**Документация**

**Дипломная работа по теме:**

**«Анализ данных футбольных событий (поиск инсайтов, составление рекомендаций стейкхолдерам)».**

**Профессия “Аналитик данных”, DA-117**

**Плотников Никита Викторович**

**г. Екатеринбург, 2025 г.**

**Содержание**

**Введение…………………………………………………………………………………………………………………..…….3**

**Блок 1. Описание исходного датасета и типов данных…………………………………..………….…4**

**Блок 2. Подготовка, очистка и преобразование данных ………………………………………….….8**

**Блок 3. Анализ данных для стейкхолдеров……………………………………………………………..……9**

**Блок 3.1 Оценка влияния пенальти на результат матча………………………………………………10**

**Блок 3.2 Рейтинг игроков по эффективности исполнения пенальти…………………………..13**

## **Блок 3.3 Индивидуальная статистика игрока………………………………………………………..…….14**

**Блок 3.4 Оценка влияния ударной ноги футболиста на качество исполнения пенальти………………………………………………………………………………………………………………………..15**

**Блок 3.5 Статистика распределения по зонам ворот ударов при пенальти левшей и правшей…………………………………………………………………………………………………………………..……16**

**Итоги проекта и заключение………………………………………..................................................18**

**Введение**

**Цели проекта:**

Цель данного исследования — провести статистический анализ исполнения и реализации пенальти (штрафного 11-метрового удара) и сформулировать практические рекомендации для тренерского штаба, направленные на повышение эффективности команды при штрафных ударах.

**Бизнес-задачи:**

1. Оценка влияния пенальти на результат матча:

* Частота назначения;
* Процент реализации;
* Процент от общего числа голов;

1. Рейтинг игроков по эффективности исполнения пенальти:

* Составить рейтинг игроков по качеству исполнения 11 - метрового удара для учета при проведении трансферов игроков;

1. Подготовка вратарей:

* Создание единого отчета по статистике исполнения 11 - метровых ударов по каждому игроку;
* Определение предпочитаемые направления ударов (левый/правый угол, центр ворот).

**Блок 1.**

**Описание исходного датасета и типов данных**

Для исследования был взят датасет “[**Football Events**](https://www.kaggle.com/datasets/secareanualin/football-events/code)” со статистикой по играм пяти крупнейших европейских футбольных лиг:

1. Англии ([*Premier League*](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B0%D1%8F_%D0%9F%D1%80%D0%B5%D0%BC%D1%8C%D0%B5%D1%80-%D0%BB%D0%B8%D0%B3%D0%B0))
2. Германии ([*Bundesliga*](https://ru.wikipedia.org/wiki/%D0%A7%D0%B5%D0%BC%D0%BF%D0%B8%D0%BE%D0%BD%D0%B0%D1%82_%D0%93%D0%B5%D1%80%D0%BC%D0%B0%D0%BD%D0%B8%D0%B8_%D0%BF%D0%BE_%D1%84%D1%83%D1%82%D0%B1%D0%BE%D0%BB%D1%83))
3. Испании ([*La Liga*](https://ru.wikipedia.org/wiki/%D0%A7%D0%B5%D0%BC%D0%BF%D0%B8%D0%BE%D0%BD%D0%B0%D1%82_%D0%98%D1%81%D0%BF%D0%B0%D0%BD%D0%B8%D0%B8_%D0%BF%D0%BE_%D1%84%D1%83%D1%82%D0%B1%D0%BE%D0%BB%D1%83))
4. Италии ([*Serie A*](https://ru.wikipedia.org/wiki/%D0%A7%D0%B5%D0%BC%D0%BF%D0%B8%D0%BE%D0%BD%D0%B0%D1%82_%D0%98%D1%82%D0%B0%D0%BB%D0%B8%D0%B8_%D0%BF%D0%BE_%D1%84%D1%83%D1%82%D0%B1%D0%BE%D0%BB%D1%83_(%D0%A1%D0%B5%D1%80%D0%B8%D1%8F_A)))
5. Франции ([*Ligue 1*](https://ru.wikipedia.org/wiki/%D0%A7%D0%B5%D0%BC%D0%BF%D0%B8%D0%BE%D0%BD%D0%B0%D1%82_%D0%A4%D1%80%D0%B0%D0%BD%D1%86%D0%B8%D0%B8_%D0%BF%D0%BE_%D1%84%D1%83%D1%82%D0%B1%D0%BE%D0%BB%D1%83))

с сезона 2011/2012 по сезон 2016/2017 по состоянию на 25.01.2017.

