**Uniwersytet Rzeszowski**

**Wydział Nauk Ścisłych i Technicznych**

**Instytut Informatyki**

**Krzysztof Motas**

***Analiza piłkarzy nominowanych do Złotej Piłki 2024 – grafowa baza danych***

**Pierwszy projekt zaliczeniowy z Sieci semantycznych**

Prowadzący: dr inż. Wojciech Kozioł

Rzeszów 2025

Spis treści

[1. Opis bazy danych 3](#_Toc197886515)

[2. Źródła danych 3](#_Toc197886516)

[3. Proces pozyskiwania i integracji danych 3](#_Toc197886517)

[4. Struktura bazy danych 5](#_Toc197886518)

[5. Nietrywialne zapytania w języku Cypher 15](#_Toc197886519)

[6. Możliwości wykorzystania bazy danych 26](#_Toc197886520)

[7. Podsumowanie i możliwości rozwoju bazy danych 27](#_Toc197886521)

1. Opis bazy danych

Baza danych zawiera szczegółowe informacje o piłkarzach nominowanych do Złotej Piłki 2024, jednej z najbardziej prestiżowych nagród indywidualnych w światowym futbolu. Dane obejmują statystyki meczowe, przebieg kariery klubowej, osiągnięcia indywidualne oraz rozszerzone informacje biograficzne zawodników.

**Zakres przechowywanych danych:**

* Statystyki z sezonu 2023/2024, obejmujące m.in. gole, asysty, expected goals (xG), expected assists (xAG) oraz inne wskaźniki.
* Dane biograficzne, takie jak data urodzenia, wzrost, narodowość, aktualny klub, pozycja na boisku oraz preferowana noga.
* Informacje o przebiegu kariery: historia transferów, szczegóły kontraktowe, agenci reprezentujący zawodników oraz sponsorzy sprzętu sportowego.
* Osiągnięcia indywidualne, w tym zdobyte nagrody oraz wyniki głosowania w plebiscycie Złotej Piłki 2024 (punkty i miejsce w rankingu).
* Relacje pomiędzy zawodnikami a klubami, krajami, agentami, sponsorami oraz miejscami urodzenia, odzwierciedlające sieciowe powiązania w karierach piłkarzy.

1. Źródła danych

Baza danych została utworzona na podstawie danych pochodzących z trzech głównych źródeł, obejmujących zarówno statystyki meczowe, informacje biograficzne i transferowe zawodników, jak i oficjalne wyniki głosowania w plebiscycie Złotej Piłki 2024.

* **Kaggle:** <https://www.kaggle.com/datasets/willianoliveiragibin/ballon-dor-2024>

Zbiór zawierający szczegółowe statystyki meczowe nominowanych zawodników za sezon 2023/2024 w największych europejskich ligach (m.in. gole, asysty, xG, xAG, progresje piłki).

* **Transfermarkt:** <https://www.transfermarkt.pl/>

Dane pobrane za pomocą dynamicznego web scrapingu, obejmujące historię transferową, dane kontraktowe, nagrody, a także profil zawodnika (m.in. wzrost, pozycja, preferowana noga, agenci, outfitterzy).

* **Wikipedia:** <https://en.wikipedia.org/wiki/2024_Ballon_d%27Or>

Oficjalne dane dotyczące liczby punktów i miejsc zajętych przez zawodników w  głosowaniu Ballon d'Or 2024.

1. Proces pozyskiwania i integracji danych

W niniejszym rozdziale przedstawiono narzędzia oraz biblioteki wykorzystane na etapie pozyskiwania, przetwarzania i importowania danych. Opisano również strukturę przygotowanych plików oraz kolejność wykonywanych operacji niezbędnych do załadowania informacji do bazy danych.

* 1. **Wykorzystane pliki**

Poniżej znajduje się lista skryptów, które zostały przygotowane i wykorzystane w projekcie do pobrania, przetwarzania oraz importu danych do bazy Neo4j.

