



Data Koalas 

1. Динмухаметов Данис
2. Иванов Герман
3. Кром Илья
4. Михайлов Андрей
5. Чугунов Арсений

Битва поколений: опыт VS молодость



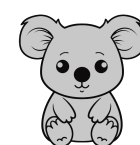
Структура данных

Информация

**Статистика АТР по теннисным матчам с
информацией об игроках**

25202 различных матча, 49 столбцов

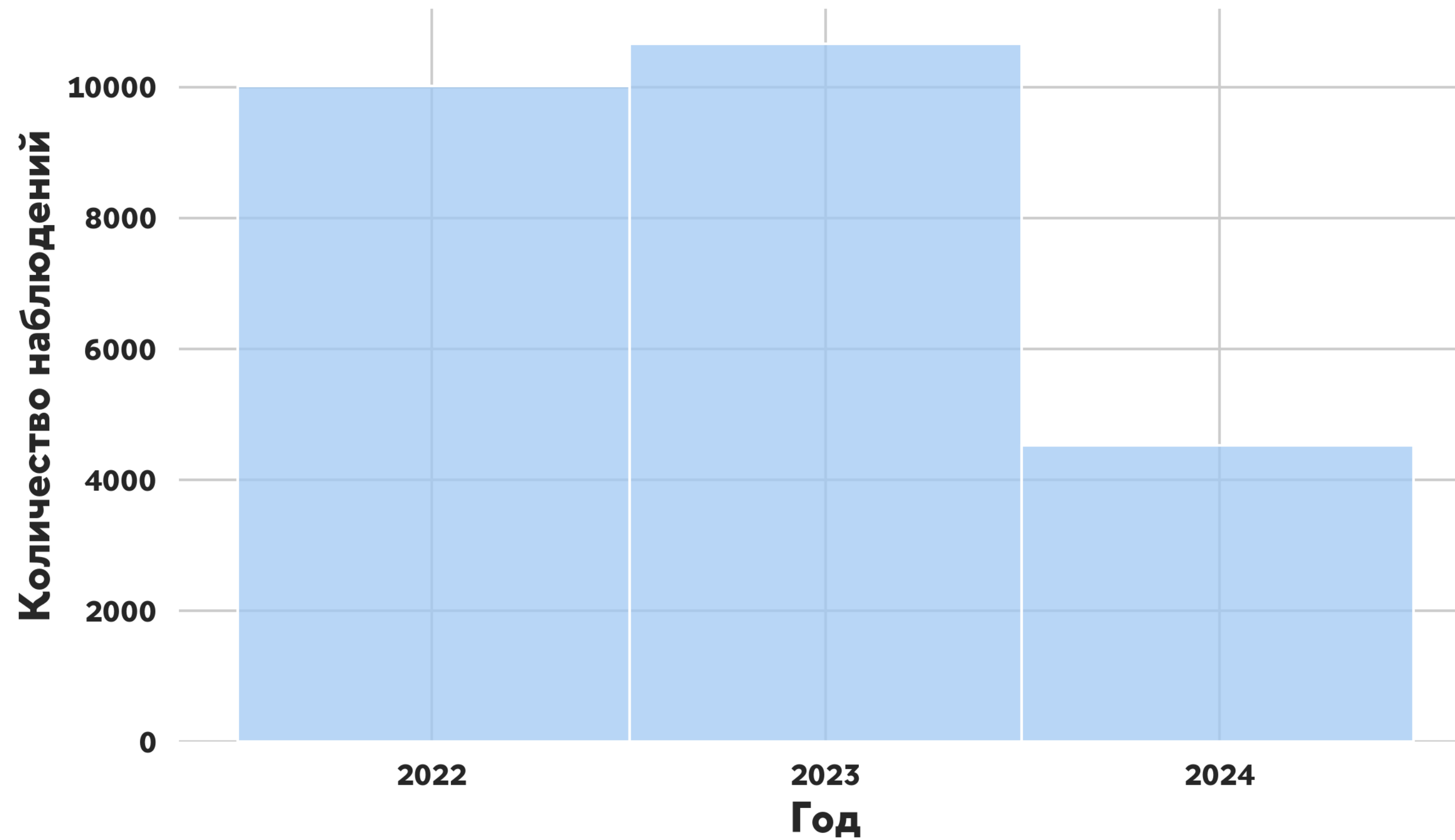
Данные за 2022-2024 годы





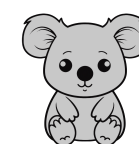
Предварительный анализ

Распределение годов



Мода:

2023 год





Обработка данных

Переход от матчей к уникальным пользователям

Наблюдение:

матч

25202



пользователь

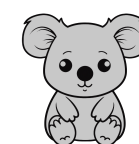
1855

Добавление доли побед от всех матчей по годам

+ r2022, r2023, r2024

Выделение полож. и отриц. динамики

+ better, worse





Обработка данных



Положительная динамика:

$$\text{ratio}_{2022} < \text{ratio}_{2023} < \text{ratio}_{2024}$$

— 88¹ человек



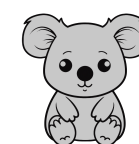
Отрицательная динамика:

$$\text{ratio}_{2022} > \text{ratio}_{2023} > \text{ratio}_{2024}$$

— 107 человек

$$\text{Ratio}_{\text{год}} = \frac{\text{количество побед в году}}{\text{количество матчей в году}}$$

