

Анализ зависимости и результативности футбольных клубов

Проект выполнили:

Гусельникова Кристина
Афанасиева Арина
Долотин Артур
Суворов Максим



Бизнес - конспект

Группа инвесторов из Ближнего Востока планирует создать футбольный клуб в Европе. Заказчики заинтересованы в понимании того, **какие финансовые вложения действительно приводят к спортивному успеху**, а какие — нет. Несмотря на то, что топ-клубы тратят сотни миллионов евро ежегодно, **высокие расходы не всегда коррелируют с победами.**

Цель исследования

Определить закономерности между уровнем инвестиций, стоимостью состава и реальными спортивными результатами, чтобы заказчики смогли сформировать стратегию построения **эффективного и устойчивого клуба.**

Формулировка задачи

Определить, как финансовые вложения влияют на спортивные результаты футбольных клубов, и выявить, какие направления инвестиций обеспечивают наибольшую отдачу — в очках, трофеях, росте рыночной стоимости состава и интересе болельщиков.

Ожидаемый результат для заказчика

1 Определить зависимость результативности от стоимости состава

Определить, что в большей степени влияет на популярность: стоимость состава, кол-во забитых голов, результативность

2



ТОП-6 лучших лиг мира на 2025 год

Бундеслига

Германия

Премьер-Лига

Англия

ЛаЛига

Испания

Лига Betclic

Португалия

Лига 1

Франция

Серия А

Италия

Для каждой конкретной команды из каждой лиги ищем их результаты (победы, ничьи, поражения, забитые голы) за последние 5 лет

Количество игроков

Средний возраст

Кол-во легионеров

Общая стоимость игроков

Разница в стоимости за сезон

Средний результат за игру

Возраст клуба



Датасеты

Целевая переменная
avg_points

Dataset 1

Matches Dataset (Scraping)

Содержимое: результаты матчей, количество побед/поражений, забитые и пропущенные мячи, сезон, лига

Squad Cost (API)

Dataset 2

Содержимое: стоимость состава (total_cost), количество игроков (players_amount), средний возраст (avg_age), число легионеров (foreign_players_amount), разница стоимости (team_cost_dif)

Dataset 3

Attendance Data (Web Scraping)

Содержимое: средняя и общая посещаемость (avg_attendance, total_attendance), аншлаги (sellouts), сезон, лига

Wikipedia API

Dataset 4

Содержимое: год основания клуба (year_of_foundation)

Dataset 5

Final Data

Объединение всех источников по ключам: team, league, season.

Финальный датасет включает финансовые, спортивные, популяризационные и исторические признаки.

Гипотезы

1. Болельщики поколениями поддерживают один и тот же клуб
2. На результативность команды преимущественно влияет стоимость состава

