

АНАЛИЗ ДАННЫХ ИГРЫ 3 В РЯД

Дипломный проект



Описание данных, целей и задач

Исходные данные имеем за 14 дней, в которых есть набор событий и параметров к ним.

Задачи и цели проекта:

- Разбить имеющийся датасет по таблицам, рассчитать недостающие параметры.
- Произвести разведочный анализ, создать систему метрик описывающих данные о первой сессии и доходе пользователей и сделать выводы.
- Провести а/в тест и сделать вывод.

Методология

Анализ данных производим в питоне, используя библиотеку pandas, numpy и stats для анализа, а также библиотеки для визуализации.

Анализ метрик проводим базируясь на юнит-экономике и статистике.

A/B тест проводим опираясь на t-тест и p-value.

Построение системы метрик первой сессии, удержания, аудитории проекта

Создаем таблицу first_session

	user_id	session_id	platform_type	type_traffic	time_start_session	time_finish_session	length_session	amount_get_level	amount_get_gold	amount_get_bronze_cup	amount_get_silver_cup	amount_get_gold_cup	level_cnt_tut	error	error_time	error_name	fps
0	1000001	a1000001ES1	android	organic	2023-02-01 08:01:42	2023-02-01 08:02:05	23.0	1	0.0	0	0	0	5	NaN	NaT	NaN	27.0
1	1000004	a1000004RU1	android	organic	2023-02-01 08:30:42	2023-02-01 08:30:42	0.0	1	0.0	0	0	0	1	NaN	NaT	NaN	49.0
2	1000007	a1000007GE1	android	organic	2023-02-01 08:51:00	2023-02-01 08:51:12	12.0	1	0.0	0	0	0	2	error_window	2023-02-01 08:51:12	network_error	24.0
3	1000009	a1000009LT1	android	paid	2023-02-01 09:10:25	2023-02-01 09:10:36	11.0	1	0.0	0	0	0	3	NaN	NaT	NaN	33.0
4	1000011	a1000011CA1	android	organic	2023-02-01 09:22:37	2023-02-01 09:22:57	20.0	1	0.0	0	0	0	3	error_window	2023-02-01 09:22:57	klient_error	36.0
...
1561	1002464	i1002464GE1	iOS	organic	2023-02-15 12:42:56	2023-02-15 12:43:03	7.0	1	0.0	0	0	0	3	NaN	NaT	NaN	53.0
1562	1002465	i1002465RU1	iOS	organic	2023-02-15 12:56:48	2023-02-15 13:00:30	222.0	5	2860.0	0	0	0	7	NaN	NaT	NaN	45.0
1563	1002466	i1002466DE1	iOS	organic	2023-02-15 12:59:09	2023-02-15 12:59:18	9.0	1	0.0	0	0	0	3	NaN	NaT	NaN	37.0
1564	1002479	i1002479CA1	iOS	organic	2023-02-15 14:47:35	2023-02-15 14:47:45	10.0	1	0.0	0	0	0	2	error_window	2023-02-15 14:47:45	network_error	24.0
1565	1002499	i1002499GE1	iOS	organic	2023-02-15 17:34:28	2023-02-15 17:35:11	43.0	1	0.0	0	0	0	7	NaN	NaT	NaN	31.0

1382 rows × 17 columns

Построение системы метрик первой сессии, удержания, аудитории проекта

Количество новых пользователей за 14 дней: **1339**

DAU 1 дня: **68**

DAU 2 дня: **86**

DAU 7 дня: **97**

WAU с 2023-02-01 по 2023-02-09: **803**

WAU с 2023-02-10 по 2023-02-16: **536**

Средняя продолжительность первой сессии: **97.60 сек**

Средняя продолжительность первой сессии (без учета 0): **108.54 сек**

Общее количество уровней полученных в первой сессии: **3101**

Среднее количество уровней полученных в первой сессии: **2**

Общее количество золота полученного в первой сессии: **1423677**

Среднее количество золота полученного в первой сессии: **1030.16**

Общее количество бронзовых кубков полученных в первой сессии: **117**

Общее количество серебряных кубков полученных в первой сессии: **0.00**

Retention Rate на 1 день (с 2023-02-01 на 2023-02-02): **0.00%**

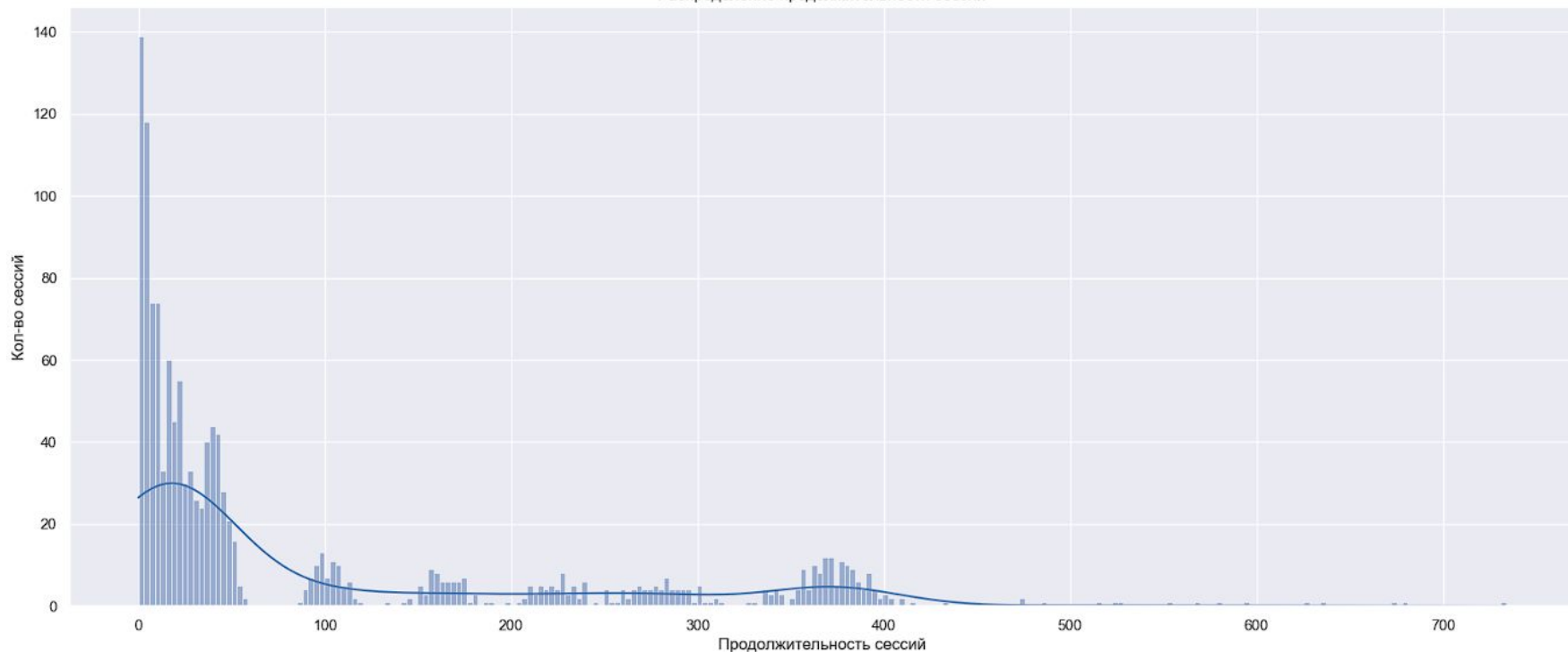
Retention Rate на 7 день (с 2023-02-01 на 2023-02-08): **0.00%**

