

Vizualizacija podataka iza timova nogometne Premier lige

Vizualizacija podataka, Filip Matić

0246087655

1. Uvod

Premier liga je najviši rang engleskog nogometa, ujedno i najelitnija nogometna liga svijeta. Sastoji se od 20 ekipa, gdje 3 najlošije ekipe svake godine ispadaju u niži rang, a 3 nove ulaze u viši. Službeno je osnovana, s ovim imenom i formatom, u 1992.oj godini, od kada se ukupno izmijenilo 50 klubova, a 7 ih je osvajalo naslov. Osim borbe za naslov, bitan je i poredak prvih 6-7 klubova (ovisno o europskom koeficijentu lige te godine) zbog plasmana u europska natjecanja.

Kao veliki fanatik engleskog nogometa, te fanatik „Fantasy football-a“, lako mi je bilo odlučiti se za ovakvu tematiku. Kroz prikazane vizualizacije sam prikazao omjere brojnih metrika pojedinih timova, što u ovoj sezoni, što u protekle 3 sezone. Glavna inspiracija za odabir ove teme mi je bio trenutni uspjeh Nottingham Forresta, koji unatoč naoko lošijim brojkama postignutih golova i utrošenog novca, u trenutku pisanja ovog rada drži drugo mjesto na tablici.

2. Podaci

Što se tiče podataka, nisam htio ići predaleko u prošlost sa starim sezonama zbog prezasićenosti grafova, a skupovi podataka s trenutnom sezonom još nisu dostupni. Zaključno s tim, odlučio sam napraviti vlastiti dataset u kojem ću skupiti neke od najbitnijih karakteristika i metrika svih ekipa. Kako ne bi ručno skupljao podatke, odradio sam jednostavan „web-scraping“ sa stranice <https://fbref.com/en/>. Podaci su skupljeni za aktualnu sezonu (s tim da u vrijeme predaje ovog projekta budu možda malo zastarjeli), te za 3 prošle završene sezone kako bi mogli usporediti međusobno uspjehe i metrike pojedinih ekipa kroz više sezona.

Podaci su skupljeni koristeći BeautifulSoup biblioteku i programski jezik Python.

```

url = "https://fbref.com/en/comps/9/Premier-League-Stats"

response = requests.get(url)

if response.status_code != 200:
    print(f"Failed to fetch data. Status code: {response.status_code}")
    exit()

soup = BeautifulSoup(response.text, 'html.parser')
table = soup.find('table', {'class': 'stats_table'})
headers = [header.text.strip() for header in table.find_all('th')[1:]]
data = []
for row in table.find_all('tr')[1:]:
    cols = row.find_all('td')
    if len(cols) > 0:
        ranking = row.find('th').text.strip()
        team_data = {
            "ranking": ranking,
            "Team": cols[0].text.strip(),
            "MP": cols[1].text.strip(),
            "W": cols[2].text.strip(),
            "D": cols[3].text.strip(),
            "L": cols[4].text.strip(),
            "GF": cols[5].text.strip(),
            "GA": cols[6].text.strip(),
            "GD": cols[7].text.strip(),
            "Pts": cols[8].text.strip(),
            "Pts/MP": cols[9].text.strip(),
            "xG": cols[10].text.strip(),
            "xGA": cols[11].text.strip(),
            "xGD": cols[12].text.strip(),
            "xGD/90": cols[13].text.strip(),
            "Last 5": cols[14].text.strip(),
            "Attendance": cols[15].text.strip(),
            "Top Team Scorer": cols[16].text.strip(),
            "Goalkeeper": cols[17].text.strip(),
            "Notes": cols[18].text.strip()
        }

        data.append(team_data)

output_file = "prem_stats_24_25.json"
with open(output_file, 'w', encoding='utf-8') as f:
    json.dump(data, f, ensure_ascii=False, indent=4)

print(f"Data saved to {output_file}")

```

Podaci su spremljeni u 4 JSON datoteke za 4 sezone, kako bi bilo lakše pristupiti podacima kroz različite sezone u Javascriptu.

