



Давайте рекомендовать!

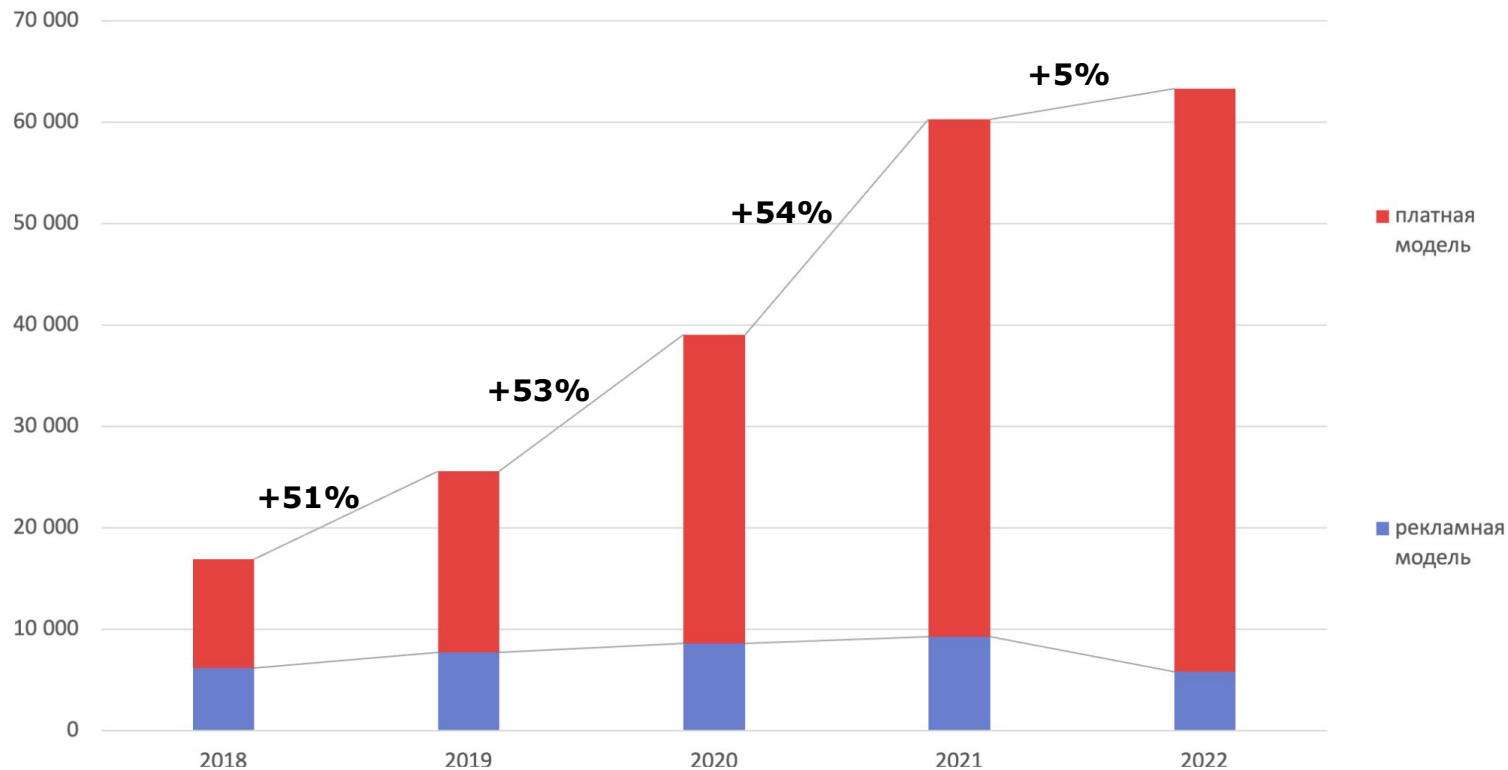
Как персонализировать
подписку KION?

Витвицкая Алиса
Гащенко Матвей
Полковникова Мария
Полянский Ян
Хомутова Александра

Инфоджежаи
12.2023

Актуальность

Выручка, млрд. рублей



* [расчёты TDaily](#)

План исследования

- 1 Исследование структуры датасета
- 2 Первичный анализ данных
- 3 Постановка исследовательского вопроса и гипотезы
- 4 Обработка данных и построение математической модели
- 5 Интерпретация результатов и выводы
- 6 Перспективы и ограничения
- 7 Практическая польза и policy implication

Структура датасета (за период 13.03.21-22.08.21)



Первичный анализ данных

Распределение по возрасту среди женщин



Распределение по возрасту среди мужчин

Средние и взаимосвязи между переменными

Среднее:

- время просмотра \approx **138 минут**
- кол-во просмотров у пользователя - **6**
- процент просмотра фильма \approx **46%**

Медиана:

- времени просмотра \approx **48 минут**
- кол-ва просмотров у пользователя - **2**
- процент просмотра фильма \approx **32%**

Мода количества просмотров у пользователя - **1**

Корреляция между total_dur и watched_pct \approx **0.189**

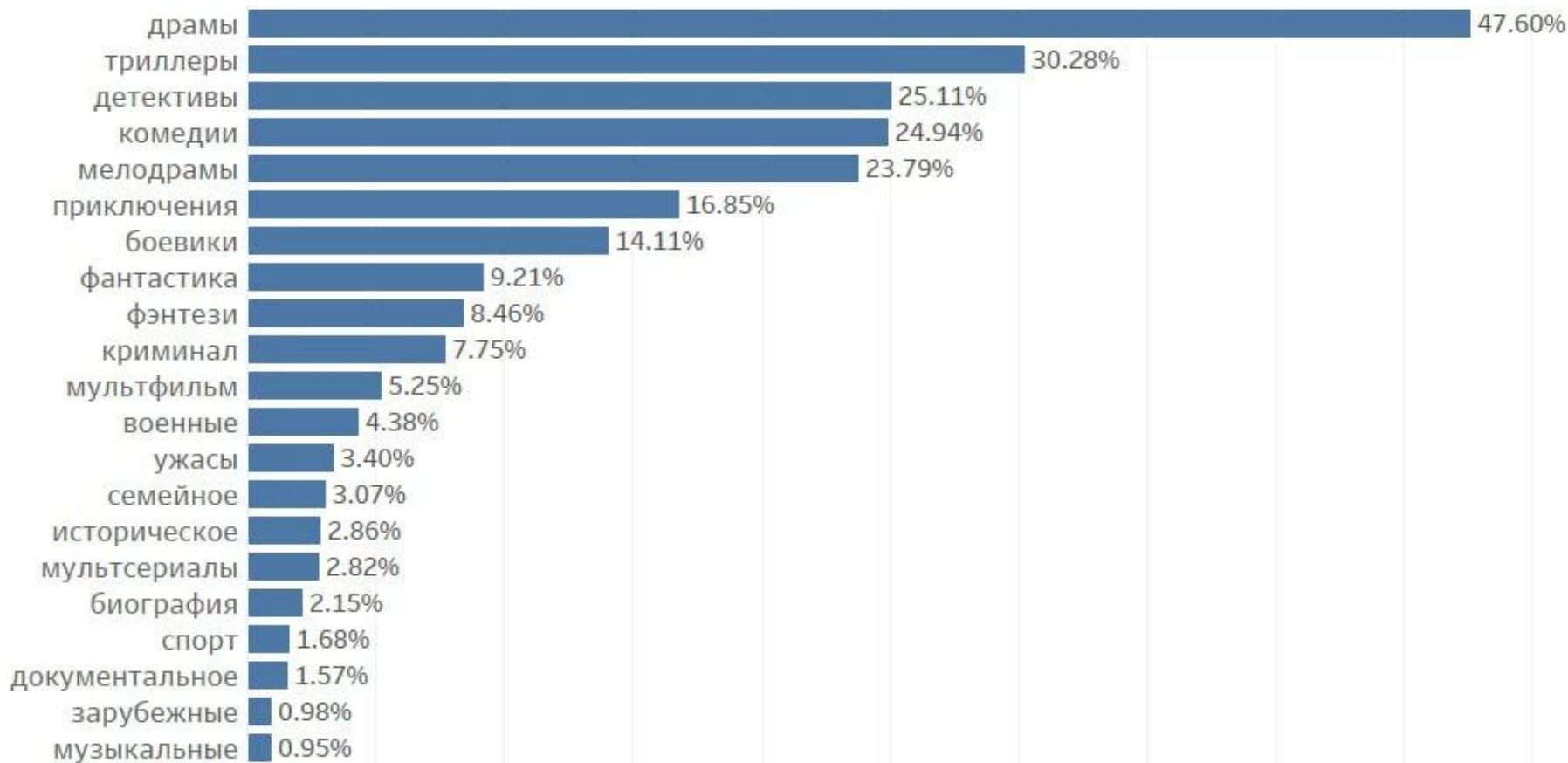
Распределение процентов просмотров контента



