**🏆 Название проекта**

**Статистический анализ факторов, влияющих на победу в матчах чемпионата мира по футболу (Катар, 2022)**

**📌 Цель проекта**

Целью проекта является выявление статистически значимых факторов, влияющих на исход матчей чемпионата мира по футболу 2022 года. Основное внимание уделяется таким показателям, как владение мячом, точность пасов и удары в створ ворот.

**📊 Краткое описание данных**

* Источник: <https://fbref.com/en/>
* Всего: 64 матча ЧМ-2022.
* Формат: веб-страницы с табличными статистиками.
* Данные: команды, счёт, владение мячом, точность пасов, удары в створ.

**🛠 Используемые инструменты**

| **Инструмент / язык** | **Назначение** |
| --- | --- |
| **C# + HtmlAgilityPack** | Парсинг HTML-страниц матчей с FBref |
| .NET Core / VSCode | Основная среда разработки |
| CSV-файл / Excel | Сохранение и последующий анализ |
| Power BI / Excel | Визуализация результатов |
| GitHub | Хранение кода |

**📋 Формулировка гипотез**

Для анализа были сформулированы следующие статистические гипотезы:

1. **H₀ (нулевая гипотеза):** показатель (например, владение мячом) не влияет на победу команды.
2. **H₁ (альтернативная гипотеза):** показатель оказывает влияние на вероятность победы.

Проверка проводится для следующих факторов:

* владение мячом,
* точность пасов,
* удары в створ.

**📘 Теоретическая база**

* **Корреляция Пирсона** — измеряет силу линейной связи между переменными.
* **t-критерий Стьюдента** — сравнение средних значений двух групп (например, победителей и проигравших).
* **p-value** — проверка статистической значимости.
* **Визуализация** — диаграммы рассеяния, боксплоты, столбчатые диаграммы.

**🪜 План действий (декомпозированный по этапам)**

**🔹 Этап 1: Постановка задачи и гипотез**

* Сформулировать цель: «Какие показатели влияют на победу?»
* Выделить ключевые переменные: владение, пассы, удары.
* Прописать нулевые и альтернативные гипотезы.

**Ответственный:** тимлид + аналитик

**🔹 Этап 2: Сбор данных**

* Спарсить 64 страницы матчей ЧМ 2022 с сайта FBref.
* Получить:
  + Названия команд;
  + Счёт;
  + Владение мячом (в процентах);
  + Точность пасов;
  + Удары в створ ворот.
  + xG(оценка вероятности забить гол в конкретном моменте игры(в итоге получается сумма за весь матч))

**Инструмент:** HtmlAgilityPack  
**Формат:** сохранение в .csv  
**Ответственный:** разработчик

**🔹 Этап 3: Обработка данных**

* Привести все строки к числовому виду (например, "80%" → 80).
* Проставить флаг победителя (победа/ничья/поражение).
* Подготовить таблицу для анализа.

**Инструмент:** C# / Excel  
**Ответственный:** разработчик + аналитик

**🔹 Этап 4: Статистический анализ**

* Разбить выборку на группы: победители и проигравшие.
* Рассчитать средние значения метрик в обеих группах.
* Провести t-тесты на значимость различий.
* Вычислить корреляцию между метриками и исходом матча.

**Инструмент:** Excel / Python (по желанию)  
**Ответственный:** аналитик

**🔹 Этап 5: Визуализация**

* Построить:
  + Столбчатые графики: средние значения метрик;
  + Диаграммы рассеяния: взаимосвязи;
  + Боксплоты: распределения по группам.

**Инструмент:** Excel / Power BI / Python  
**Ответственный:** визуализатор

**🔹 Этап 6: Подготовка отчёта**

* Титульный лист.
* Постановка задачи и гипотезы.
* Описание данных и методов.
* Теория (кратко: t-тест, корреляция, p-value).
* Результаты (таблицы, графики).
* Обсуждение и выводы.
* Ссылка на GitHub с кодом.

**Формат:** PDF  
**Ответственный:** тимлид + все

**📈 Результат**

* Финальный .csv со всеми 64 матчами.
* Мини-анализ в духе: «владение > 55% увеличивает шанс победы».
* Простые, понятные графики.
* Оформленный отчёт по шаблону.

**💬 Обсуждение и возможные расширения**

* Можно подключить больше метрик: угловые, xG (если найдём).
* Можно построить простую модель предсказания (опционально).
* Можно реализовать всё в Python позже, чтобы сравнить подходы.