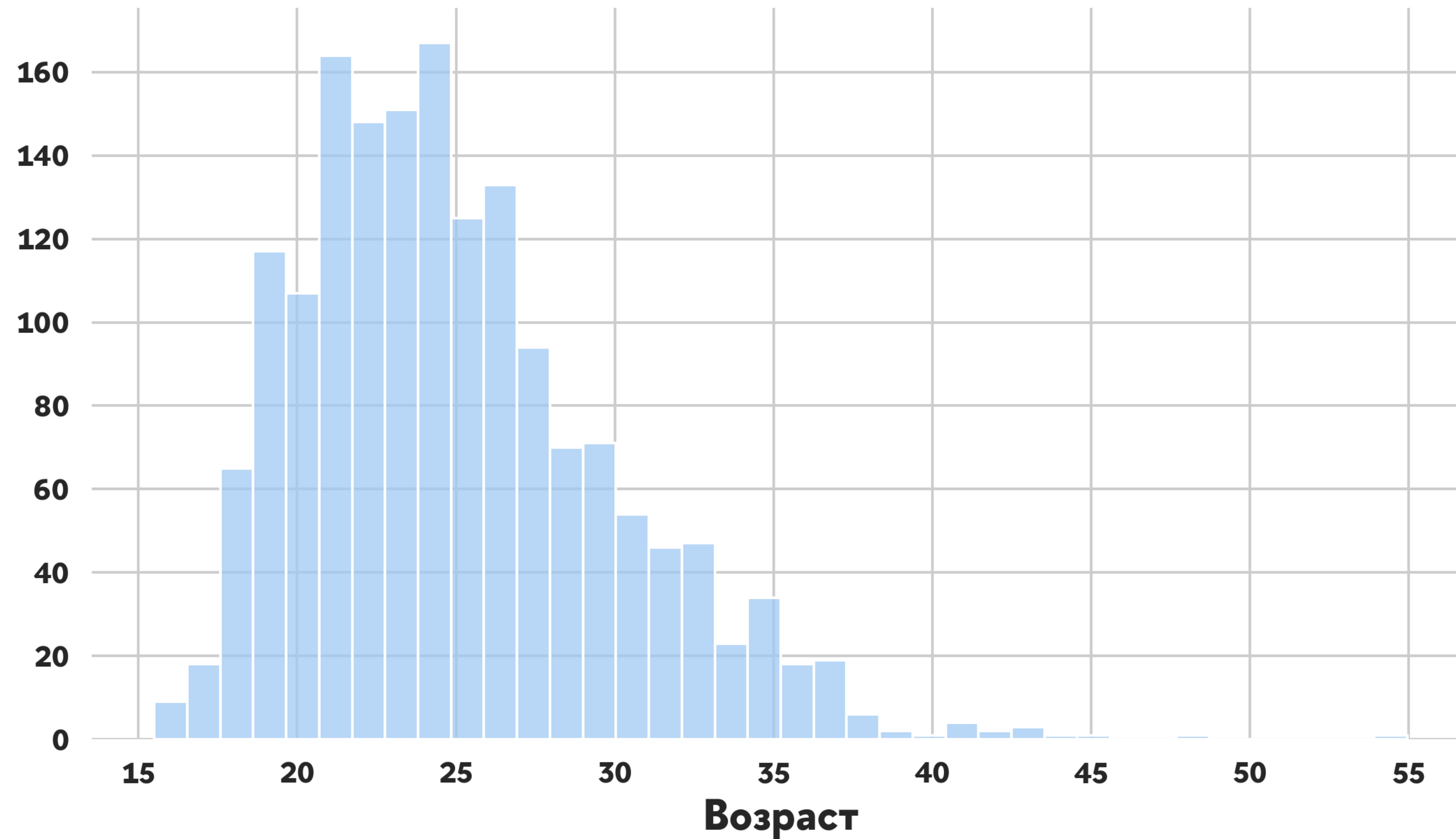
¹Исключили 3 человека, у которых сильный разброс в количестве матчей по годам.





Предварительный анализ

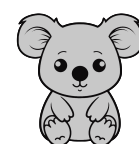
Распределение возраста игроков



Среднее: 25.0

Медиана: 24.4

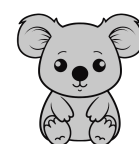
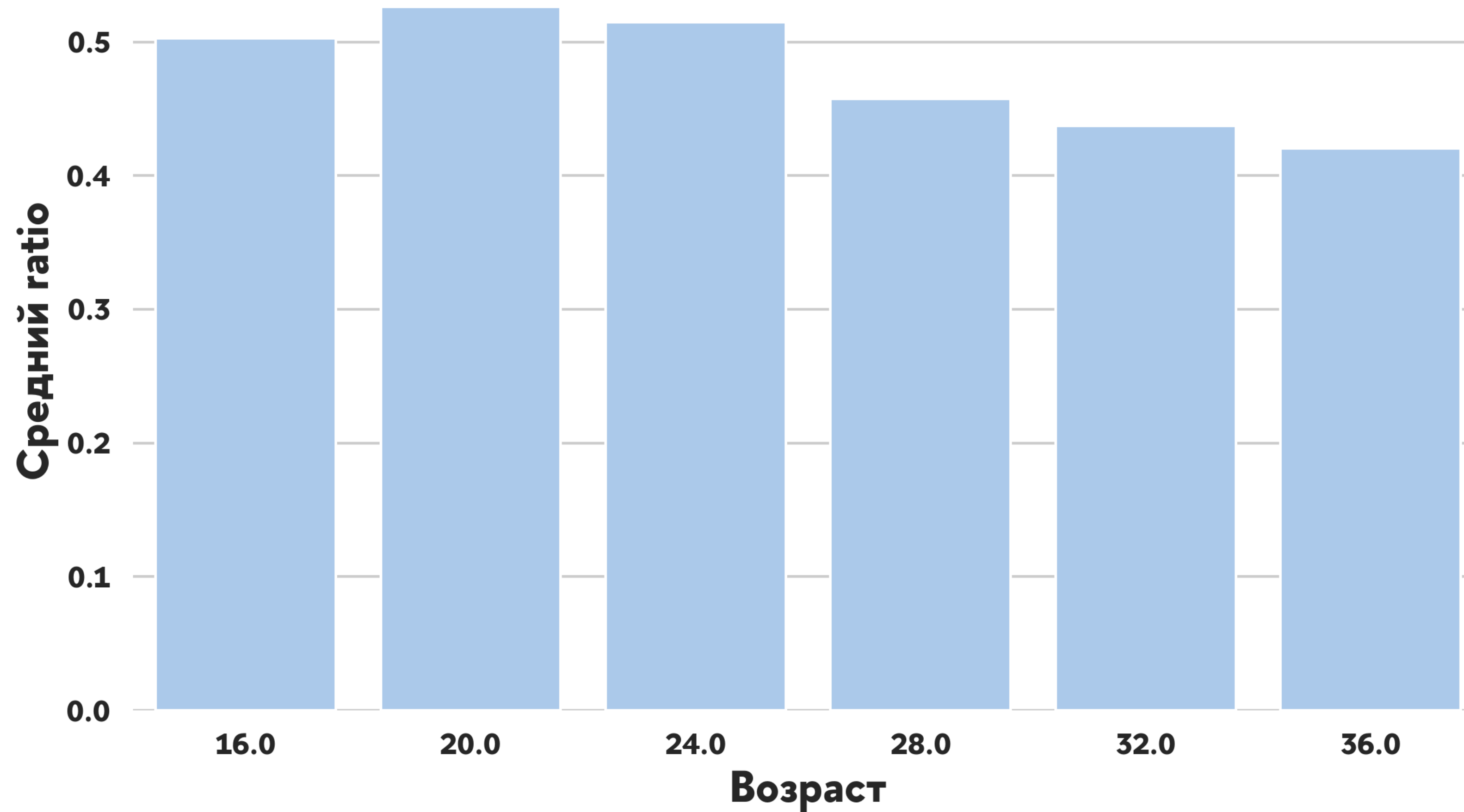
Стандартное
отклонение: 4.9





