

# Анализ влияния технических параметров на восприятие качества связи

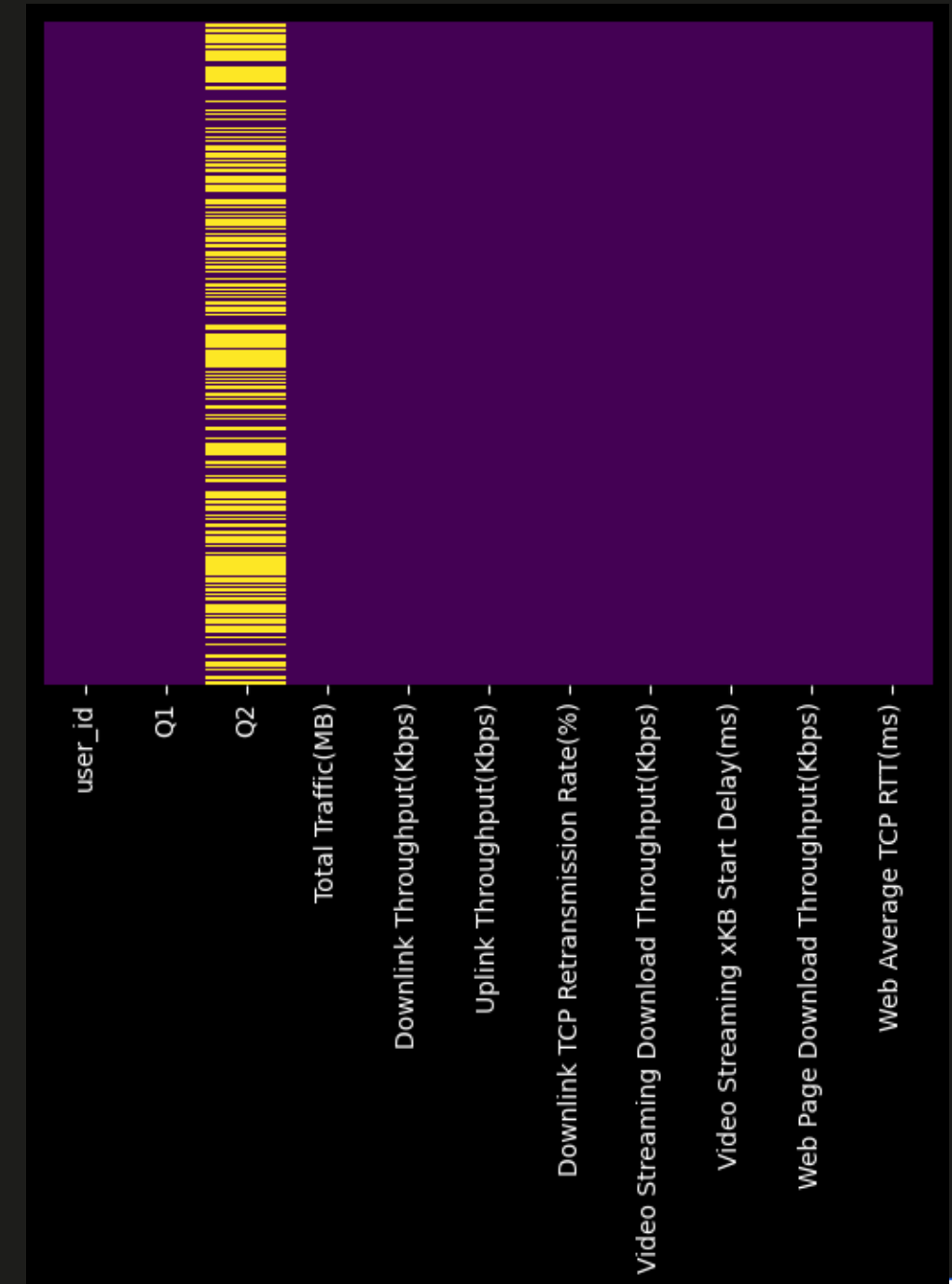
В рамках исследования клиентского опроса и технических метрик связи от компании «Мегафон» мы анализируем, какие технические параметры оказывают влияние на восприятие качества связи пользователями (оценка Q1).

Цель анализа — выявить значимые факторы, связанные с неудовлетворённостью пользователей, чтобы определить приоритетные направления для технических улучшений сети.

