

## Fantasy Premier League. Мастерство или удача?

Меня этот вопрос волнует давно. Чтобы разобраться, я предпринял серьёзное исследование, по ходу которого пришлось преодолеть ряд технических трудностей. Краткий ответ звучит так: хороший результат в одном сезоне может быть получен благодаря удаче. Закономерно из сезона в сезон занимать высокие места могут лишь мастера. Давайте начнем с представления четырех выдающихся игроков Fantasy Premier League:

entry	23984	entry	728021	entry	1707	entry	116739
Rank	Paul Marshman	Rank	Fábio Borges	Rank	richard clarke	Rank	Mark Sutherns
2006/07	787	2014/15	17 996	2006/07	13 425	2006/07	2 217
2007/08	512 345	2015/16	4 682	2007/08	29 516	2007/08	2 392
2008/09	101 182	2016/17	1 145	2008/09	5 184	2008/09	268
2009/10	2 797	2017/18	671	2009/10	4 750	2009/10	29 460
2010/11	196	2018/19	95	2010/11	7 422	2010/11	3 285
2011/12	44 332	2019/20	1 965	2011/12	303	2011/12	382
2012/13	4 160	2020/21	1 828	2012/13	17 902	2012/13	2 044
2013/14	275 090	2021/22	222	2013/14	11 268	2013/14	192 837
2014/15	427			2014/15	8 237	2014/15	42
2015/16	229			2015/16	32 547	2015/16	28 329
2016/17	523			2016/17	1 354	2016/17	115
2017/18	308 969			2017/18	263	2017/18	11 458
2018/19	24 623			2018/19	14 517	2018/19	1 367
2019/20	10 115			2019/20	7 800	2019/20	177 309
2020/21	1 731			2020/21	13 987	2020/21	2 061
2021/22	34 060			2021/22	48 076	2021/22	156 321

Рис. 1. Четыре потрясающих достижения: Paul Marshman наряду с тремя другими игроками 5 раз финишировал в TOP-1k, Fábio Borges с большим отрывом от преследователей имеет самое низкое среднее значения ранга за последние пять сезонов, Richard Clarke единственный, кто во всех 16 сезонах финишировал в TOP-50k, Mark Sutherns больше всех – 10 раз – финишировал в TOP-5k

### Завязка

Я увлекаюсь Fantasy Premier League, статистикой и Excel. Объединение этих интересов и позволило получить результаты, представленные в заметке.

Уже много лет я строю модели на основе больших данных FPL (см. [Лучшая команда Fantasy Premier League сезона 2017/18](#)). Программирование не является моим коньком, и для скрапинга данных с сайта FPL я заказал программу у профессионалов. Но со временем условия изменились. Сначала сайт стал запрашивать подтверждение для входа, а затем потребовалось работать под VPN. В этих условиях скрапинг был затруднен.

Я не унывал, и начал гуглить. И удача! Нахожу [описание](#) и [библиотеку](#) python для скрапинга сайта FPL. Прошу моего друга Вадима Миняйло помочь с python. После первых пробных попыток Вадим сообщает, что оказывается FPL предоставляет доступ к данным по API. Гуглю более прицельно, и нахожу [перечень](#) точек доступа. Настраиваю запросы с помощью Power Query в Excel, и получаю массив данных – около 100М строк. Загружаю данные в модель Power Pivot и... комп не справляется. Покупаю новый в максимально доступной конфигурации – с процессором Intel Core i9-12900K, 3.20 GHz и оперативной памятью 32GB.

### Методика

В рамках подготовки к сезону 2022/23 я прочитал несколько книг, и у [Пола Роджерса](#) нашел ссылку на настоящее научное исследование опубликованное в рецензируемом электронном журнале [Джозеф Д. О'Брайен, Джеймс П. Глисон, Дэвид Дж.П. О'Салливан. Идентификация мастерства в Fantasy Premier League](#). Исследование весьма интересное, но у меня к нему есть одно замечание.

Авторы изучали TOP-1М игроков по итогам сезона 2018/19. Проблема в том, что часть игроков из этой выборки попали на верхние строчки благодаря везению, а не подтвержденному мастерству. На мой взгляд интереснее выделить элитных игроков на основании многолетних результатов, и именно эту группу сравнивать с «поляной».