Retention 2 сессии: **10.75%**

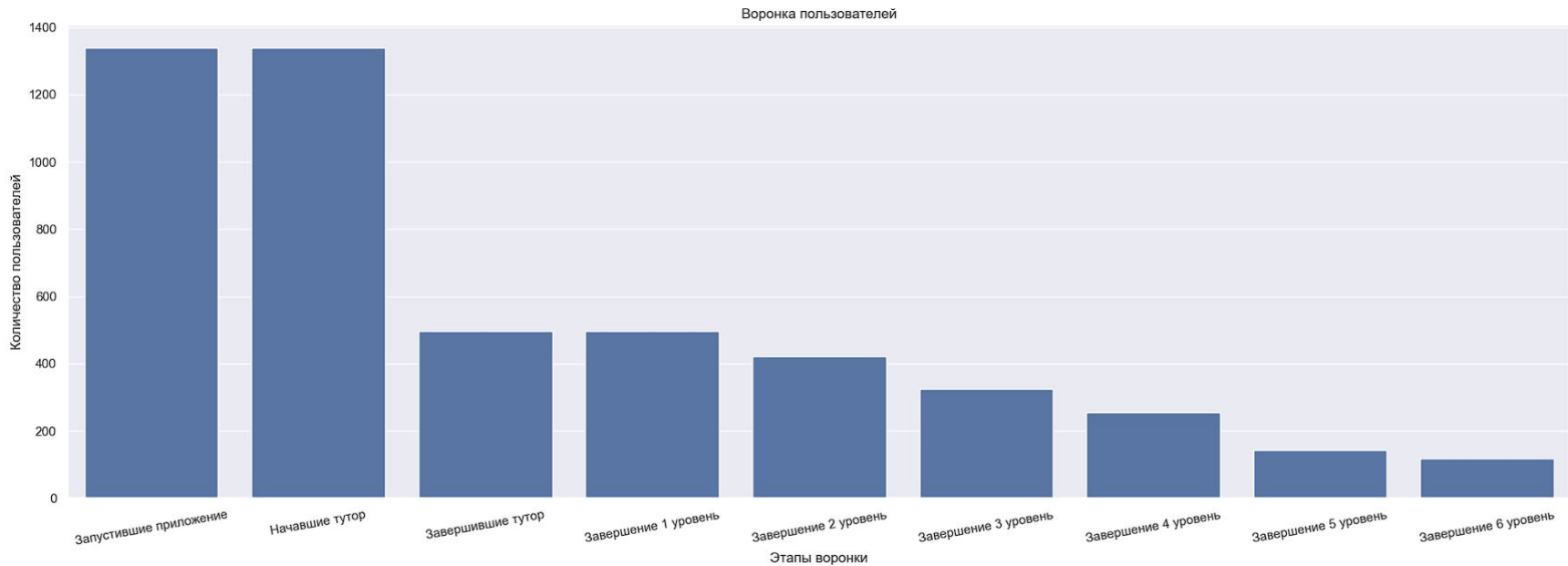
Retention 3 сессии: **2.99%**

Построение системы метрик первой сессии, удержания, аудитории проекта

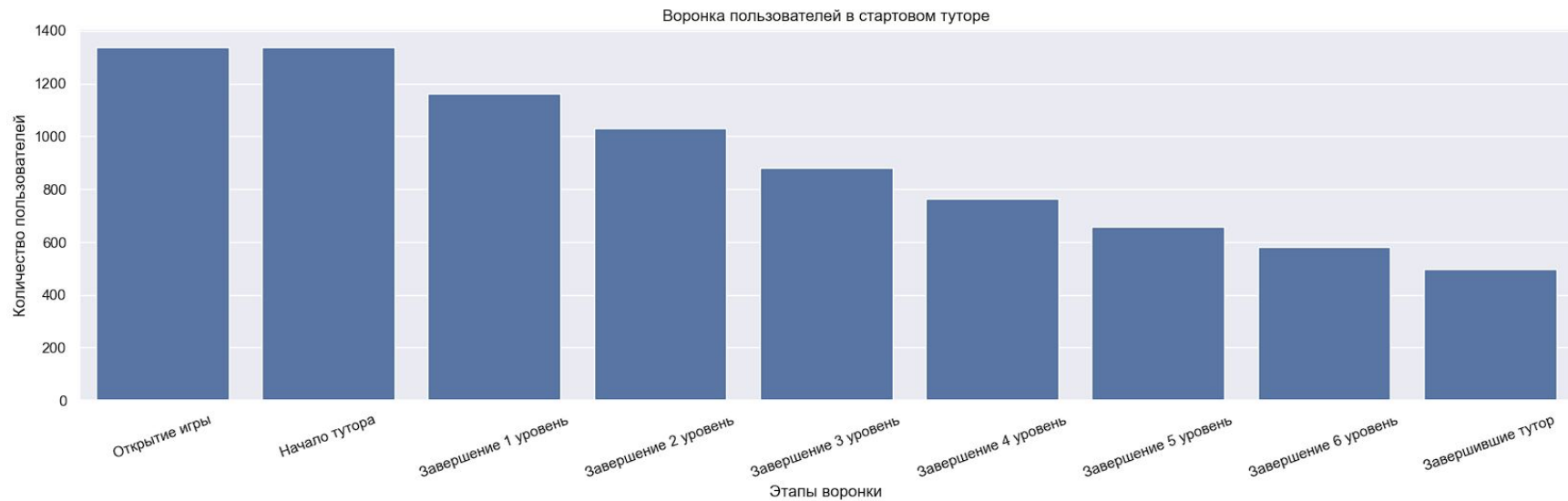
Распределение продолжительности сессии