Обработка данных

- Соединили **три таблицы** в одну на основе таблицы interactions
- Удалили данные взаимодействий, полученные раньше релиза KION (**13 марта - 20 апреля 2021**)
- Убрали людей с **отсутствием информации** по полу и возрасту
- Убрали взаимодействия с процентом просмотра менее **80%** по фильмам и **60%** по сериалам
- Удалили пользователей, которые смотрели меньше **3 фильмов**
- Удалили пользователей, которые посмотрели **более 12 суток** контента
- **Разбили списки** жанров по одному на каждый контент

Процент от общего количества просмотров



Рассматриваемые данные

Окончательная выборка (без разбиения жанров) содержит:

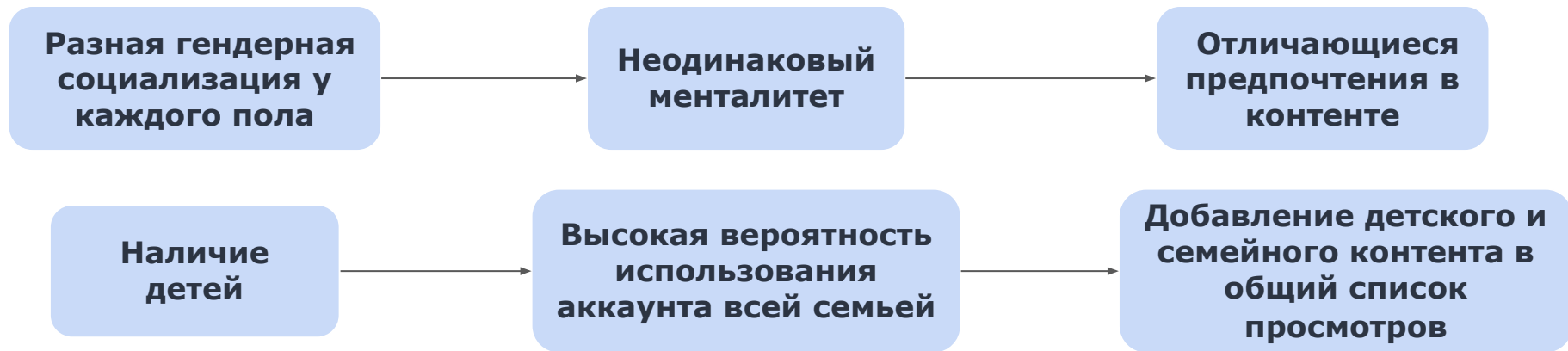
- **25.6%** (215 136) от исходного количества пользователей
- **20.1%** (1 147 440) от всех взаимодействий
- **67.2%** (10 737) от всего контента (фильмов и сериалов)

**Как пол и наличие детей влияют на
распределение просмотров по жанрам
фильмов и сериалов?**

KION

Для популярных жанров существует свой преобладающий портрет зрителя

Механизм:



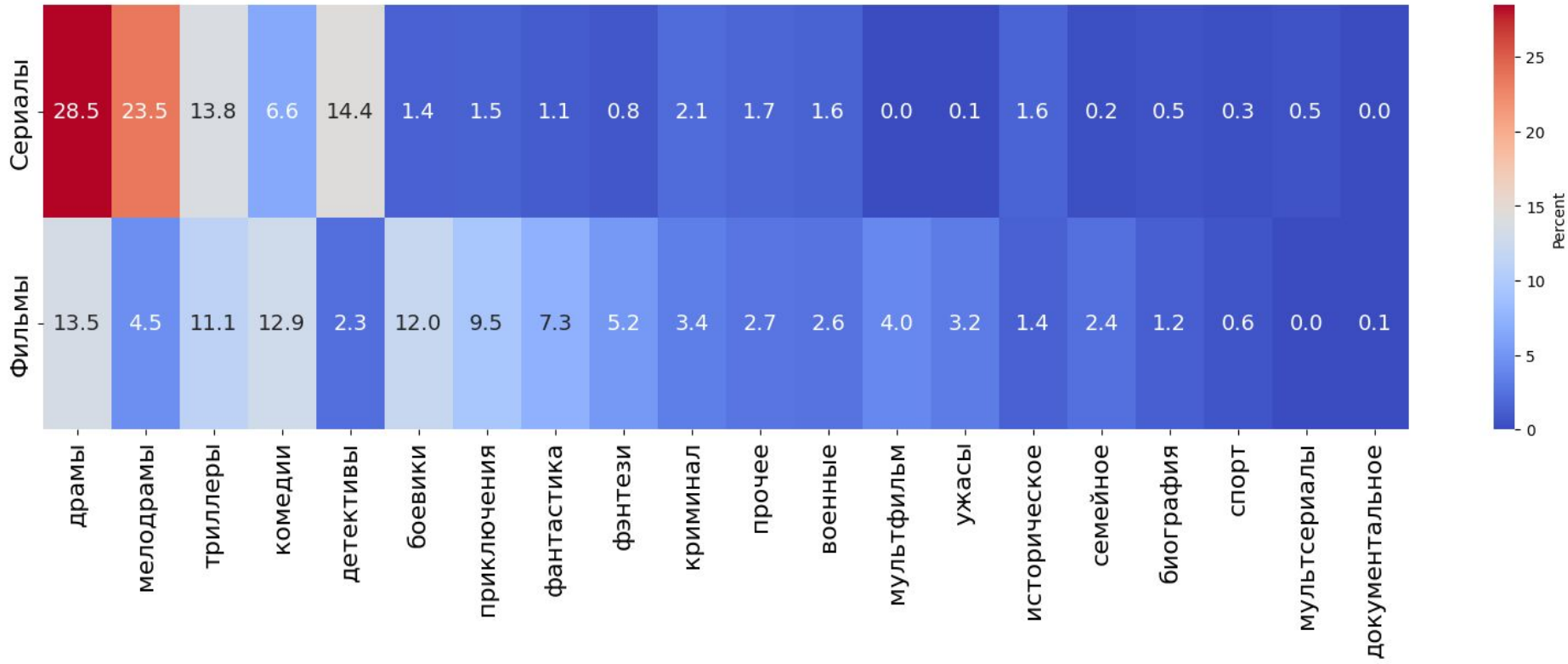
Гипотеза H_0

Для популярных жанров **не существует** выраженного портрета зрителя

Гипотеза H_1

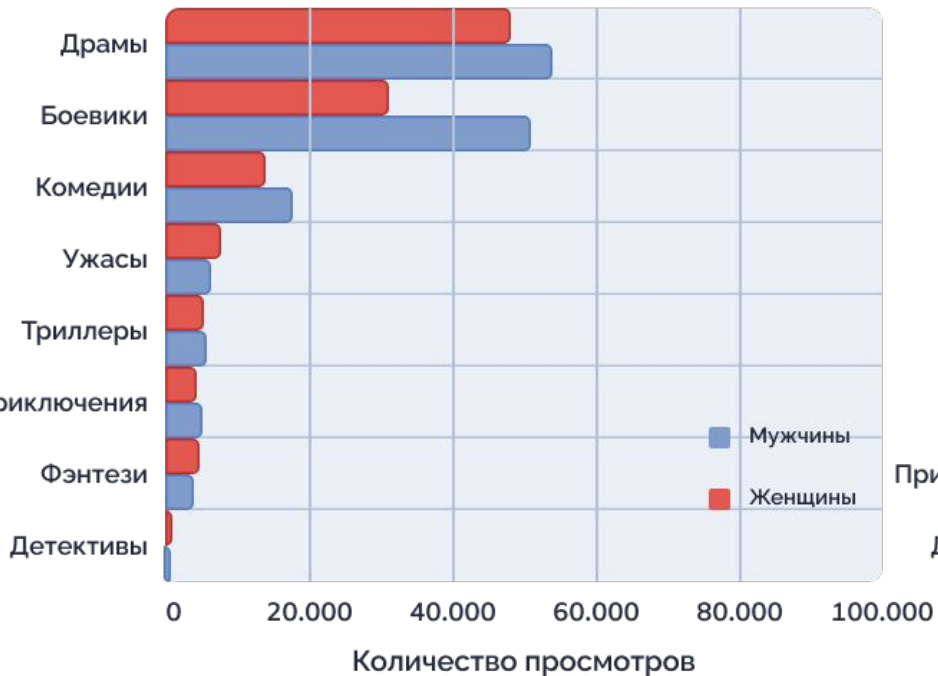
Для популярных жанров **существует** преобладающий портрет зрителя

Распределение просмотров жанров по типу контента (в процентах)

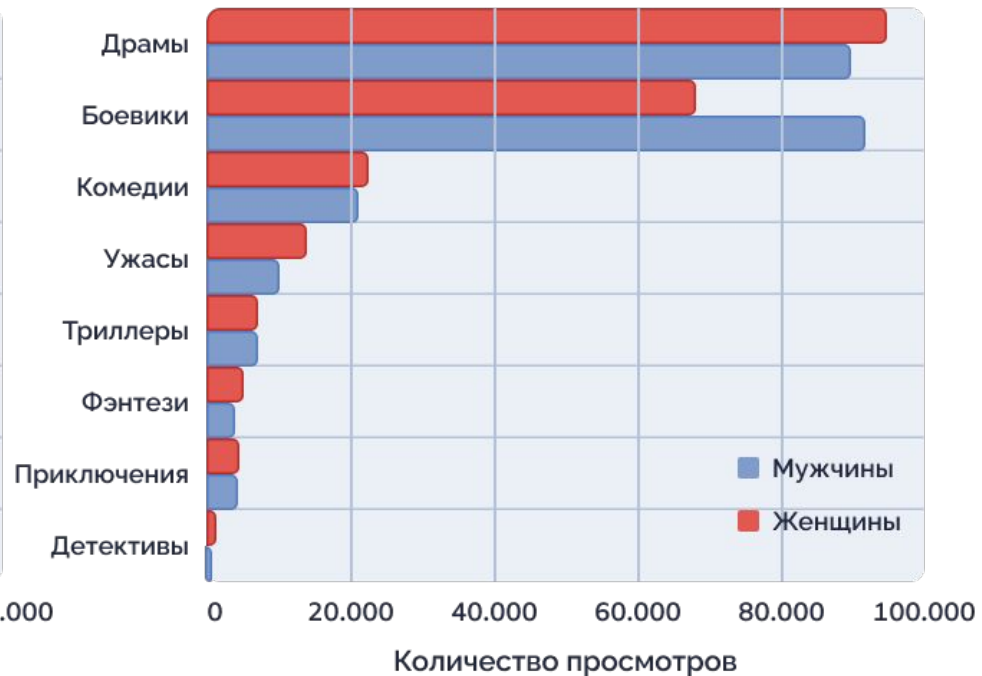


Распределение количества просмотров по фильмам в зависимости от пола и наличия детей