* **main\_scrapper.py** – główny skrypt pobierający szczegółowe dane zawodników ze strony Transfermarkt (profil zawodnika, historia transferów, nagrody).
* **add\_stats.py** – skrypt łączący podstawowe dane zawodników z rozszerzonymi statystykami z serwisu Kaggle.
* **add\_rank.py** – skrypt dodający dane o liczbie punktów i miejscu w głosowaniu Ballon d'Or 2024 z Wikipedii.
* **import\_players\_to\_neo4j.py** – skrypt ładujący przetworzone dane do bazy danych Neo4j i tworzący odpowiednie węzły oraz relacje.
  1. **Wykorzystane biblioteki i technologie**

Do realizacji projektu niezbędne było wykorzystanie bibliotek wspomagających pobieranie, przetwarzanie oraz zapis danych. Poniżej przedstawiono wszystkie biblioteki, które zostały wykorzystane.

* **requests** – pobieranie danych z internetu (Wikipedia, Transfermarkt)
* **BeautifulSoup (bs4)** – parsowanie HTML, wyszukiwanie danych w strukturze stron
* **pandas** – obsługa i przetwarzanie danych z plików CSV
* **selenium** – automatyzacja przeglądarki do pobierania dynamicznie generowanych danych (Transfermarkt)
* **Neo4j Python Driver (neo4j)** – bezpośrednia komunikacja z bazą grafową Neo4j
* **json** – zapis i odczyt danych w formacie JSON
* **re** – obsługa wyrażeń regularnych (czyszczenie danych tekstowych)
* **unicodedata** – usuwanie akcentów w nazwach zawodników dla lepszego dopasowania danych
  1. **Opis procesu pozyskiwania i przetwarzania danych**

Przygotowanie i przetworzenie odbywa odbywało się etapami, zgodnie z poniższym planem. Każdy krok procesu opiera się na osobnym skrypcie odpowiedzialnym za konkretny zakres operacji.

1. **Pobranie podstawowych danych** za pomocą skryptu main\_scrapper.py, który pobiera szczegółowe dane zawodników z Transfermarkt, w tym dane biograficzne, historię transferów i nagrody.
2. **Dodanie statystyk meczowych** przy użyciu skryptu add\_stats.py, który uzupełnia informacje o zawodnikach na podstawie danych z serwisu Kaggle.
3. **Dodanie danych rankingowych** z Wikipedii przy wykorzystaniu skryptu add\_rank.py.
4. **Import danych do bazy Neo4j** przy użyciu skryptu import\_players\_to\_neo4j.py, który tworzy węzły i relacje w bazie danych.
   1. **Instrukcja uruchomienia**

Aby poprawnie zbudować bazę danych i załadować do niej przetworzone informacje, należy wykonać poniższe polecenia w terminalu. Każde z nich uruchamia osobny skrypt odpowiedzialny za wybrany etap przetwarzania danych.

* python main\_scrapper.py
* python add\_stats.py
* python add\_rank.py
* python import\_players\_to\_neo4j.py

Przed uruchomieniem ostatniego skryptu (import\_players\_to\_neo4j.py) należy upewnić się, że baza danych Neo4j jest uruchomiona i dostępna pod adresem bolt://localhost:7687, a  użytkownik i hasło są ustawione zgodnie z konfiguracją (domyślnie neo4j / the-golden-ball).

1. Struktura bazy danych

Baza danych zawiera informacje, które zostały odwzorowane w formie grafu, w którym poszczególne elementy reprezentowane są jako węzły, a relacje między nimi jako krawędzie. Baza umożliwia analizę karier zawodników, ich osiągnięć oraz transferów w ujęciu sieciowym.

* 1. **Węzły**

Każdy z opisanych typów węzłów został zilustrowany przykładem pochodzącym bezpośrednio z bazy danych Neo4j, prezentującym jego właściwości.

* **Player** – zawodnik. Jego właściwości to np. imię i nazwisko, data urodzenia, wzrost, aktualny klub, punkty zdobyte w głosowaniu, miejsce w rankingu Ballon d'Or, statystyki sezonowe (gole, asysty, xG, xAG itp.).