Objašnjenja podataka:

„**ranking**“: Pozicija tima na tablici

„**team**“: Ime tima

„**MP**“: Odigran broj utakmica

„**W**“: Broj pobjeda

„**D**“: Broj remija

„**L**“: Broj poraza

„**GF**“: Broj postignutih golova

„**GA**“: Broj primljenih golova

„**GD**“: Gol razlika

„**Pts**“: Broj osvojenih bodova

„**Pts/MP**“: Prosječan broj osvojenih bodova po meču

„**xG**“: Očekivan broj golova

„**xGA**“: Očekivan broj primljenih golova

„**xGD**“: Očekivana gol razlika

„**xGD/90**“: Očekivana gol razlika po meču

„**Last 5**“: Trenutna forma, odnosno forma u proteklih 5 utakmica

„**Attendance**“: Prosječna posjećenost stadiona

„**Top Team Scorer**“: Najbolji strijelac tima

„**Goalkeeper**“: Golman

Iako je u današnjim vremenima gdje se prati svaki pokret svakog igrača na terenu dostupan puno veći broj raznih metrika, odlučio sam da su ove metrike sasvim dovoljne kako bi se prikazao omjer snaga svih ekipa, što ofanzivno što defanzivno.

3. Vizualizacije

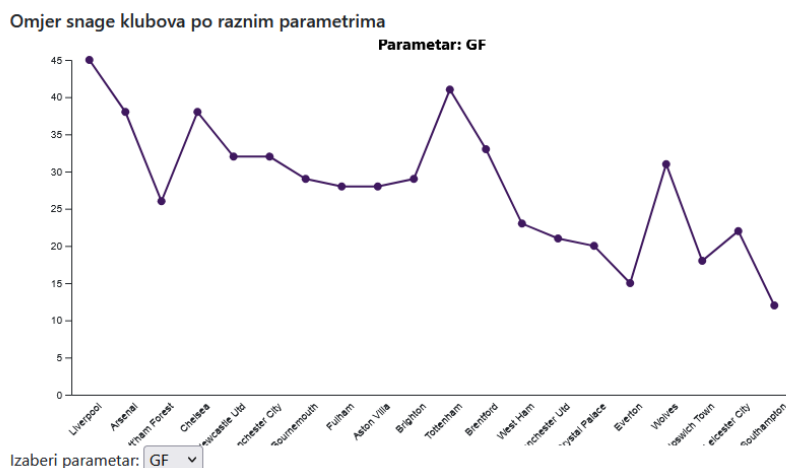
3.1. S lijeva na desno, prva prikazana vizualizacija je linijski graf. Radi se o grafu koji u isto vrijeme prikazuje odnos svih 20 timova lige po pojedinim parametrima. Korisnik ima mogućnost birati koji parametar želi da se uspoređi, a omogućeno mu je biranje između postignutih, primljenih, očekivano postignutih i očekivano primljenih golova. To su neki od najosnovnijih parametara kojima možemo vidjeti koliko su timovi orijentirani ofanzivno, odnosno defanzivno. Iznad grafa je naslov koji prikazuje trenutno odabranu metriku, a dropdown menu ispod grafa sugerira korisniku da je graf interaktivan.

Na X-osi su prikazani timovi, poredani po njihovom tabličnom poretku u tom trenutku. Na Y-osi su numeričke vrijednosti, kreće se okvirno do maksimalne vrijednosti svih metrika, konkretno do broja 45.

Svaka točka na grafu predstavlja broj golova koje je određeni gol zabio/primio i slično. Prelaskom kursorom preko točaka dobije se točna informacija o broju ako nije najvidljivije iz grafa.

Vizualno gledano graf je jednostavan i lako čitak. Linija i točke su iste boje, boje #3f185f koja je boja na službenom logotipu Premier lige. Prilikom mijenjanja parametara, odnosno prilikom mijenjanja linijskog grafa, postoji blaga animacija kako bi korisniku bilo jednostavnije pratiti kakve promjene se događaju.

Ovom vizualizacijom vidimo neke zanimljive stvari poput Nottinghama i Tottenhama te njihovih uspjeha. Vidimo da se Nottingham nalazi na trećem mjestu, ali ima znatno niži broj postignutih pogodaka nego npr. Tottenham i Wolverhampton, koji su po broju pogodaka u rang s liderima, ali su dosta nisko na tablici.



3.2. Druga vizualizacija je jednostavna tablica koja predstavlja poredak timova u ligi. U tablici su prikazane osnovne informacije poput poretka, imena ekipe i broja osvojenih bodova, ali i neke detaljnije poput broja golova primljenih i postignutih, očekivanih golova, broja pobjeda poraza i remija te trenutnu formu ekipe. Estetski gledano, nema se puno toga za reći o ovoj vizualizaciji jer se radi o običnoj nogometnoj tablici. Kako ne bi tablica samo bila tablica, dodane su neke interaktivne funkcije.

