

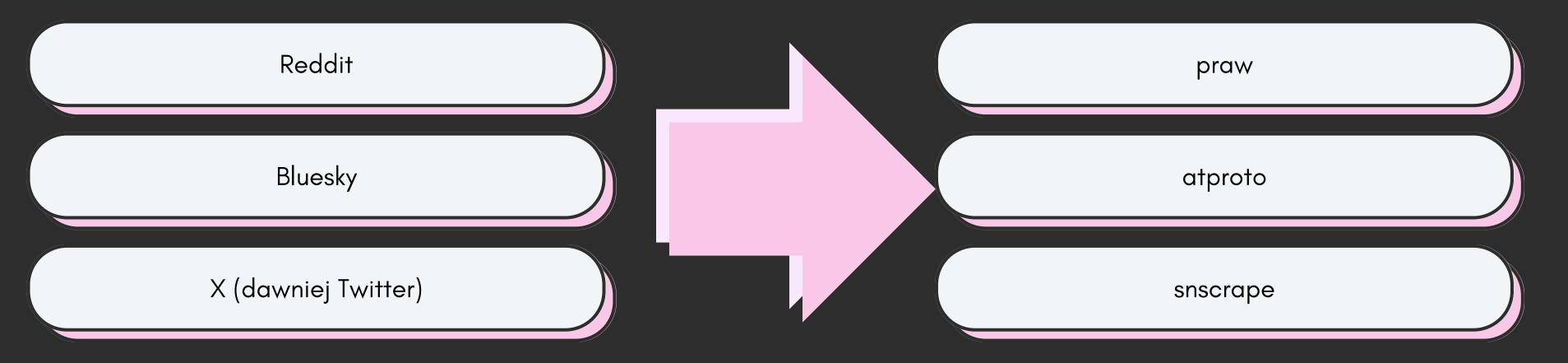
Analiza tekstu z mediów społecznościowych:

SONDAŻE
WYBORCZE 2025

Autor: Marta Piotrowska

Platformy, z których pobrałam posty i biblioteki, które to umożliwiły







Wystąpiły jednak pewne problemy z paczką snscrape...

```
retrieving <a href="https://twitter.com/search?f=live&lang=en&q=sonda%C5%BC+since%3A2025-05-04&src=spelling_expansion_revert_click">https://twitter.com/search?f=live&lang=en&q=sonda%C5%BC+since%3A2025-05-04&src=spelling_expansion_revert_click</a>: SSLError(MaxRetryError("HTTPSConnectionPool(https://doi.org/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016/10.1016
requests to https://twitter.com/search?f=live&lang=en&q=sonda%C5%BC+since%3A2025-05-04&src=spelling_expansion_revert_click failed, giving up.
rrors: SSLError(MaxRetryError("HTTPSConnectionPool(host='twitter.com', port=443): Max retries exceeded with url: /search?f=live&lang=en&g=sonda%C5%BC+since%3A2025-05-04&src=
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\twitter\pobieranie_twitter.py", line 35, in <module>
   tweets = pobierz_tweety(search_term, max_tweets, date_since)
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\twitter\pobieranie_twitter.py", line 18, in pobierz_tweety
   for i, tweet in enumerate(sntwitter.TwitterSearchScraper(query).get_items()):
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\modules\twitter.py", line 1763, in get_items
   for obj in self._iter_api_data('https://twitter.com/i/api/graphql/7jT5GT59P8IFjgxwqnEdQw/SearchTimeline', _TwitterAPIType.GRAPHQL, params, paginationParams, cursor = self.
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\modules\twitter.py", line 915, in _iter_api_data
   obj = self._get_api_data(endpoint, apiType, reqParams, instructionsPath = instructionsPath)
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\modules\twitter.py", line 883, in _get_api_data
   self._ensure_guest_token()
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\modules\twitter.py", line 825, in _ensure_guest
File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\base.py", line 275, in _get
   return self._request('GET', *args, **kwargs)
 File "C:\Marta\Studia\semestr_4\inteligencja_obliczeniowa\projekt-analiza-sondazy-wyborczych\.venv\lib\site-packages\snscrape\base.py", line 271, in _request
  scrape.base.ScraperException: 4 requests to https://twitter.com/search?f=live&lang=en&q=sonda%C5%BC+since%3A2025-05-04&src=spelling_expansion_revert_click failed, giving up
     ess finished with exit code 1
```