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 1 Przykład węzła Player

* **Club** – klub piłkarski

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 2 Przykład węzła Club

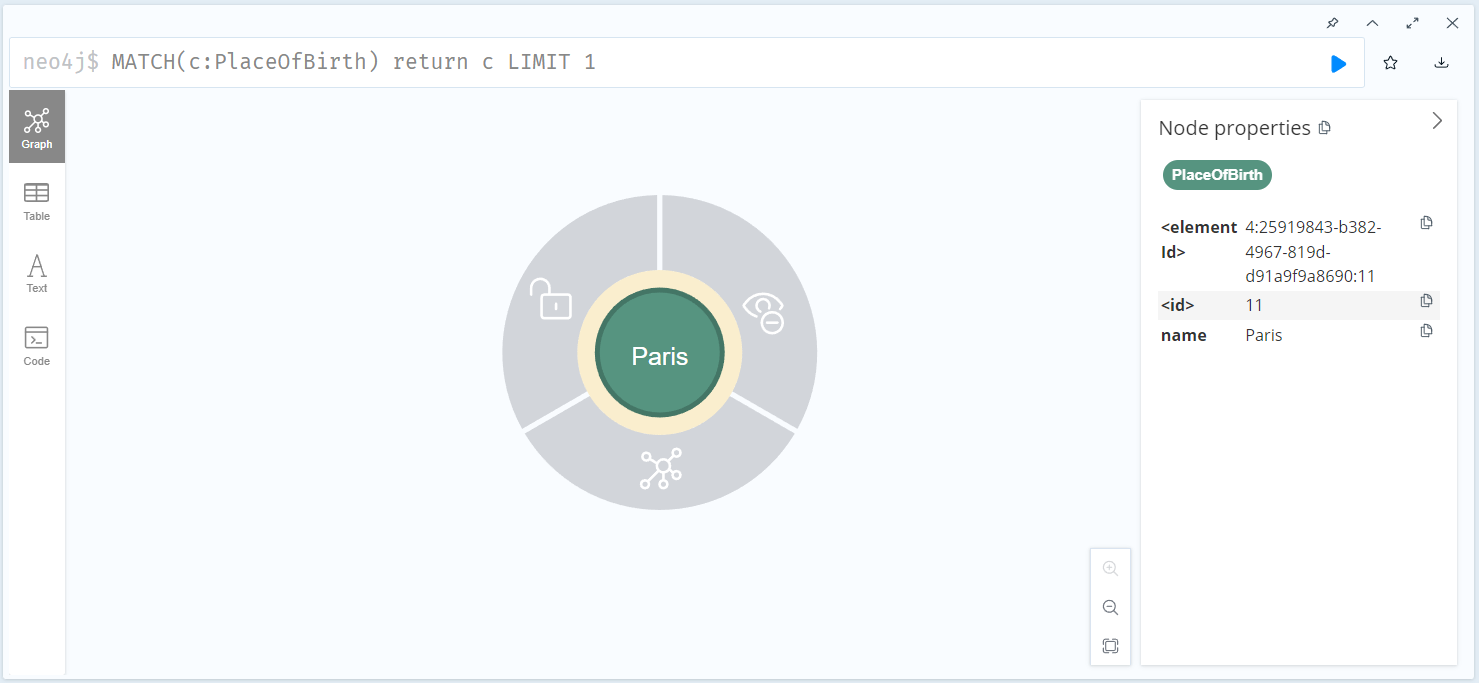
* **Country** – kraj, reprezentujący obywatelstwo zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 3 Przykład węzła Country

* **PlaceOfBirth** – miejsce urodzenia zawodnika



Rysunek. 4 Przykład węzła PlaceOfBirth

* **Position** – pozycja, na której zawodnik występuje na boisku

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 5 Przykład węzła Position

* **Foot** – dominująca noga zawodnika (lewa lub prawa)