**Датасет состоит из 3 файлов:**

**events.csv содержит данные о каждой игре. Текстовые комментарии были взяты с сайтов:** [**bbc.com**](http://bbc.com)**,** [**espn.com**](http://espn.com) **и** [**onefootball.com**](http://onefootball.com)

| № | Имя Столбца | Описание | Тип данных |
| --- | --- | --- | --- |
| 1 | id odsp | Уникальный идентификатор матча | object |
| 2 | id\_event | Уникальный идентификатор события в матче | object |
| 3 | sort\_order | Порядковый номер события в матче | int64 |
| 4 | time | Время события (минута) | int64 |
| 5 | text | Текстовое описание события | object |
| 6 | event\_type | Тип события (удар, передача, пенальти и т.д.) | int64 |
| 7 | event\_type 2 | Дополнительный тип события | float64 |
| 8 | side | Сторона (хозяева/гости) | int64 |
| 9 | event team | Команда, совершившая событие | object |
| 10 | opponent | Команда-соперник | object |
| 11 | player | Игрок, совершивший событие | object |
| 12 | player2 | Второй игрок, связанный с событием (ассистент) | object |
| 13 | player\_in | Игрок, который вышел на замену | object |
| 14 | player\_out | Игрок, который был заменен | object |
| 15 | shot place | Место удара | float64 |
| 16 | shot outcome | Результат удара | float64 |
| 17 | is\_goal | Был ли забит гол (0 - нет, 1 - да) | int64 |
| 18 | location | Локация на поле, где произошло событие | float64 |
| 19 | bodypart | Часть тела, использованная для удара или передачи | float64 |
| 20 | assist\_method | Метод ассиста | int64 |
| 21 | situation | Ситуация, в которой произошло событие | float64 |
| 22 | fastbreak | Быстрая контратака (0 - нет, 1 - да) | int64 |

**dictionary.txt содержит словарь с текстовым описанием каждой категориальной переменной, закодированной целыми числами**

eventtype (Тип события)

| Код | Значение | Описание |
| --- | --- | --- |
| 0 | Announcement | Объявление |
| 1 | Attempt | Попытка (удар по воротам) |
| 2 | Corner | Угловой |
| 3 | Foul | Фол |
| 4 | Yellowcard | Жёлтая карточка |
| 5 | Second yellow card | Вторая желтая карточка (и удаление) |
| 6 | Red card | Красная карточка |
| 7 | Substitution | Замена |
| 8 | Free kick won | Выигранный штрафной удар |
| 9 | Offside | Офсайд |
| 10 | Hand ball | Игра рукой |
| 11 | Penalty conceded | Пропущенный пенальти |

#### event\_type 2 (Дополнительный тип события)

| Код | Значение | Описание |
| --- | --- | --- |
| 12 | Key Pass | Ключевой пас |
| 13 | Failed through ball | Сквозной пас |
| 14 | Sending off | Удаление |
| 15 | Own goal | Автогол |

#### side (Сторона)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | Home | Домашняя команда |
| 2 | Away | Гостевая команда |

#### shot place (Место удара)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | Bit too high | Немного выше ворот |
| 2 | Blocked | Блокирован |
| 3 | Bottom left corner | Нижний левый угол |
| 4 | Bottom right corner | Нижний правый угол |
| 5 | Centre of the goal | Центр ворот |
| 6 | High and wide | Выше и мимо ворот |
| 7 | Hits the bar | Попадание в штангу |
| 8 | Misses to the left | Мимо ворот (влево) |
| 9 | Misses to the right | Мимо ворот (вправо) |
| 10 | Too high | Слишком высоко |
| 11 | Top centre of the goal | Верхний центр ворот |
| 12 | Top left corner | Верхний левый угол |
| 13 | Top right corner | Верхний правый угол |

#### shot outcome (Результат удара)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | On target | В створ ворот |
| 2 | Off target | Мимо ворот |
| 3 | Blocked | Блокирован защитником |
| 4 | Hit the bar | Попадание в штангу или перекладину |