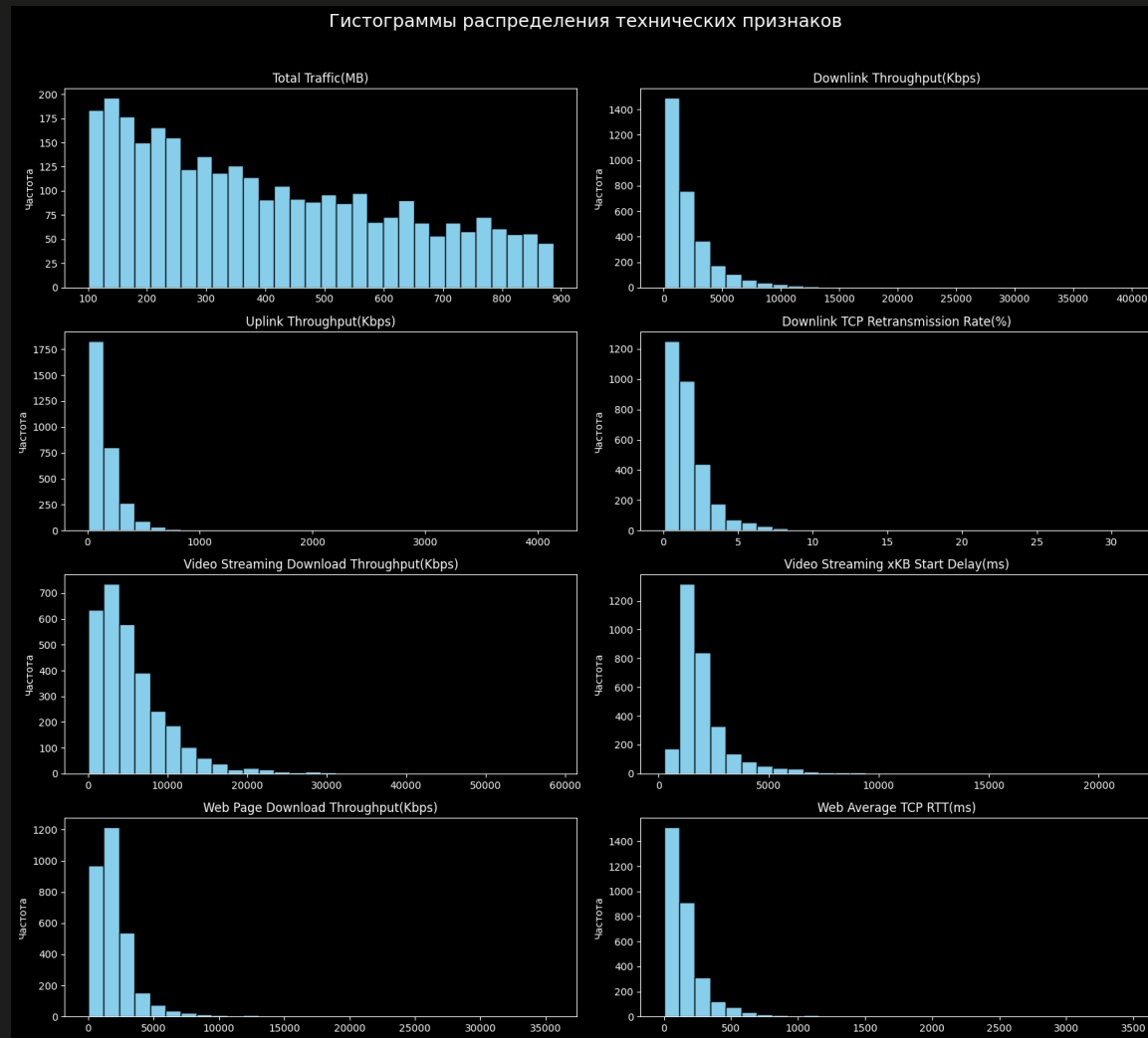
# Проверка и предобработка данных

Нулевые значения встречаются только в Q2 — это нормально, не всем пользователям задавали второй вопрос. Все признаки приведены к корректным типам данных.

Столбец Q1 изначально содержал строковые и другие некорректные значения — оставлены только числовые значения в диапазоне от 1 до 10.



# Разведочный анализ

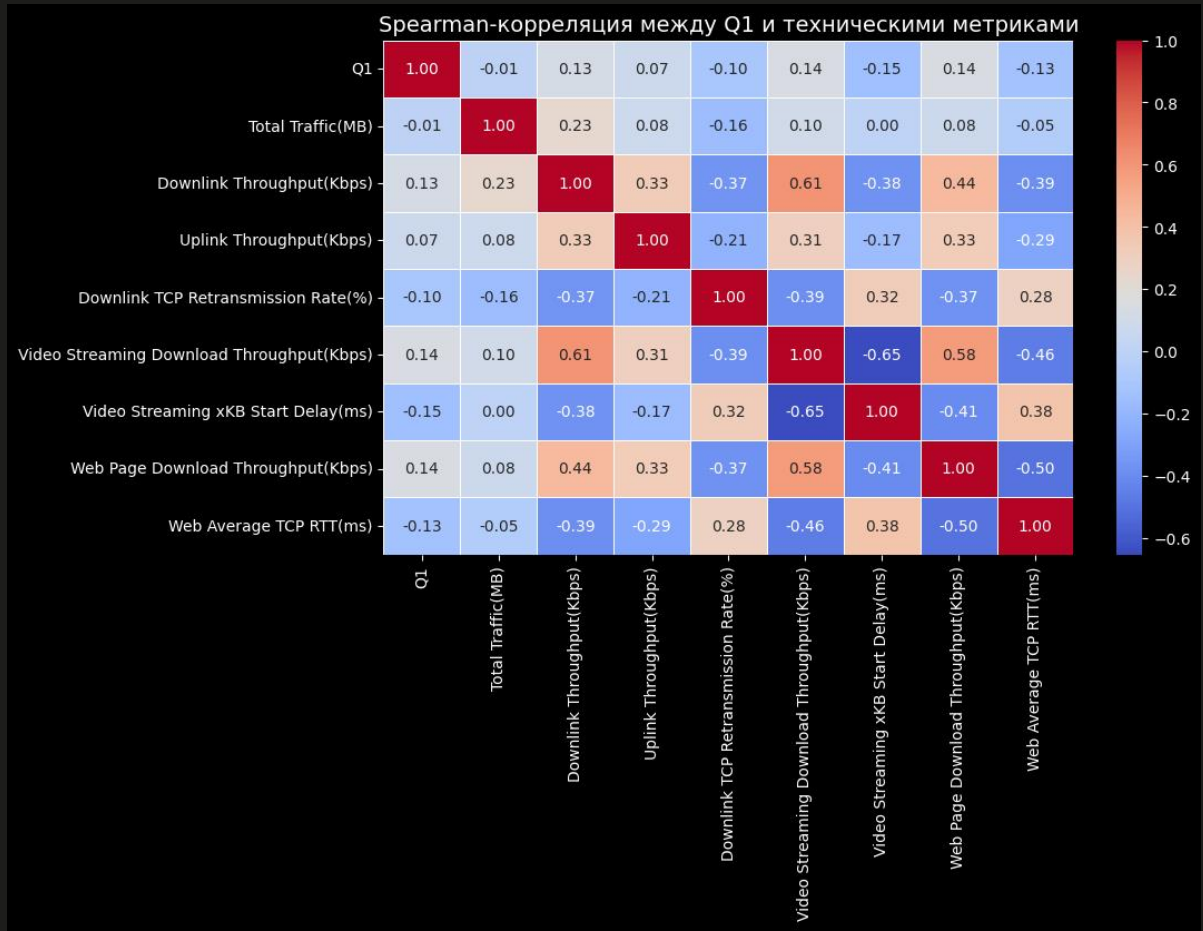


Большинство технических метрик (скорости, задержки, пинг) имеют асимметричное распределение с выраженной правой скошенностью.