Obraz zawierający zrzut ekranu, tekst, oprogramowanie, Oprogramowanie multimedialne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 6 Przykład węzła Foot

* **Outfitter** – sponsor sprzętu sportowego zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 7 Przykład węzła Outfitter

* **Agent** – agent piłkarski reprezentujący zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 8 Przykład węzła Agent

* **Award** – nagroda. Jego właściwości to tytuł nagrody, rok zdobycia, klub (jeżeli dotyczy)

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 9 Przykład węzła Award

* 1. **Relacje**

Dla każdego z wymienionych typów relacji przedstawiono przykład bezpośrednio z bazy danych Neo4j, obrazujący sposób powiązania węzłów danego rodzaju.

* **(:Player)-[:BORN\_IN]->(:PlaceOfBirth)** – miejsce urodzenia zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 10 Przykład relacji BORN\_IN

* **(:Player)-[:HAS\_CITIZENSHIP]->(:Country)** – obywatelstwo zawodnika

Obraz zawierający tekst, oprogramowanie, Ikona komputerowa, System operacyjny

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 11 Przykład relacji HAS\_CITIZENSHIP

* **(:Player)-[:PLAYS\_AS]->(:Position)** – pozycja na boisku

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 12 Przykład relacji PLAYS\_AS

* **(:Player)-[:PLAYS\_WITH]->(:Foot)** – preferowana noga

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 13 Przykład relacji PLAYS\_WITH

* **(:Player)-[:SPONSORED\_BY]->(:Outfitter)** – sponsor sprzętu

Obraz zawierający tekst, oprogramowanie, Ikona komputerowa, Oprogramowanie multimedialne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 14 Przykład relacji SPONSORED\_BY

* **(:Player)-[:REPRESENTED\_BY]->(:Agent)** – agent reprezentujący zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 15 Przykład relacji REPRESENTED\_BY

* **(:Player)-[:WON]->(:Award)** – zdobyte nagrody

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 16 Przykład relacji WON

* **(:Player)-[:TRANSFERRED]->(:Club)** – historia transferów

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 17 Przykład relacji TRANSFERED

* **(:Player)-[:STARTED\_CAREER\_AT]->(:Club)** – pierwszy klub w karierze zawodnika

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 18 Przykład relacji STARTED\_CAREER\_AT

Relacje często zawierają **dodatkowe właściwości,** takie jak np. sezon transferu, data transferu, wartość rynkowa lub kwota transferu (dla TRANSFERRED).

* 1. **Liczebność bazy danych**

Liczebność poszczególnych węzłów i relacji przedstawia się następująco:

**Węzły:**

* Award: 580
* Club: 162
* Player: 30
* PlaceOfBirth: 27
* Agent: 23
* Country: 18
* Position: 11
* Outfitter: 6
* Foot: 2

**Całkowita liczba węzłów:** 859

**Relacje:**

* WON: 824
* TRANSFERRED: 221
* HAS\_CITIZENSHIP: 43
* BORN\_IN: 30
* PLAYS\_AS: 30
* PLAYS\_WITH: 30
* STARTED\_CAREER\_AT: 30
* SPONSORED\_BY: 29
* REPRESENTED\_BY: 27

**Całkowita liczba relacji:** 1264

* 1. **Ogólny wygląd bazy danych**

Na poniższym rysunku przedstawiono pełną strukturę bazy danych w formie grafu. Układ grafu przypomina rozległą sieć skupisk (klastrów), gdzie zawodnicy znajdują się w centrum swoich powiązań, a wokół nich rozmieszczone są powiązane podmioty.

Obraz zawierający fajerwerki, światło

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek. 19 Wygląd bazy danych

1. Nietrywialne zapytania w języku Cypher

W ramach projektu przygotowano zestaw nietrywialnych zapytań w języku Cypher, umożliwiających analizę zgromadzonych danych. Zapytania te pozwalają m.in. na ocenę efektywności zawodników, analizę ścieżek kariery, wpływu agentów na sukcesy piłkarzy, a także porównywanie osiągnięć sponsorów sprzętu sportowego.