### location (Локация на поле)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | Attacking half | Атакующая половина поля |
| 2 | Defensive half | Защитная половина поля |
| 3 | Centre of the box | Центр штрафной площади |
| 4 | Left wing | Левый фланг |
| 5 | Right wing | Правый фланг |
| 6 | Difficult angle and long range | Сложный угол и дальний удар |
| 7 | Difficult angle on the left | Сложный угол слева |
| 8 | Difficult angle on the right | Сложный угол справа |
| 9 | Left side of the box | Левая сторона штрафной площади |
| 10 | Left side of the six yard box | Левая сторона десятиярдовой зоны |
| 11 | Right side of the box | Правая сторона штрафной площади |
| 12 | Right side of the six yard box | Правая сторона десятиярдовой зоны |
| 13 | Very close range | Очень близко к воротам |
| 14 | Penalty spot | Точка пенальти |
| 15 | Outside the box | За пределами штрафной площади |
| 16 | Long range | Дальний удар |
| 17 | More than 35 yards | Более 35 ярдов (≈32 м) |
| 18 | More than 40 yards | Более 40 ярдов (≈36.5 м) |
| 19 | Not recorded | Не записано |

#### bodypart (Часть тела)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | right foot | Правая нога |
| 2 | left foot | Левая нога |
| 3 | head | Голова |

#### assist method (Метод ассиста)

| Код | Значение | Описание |
| --- | --- | --- |
| 0 | None | Без ассиста |
| 1 | Pass | Пас |
| 2 | Cross | Навес |
| 3 | Headed pass | Пас головой |
| 4 | Through ball | Сквозной пас |

#### situation (Ситуация)

| Код | Значение | Описание |
| --- | --- | --- |
| 1 | Open play | Игра в открытой атаке |
| 2 | Set piece | Стандартное положение |
| 3 | Corner | Угловой |
| 4 | Free kick | Штрафной удар |

**ginf.csv — содержит метаданные и рыночные коэффициенты по каждой игре. Коэффициенты были собраны с сайта** [**oddsportal.com**](http://oddsportal.com)

| **№** | **Имя Столбца** | **Описание** | **Тип данных** |
| --- | --- | --- | --- |
| 1 | id odsp | Уникальный идентификатор матча | object |
| 2 | link\_odsp | Ссылка на матч | object |
| 3 | adv\_stats | Наличие расширенной статистики | object |
| 4 | date | Дата матча (YYYY-MM-DD) | object |
| 5 | league | Лига | object |
| 6 | season | Сезон | object |
| 7 | country | Страна | object |
| 8 | ht | Домашняя команда (Home Team) | object |
| 9 | at | Гостевая команда (Away Team) | object |
| 10 | fthg | Голы домашней команды (Full Time Home Goals) | int64 |
| 11 | ftag | Голы гостевой команды (Full Time Away Goals) | int64 |
| 12 | odd h | Коэффициент на победу домашней команды | float64 |
| 13 | odd d | Коэффициент на ничью | float64 |
| 14 | odd\_a | Коэффициент на победу гостей | float64 |
| 15 | odd\_over | Коэффициент на тотал больше (например, 2.5) | float64 |
| 16 | odd\_under | Коэффициент на тотал меньше | float64 |
| 17 | odd\_bts | Коэффициент на "Обе забьют" (Both Teams to Score) | float64 |
| 18 | odd\_bts\_n | Коэффициент на "Обе НЕ забьют" | float64 |

**Блок 2.**

**Подготовка, очистка и преобразование данных**

**2.1 Подготовка**

1. В столбце 'date' изменен тип данных с object на datetime64
2. Не все матчи из файла gif есть в events. 1038 матчей не имеют расширенной статистики

**2.2 Преобразование**

1. Отфильтруем таблицу df\_events оставив только столбцы необходимые для анализа

| Имя Столбца | Описание | Тип данных |
| --- | --- | --- |
| id odsp | Уникальный идентификатор матча | object |
| time | Время события (минута) | int64 |
| text | Текстовое описание события | object |
| event\_type | Тип события (удар, передача, пенальти и т.д.) | int64 |
| event team | Команда, совершившая событие | object |
| opponent | Команда-соперник | object |
| player | Игрок, совершивший событие | object |
| player2 | Второй игрок, связанный с событием (ассистент) | object |
| shot place | Место удара | float64 |
| shot outcome | Результат удара | float64 |
| is\_goal | Был ли забит гол (0 - нет, 1 - да) | int64 |
| bodypart | Часть тела, использованная для удара или передачи | float64 |