Iznad tablice se nalazi dropdown menu gdje korisnik može izabrati za koju sezonu se prikazuju podaci, gdje naravno trenutna nije završena. Prilikom biranja jedne od proteklih sezona, stupac s trenutnom formom ekipe se briše.

Osim biranja sezona, korisnik je u mogućnosti poredati tablicu po pojedinim metrikama. Odnosno, klikom na npr. vrh stupca za formu, tablica se preformulira tako da na prvom mjestu bude tim koji je trenutno u najboljoj formi. To je omogućeno funkcijom koja računa formu svake ekipe po bodovima (3 boda za pobjedu, 1 za remi, 0 za poraz). Osim po formi, korisnik može prerasporediti tablicu i klikom na ostale parametre te tako poredati tablicu npr. po broju pogodaka itd.

Trenutno stanje tablice

2024-25 ▾

Pos	Team	MP	W	D	L	GF
1	Liverpool	18	14	3	1	45
2	Arsenal	19	11	6	2	38
3	Nott'ham Forest	19	11	4	4	26
4	Chelsea	19	10	5	4	38
5	Newcastle Utd	19	9	5	5	32
6	Manchester City	19	9	4	6	32
7	Bournemouth	19	8	6	5	29

3.3. Treća vizualizacija je stupčasti graf, odnosno bar chart. Radi se o jednostavnoj vizualizaciji, koja se sastoji samo od dva stupca. Cilj ovog grafa je usporedba očekivanih postignutih golova svakog tima, i koliko su ih zapravo postigli. Na taj način vidimo koliko timovi zapravo iskorištavaju svoje prilike, a koliko su neki timovi „sretniji“ s postizanjem golova.

Na Y-osi se nalaze vrijednosti metrika, raspoređeni u razmacima od 5 golova. Ako je odabrana trenutna sezona, maksimalna vrijednost je 50 golova, a odabirom proteklih sezona se maksimum diže na 100 golova. To je učinjeno zbog čitkosti grafa, kako ne bi npr. stupac bio prenizak za određene timove kad je maksimum 100, ili kad je maksimum 50 da ne prelaze timovi preko limita.

Boja stupaca je ponovno ista kako bi se zadržao neki vizualni trend na cijeloj stranici.

Ako iz prvotnog prikaza nije jasno očitati točan broj golova, prelaskom kursora preko stupca se dobiju detaljnije informacije o broju golova i očekivanih golova.

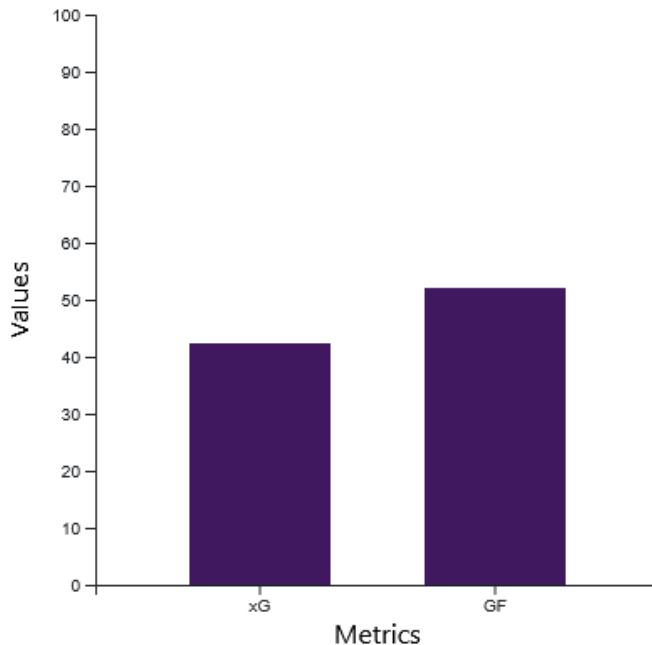
Graf je ponovno interaktivan, odnosno korisnik može birati koji će tim biti prikazan, te u kojoj sezoni. Ako je slučajno odabran tim koji u odabranoj sezoni nije bio u Premier ligi, ostaje prikazana posljednja uspješna vizualizacija tog tima.

Neke zanimljive stvari koje možemo vidjeti su npr. omjer postignutih i očekivanih golova kod ekipa za koje smo navikli da inače imaju bolji plasman. Npr. vidimo kod Manchester Citya i West Hama, koji imaju lošije sezone nego smo navikli, kako ne iskorištavaju prilike koje im se nude, dok pak kod ekstremno ofanzivnih timova poput Wolverhamptona, vidimo ogromnu pozitivnu razliku postignutih golova i očekivanih, zbog individualnosti ekipe.