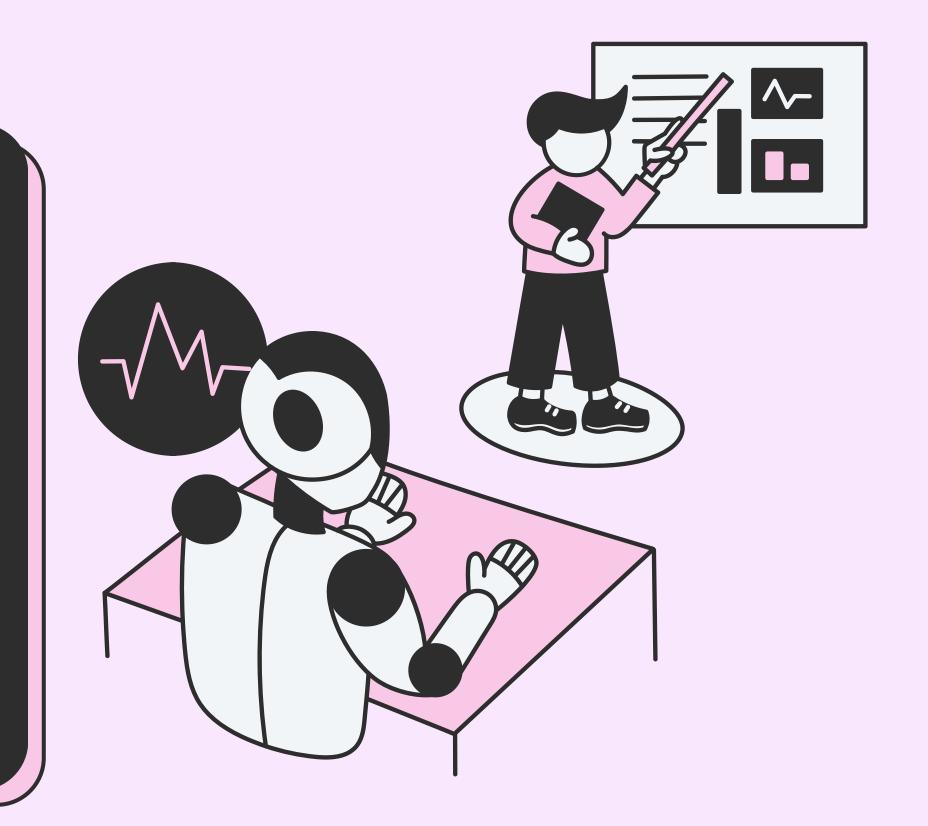
ŁĄCZNIE UDAŁO MI SIĘ POBRAĆ 7456 POSTÓW NA TEMAT TEGOROCZNYCH SONDAZY WYBORCZYCH

- Zliczanie postów z plików Reddit:
- reddit_bartoszewicz_Polska.csv: 9 postów
- reddit_biejat_Polska.csv: 48 postów
- reddit_braun_Polska.csv: 49 postów
- reddit_hołownia_Polska.csv: 10 postów
- reddit_jakubiak_Polska.csv: 14 postów
- reddit_maciak_Polska.csv: 15 postów
- reddit_mentzen_Polska.csv: 52 postów
- reddit_nawrocki_Polska.csv: 179 postów
- reddit_prezydent_Polska.csv: 27 postów
- reddit_senyszyn_Polska.csv: 31 postów
- reddit_sondaże_Polska.csv: 14 postów
- reddit_sondaż_Polska.csv: 14 postów
- reddit_stanowski_Polska.csv: 26 postów
- reddit_trzaskowski_Polska.csv: 100 postów
- reddit_woch_Polska.csv: 11 postów
- reddit_wybory_Polska.csv: 73 postów
- reddit_zandberg_Polska.csv: 51 postów
- Zliczanie postów z plików Bluesky:
- bluesky_nawrocki.csv: 3382 postów
- bluesky_sondaż.csv: 593 postów
- bluesky_trzaskowski.csv: 995 postów
- bluesky_wybory prezydenckie.csv: 497 postów
- bluesky_wybory2025.csv: 1266 postów
- Podsumowanie:
- Łączna liczba postów z Reddita: 723
- Łączna liczba postów z Bluesky: 6733
- ☑ Całkowita liczba postów: 7456

Process finished with exit code 0

Kolejny problem - a mianowicie boty...