Предварительный анализ

Распределение доли побед в зависимости от возраста





Очистка датасета

Удаление выбросов

По возрасту

Чистим по IQR

1855 строк → 1680 строк

Удаление NaN

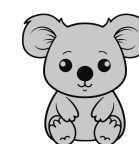
Отбрасываем всех людей, имеющих
NaN в доле побед хотя бы в один год

1680 строка → 627 строк

Выделение динамики

Отбрасываем людей, не
обладающих динамикой в
течение всех трёх лет

627 строк → 195 строк





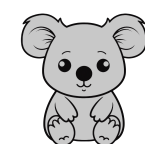
Исследовательский вопрос и гипотеза

Исследовательский вопрос

Что влияет на **динамику процента побед** игрока в теннисе?

Гипотеза

Среди **молодых** игроков **чаще** наблюдается тенденция к росту доли побед, чем у более **старших**





Механизмы

**Первый
механизм**

Старшие игроки дольше играют в профессиональных матчах

Больше травм

Тяжелее и дольше восстановление

Чем старше становятся, тем реже выигрывают в матчах

Уже побеждали на соревнованиях

Могут переоценивать себя

**Второй
механизм**

У более возрастных игроков реже наблюдается
тенденция роста, чем у молодых





Альтернативные механизмы

Старшие игроки уже побеждали на соревнованиях



Уменьшается мотивация



Начинают менее усердно тренироваться

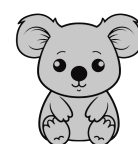


Становятся менее конкурентно-способными по сравнению с более молодыми игроками

У старших игроков более развитая личная жизнь



Меньше свободного времени





Математическая модель

Итоговый объем выборки

198

Уровень стат. значимости

5%

H0

Тенденции роста побед у игроков более молодого и старшего
возраста **равны**

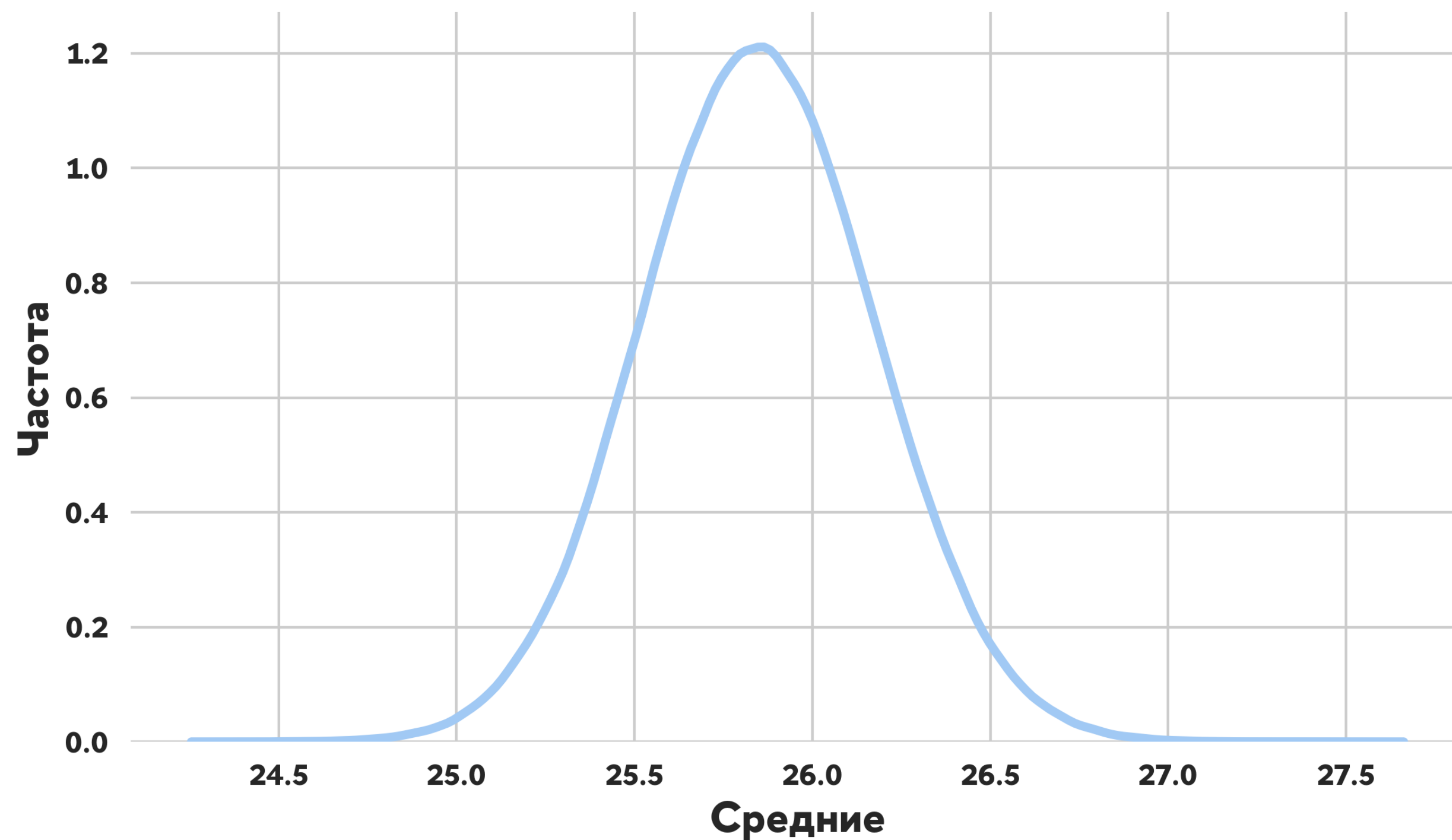
H1

Тенденции роста побед у игроков более молодого и старшего
возраста **не равны**



Проверка нормальности выборочных средних

Распределение выборочных средних возраста



Результат

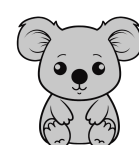
Показатель	Значение	Распред.
p-value	0.51	Нормальное

Вывод

Тест Шапиро-Уилка на нормальное распределение пройден



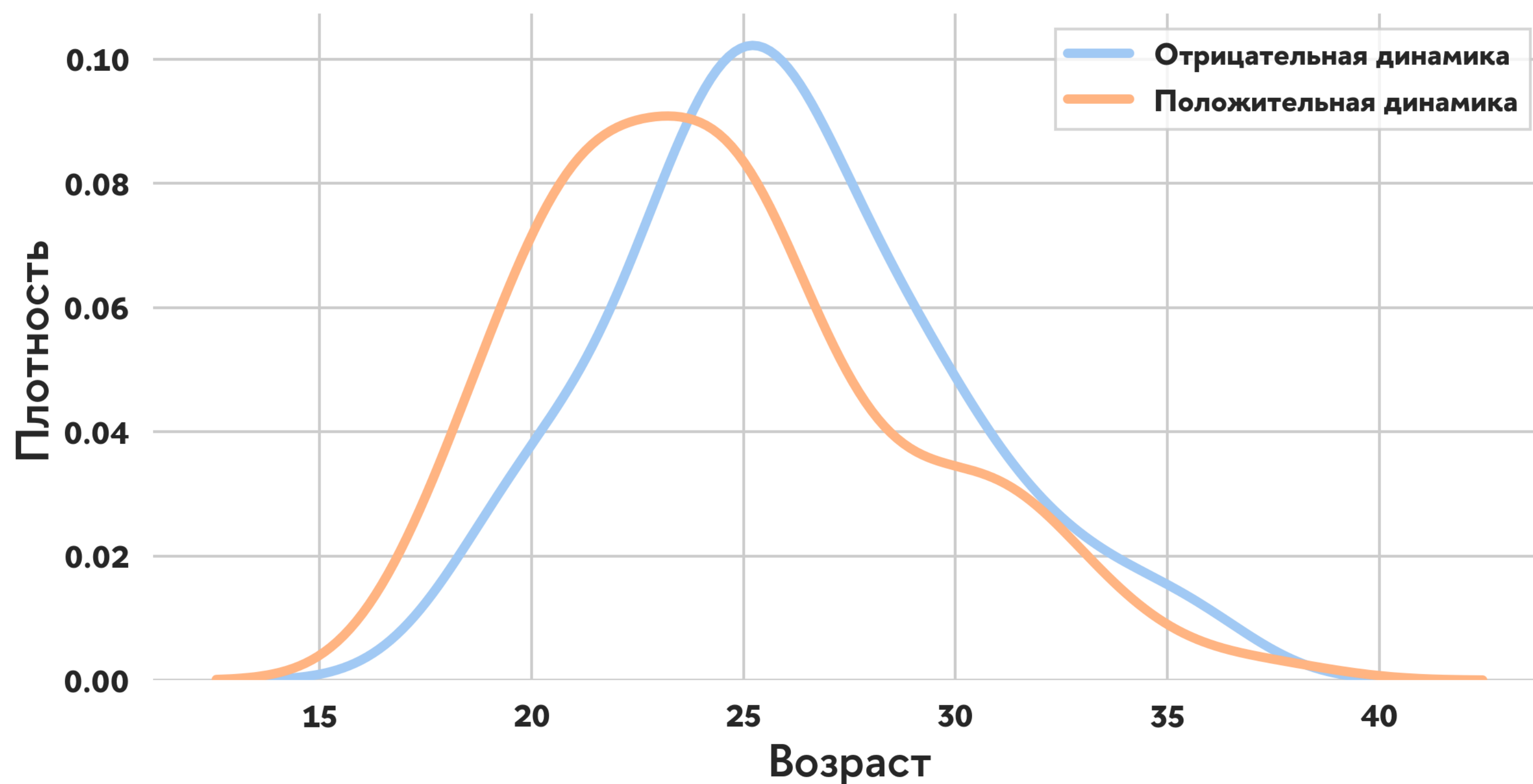
Мы будем использовать
параметрический Т-критерий
Стьюдента





