**Raport z projektu:**Wpływ mediów społecznościowych na zdrowie psychiczne

Analiza i wizualizacja danych - Pandas, Data Frame

Semestr 3, 2024-2025

**1. Skład sekcji projektowej:**

* Chmielewska Alicja
* Michna Natalia

**2. Opis analizowanego problemu:**

W dobie powszechnego dostępu do Internetu i rozwoju mediów społecznościowych coraz więcej osób korzysta z platform takich jak Facebook, Instagram czy Twitter. Chociaż media społecznościowe umożliwiają łatwiejszy kontakt z innymi oraz dostęp do informacji, ich wpływ na zdrowie psychiczne budzi obawy. Celem projektu jest analiza, czy i w jakim stopniu korzystanie z mediów społecznościowych może wpływać na zdrowie psychiczne ich użytkowników.

**3. Szczegółowy opis zbioru danych:**

**Źródło danych**: Zbiór danych pochodzi z platformy Kaggle: [Social Media and Mental Health Dataset](https://www.kaggle.com/datasets/souvikahmed071/social-media-and-mental-health).

Zbiór ten został pierwotnie zebrany na potrzeby projektu z zakresu uczenia maszynowego, którego celem było zbadanie potencjalnej korelacji między czasem spędzanym na mediach społecznościowych a jego wpływem na zdrowie psychiczne. Dane zostały zebrane poprzez przeprowadzenie ankiety i udostępnione jako plik CSV.

**Opis kolumn/zmiennych**:

* **Timestamp**Znacznik czasu (data + czas)
* **1. What is your age?**   
  Jaki jest Twój wiek?
* **2. Gender**Płeć
* **3. Relationship Status**   
  Stan cywilny
* **4. Occupation Status**  
  Status zawodowy
* **5. What type of organizations are you affiliated with?**  
  Z jakiego rodzaju organizacjami jesteś związany?
* **6. Do you use social media?**   
  Czy korzystasz z mediów społecznościowych?
* **7. What social media platforms do you commonly use?**   
  Z jakich platform mediów społecznościowych najczęściej korzystasz?
* **8. What is the average time you spend on social media every day?**Ile średnio czasu spędzasz na mediach społecznościowych każdego dnia?
* **9. How often do you find yourself using Social media without a specific purpose?**   
  Jak często korzystasz z mediów społecznościowych bez konkretnego celu?
* **10. How often do you get distracted by Social media when you are busy doing something?**   
  Jak często rozpraszają Cię media społecznościowe, gdy jesteś zajęty?
* **11. Do you feel restless if you haven't used Social media in a while?**Czy czujesz się niespokojny, jeśli przez jakiś czas nie korzystasz z mediów społecznościowych?
* **12. On a scale of 1 to 5, how easily distracted are you?**   
  W skali od 1 do 5, jak łatwo się rozpraszasz?
* **13. On a scale of 1 to 5, how much are you bothered by worries?**W skali od 1 do 5, jak bardzo odczuwasz zmartwienia?
* **14. Do you find it difficult to concentrate on things?**Czy masz trudności z koncentracją?
* **15. On a scale of 1-5, how often do you compare yourself to other successful people through the use of social media?**W skali od 1 do 5, jak często porównujesz się do innych odnoszących sukcesy osób za pośrednictwem mediów społecznościowych?
* **16. Following the previous question, how do you feel about these comparisons, generally speaking?**W odniesieniu do poprzedniego pytania, jak ogólnie czujesz się w związku z tymi porównaniami?
* **17. How often do you look to seek validation from features of social media?**Jak często szukasz akceptacji/walidacji za pomocą funkcji mediów społecznościowych?
* **18. How often do you feel depressed or down?**Jak często czujesz się przygnębiony lub smutny?
* **19. On a scale of 1 to 5, how frequently does your interest in daily activities fluctuate?**   
  W skali od 1 do 5, jak często zmienia się Twoje zainteresowanie codziennymi aktywnościami?
* **20. On a scale of 1 to 5, how often do you face issues regarding sleep?**W skali od 1 do 5, jak często masz problemy ze snem?

