



Анализ поведения пользователей Netflix

Понимание вовлечённости, географии и предпочтений аудитории

Выполнил: Сингаевский Андрей Николаевич

Дата: Сентябрь 2025

Источник данных: Kaggle – симулированный датасет пользовательской активности

Цель исследования и описание данных

Основная цель

Выявить паттерны поведения подписчиков Netflix, определить ключевые регионы роста и понять, какие категории контента удерживают пользователей дольше всего. Исследование направлено на оптимизацию стратегии удержания клиентов и развития платформы.

Структура датасета

- `userId`, `subscriptionType` (Basic, Standard, Premium)
- `country`, `state`, `city` – географические данные
- `joinDate`, `lastActiveDate` – временные метки
- `contentCategory` (Movies, TV Shows, Documentaries)
- `viewingHours`, `deviceType` (Mobile, TV, Tablet)

Объём: ~50 000 записей, 10+ признаков



Данные синтезированы на основе публичных отчётов Netflix и поведенческих паттернов стриминговых сервисов



Этапы предобработки данных

01

Стандартизация структуры

Приведение названий столбцов к стилю `camelCase` для единообразия кода и улучшения читаемости

02

Преобразование дат

Конвертация `joinDate` и `lastActiveDate` в формат `datetime` для корректного расчёта активности и срока подписки

03

Очистка дубликатов

Удаление повторяющихся записей, поскольку `userId` должен быть уникальным идентификатором

04

Заполнение пропусков

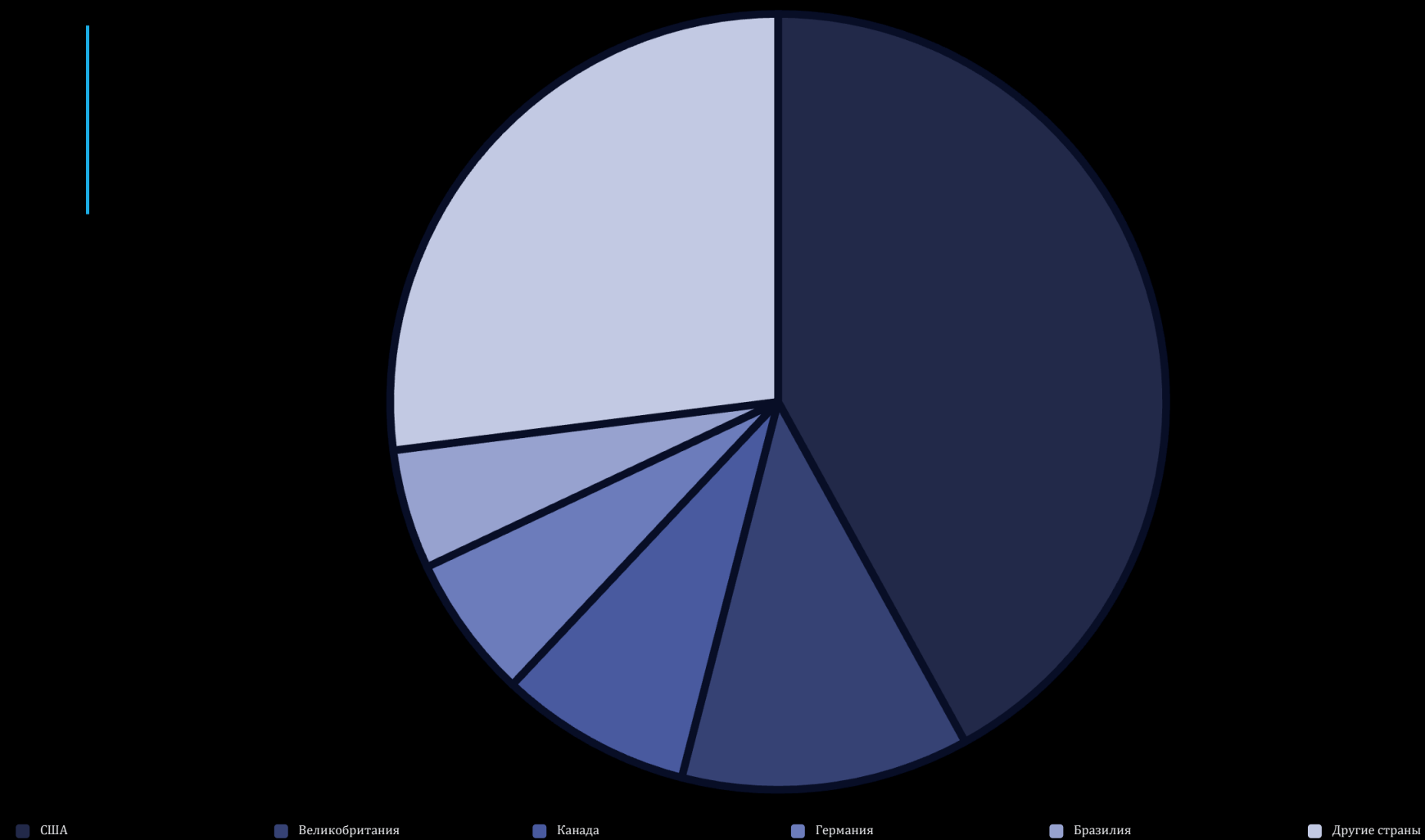
Восстановление отсутствующих значений в полях `state` и `city` на основе данных `country`

