

Серталп Б. Кей. Аналитика Fantasy Premier League

В рамках подготовки к сезону FPL 2022/23 продолжаю изучать книги и материалы по теме. Недавно я опубликовал [Пол Роджерс. Fantasy Premier League. Как преуспеть с платными трансферами](#).

Настоящая заметка – это перевод статьи Sertalp B. Cay. A Friendly Introduction to FPL Analytics, опубликованной 4 января 2022 г. в блоге [alpscode](#). В заметке представлены основные аналитические инструменты, используемые в Fantasy Premier League (FPL).

This Season

GW	OPP	Pts	MP	GS	A	CS	GC	OG	PS	PM	YC	RC	S	B
13	WHU (H) 2-1	2	90	0	0	0	1	0	0	0	0	0	0	0
14	AVL (A) 1-2	10	90	1	0	0	1	0	0	0	0	0	0	3
15	WAT (A) 1-3	15	90	2	0	0	1	0	0	0	0	0	0	3
16	WOL (H) 1-0	6	90	0	1	1	0	0	0	0	0	0	0	0
17	LEE (H) 7-0	1	45	0	0	0	0	0	0	0	0	0	0	0
18	NEW (A) 0-4	2	90	0	0	1	0	0	0	0	1	0	0	0

Рис. 1. Базовая статистика Бернарду Силвы, сезон 2021/22, GW13–18

Часть 1. За кулисами базовой статистики

Как планировать трансферы? Можно обратиться к статистике FPL и выбрать лучшего игрока в соответствующей позиции и ценовой категории. Однако [стандартная статистика FPL](#) – очень шумный показатель будущей отдачи игрока. Шумный означает, что существует большая неопределенность (дисперсия). Вы также знаете, что результат зависит от сложности матча: иногда команды проводят сложные выездные игры, а иногда очень легкие домашние. Таким образом, не все выступления одинаковы. Так что предсказания стандартной статистики могут вводить в заблуждение.

Давайте начнем с примера. Вы хотите оценить, сколько голов забьет Бернарду Силва в следующем матче. Для этого вы просматриваете последние игры Силвы (см. рис. 1). В 6 матчах он забил 3 гола. Но его ожидаемое число голов ([xG](#)) было ниже. Это означает, что средний игрок с такими же атакующими действиями, как у Силвы, скорее всего, забил бы меньше 3 голов. Вы можете подумать: «Ну, он не обычный игрок!». Это может быть правдой, но игроки, которые могут последовательно превосходить свой показатель xG встречаются редко (очень редко... единственный, кто известен – это Лионель Месси!).

Date	Day	Round	Venue	Result	Opponent	Pos	Min	Performance				Expected		
								Gls	Ast	Sh	SoT	xG	npG	xA
2021-11-28	Sun	Matchweek 13	Home	W 2-1	West Ham	CM,FW	90	0	0	1	1	0.2	0.2	0.1
2021-12-01	Wed	Matchweek 14	Away	W 2-1	Aston Villa	AM	90	1	0	1	1	0.2	0.2	0.2
2021-12-04	Sat	Matchweek 15	Away	W 3-1	Watford	CM	90	2	0	3	2	0.4	0.4	0.8
2021-12-11	Sat	Matchweek 16	Home	W 1-0	Wolves	CM	90	0	0	3	2	0.5	0.5	0.1
2021-12-14	Tue	Matchweek 17	Home	W 7-0	Leeds United	CM	45	0	0	1	0	0.6	0.6	0.0
2021-12-19	Sun	Matchweek 18	Away	W 4-0	Newcastle Utd	DM	90	0	0	1	0	0.1	0.1	0.5

Рис. 2. Расширенная статистика по данным сайта [fbref](#)

В первых трех играх Бернарду забил 3 гола, а его xG = 0,8. Он превзошел ожидания более, чем в три раза. Таким образом, несмотря на то, что он забил 3 гола в 3 матчах, это не выглядит устойчивой тенденцией. В следующих трех матчах Силве не удалось забить, а его xG = 1,2. Хотя ему предоставилось два неплохих шанса, он не смог их реализовать.

Исследования показали, что xG является лучшим предиктором будущих голов, чем базовая статистика FPL. Аналогично, в качестве предиктора xA (ожидаемые голевые передачи) работает лучше, чем реализованные ассисты. Интуитивно понятно, что если игрок продолжит наносить удары или делать отличные передачи, он рано или поздно забьет или отдаст голевую передачу.

Таким образом, лучше сравнивать игроков по прхG (ожидаемые голы без пенальти, поскольку количество пенальти искажает цифры) и xA.