Выбросы присутствуют практически во всех технических признаках, особенно в задержках и скоростях. Это следует учитывать при дальнейшем анализе и при выборе статистических методов.

Для дальнейших шагов (например, проверки гипотез и построения моделей) будет разумно применить логарифмическое преобразование или непараметрические методы.

Самую высокую негативную корреляцию с Q1 наблюдаем у признака **Video Streaming Start Delay** (задержка старта видео). Признак имеет сильно скошенное распределение с длинным правым хвостом и множеством выбросов.

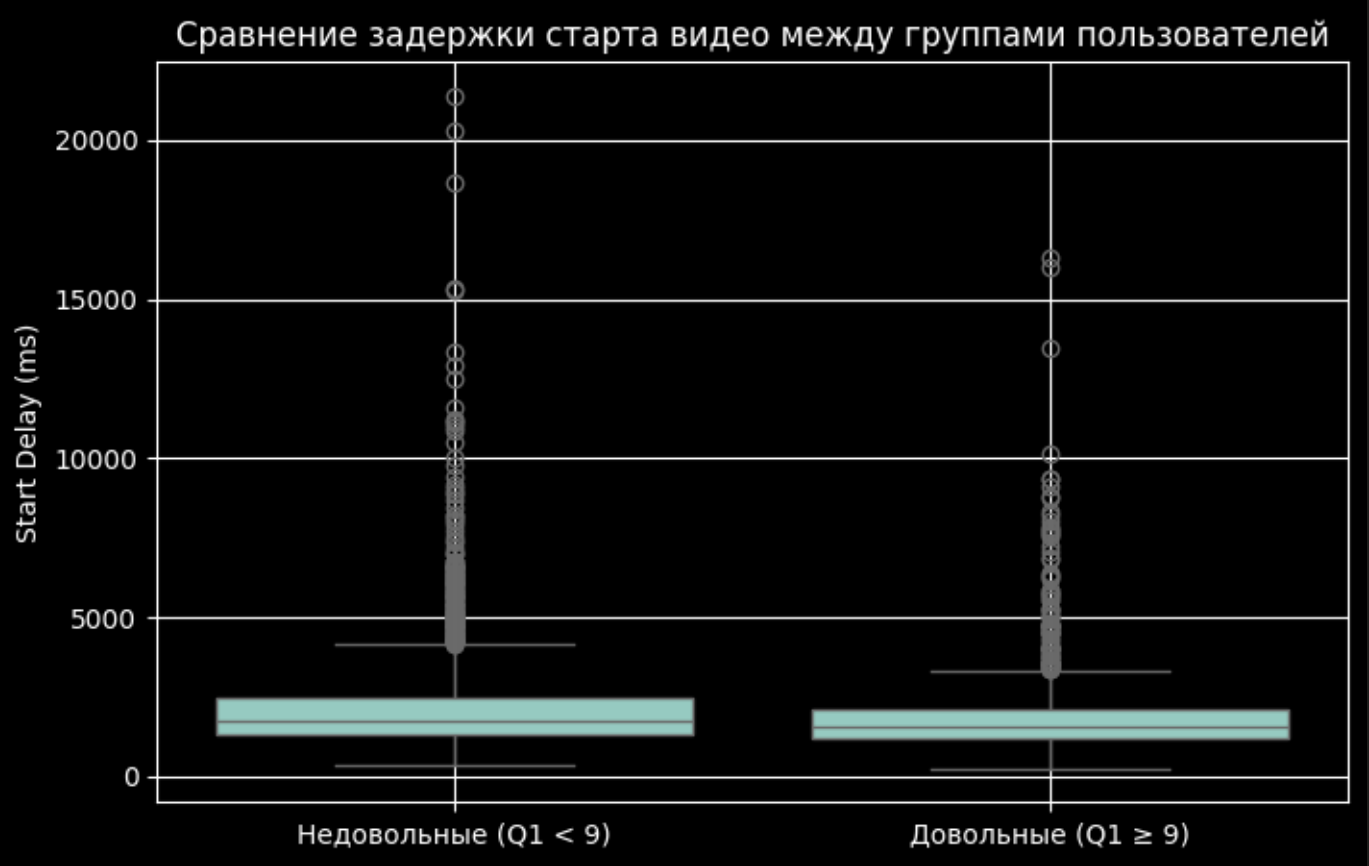


### Корреляция

Слабая отрицательная корреляция с оценкой Q1 ( $\rho \approx -0.15$  по Спирману)

### Различие между группами

Средняя задержка у недовольных  $\approx 2217$  мс, у довольных  $\approx 1918$  мс. Медиана отличается на  $\approx 172$  мс



### Предварительный вывод

Систематическое смещение в значениях задержки между группами указывает на предпосылки к статистически значимому различию

# Проверка первой гипотезы: задержка старта видео

## Формулировка гипотезы

**H<sub>0</sub>:** распределения **Video Streaming Start Delay** не отличаются между довольными и недовольными пользователями

**H<sub>1</sub>:** **Video Streaming Start Delay** у недовольных пользователей статистически значимо ниже, чем у довольных