С детьми

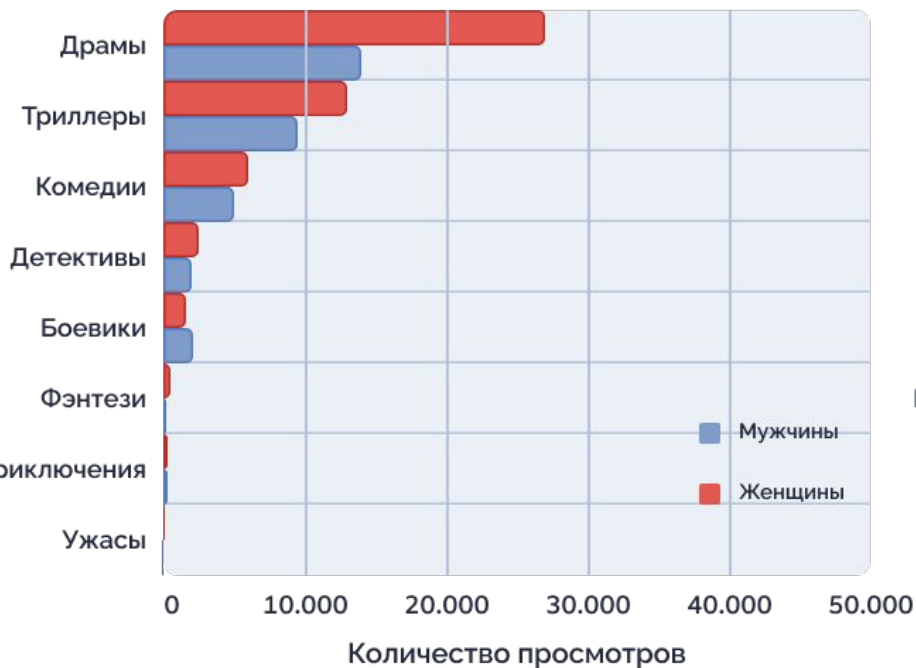


Без детей

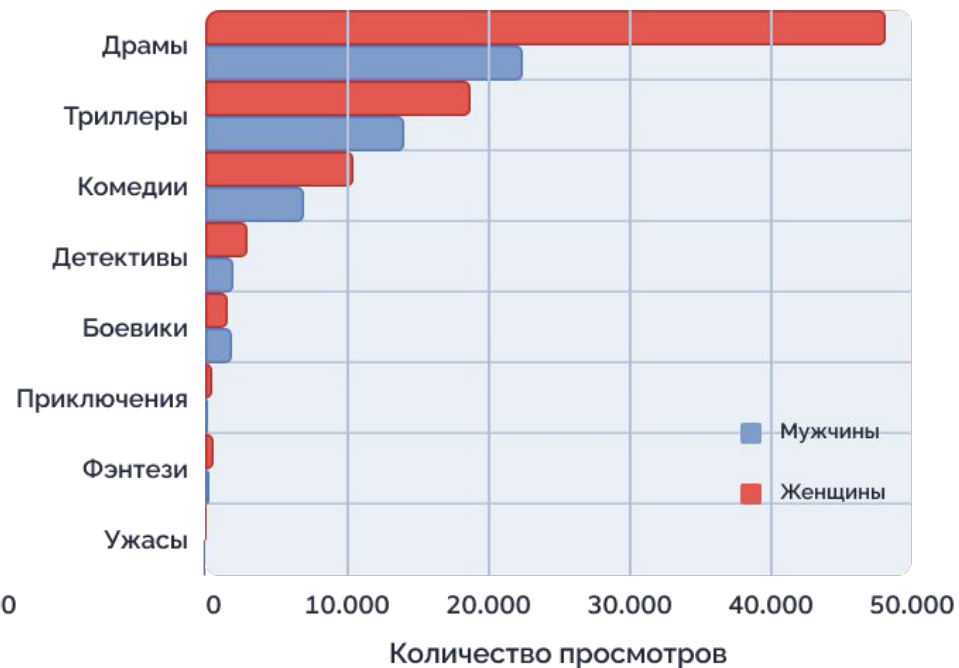


Распределение количества просмотров по сериалам в зависимости от пола и наличия детей

С детьми



Без детей



Стат.значимость для фильмов

	Z(U)-значимость без детей		Z(U)-значимость с детьми	
Драмы	1.00	В пользу женщин	1.00	Ужасы
Фэнтези	0.999		0.999	Фэнтези
Комедии	0.998		0.54	Детективы
Приключения	0.698	Отсутствует	0.588	Приключения
Детективы	0.722		0.711	Триллеры
Триллеры	0.801		0.953	Комедии
Ужасы	0.999	В пользу мужчин	0.999	Драмы
Боевики	1.00		1.00	Боевики

Стат.значимость для сериалов

	Z(U)-значимость без детей		Z(U)-значимость с детьми	
Драмы	1.00	В пользу женщин	1.00	Драмы
Триллеры	0.999		0.999	Комедии
Комедии	0.998		0.664	Детективы
Детективы	0.671	Отсутствует	0.685	Фэнтези
Фэнтези	0.737		0.764	Ужасы
Ужасы	0.689		0.797	Приключения
Приключения	0.931		0.999	Триллеры
Боевики	1.00	В пользу мужчин	1.00	Боевики

Результат работы модели

Фильмы

p-value = 0.05

С детьми

Без детей

	С детьми						Без детей					
	Возраст						Возраст					
	18-24	25-34	35-44	45-54	55-64	65+	18-24	25-34	35-44	45-54	55-64	65+
Драмы	0.999	1.00	0.999	0.999	0.54	0.85	0.999	1.00	0.999	0.999	0.785	0.646
Ужасы	0.996	0.997	0.999	0.999	0.849	0.504	0.998	0.999	0.998	0.999	0.98	0.795
Комедии	0.89	0.951	0.954	0.999	0.792	0.566	0.872	0.999	0.999	1.00	0.981	0.984
Боевики	1.00	1.00	0.999	1.00	0.999	1.00	0.999	0.999	1.00	0.999	0.999	1.00
Триллеры	0.582	0.519	0.512	0.626	0.587	0.611	0.997	0.818	0.585	0.899	0.738	0.58
Фэнтези	0.85	0.997	0.996	0.962	0.64	0.787	0.951	0.999	0.953	0.995	0.952	0.798
Приключения	0.52	0.655	0.58	0.608	0.677	0.554	0.75	0.702	0.77	0.556	0.746	0.689
Детективы	0.582	0.518	0.512	0.626	0.587	0.612	0.636	0.52	0.574	0.6	0.512	0.631



Модель не ошиблась



Модель ошиблась

Результат работы модели

Сериалы

p-value = 0.05

С детьми

Без детей



Гипотеза: Для популярных жанров существует свой портрет зрителя

фильмы

Приняли **H₁** для:

- 12 жанров из 17, которые смотрят пользователи с детьми
- 13 жанров из 17, которые смотрят пользователи без детей