**Rozmiar danych**: Zbiór zawiera dane od 481 uczestników.

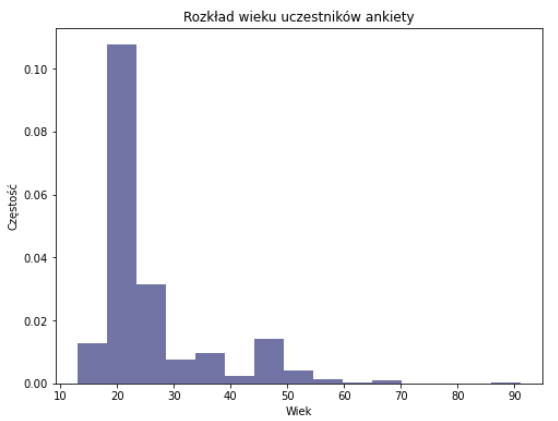
**4. Obróbka danych (pre-processing):**

1. **Zamieniono nazwy kolumn** – nazwy zostały skrócone i uproszczone, dzięki czemu późniejszy kod jest czytelniejszy:
   * timestamp
   * age
   * gender
   * relationship\_status
   * occupation\_status
   * affiliated\_organizations
   * sm\_use\_YN
   * sm\_platforms
   * sm\_avg\_time
   * scale\_sm\_wo\_purpose
   * scale\_sm\_freq\_distracted
   * scale\_sm\_freq\_restless
   * scale\_distraction
   * scale\_bothered
   * scale\_difficult\_to\_concentrate
   * scale\_compare\_to\_sm
   * scale\_attitude\_sm\_comparison
   * scale\_sm\_freq\_validation
   * scale\_freq\_depressed
   * scale\_freq\_fluctate\_interest
   * scale\_freq\_sleep\_issues
2. **Znaleziono i uzupełniono brakujące wartości:**
   * Brakujące wartości (NaN) pojawiały się jedynie w kolumnie „affiliated\_organizations”. Pozostałe kolumny takowych wartości nie posiadały.
   * Jako, że kolumna „affiliated\_organizations” zawiera listę powiązanych organizacji, gdzie jedną z opcji jest wartość ‘N/A’, puste komórki zostały uzupełnione tą wartością.
3. **Zmieniono typ danych wybranych kolumn:**
   * Kolumna „timestamp” została zamieniona na typ datetime.
   * Kolumna „age” została zamieniona na typ Int64, a wartości dziesiętne zaokrąglone do liczby całkowitej.
   * Kolumny „affiliated\_organizations” i „sm\_platforms” zostały zamienione z pojedynczego łańcucha znaków na listy zawierające składowe oryginalnego łańcucha znaków.
4. **Dodano nowe kolumny:**
   * Kolumna „sm\_avg\_time\_num” – zawiera przydzieloną numeryczną wartość dla kolumny „sm\_avg\_time” (średni czas spędzany na mediach społecznościowych każdego dnia).
   * Kolumna „affiliated\_organizations\_count” – określa liczbę powiązanych organizacji zaznaczonych przez każdą osobę ankietowaną.
   * Kolumna „sm\_platforms\_count” – określa liczbę używanych platform społecznościowych zaznaczonych przez każdą osobę ankietowaną.
   * Kolumna „tech\_ID” – zawiera techniczne ID będące unikalną liczbą całkowitą przydzieloną każdemu rekordowi.
5. **Analiza danych:**

**Czas przeprowadzania badania.**

* Ankieta przeprowadzona została w **2022 roku**.
* Pierwsza odpowiedź została zarejestrowana **18 kwietnia**, ostatnia – **12 listopada**.

**Wiek uczestników (rozkład).**

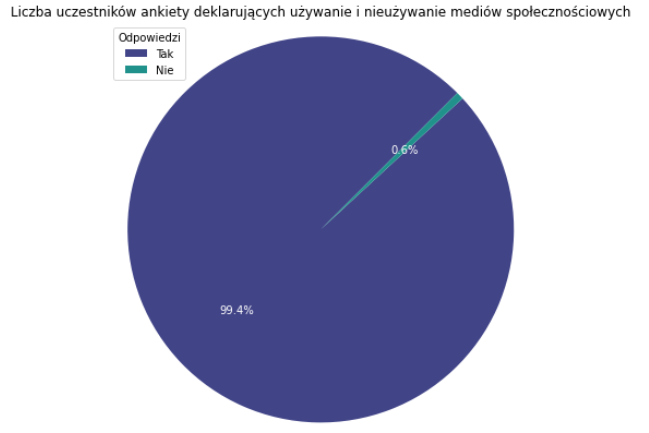
****

Rys. 1 Rozkład wieku uczestników ankiety

* Minimalny wiek uczestnika ankiety – **13 lat**.
* Maksymalny wiek uczestnika ankiety – **91 lat.**
* Najwięcej uczestników znajdowało się w przedziale wiekowym **20-30 lat.**

Na podstawie otrzymanego rozkładu wieku uczestników oraz opisu zbioru danych można wnioskować, że największą grupą uczestników byli studenci uczestniczący w kursie oraz ich znajomi.

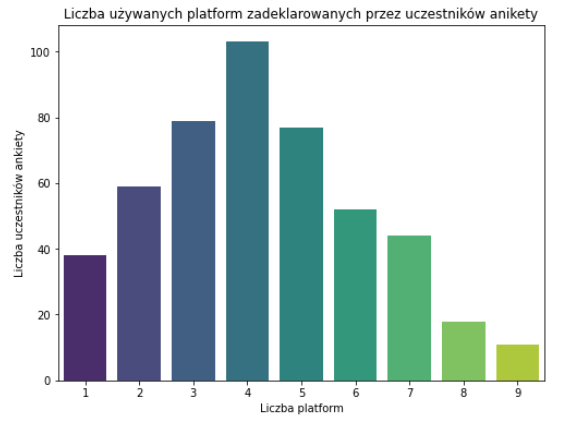
**Liczba uczestników ankiety deklarujących używanie i nieużywanie mediów społecznościowych.**

****

Rys. 2 Liczba uczestników ankiety deklarujących używanie i nieużywanie mediów społecznościowych

* Znacząca większość uczestników ankiety zadeklarowała korzystanie z mediów społecznościowych.
* Osoby korzystające stanowiły **99.4%** wszystkich uczestników, a osoby niekorzystające – jedynie **0.6%.**