**Блок 3.**

**Анализ данных для стейкхолдеров**

***Целью блока является поиск тенденций и инсайтов для составления рекомендаций стейкхолдерам.***

**График распределения голов**

С целью знакомства с данными построим график распределения голов по времени матча



Голов в первом тайме: 10744

Голов во втором тайме: 13702

Проведя анализ статистики распределения голов по временным интервалам можно сделать следующие выводы:

* Больше голов забивается во втором тайме
* Самое большое количество голов забивается после 75 минуты (с учетом дополнительного времени)

**3.1 Оценка влияния пенальти на результат матча.**

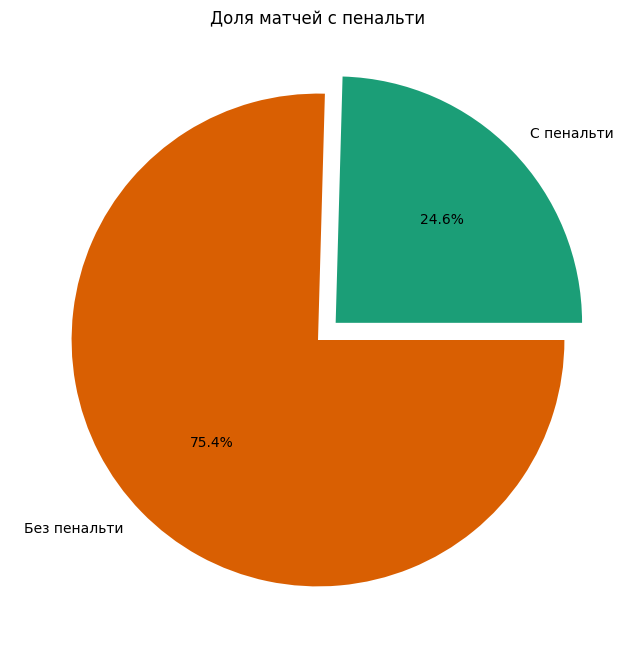
***Общее количество пенальти: 2596***

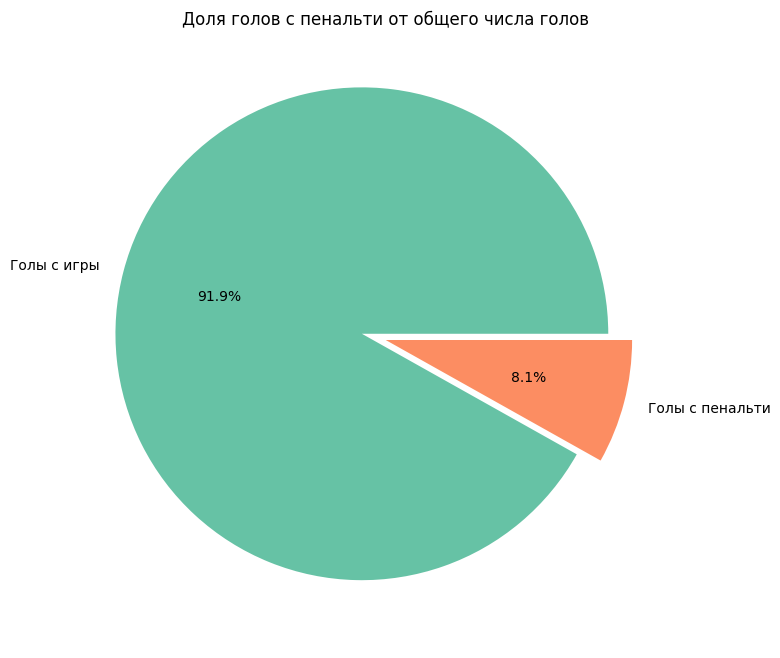
***Количество забитых пенальти: 1989***

***Процент реализации пенальти: 76.6%***

***Общее количество голов: 24446***

***Доля пенальти от общего числа голов: 8.14%***

****

****

**Проведенный анализ статистических данных позволяет сделать следующие выводы:**

**- Высокая частота назначения пенальти**

В 24,6% футбольных матчей назначается минимум один 11-метровый удар, что свидетельствует о значительном влиянии данного игрового эпизода на ход встречи.

