

Проект по оценке важности статистических показателей при подсчёте стоимости полевых игроков топ-5 футбольных лиг Европы

Выполнил: Роберт Сулоев Амарович*

December 2024

*Отдельное спасибо Савиной Анастасии. Изначально мы были в одной группе, и она уже помогла мне найти некоторые ссылки на источники, однако на последующих этапах она поняла, что ей тяжело даётся футбольная тема, и мы разделились.

1 Цели и задачи проекта

1.1 Цели

- Определить и проинтерпретировать значимость статистических показателей в стоимости футболистов.
- Проверить ряд гипотез.

1.2 Задачи

- Собрать начальные данные по стоимости и игровой статистике, а также национальности и возрасту.
- Провести EDA и, возможно, выявить на начальном этапе какие-то интересные закономерности.
- Построить модель для оценки коэффициентов, а также для их последующей интерпретации.

1.3 Актуальность

Довольно простая: я просто играю в футбол с детства и люблю эту игру, в вышкинской футбольной лиге играю, и мне просто интересна эта тема.

1.4 Описание данных

Данные были собраны вручную с помощью Python. Код для парсинга можно найти в репозитории в ветке "парсинг данных готовые таблицы Excel для удобства восприятия лежат там же.

- **Transfermarkt**: Первый источник для данных по стоимости и остальных важных данных, таких как национальность, позиция и др. Transfermarkt.
- **FBref**: Вторым источником данных выступил сайт FBref, на котором я собрал основные игровые статистики для полевых игроков FBref.

Данные собирались непосредственно с источника FBref совместно с ОРТА. Так как статистика за сезон 2023/2024, данные обновлялись в начале лета 2024 года.

Конечное количество наблюдений, которое я буду анализировать, составляет 297.

Статья 1 и статья 2 , на которые можно было ориентироваться при проведении работы

2 Экономическая модель

2.1 Описание данных

- **Стоимость футболистов** — измерена в млн евро. В моей модели выступает объясняемая переменная (TARGET).
- **Возраст игрока** — в годах.
- **Национальность** — измерена в национальностях.
- **Позиция на поле**
- **Голы** — в штуках.
- **Ассисты (пас, после которого был забит гол)** — в штуках.
- **Голы с игры** — голы за вычетом голов, забитых с пенальти — в штуках.
- **Желтые карточки** — количество желтых карточек (предупреждений).
- **Красные карточки** — количество удалений.
- **Заработанные пенальти** — количество заработанных пенальти.
- **Ожидаемые голы** — количество ожидаемых голов, метрика рассчитывается примерно как описано в статье.
- **Ожидаемые ассисты** — в штуках.
- **Прогрессивные переводы** — количество в штуках.
- **Прогрессивные пасы** — в штуках.
- **Принятые игроком прогрессивные пасы** — в штуках.
- **Попытки отбора** — в штуках.
- **Успешные попытки отбора** — в штуках.

- **Количество остановок дриблинга + неудачных попыток остановить дриблинг** — в штуках.
- **Неудачные попытки остановки дриблинга** — в штуках.
- **Ошибки, приводящие к ударам по воротам** — в штуках.
- **Выносы из штрафной** — в штуках.
- **Прерванные пасы** — в штуках.
- **Прерванные удары** — в штуках.
- **Количество матчей** — в штуках (рассчитано как количество минут/90).

2.2 Выбор переменных

Я выбрал именно эти переменные, так как они описывают основные футбольные характеристики полевых игроков: как защитные, так и атакующие метрики. Конечно, эти данные не учитывают влияния, например, популярности игрока в медийном поле, однако позволяют оценить, как игра на поле влияет на стоимость игрока на трансферном рынке. Каждая из выбранных мной метрик часто обсуждается в контексте уровня игрока и важности для команды.

2.3 Предположения влияния переменных

Сначала перечислю переменные, которые, по моему мнению, будут *положительно влиять*:

- Голы
- Ассисты
- Голы с игры
- Заработанные пенальти
- Ожидаемые голы и ассисты
- Все прогрессивные метрики
- Успешные отборы
- Выносы из штрафной

- Прерванные удары и пасы
- Количество сыгранных матчей

Также предполагаю, что если игрок из английской Премьер-лиги, то это будет положительно влиять на его стоимость, а также если игрок англичанин или француз — это будет положительно сказываться на его стоимости.

