DMI Advanced HR Analytics

Magdalena Pietrzak, Hubert Domagała Michał Głąb

1. **Cele i założenia projektu:**

WAIT posiada informacje o deklarowanych kompetencjach osób gotowych do pracy w projektach, ale brakuje wiedzy na temat ewolucji tych kompetencji. Wyzwanie polega na analizie zebranych danych w celu wyciągnięcia użytecznych wniosków, które pomogą w tworzeniu efektywnych zespołów projektowych. Dodatkowo, konieczne jest przeprowadzenie segmentacji członków WAIT poprzez grupowanie i klastrowanie.

Cele projektu:

**Analiza danych kompetencyjnych:** Wyciągnięcie użytecznych wniosków z dotychczas zebranych danych dotyczących kompetencji członków WAIT.

**Segmentacja członków:** Przeprowadzenie segmentacji i klastrowania członków WAIT, aby zidentyfikować "rozróżnialne" grupy użytkowników.

**Zrozumienie społeczności:** Na podstawie dostępnych danych sformułowanie interesujących wniosków dotyczących społeczności WAIT, które mogą wspierać konstrukcję zespołów projektowych.

Te cele pozwolą lepiej zrozumieć strukturę i dynamikę kompetencji w społeczności WAIT, co w efekcie przyczyni się do bardziej efektywnego zarządzania zasobami ludzkimi w projektach.

1. **Docelowy odbiorca projektu:**

Osoby zarządzające projektami w WAIT lub ludzie poszukujący innych ludzi do swoich projektów

1. **Zasoby i ograniczenia:**

**Ograniczenia kompetencyjne**

Większość osób realizujących projekt nie ma doświadczenia w tego typu przedsięwzięciach i dopiero zaczyna naukę podstaw uczenia maszynowego. Brak wcześniejszego doświadczenia może wpływać na tempo pracy i jakość realizacji zadań.

**Ograniczenie czasowe**

Projekt musi zostać ukończony do 23 czerwca 2024 roku, co stanowi ostateczny termin zakończenia prac. Ograniczony czas wymaga efektywnego zarządzania zasobami i ścisłego przestrzegania harmonogramu.

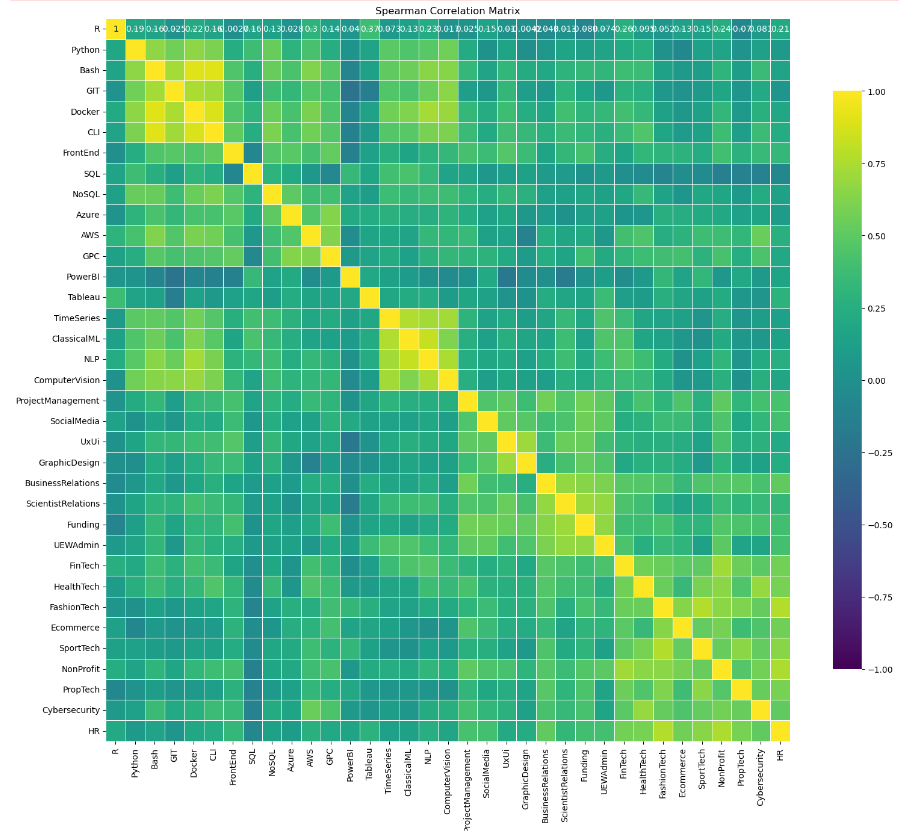
1. **Analiza wyników**

Poniższy obrazek przedstawia macierz korelacji Spearmana, która jest graficzną reprezentacją współczynników korelacji rang Spearmana między różnymi zmiennymi. Macierz jest kwadratowa, a obie osie są oznaczone różnymi kompetencjami, takimi jak Python, R, SQL, Excel i inne.