Добро пожаловать в аналитику FPL!

Часть 2: Измерение сложности матча

Одним из наиболее известных показателей сложности матча в FPL является Рейтинг сложности матчей ([Fixture Difficulty Ratings](#), FDR). FDR один из самых важных инструментов принятия решений для многих менеджеров.

This Season			
Date	GW	Opponent	FDR
Sun 26 Dec 10:00	19	NOR (A)	2
Tue 28 Dec 07:30	20	WOL (H)	3
Sat 1 Jan 07:30	21	MCI (H)	4
Sun 16 Jan 11:30	22	TOT (A)	3
Sat 22 Jan 10:00	23	BUR (H)	2

Рис. 3. Пример FDR для предстоящих матчей Арсенала сезона 2021/22

Возможно, FDR – хорошая отправная точка для понимания того, как ориентироваться в расписании, но его точно недостаточно. Вполне закономерно, мы можем спросить, почему и WOL, и TOT имеют рейтинг 3. И хорошо ли иметь защитника Арсенала в этих матчах, или лучше иметь нападающего Арсенала? И насколько велика разница между оценкой FDR 2 и 3?

TEAM	TEAM RATING	
	SPI	OFF. DEF.
 Tottenham 26 pts	72.1	2.0 0.8
 Wolves 25 pts	72.4	1.7 0.6

Рис. 4. Сравнение рейтингов команд от FiveThirtyEight¹

Если вы зайдете на страницу прогнозов Премьер-лиги [FiveThirtyEight](#), вы сможете увидеть разницу между этими двумя командами. У TOT сила атаки 2,0, а у WOL – 1,7. То есть, можно ожидать, что TOT забьет 2 гола против среднего соперника, а WOL – 1,7. С другой стороны, WOL имеют лучший оборонительный рейтинг, чем TOT. Они пропускают против среднего соперника 0,6 гола за игру, а TOT – 0,8. Что касается матчей Арсенала, то лучше иметь защитника для GW20 и нападающего для GW22, чем наоборот.

Но, и это все еще слишком упрощенный подход.

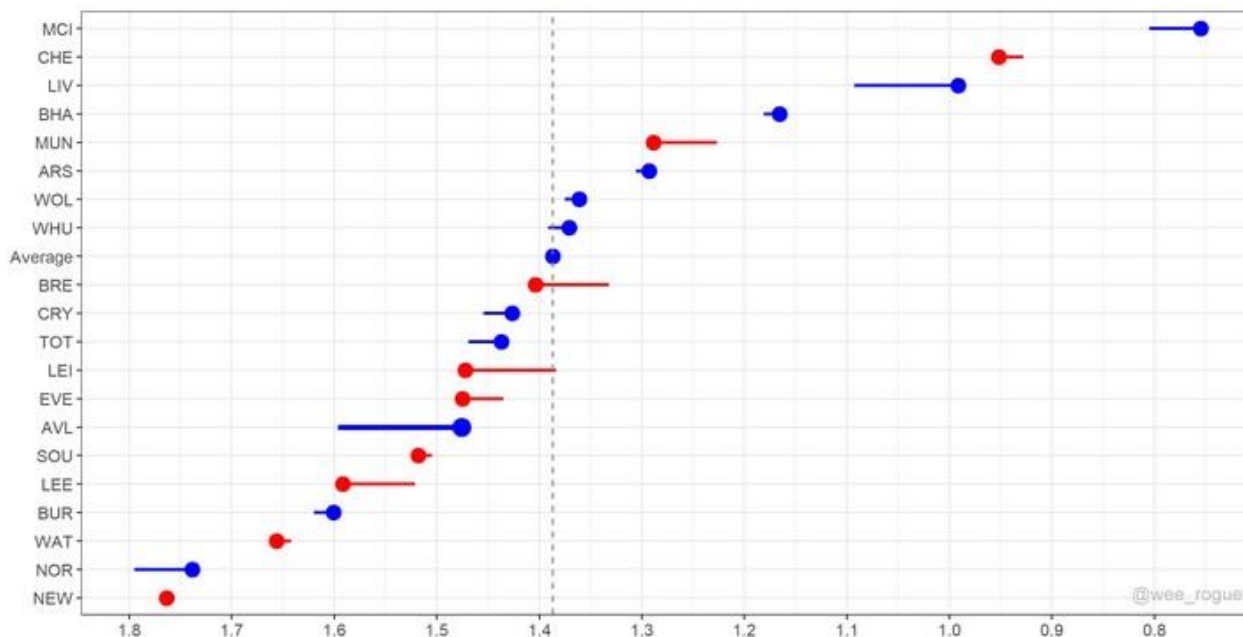
Wee Rogue ([@wee_rogue](#)) использует аналогичные методы и генерирует показатель *Прогнозируемые пропущенные / забитые голы за матч против среднего соперника Премьер-лиги*. Эта модель лучше подходит для целей FPL, чем использование FDR или данные с сайта 538.