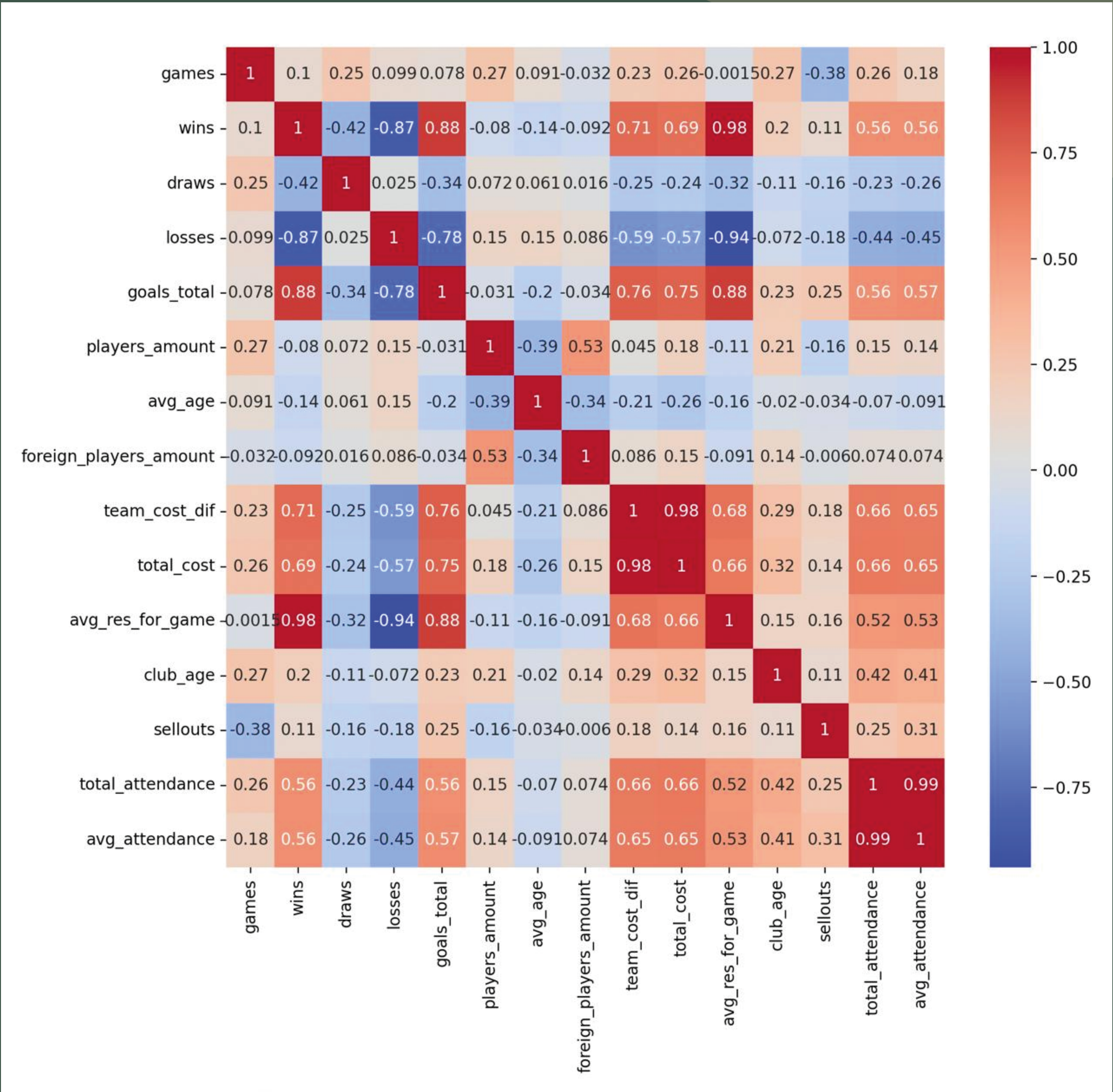
3. Посещаемость зависит только от результативности



Типы признаков в финальном датасете

Категория признаков	Название столбцов	Тип данных	Описание
Целевая переменная	avg_pts	float	Среднее количество очков за матч (цель анализа)
Спортивные	wins, draws, losses, games, goals_total	int	Количество побед, ничьих, поражений, матчей и голов
Финансовые	total_cost, team_cost_dif, avg_age	float	Стоимость состава, изменение стоимости, средний возраст игроков
Состав команды	players_amount, foreign_players_amount	int	Количество игроков и легионеров в команде
Популярность	avg_attendance, total_attendance, sellouts	float	Средняя и общая посещаемость, количество аншлагов
Исторические	club_age	int	Года существования клуба
Категориальные (текстовые)	team, league, season	string	Название клуба, лиги и сезона
Временные (производные)	season_year_start, season_year_end (если добавлено)	int / datetime	Начальный и конечный годы сезона