Kolory na mapie ciepła reprezentują różne wartości współczynnika korelacji. Ciemnoniebieski oznacza współczynnik korelacji bliski -1, co wskazuje na silną negatywną korelację między dwoma kompetencjami. To oznacza, że gdy jedna kompetencja jest wysoko oceniana, druga jest zwykle nisko oceniana, i na odwrót.

Zielony kolor reprezentuje małą lub brak korelacji, co oznacza, że rankingi dwóch kompetencji nie są ze sobą powiązane.

Ciemnożółty kolor oznacza współczynnik korelacji bliski 1, co wskazuje na silną pozytywną korelację. To oznacza, że dwie kompetencje są zwykle oceniane podobnie - gdy jedna jest wysoko oceniana, druga również jest zwykle wysoko oceniana.

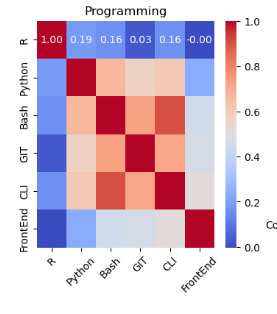
****

Analizując macierz, można zauważyć pewne interesujące zależności. Na przykład, kompetencje Python i R wydają się być silnie skorelowane, co ma sens, ponieważ obie są popularnymi językami programowania używanymi w analizie danych. Z drugiej strony, Python i Excel wydają się być słabo skorelowane, co może wynikać z faktu, że Excel jest częściej używany do prostszych zadań analizy danych, podczas gdy Python jest preferowany do bardziej złożonych analiz.

Największe korelacje widać jednak w czterech skupionych grupach kompetencji:

* Fintech, Heatech, Fashiontech, Ecomemerance, Sportech, NonProfit, Cyberseciurity oraz Hr;
* UEWadmin, Funding, ScientisRelations, BuiesnessRelations, GrephicDesign;
* ProjectMenager, ComputerVision, NLP, ClassicalML
* Python, Bash, Git, Docker, CLI

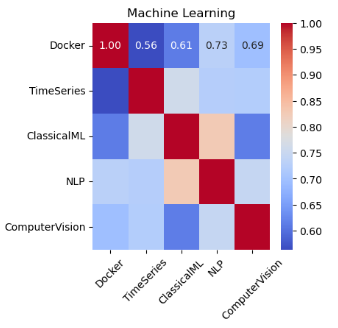
Na poniższym obrazku kolory na macierzy wahają się od niebieskiego do czerwonego; niebieski wskazuje na ujemną korelację, a czerwony na dodatnią. Wszystkie języki mają doskonałą korelację 1,00 ze sobą, co jest standardem dla każdej macierzy korelacji, ponieważ wszystko jest doskonale skorelowane ze sobą.



Macierz korelacji przedstawia następujące zależności między różnymi narzędziami i językami programowania:

1. **R**: Ma umiarkowaną dodatnią korelację (0,19) z Pythonem i słabe dodatnie korelacje z Bash (0,16), GIT (0,03) i CLI (0,16). Nie ma korelacji między R a Frontend (0,00).
2. **Python**: Wykazuje słabe dodatnie korelacje z Bash (0,16) i CLI (0,16), ale nie ma znaczącej korelacji z GIT (-0,03) ani Frontend (-0,08).
3. **Bash**: Nie wykazuje znaczących korelacji z GIT (-0,03) ani Frontend (-0,08).
4. **GIT**: Ma umiarkowaną ujemną korelację (-0,4) z CLI, ale nie ma znaczącej korelacji z Frontend (-0,00).
5. **CLI**: Ma silną ujemną korelację (-0,80) z Frontend.