¹ FiveThirtyEight – это сайт аналитики в области политики, спорта и др. 538 – число выборщиков президента США.

Premier League 2021-22 - Defence Ratings

Dots: Estimated Rating at 2021-12-14 After Round 17

Lines: Movement in Rating Since Gerard Took Over (Last 6 Matches)



Predicted Goals Conceded per Match vs Average Premier League Opponent

Proprietary Model: @wee_rogue
Data: @fivethirtyeight



Рис. 5. Оценки атаки и защиты от @wee_rogue

Данные от Wee Rogue в ином виде также представлены на сайте [Elevenify](https://www.elevenify.com/).

Team	Projected Goals Scored (next 8 GWs)	GW19	GW20	GW21	GW22	GW23
Arsenal		1.6	1.5	0.9	1.4	1.8

Team	Projected Goals Conceded (next 8 GWs)	GW19	GW20	GW21	GW22	GW23
Arsenal		1.1	0.9	1.9	1.4	1.0

Рис. 6. Прогнозы на предстоящие матчи Арсенала на сайте Elevenify

В табличном виде показано прогнозируемое количество голов. В верхней части – ожидаемые забитые голы, в нижней – ожидаемые пропущенные. Теперь у нас больше данных, чтобы оценить сложность ближайших матчей.

Часть 3: Вероятности и выигрыши (математическое ожидание)

Любое решение ведет к ряду возможных результатов. Каждый отдельный результат (клиншит, пропуск одного мяча, пропуск двух мячей и т.д.) имеет два свойства: вероятность события и результат события (очки FPL). Объединить вероятность и результат можно с помощью математического ожидания – Expected Value, EV:

$$(1) EV = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n$$

где x_1, x_2, \dots, x_n – очки, набранные при наступлении событий 1, 2, ..., n ; p_1, p_2, \dots, p_n – вероятность наступления событий.

Вероятность пропустить (или забить) 0, 1, 2 и т.д. голов лучше всего описывается [распределением Пуассона](#), которое подходит для редких событий:

$$(2) P(X) = \frac{e^{-\lambda} \lambda^X}{X!}$$

где $P(X)$ – вероятность X успешных испытаний, λ – ожидаемое количество успехов, e – основание натурального логарифма, равное 2,71828, X – количество успехов в единицу времени.

Под успехом может подразумеваться, как забитый гол, так и пропущенный. Например, в игре против WOL в GW20 мы ожидаем, что Арсенал пропустит $\lambda = 0,9$ гола (см. рис. 6). Тогда в Excel можно построить распределение Пуассона (формулы см. в приложенном файле Excel):

