

# Introwertycy i Ekstrawertycy

Zależność między czasem  
spędzanym samotnie i  
aktywnością w Social  
Mediach, a ilością znajomych



# Zbiór Danych

## Wybór zbioru, cel analizy badawczej

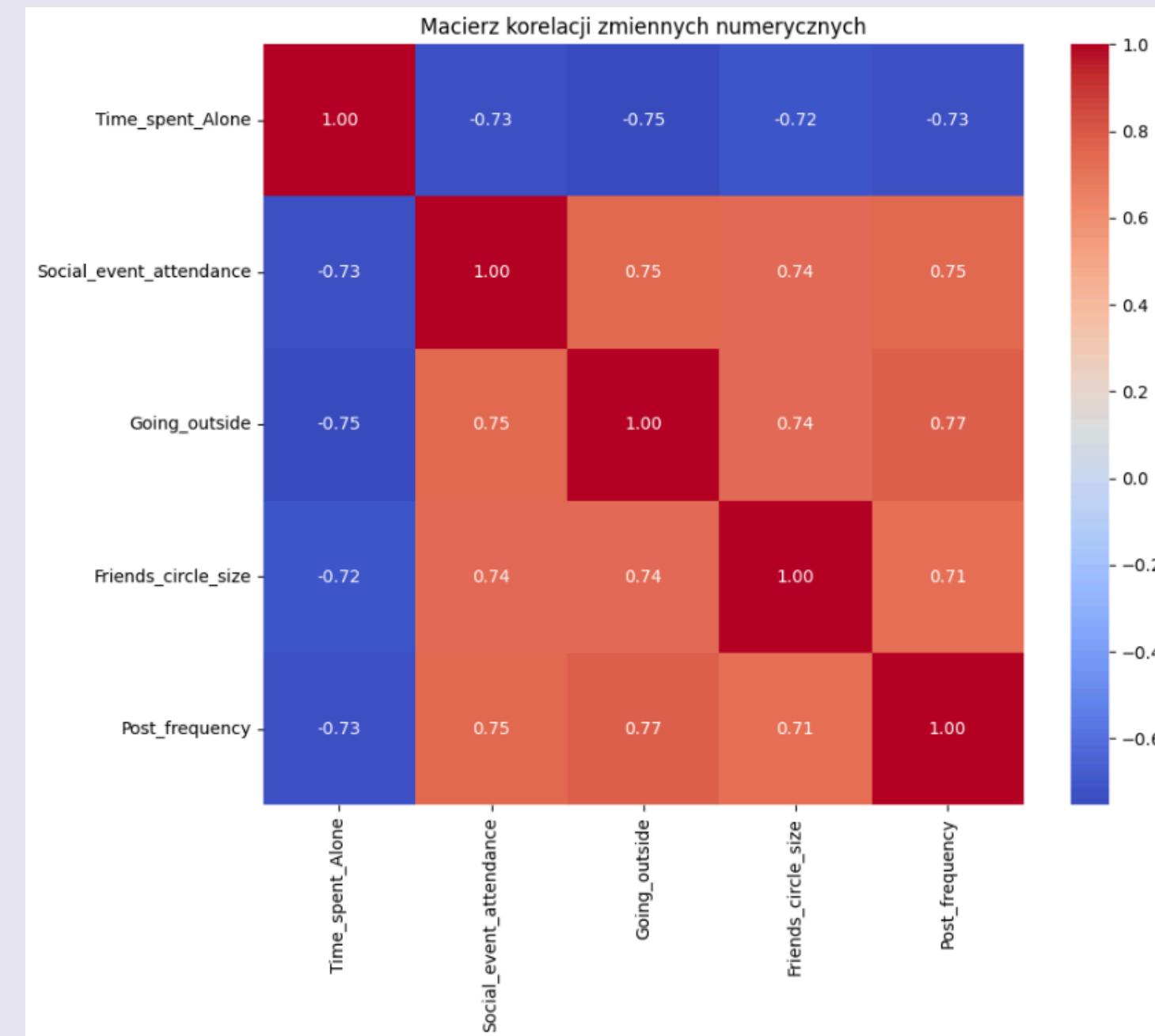
Wybrany przez nas temat reprezentowany przez zbiór danych “*Extrovert vs. Introvert Behavior Data*” autorstwa Rakesha Kapilavayi zawierający **2900** rekordów danych dotyczących zachowań społecznych dla dwóch typów osobowości - **introwertyków i ekstrawertyków**. Zawiera on pięć zmiennych ilościowych i trzy jakościowe. Zmienne wybrane przez nas do analizy to:

- **Time Spent Alone** (czas spędzany samotnie) - zmienna zależna, liczba godzin spędzanych w samotności dziennie.
- **Friends Circle Size** (wielkość kręgu przyjaciół) - zmienna objaśniająca, liczba osób w bliskim kręgu znajomych.
- **Post Frequency** (częstotliwość “postowania”) - zmienna objaśniająca, liczba publikowanych treści w Social Mediach dziennie.

Celem naszej analizy było zbadanie, **czy w erze Internetu, komunikatorów online i pracy zdalnej osoby z osobowością introwertyczną wciąż mają problemy z nawiązywaniem relacji w zestawieniu z ekstrawertykami**. Model regresji przewiduje, **jak wielkość kręgu znajomych i częstotliwość dodawania treści w Social Mediach wpływa na ilość czasu spędzanego w samotności**.

# Zbiór Danych

## Macierz korelacji



# Zbiór Danych

## Środowisko badawcze, czyszczenie danych

Analizę danych przeprowadziliśmy w środowisku Google Colab z wykorzystaniem języka Python oraz bibliotek NumPy, Pandas, Matplotlib oraz Seaborn, co ułatwiło kolaborację i sprawne uzyskiwanie informacji ze zbioru danych. Aby to nastąpiło, trzeba było je najpierw wyczyścić, ponieważ zbiór zawierał niepełne informacje i wartości NaN / null. Po wyczyszczeniu zbiór pomniejszył się o 523 rekordy.

```
[5] print("Rozmiar zbioru:", df.shape)
→ Rozmiar zbioru: (2900, 8)

Usuwamy wiersze dla których jest więcej niż 1 NaN i resetujemy index

[6] df = df[nan_per_row < 1].reset_index(drop=True)

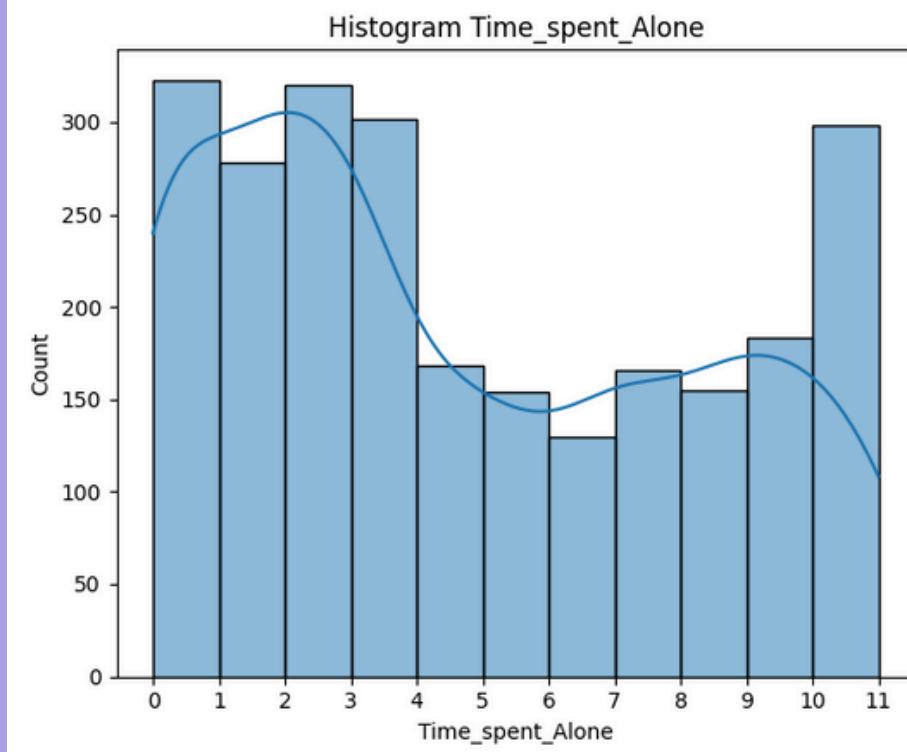
[7] print("Rozmiar zbioru:", df.shape)
→ Rozmiar zbioru: (2477, 8)

Sprawdzamy czy wszystkie NaNy zostały usunięte

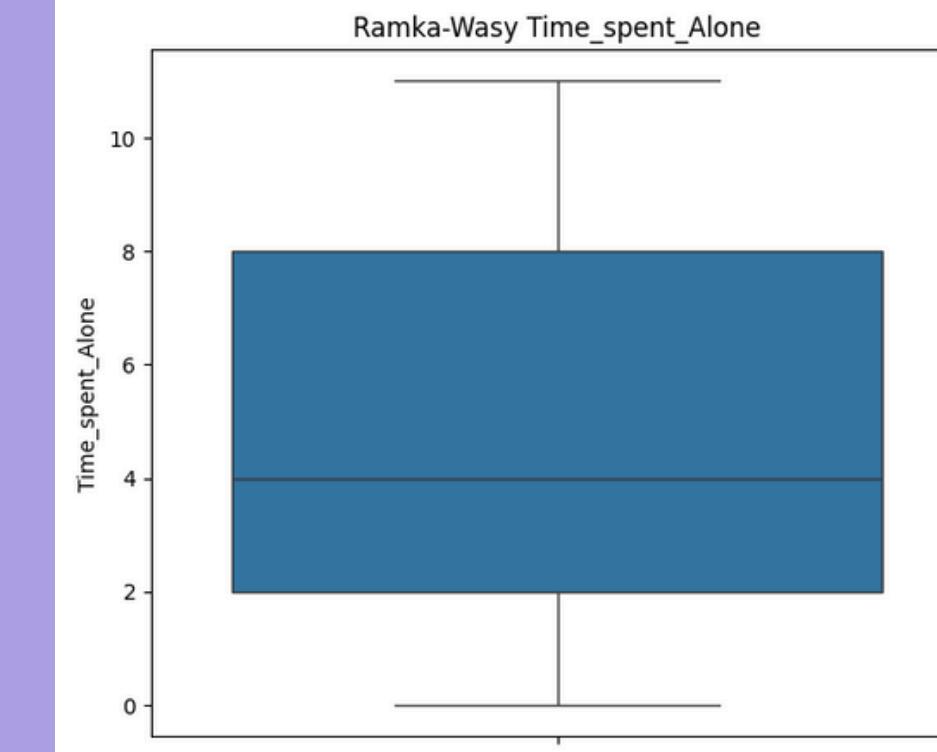
[8] print("\nLiczba NaN-ów w kolumnach:\n", df.isnull().sum())
→ Liczba NaN-ów w kolumnach:
    Time_spent_Alone      0
    Stage_fear              0
    Social_event_attendance 0
    Going_outside            0
    Drained_after_socializing 0
    Friends_circle_size      0
    Post_frequency            0
    Personality                0
    dtype: int64
```

# Analiza Danych

Wykresy zmiennej objaśnianej (nieskategoryzowane)