Остальные исходные метрики, скорее всего, влияют на стоимость отрицательно.

Говорить, как влияет, например, португальское гражданство, не буду, так как здесь даже предположений нет.

2.4 Формулировка гипотез

1. **Нулевая гипотеза (H0):** Игроки в АПЛ оцениваются адекватно и их рыночная стоимость соответствует той пользе, которую они приносят на поле. Данная гипотеза пересекается с вот этой статьей.

Альтернативная гипотеза (H1): Игроки в АПЛ переоценены, их рыночная стоимость значительно превышает их реальную производительность на поле.

2. **Нулевая гипотеза (H0):** Атакующие игроки являются более ценными/важными игроками на трансферном рынке, нежели защитники и полузащитники.

Альтернативная гипотеза (H1): Атакующие игроки не являются более ценными на трансферном рынке по сравнению с защитниками и полузащитниками.

2.5 Результаты регрессионного анализа

Таблица 1: Общие результаты регрессионной модели

Dep. Variable:	Стоимость	R-squared:	0.776
Model:	OLS	Adj. R-squared:	0.593
Method:	Least Squares	F-statistic:	4.244
Date:	Wed, 04 Dec 2024	Prob (F-statistic):	4.41e-18
Time:	23:04:59	Log-Likelihood:	378.38
No. Observations:	297	AIC:	-488.8
Df Residuals:	163	BIC:	6.203
Df Model:	133		
Covariance Type:	nonrobust		

Таблица 2: Коэффициенты регрессии (Часть 1)