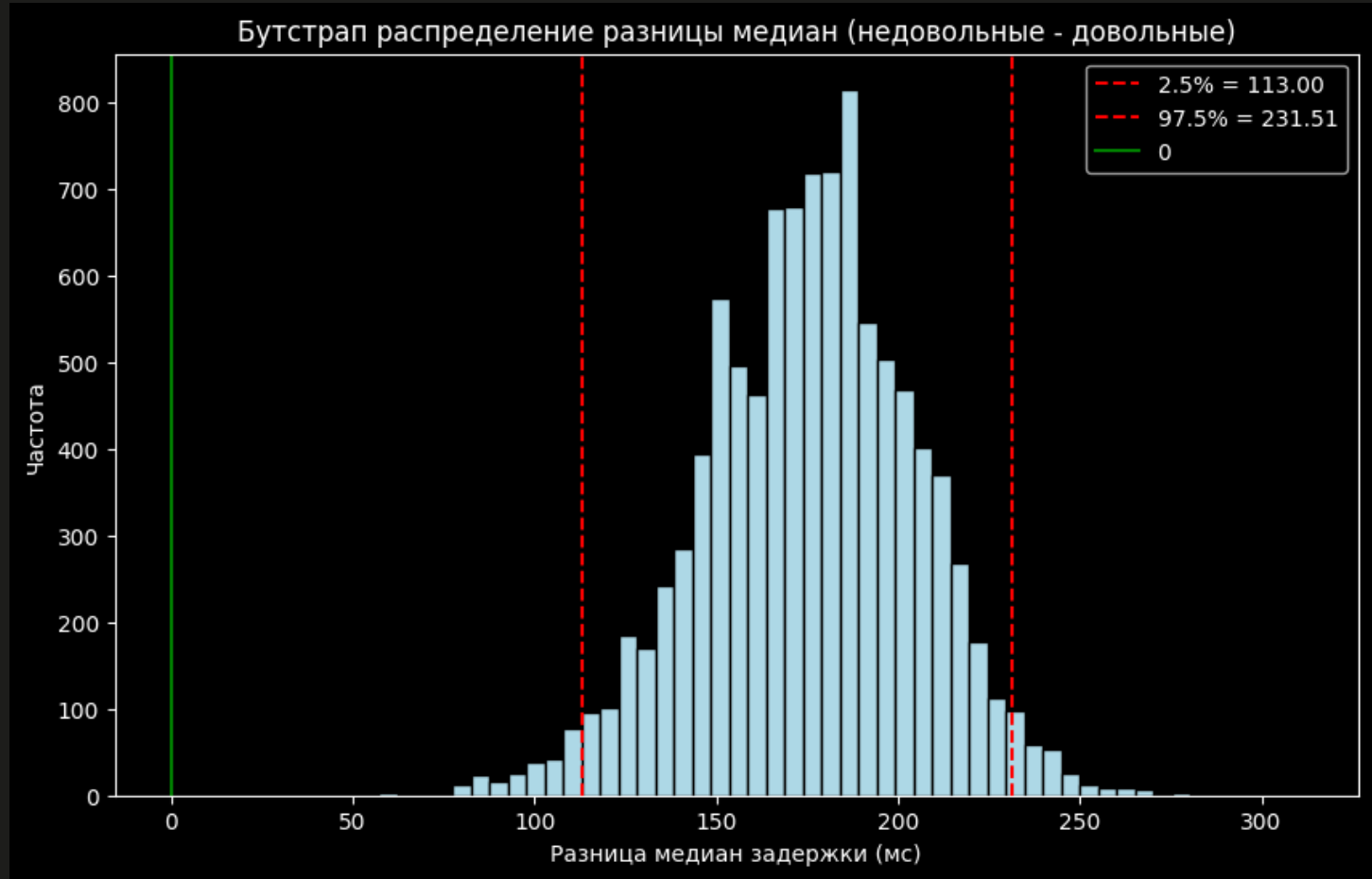
☒ Выбрана односторонняя проверка на основе предварительного анализа отрицательной корреляции

## Результаты теста

**U-статистика:** 1243574.00

**p-value:** 0.0000

Статистически значимое различие подтверждено



95% доверительный интервал разницы медиан: [113.00, 231.51] мс

## Вывод по первой гипотезе

Увеличение времени старта воспроизведения видео оказывает негативное влияние на восприятие качества связи со стороны пользователей.