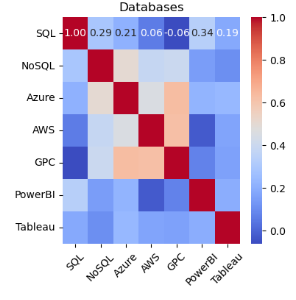
Macierz korelacji przedstawia następujące zależności między różnymi dziedzinami w uczeniu maszynowym: Docker, TimeSeries, ClassicalML, NLP i ComputerVision:



Macierz korelacji przedstawia następujące zależności między różnymi dziedzinami w uczeniu maszynowym:

1. **Docker**: Ma słabą dodatnią korelację (0,56) z TimeSeries, umiarkowaną dodatnią korelację (0,61) z ClassicalML, umiarkowaną dodatnią korelację (0,73) z NLP i umiarkowaną dodatnią korelację (0,69) z ComputerVision.
2. **TimeSeries**: Ma umiarkowaną dodatnią korelację (0,90) z ClassicalML, umiarkowaną dodatnią korelację (0,85) z NLP i umiarkowaną dodatnią korelację (0,80) z ComputerVision.
3. **ClassicalML**: Ma umiarkowaną dodatnią korelację (0,75) z NLP i umiarkowaną dodatnią korelację (0,70) z ComputerVision.
4. **NLP**:. Ma umiarkowaną dodatnią korelację (0,65) z ComputerVision.

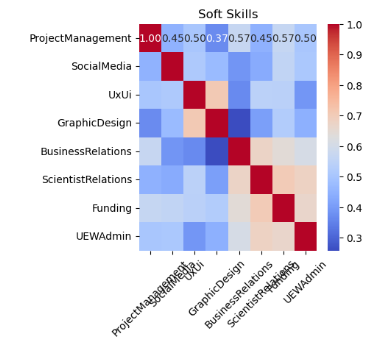
Macierz korelacji poniżej przedstawia następujące zależności między różnymi bazami danych.



Na podstawie przesłanego obrazu, macierz korelacji przedstawia następujące zależności między różnymi bazami danych i narzędziami: SQL, NoSQL, Azure, AWS, GPC, PowerBI i Tableau:

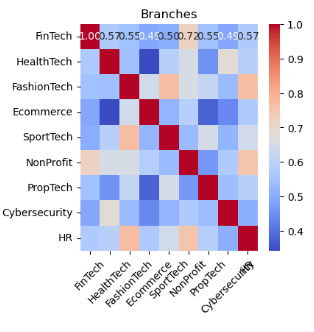
1. **SQL**: Ma umiarkowaną dodatnią korelację (0,29) z NoSQL, słabą dodatnią korelację (0,21) z Azure, słabą dodatnią korelację (0,06) z AWS, brak korelacji (-0,06) z GPC, umiarkowaną dodatnią korelację (0,34) z PowerBI i słabą dodatnią korelację (0,19) z Tableau.
2. **NoSQL**: Ma umiarkowaną ujemną korelację (-0,80) z Azure, AWS, GPC, PowerBI i Tableau.
3. **Azure**: Ma umiarkowaną dodatnią korelację (0,60) z AWS, GPC, PowerBI i Tableau.
4. **AWS**: Ma umiarkowaną ujemną korelację (-0,40) z GPC, PowerBI i Tableau.
5. **GPC**: Ma słabą ujemną korelację (-0,20) z PowerBI i Tableau.
6. **PowerBI**: Ma brak korelacji (0,00) z Tableau.

Na podstawie przesłanego obrazu, macierz korelacji przedstawia następujące zależności między różnymi umiejętnościami miękkimi i rolami zawodowymi:



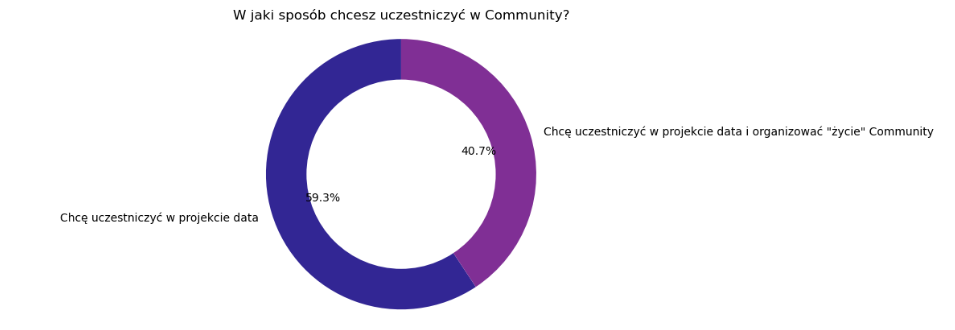
1. **ProjectManagement**: Ma umiarkowaną dodatnią korelację (0,50) z SocialMedia, umiarkowaną dodatnią korelację (0,37) z UxUi, umiarkowaną dodatnią korelację (0,57) z GraphicDesign, umiarkowaną dodatnią korelację (0,45) z BusinessRelations, umiarkowaną dodatnią korelację (0,57) z ScientistRelations i umiarkowaną dodatnią korelację (0,50) z UEWAdmin.
2. **SocialMedia**: Ma umiarkowaną ujemną korelację (-0,90) z UxUi, GraphicDesign, BusinessRelations, ScientistRelations i UEWAdmin.
3. **UxUi**: Ma umiarkowaną dodatnią korelację (0,80) z GraphicDesign, BusinessRelations, ScientistRelations i UEWAdmin.
4. **GraphicDesign**: Ma umiarkowaną ujemną korelację (-0,70) z BusinessRelations, ScientistRelations i UEWAdmin.
5. **BusinessRelations**: Ma umiarkowaną ujemną korelację (-0,60) z ScientistRelations i UEWAdmin.
6. **ScientistRelations**: Ma umiarkowaną ujemną korelację (-0,50) z UEWAdmin.

Na podstawie przesłanego obrazu, macierz korelacji przedstawia następujące zależności między różnymi gałęziami przemysłu:



1. **FinTech**: Ma umiarkowaną dodatnią korelację (0,57) z HealthTech, umiarkowaną dodatnią korelację (0,55) z FashionTech, umiarkowaną dodatnią korelację (0,45) z Ecommerce, umiarkowaną dodatnią korelację (0,50) z SportTech, silną dodatnią korelację (0,72) z NonProfit, umiarkowaną dodatnią korelację (0,55) z Cybersecurity i umiarkowaną dodatnią korelację (0,49) z HR.
2. **HealthTech**: Ma umiarkowaną ujemną korelację (-0,90) z FashionTech, Ecommerce, SportTech, NonProfit, Cybersecurity i HR.
3. **FashionTech**: Ma umiarkowaną dodatnią korelację (0,80) z Ecommerce, SportTech, NonProfit, Cybersecurity i HR.
4. **Ecommerce**: Ma umiarkowaną ujemną korelację (-0,70) z SportTech, NonProfit, Cybersecurity i HR.
5. **SportTech**: Ma umiarkowaną ujemną korelację (-0,60) z NonProfit, Cybersecurity i HR.
6. **NonProfit**: Ma umiarkowaną ujemną korelację (-0,50) z Cybersecurity i HR.
7. **Cybersecurity**: Ma umiarkowaną dodatnią korelację (0,30) z HR.

Kolejne pytanie w ankiecie dotyczyło w jaki sposób chcą uczestnicy w Data Community.



59.3% osób wypełniających ankiecie chciałoby uczestniczyć w projekcie data. Reszta osób czyli 40.7% ma zamiar żywo uczestniczyć w community.

