**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**

**(РУТ (МИИТ)**

Институт/факультет «Академия “Высшая Инженерная Школа”, АВИШ»

Специальность/Направление подготовки Информатика и вычислительная техника

**КУРСОВАЯ РАБОТА**

**по дисциплине:** Технологии хранения больших данных

|  |  |
| --- | --- |
| **на тему:** | “Аналитика параметров спортсменов с помощью методов Big Data” |

|  |  |
| --- | --- |
| **Студент группы ШАД 212** | **(Воронин А.А. )** |
| **Научный руководитель** | **(Зуев И.А. )** |

**Москва 2023 г.**

Оглавление

[Основные термины и сокращения 3](#_Toc153559847)

[Введение 4](#_Toc153559848)

[Исследование данных 5](#_Toc153559849)

[Заключение 17](#_Toc153559850)

[Список использованных источников и программ 18](#_Toc153559851)

[Приложение 19](#_Toc153559852)

# Основные термины и сокращения

ФИФА (федеральная интернациональная футбольная ассоциация) — главная футбольная организация, являющаяся крупнейшим международным руководящим органом в футболе, футзале и пляжном футболе.

Дриблинг – мастерство ведения мяча.

ИМТ (индекс массы тела) — это числовой показатель, который используется для оценки соотношения между массой тела и ростом человека. Он является одним из методов оценки нормальности или избыточности веса.

# Введение

Современный спорт становится все более зависимым от науки и технологий. Анализ данных играет большую роль в оценке способностей, выявлении сильных и слабых сторон спортсменов. Целью данной курсовой работы является детальное изучение параметров и ключевых факторов, влияющих на успех при участии в соревнованиях.

В ходе работы будет проведена аналитика игровых и физических показателей. Был выбран набор данных профессиональных футболистов ФИФА со всего мира. В нем есть множество атрибутов, относящихся к каждому игроку, что делает его ценным ресурсом для различных анализов. Для проведения исследования были выбраны такие характеристики спортсмена, как страна, клуб, рост, вес, возраст, а также основные параметры и показатели: реакция, выносливость, скорость бега, дриблинг, хладнокровность (самообладание игрока под давлением), сила удара мяча, окончание момента или ситуации (навык игрока в добивании мяча), уровень агрессии, процент успешного выполнения пенальти, уровень ловкости маневрирования.

С помощью результатов исследования удастся найти ключевые факторы, которые прямо или косвенно влияют на проведение спортивных мероприятий. Такую информацию можно использовать как для совершенствования, оттачивания навыков и набора сильной команды, так для оценки последующих исходов матчей.

# Исследование данных

Для проведения анализа был взят набор данных профессиональных футболистов за 2023 год. Также в качестве аналитик было выбрано построение визуализации данных в виде графиков. Опишем эти графики:

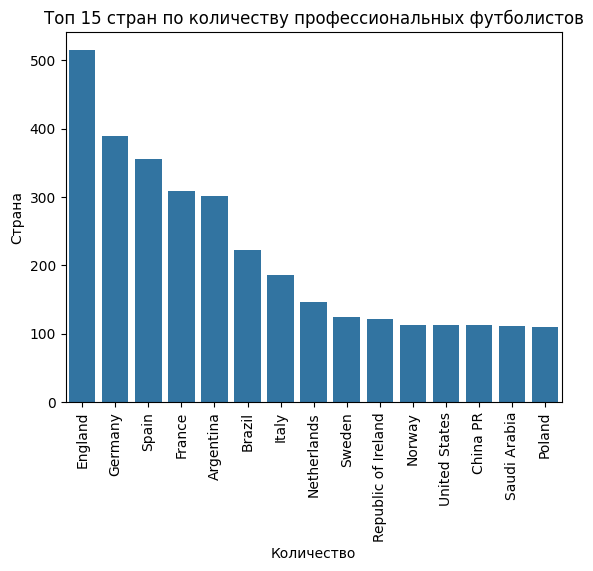


Рис. 1

Видно, что в Англии, Германии, Испании, Франции и Аргентине наблюдается наибольшее количество профессиональных футболистов. Это может означать, что именно в этих странах уделяется достаточно внимания к организации и подходу тренировок, а также футбол является одной из любимых игр.