**Liczba platform społecznościowych wykorzystywanych przez użytkowników.**

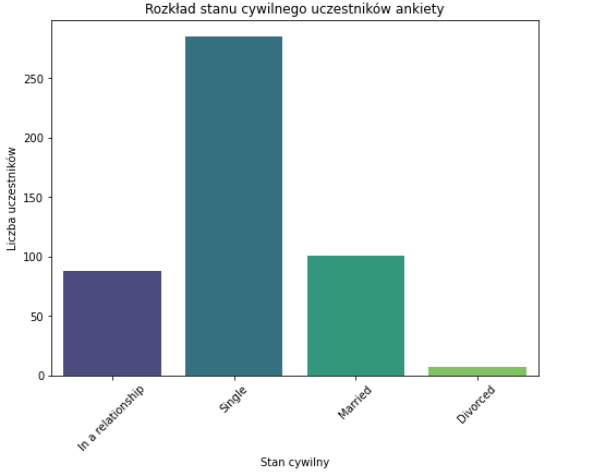
****

Rys. 3 Liczba używanych platform zadeklarowanych przez uczestników anikety

* Największa liczba uczestników korzysta z **4 platform mediów społecznościowych** (ponad 100 osób).
* Wraz ze wzrostem liczby platform (powyżej 4), liczba użytkowników stopniowo maleje. Uczestnicy deklarujący użycie 8 lub 9 platform stanowią niewielką grupę.
* Mniej osób korzysta z 1, 2 lub 3 platform niż z 4, ale nadal są to liczebnie znaczące grupy.

Najpopularniejsze jest korzystanie z kilku platform, a zwłaszcza 4 platform, co może świadczyć o tym, że użytkownicy angażują się w kilka różnych środowisk mediów społecznościowych. Użycie dużej liczby platform (powyżej 7) jest rzadsze, co może wskazywać na wyczerpanie zasobów czasowych lub mniejsze zainteresowanie większą liczbą środowisk cyfrowych.

**Rozkład stanu cywilnego uczestników ankiety.**



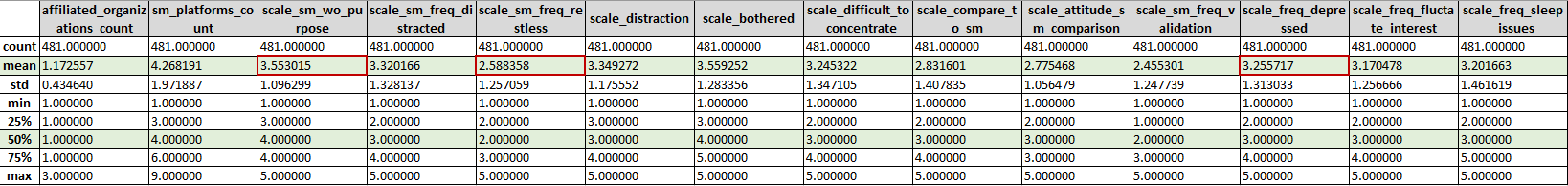
Rys. 4 Rozkład stanu cywilnego uczestników ankiety

* Największa grupa uczestników ankiety to single (*Single*) – **ponad 250 osób**.
* Znacznie mniej uczestników zadeklarowało bycie w związku (*In a relationship*) – **około 90 osób.**
* **Około 100 uczestników** zadeklarowało bycie w związku małżeńskim (*Married*), co stanowi drugą największą grupę.
* Kategorię osób po rozwodzie (*Divorced*) reprezentuje najmniejsza liczba uczestników – **poniżej 20 osób**.

Różnice w liczebności pomiędzy kategoriami mogą odzwierciedlać strukturę demograficzną grupy badanej (np. wiek uczestników lub etap życia).

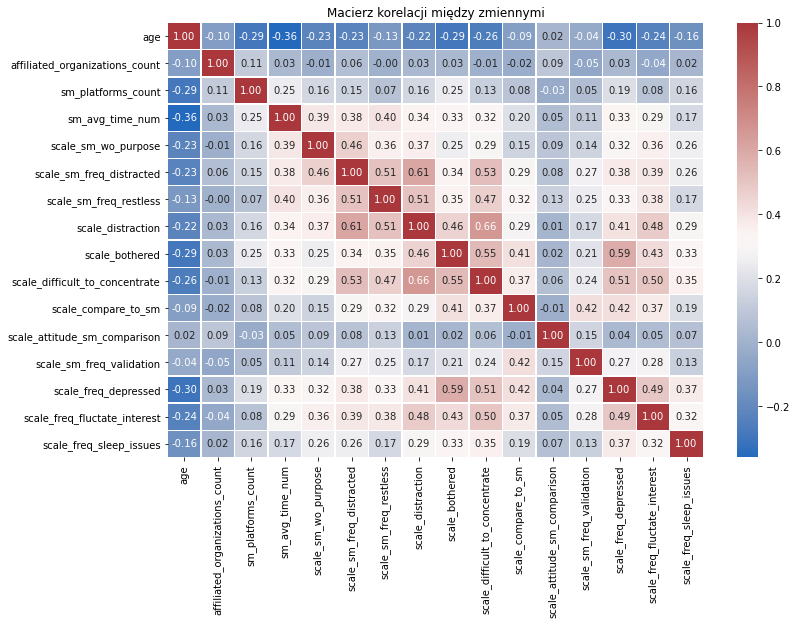
Korelacja między zmiennymi, w szczególności między korzystaniem z mediów a problemami psychicznymi.