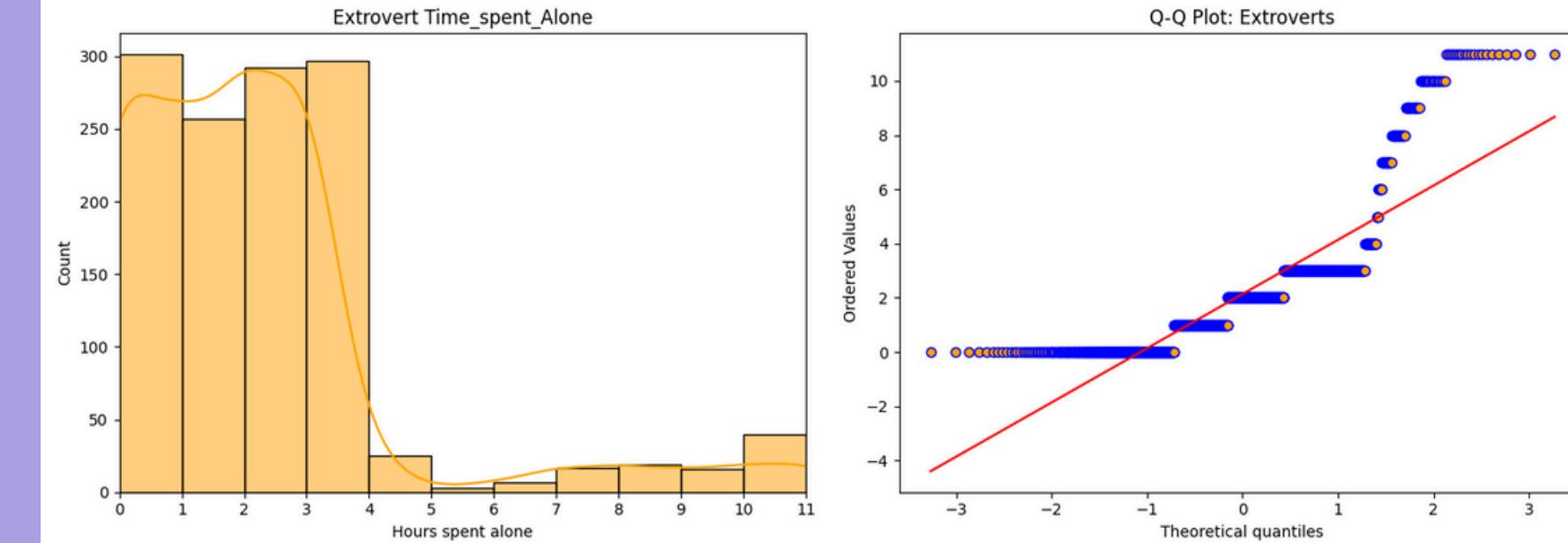
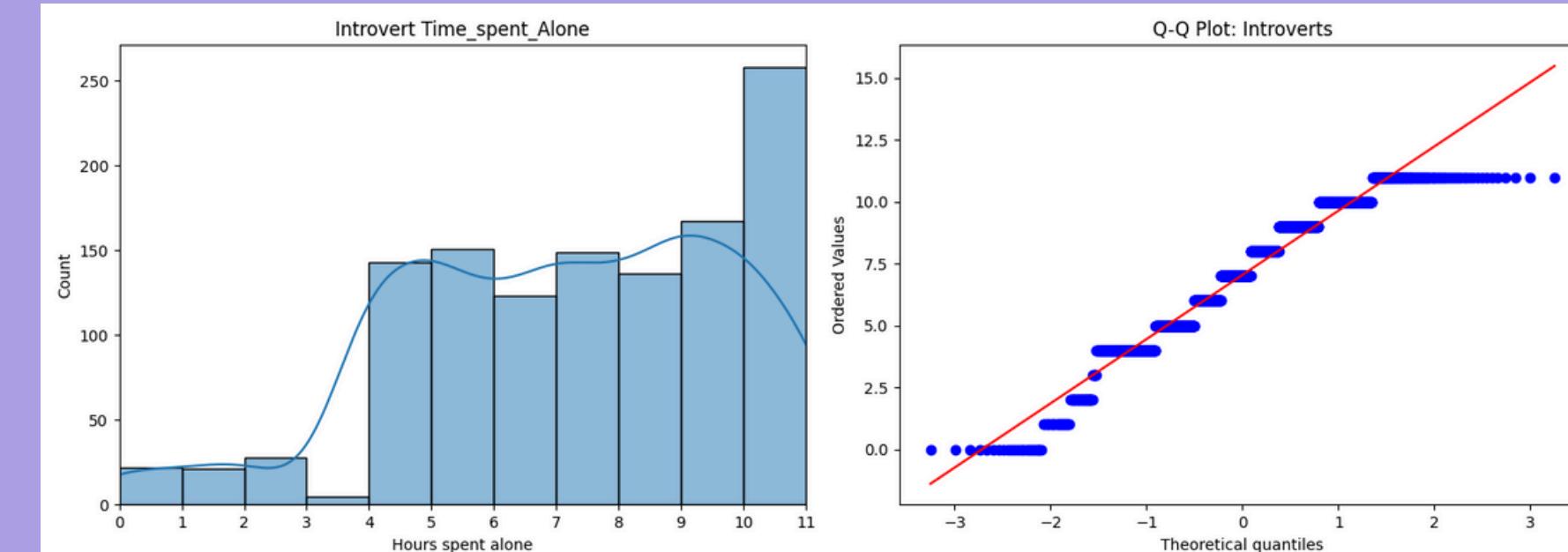