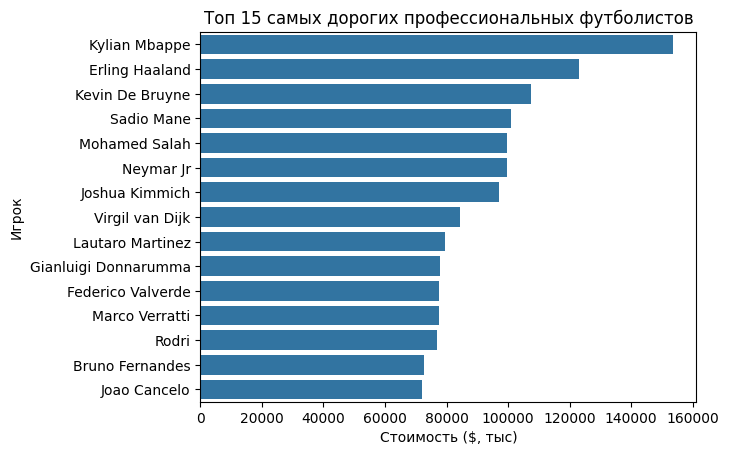


Рис. 2

Стоимость футболиста определяется за счёт его физических показателей, игровых навыков и результатов на играх. Поэтому было решено взять самых дорогих футболистов, чтобы детальнее взглянуть на отличия и сопоставить между высоко оцениваемыми и низко оцениваемыми игроками.

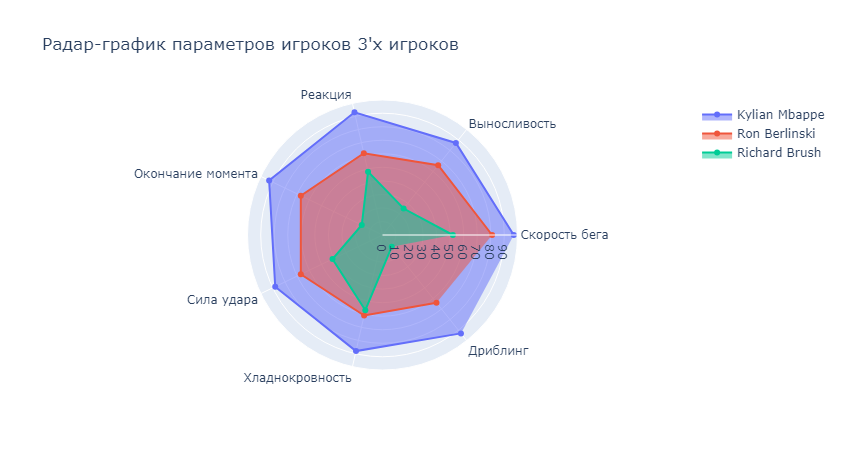


Рис. 3

Было выбрано 7 параметров, они оказались основополагающими среди остальных, и 3 футболиста: самый дорогостоящий – Килиан Мбаппе, средний, имеющий медианное значение – Рон Берлински и самый низко стоящий – Ричард Бруш. Видно, что такие показатели как реакция, выносливость, скорость бега, дриблинг, хладнокровность, сила удара и окончание момента (добивание мяча) у Килиана почти достигают максимальных значений, в то время как у Ричарда едва доходят до половины. У Рона неплохие характеристики. Учтем эти параметры и в дальнейшем рассмотрим их.