# Анализ жалоб пользователей Q2 (топ 3)



## Причина №3

Плохое качество связи в зданиях, ТЦ —  
*невозможно исследовать количественно*



## Причина №1

Долгое ожидание гудков при звонке —  
*отсутствуют метрики голосовой связи*



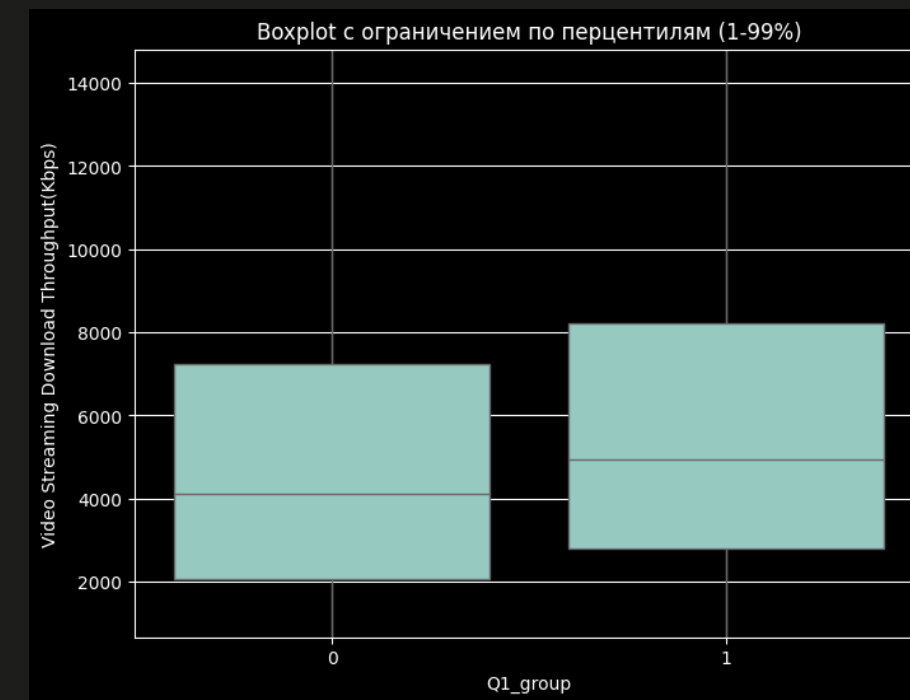
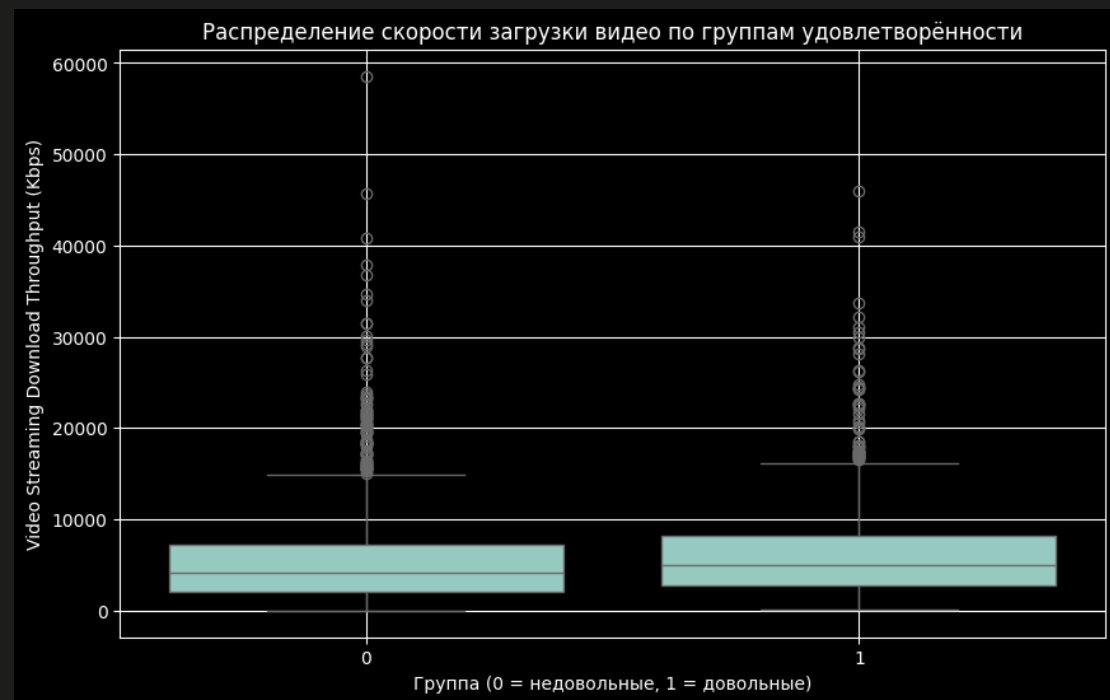
## Причина №4

Медленный мобильный интернет — приоритет для  
анализа

По корреляционной матрице после задержки самую высокую имеют:

- Video Streaming Download Throughput
- Web Page Download Throughput

Начнём с пропускной способности потоковой загрузки видео



Видим явное различие в распределении между группами довольных и недовольных пользователей. Недовольные пользователи имели более низкие значения метрики.

Различие между группами

Средняя пропускная способность у недовольных  $\approx 4100$  kbps у довольных  $\approx 4938$  kbps. Медиана отличается на  $\approx 838$  kbps

## Формулировка гипотезы

**H<sub>0</sub>:** распределения Video Streaming Download Throughput не отличаются между довольными и недовольными пользователями

**H<sub>1</sub>:** Video Streaming Download Throughput у недовольных пользователей статистически значимо ниже, чем у довольных

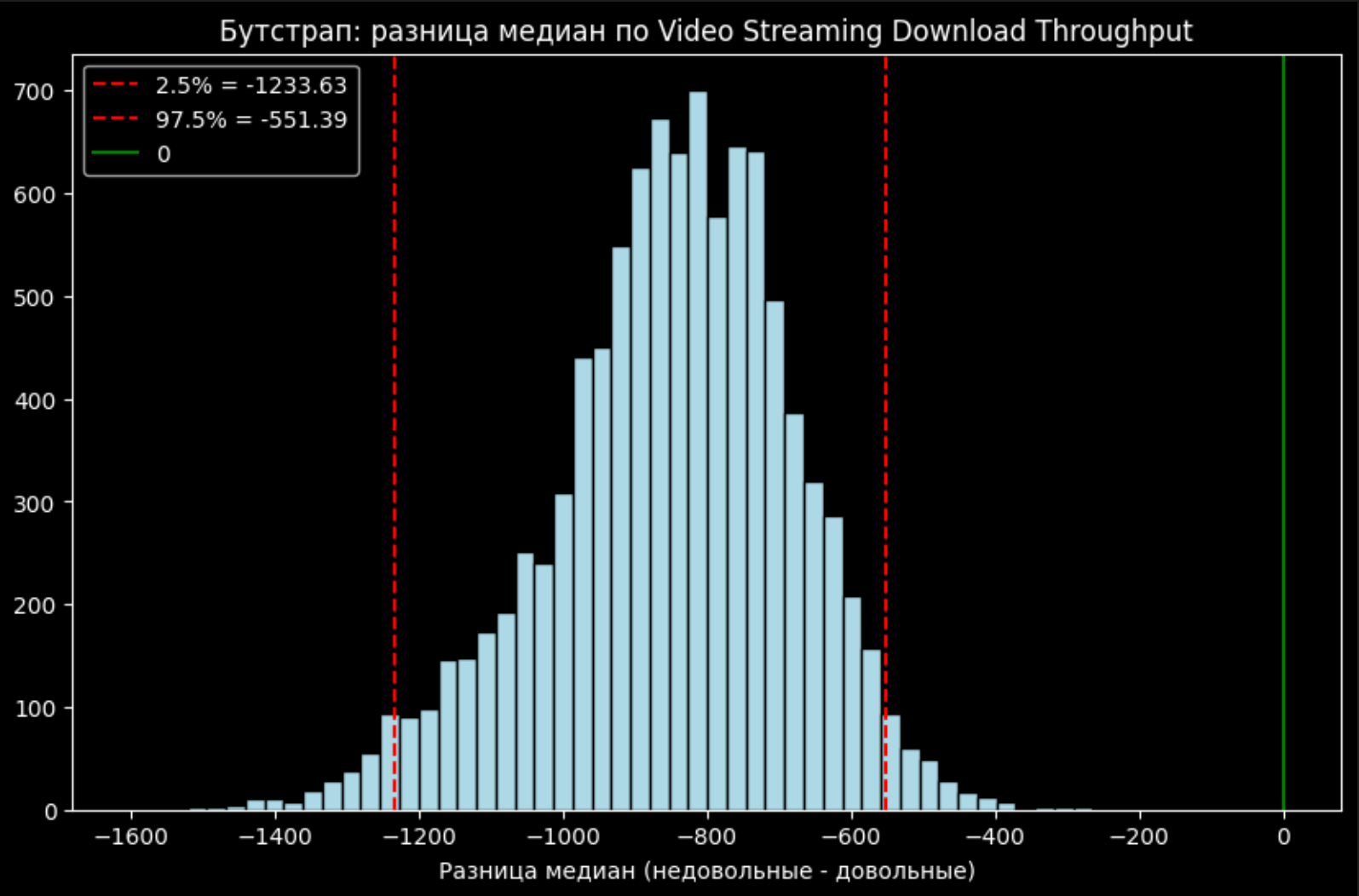
☒ Выбрана односторонняя проверка на основе предварительного анализа отрицательной корреляции