05

Кодирование переменных

Преобразование категориальных переменных для последующего статистического анализа

Географическое распределение пользователей

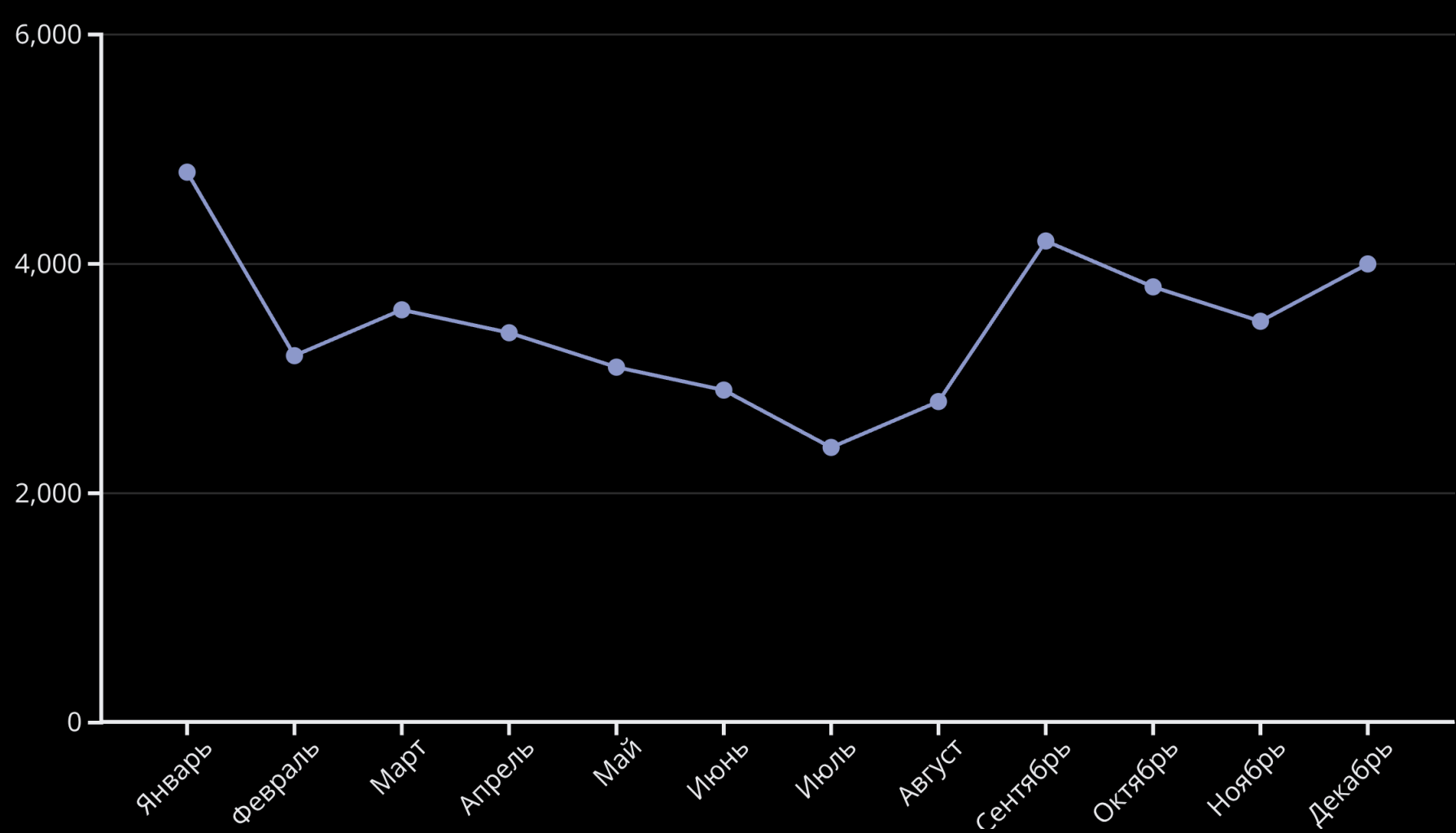


Ключевые наблюдения

- Северная Америка доминирует с долей 50%+ пользователей
- Латинская Америка и Азия показывают низкое проникновение – потенциал роста
- Города-миллионники демонстрируют самую высокую плотность пользователей