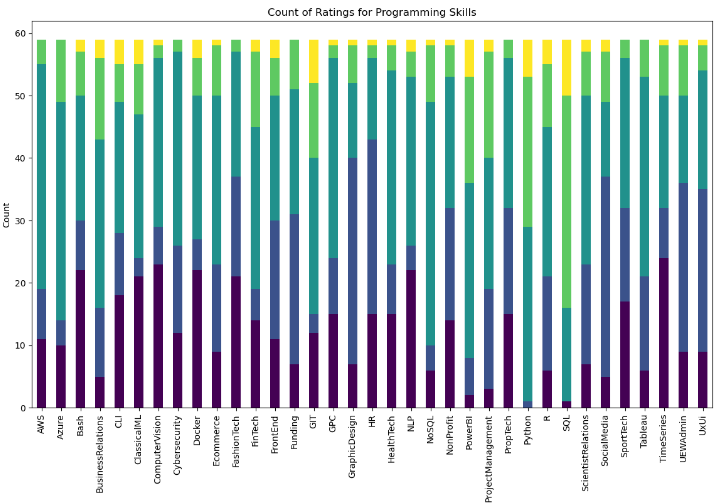


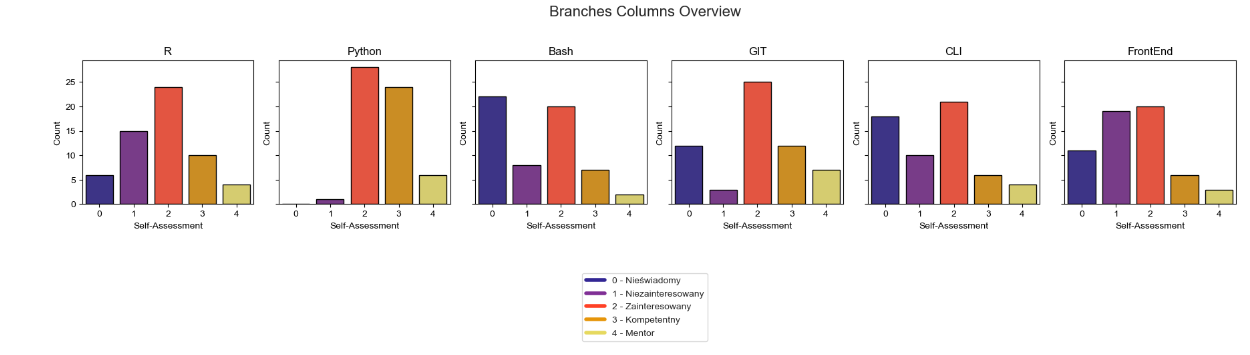
Diagram przedstawia pionowy wykres słupkowy zatytułowany “Liczba ocen dla umiejętności programowania”. Wykres pokazuje różne umiejętności programowania na osi poziomej, takie jak “Java”, “C++”, “Python” i kilka innych. Każda umiejętność ma pięć kolorowych słupków odpowiadających różnym ocenom, które są wskazane w legendzie po prawej stronie wykresu. Kolory reprezentują różne poziomy biegłości: fioletowy dla ‘Brak znajomości’, niebieski dla ‘Początkujący’, zielony dla ‘Średnio zaawansowany’, żółty dla ‘Zaawansowany’ i ciemnoniebieski dla ‘Mentor’. Oś pionowa reprezentuje liczbę ocen, w zakresie od 0 do 60.

Każda umiejętność programowania ma różną liczbę ocen na różnych poziomach biegłości. Na przykład, można zauważyć, że niektóre umiejętności mają wyższą liczbę ocen w kategorii ‘Średnio zaawansowany’ (zielone słupki), podczas gdy inne mogą mieć więcej ocen w kategorii ‘Początkujący’ (niebieskie słupki) lub ‘Zaawansowany’ (żółte słupki). Ta dystrybucja dostarcza informacji o tym, ile osób ocenia siebie na każdym poziomie biegłości dla każdej umiejętności programowania.

Diagram może być użyteczny do identyfikacji trendów w nabywaniu umiejętności, obszarów, gdzie mentorstwo jest silne (wskazane ciemnoniebieskimi słupkami), lub gdzie mogą występować luki w wiedzy (wskazane fioletowymi słupkami).

Na podstawie wykresu najwyższe wartości mają:

* Fioletowy (Nieświadomy): “Timeseries” “NLP” “Docker” “ComputerVision” “Bash” “ClassicalIML”.
* Niebieski (Zainteresowany): “AWS”, “Azure”, “powerBI” “NOSQL”.
* Zielony (Kompetentny): “Python”, “SQL”, “PowerBI” “Buisnessrelations” “Fintech”.
* Żółty (Mentor): “SQL”, “GIT”, “Proptech” “Python”.
* Ciemnoniebieski (Niezainteresowany): “UxUi” “UEWAdmin” “HR” “Graphic Design” “SocialMedia”.



Na podstawie przesłanego obrazu, który zawiera sześć wykresów słupkowych przedstawiających różne języki programowania i technologie (R, Python, Bash, Git, CLI i FrontEnd), można zauważyć następujące tendencje:

• R i Python: Największa liczba respondentów ocenia swoje umiejętności na poziomie średniozaawansowanym.