Bluesky, choć okazało się cennym źródłem danych do analizy, niestety charakteryzuje się znaczną aktywnością ze strony botów. Tę dużą obecność automatycznie generowanych lub powielanych treści należało wziąć pod uwagę, wdrażając dodatkowe mechanizmy filtrowania. Bez tego działania istniało ryzyko, że zniekształcone dane mogłyby wprowadzić błędy w wynikach analizy sentymentu i tematów.



PO PRZEFILTROWANIU BOTÓW ZOSTAJEMY Z WYNIKIEM 6480 POSTÓW NA TEMAT TEGOROCZNYCH SONDAZY WYBORCZYCH

- Zliczanie postów z plików Reddit:
- reddit_bartoszewicz_Polska.csv: 9 postów
- reddit_biejat_Polska.csv: 48 postów
- reddit_braun_Polska.csv: 49 postów
- reddit_hołownia_Polska.csv: 10 postów
- reddit_jakubiak_Polska.csv: 14 postów
- reddit_maciak_Polska.csv: 15 postów
- reddit_mentzen_Polska.csv: 52 postów
- reddit_nawrocki_Polska.csv: 179 postów
- reddit_prezydent_Polska.csv: 27 postów
- reddit_senyszyn_Polska.csv: 31 postów
- reddit_sondaże_Polska.csv: 14 postów
- reddit_sondaż_Polska.csv: 14 postów
- reddit_stanowski_Polska.csv: 26 postów
- reddit_trzaskowski_Polska.csv: 100 postów
- reddit_woch_Polska.csv: 11 postów
- reddit_wybory_Polska.csv: 73 postów
- reddit_zandberg_Polska.csv: 51 postów
- Zliczanie postów z plików Bluesky:
- przefiltrowany_bluesky.csv: 5757 postów
- Podsumowanie:
- · Łączna liczba postów z Reddita: 723
- Łączna liczba postów z Bluesky: 5757
- Całkowita liczba postów: 6480

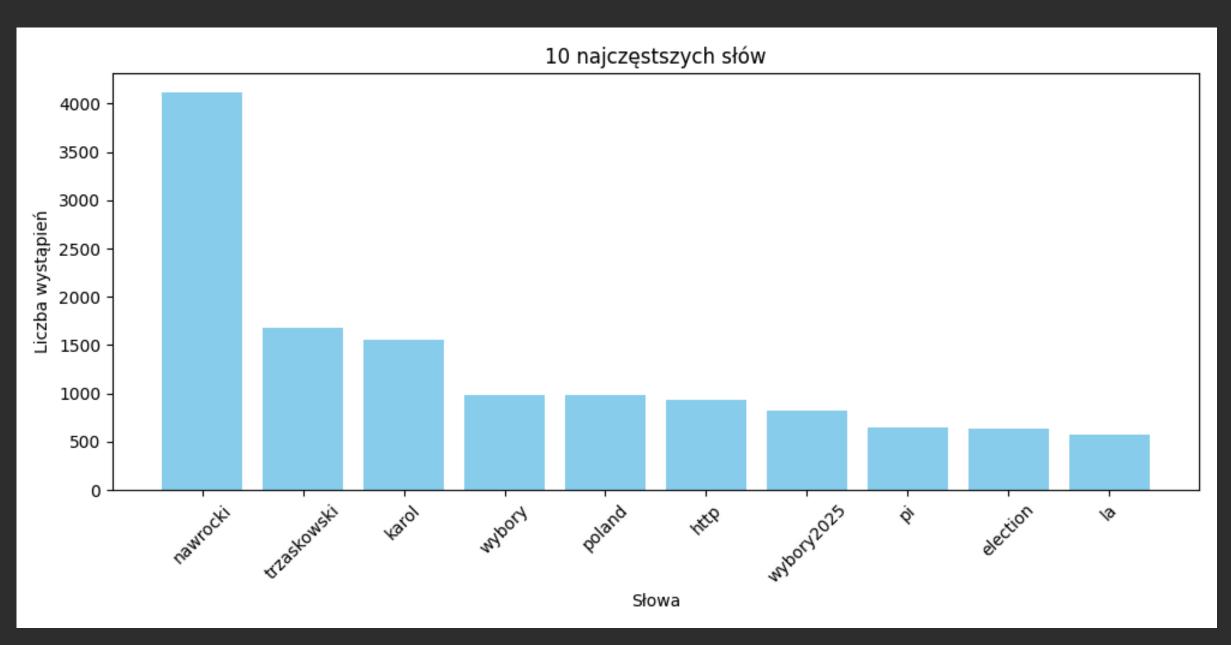
Process finished with exit code 0

Tokenizacja, lematyzacja, częstość występowania słów, chmura tagów...