**- Качество реализации пенальти**

Точность исполнения пенальти составляет 76,6%, что означает высокую эффективность. Эффективность пенальти зависит не только от точности бьющего, но и от действий вратаря. Это подтверждает необходимость специализированных тренировок вратарей в отражении 11-метровых ударов и психологической подготовки исполнителей.

**- Вклад пенальти в общую результативность**

Доля голов, забитых с пенальти, составляет 8,1% от общего количества забитых мячей. Этот показатель демонстрирует, что пенальти оказывают существенное влияние на набор очков командой в течении сезона.

**Заключение:**

Пенальти являются значимым фактором, влияющим на результат футбольных матчей. Их высокая реализация (76,6%) в сочетании с заметной долей в общей статистике голов (8,1%) подчеркивает необходимость совершенствования методик подготовки игроков и вратарей к данным игровым ситуациям. Дальнейшее исследование направлено на выявление и проверку метрик для оценки индивидуальных показателей игроков.

**3.2 Рейтинг игроков по эффективности исполнения пенальти**

| **Футболист** | **Пробил пенальти** | **Реализовано** | **Точность исполнения, %** |
| --- | --- | --- | --- |
| **cristiano ronaldo** | **51** | **43** | **84.31** |
| **zlatan ibrahimovic** | **38** | **35** | **92.11** |
| **lionel messi** | **37** | **30** | **81.08** |
| **edinson cavani** | **27** | **20** | **74.07** |
| **falcao** | **22** | **20** | **90.91** |
| **antonio candreva** | **22** | **18** | **81.82** |
| **sejad salihovic** | **21** | **19** | **90.48** |
| **alexandre lacazette** | **21** | **18** | **85.71** |
| **mauricio pinilla** | **20** | **14** | **70.00** |
| **eden hazard** | **19** | **18** | **94.74** |

## **3.3 Индивидуальная статистика игрока**

Для примера приведена статистика [Криштиану Роналду](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%B8%D1%88%D1%82%D0%B8%D0%B0%D0%BD%D1%83_%D0%A0%D0%BE%D0%BD%D0%B0%D0%BB%D0%B4%D1%83)