**Podstawowe statystyki.**



* **Średnie** (*mean*) dla większości skal (1-5) są w przedziale 2.5-3.6, co wskazuje na umiarkowane nasilenie objawów lub zachowań związanych z mediami społecznościowymi i samopoczuciem.
  + **scale\_sm\_wo\_purpose (3.55)**: Uczestnicy często korzystają z mediów społecznościowych bez konkretnego celu.
  + **scale\_freq\_depressed (3.26)**: Użytkownicy umiarkowanie często odczuwają przygnębienie lub smutek.
  + **scale\_sm\_freq\_restless (2.59)**: Trudności z niepokojem związanym z niekorzystastaniem z mediów społecznościowych są rzadziej raportowane.
* **Odchylenie standardowe** (*std*).
  + Wartości odchylenia standardowego **(1.06–1.65)** wskazują na spore zróżnicowanie odpowiedzi uczestników – szczególnie biorąc pod uwagę że skala odpowiedzi to liczby od 1 do 5.
* **Wartości minimalne i maksymalne** (*min, max*).
  + Wszystkie kolumny zawierają skrajne wartości skali – wartość **minimalną 1** i **maksymalną 5**.
* **Kwartyle** (25%, 50%, 75%):
  + W większości zmiennych **mediana (50%) wynosi 3**, co wskazuje na centralną tendencję w odpowiedziach na poziomie umiarkowanym.
  + Wartości median zbliżone są do wartości średniej, co może świadczyć o względnie symetrycznym rozłożeniu odpowiedzi.

**Korelacja między zmiennymi.**



Rys. 5 Macierz korelacji między zmiennymi

Korelacje, których siła jest większa lub równa 0.5:

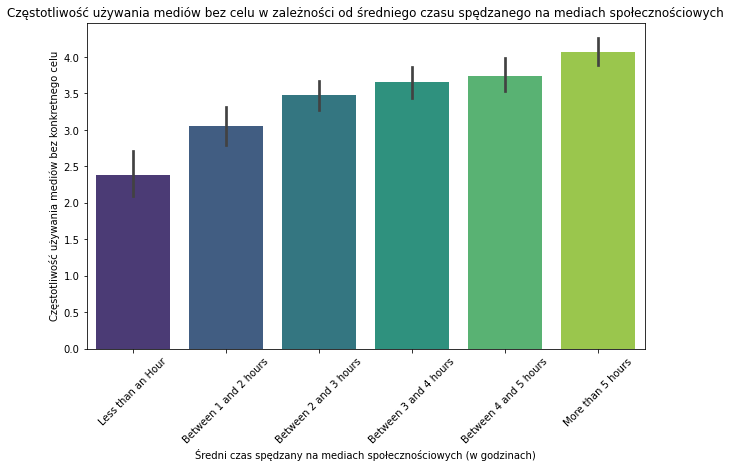
* scale\_sm\_freq\_distracted i scale\_sm\_freq\_restless (**0.51**)
  + Jak często rozpraszają Cię media społecznościowe, gdy jesteś zajęty?
  + Czy czujesz się niespokojny, jeśli przez jakiś czas nie korzystasz z mediów społecznościowych?
* scale\_sm\_freq\_distracted i scale\_distraction (**0.61**)
  + Jak często rozpraszają Cię media społecznościowe, gdy jesteś zajęty?
  + W skali od 1 do 5, jak łatwo się rozpraszasz?
* scale\_sm\_freq\_distracted i scale\_difficult\_to\_concentrate (**0.53**)
  + Jak często rozpraszają Cię media społecznościowe, gdy jesteś zajęty?
  + Czy masz trudności z koncentracją?
* scale\_sm\_freq\_restless i scale\_distraction (**0.51**)
  + Czy czujesz się niespokojny, jeśli przez jakiś czas nie korzystasz z mediów społecznościowych?
  + W skali od 1 do 5, jak łatwo się rozpraszasz?
* scale\_distraction i scale\_difficult\_to\_concentrate (**0.66**)
  + W skali od 1 do 5, jak łatwo się rozpraszasz?
  + Czy masz trudności z koncentracją?
* scale\_bothered i scale\_difficult\_to\_concentrate (**0.55**)
  + W skali od 1 do 5, jak bardzo odczuwasz zmartwienia?
  + Czy masz trudności z koncentracją?
* scale\_bothered i scale\_freq\_depressed (**0.59**)
  + W skali od 1 do 5, jak bardzo odczuwasz zmartwienia?
  + Jak często czujesz się przygnębiony lub smutny?
* scale\_difficult\_to\_concentrate i scale\_freq\_depressed (**0.51**)
  + Czy masz trudności z koncentracją?
  + Jak często czujesz się przygnębiony lub smutny?
* scale\_difficult\_to\_concentrate i scale\_freq\_fluctate\_intrest (**0.50**)
  + Czy masz trudności z koncentracją?
  + W skali od 1 do 5, jak często zmienia się Twoje zainteresowanie codziennymi aktywnościami?