* **Wyszukaj zawodników, którzy mają więcej niż jedno obywatelstwo i przynajmniej jedno z krajów to kraj spoza Europy.**

**MATCH** (p:Player)-[:HAS\_CITIZENSHIP]->(c:Country)

**WITH** p, collect(c.name) **AS** citizenships

**WHERE** size(citizenships) > 1

**AND** any(country **IN** citizenships **WHERE** **NOT** country **IN** [

'England', 'Spain', 'Germany', 'France', 'Italy', 'Portugal', 'Netherlands', 'Belgium', 'Norway',

'Sweden', 'Denmark', 'Switzerland', 'Austria', 'Croatia', 'Poland', 'Serbia', 'Ukraine', 'Turkey'

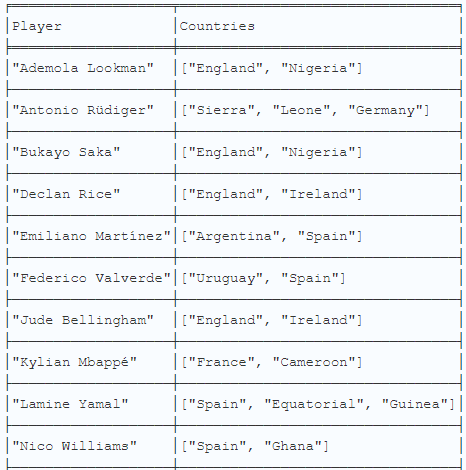
])

**RETURN**

p.name **AS** Player,

citizenships **AS** Countries

**ORDER** **BY** p.name



Rysunek 20. Rezultat działania zapytania

* **Wypisz sponsorów, których zawodnicy zdobyli największą liczbę nagród w karierze.**

**MATCH** (o:Outfitter)<-[:SPONSORED\_BY]-(p:Player)-[:WON]->(a:Award)

**WITH**

o.name **AS** Outfitter,

p.name **AS** Player,

count(a) **AS** PlayerAwardCount

**WITH**

Outfitter,

collect({player: Player, awards: PlayerAwardCount}) **AS** PlayersAwards,

sum(PlayerAwardCount) **AS** TotalAwards,

avg(PlayerAwardCount) **AS** AvgAwardsPerPlayer

**UNWIND** PlayersAwards **AS** pa

**WITH**

Outfitter,

TotalAwards,

AvgAwardsPerPlayer,

pa.player **AS** PlayerName,

pa.awards **AS** PlayerAwards

**WITH**

Outfitter,

TotalAwards,

AvgAwardsPerPlayer,

collect({name: PlayerName, awards: PlayerAwards}) **AS** AllPlayers,

max(PlayerAwards) **AS** MaxAwards

**WITH** Outfitter,

TotalAwards,

AvgAwardsPerPlayer,

[p **IN** AllPlayers **WHERE** p.awards = MaxAwards][0] **AS** TopPlayer

**RETURN**

Outfitter,

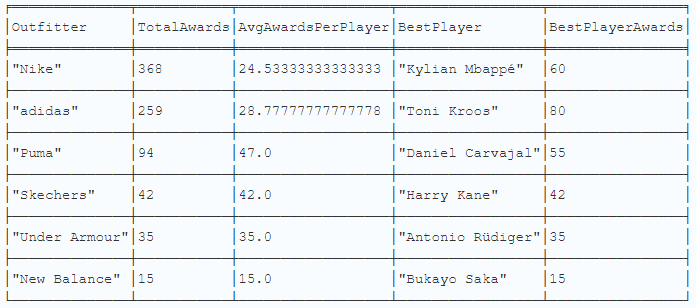
TotalAwards,

AvgAwardsPerPlayer,

TopPlayer.name **AS** BestPlayer,

TopPlayer.awards **AS** BestPlayerAwards

**ORDER** **BY** TotalAwards **DESC**



Rysunek 21. Rezultat działania zapytania

* **Wypisz kraje, których zawodnicy charakteryzują się najwyższym średnim wzrostem, podając przy tym liczbę zawodników, minimalny i maksymalny wzrost oraz rozrzut wzrostu w danym kraju.**