## Результаты теста

**U-статистика:** 938520.00

**p-value:** 0.0000

Статистически значимое различие подтверждено



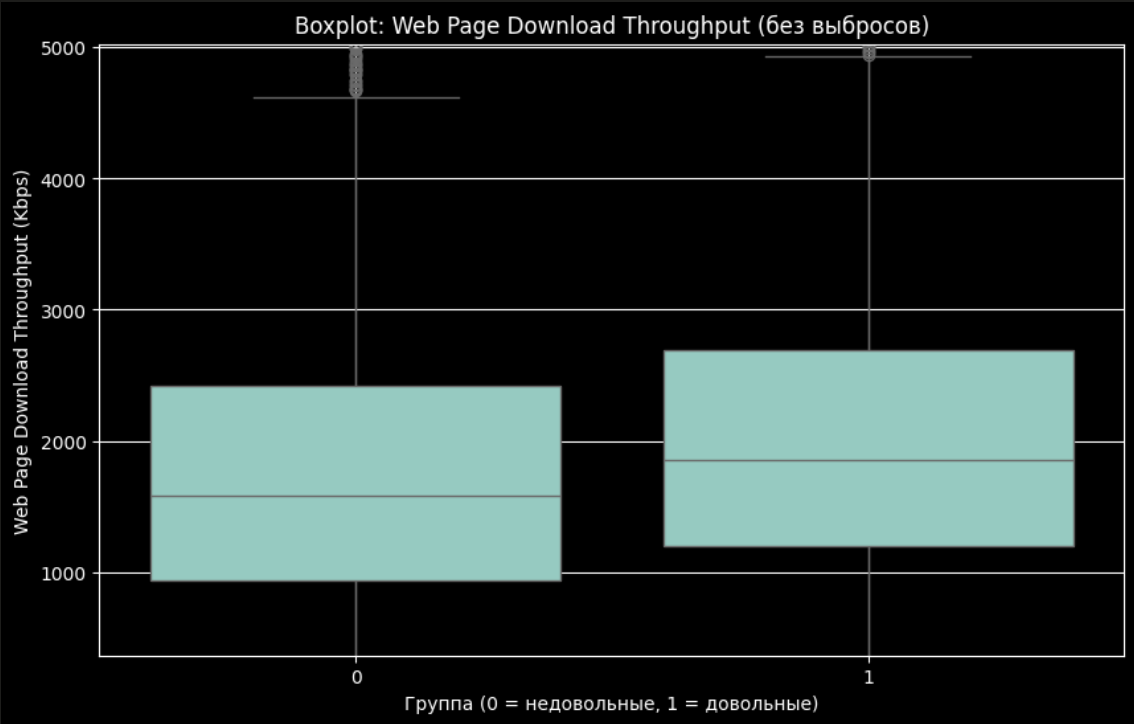
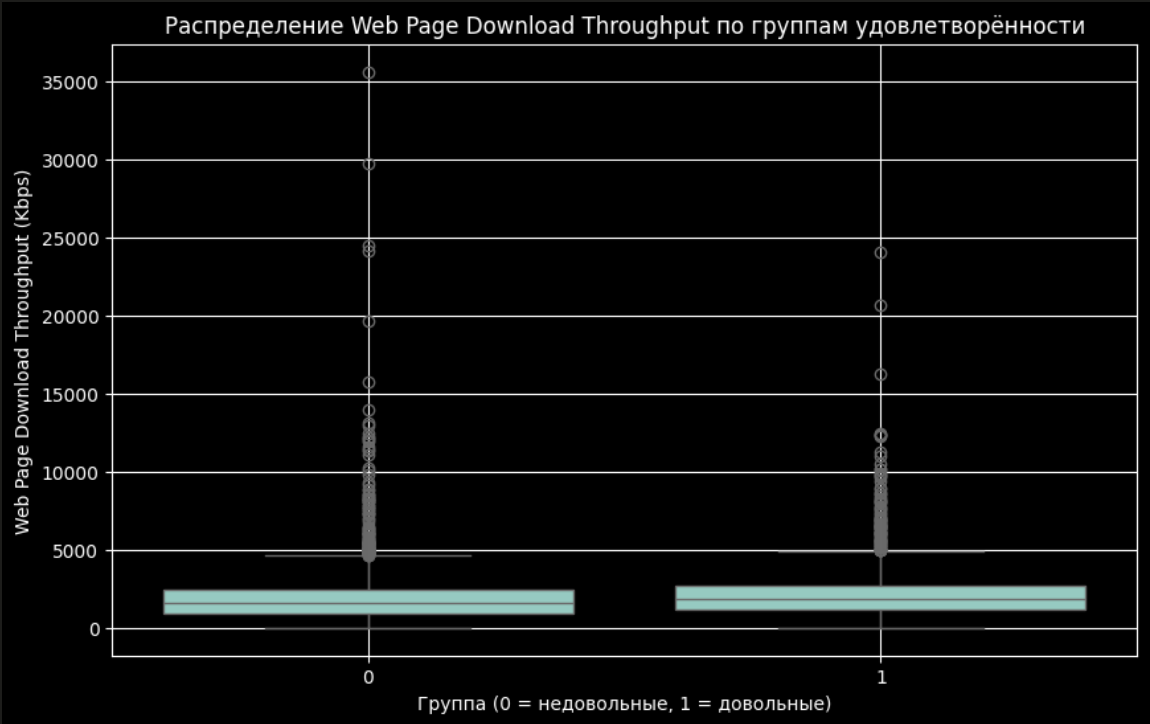
95% доверительный интервал разницы медиан: [-1233.63, -551.39] Kbps