Динамика подписок и активности пользователей



14

Месяцев

Средняя продолжительность подписки

28%

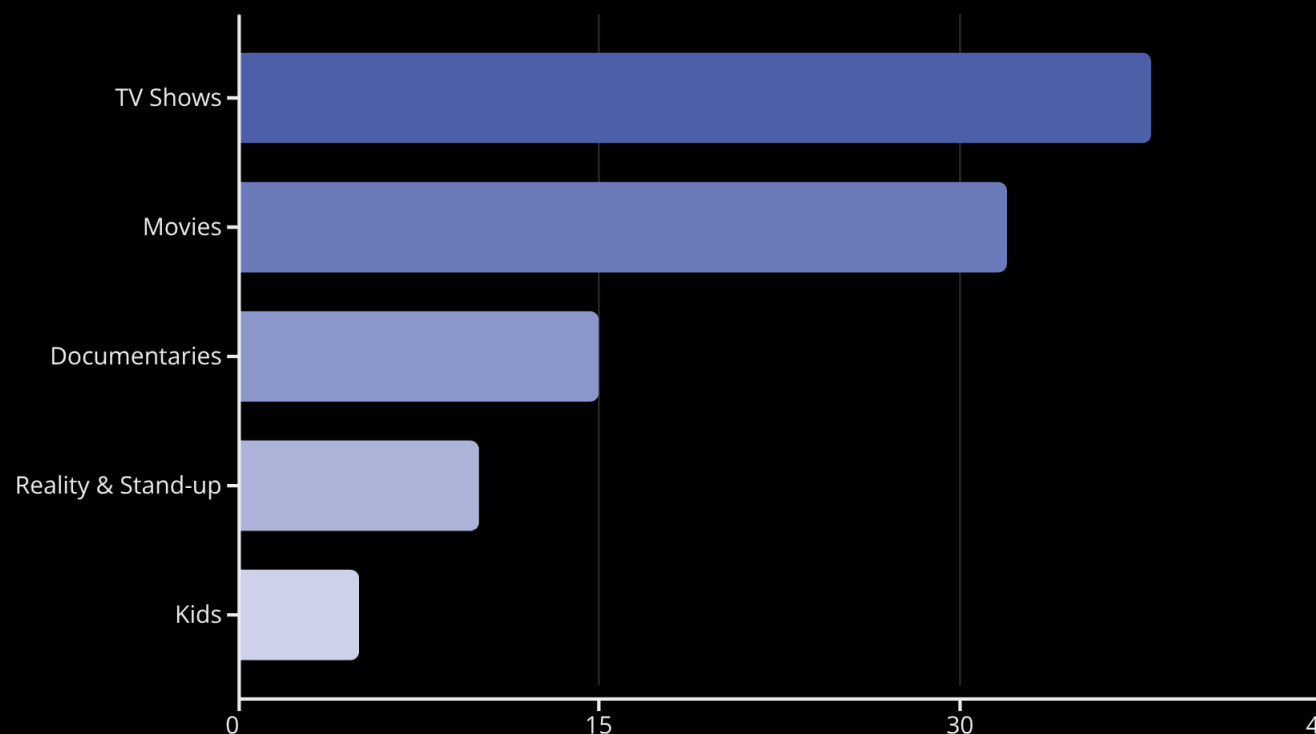
Неактивных

Пользователей более 30 дней – риск оттока

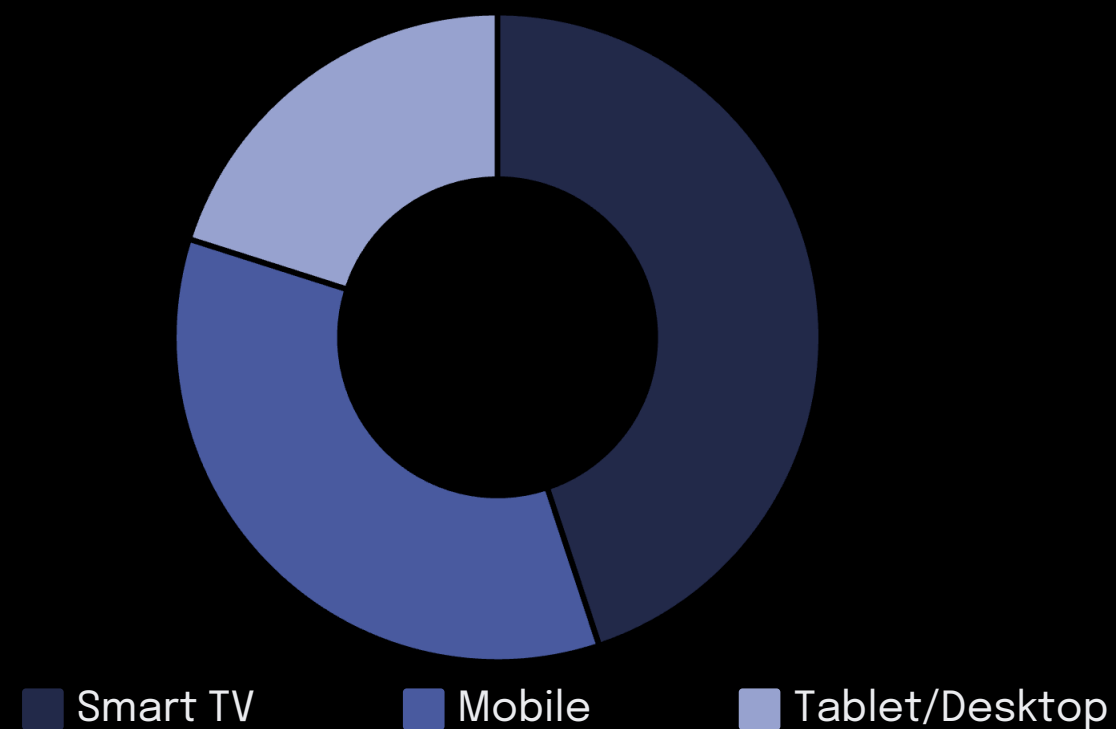
Пик подключений приходится на **январь** и **сентябрь** (новогодние подарки, возвращение после лета), минимум – июль (высокая мобильность, меньше времени на просмотр). Сезонность сильно влияет на привлечение, необходимы таргетированные кампании в «низкие» месяцы.

Предпочтения в контенте и устройствах

Популярные категории контента



Предпочитаемые устройства



📌 **Интересный паттерн:** Пользователи Premium чаще смотрят на Smart TV (домашний просмотр), Basic – на мобильных устройствах (мобильность, молодёжь)

Анализ по типу подписки и удержанию

Тип подписки	Доля пользователей	Часы просмотра/неделя	Отток за 6 мес
Basic	40%	6.2	22%
Standard	35%	8.7	14%
Premium	25%	11.3	8%

Лояльность Premium

Премиум-подписчики в 3 раза лояльнее базовых пользователей благодаря расширенному функционалу

Вовлечённость = Удержание

Прямая корреляция между временем просмотра и снижением оттока пользователей

Потенциал апселла

Возможность предложить Standard активным пользователям Basic с высокой активностью

Ключевые инсайты и стратегические рекомендации



Географическая стратегия

Фокус на США и Европу для максимизации доходов, но активное тестирование локализованного контента в Бразилии и Индии для захвата новых рынков



Контентная стратегия

Инвестировать в производство сериалов и документальных фильмов – они демонстрируют наивысший уровень удержания пользователей