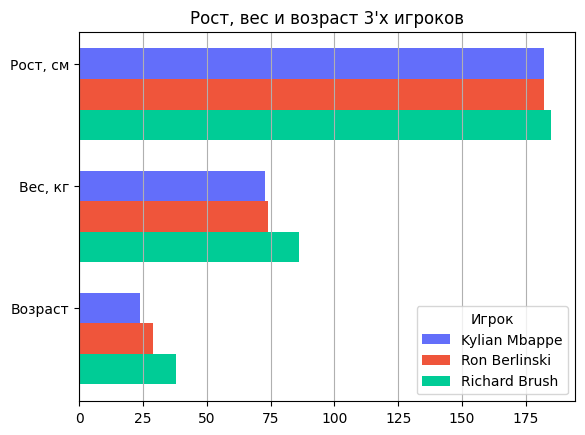


Рис. 4

Заметим, что показатели рост и вес у этих игроков практические близки, а возраста отличаются. Килиан – самый молодой среди них, далее идёт Рон, а затем Ричард. Исходя из рис. (4), можно сделать вывод, что у Бруша многие параметры с возрастом стали слабеть. У Берлинского также замечается подобное явление. Возраст оказывает сильное влияние на многие показатели, это также будем учитывать дальше.



Рис. 5

В основном у профессиональных футболистов рост расположен в диапазоне от 180 до 190 и от 170 до 180 (см). Однако для позиции вратаря лучше всего выбирать высоких игроков, в основном такие попадают в диапазон от 190 до 200 (см).

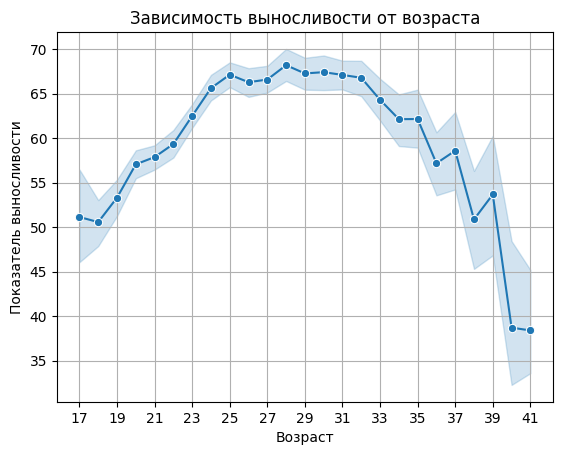


Рис. 6

Наилучшая выносливость спортсменов наблюдается в промежутке от 25 до 32 лет. До 25 лет этот показатель растёт, а после 32 выносливость стремительно падает.

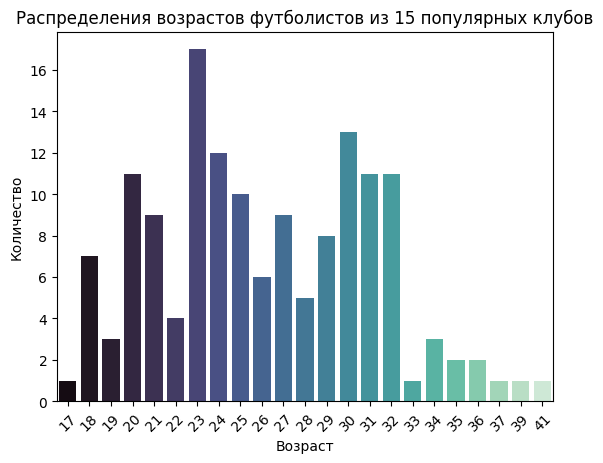


Рис. 7

Было выбрано 15 самых популярных клубов, и по ним построить как распределяется возраст. Видно, что в основном в клубах собраны молодые игроки от 23 до 32 лет.