	coef	std err	t	P > t	[0.025	0.975]
const	-127.5751	67.006	-1.904	0.059	-259.803	4.653
Возраст игрока	13.8759	5.578	2.488	0.014	2.868	24.884
Голы	1.6164	0.600	2.695	0.008	0.433	2.800
Ассисты	0.1465	0.822	0.178	0.859	-1.476	1.769
Голы с игры	0.6272	0.508	1.236	0.218	-0.374	1.629
Желтые карточки	0.7257	0.513	1.416	0.159	-0.286	1.737
красные карточки	2.3406	3.410	0.686	0.493	-4.390	9.071
заработанные пенальти	0.9892	0.939	1.053	0.294	-0.864	2.842
Ожидаемые голы	1.4423	0.794	1.816	0.071	-0.125	3.010
Ожидаемые ассисты	-0.0356	1.165	-0.031	0.976	-2.335	2.264
Прогрессивные переводы	0.0806	0.071	1.128	0.261	-0.060	0.222
Прогрессивные пасы	0.1554	0.038	4.118	0.000	0.081	0.230
Принятые игроком прогрессивные пасы	0.0270	0.040	0.675	0.501	-0.052	0.106
Попытки отбора	0.0063	0.316	0.020	0.984	-0.617	0.630
успешные попытки отбора	0.5614	0.451	1.245	0.215	-0.328	1.451
количество остановок дриблинга+неудачных попыток остановить дриблинг	-0.2271	0.273	-0.832	0.406	-0.766	0.311
неудачные попытки остановки дриблинга	0.2280	0.352	0.648	0.518	-0.467	0.923
Ошибки приведшие к ударам по воротам	-0.3521	1.358	-0.259	0.796	-3.031	2.327
Выносы из штрафной	0.0253	0.081	0.312	0.755	-0.135	0.185
Прерванные пасы	-0.1766	0.219	-0.806	0.422	-0.609	0.256
Прерванные удары	0.6067	0.363	1.671	0.096	-0.110	1.323
колво матчей (колво сыгранных минут/90)	-1.3660	0.386	-3.541	0.001	-2.127	-0.605
Прочитанных передач	0.0022	0.186	0.012	0.991	-0.366	0.370
возраст ²	-0.3208	0.113	-2.846	0.005	-0.543	-0.098
log_ ожидаемые голы	0.1811	0.512	0.354	0.724	-0.829	1.191
log_ ожидаемые ассисты	-0.5099	0.827	-0.617	0.538	-2.141	1.122
log_ принятые прогрессивные пасы	-0.0342	1.093	-0.031	0.975	-2.191	2.123
log_ выносы из штрафной	-0.8665	0.433	-2.002	0.047	-1.721	-0.012
Национальность_Англия	7.1124	12.600	0.564	0.573	-17.752	31.976
Национальность_Аргентина	12.5434	12.950	0.969	0.334	-13.011	38.098
Национальность_Бельгия	0.0977	14.541	0.007	0.995	-28.597	28.793
Национальность_Бразилия	5.1334	12.678	0.405	0.686	-19.885	30.152
Национальность_Германия	6.5039	12.664	0.514	0.608	-18.486	31.494
Национальность_Грузия	9.5998	24.376	0.394	0.694	-38.503	57.703
Национальность_Ирландия	15.0181	17.623	0.852	0.395	-19.760	49.796
Национальность_Испания	8.9910	13.329	0.675	0.501	-17.312	35.294
Национальность_Италия	6.1189	13.576	0.451	0.653	-20.672	32.910
Национальность_Нидерланды	5.5770	12.911	0.432	0.666	-19.900	31.054
Национальность_Польша	-11.0663	17.706	-0.625	0.533	-46.007	23.874
Национальность_Португалия	9.4548	13.539	0.698	0.486	-17.263	36.173
Национальность_Россия	4.5145	22.794	0.198	0.843	-40.468	49.496
Национальность_Румыния	-4.5449	26.627	-0.171	0.865	-57.090	48.000
Национальность_Сербия	6.4940	36.474	0.178	0.859	-65.484	78.472
Национальность_Словакия	15.5170	23.790	0.652	0.515	-31.430	62.464
Национальность_Турция	-27.7308	19.018	-1.458	0.147	-65.260	9.799
Национальность_Украина	5.0142	19.458	0.258	0.797	-33.384	43.413
Национальность_Уэльс	14.9656	22.362	0.669	0.504	-29.164	59.095
Национальность_Франция	10.8470	12.650	0.857	0.392	-14.117	35.811
Национальность_Чехия	15.2851	21.654	0.706	0.481	-27.446	58.016
Национальность_Швейцария	-0.6098	18.736	-0.033	0.974	-37.584	36.364
Национальность_Шотландия	0.5876	15.135	0.039	0.969	-29.280	30.455
Национальность_Япония	7.4693	15.514	0.481	0.631	-23.146	38.084
Позиция на поле_Левый Вингер	-2.2052	5.521	-0.399	0.690	-13.099	8.689
Позиция на поле_Левый защитник	-0.2983	6.400	-0.047	0.963	-12.927	12.330
Позиция на поле_Опорный полузащитник	5.8125	6.513	0.892	0.373	-7.040	18.665
Позиция на поле_Оттянутый нап.	-28.2516	14.546	-1.942	0.054	-56.957	0.454
Позиция на поле_Правый Вингер	-5.3166	5.365	-0.991	0.323	-15.903	5.270
Позиция на поле_Правый защитник	-2.8342	6.833	-0.415	0.679	-16.319	10.650
Позиция на поле_Правый полузащитник	0.2832	14.019	0.020	0.984	-27.381	27.947
Позиция на поле_Центр. защитник	3.8607	7.489	0.515	0.607	-10.919	18.640
Позиция на поле_Центр. полузащитник	-2.2717	5.411	-0.420	0.675	-12.950	8.407
Позиция на поле_Центральный нап.	-7.3539	6.256	-1.175	0.241	-19.700	4.992
Лига_АПЛ	1.7150	4.662	0.368	0.713	-7.486	10.916
Лига_Бундеслига	-3.1863	12.277	-0.260	0.796	-27.413	21.040
Лига_Ла лига	4.5070	5.460	0.825	0.410	-6.269	15.283
Лига_Лига 1	-10.7035	5.791	-1.848	0.066	-22.131	0.724
Лига_Серия А	-2.9765	6.150	-0.484	0.629	-15.112	9.159