Т-критерий Стьюдента

Распределение возрастов среди групп
с положительной и отрицательной динамиками долей побед



Результат

Показатель	Значение	Гипотеза
p-value	0.008	✓





Проверка модели на устойчивость

Алгоритм

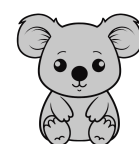
Разделение на группы отдельно по материкам и по
среднему числу минут



Построение Т-критерия для каждой группы



Проверка результата

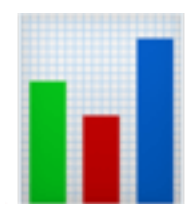




Разделение на группы по материкам

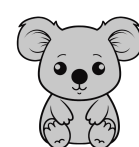
Результат

Группа	Размер группы	p-value	Гипотеза
Европа	96	0.009	✓
Другие континенты	99	0.006	✓



Разделение на группы по среднему числу минут в матче

Результат			
Группа	Размер группы	p-value	Гипотеза
Первая треть	65	0.007	✓
Вторая треть	65	0.008	✓
Третья треть	65	0.006	✓





Интерпретация



Гипотеза подтвердилась

У более молодых игроков чаще имеется
возрастающая тенденция процента
побед, чем у возрастных

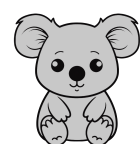
Произведена проверка устойчивости



Разделение по
континентам



Разделение по среднему
числу минут в матче





Ограничения и перспективы

Ограничения

Исследование распространяется только на чемпионаты, отслеживаемые АТР и представлены не все данные

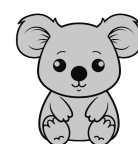
Только 2022–2024 год

Перспективы

Добавить больше биологических и исторических характеристик для каждого игрока

Отслеживать и более низкоуровневые турниры для выявления более общих паттернов

Провести исследование на данных с более большим временным отрезком





Policy implication

Букмекерам:

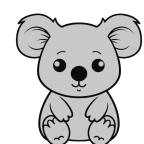
Учитывать возраст игроков при подсчёте коэффициентов.
Придумать новые виды ставок, например "Кто выстрелит в новом сезоне?"

Теннисным сборным:

Стоит чаще менять состав и брать молодых игроков

Спортсменам:

Стараться как можно раньше выходить на международные соревнования
Не терять мотивацию, даже если уже добились успеха