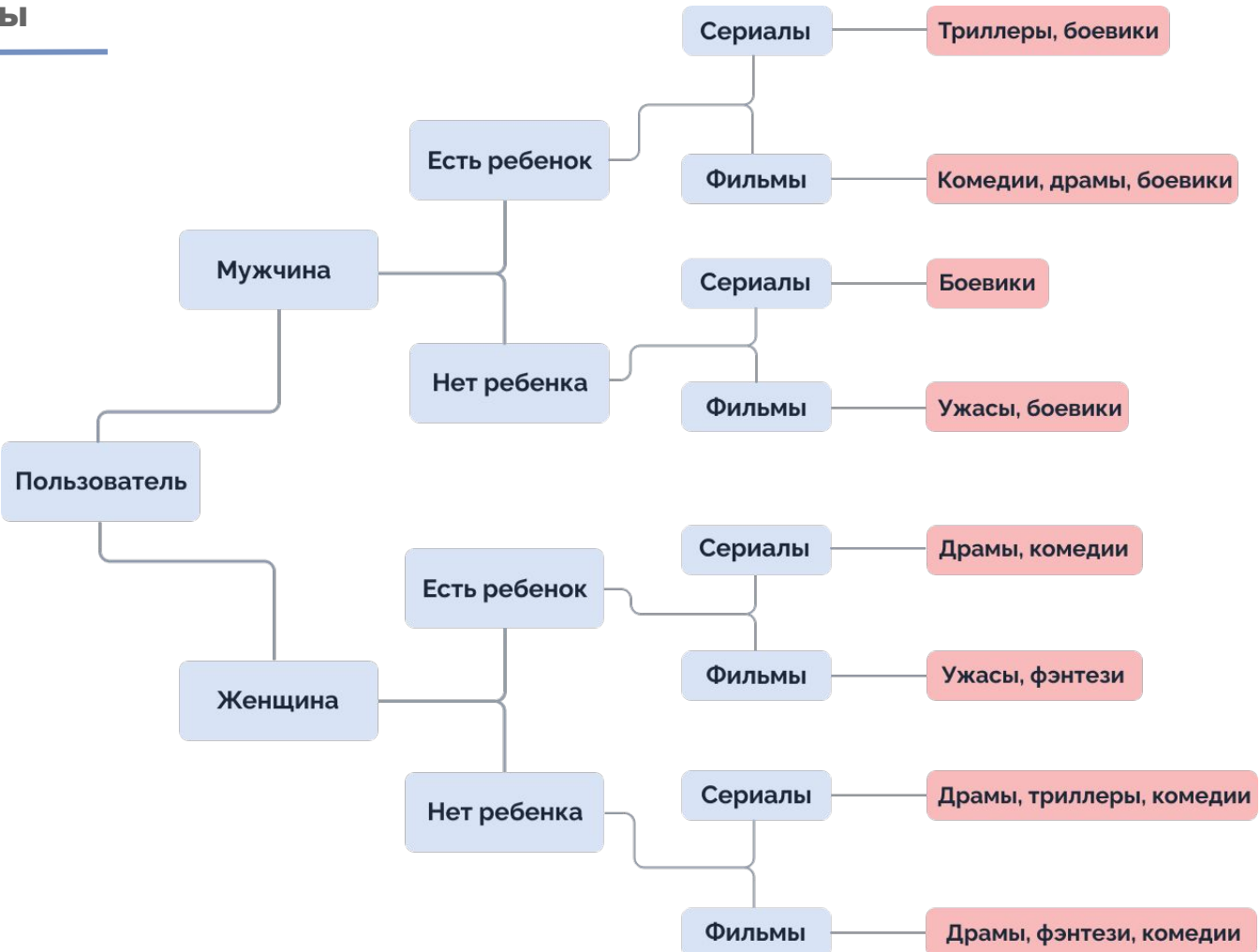
сериалы

Приняли **H₀** для:

- 12 жанров из 18, которые смотрят пользователи с детьми
- 12 жанров из 18, которые смотрят пользователи без детей

Для старших возрастных групп (от 55 лет) матмодель часто ошибается

Выводы



Перспективы

- Собрать данные с KION по пользователям, в том числе пользовательские оценки контента (нам даны только рейтинги с Кинопоиска)
- Собрать данные по взаимодействиям за более долгий период времени, чтобы для каждого жанра было достаточное количество взаимодействий, и математическая модель работала хорошо и на них

Ограничения

- Не понятно, пользуется аккаунтом один человек или несколько
- Выводы исследования нельзя обобщить на настоящее время
- Нельзя точно сказать, понравился ли пользователю просмотренный контент

Практическая польза

KION может предлагать более персонализированный контент пользователям, тем самым привлекая большее количество людей к покупке подписки

Policy implication

Наше исследование помогает понять, что в первую очередь стоит рекомендовать новым пользователям

Для уже существующих пользователей можно скорректировать рекомендательную систему