Таблица 3: Коэффициенты регрессии (Часть 2)

	coef	std err	t	P > t	[0.025	0.975]
Клуб_Аль-Наср Эр-Рияд	-16.9210	17.579	-0.963	0.337	-51.611	17.769
Клуб_Аль-Хилаль Эр-Рияд	-3.1883	17.914	-0.178	0.859	-38.539	32.162
Клуб_Арсенал	21.7832	5.935	3.670	0.000	10.072	33.495
Клуб_Астон Вилла	-0.0426	5.528	-0.008	0.994	-10.951	10.865
Клуб_Аталанта	-6.5656	8.928	-0.735	0.463	-24.184	11.053
Клуб_Атлетик Бильбао	-18.3375	12.722	-1.441	0.151	-43.443	6.768
Клуб_Атлетико Мадрид	-9.2150	6.833	-1.349	0.179	-22.699	4.269
Клуб_Бавария Мюнхен	12.4818	13.809	0.904	0.367	-14.768	39.731
Клуб_Байер 04 Леверкузен	-7.4163	14.429	-0.514	0.608	-35.890	21.057
Клуб_Барселона	13.5737	6.956	1.951	0.053	-0.152	27.300
Клуб_Бетис	-9.5688	23.365	-0.410	0.683	-55.678	36.540
Клуб_Болонья	-4.5058	14.224	-0.317	0.752	-32.576	23.564
Клуб_Борнмут	-20.7794	9.434	-2.203	0.029	-39.396	-2.163
Клуб_Боруссия Дортмунд	-11.3796	13.878	-0.820	0.413	-38.766	16.007
Клуб_Брайтон энд Хоув Альбион	-6.6045	5.919	-1.116	0.266	-18.285	5.076
Клуб_Бrentфорд	-5.2116	9.141	-0.570	0.569	-23.250	12.827
Клуб_Валенсия	-12.5017	21.175	-0.590	0.556	-54.289	29.285
Клуб_Вест Хэм Юнайтед	-2.4385	8.770	-0.278	0.781	-19.745	14.868
Клуб_Вулверхэмптон Уондерерс	-8.6834	10.502	-0.827	0.409	-29.408	12.042
Клуб_Галатасарай	21.2455	16.808	1.264	0.208	-11.924	54.415
Клуб_Жирона	-18.2353	13.002	-1.403	0.163	-43.893	7.422
Клуб_Интернационале	16.1586	7.356	2.197	0.029	1.643	30.674
Клуб_Ипсвич Таун	-16.4355	22.958	-0.716	0.475	-61.739	28.868
Клуб_Иттихад	15.5940	17.456	0.893	0.373	-18.852	50.041
Клуб_Кристал Пэлас	2.0401	6.896	0.296	0.768	-11.569	15.649
Клуб_Ланс	-8.5690	11.646	-0.736	0.463	-31.552	14.414
Клуб_Лацио	-0.9001	16.245	-0.055	0.956	-32.958	31.158
Клуб_Ливерпуль	-0.6238	5.019	-0.124	0.901	-10.528	9.280
Клуб_Лилль	-8.8899	10.427	-0.853	0.395	-29.467	11.687
Клуб_Манчестер Сити	32.1941	5.685	5.663	0.000	20.976	43.412
Клуб_Манчестер Юнайтед	8.2873	5.383	1.539	0.125	-2.336	18.910
Клуб_Милан	-4.2449	7.431	-0.571	0.569	-18.909	10.419
Клуб_Монако	-9.7296	8.688	-1.120	0.264	-26.875	7.416
Клуб_Наполи	7.8370	10.514	0.745	0.457	-12.911	28.585
Клуб_Ноттингем Форест	-5.8469	7.664	-0.763	0.447	-20.971	9.277
Клуб_Ньюкасл Юнайтед	-4.4415	5.665	-0.784	0.434	-15.621	6.738
Клуб_Олимпик Марсель	-11.2815	12.448	-0.906	0.366	-35.847	13.284
Клуб_Пари Сен-Жермен	16.4851	6.408	2.573	0.011	3.840	29.130
Клуб_РБ Лейпциг	-3.8147	14.823	-0.257	0.797	-33.067	25.438
Клуб_Реал Мадрид	30.8744	7.217	4.278	0.000	16.632	45.117
Клуб_Реал Сосьедад	0.1132	8.734	0.013	0.990	-17.123	17.350
Клуб_Рома	-24.3617	12.452	-1.956	0.052	-48.935	0.212
Клуб_Стад Ренн	-5.4303	16.910	-0.321	0.748	-38.799	27.939
Клуб_Страсбур	-11.0478	17.315	-0.638	0.524	-45.216	23.121
Клуб_Торино	-6.6062	16.161	-0.409	0.683	-38.499	25.286
Клуб_Тоттенхэм Хотспур	-6.9877	5.823	-1.200	0.232	-18.479	4.504
Клуб_Фенербахче	-1.1739	12.702	-0.092	0.926	-26.240	23.892
Клуб_Фиорентина	-3.0514	16.456	-0.185	0.853	-35.526	29.423
Клуб_Фламенго	2.7769	17.261	0.161	0.872	-31.286	36.840
Клуб_Фулхэм	-4.8977	7.987	-0.613	0.541	-20.659	10.863
Клуб_Челси	-2.0227	4.597	-0.440	0.661	-11.095	7.050
Клуб_Штуттгарт	-21.2792	12.578	-1.692	0.092	-46.101	3.542
Клуб_Эвертон	-5.0679	7.753	-0.654	0.514	-20.368	10.232
Клуб_Ювентус	-7.6637	7.896	-0.971	0.333	-23.246	7.918
Национальность_Скандинав	12.6321	13.630	0.927	0.355	-14.264	39.528
Национальность_Северная_Америка	7.5579	13.884	0.544	0.587	-19.841	34.957
Южная_Америка	10.3222	13.173	0.784	0.434	-15.674	36.318
Африканец	-1.7890	12.914	-0.139	0.890	-27.274	23.696