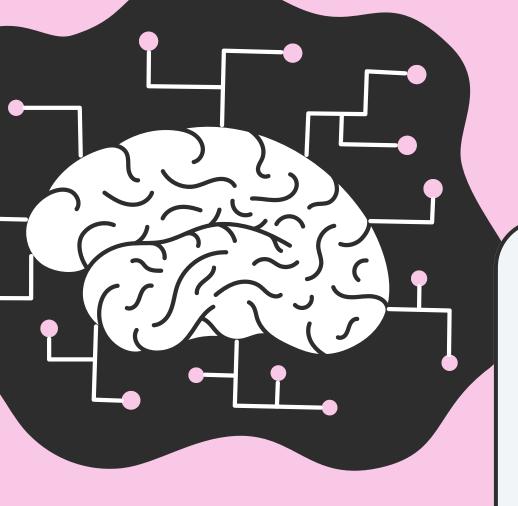
Na tym etapie skupiłam się na przetworzeniu tekstów zebranych z **Reddita** i **Bluesky**. Po wczytaniu danych połączyłam wszystkie treści i wykonałam **tokenizację**, czyli podział treści na zdania i pojedyncze słowa. Następnie oczyściłam dane, usuwając tzw. **stopwords** – najczęściej występujące słowa, które nie mają większego znaczenia analitycznego. W tym celu wykorzystałam wbudowaną listę angielskich stopwords z biblioteki NLTK oraz dodatkową listę polskich słów pobraną z GitHuba. Uzupełniłam ją także o znaki przestankowe i popularne słowa neutralne. Kolejnym krokiem była **lematyzacja**, dzięki której sprowadziłam wyrazy do ich podstawowych form. Na koniec policzyłam najczęściej występujące słowa i przygotowałam wizualizacje – **wykres słupkowy** i **chmurę słów (word cloud)** – które ułatwiają interpretację najważniejszych tematów pojawiających się w postach.

Wykres słupkowy (word frequency)



Chmura słów (word cloud)







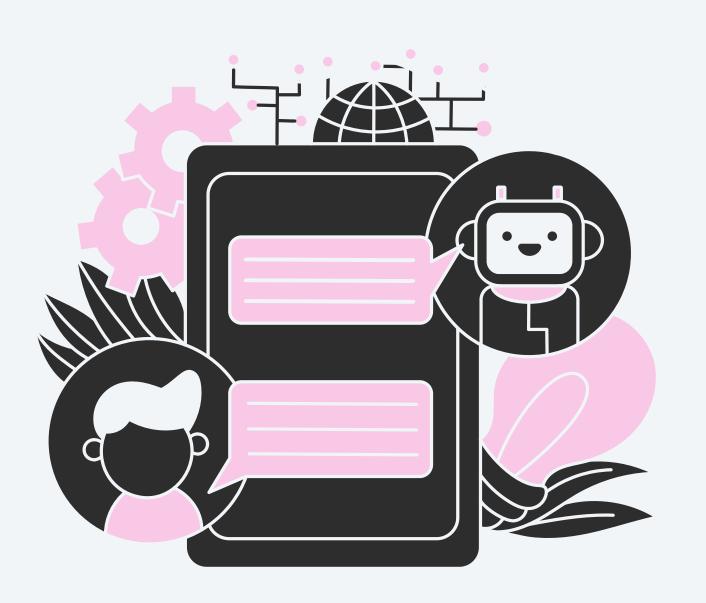
zrozumienie nastrojów

wyborców online









W dobie kampanii wyborczych, internet staje się główną areną debaty. Aby zrozumieć, co naprawdę myślą i czują wyborcy, przeprowadzamy analizę sentymentu postów z mediów społecznościowych.

Używamy do tego **VADER**, wyspecjalizowanego narzędzia z biblioteki NLTK, które doskonale radzi sobie z niuansami języka internetowego.