Воронка первой сессии

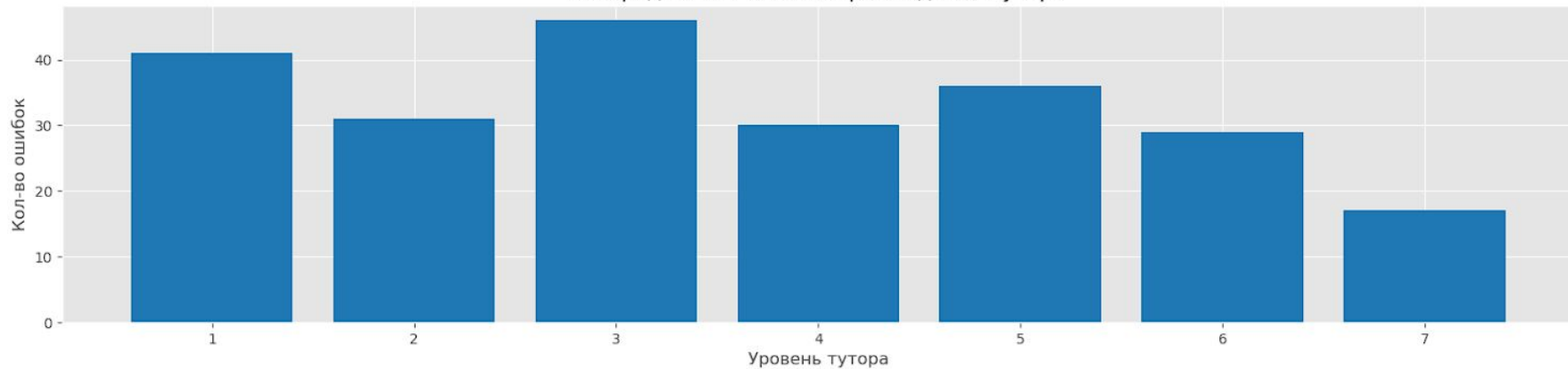


Воронка стартового тьютора



Ошибки у пользователей по уровням тьютора

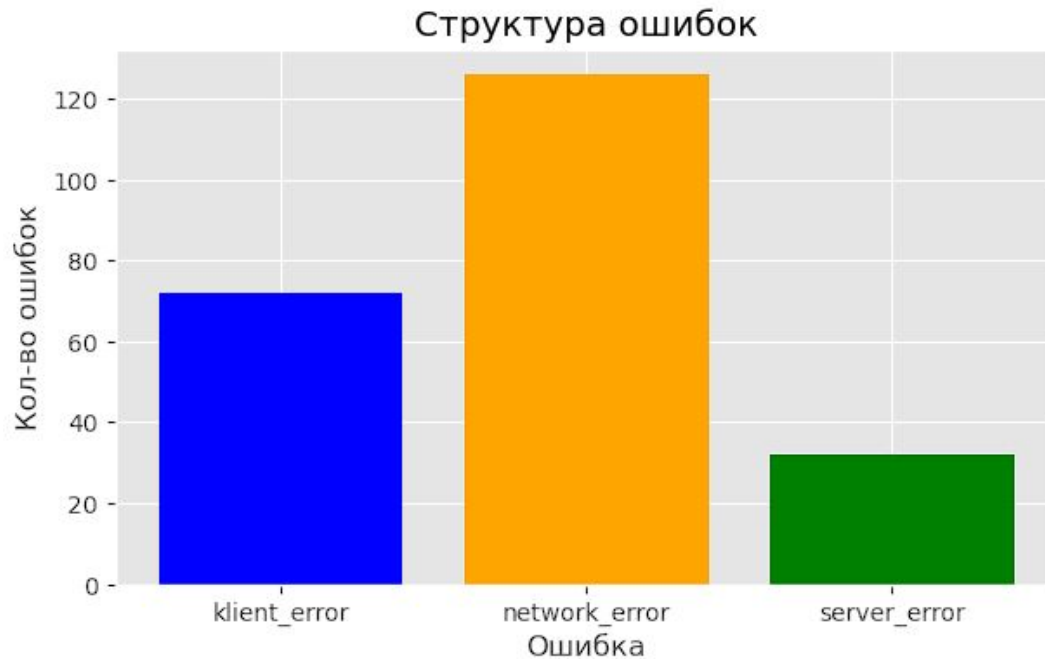
Распределение ошибок прохождения тьютора