**MATCH** (p:Player)-[:HAS\_CITIZENSHIP]->(c:Country)

**WHERE** p.height IS **NOT** NULL

**WITH** c.name **AS** country,

toFloat(REPLACE(REPLACE(p.height, " m", ""), ",", ".")) **AS** height\_numeric

**WITH** country,

count(height\_numeric) **AS** players\_count,

avg(height\_numeric) **AS** avg\_height,

min(height\_numeric) **AS** min\_height,

max(height\_numeric) **AS** max\_height,

max(height\_numeric) - min(height\_numeric) **AS** height\_range

**RETURN** country, players\_count, avg\_height, min\_height, max\_height, height\_range

**ORDER** **BY** avg\_height **DESC**;

Obraz zawierający tekst, numer, Równolegle, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, numer, Równolegle, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rysunek 22. Rezultat działania zapytania

* **Wypisz zawodników, którzy rozegrali najwięcej minut, ale mieli minimalny lub żaden udział w golach i asystach. Zapytanie pozwala wykryć mało efektywnych ofensywnie graczy lub zidentyfikować zawodników o defensywnym profilu.**

**MATCH** (p:Player)

**WHERE** p.playing\_time\_min IS **NOT** NULL **AND** p.performance\_gls IS **NOT** NULL **AND** p.performance\_ast IS **NOT** NULL

**WITH** p, (toFloat(p.performance\_gls) + toFloat(p.performance\_ast)) **AS** ga, toFloat(p.playing\_time\_min) **AS** minutes

**WITH** p.name **AS** player, ga, minutes, (ga / minutes) **AS** ga\_per\_minute

**WHERE** ga\_per\_minute < 0.001

**RETURN** player, ga, minutes, ga\_per\_minute

**ORDER** **BY** ga\_per\_minute **ASC**

**LIMIT** 10;

**Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 23. Rezultat działania zapytania

* **Znajdź zawodników z największą liczbą kluczowych podań (key passes) per 90 minut, co pozwala określić, kto jest najlepszym kreatorem gry.**

**MATCH** (p:Player)

**WHERE** p.kp IS **NOT** NULL **AND** p.playing\_time\_90s IS **NOT** NULL

**WITH** p.name **AS** player,

toFloat(p.kp) **AS** key\_passes,

toFloat(p.playing\_time\_90s) **AS** games\_played,

toFloat(p.kp) / toFloat(p.playing\_time\_90s) **AS** kp\_per\_90

**RETURN** player, key\_passes, games\_played, kp\_per\_90

**ORDER** **BY** kp\_per\_90 **DESC**

**LIMIT** 10;

**Obraz zawierający tekst, numer, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 24. Rezultat działania zapytania

* **Wypisz zawodników o najbardziej wszechstronnym profilu: ci, którzy łączą skuteczność ofensywną, kreatywność, progresję piłki, grę w defensywie oraz dryblingi. Zapytanie pokazuje kompletne profile piłkarskie.**

**MATCH** (p:Player)

**WHERE** p.performance\_gls IS **NOT** NULL

**AND** p.performance\_ast IS **NOT** NULL

**AND** p.kp IS **NOT** NULL

**AND** p.carries\_prgdist IS **NOT** NULL

**AND** p.int IS **NOT** NULL

**AND** p.blocks\_blocks IS **NOT** NULL

**AND** p.take\_ons\_succ IS **NOT** NULL

**AND** p.playing\_time\_90s IS **NOT** NULL

**WITH** p.name **AS** player,

toFloat(p.performance\_gls) **AS** goals,

toFloat(p.performance\_ast) **AS** assists,

toFloat(p.kp) **AS** key\_passes,

toFloat(p.carries\_prgdist) **AS** progressive\_carries,

toFloat(p.int) **AS** interceptions,

toFloat(p.blocks\_blocks) **AS** blocks,

toFloat(p.take\_ons\_succ) **AS** successful\_dribbles,

toFloat(p.playing\_time\_90s) **AS** games\_played,

(toFloat(p.performance\_gls) + toFloat(p.performance\_ast)) / toFloat(p.playing\_time\_90s) **AS** ga\_per\_90,