## Вывод по второй гипотезе

Низкая пропускная способность потоковой загрузки видео негативно влияет на восприятие качества связи со стороны пользователей, и этот параметр должен быть приоритетной целью для технических улучшений.



### Третья метрика: Web Page Download Throughput (пропускная способность загрузки веб-страниц)



Видим ту же картину, что и с пропускной способностью загрузки видео. Недовольные пользователи имели более низкие значения метрики.

#### Различие между группами

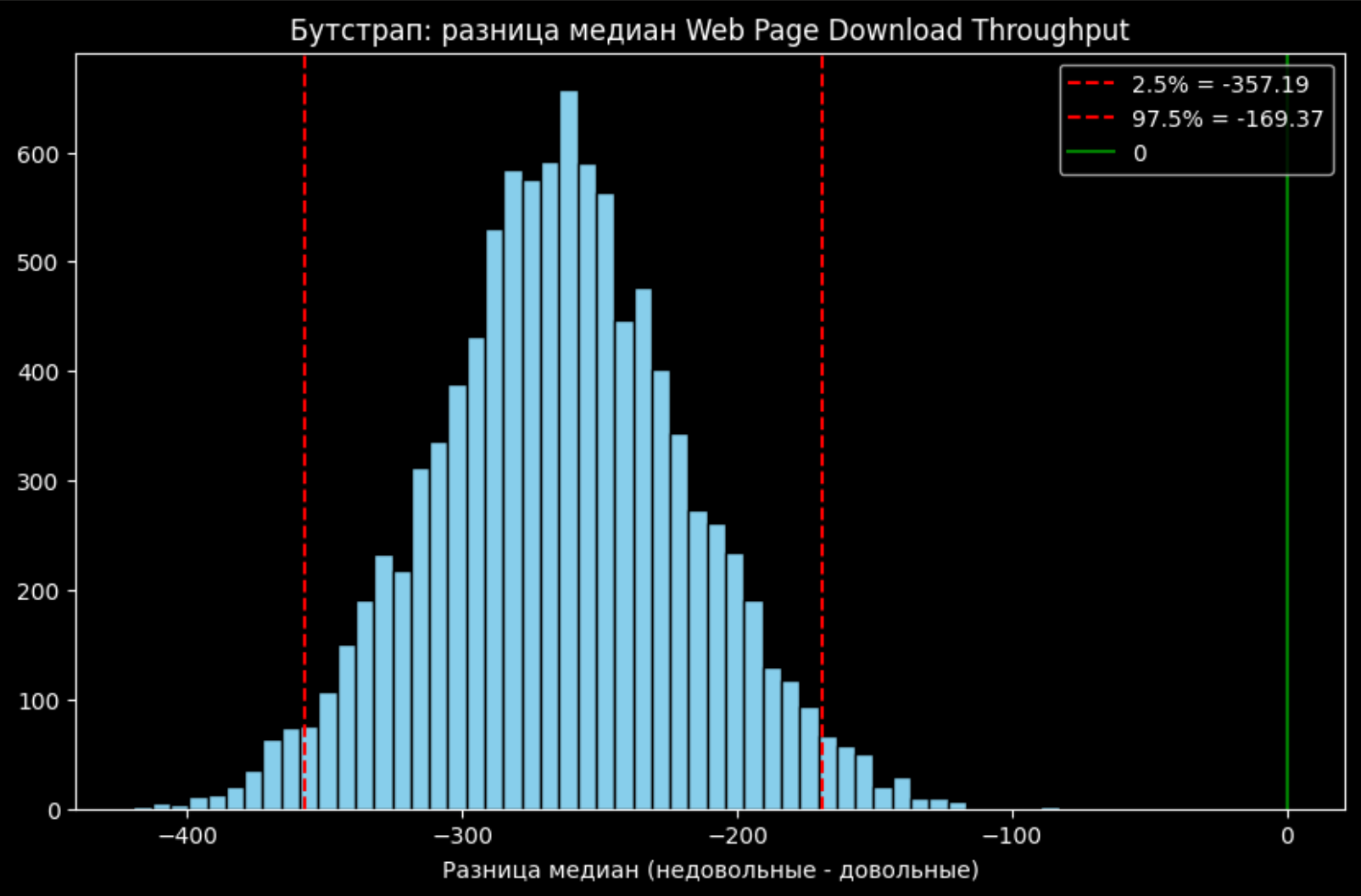
Средняя пропускная способность у недовольных  $\approx 1584$  kbps у довольных  $\approx 1849$  kbps. Медиана отличается на  $\approx 265$  kbps

## Формулировка гипотезы

**H<sub>0</sub>:** Распределения Web Page Download Throughput не отличаются

**H<sub>1</sub>:** У недовольных пользователей Web Page Download Throughput статистически значимо ниже

☐ Выбрана односторонняя проверка на основе предварительного анализа отрицательной корреляции



## Результаты теста

**U-статистика:** 926 878.00

**p-value:** 0.0000

Статистически значимое различие подтверждено

95% доверительный интервал разницы медиан: [-357.19, -169.37] Kbps

## Вывод по третьей гипотезе

У недовольных пользователей скорость загрузки веб-страниц статистически значимо ниже