Ошибки у пользователей

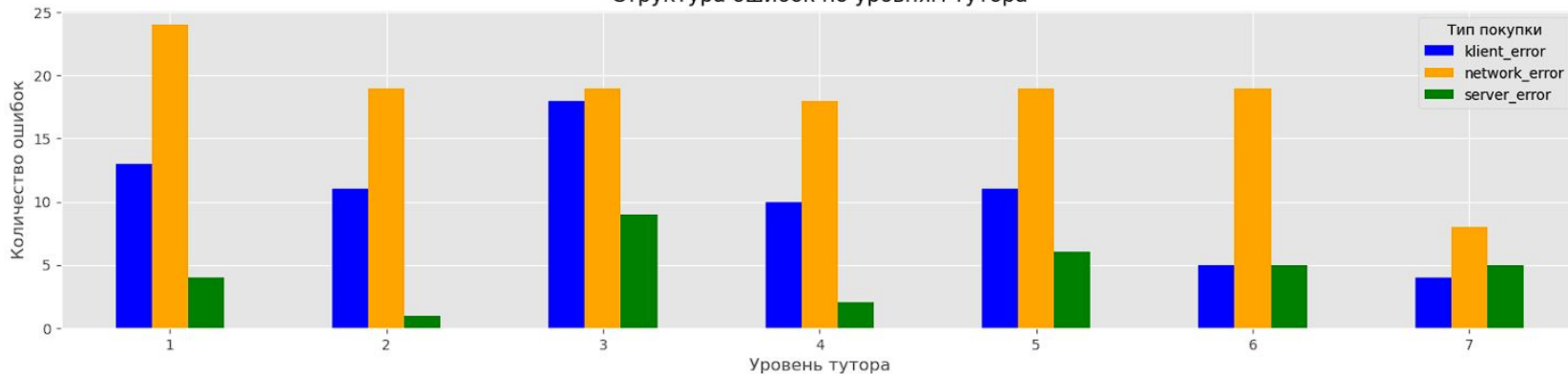
Количество вернувшихся пользователей после получения любой ошибки: **0**

Общее кол-во ошибок: **230**

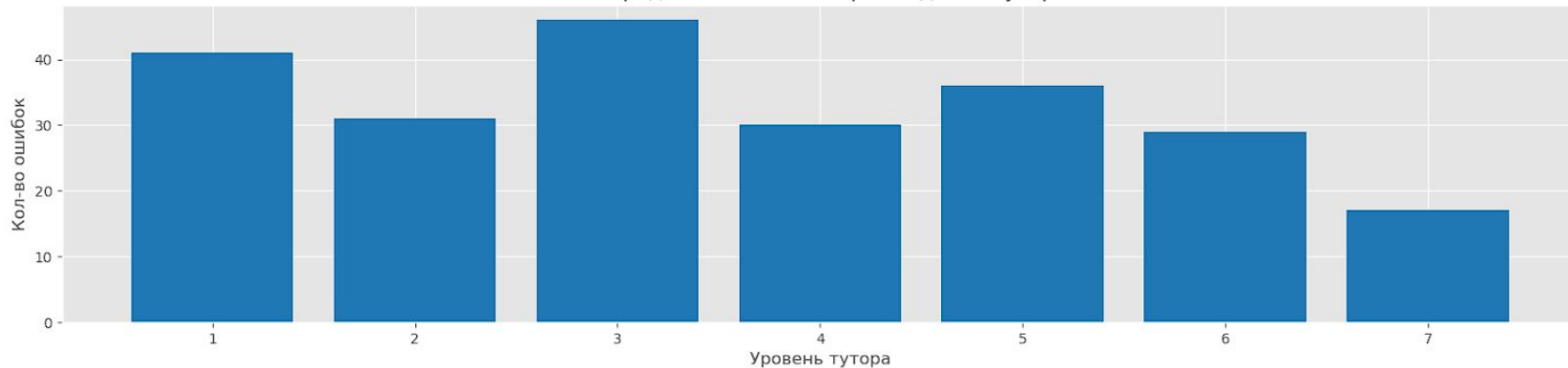


Ошибки у пользователей

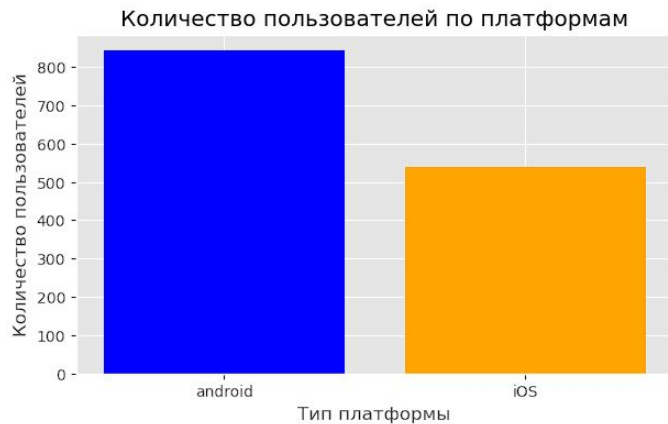
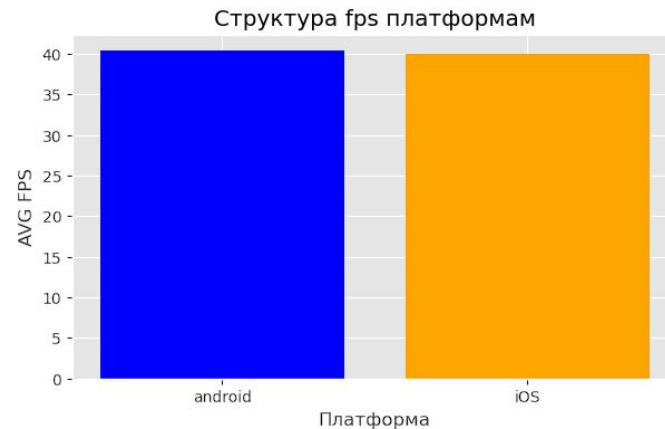
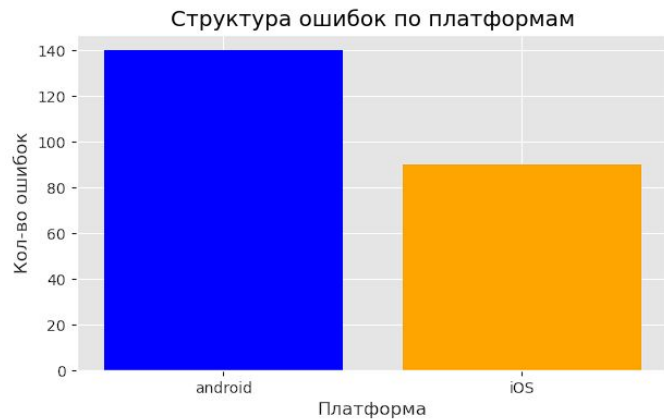
Структура ошибок по уровням тьютора