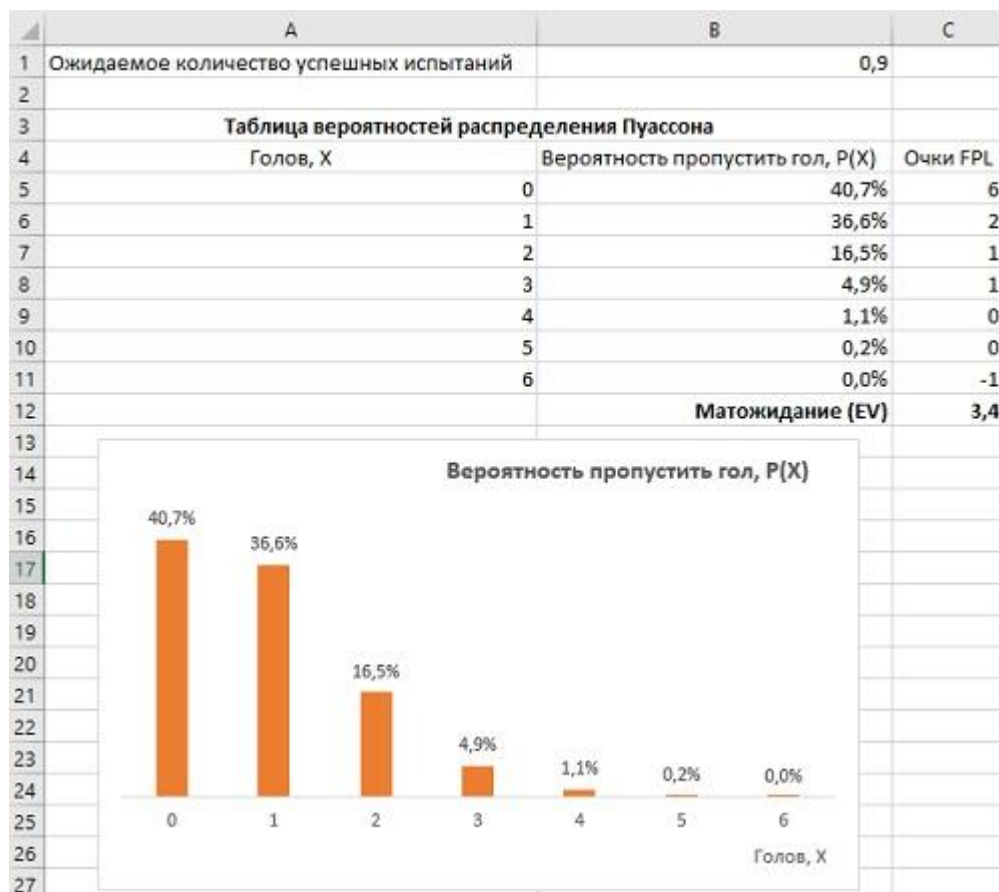


Рис. 7. Распределение вероятностей ожидаемых пропущенных голов Арсенала в матче с Волками

Вероятность клиншита = 40,7%, пропущенного гола = 36,6%, двух голов = 16,5% и т.д. Математическое ожидание очков защитника в матче против Волков составит:

$$(3) EV = 0,407 \cdot 6 + 0,366 \cdot 2 + 0,165 + 0,049 = 3,4$$

В матче против Шпор ожидается, что Арсенал пропустит $\lambda = 1,4$ гола (см. рис. 6). Математическое ожидание очков защитника соответственно меньше: $EV = 2,5$.

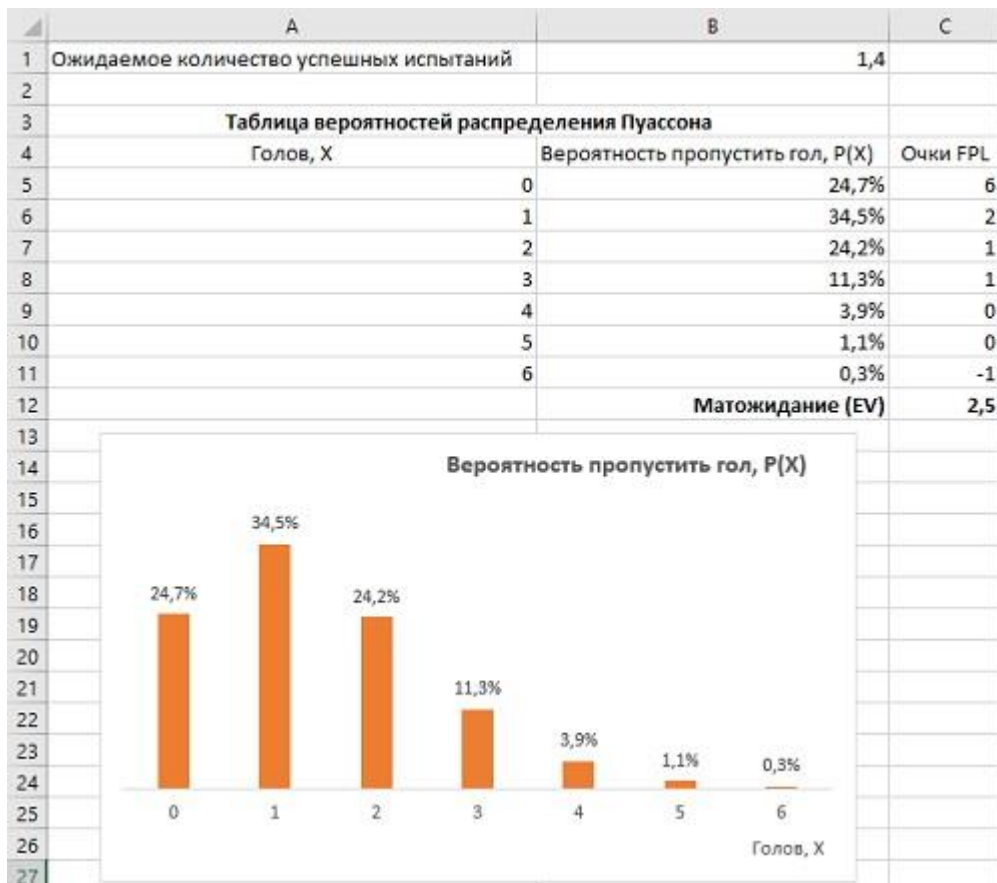


Рис. 8. Распределение вероятностей ожидаемых пропущенных голов Арсенала в матче со Шпорами
 Вы можете воспользоваться [калькулятором](#), чтобы определить ожидаемые пункты (xP) для одного или нескольких игроков:

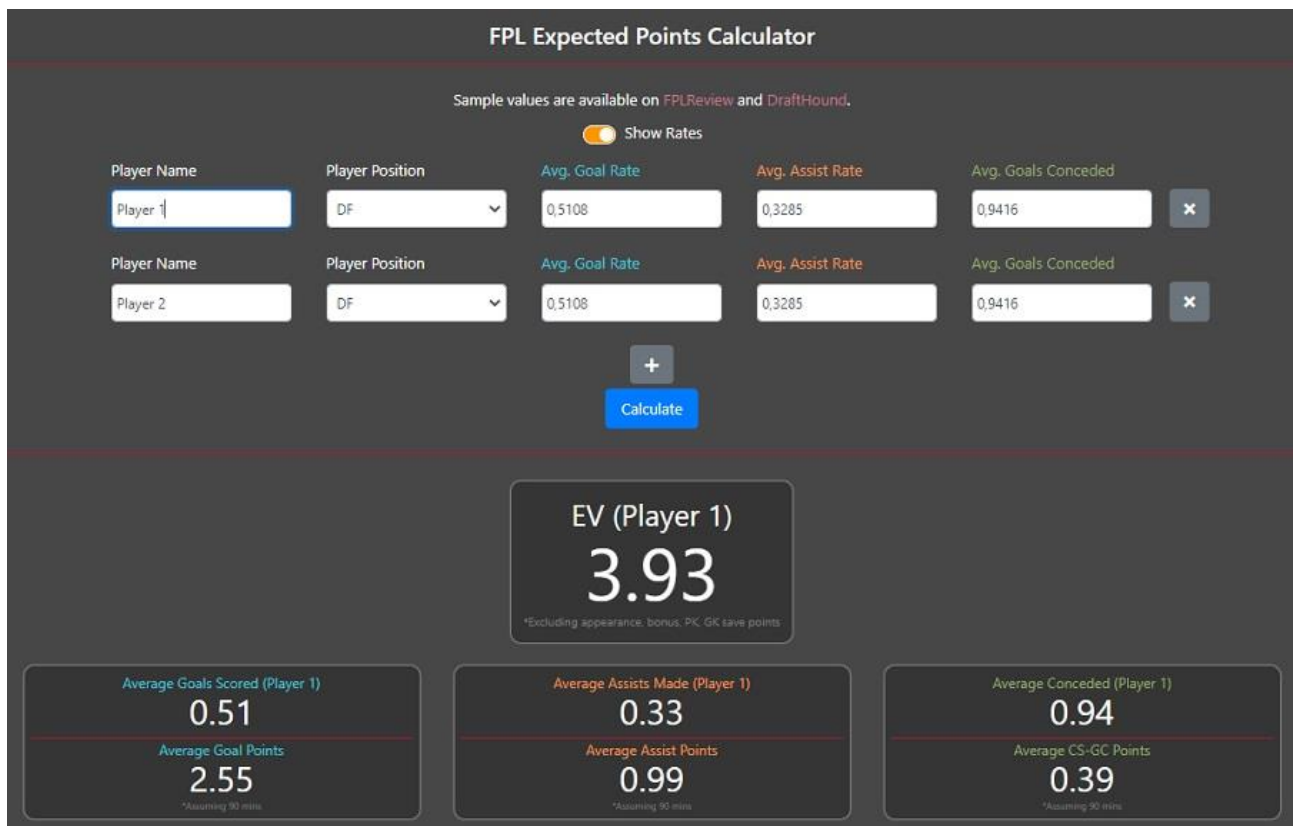


Рис. 9. Сравнение матожидания (EV) двух игроков

Существуют и иные прогнозы для действий игроков. Их можно найти в Интернете. Они также регулярно публикуются в Twitter ([@FPL_Salah](#)).

Часть 4: Прогнозирование будущего (прогнозная аналитика)

Если мы добавляем все больше и больше деталей в нашу модель EV (голы, передачи, бонусы, сейвы, пенальти), а также пытаемся предсказать стартовые составы команд, все становится сложнее. Теперь у нас есть неопределенность не только в отношении редких событий, таких как голы, но и в отношении некоторых уникальных аспектов FPL, таких как система бонусных очков. Вдобавок стартовые составы некоторых команды с трудом поддаются прогнозированию, например, МанСити (рулетка Пепа), и Челси (тасовка Тухеля).