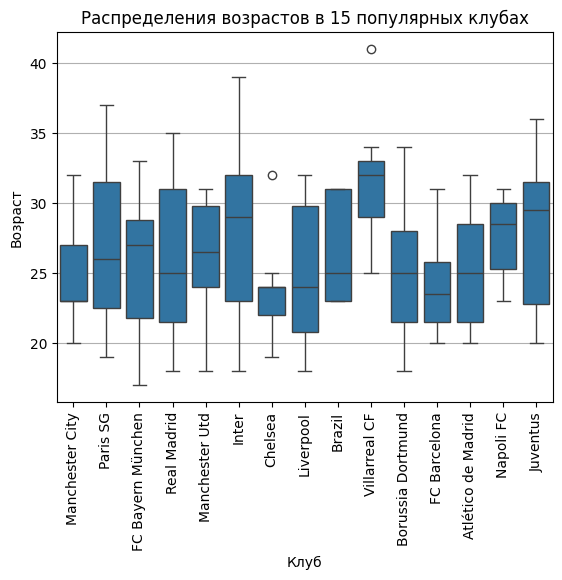


Рис. 8

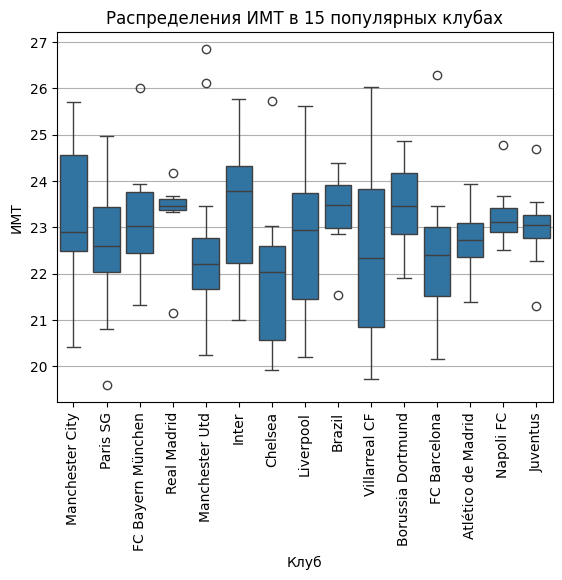


Рис. 9

В 15 популярных клубах у большинства игроков ИМТ находится в норме, однако есть футболисты, с избыточной массой тела, но их лучше не учитывать.

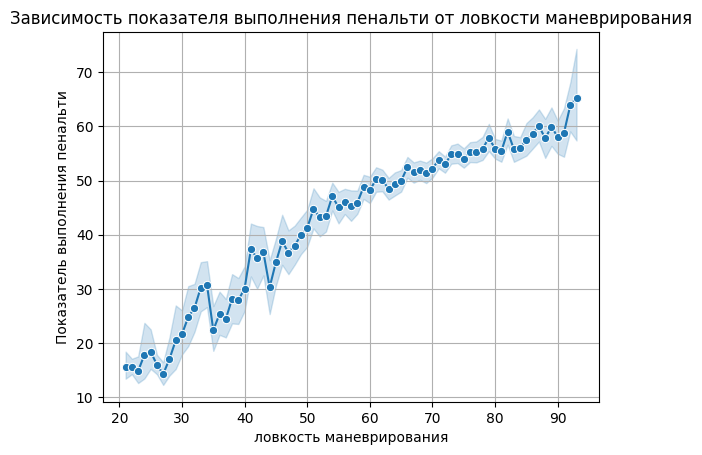


Рис. 10

Чем ловчее игрок, тем он успешнее выполняет пенальти, так он больше может запутать вратаря и совершить корректный удар в футбольные ворота.

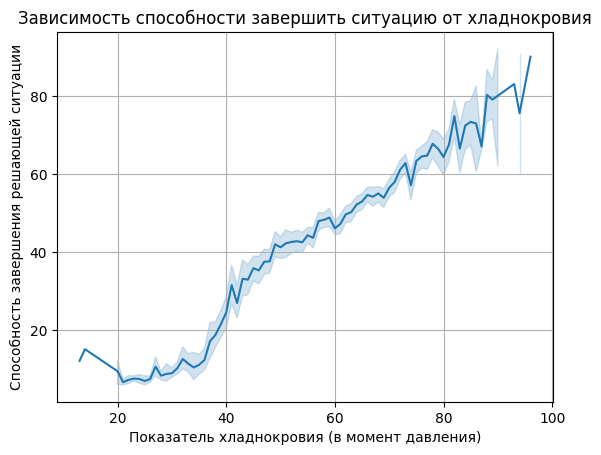


Рис. 11

Чем игрок хладнокровнее, спокойнее относится к ситуации, тем он лучше может завершить решающий момент, например, отдать важную передачу или попытаться забить в футбольные ворота соперника более точно.