В цикле заметок я планирую представить вам интересные выводы о том, что отличает элитных игроков, а начну с выделения элиты.

Я скачал и проанализировал данные о истории выступлений 9,6М менеджеров, зарегистрировавшихся на сезон 2022/23 к 1 сентября 2022 г.

### *Статистика участия*

Точные данных о числе участников за прошлые годы я не нашел, но в качестве оценки выбрал наихудший рейтинг в соответствующем сезоне среди менеджеров, зарегистрировавшихся на сезон 2022/23.

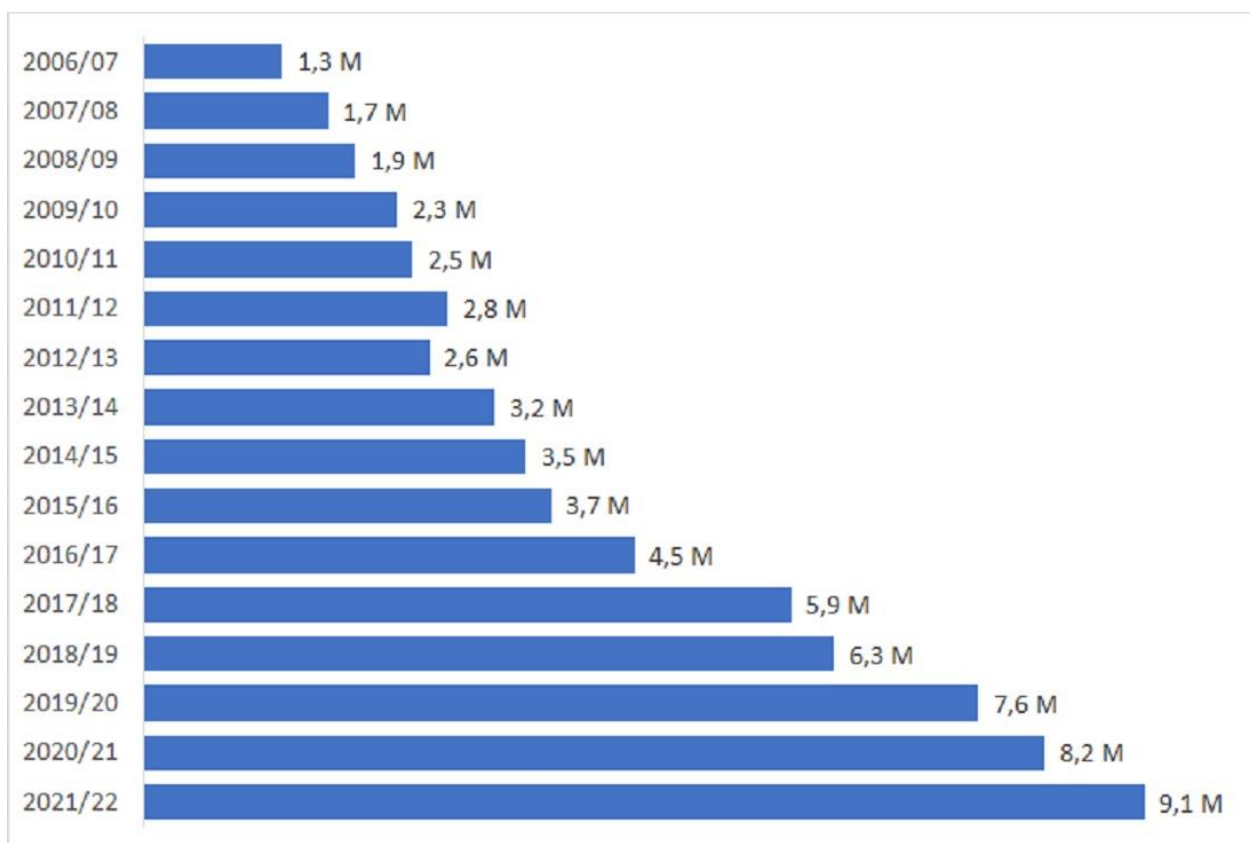


Рис. 2. За 16 сезонов количество аккаунтов увеличилось в 7 раз

Современную базу данных FPL ведет с сезона 2006/07. За более ранние периоды доступны лишь разрозненные данные.