Распределение ошибок прохождения тьютора

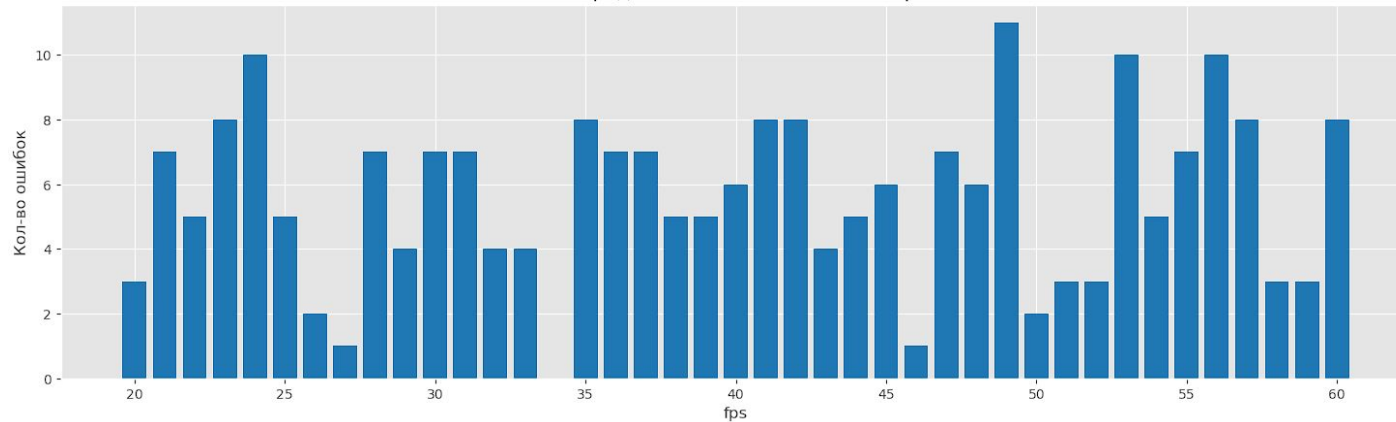


Ошибки у пользователей

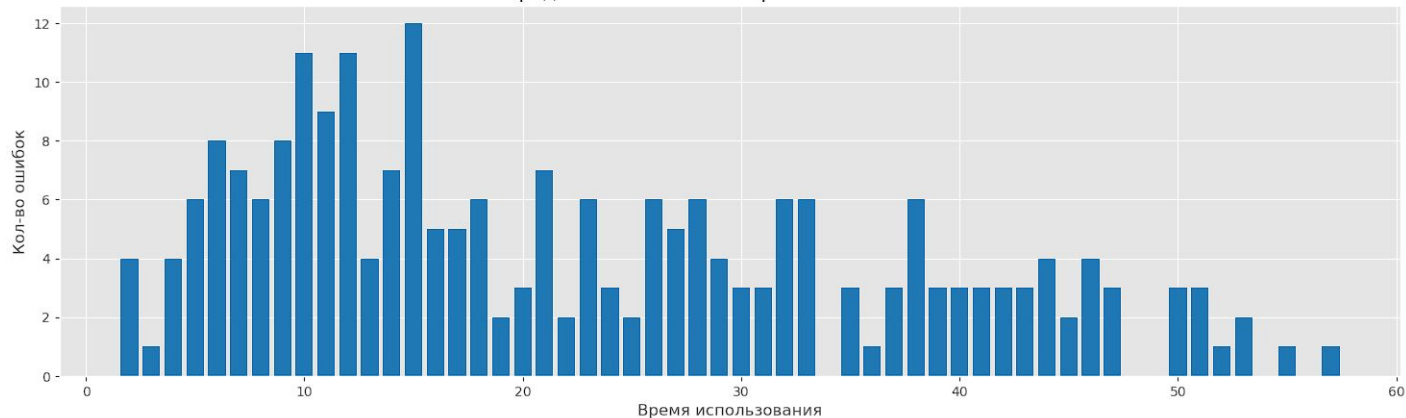


Ошибки у пользователей по fps и времени

Распределение кол-ва ошибок по fps

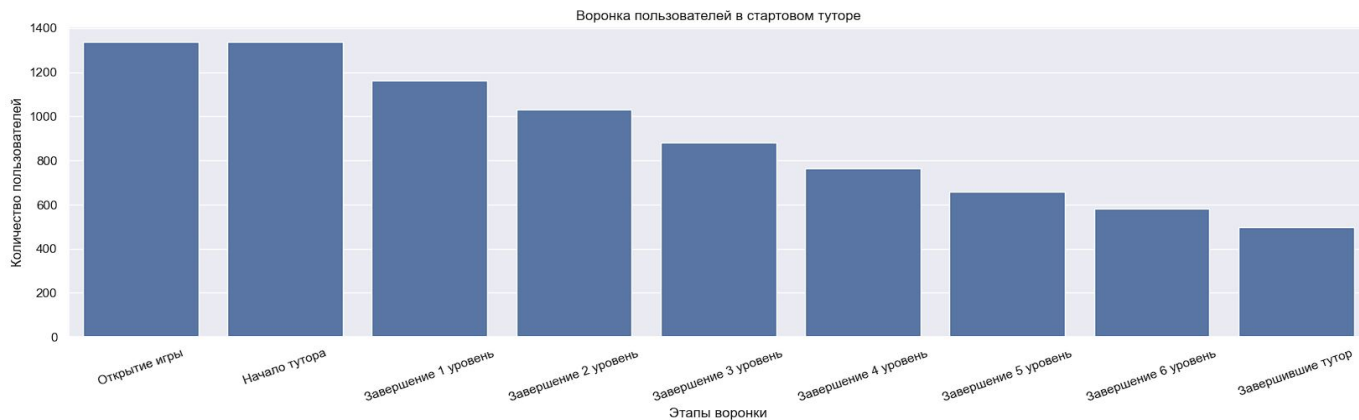
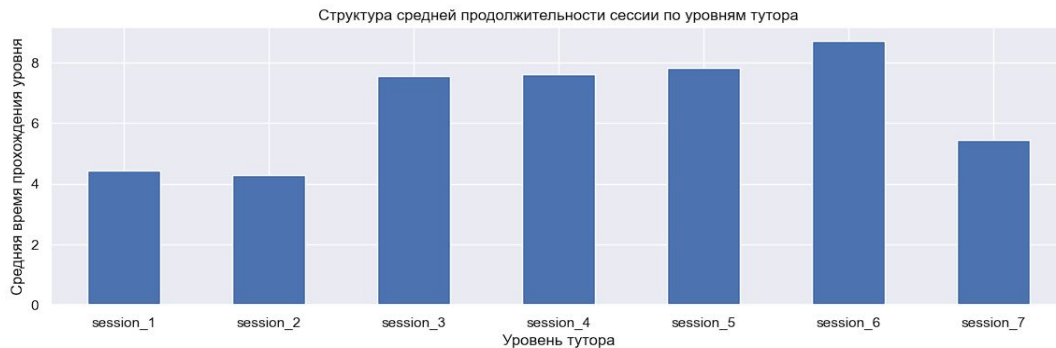