Этот параметр также должен быть учтён в приоритизации технических улучшений

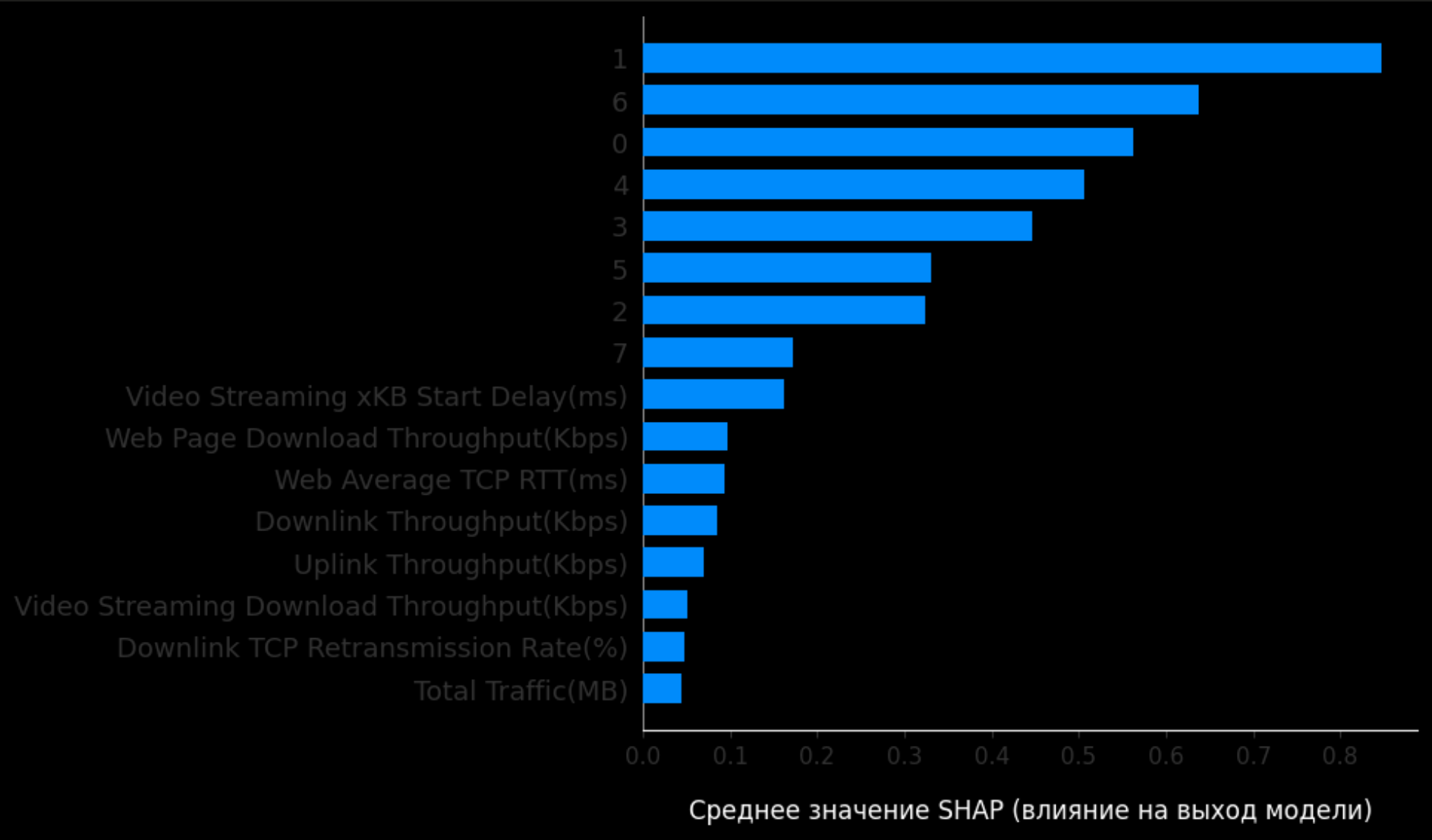
# SHAP-анализ для выявления важности признаков

Для дополнительной валидации результатов была обучена модель CatBoost с целью выявления наиболее важных факторов, влияющих на оценку Q1. SHAP-анализ позволил оценить вклад каждого признака в предсказание модели.

## Ключевые выводы





- Наиболее важными оказались бинарные жалобы из Q2
- Технические метрики с высокой важностью уже были проверены гипотезами
- Причины №1, №3 и №6 невозможно исследовать количественно
- Причина №0 представляет собой мусор в данных
- Причина №4 была детально исследована ранее

SHAP-анализ не выявил новых значимых признаков для исследования, но подтвердил важность ранее идентифицированных факторов.



# Заключение и рекомендации

В ходе комплексного анализа были изучены факторы, влияющие на пользовательскую оценку качества связи (Q1), с целью выявления приоритетных направлений для технических улучшений инфраструктуры «Мегафон».

	<p>Предобработка данных</p> <p>Пользователи разделены на группы: недовольные (<math>Q1 &lt; 9</math>) и довольные (<math>Q1 \geq 9</math>). Выделены бинарные признаки Q2 с причинами низкой оценки</p>
	<p>Разведочный анализ</p> <p>Выявлены значимые различия между группами по ключевым техническим метрикам, связанным с видео и веб-загрузкой</p>
	<p>Проверка гипотез</p> <p>Подтверждено влияние задержки старта видео, скорости загрузки видео и скорости загрузки веб-страниц на удовлетворённость</p>
	<p>Машинное обучение</p> <p>Модель CatBoost с точностью ~80% подтвердила значимость выявленных признаков через SHAP-анализ</p>

## Основные выводы

<p>Приоритет №1</p> <p>Оптимизация скорости загрузки видеоконтента и снижение задержки старта воспроизведения</p>	<p>Приоритет №2</p> <p>Улучшение скорости загрузки веб-страниц для повышения качества интернет-сервисов</p>	<p>Стратегия мониторинга</p> <p>Постоянный контроль выявленных метрик как ключевых индикаторов удовлетворённости клиентов</p>
---	---	---