Korelacje między intensywnością korzystania z social mediów a zdrowiem psychicznym są na mniej znaczącym poziomie:

* Korelacja uwzględniająca liczbę używanych platform społecznościowych osiąga maksymalnie wartość **0.25** – korelacja z wartością scale\_bothered (*W skali od 1 do 5, jak bardzo odczuwasz zmartwienia?*)
* Korelacja uwzględniająca średni dzienny czas korzystania z social mediów osiąga maksymalnie wartość **0.4** – korelacja z wartością scale\_sm\_freq\_restless (*Czy czujesz się niespokojny, jeśli przez jakiś czas nie korzystasz z mediów społecznościowych?*). Zbliżone wartości osiąga również dla zmiennych
  + scale\_sm\_freq\_distracted (*Jak często rozpraszają Cię media społecznościowe, gdy jesteś zajęty?):* **0.38**
  + scale\_sm\_wo\_purpose (*Jak często korzystasz z mediów społecznościowych bez konkretnego celu?):* **0.39**

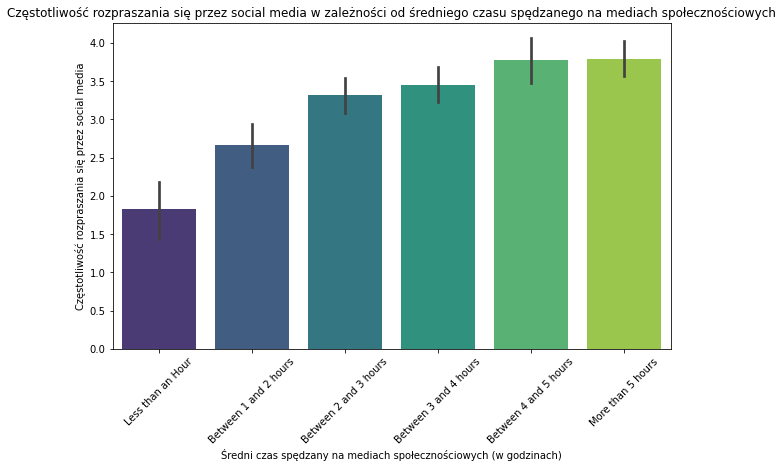
Obserwacje wyciągnięte na podstawie korelacji można również zaobserwować na wykresie słupkowym korzystając ze zmiennej sm\_avg\_time (*Ile średnio czasu spędzasz na mediach społecznościowych każdego dnia?*). Na wykresach można zauważyć ogólną tendencję do obniżania się zdrowia psychicznego wraz ze wzrostem dziennego czasu spędzanego w social mediach. Nie jest to natomiast znacząca zależność.

**Częstotliwość używania mediów bez celu w zależności od średniego czasu spędzanego na mediach społecznościowych.**



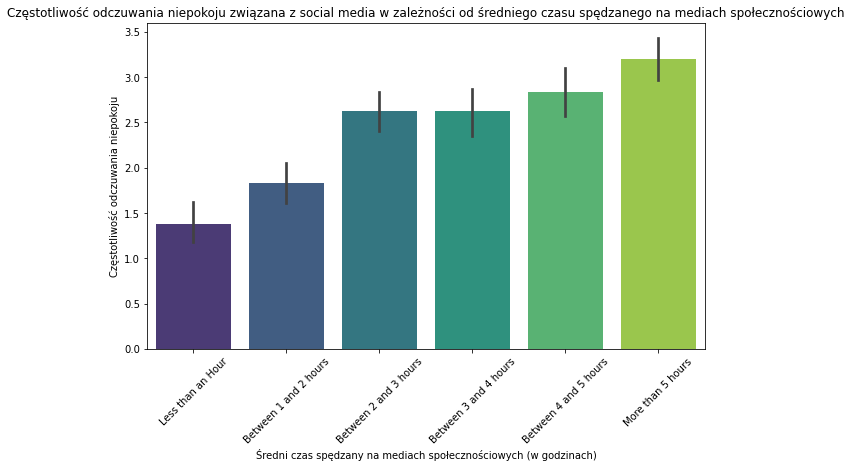
Rys. 6 Częstotliwość używania mediów bez celu w zależności od średniego czasu spędzanego na mediach społecznościowych

**Częstotliwość rozpraszania się przez social media w zależności od średniego czasu spędzanego na mediach społecznościowych.**



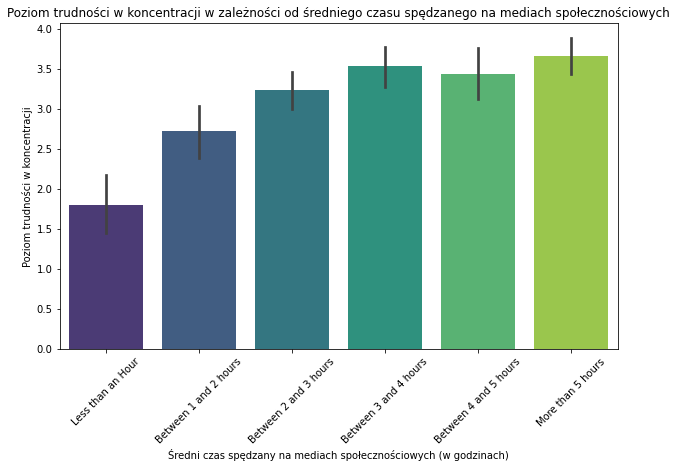
Rys. 7 Częstotliwość rozpraszania się przez social media w zależności od średniego czasu spędzanego na mediach społecznościowych