**Статистика пенальти cristiano ronaldo:**

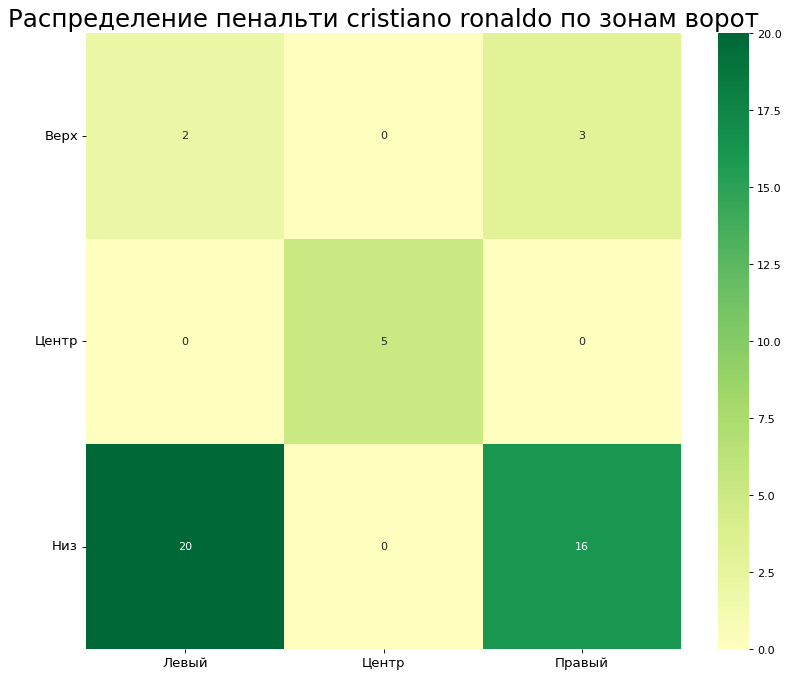
***--------------------------------***

***Всего пробито пенальти: 51***

***Реализовано: 43 (84%)***

***Промахов: 8***

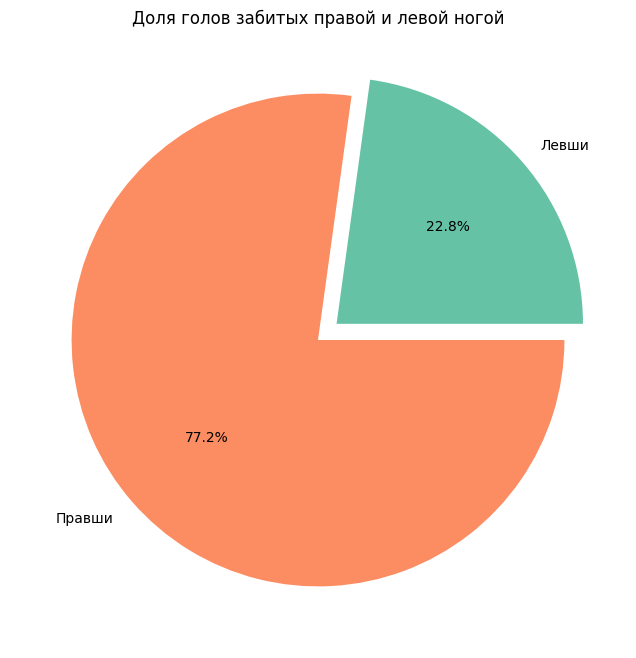
***Бьющая нога: правая***

****

Составление рейтинга эффективности исполнения пенальти игроков, а также подробная информация по индивидуальной статистике игрока позволит тренерскому штабу футбольного клуба:

* Выявлять игроков с высокой реализацией пенальти для усиления состава.
* Сравнивать игроков одной позиции по дополнительному критерию (% реализации, ударная нога).
* Более качественно проводить тренировки вратарей делая акцент на индивидуальных особенностях соперника.