Распределение ошибок по времени пользования

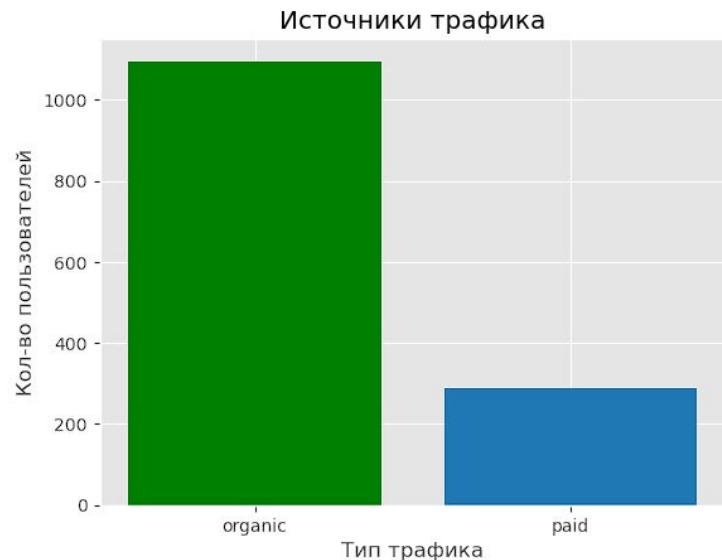


Структура средней продолжительности прохождения

Среднее время прохождения
тутора: **45.55 сек**



Построение системы метрик первой сессии, удержания, аудитории проекта



Выводы по первой сессии

Замеченные аномалии:

1. Низкий retention первого дня
2. Низкая продолжительность 1 сессии

Данные проблемы могут быть связаны с прохождением тьютора у части пользователей возникают ошибки различных видов, которые препятствуют прохождению, а также отвал пользователей с тьютора возможен из-за его сложности или проблем взаимодействия с интерфейсом.

Анализ дохода с пользователей, конверсии и структуры покупок

Общее кол-во пользователей: **1382**

Общее кол-во платящих пользователей: **36**

Общее кол-во покупок: **36**

Общий доход: **105.64**

Средний доход на пользователя (ARPU): **0.08**

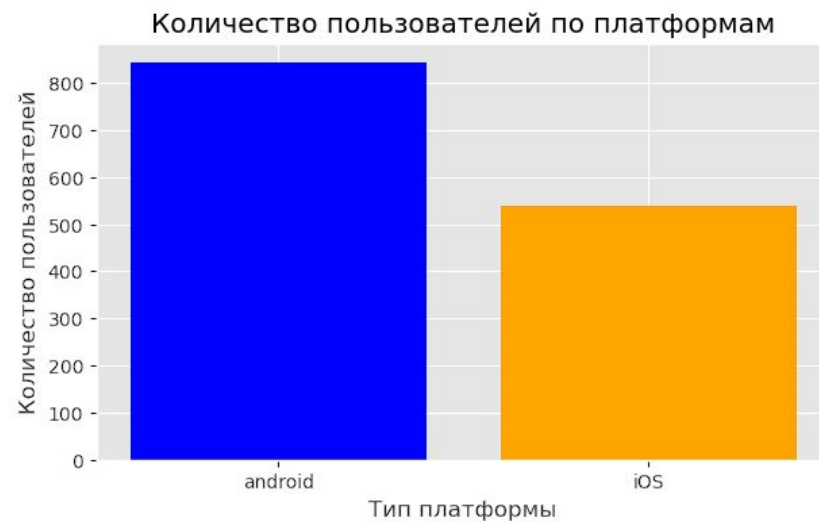
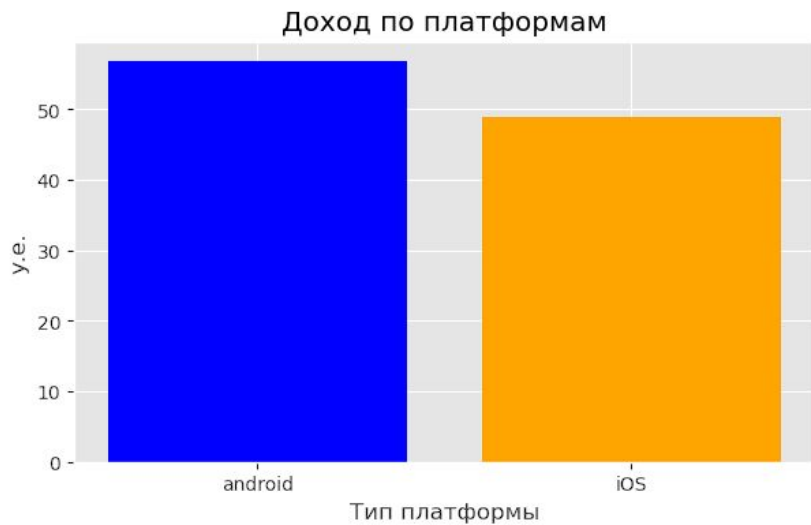
Средний доход на платящего пользователя (ARPPU): **2.93**