KION

KION



KION

KION

Наша команда

Ян



Прогер

Алиса



Мастер
статистики

Маша



Тимлид

Саша



Дизайнер

Матвей



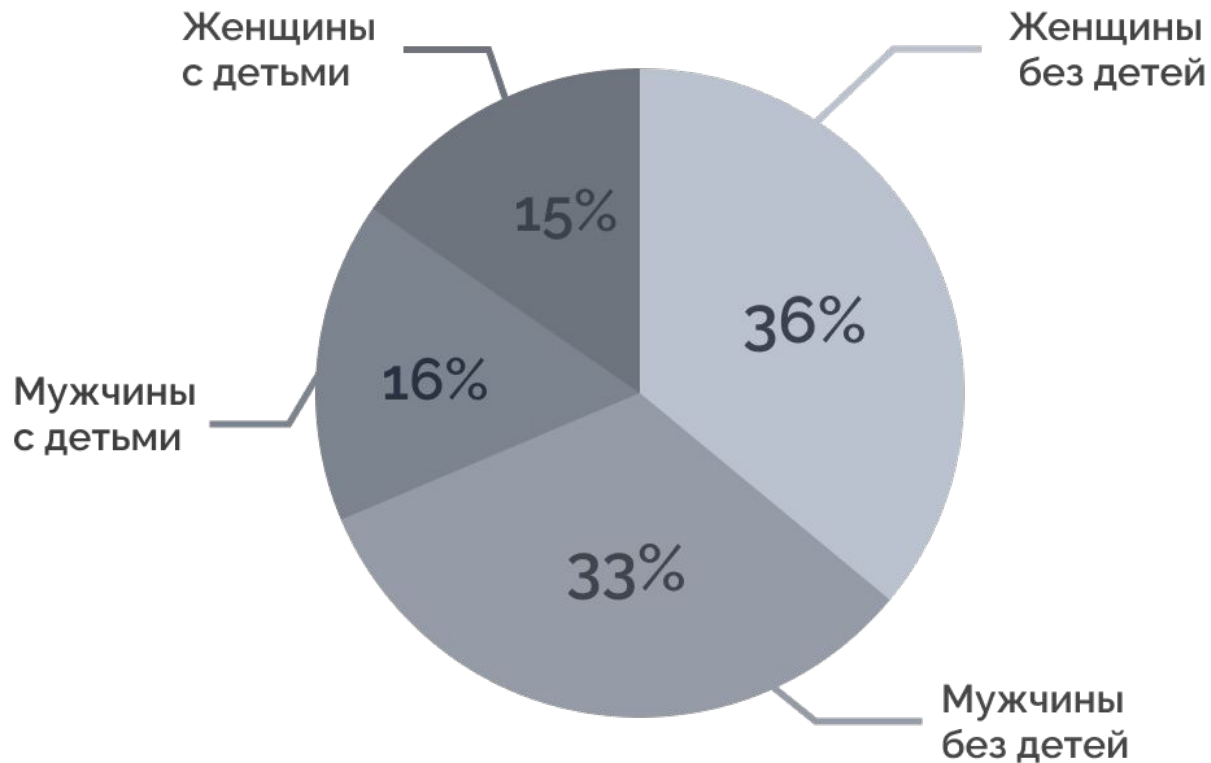
Прогер

ПРИЛОЖЕНИЕ

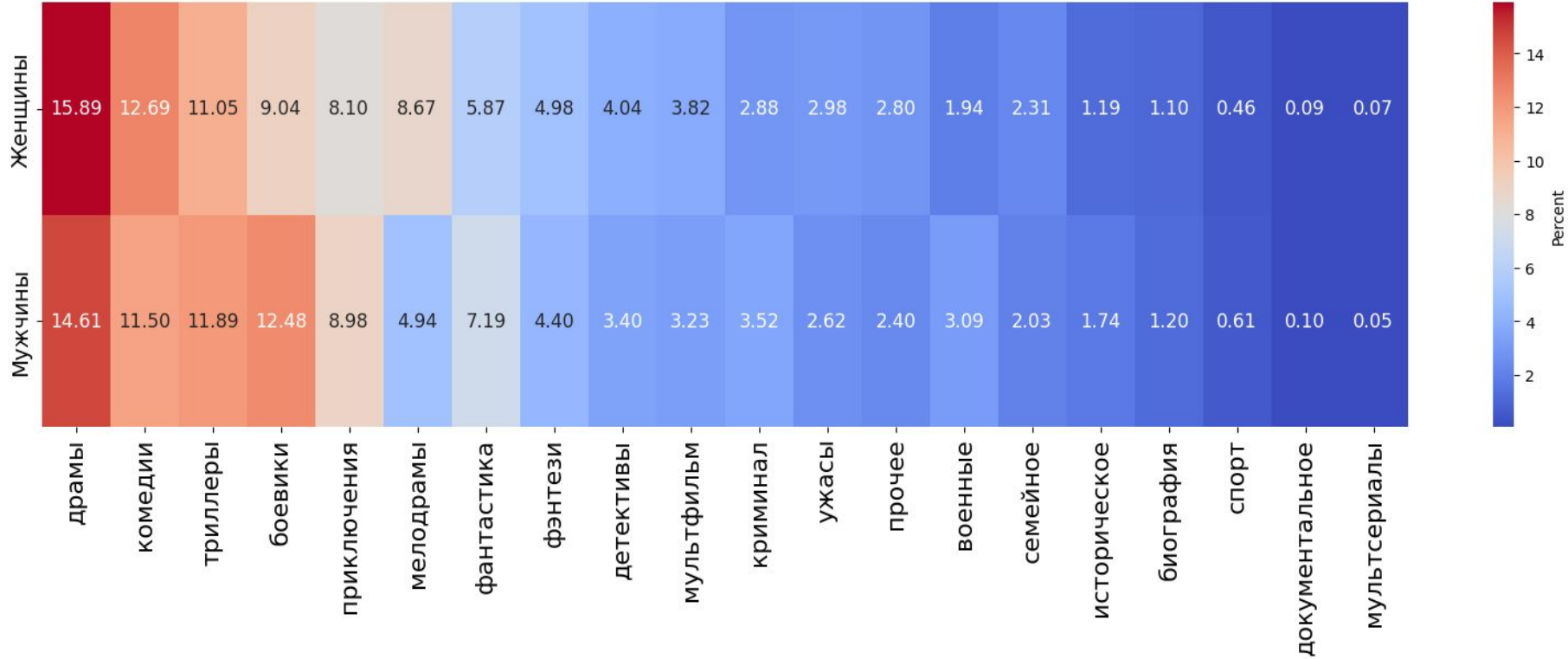
Источники информации:

- <https://web.archive.org/web/20150114182137/http://psystatistica.ru/index.php/stati/neparametricheskie-kriterii> - непараметрические критерии
- <https://scholar.harvard.edu/files/shleifer/files/beliefsaboutgender2.2019.pdf> - гендерная социализация