toFloat(p.kp) / toFloat(p.playing\_time\_90s) **AS** kp\_per\_90,

toFloat(p.carries\_prgdist) / toFloat(p.playing\_time\_90s) **AS** carries\_per\_90,

(toFloat(p.int) + toFloat(p.blocks\_blocks)) / toFloat(p.playing\_time\_90s) **AS** defensive\_actions\_per\_90,

toFloat(p.take\_ons\_succ) / toFloat(p.playing\_time\_90s) **AS** dribbles\_per\_90

**WITH** player, ga\_per\_90, kp\_per\_90, carries\_per\_90, defensive\_actions\_per\_90, dribbles\_per\_90,

(ga\_per\_90 + kp\_per\_90 + carries\_per\_90 + defensive\_actions\_per\_90 + dribbles\_per\_90) **AS** total\_score

**ORDER** **BY** total\_score **DESC**

**LIMIT** 10;

**Obraz zawierający tekst, numer, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 25. Rezultat działania zapytania

* **Wypisz top 5 zawodników, którzy najczęściej wykonują progresywne podania oraz utrzymują wysoką skuteczność podań. Pokazuje kluczowych graczy odpowiedzialnych za kontrolę tempa gry.**

**MATCH** (p:Player)

**WHERE** p.total\_att IS **NOT** NULL

**AND** p.total\_cmp IS **NOT** NULL

**AND** p.progression\_prgp IS **NOT** NULL

**AND** p.playing\_time\_90s IS **NOT** NULL

**WITH** p.name **AS** player,

toFloat(p.total\_att) **AS** passes\_attempted,

toFloat(p.total\_cmp) **AS** passes\_completed,

toFloat(p.progression\_prgp) **AS** progressive\_passes,

toFloat(p.playing\_time\_90s) **AS** games\_played

**WITH** player, passes\_attempted, passes\_completed, progressive\_passes, games\_played,

passes\_completed / passes\_attempted **AS** pass\_accuracy,

progressive\_passes / games\_played **AS** progressive\_passes\_per\_90

**RETURN** player, pass\_accuracy, progressive\_passes\_per\_90

**ORDER** **BY** progressive\_passes\_per\_90 **DESC**, pass\_accuracy **DESC**

**LIMIT** 5;

**Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 26. Rezultat działania zapytania

* **Wypisz zawodników, którzy zdobyli najwięcej nagród grając w najmniejszej liczbie klubów w karierze.**

**MATCH** (p:Player)-[:TRANSFERRED]->(c:Club)

**WITH** p, count(DISTINCT c) **AS** clubs\_played

**MATCH** (p)-[:WON]->(a:Award)

**WITH** p.name **AS** player, clubs\_played, count(a) **AS** awards\_won

**RETURN** player, clubs\_played, awards\_won

**ORDER** **BY** awards\_won **DESC**, clubs\_played **ASC**

**LIMIT** 10;

**Obraz zawierający tekst, zrzut ekranu, numer, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 27. Rezultat działania zapytania

* **Wypisz zawodników, których liczba zdobytych bramek i asyst (G+A) jest niższa niż przewidywana suma Expected Goals i Expected Assists (xG + xA). Zapytanie pokazuje także ile gracz "powinien mieć" (xG i xA osobno) oraz ile faktycznie ma (gole i asysty).**