Wyniki analizy

1. Stosunek wyborców online do każdego z kandydatów (dane z Reddita)

2. Stosunek wyborców online dla kandydatów II tury

3. Ogólny stosunek internautów do wyborów

```
=== BARTOSZEWICZ_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 7
negatywny 2
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.292

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.008
Negatywny (neg): 0.031
Neutralny (neu): 0.961
Zagregowany wynik (compound): -0.992
```

Bartoszewicz Artur

```
=== BIEJAT_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 1
neutralny 28
negatywny 19
Name: count, dtype: intó4
Średni compound_score (średnia z każdego posta): -0.354

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.006
Negatywny (neg): 0.027
Neutralny (neu): 0.967
Zagregowany wynik (compound): -0.999
```

Biejat Magdalena

```
=== BRAUN_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 3
neutralny 36
negatywny 10
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.212

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.005
Negatywny (neg): 0.022
Neutralny (neu): 0.972
Zagregowany wynik (compound): -0.998
```

Braun Grzegorz

```
=== HOŁOWNIA_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 6
negatywny 4
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.401

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.009
Negatywny (neg): 0.036
Neutralny (neu): 0.955
Zagregowany wynik (compound): -0.997
```

Hołownia Szymon

```
=== JAKUBIAK_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 13
negatywny 1
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.165

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.008
Negatywny (neg): 0.032
Neutralny (neu): 0.961
Zagregowany wynik (compound): -0.990
```

Jakubiak Marek

```
=== MACIAK_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 1
neutralny 10
negatywny 4
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.266

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.008
Negatywny (neg): 0.028
Neutralny (neu): 0.964
Zagregowany wynik (compound): -0.996
```

Maciak Maciej

```
=== MENTZEN_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 1
neutralny 47
negatywny 4
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.129

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.005
Negatywny (neg): 0.020
Neutralny (neu): 0.974
Zagregowany wynik (compound): -0.990
```

Mentzen Sławomir

```
=== NAWROCKI_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 12
neutralny 120
negatywny 47
Name: count, dtype: intó4
Średni compound_score (średnia z każdego posta): -0.242

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.007
Negatywny (neg): 0.025
Neutralny (neu): 0.967
Zagregowany wynik (compound): -1.000
```

Nawrocki Karol

```
=== SENYSZYN_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 24
negatywny 7
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.265

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.007
Negatywny (neg): 0.032
Neutralny (neu): 0.961
Zagregowany wynik (compound): -0.995
```

Senyszyn Joanna

```
=== STANOWSKI_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 21
negatywny 5
Name: count, dtype: intó4
Średni compound_score (średnia z każdego posta): -0.316

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.007
Negatywny (neg): 0.028
Neutralny (neu): 0.966
Zagregowany wynik (compound): -0.996
```

Stanowski Krzysztof

```
=== TRZASKOWSKI_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 7
neutralny 68
negatywny 25
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.213

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.009
Negatywny (neg): 0.023
Neutralny (neu): 0.968
Zagregowany wynik (compound): -0.999
```

Trzaskowski Rafał

```
=== WOCH_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 1
neutralny 9
negatywny 1
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.053

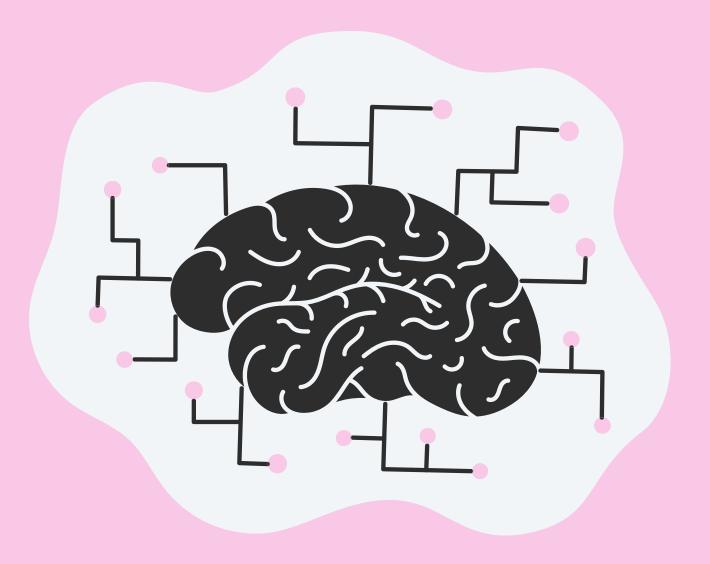
Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.010
Negatywny (neg): 0.031
Neutralny (neu): 0.959
Zagregowany wynik (compound): -0.987
```

Woch Marek

```
=== ZANDBERG_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 3
neutralny 39
negatywny 9
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.202

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.005
Negatywny (neg): 0.018
Neutralny (neu): 0.977
Zagregowany wynik (compound): -0.995
```

Zandberg Adrian



2.