Конверсия: **2.6%**

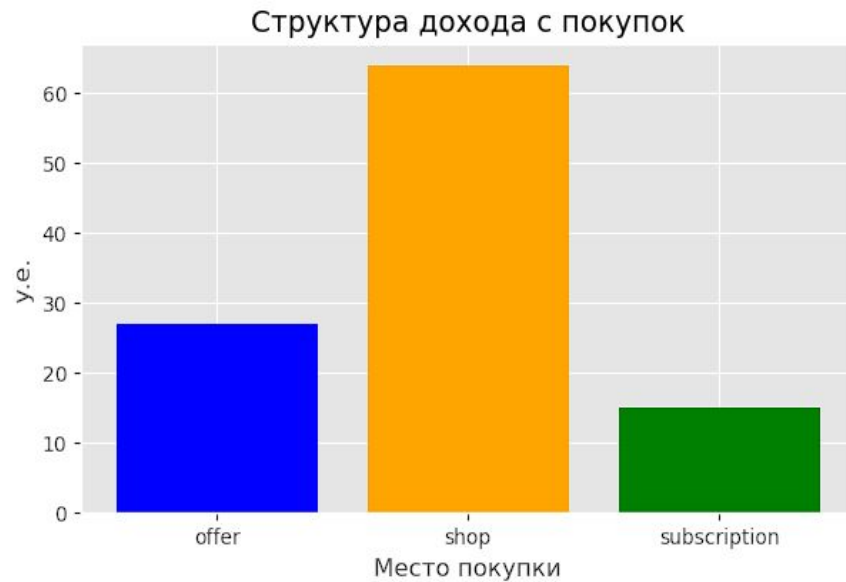
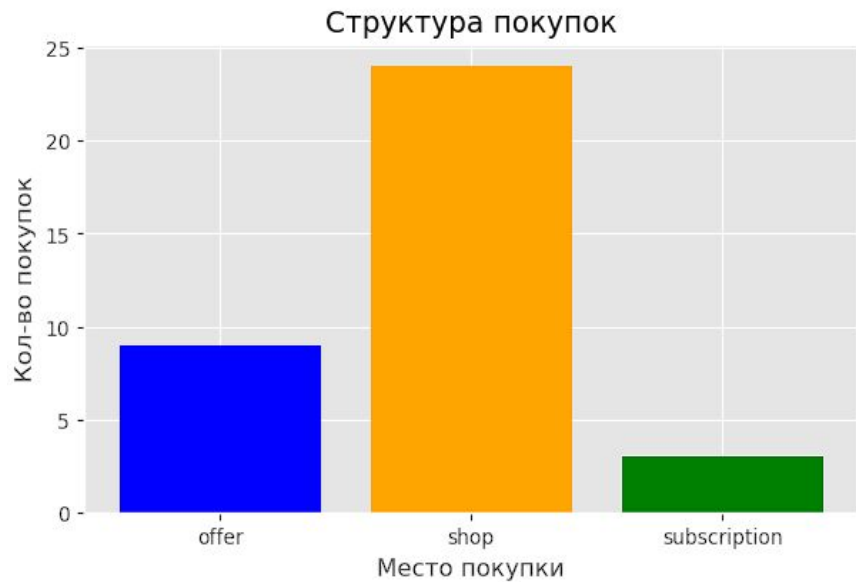
Среднее кол-во покупок на пользователя: **0.026**

Доля платящих пользователей: **0.03**

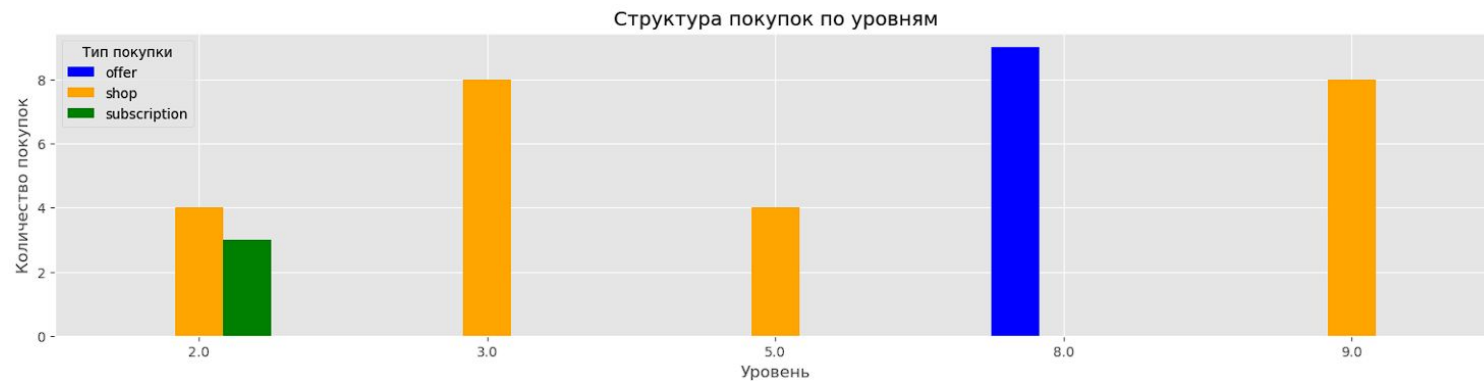
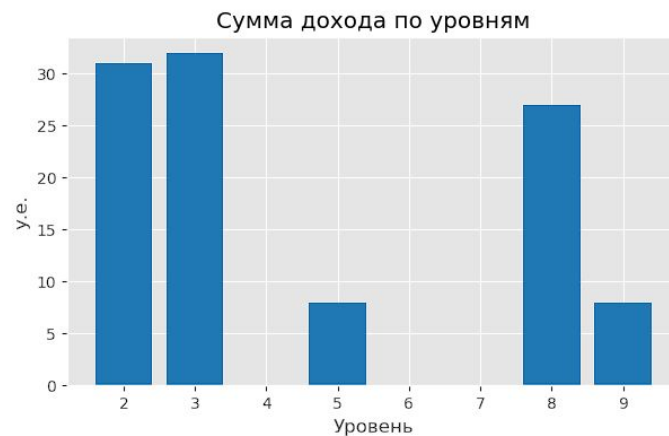
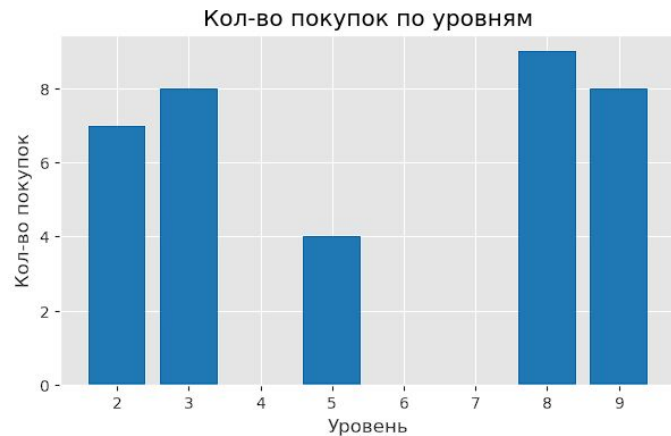
Структура покупок по платформе