Более трети аккаунтов – 3 464 000 – впервые зарегистрировались на сезон 2022/23. Но также есть и 52 тысячи менеджеров, которые приняли участие во всех 16 сезонах.

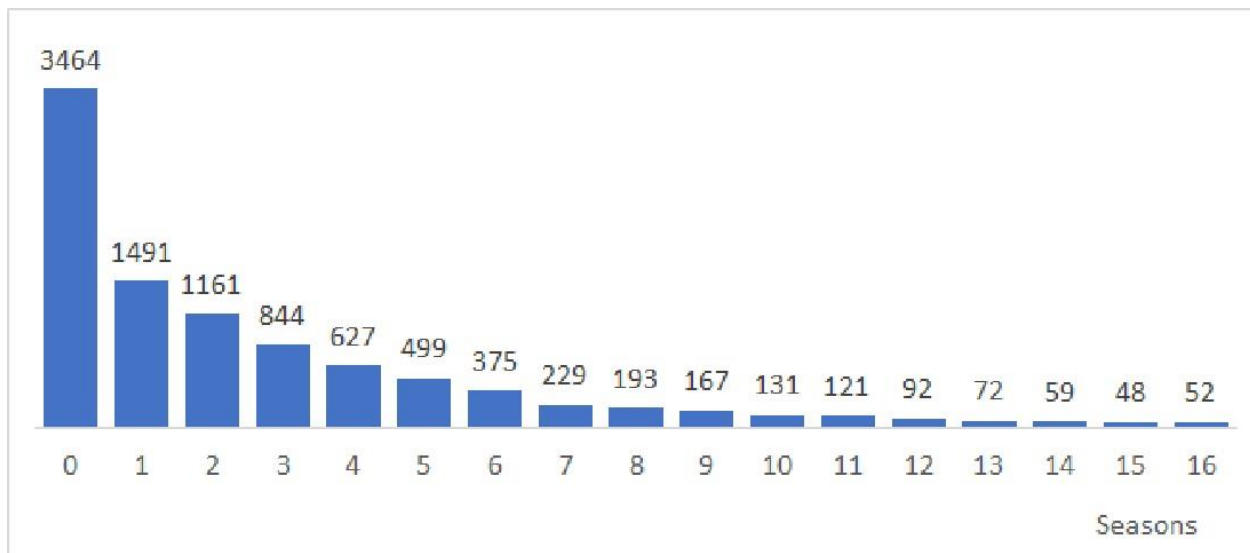


Рис. 3. Распределение аккаунтов по количеству участия в предыдущих сезонах, x1000  
 Естественно, что опытные игроки регистрируются раньше))