**3.4 Оценка влияния ударной ноги футболиста на качество исполнения пенальти**

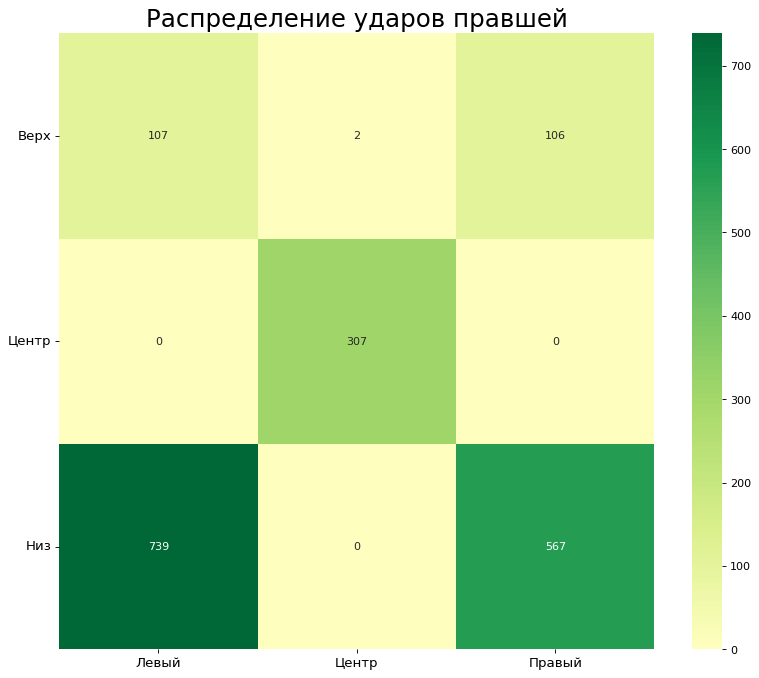
****

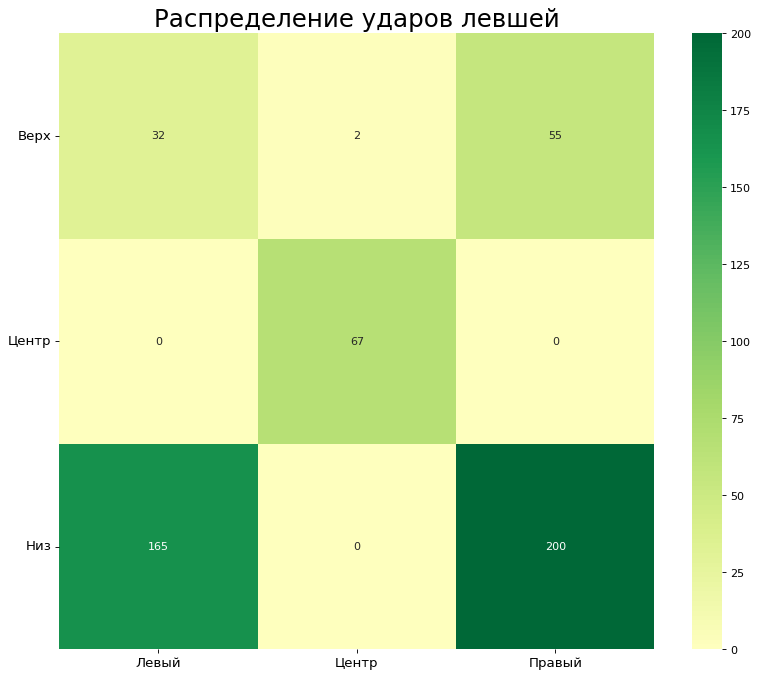
**Процент реализации правшей: 76.0 %**

**Процент реализации левшей: 78.8 %**

**Вывод: несмотря на преобладание правшей над левшами (77,2% и 22,8%) качество реализации практически не отличается (76,0% и 78,8%), поэтому данный критерий не должен учитываться при выборе исполнителя пенальти.**

**3.5 Статистика распределения по зонам ворот ударов при пенальти левшей и правшей**

****

****

**Выводы:**

* **Проведенный анализ показывает, что в большинстве случаев правши бьют в левый от себя угол ворот, а левши — в правый.**
* **Результативнее удары в нижние углы ворот**

# **Выводы**

**Принято считать, что исход пенальти (штрафного 11-метрового удара) преимущественно определяется фактором везения, приближая его к лотерее. Однако проведенное исследование на основе анализа 2596 пенальти из пяти ведущих европейских лиг с 2011 по 2017 годы демонстрирует статистически значимые закономерности, опровергающие данное предположение.**

1. **Пенальти влияют на 24.6% матчей, их реализация — 76.6% - существенно превышает случайное распределение**
2. **Выявлена закономерность по ударам правшей и левшей (правши чаше бьют в левый от себя угол ворот, а левши — в правый)**
3. **Удары в нижние углы ворот результативнее, чем в другие зоны ворот.**

**По бизнес - задачам:**

1. **Исходные данные были преобразованы и отфильтрованы.**
2. **На основе анализа данных были приведены обоснования о значимости пенальти как для отдельного матча, так и на набор очков командой в течении сезона.**
3. **Составлен рейтинг игроков по эффективности исполнения пенальти.**
4. **Разработан алгоритм и форма отчета для определения индивидуальных характеристик исполнителя пенальти.**

**Результаты анализа данных по бизнес - задачам будут полезны тренерским штабам футбольных команд при подготовке команды к сезону.**

**Практическая значимость для стейкхолдеров:**

* **Научно обоснованный выбор исполнителей пенальти.**
* **Объективные критерии оценки игроков при трансферах.**
* **Оптимизация тренировочного процесса по отработке 11-метровых ударов.**
* **Персонализированная подготовка вратарей.**
* **Методика сбора и обработки данных по пенальти.**

**В ходе анализа была выявлена недостаточность информативности данных для предоставления четких выводов, поэтому для улучшения качества данных необходимо доработать отчет и дополнить колонками со следующей информацией:**

* **Позиция игрока на поле.**
* **Фамилия и имя вратаря при розыгрыше пенальти.**
* **Антропометрические данные вратаря.**
* **Направление прыжка вратаря.**
* **Погодные условия.**

# **Заключение**

# **В рамках дипломной работы была рассмотрена лишь малая часть аналитики футбольных событий, но даже она позволяет сделать значимые выводы. Пенальти оказывают существенное влияние на результат матчей, а их реализация зависит от множества факторов, включая индивидуальные особенности игроков. Результаты исследования демонстрируют, что анализ данных может стать мощным инструментом для тренерского штаба, помогая оптимизировать подготовку игроков, выбирать исполнителей пенальти и улучшать тренировочный процесс. Это подчеркивает важность дальнейшего углубления в аналитику футбольных событий для достижения более точных и практических рекомендаций.**