**Częstotliwość odczuwania niepokoju związana z social media w zależności od średniego czasu spędzanego na mediach społecznościowych.**



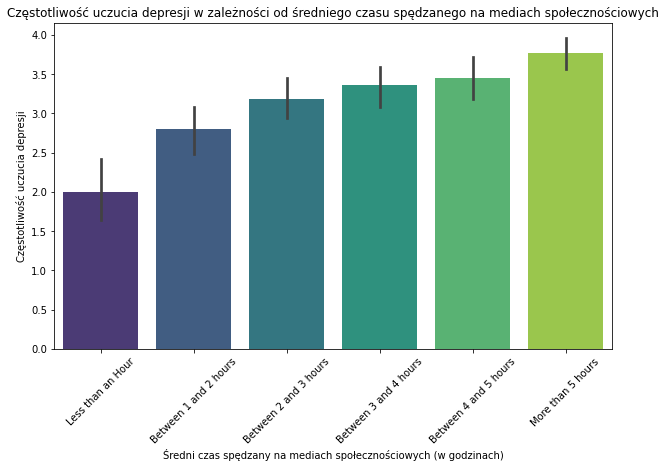
Rys. 8 Częstotliwość odczuwania niepokoju związana z social media w zależności od średniego czasu spędzanego na mediach społecznościowych

**Poziom trudności w koncentracji w zależności od średniego czasu spędzanego na mediach społecznościowych.**



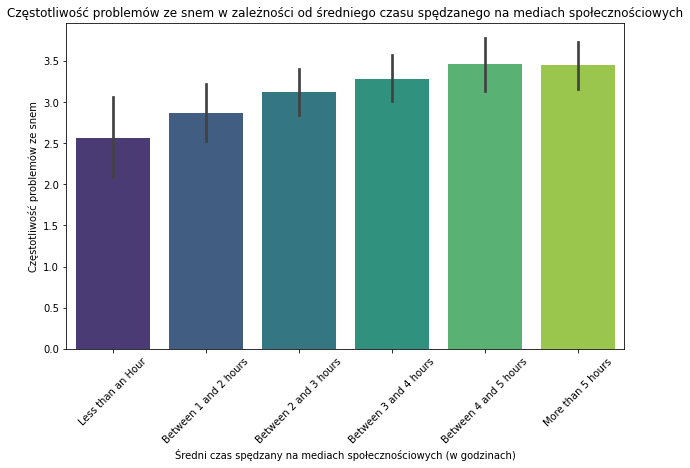
Rys. 9 Poziom trudności w koncentracji w zależności od średniego czasu spędzanego na mediach społecznościowych

**Częstotliwość uczucia depresji w zależności od średniego czasu spędzanego na mediach społecznościowych.**



Rys. 10 Częstotliwość uczucia depresji w zależności od średniego czasu spędzanego na mediach społecznościowych

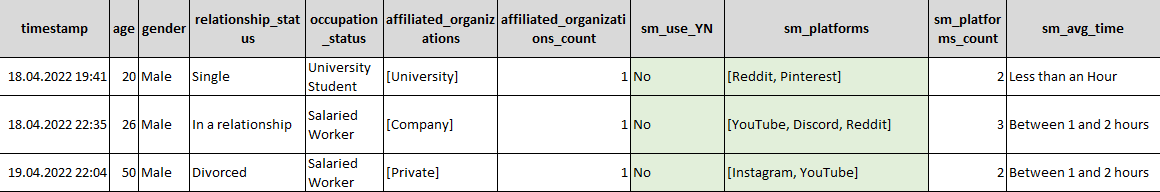
**Częstotliwość problemów ze snem w zależności od średniego czasu spędzanego na mediach społecznościowych.**



Rys. 11 Częstotliwość problemów ze snem w zależności od średniego czasu spędzanego na mediach społecznościowych

1. **Analiza skrajnych przypadków:**

W ankiecie, trzech uczestników zadeklarowało nieużywanie mediów społecznościowych. Analizując te przypadki, stwierdzono sprzeczność tej deklaracji – wszyscy trzej uczestnicy, wybrali platformy społecznościowe, z których korzystają. Dodatkowo dwóch z nich, zadeklarowało spędzanie ponad godziny dziennie w mediach społecznościowych.