3 Основные результаты и выводы

3.1 Метрики и результаты модели

R^2 получился достаточно высокий ($R^2 = 0.776$), что указывает на то, что модель объясняет значительную часть вариации зависимой переменной **Стоимость**.

F-статистика в 4.2 также свидетельствует о значимости нашей получившейся модели

3.2 Определим наиболее значимые коэффициенты, на разных уровнях значимости

Наиболее значимые коэффициенты P-Value < 0.001 (***)

- Прогрессивные пасы
- Клуб Арсенал
- клуб Манчестер Сити
- Реал Мадрид
- колво матчей

Коэффициенты, значимые на уровне P-Value < 0.01 (**)

- *Возраст*²
- *Голы*

Коэффициенты, значимые на уровне P-Value < 0.05 (*)

- log
- Клуб Борнмут
- Клуб Интернационале
- Клуб ПСЖ

Коэффициент значим на уровне $P\text{-value} < 0.1$

- Ожидаемые голы
- Оттянутый нападающий
- Лига 1
- Барселона
- Рома
- Штутгарт

3.3 Интерпритация основных коэффициентов

Начнем с интерпритации чисто игровых показателей и их коэффициентов

- **Каждый дополнительный гол в сезоне в национальной лиге добавляет 1.6 миллионов евро к оценочной стоимости игрока.** Это хорошо сочетается с общим представлением о футболе . В целом, считается , что самые дорогие игроки -это атакующие игроки с наибольшим количеством голов , либо атакующие игроки с самыми важными голами (Победные голы в топовых матчах)
- **Каждый ожидаемый гол прибавляет 1.44 миллиона евро к стоимости.** Ожидаемые голы важный показатель для нападающих, так как реализация голов в определенной степени удача, а вот выбор позиции и грамотные предупредительные действия показывают высокий уровень нападающего
- **Каждые 90 минут на поле забирают 1.366 миллиона евро от стоимости игрока.** Тяжело сказать почему так. Ведь топовые игроки проводят максимально возможное количество матчей . Возможно, модель оценивает
- **Каждый прогрессивный пас поднимает цену игрока на 0.155 миллионов долларов.** Прогрессивные пасы важная метрика для игроков, которые создают. Созиданием в футболе называются игровые действия, приводящие к опасным моментам у вражеских ворот.