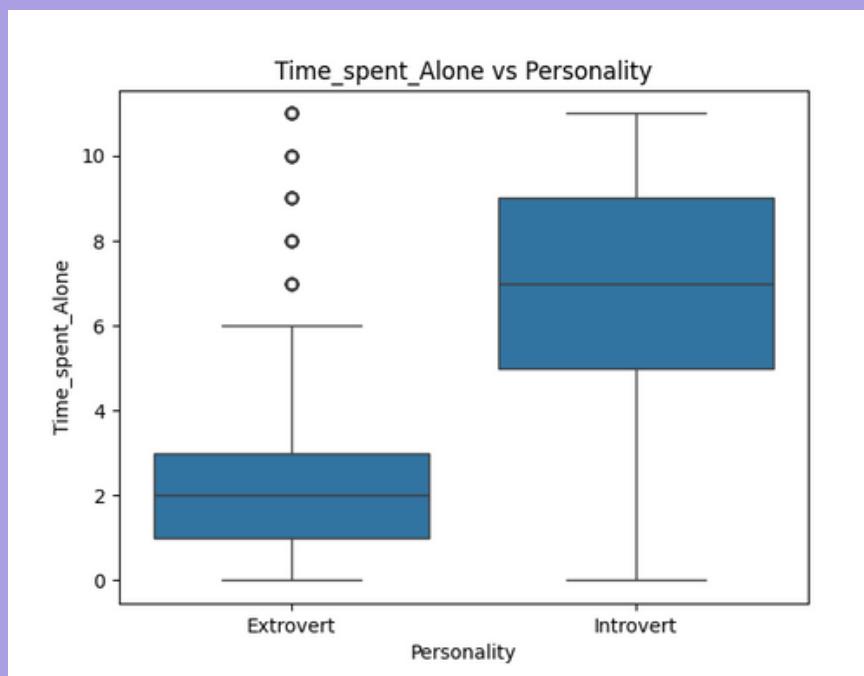
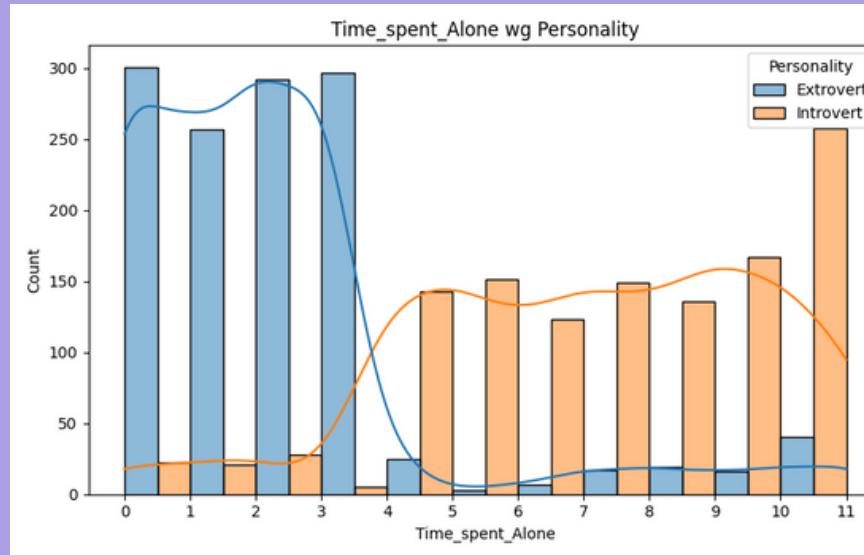