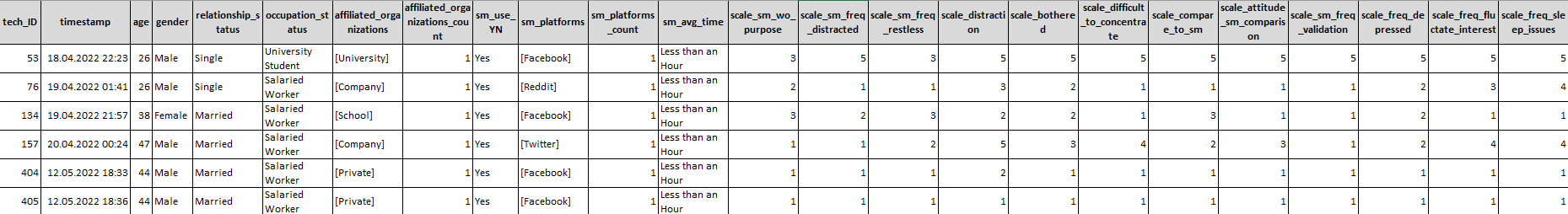
Rys. 12 Uczestnicy "niekorzystający" z mediów społecznościowych

W związku z powyższym, jako skrajną grupę wybrano uczestników na podstawie zadeklarowanej liczby wykorzystywanych platform społecznościowych oraz średniego czasu korzystania z social mediów na dzień.

Filtrowanie rekordów, gdzie:

* uczestnicy korzystają tylko z jednej platformy społecznościowej oraz
* uczestnicy spędzają mniej niż godzinę dziennie w mediach społecznościowych

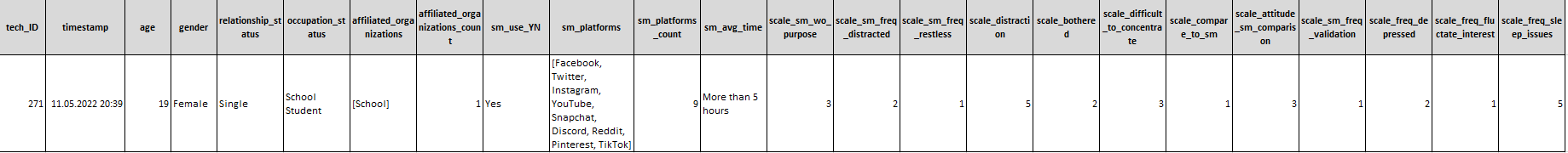
ujawniło sześć osób, które będą stanowiły jedną z grup skrajnych (uczestnicy najmniej korzystający z mediów społecznościowych).



Rys. 13 Grupa skrajna - uczestnicy najmniej korzystający z mediów społecznościowych

Do drugiej grupy skrajnej zaliczono uczestników wybranych na podstawie tych samych dwóch zmiennych (liczba używanych platform społecznościowych oraz średni dzienny czas korzystania z social mediów) – szukano tu jednak osób z maksymalnymi ich wartościami.

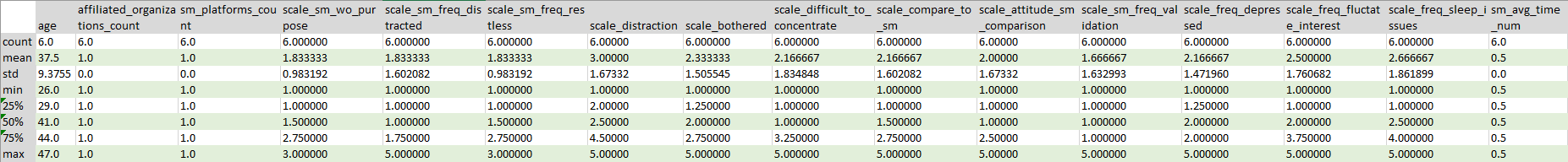
Wynikiem filtrowania była jedna osoba:



Rys. 14 Grupa skrajna - uczestnicy najwięcej korzystający z mediów społecznościowych

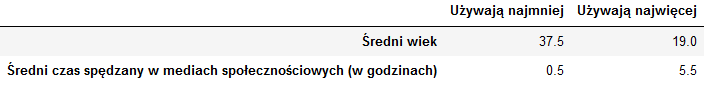
Korzystała ona z dziewięciu platform społecznościowych średnio ponad 5 godzin dziennie.

**Podstawowe statystyki dla uczestników najmniej korzystających z mediów społecznościowych.**



Rys. 15 Podstawowe statystyki dla uczestników najmniej korzystających z mediów społecznościowych.

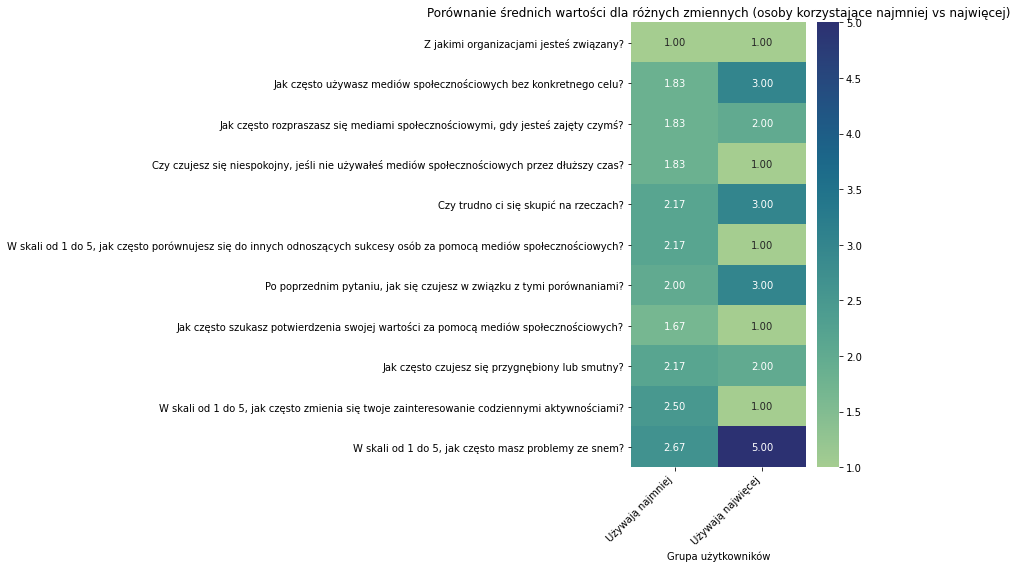
**Porównanie średnich wartości dla obu grup skrajnych – Wiek.**



Rys. 16 Porównanie średnich wartości dla obu grup skrajnych - wiek i średni czas spędzany w social mediach.