Можно добавить больше деталей к нашей модели, рассчитав вероятность появления игрока (например, Фодена) в следующей игре. Если мы знаем, что существует 70%-ная вероятность того, что Фоден сыграет все 90 минут, 20%-ная вероятность того, что он выйдет на позднюю замену и получит 30 минут, и 9%-ная вероятность того, что он выйдет в стартовом составе, но будет заменен, опять же, мы можем рассчитать EV, используя потенциальные исходы и эти вероятности.

Для расчета EV можно добавить еще больше деталей: количество касаний в штрафной, кроссы и ключевые пасы. Сбор исходных данных и сохранение всех этих деталей может оказаться сложной задачей. Но уже есть энтузиасты, занимающиеся этим. Они следуют шагам, описанным выше: считывают данные о прошлых событиях, создают модель прогнозирования, предсказывают будущие вероятности и вычисляют EV для каждого игрока. Прогнозная аналитика для FPL переживает бум, и вы можете найти множество людей, делящихся результатами своих моделей. Некоторые из них бесплатны, других предоставляются по подписки. Наиболее известные: [FPLReview](#), the Transfer Algorithm от [@MikkelTokvam](#) и [FPL Kiwi](#), но есть и многие другие. Мы храним обновленный список моделей прогнозирования на нашем сервере Analytics Discord.

Pos	Name	Cost	Ave xMin	GW19	GW20	GW21	GW22	GW23	GW24	GW25	GW26
M	Salah	13.1	76	7.41	2.21	0.00	0.00	0.00	3.42	7.14	8.72
M	Mané	11.9	78	5.77	2.20	0.00	0.00	0.00	4.74	5.39	6.50
D	Alexander-Arnold	8.3	80	5.29	3.62	2.54	6.11	4.78	5.36	5.19	6.33
F	Kane	12.1	80	4.46	3.90	3.97	4.58	3.26	5.67	4.91	3.07
F	Ronaldo	12.5	75	4.34	4.93	4.21	5.08	5.65	5.41	6.10	5.61

Рис. 10. Пример прогноза с FPLReview

The FPL Kiwi - FPL Projections

Show entries

ID	Name	Pos	Team	Price	18		19		20		21	
					xPts	xMin	xPts	xMin	xPts	xMin	xPts	xMin
256	Cancelo	DEF	MCI	6.7	5.42	82.58	5.05	79.57	4.88	78.50	4.28	78.50
233	Salah	MID	LIV	13	5.40	75.50	7.22	76.50	2.90	40.97	0.00	0.00
237	Trent	DEF	LIV	8	4.69	79.79	6.01	79.79	4.68	79.79	3.84	79.79
40	Watkins	FWD	AVL	7.4	4.53	80.40	3.37	79.98	3.93	78.69	3.75	78.07
196	Raphinha	MID	LEE	6.7	4.46	81.02	2.60	65.50	3.79	67.50	4.75	76.90
251	De Bruyne	MID	MCI	11.8	4.43	63.50	4.73	63.50	4.48	63.50	4.23	63.50
30	Martinez	GK	AVL	5.5	4.41	87.81	3.69	87.81	3.92	87.81	4.06	87.81

Рис. 11. Пример прогноза с FPL Kiwi

Использование прогностической модели упрощает процесс принятия решений. Если вы заинтересованы в создании собственной модели, послушайте эпизод подкаста [Коридор неопределенности](#). Также подумайте о присоединении к нашему серверу Discord, если вы хотите поделиться своей моделью прогнозирования или у вас есть какие-либо вопросы.

Часть 5: Поиск пути внутри хаоса (оптимизация)

Когда у вас есть прогнозные данные, последним шагом процесса анализа становится поиск трансфера, который максимизирует общее количество ожидаемых баллов вашей команды (xP). Доступно несколько вариантов оптимизации. Первый основан на линейном программировании (Mixed-Integer Linear Programming, MILP), где каждое решение (какого игрока покупать или продавать) моделируется как отдельная двоичная переменная (да / нет). Если вас интересует, как такие модели пишутся в [Excel](#) и [Python](#) посмотрите на youtube.

Относительно несложная модель линейного программирования приведена мною в заметке [Лучшая команда Fantasy Premier League сезона 2017/18](#). – Прим. Багузина



Рис. 12. План трансферов на пять недель

Поскольку некоторые аспекты FPL не могут быть записаны линейно (например, автоматические замены), были попытки решить эту проблему иначе. Например, на сайтах FPLReview и FantasyFootballFix используются эвристические методы.