**MATCH** (p:Player)

**WHERE** p.expected\_xg IS **NOT** NULL

**AND** p.expected\_xag IS **NOT** NULL

**AND** p.performance\_gls IS **NOT** NULL

**AND** p.performance\_ast IS **NOT** NULL

**WITH** p.name **AS** player,

toFloat(p.expected\_xg) **AS** expected\_goals,

toFloat(p.expected\_xag) **AS** expected\_assists,

toFloat(p.performance\_gls) **AS** actual\_goals,

toFloat(p.performance\_ast) **AS** actual\_assists,

(toFloat(p.expected\_xg) + toFloat(p.expected\_xag)) **AS** expected\_contributions,

(toFloat(p.performance\_gls) + toFloat(p.performance\_ast)) **AS** actual\_contributions

**WITH** player, expected\_contributions, actual\_contributions,

expected\_goals, expected\_assists, actual\_goals, actual\_assists,

actual\_contributions - expected\_contributions **AS** difference

**WHERE** difference < 0

**RETURN** player, expected\_goals, expected\_assists, actual\_goals, actual\_assists, expected\_contributions, actual\_contributions, difference

**ORDER** **BY** difference **ASC**

**LIMIT** 10;

**Obraz zawierający tekst, numer, zrzut ekranu, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 28. Rezultat działania zapytania

* **Wypisz zawodników, którzy zdobyli nagrody w sezonie 23/24 i posortuj ich według efektywności — czyli liczby minut przypadających na jedną nagrodę. Zapytanie pokazuje także listę zdobytych nagród oraz łączną liczbę minut zawodnika. Im niższa wartość minut na nagrodę, tym bardziej zawodnik był doceniany mimo mniejszej liczby minut na boisku.**

**MATCH** (p:Player)-[:WON]->(a:Award)

**WHERE** p.playing\_time\_min IS **NOT** NULL

**AND** a.year = "23/24"

**WITH** p, collect(DISTINCT a.title) **AS** awards\_titles, count(a) **AS** awards\_count, toFloat(p.playing\_time\_min) **AS** minutes

**WITH** p.name **AS** player, awards\_titles, awards\_count, minutes, minutes / awards\_count **AS** minutes\_per\_award

**RETURN** player, awards\_titles **AS** awards, awards\_count, minutes, minutes\_per\_award

**ORDER** **BY** minutes\_per\_award **ASC**

**LIMIT** 10;

**Obraz zawierający tekst, zrzut ekranu, linia, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Rysunek 29. Rezultat działania zapytania

1. Możliwości wykorzystania bazy danych

Baza danych umożliwia analizę i porównywanie statystyk zawodników nominowanych do  Złotej Piłki 2024. Pozwala badać wpływ transferów oraz zmian klubów na osiągnięcia zawodników w sezonie. Umożliwia śledzenie przebiegu kariery zawodników w ujęciu czasowym.  
Na podstawie zgromadzonych danych możliwe jest przewidywanie przyszłych wyników i  rankingów. Baza daje możliwość analizy powiązań pomiędzy agentami a sukcesami reprezentowanych zawodników. Pozwala na wizualizację ścieżek kariery zawodników w postaci grafu. Może być również podstawą do budowy aplikacji rekomendacyjnych, np. wskazujących zawodników najlepiej pasujących do określonego klubu.

1. Podsumowanie i możliwości rozwoju bazy danych

Projekt umożliwia szczegółową analizę danych o czołowych piłkarzach sezonu 2023/2024, obejmującą zarówno ich osiągnięcia, jak i przebieg kariery. Dzięki integracji danych statystycznych, biograficznych i transferowych w modelu grafowym możliwe jest przeprowadzanie analiz oraz zapytań.

W ramach dalszego rozwoju bazy danych możliwe jest:

* Automatyczne aktualizowanie danych o zawodnikach wraz z pojawianiem się nowych sezonów.
* Rozszerzenie bazy o dane dotyczące drużyn narodowych oraz występów w  międzynarodowych turniejach, takich jak Euro 2024 czy Copa América.
* Dodanie informacji o historii kontuzji zawodników.
* Integracja z zewnętrznymi źródłami danych, takimi jak API FIFA czy UEFA.