count	2477.000000
mean	4.522406
std	3.491162
min	0.000000
25%	2.000000
50%	4.000000
75%	8.000000
max	11.000000



# Analiza Danych

## Wykresy zmiennej objaśnianej (skategoryzowane)



# Analiza Danych

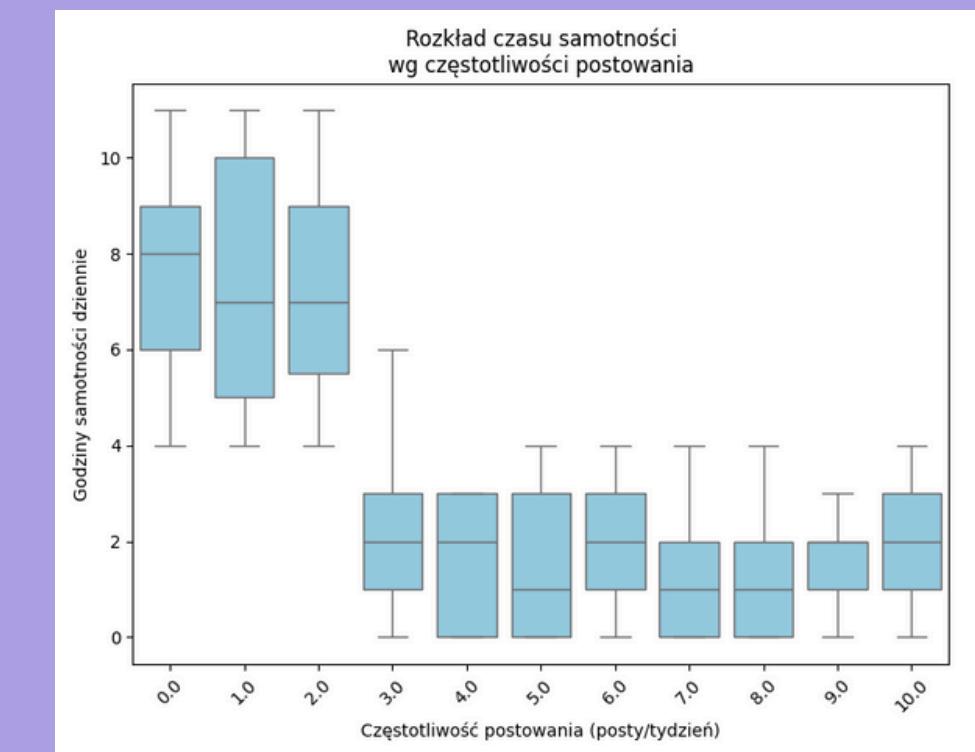
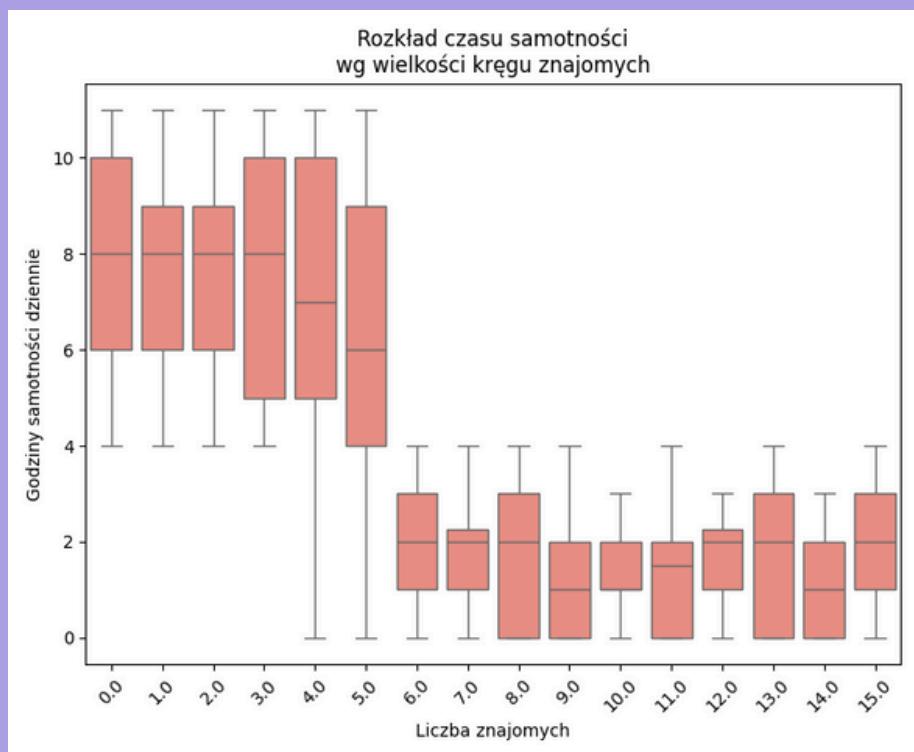
## Interpretacja wykresów (nieskategoryzowane)