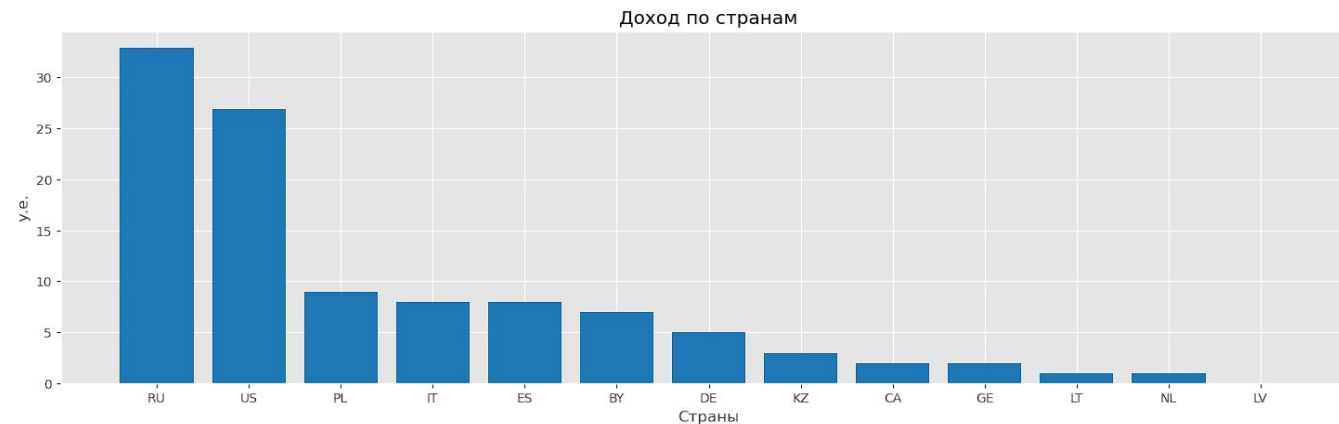
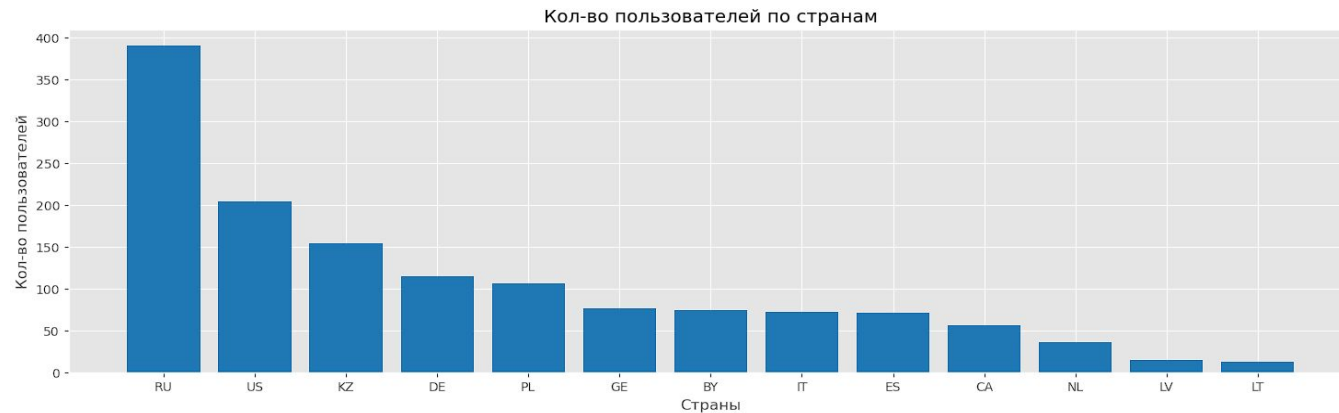
Структура видов покупок



Структура покупок по уровням



Структура покупок по странам



A/B тест

На основе данных A/B-тестирования необходимо проанализировать итоги эксперимента и сделать вывод стоит ли заменить стартовый тутор новым.

Тест проходил с 01.02.2023 по 07.02.2023

Группа А юзера чьи user_id – четные

Группа В юзера чьи user_id – нечетные

При уровне значимости 0.05

A/B тест проводился по 2 метрикам: средняя продолжительность сессии и среднее кол-во полученных уровней.

Результаты А/В тестирования

По метрике средняя продолжительность сессии:

- Средняя продолжительность сессий группы А: **132.94 сек**
- Средняя продолжительность сессий группы В: **119.45 сек**
- t-статистика: **1.02**, p-значение: **0.3099**

Статистическая разница между группами **отсутствует**.

По метрике среднее количество полученных уровней:

- Среднее количество полученных уровней группы А: **2.60**
- Среднее количество полученных уровней группы В: **2.44**
- t-статистика: **0.98**, p-значение: **0.3294**

Статистическая разница между группами **отсутствует**.

Вывод: замена стартового тьютора **не требуется**