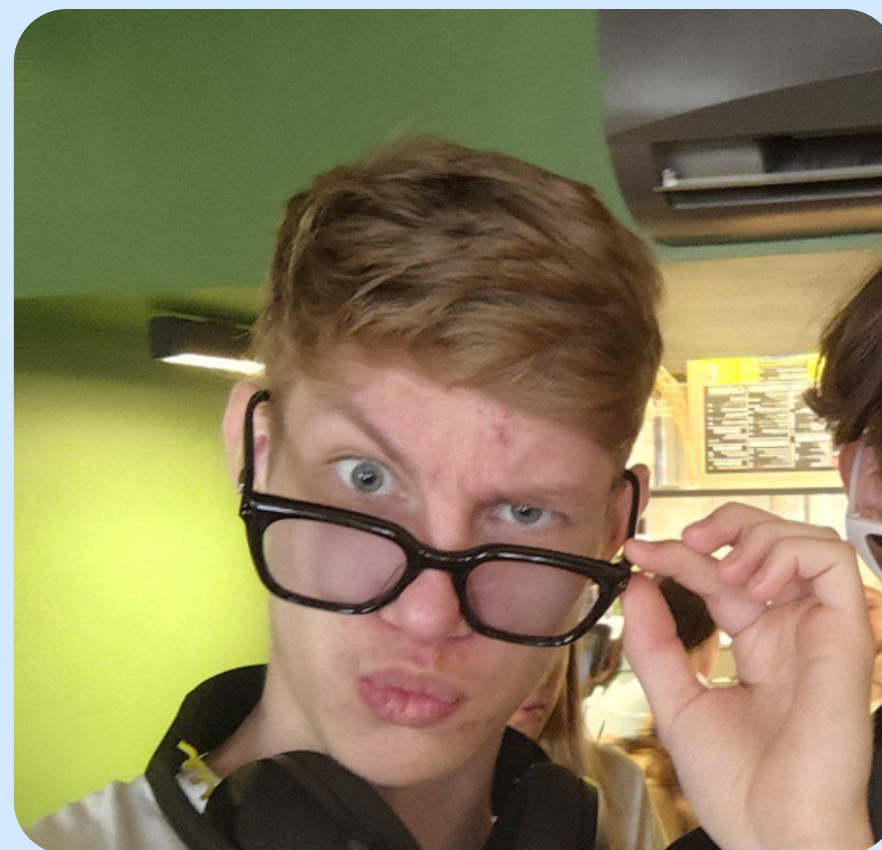
Команда



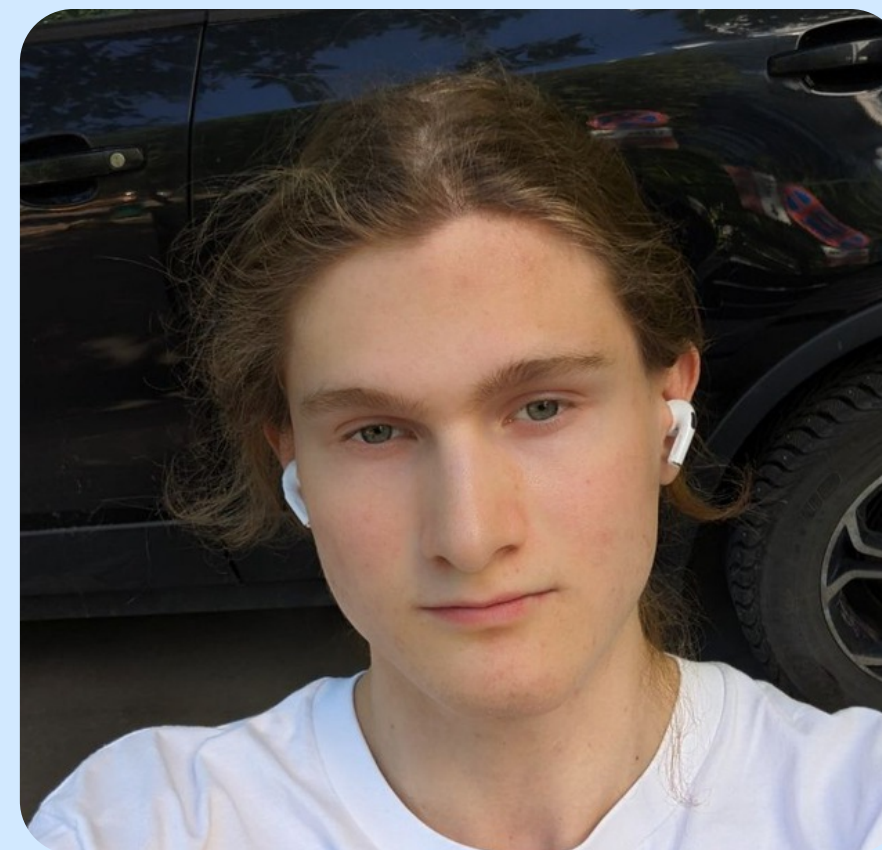
Данис



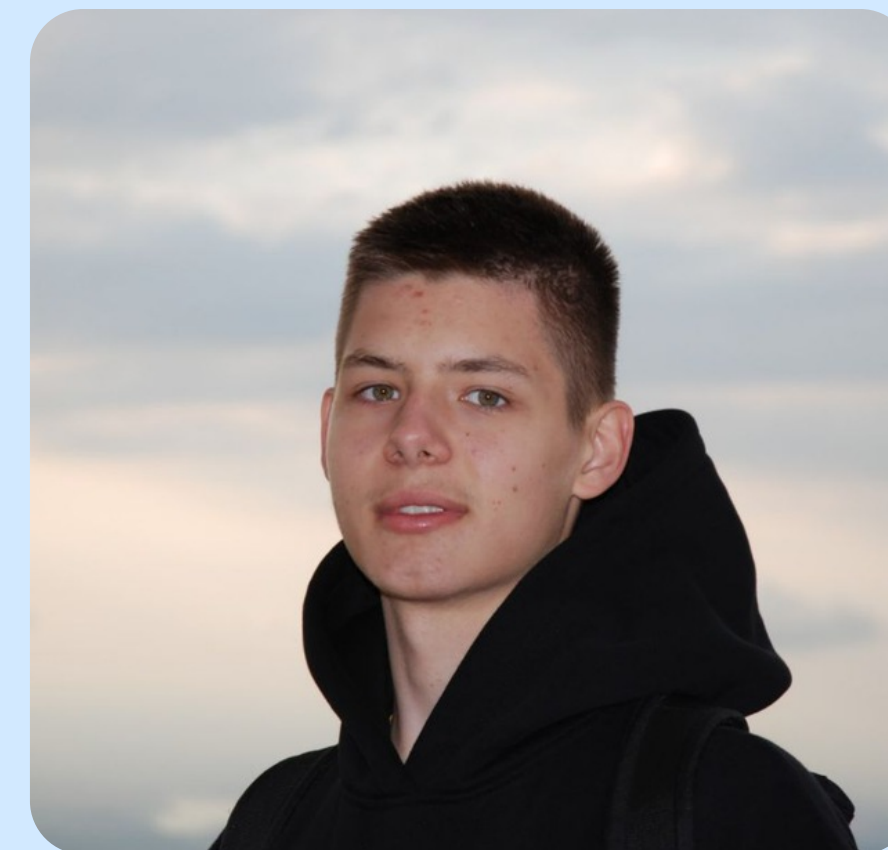
Герман