Rafał Trzaskowski

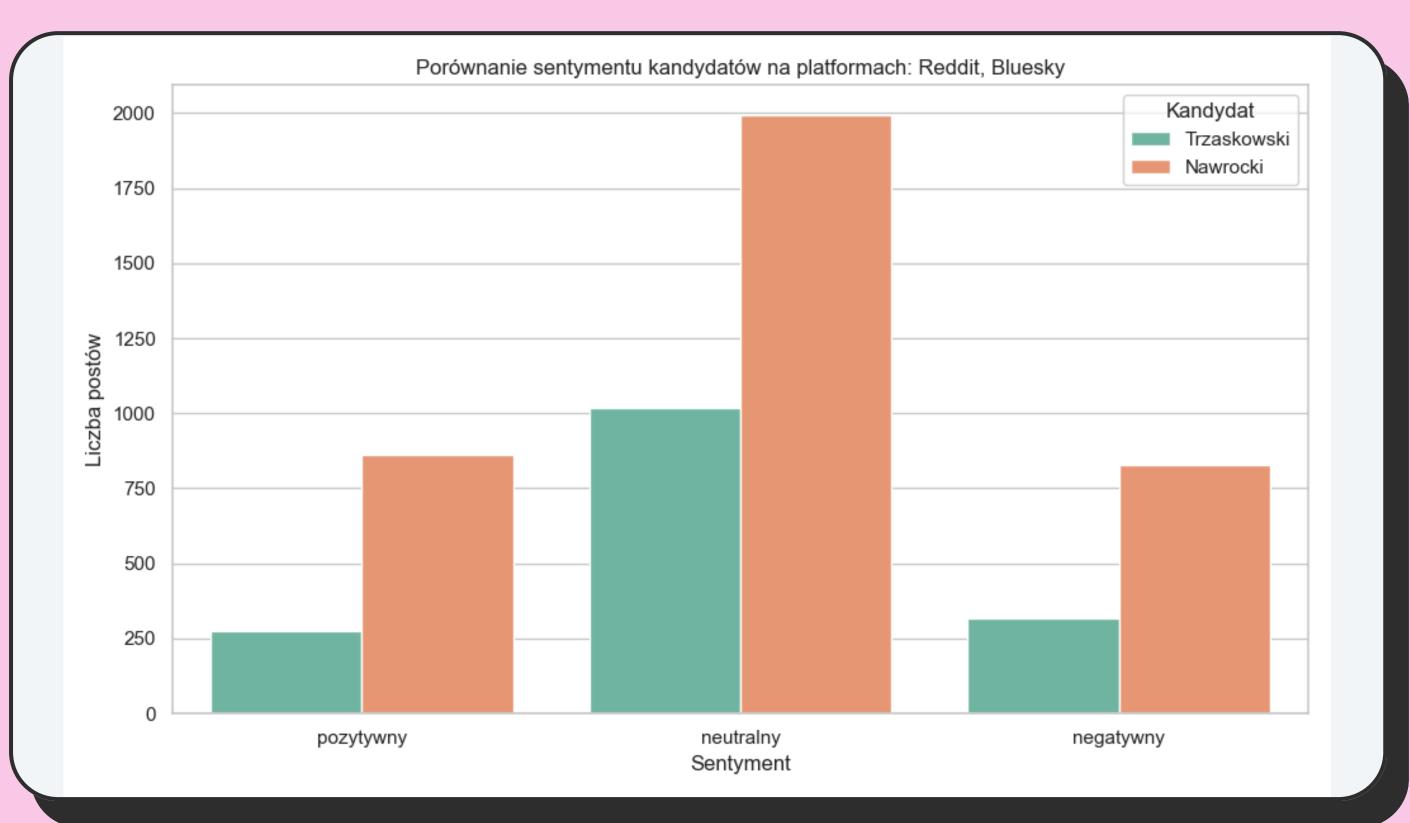
```
=== TRZASKOWSKI_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
             7
pozytywny
neutralny
            68
            25
negatywny
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.213
Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.009
Negatywny (neg): 0.023
Neutralny (neu): 0.968
Zagregowany wynik (compound): -0.999
```