Техническая оптимизация

Приоритетная оптимизация интерфейса под Smart TV – основной экран премиум-аудитории с высокой покупательной способностью



Программа удержания

Запуск персонализированных напоминаний неактивным пользователям через 21 день для предотвращения оттока



Монетизация

Программа апгрейда для активных пользователей Basic → Standard на основе поведенческих паттернов

Методология и технический стек

Инструменты анализа

- **Python:** pandas, numpy для обработки данных
- **Визуализация:** matplotlib, seaborn, plotly
- **Статистика:** scipy, scikit-learn
- **Среда разработки:** Jupyter Notebook

Ключевые метрики

- Churn Rate (коэффициент оттока)
- Customer Lifetime Value (CLV)
- Engagement Score (индекс вовлечённости)
- Geographic Penetration (географическое проникновение)



Анализ проводился с использованием современных методов data science, включая корреляционный анализ, кластеризацию пользователей и прогнозное моделирование поведения

Контакты и ресурсы проекта

Исходный код

https://github.com/singaevsky/portfolio_Python.git

Полный код анализа с комментариями и визуализациями

Датасет

Синтезирован на основе публичных данных и поведенческих паттернов (Kaggle-подобный формат)

Объём: 50,000+ записей пользовательской активности

Связаться со мной

Email: singaevsky@yandex.ru

GitHub: [@singaevsky](https://github.com/singaevsky)

LinkedIn: Сингаевский Андрей Николаевич

Спасибо за внимание! Готов ответить на вопросы и обсудить результаты исследования.