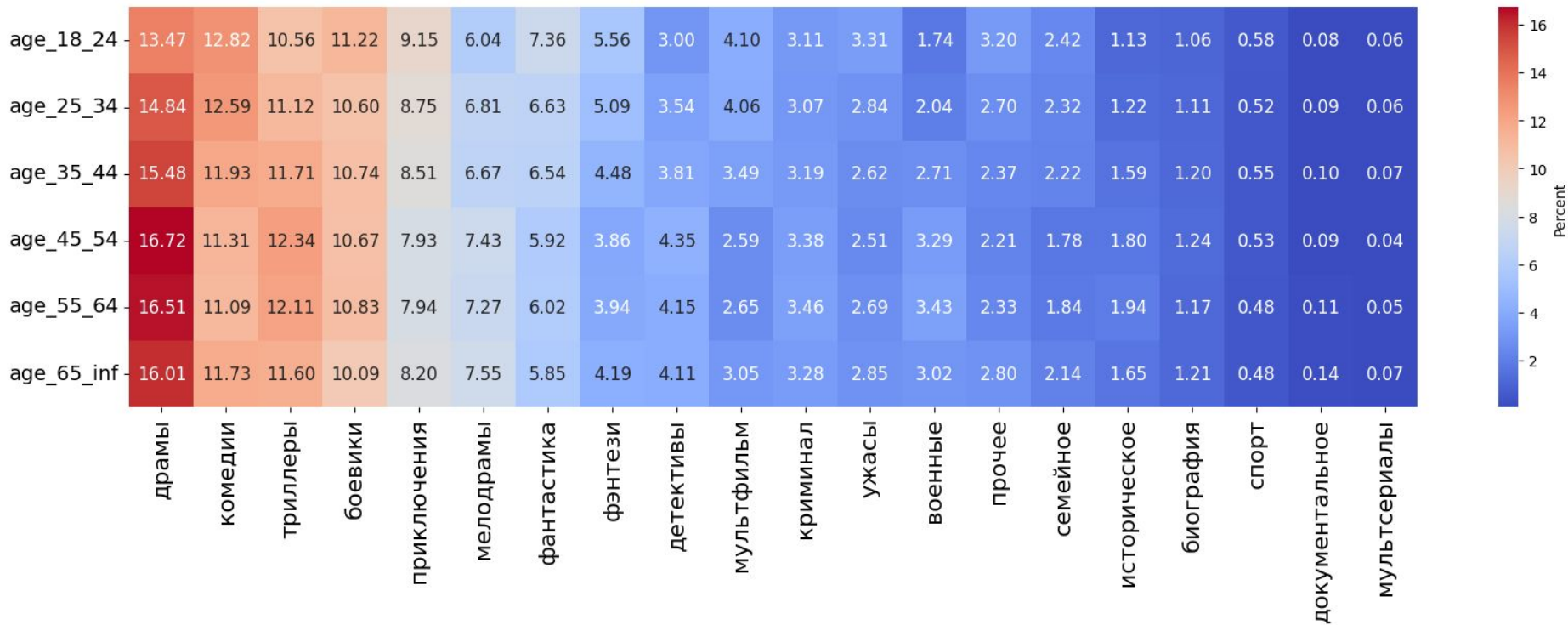
Распределение по детям и полу



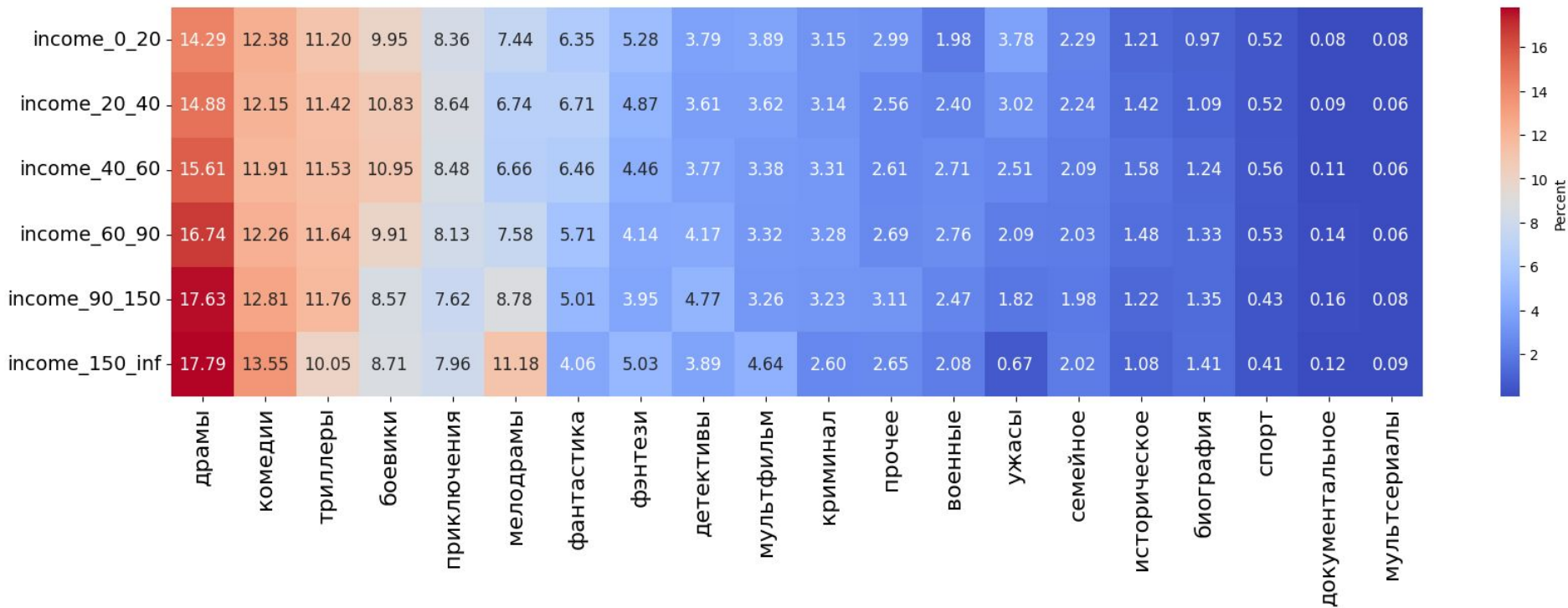
Распределение просмотров жанров по полу (в процентах)



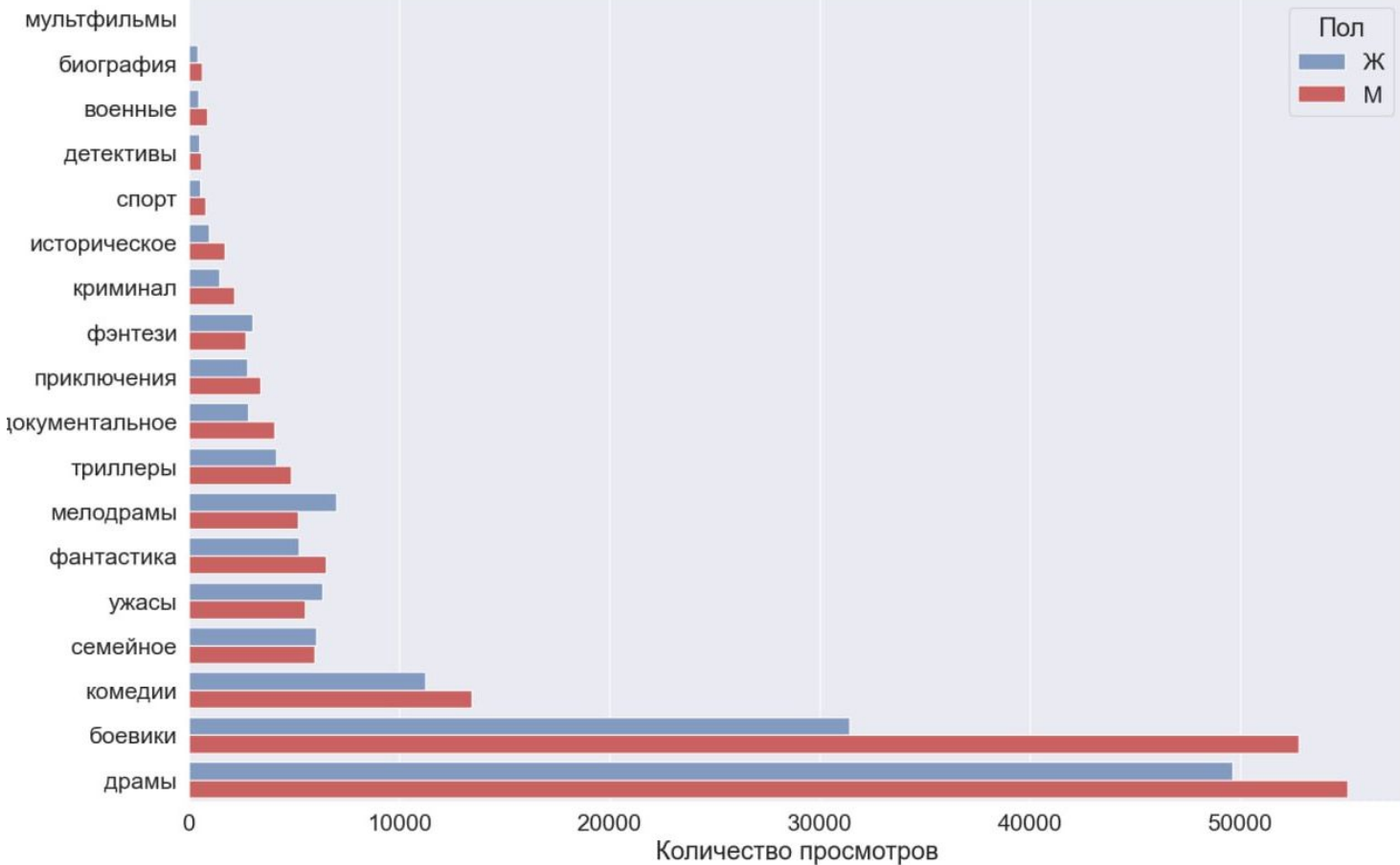
Распределение просмотров жанров по возрасту (в процентах)