Wartości na wykresach jasno pokazują, że introwertycy spędzają zdecydowanie więcej czasu w samotności niż ekstrawertycy, co jest zgodne z ogólnie przyjętymi społecznie założeniami. Rozkład wyników również zdecydowanie bardziej przypomina normalny po skategoryzowaniu go na dwie grupy, chociaż dla żadnego podziału wartość  $p$  nie wyszła powyżej  $\alpha = 0,05$ .



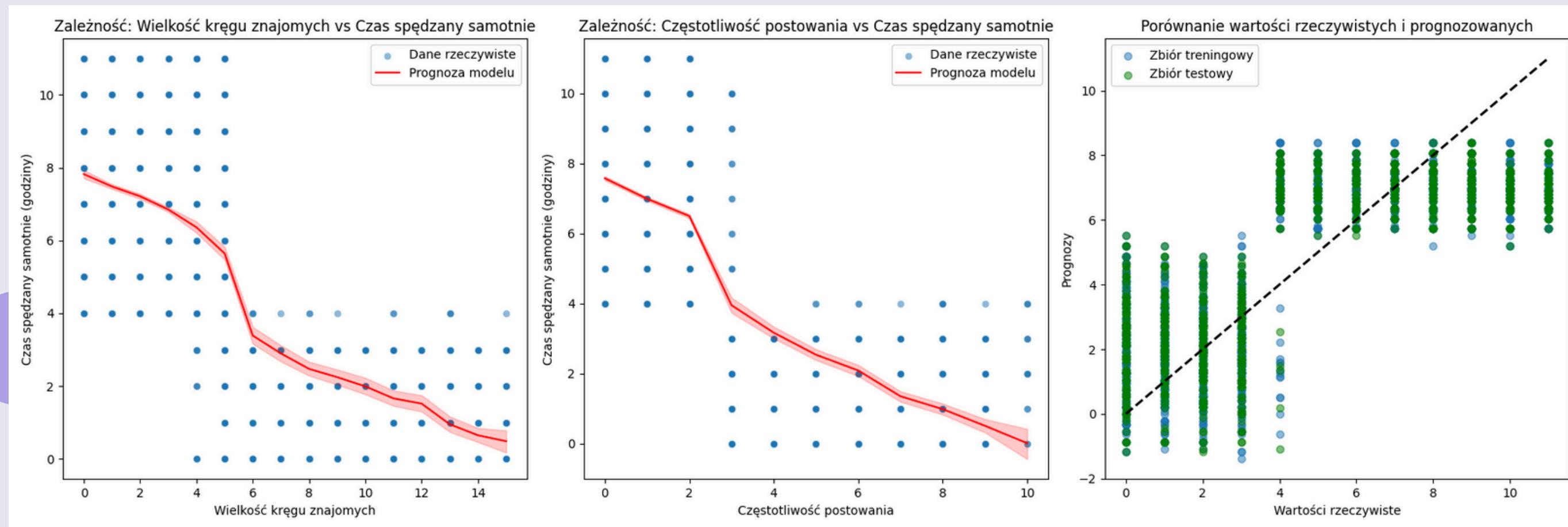
# Analiza Danych

Wykresy pozostałych zmiennych (skategoryzowane)



# Analiza Danych

## Model regresji



Równanie regresji:  
 $Time\_spent\_Alone = 8.38 + -0.32 * Friends\_circle\_size + -0.54 * Post\_frequency$

Ocena jakości modelu:

Metryki dla zbioru treningowego:  
R2: 0.6196  
MAE: 1.7547  
MSE: 4.5570  
RMSE: 2.1347

Metryki dla zbioru testowego:  
R2: 0.6062  
MAE: 1.8191  
MSE: 4.9741  
RMSE: 2.2303

# Analiza Danych

## Interpretacja modelu regresji

Wyniki wskazują, że przyrost liczby znajomych o jedną jednostkę obniża przewidywany czas samotności średnio o 0.51 jednostki, natomiast wzrost częstotliwości publikowania postów o jednostkę zwiększa go o 0.38 jednostki.