Heatmap



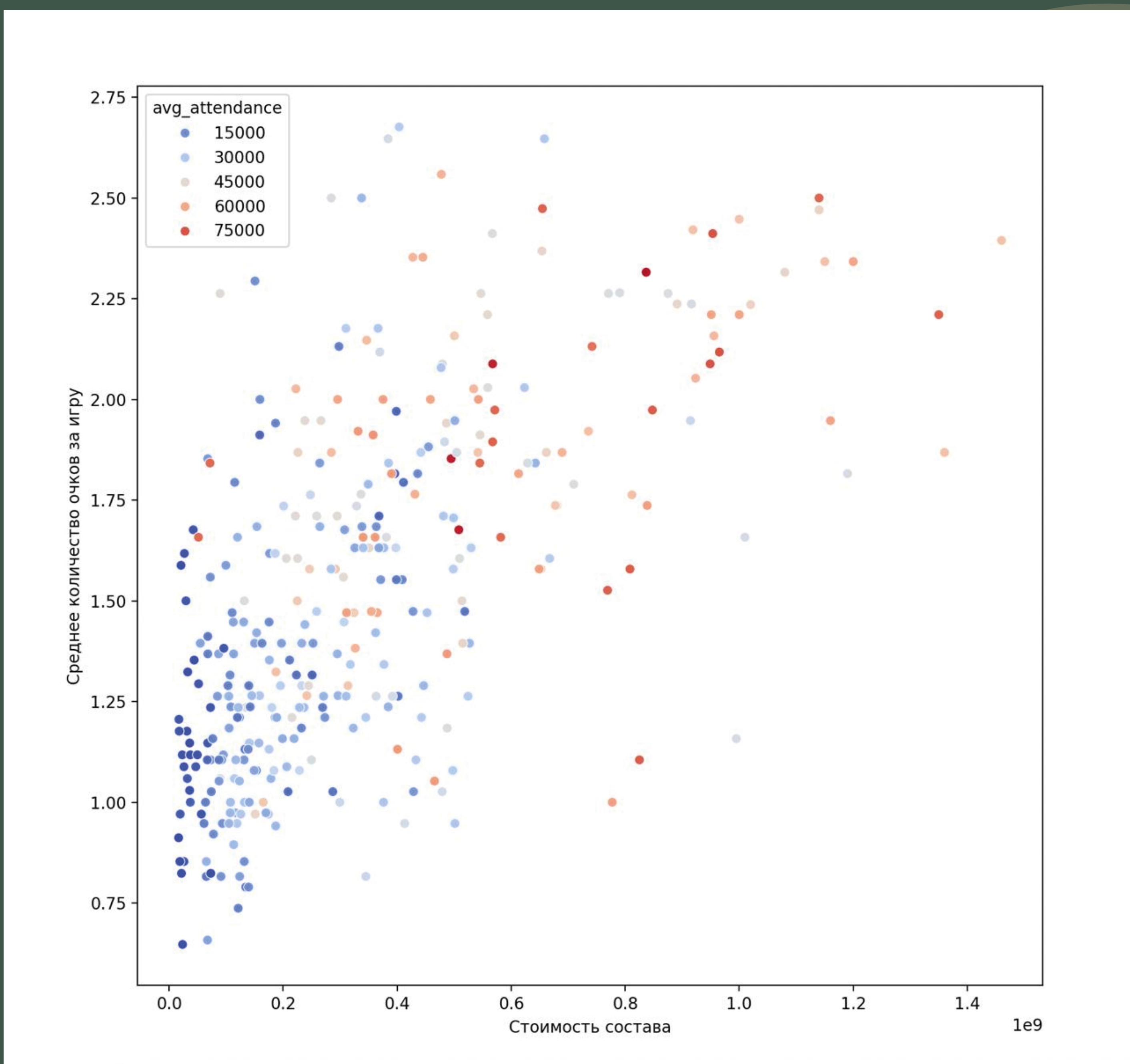
Интересные корреляции



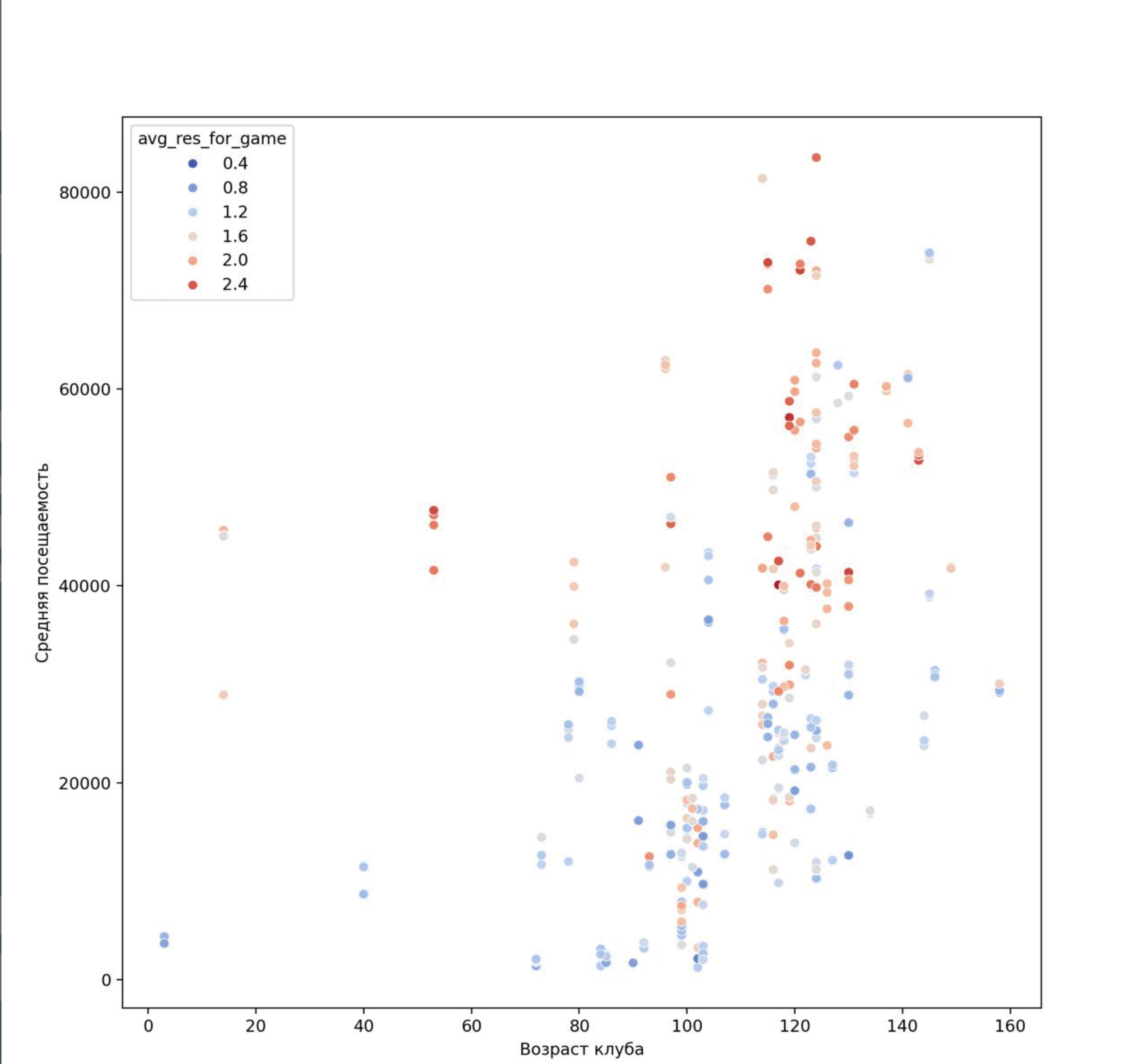
total_cost - avg_res_for_game - 0,66
total_attendance - total_cost - 0,66
avg_res_for_game - avg_attendance - 0,52
club_age - avg_attendance - 0,4



Зависимость посещаемости от результативности и стоимости состава



Влияние возраста клуба на посещаемость и результат



Коэффициент каждой лиги

league	total_cost
Premier League	2.890350
Serie A	5.508863
La Liga	5.920398
Bundesliga	7.594349
Ligue 1	8.801727
Liga Portugal	17.153533

Клубы с наибольшими результатами при наименьшей стоимости состава

season	league	team
2024-2025	Serie A	AC Milan
2022-2023	Serie A	AC Milan
2021-2022	Serie A	AC Milan
2022-2023	Liga Portugal	FC Arouca
2022-2023	La Liga	CA Osasuna
2022-2023	Liga Portugal	Boavista FC
2022-2023	Liga Portugal	Casa Pia AC
2021-2022	Liga Portugal	CS Marítimo

Выводы

1. Популярность клубов напрямую связана с их инвестиционной привлекательностью.
2. Результативность имеет чёткую линейную зависимость от стоимости состава.
3. Возраст клуба не оказывает значимого влияния на посещаемость.
4. Фанаты ориентируются на динамику успеха, а не на наследие.
5. Выявлены клубы с наибольшей эффективностью состава.

Инструменты извлечения данных

Web Scrapping (requests, BeautifulSoup, Selenium, AJAX, CSRF-токеном, cookies, JSON и HTTP-заголовками)
источники: transfermarket.world, globalsportsarchive.com

API-запросы
источники:
field-hub.online,
wikipedia API

Источник	Признаки
Transfermarkt	Сезон, Лига, Матчи, Аншлаг, Зрители, Средняя посещаемость, Команда
Globalsportsarchive	Команда, Соперник, Дата, Результат, Лига, Сезон
Field-hub	Рыночная стоимость команды
Wikipedia API	Возраст клуба

Спасибо за внимание