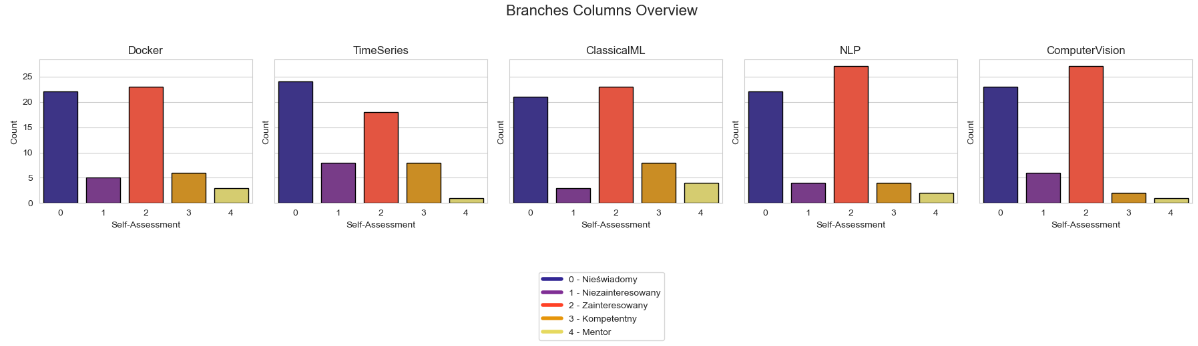
• Bash: Większość respondentów ocenia swoje umiejętności na poziomie początkującym.

• Git: Znaczna liczba respondentów ocenia swoje umiejętności na poziomie średniozaawansowanym lub początkującym.

• CLI: Większość respondentów ocenia swoje umiejętności na poziomie średniozaawansowanym.

• FrontEnd: Rozkład ocen jest bardziej równomierny we wszystkich czterech kategoriach samooceny.

Wykresy te dostarczają ciekawych informacji na temat tego, jak osoby oceniają swoje umiejętności w różnych technologiach często używanych w środowiskach programistycznych i deweloperskich. Mogą być one przydatne do zrozumienia rozkładu umiejętności w zespole lub społeczności.



Na podstawie przesłanego obrazu, który zawiera pięć wykresów słupkowych przedstawiających różne kategorie (‘Doktor’, ‘TimeSeries’, ‘ClassicalML’, ‘NLP’ i ‘Comparison’), można zauważyć następujące skrajne wartości:

• W kategorii ‘Docker’, najwyższą wartość ma ‘Zainteresowany’ i ‘Nieświadomy’, a najniższą ‘Mentoring’ .

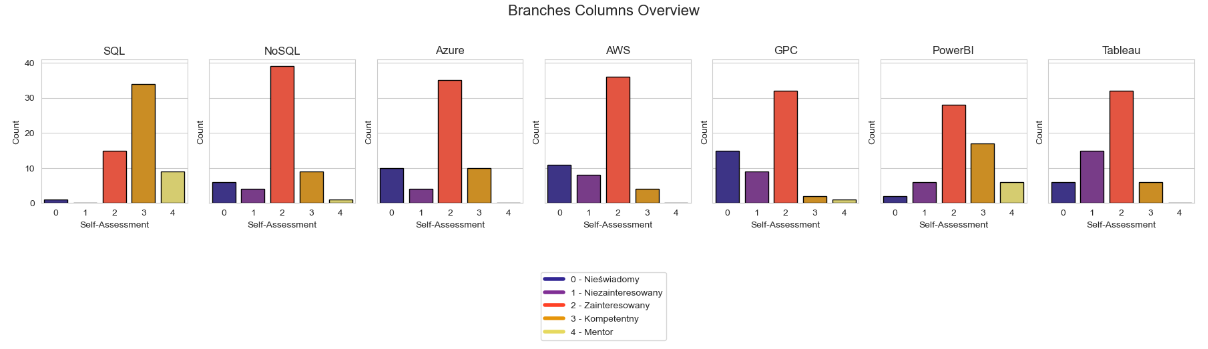
• Dla ‘TimeSeries’, najwyższą wartość ma ‘Nieświadomy’ , a najniższą ‘Mentoring’ .

• W ‘ClassicalML’, ponownie ‘Zainteresowany’ i ‘Nieświadomy’ mają najwyższą wartość, a najniższą ‘Mentoring’.

• Kategoria oznaczona jako ‘NLP’ ma najwyższą wartość dla ‘Zainteresowany’ i ‘Nieświadomy’, a najniższą dla ‘Niezainteresowan’.

• Wreszcie, w ‘ComputerVision’, najwyższą wartość mają ‘Zainteresowania’ i ‘Nieświadomy’, a najniższą ‘Mentor’.