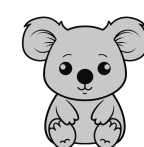
Илья



Андрей



Арсений



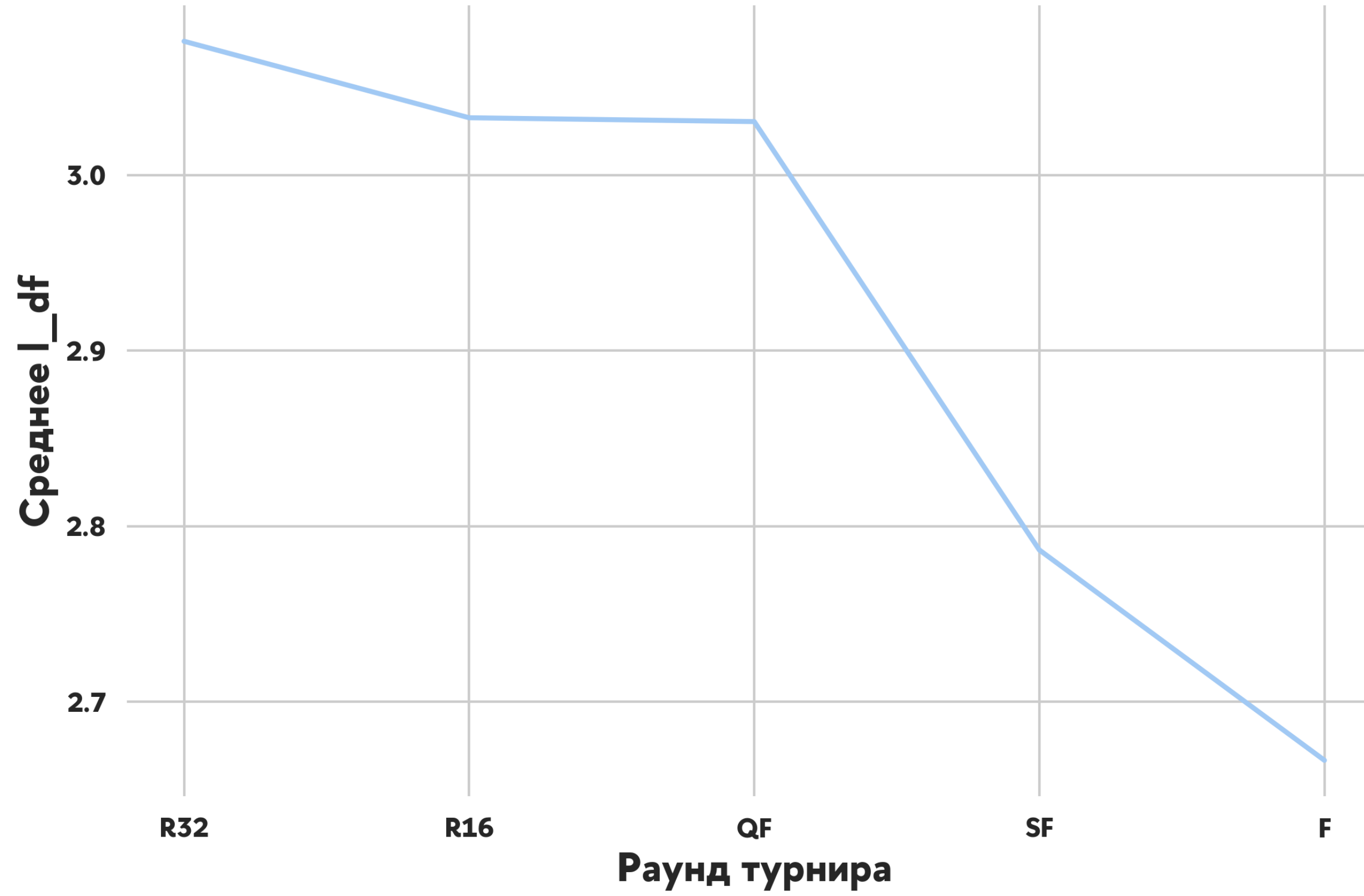
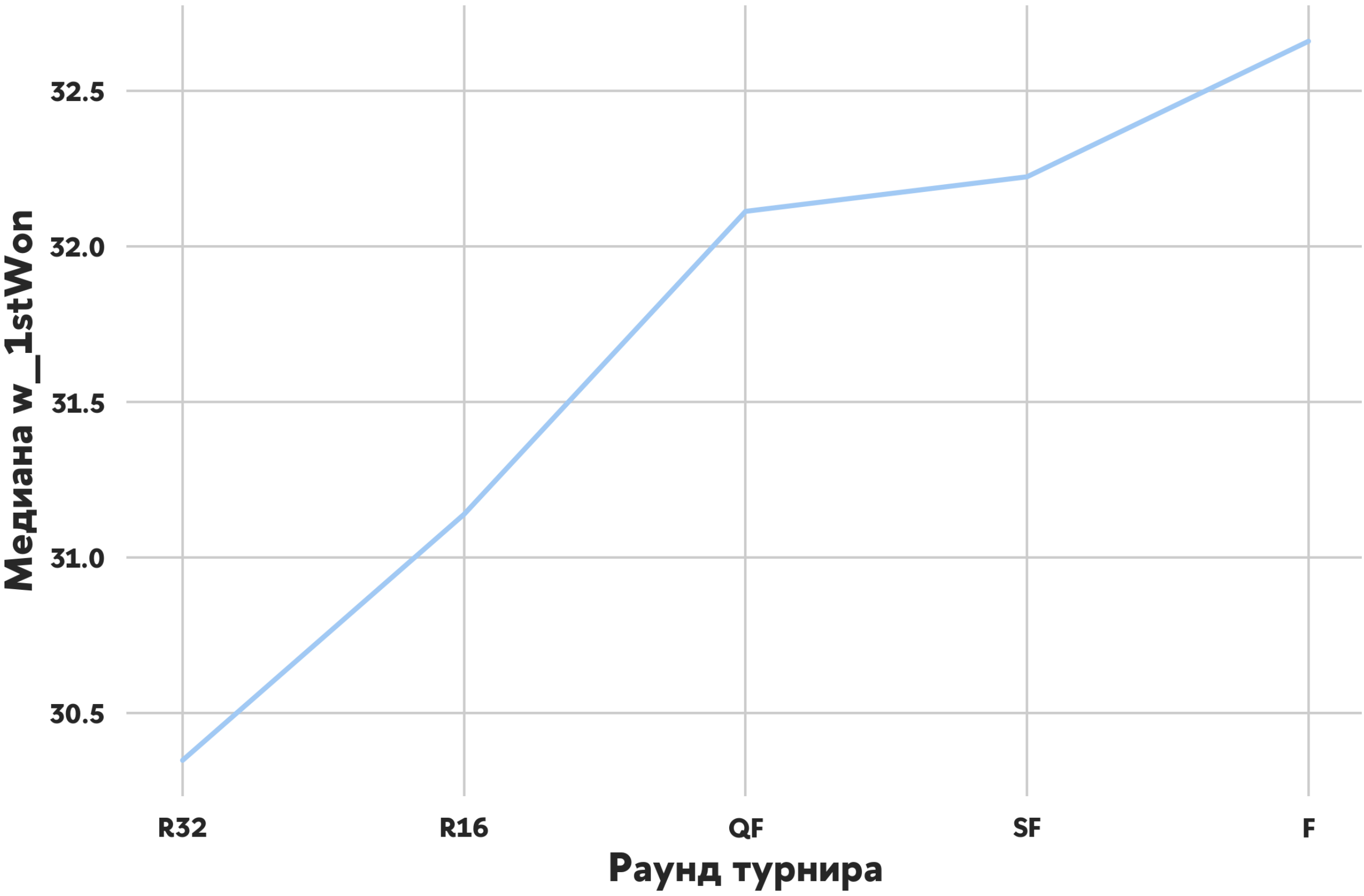


Спасибо за внимание!