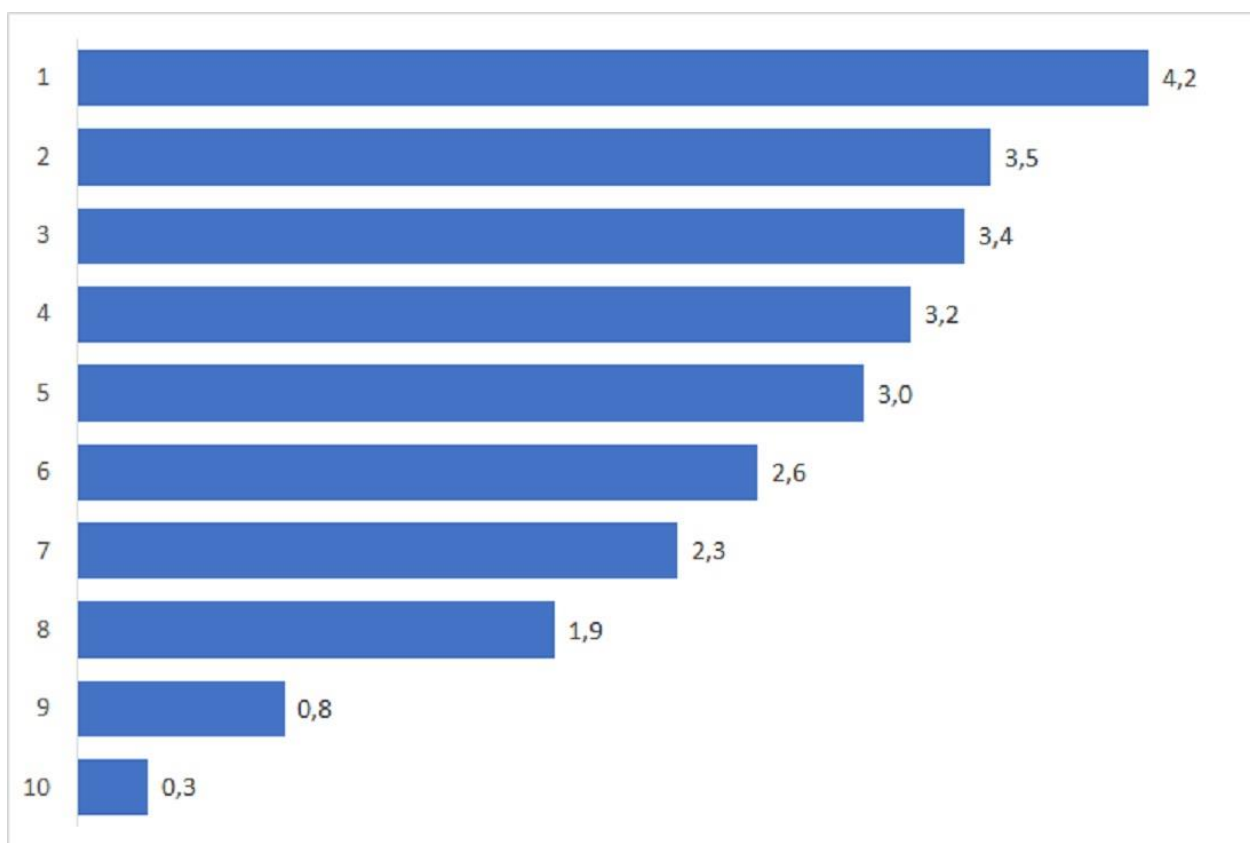


Рис. 4. Среднее число сезонов по группам аккаунтов; 1 по оси ординат соответствует аккаунтам с номерами от 1 до 1 000 000, 2 – аккаунты с номерами от 1 000 001 до 2 000 000 и т.д.

### Элита

Чтобы определить, кого отнести к элите, мне предстояло ответить на ряд вопросов. Учитывать ли все выступления или только несколько последних? Отбирать элиту по абсолютным или относительным рейтингам? Ориентироваться на средние показатели или на число вхождений в ТОП? И наконец, сколько игроков отбирать? 100, 1000, 10 000?

Надо сказать, что элитные игроки представлены на ряде сайтов. Например, [livefpl.net](http://livefpl.net), [fplstatistics](http://fplstatistics.com), [fplreview](http://fplreview.com), [myfplanalysis.co.in](http://myfplanalysis.co.in) и [premierfantasytools.com](http://premierfantasytools.com). Я не думаю, что тот или иной список лучше или хуже. Важен общий принцип – высокие результаты на протяжении нескольких сезонов.

Чтобы не было перекоса в сторону игроков с большим числом участий, я решил считать результаты только за пять последних сезонов. Дабы отсеять случайные успехи, я не учитывал игроков с небольшим числом участий. За последние пять сезонов число аккаунтов увеличилось в полтора раза с 5,9М до 9,1М. С одной стороны, относительный рейтинг более справедлив. Ведь попасть в ТОП-10k при 5,9М участников легче, чем при 9,1М. С другой стороны, абсолютные рейтинги видны на сайте, и именно они фигурируют в обсуждениях. Что такое ТОП-10k понятно всем, а что такое ТОП-0,1%?.. На мой взгляд центральная мера (среднее) лучше отражает стабильность результатов, чем число попаданий в ТОП. И наконец медиана более робастна,<sup>1</sup> чем среднее арифметическое, подверженное влиянию даже одного относительно неудачного сезона. Я включил в элиту 10 000 игроков:

1. 9318 лучших по медианному значению рейтинга, принявших участие во всех пяти сезонах 2017/2018 – 2021/22; медианное значение не более 55 000.
2. 537 лучших, принявших участие в четырех из пяти сезонов; медианное значение не более 40 000.
3. 145 лучших, принявших участие в трех из пяти сезонов; медианное значение не более 13 300.