Рис. 12

В основном показатель агрессии у всех одинаковый, но начиная с веса 80 кг, он сильно скачет, а затем стремительно растет. За счёт своей массы, строения скелета и мышц, игрок может себя вести чересчур агрессивно к другим игрокам: толкать, делать опасные подкаты, вызывать конфликты – это очень плохо может сказаться на игре и репутации команды.

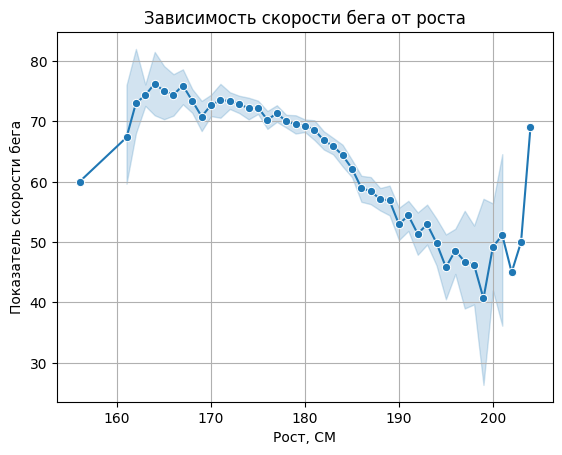


Рис. 13

Как оказалось, не очень высокие спортсмены быстрее бегают. Также наблюдается, что при увеличении роста, уменьшается показатель скорости бега. При росте от 170 до 180 (см) достигается наиболее равномерный показатель, который подходит нам при выборе игроков.

# Заключение

Проведя аналитику параметров спортсменов, удалось выяснить, какие факторы влияют на построение сильной команды и успешное участие в спортивных мероприятиях. В ходе работы была проведена аналитика игровых и физических показателей по выбранному набору данных.

1. Удалось выделить 5 стран: Англия, Германия, Испания, Франция и Аргентина, в которых наблюдается большое количество профессиональных футболистов, и из которых стоило бы рассматривать игроков в команду.
2. Если футболист хочет, чтобы его стоимость высоко оценивали, он должен прилагать много усилий для игровых и физических тренировок. Однако с возрастом совершенствоваться становится тяжелее. Самый лучший возраст игрока находится в диапазоне от 23 до 32 лет.
3. Среди профессиональных игроков в футбол получили 3 основных диапазона ростов, от 170-180, 180-190 и 190-200 (см)
4. Чем выше рост игрока, тем он медленнее бегает, поэтому игроков с ростом 190-200 (см) лучше всего ставить на позицию вратаря.
5. Возраст играет большую роль в физических параметрах футболиста, так, например, молодые игроки являются самыми выносливыми и обладают лучшей реакцией.
6. При выборе футболиста нужно посмотреть на его ИМТ и отдельно массу тела. Игроки с избыточной или высокой массой тела склонны чересчур агрессивно себя вести на поле. Из-за этого могут возникать лишние конфликты и удаления, что сильно может оказать отрицательное влияние на результат матча и репутацию команды.
7. Необходимо отдельно уделить внимание психологическому состоянию игрока. Излишняя нервность, неспокойность плохо сказывается на итоговом решении в завершении ситуации. Чем хладнокровнее игрок относится к выбору, тем лучше у него выходит реализовать важный момент.
8. Чтобы футболист исполнил пенальти с наилучшим исходом, забил мяч в ворота соперника, нужно уделить достаточное время тренировка маневрирования. Игрок должен быть очень ловким, тогда он сможет лучше всего запутать вратаря.

Поставленная цель была достигнута. Полученную информацию можно использовать для подбора игроков, совершенствования навыков спортсменов и оценивания исхода матча.

# Список использованных источников и программ

1. Набор данных с сайта Kaggle –<https://www.kaggle.com/datasets/rehandl23/fifa-24-player-stats-dataset>
2. Библиотеки для проведения анализа и построения аналитик: numpy, pandas, fuzzywuzzy, matplotlib, seaborn, plotty, wordcloud.

# Приложение

Файл с кодом – <https://github.com/jsinkx/bigdata-player-stats-coursework>