- **При увеличении выносов из штрафной на процент , стоимость игрока уменьшается на 0.00867.** Тяжело объяснить почему так происходит , ведь вынос из штрафной для любого футболиста хороший показатель. Возможно, это дает минус к стоимости, хоть и не значительный , так как большое количество выносов в основном у защитников , однако и высокие нападающие , часто делают выносы из штрафной . И так как в основном защитники дешевле , чем атакующие игроки , выносы задают минус в стоимость

Попробуем проинтерпритировать остальные показатели

- **Каждый год жизни прибавляет к стоимости игрока 13.88 миллионов евро.** Тут есть некоторая странность в интерпретации, так как очевидно, что в определенном возрасте игрок выходит на свой пик , после чего его физические способности начинают падать. То есть скорее всего у нас получится квадратичная зависимость, о ней в следующем пункте
- **Убывающее Влияние с Увеличением Возраста: Отрицательный коэффициент (-0.32) на квадратичный член означает, что по мере роста возраста влияние возраста на стоимость становится всё более негативным.** Это указывает на то, что после определённого возраста стоимость футболиста начинает снижаться, и скорость этого снижения увеличивается с каждым годом.
- **Если игрок является игроком клуба Арсенал, Мадридский Реал или Манчестер Сити , то факт его пребывания в этом клубе прибавляет к его стоимости 21.78, 32.19, 30,87 миллионов евро соответственно.** Это вполне логично, так как в прошлом сезоне это были 3 клуба - главных претендента на Лигу Чемпионов , а так же 3 клуба с самыми дорогими составами . Эти клубы либо берут готовых звезд, стоимость которых уже велика, либо берут молодых игроков и делают из них звезд, то есть дорогих игроков .
- **Если игрок играет в Лиге 1 (Французская Высшая Лига) это убавляет от его стоимости 10.7 миллионов евро.** Французская лига считается слабейшей из топ-5 лиги европы (во время летнего трансферного окна, французская лига даже на некоторое время официально выпала из списка топ-5 лиг, уступив свое место Нидерландской Эредивизи) Во французской лиге есть только один клуб,

постоянно претендующий на ЛЧ. Остальные клубы же, по меркам главного турнира Европы, считаются слабыми. Так же французская лига достаточно жесткая и травмоопасная, поэтому элитного уровня игроки не особо стремятся туда. Французскую лигу в футбольном сообществе даже пренебрежительно называют "Лигой фермеров"

- **Клубы предтопы-Интернационале, ПСЖ, Барселона :** играя в этих командах, игроки прибавляют в стоимости 16,16 миллионов евро, 16,5 миллионов евро и 13,6 миллионов евро соответственно. Барселона и Интернационале — это клубы с выстроенной игровой системой, однако проблемой с финансами, поэтому они не могут позволить большие трансферы и покупать готовых больших игроков. Интернационале строит свою систему на покупке уже готовых топовых игроков, однако в более позднем возрасте, так сказать "Звезд ветеранов" то есть игроков уже на спаде их карьеры. Барселона же наоборот, берет очень молодых игроков из своей академии и выращивает из них топовых, зрелых игроков. ПСЖ же до нынешнего сезона строил свою модель на том, что просто покупал больших игроков с большой наценкой и предлагал им большую зарплату.
- **Средние клубы -Штутгарт, Рома, Борнмут :** это клубы, которые в прошлом сезоне сильно сдали в своих чемпионатах. Игроки этих клубов стоят на 21.27, 24.36, 20.78 дешевле.
- **Если игрок играет на позиции оттянутого нападающего — это снижает его стоимость на 28.3 миллиона евро.**

3.4 Итоги тестов для гипотез

Первая гипотеза о переоценненности игроков АПЛ **Отвергается** при любых адекватных уровнях значимости. Предположение о переоценке игроков АПЛ было сформулировано на основе того, что английские клубы часто сильно переплачивают за игроков. Наша гипотеза отверглась скорее всего потому, что мы брали как показатель именно оценочную стоимость от трансфермаркета, а не реальную стоимость перехода в клуб АПЛ. Как раз таки мнение о переплате за игроков существует на основе того, что клубы из АПЛ покупают игроков дороже их оценочной стоимости. Но благодаря этому тесту мы выяснили, что трансфермаркет не переоценивает стоимость игроков из АПЛ.

Вторая гипотезу **подтверждается частично**. Рельные статистически значимые различия между группами позиций наблюдаются только между защитниками и атакующими игроками. Статистически значимых различий между защитниками и полузащитниками, а также полузащитниками и атакующими не наблюдается