Я решил на этом этапе отобрать в элиту 10k. А на последний важный вопрос, сколько игроков включать в элиту, ответить в рамках последующих исследований.

Чтобы оценить уровень игроков, посмотрите на показатели лучших и «худших» по категориям:

entry 728021		entry 6810943		entry 8807	
<b>Rank</b>	1)	<b>Rank</b>	2)	<b>Rank</b>	3)
<b>Fábio Borges</b>		<b>Magne Loftesnes</b>		<b>Gio Alarcon</b>	
2017/18	671	2018/19	277	2019/20	10 089
2018/19	95	2019/20	103 769	2020/21	457
2019/20	1 965	2020/21	366	2021/22	325
2020/21	1 828	2021/22	623		
2021/22	222				
entry 5188805		entry 1351403		entry 18411	
<b>Rank</b>	4)	<b>Rank</b>	5)	<b>Rank</b>	6)
<b>Gaurav Saha</b>		<b>Imraan Kaprey</b>		<b>Wissam Jawher</b>	
2017/18	11 433	2018/19	14 997	2019/20	3 039 391
2018/19	34 366	2019/20	220 087	2020/21	13 295
2019/20	241 488	2020/21	323	2021/22	2 238
2020/21	54 984	2021/22	64 982		
2021/22	149 310				

Рис. 5. Показатели игроков: 1) 1-й в категории «5 сезонов», 2) 1-й в категории «4 сезона», 3) 1-й в категории «3 сезона», 4) 9318-й в категории «5 сезонов», 5) 537-й в категории «4 сезона», 6) 145-й в категории «3 сезона»

Полный список элитных аккаунтов можно найти в приложенном Excel-файле. Также для дальнейшего исследования выделил три группы. Большая группа включает меньшие.

Group	AvgMedian	Entries
TOP-1k	10 455	1 000
TOP-5k	23 326	5 000
TOP-10k	33 979	10 000

Рис. 6. Средняя медиана в сезонах 2017/2018 – 2021/22 и количество аккаунтов по элитным группам

### Группа сравнения

Поскольку из тура в тур изучать все аккаунты (сейчас их более 10М) не представляется возможным, я выделил группу сравнения, и назвал ее *поляной*. Эту группу я сформировал на основе стратифицированной выборки:

<sup>1</sup> [Робастность](#) — свойство статистического метода, характеризующее устойчивость к помехам.

Страта	Entries		Размер выборки
	с	по	
1	1	1 000 000	22 000
2	1 000 001	2 000 000	17 000
3	2 000 001	3 000 000	13 000
4	3 000 001	4 000 000	10 000
5	4 000 001	5 000 000	8 000
Итого			70 000

Рис. 7. Правила формирования поляны

Такой выбор связан с тем, что аккаунты с меньшими номерами формируются более опытными и вовлеченными менеджерами, поэтому стратифицированная выборка больше соответствует моей цели – сравнить игру элитного игрока и среднего. Среди аккаунтов с большими номерами выше процент фейковых и случайных игроков.

Поскольку выбор поляны был случайным, в ней присутствует и некоторое число элитных игроков. Полный список аккаунтов, включенных в поляну, можно найти в приложенном Excel-файле.

В целом параметры выборок по всем участиям (а не только в сезонах 2017/2018 – 2021/22) выглядят следующим образом:

Group	AvgMedian	Seasons	Entries
TOP-1k	31 085	9,93	1 000
TOP-5k	45 382	9,80	5 000
TOP-10k	56 888	9,73	10 000
Поляна	1 445 437	3,60	70 000
Все аккаунты	1 575 639	2,64	9,6M

Рис. 8. Параметры выборок