Основной проблемой оптимизации является ее высокая чувствительность к исходным прогнозам. При незначительном изменении последних картина предлагаемых трансферов может значительно измениться. Чтобы решить эту проблему, мы с [@FF Trout](#) протестировали прошлогодние изменения в прогнозах FPLReview. Вместо того, чтобы получить один план трансферов, мы выполняли оптимизацию несколько раз (методом MILP). В итоге получили набор данных, как часто тот или иной трансфер является оптимальным. Если вы будете выполнять оптимизацию много раз на слегка изменяющихся данных, вы сможете увидеть все интересные варианты трансферов.

	Total	⚽	⚽	⚽	Move	Score	Lineup xP - Hit
☐ X	24	19	4	1	Roll	288.09 ₣ 15.57	530.14 ₣ 22.72
☐ X	7	2	4	1	Brownhill → Martinelli	289.35 ₣ 21.88	534.16 ₣ 30.29
☐ X	5	-	2	3	Davis → Jiménez	287.99 ₣ 12.28	532.95 ₣ 7.74
☐ X	4	-	1	3	Brownhill → Saka	284.84 ₣ 8.28	525.21 ₣ 11.79
☐ X	3	-	3	-	Ramsdale → de Gea	283.71 ₣ 12.56	526.58 ₣ 23.73
☐ X	3	-	2	1	Brownhill → Lanzini	288.97 ₣ 15.63	530.59 ₣ 14.15
☐ X	3	-	1	2	Brownhill → Raphinha	283.74 ₣ 11.11	527.28 ₣ 19.60
☐ X	3	-	1	2	Brownhill → Gray	283.77 ₣ 8.17	526.69 ₣ 22.45

Рис. 13. Анализ чувствительности вариантов трансфера, предложенных оптимизационной моделью

Оптимизация на основе имитационного моделирования – еще один метод, который я нахожу очень многообещающим. Мы начали делать первые шаги в этом направлении на [FPLOptimized](#). Это дает лучшее понимание контекст борьбы в мини-лигах или на высших уровнях общего рейтинга. Моделирование особенно полезно для анализа эффективного владения (EO) и того, как оно может влиять на нашу оптимальную стратегию.

Часть 6: Современное состояние и будущее (технологии ИИ)

Я часто обсуждаю, насколько полезны эти методы. Некоторые утверждают, что они никогда не смогут превзойти топ-менеджеров, в то время как другие с подозрением относятся к качеству моделей. Я понимаю эти опасения, но методы анализа приведены здесь не для того, чтобы заставить вас играть как робота. Есть множество мест, где менеджеры FPL могут превзойти модели. Например, вы могли бы предсказать состав МанСити, что позволит вам опередить других в выборе «правильного» игрока.

Методы аналитики предназначены для того, чтобы упростить вам процесс принятия решений. Вот почему мы называем их *Системами поддержки принятия решений*. Вместо того, чтобы делать FPL скучным для вас, эти инструменты сузят варианты трансферов до разумного числа, предоставив вам окончательный выбор.