Promatrajući pojedine timove i njihove plasmane, dolazimo do zaključka da ovakve metrike nisu dovoljne za primjećivanje nekog trenda u odnosu na njihovu uspješnost u ligi, ali je zanimljivo za praćenje stila igre pojedinih timova i njihovog stanja na tablici.

Omjer očekivanih i postignutih golova

2023-24 ▾ Luton Town ▾



3.4. Četvrta vizualizacija je radarski graf, poznat kao i paukov graf, koji vizualno prikazuje više varijabli za određeni klub u određenoj sezoni.

Prikazane su 4 neke osnovne metrike i njihov međusobni odnos : očekivani golovi postignuti i primljeni, gol razlika i osvojeni bodovi.

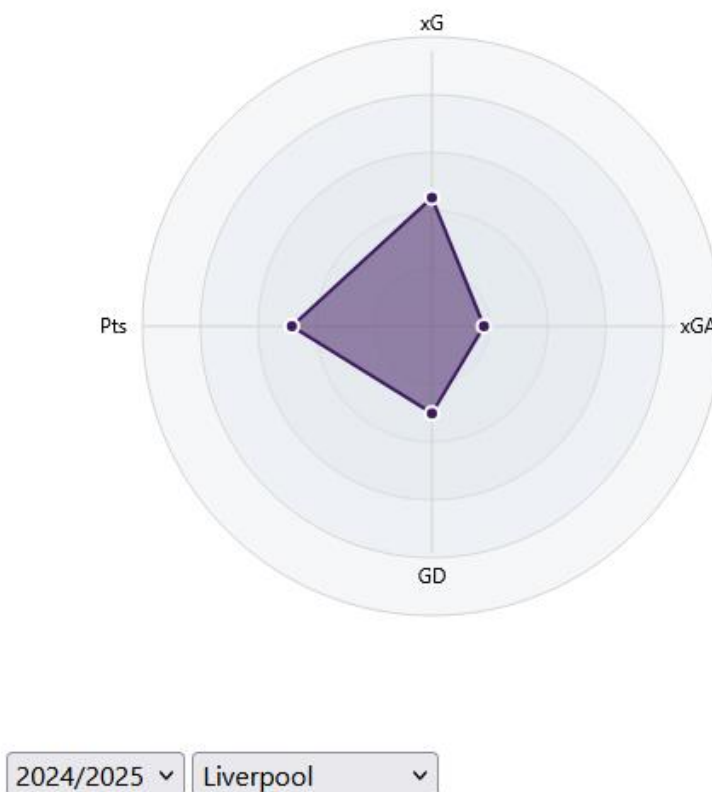
Graf ima više osi, svaka predstavlja jednu od varijabli. Vrijednosti za svaku varijablu su prikazane kao točke na odgovarajućim osima. Točke su povezane linijama koje formiraju poligon, čim se omogućava vizualno uspoređivanje varijabli.

Što se tiče boje, nastavljeno je istim tonom, uz to da su poligoni ispunjeni prozirnijom ljubičastom bojom. Koncentrični krugovi pružaju referentne okvire za poređenje vrijednosti različitih varijabli. Točke

I još jednom je graf interaktivan, ispod grafa se nalazi dropdown menu za izbor sezone i ekipe.

Za bolje vizualno razumijevanje promjena prilikom interaktivnih mijenjanja, dodana je animacija koja uspoređuje promjene pozicije točaka.

S radarskim grafom sam postigao kompaktnu vizualizaciju višedimenzionalnih podataka, te omogućio korisniku da brzo uspoređi ključne metrike ekipe.



3.5. Posljednja vizualizacija je linijski graf koji prikazuje promjenu bodova za različite timove Premier lige kroz prošle 3 sezone. Ovaj graf je dodan kako bi prikazao rast i razvoj, ili suprotno, pojedinih timova kroz više sezona.

X-os predstavlja sezone koje su linearno raspoređene predstavljajući kontinuitet između sezona. Y-os predstavlja broj osvojenih bodova.

Svaka linija predstavlja jedan tim i prikazuje kako se broj bodova mijenjao kroz tri sezone. Timovi su prikazani različitim bojama za lako razlikovanje.

