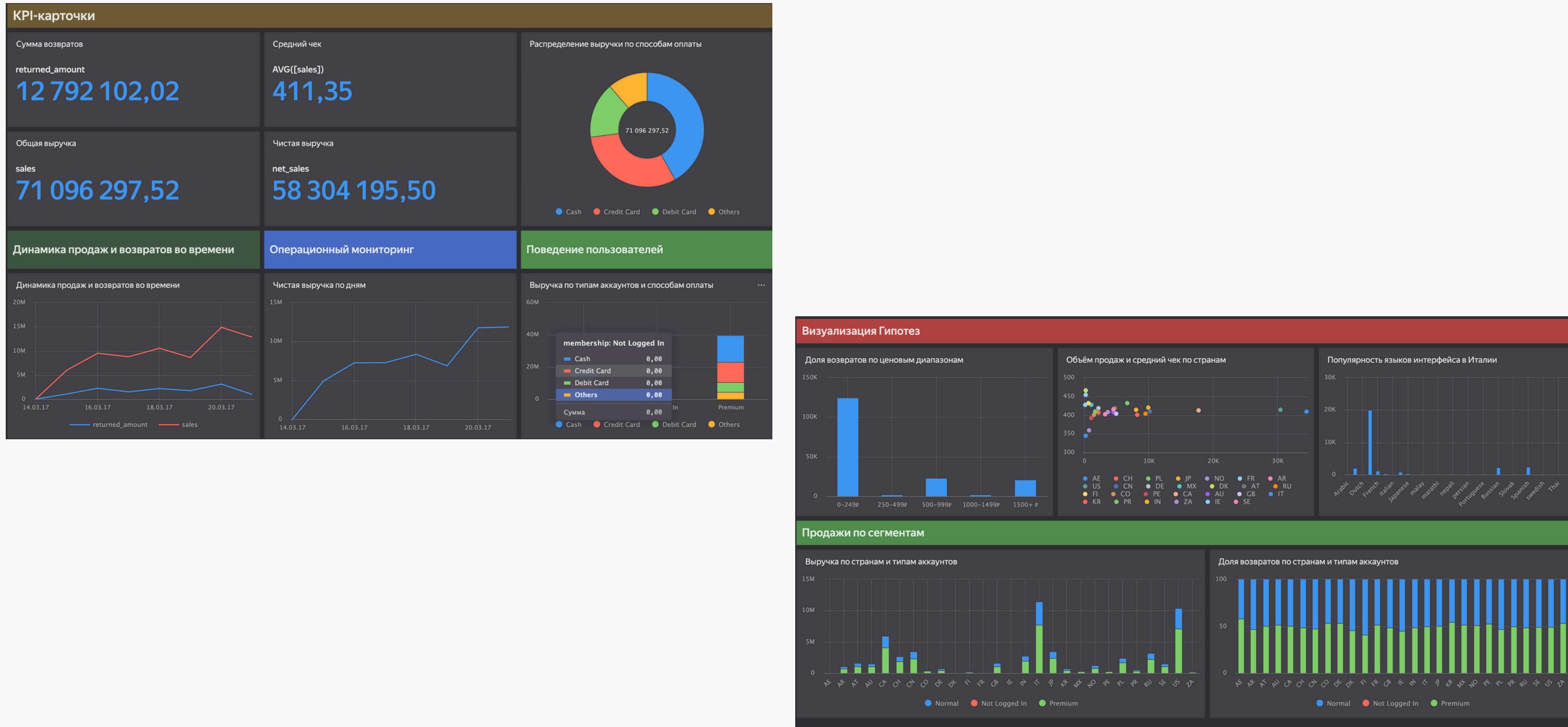
А - контроль: стандартный пользовательский опыт (без изменений)

В - эксперимент: логика взаимодействия строилась по результатам скоринга (Р2В и R2R)

Динамика конверсии за 4 недели



Финальный дашборд проекта



Финальный дашборд проекта