Model został zastosowany do prognoz na zbiorze treningowym i testowym. Na zbiorze treningowym uzyskano  $R^2$  na poziomie 0.62, co oznacza, że model wyjaśnia 62% zmienności zmiennej zależnej. Dla zbioru testowego  $R^2$  wyniosło 0.58, co świadczy o umiarkowanej, ale stabilnej jakości predykcji. Błędy średnie (MAE) i błędy średniokwadratowe (RMSE) były na zblżonym poziomie dla obu zbiorów, co potwierdza, że model generalizuje w miarę dobrze i nie jest nadmiernie dopasowany.



# Wnioski

## Zależności, wyniki analizy danych

Nasza analiza danych pokazała, że istnieje realna korelacja między typem osobowości, stylem życia (również tego w sieci) a realną liczbą bliskich znajomych. Osoby z osobowością introwertyczną pomimo rosnącej liczby narzędzi komunikacyjnych wciąż mają problemy z nawiązywaniem relacji i dużą część czasu w trakcie dnia spędzają sami.



# Pytania?

# Dziękujemy za uwagę!

Mikołaj Bednarczyk, Kacper Golmento



# Źródła

## Wybór zbioru, cel analizy badawczej

- materiały wykładowe,
- <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data>
- [https://upel.agh.edu.pl/pluginfile.php/484833/mod\\_page/content/2/HIPOTEZY-pigu%C5%82ka.pdf](https://upel.agh.edu.pl/pluginfile.php/484833/mod_page/content/2/HIPOTEZY-pigu%C5%82ka.pdf)
- <https://colab.research.google.com/drive/1fvlfqykLSU45cxZRWlwhOrWiT-IgV6?usp=sharing>
- <https://www.youtube.com/watch?v=MG7DRpf-lfk>