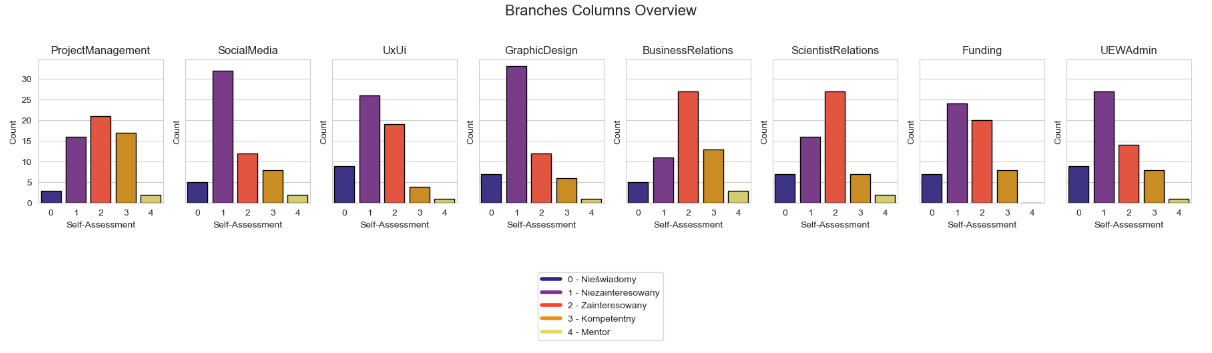
Ten obraz wydaje się być istotny dla oceny wyników w różnych obszarach ekspertyzy lub gałęziach w takich dziedzinach jak nauka danych czy nauka akademicka.



Na podstawie przesłanego obrazu, który zawiera siedem wykresów słupkowych przedstawiających różne technologie lub platformy (SQL, NoSQL, Azure, AWS, GCP, PowerBI i Tableau), można zauważyć następujące szczegółowe obserwacje:

* SQL: Większość osób oceniło swoje umiejętności jako ‘Zaawansowane’ zarówno w samoocenie, jak i ocenie mentora. Jednak ocena mentora jest nieco niższa dla poziomu ‘Ekspert’.
* NoSQL: Większość osób oceniło swoje umiejętności jako ‘Znajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Zaawansowany’.
* Azure: Większość osób oceniło swoje umiejętności jako ‘Znajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Zaawansowany’.
* AWS: Większość osób oceniło swoje umiejętności jako ‘Znajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Zaawansowany’.
* GCP: Większość osób oceniło swoje umiejętności jako ‘Nieznajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Znajomy’.
* PowerBI: Większość osób oceniło swoje umiejętności jako ‘Znajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Zaawansowany’.
* Tableau: Większość osób oceniło swoje umiejętności jako ‘Znajomy’ w samoocenie, podczas gdy ocena mentora była wyższa, z większością na poziomie ‘Zaawansowany’.

Wszystkie te obserwacje sugerują, że oceny mentorów są generalnie wyższe niż samooceny, co może wskazywać na skromność osób oceniających siebie lub na to, że mentory mają wyższe oczekiwania. Może to również sugerować, że osoby te mają potencjał do dalszego rozwoju swoich umiejętności w tych technologiach.



Na podstawie przesłanego obrazu, który zawiera serię wykresów słupkowych zatytułowanych “Branches Columns Overview”, przedstawiających różne kategorie takie jak ProjectManagement, SocialMedia, GraphicDesign, BusinessStories, ScientistRelations, Funding i LEWiatmin, można zauważyć następujące szczegółowe obserwacje:

* ProjectManagement: Większość osób oceniło swoje umiejętności jako ‘Zainteresowany’. Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* SocialMedia: Większość osób oceniło swoje umiejętności jako ‘Niezainteresowany’ Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* GraphicDesign: Większość osób oceniło swoje umiejętności jako ‘niezainteresowany’ Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* Businessrelations: Większość osób oceniło swoje umiejętności jako ‘Zainteresowany’ Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* ScientistRelations: Większość osób oceniło swoje umiejętności jako ‘Zainteresowany’ Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* Funding: Większość osób oceniło swoje umiejętności jako ‘Niezainteresowany’. Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.
* UEWAdmin: Większość osób oceniło swoje umiejętności jako ‘Niezainteresowany’ Natomiast najmniej osób oceniło swoje umiejętności ‘Mentor’.