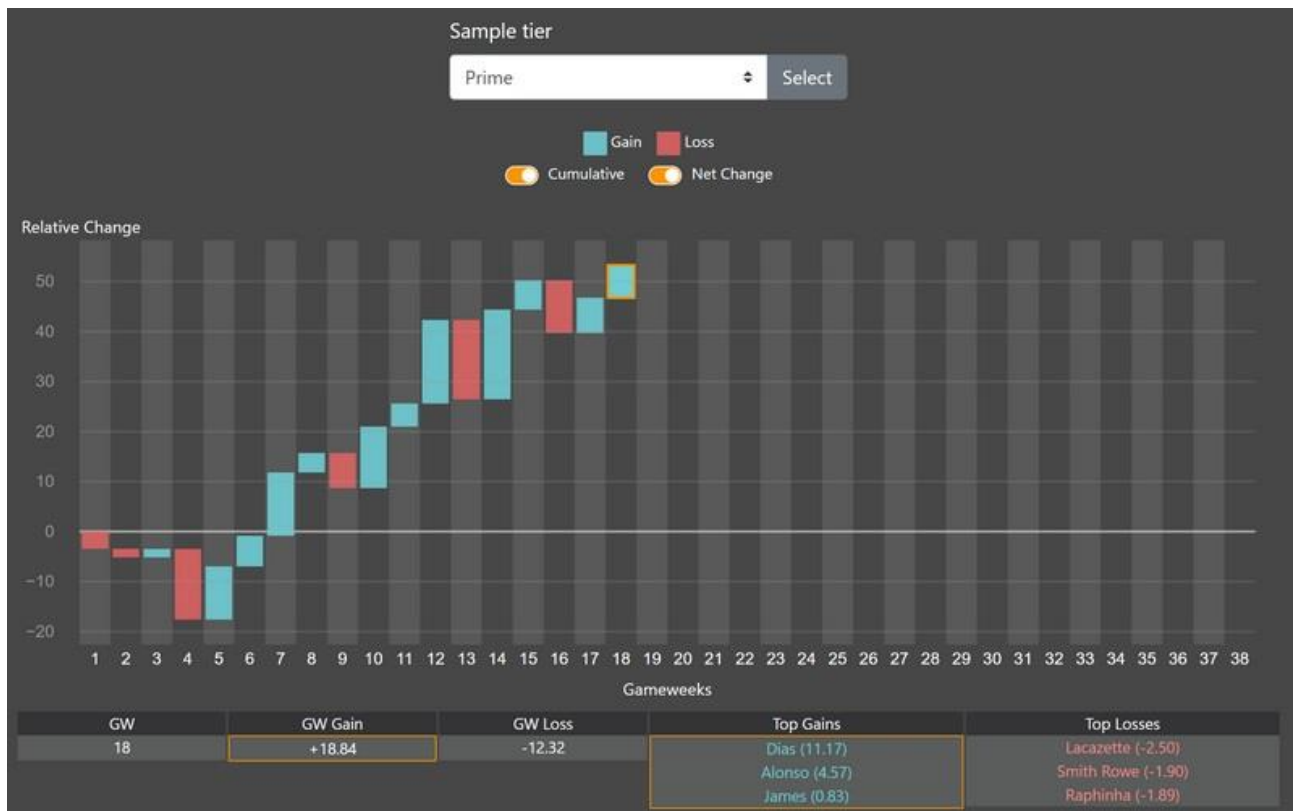


Рис. 14. Инструменты помогут проанализировать прошлое, прогнозировать будущее и оптимизировать решения

Я твердо верю, что инструменты аналитики позволят вам легко превзойти средний показатель по лиге. Одержимый менеджер FPL, который отслеживает каждую новость и тратит много времени на FPL, может победить полностью автоматизированного бота FPL. Однако достаточно хороший менеджер, использующий аналитические подходы, чаще оказывается на вершине в долгосрочной перспективе.

Если вы опытный специалист по обработке данных, впервые играющий в FPL, или менеджер FPL с многолетним стажем, никогда не использовавший аналитические подходы, подумайте о том, чтобы присоединиться к нам [FPL Analytics Discord Server](#) для обсуждения аналитики в FPL.

Часть 7: Резюме. Аналитика FPL в двух словах

Не используйте текущую статистику для прогнозирования будущих результатов. Используйте prxG и xA с сайта [fbref.com](#).

Не используйте FDR с сайта FPL. Используйте более тонкий анализ: [538](#), [FPLOptimized](#), [Elevenify](#), [Wee Rogue's PL model](#).

Результаты являются вероятностными и следуют простому (но зашумленному) распределению. Коэффициенты ставок являются лучшим показателем будущей доходности. [EV Calculator](#).

Для прогнозов более чем на 1GW используйте бесплатную / премиум модель, или разработайте собственную. Список доступных моделей прогнозирования можно найти на сервере Discord.

Используйте оптимизацию на основе ожидаемых результатов, чтобы минимизировать число трансферов, которое вы будете оценивать в процессе принятия решения:

- Изучите видеопособия по [Excel](#) и [Python](#) для разработки своих моделей.
- Воспользуйтесь [FPLNathan's FPL Solver](#), если вам нужно что-то попроще, чтобы использовать прямо сейчас.
- Я также регулярно пишу и делюсь советами о моделях оптимизации в Twitter.

Знайте, что существует множество аналитических инструментов, и их количество растет с каждым днем. Ознакомьтесь с тем, что доступно.

Хотите узнать больше или у вас есть дополнительные вопросы? У нас есть сервер Discord, где можно общаться и задавать вопросы: [FPL Analytics Discord server](#), [Analytics xP League](#).

Каждая часть этой заметки подготовлена совместно со следующими замечательными людьми:

- Part 1: @FPLSimpson and @dilksybab
- Part 2: @wee_rogue
- Part 3: @FPLSigurd
- Part 4: @theFPLkiwi 🌀
- Part 5: @FF_Trout
- Part 6: Allan Dann Under