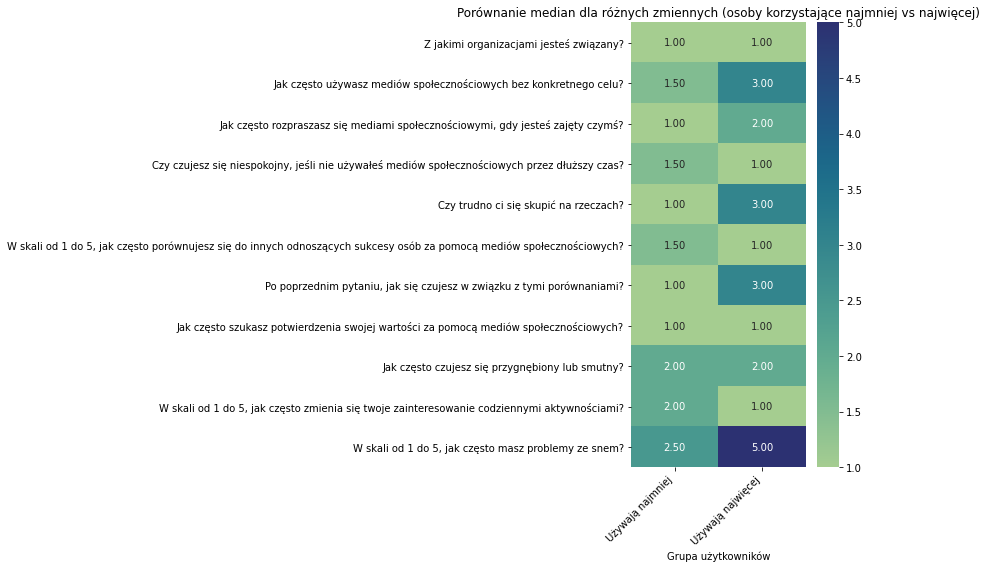
Porównując średni wiek obu grup można zauważyć, że wraz ze wzrostem używania social mediów maleje wiek uczestników ankiety.

**Porównanie średnich wartości dla obu grup skrajnych.**



Rys. 17 Porównanie średnich wartości dla różnych zmiennych (osoby korzystające najmniej vs najwięcej)

**Porównanie median wartości dla obu grup skrajnych.**



Rys. 18 Porównanie median dla różnych zmiennych (osoby korzystające najmniej vs najwięcej)

Różnica między wartościami median a średnimi dla grupy użytkowników korzystających najmniej z mediów społecznościowych spowodowana jest bardzo wysokimi odpowiedziami (w skali od 1 do 5) jednego z uczestników ankiety.

1. **Podsumowanie i wnioski:**

Analiza wyników ankiety wskazuje, że korelacje między intensywnością korzystania z mediów społecznościowych (mierzonych poprzez dzienny czas spędzany online oraz liczbę używanych platform) a zdrowiem psychicznym uczestników nie osiągają znaczących wartości. Chociaż istnieją pewne zależności, są one stosunkowo słabe i nie pozwalają jednoznacznie stwierdzić, że intensywne korzystanie z mediów społecznościowych ma bezpośredni wpływ na zdrowie psychiczne.

Warto zauważyć, że do pełniejszego zrozumienia tej relacji przydałaby się większa i bardziej zróżnicowana grupa badawcza, obejmująca osoby o różnym wieku, pochodzeniu, stylu życia i poziomie zaangażowania w media społecznościowe. Obecna próba mogła nie być wystarczająco reprezentatywna, co ogranicza możliwość uogólnienia wyników na całą populację.

Ponadto, przeprowadzenie badań w innej formie niż ankieta mogłoby dostarczyć bardziej precyzyjnych danych. Na przykład badania eksperymentalne, długoterminowe obserwacje czy analizy zachowań w rzeczywistych warunkach pozwoliłyby na bardziej dokładne ustalenie zależności przyczynowo-skutkowych. Sama ankieta, choć użyteczna, ma swoje ograniczenia, szczególnie jeśli pytania nie są wystarczająco precyzyjne.

W przyszłości warto rozważyć opracowanie bardziej szczegółowego kwestionariusza, który dokładniej uchwyci specyfikę korzystania z mediów społecznościowych oraz ich potencjalny wpływ na różne aspekty zdrowia psychicznego. Lepsza konstrukcja pytań oraz uzupełnienie badania o inne metody analizy mogłyby znacząco poprawić jakość i wiarygodność uzyskanych wyników.