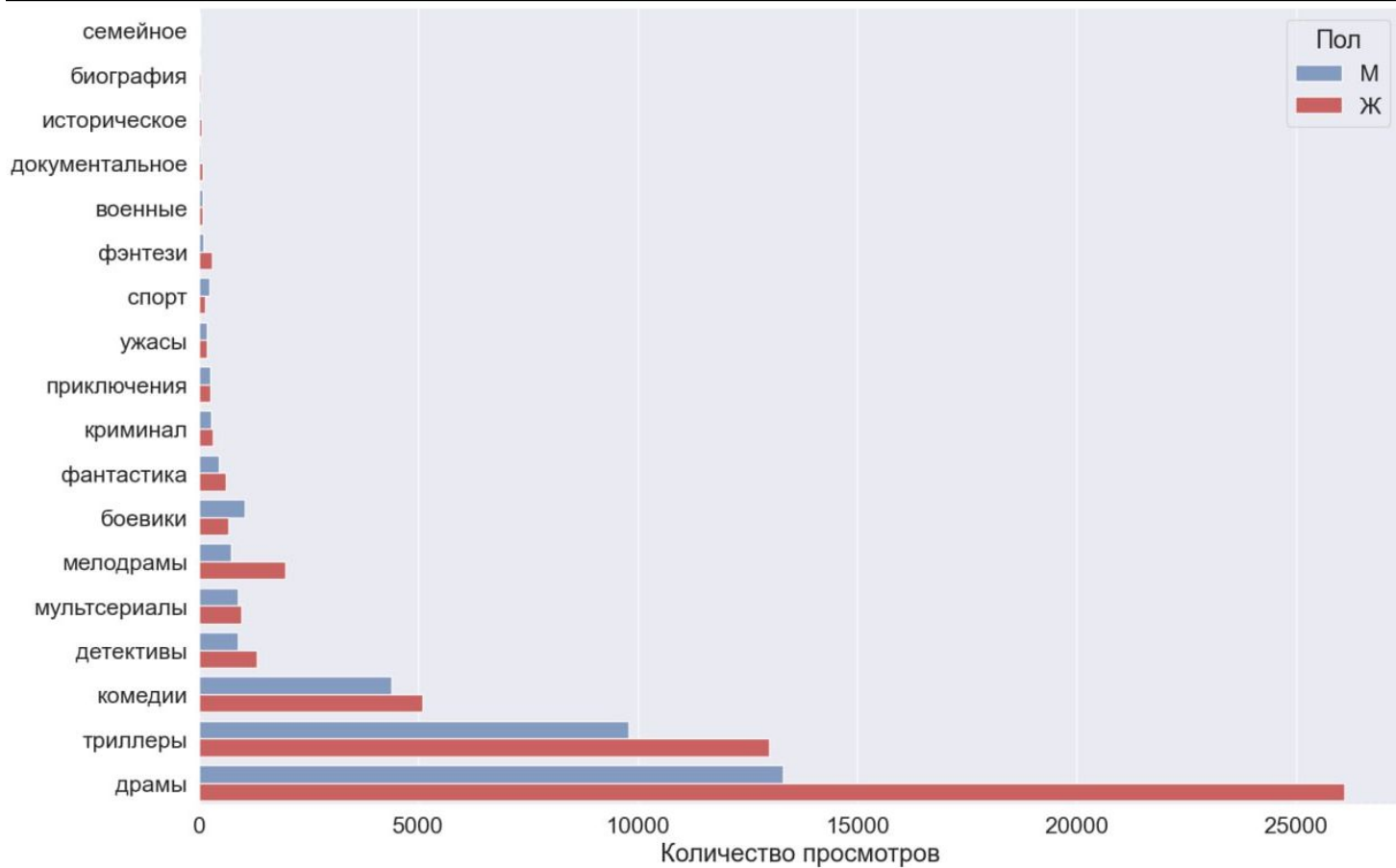
Распределение просмотров жанров по зарплате (в процентах)



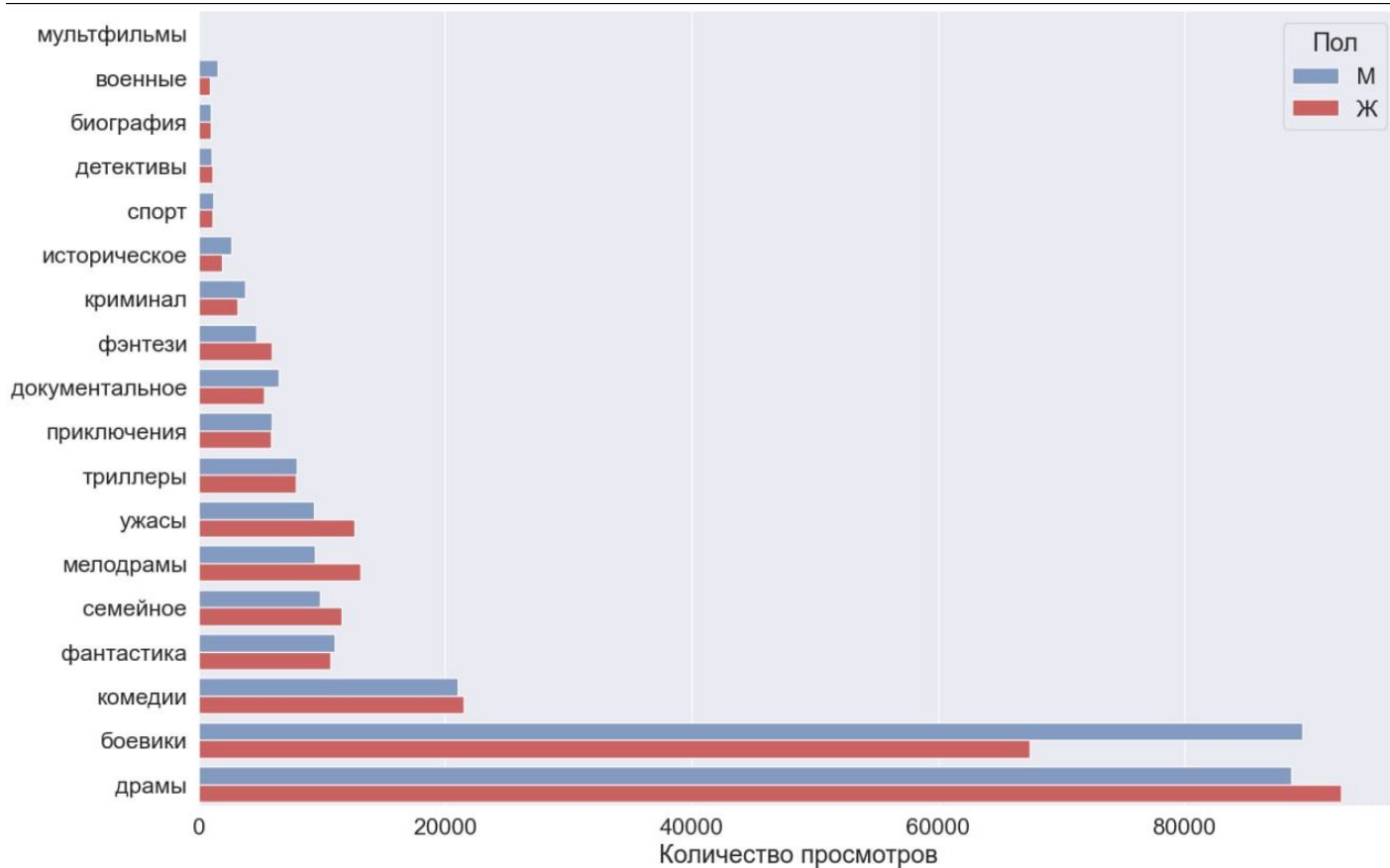
Фильмы с детьми



Сериалы с детьми



Фильмы без детей



Сериалы без детей