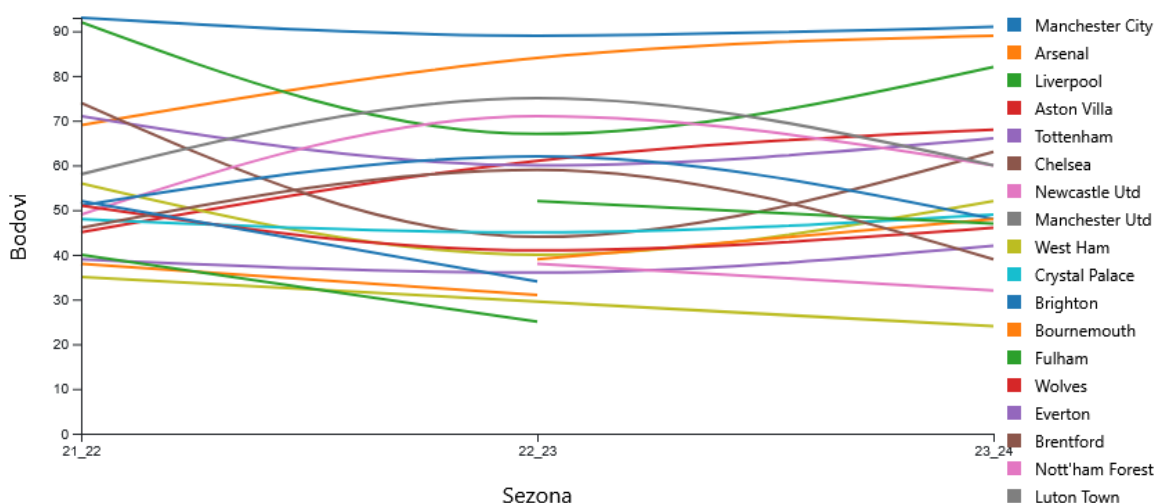
Na desnoj strani je legenda koja jasno identificira timove i njihove odgovarajuće boje linija. U slučaju da korisnik zbog velikog broja timova ne može dobro razaznati kretanje linija i broj bodova, dodan je Tooltip, pa tako prelaskom kursora preko svake linije korisnik dobiva detaljne informacije o timu i njegovim bodovima kroz sve sezone.

S obzirom da se radi o grafu s većim brojem varijabli, odnosno timova, nije se mogao zadržati trend jedne primarne boje kroz cijelu stranicu, nego je bilo potrebno izabrati što različiti boje kako bi se linije mogle vizualno razaznavati.

Za timove koji nisu sudjelovali sve tri sezone u ligi, vidi se kako je linija prekinuta, ili počinje tek na polovici. Prelaskom kursorom preko tih timova prikazuju se bodovi samo za sezone u kojima su sudjelovali.

Ovakav graf omogućava korisniku lako razumijevanje omjera snaga timova kroz prošle sezone. Npr. lako je vidljiva dominacija Manchester Cityja kroz prošle 3 sezone, kao i uspon Arsenala, ali i npr. pad timova poput Leicestera koji su ispali iz lige nedugo nakon njenog osvajanja.

Mana grafa je što prikazuje samo osvojene bodove na kraju sezone, dok bi bilo još zanimljivije pratiti kretanje broja bodova kroz sezonu, no to bi zahtjevalo puno kompliciraniji skup podataka.



4. Moguće nadogradnje

Zbog pregustog rasporeda na faksu, nisam bio u mogućnosti realizirati još neke ideje koje sam imao kako bi poboljšao ovu stranicu kako bi ona bila što zanimljivija i zapravo korisna u stvarnom životu.

Prva potencijalna nadogradnja može biti proširenje skupa podataka tako da se skupe podaci po svim kolima, kako bi se mogao pratiti razvoj tima kroz sezonu. Također bi bilo zanimljivo uspoređivati gdje se koji tim nalazio u prošlim sezonama u odnosu na trenutnu.

Nadalje, što se tiče samog izgleda stranice, reaktivnost nije optimizirana, pa bi jedan od prioritetnih nadogradnji bio CSS, točnije prilagođavanje pozicije elemenata za uređaje različitih dimenzija.

Ono što bi korisnicima moglo biti korisno i zanimljivo su i podaci o individualnim igračima, odnosno dodatni skup podataka koji opisuje igrače i njihove statistike. Fokusirao bih se na neke korisne stvari za igrače Fantasy footballa, poput osvojenih bodova, forme igrača, rasporeda nadolazećih utakmica, broja kartona, "clean sheetova", metrika za opasnost po gol i slično. Uspoređujući takve metrike s cijenom igrača na Fantasyu se mogu kreirati neke zanimljive vizualizacije koje bi bile od pomoći korisnicima.