Остались ли вопросы?



Тенденции по раундам турнира



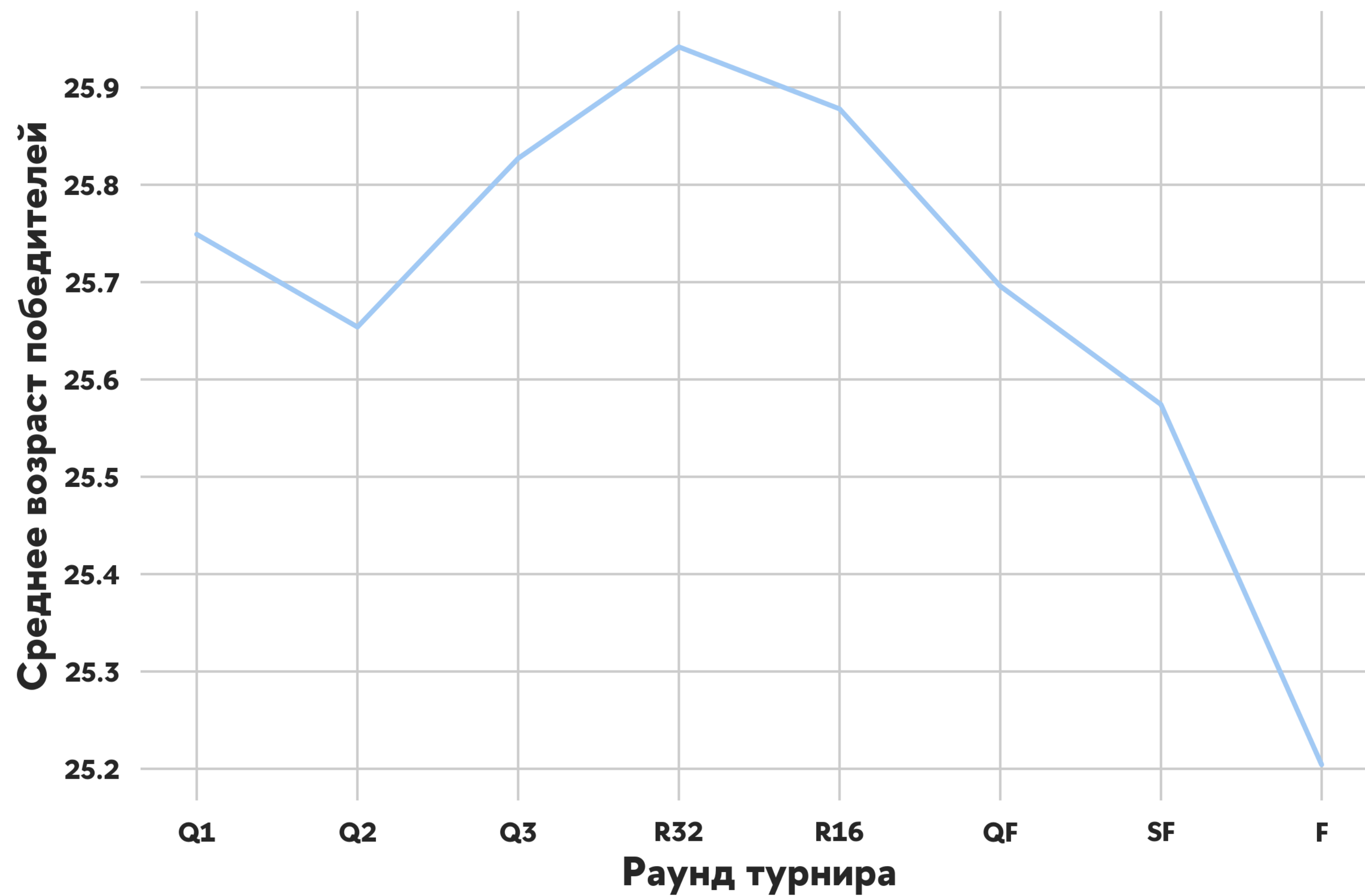


Вариант формулы для вычисления рейтинга

$$\frac{a}{a+b} - 1,65 \sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$$



Средний возраст по раундам турнира





Матрица корреляций

Возраст выигравшего	1	0.0079	0.036	-0.027	-0.018	-0.023	-0.012
Возраст проигравшего	0.0079	1	0.035	-0.023	-0.013	-0.02	-0.01
Кол-во игроков в сетке	0.036	0.035	1	0.00026	0.01	-0.029	-0.015
BP-Faced y выигравшего	-0.027	-0.023	0.00026	1	0.9	0.092	0.099
BP-Saved y выигравшего	-0.018	-0.013	0.01	0.9	1	0.097	0.093
BP-Faced y проигравшего	-0.023	-0.02	-0.029	0.092	0.097	1	0.9
BP-Saved y проигравшего	-0.012	-0.01	-0.015	0.099	0.093	0.9	1
	Возраст выигравшего	Возраст проигравшего	Кол-во игроков в сетке	BP-Faced y выигравшего	BP-Saved y выигравшего	BP-Faced y проигравшего	BP-Saved y проигравшего