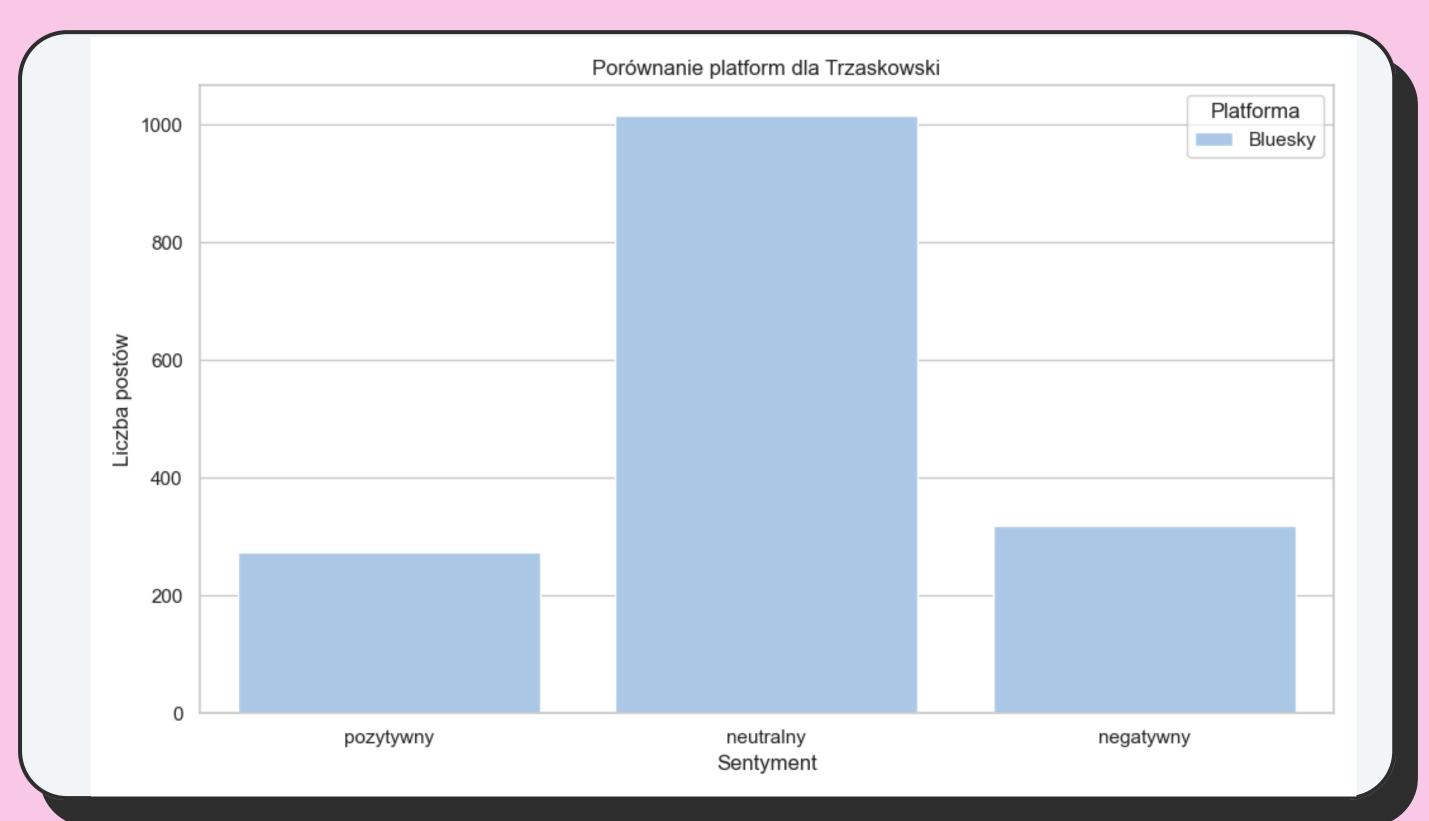
```
=== TRZASKOWSKI (BLUESKY) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny
              274
            1016
neutralny
              318
negatywny
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.014
Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.049
Negatywny (neg): 0.052
Neutralny (neu): 0.898
Zagregowany wynik (compound): -0.999
```



Wykresy



Wykresy



Karol Nawrocki

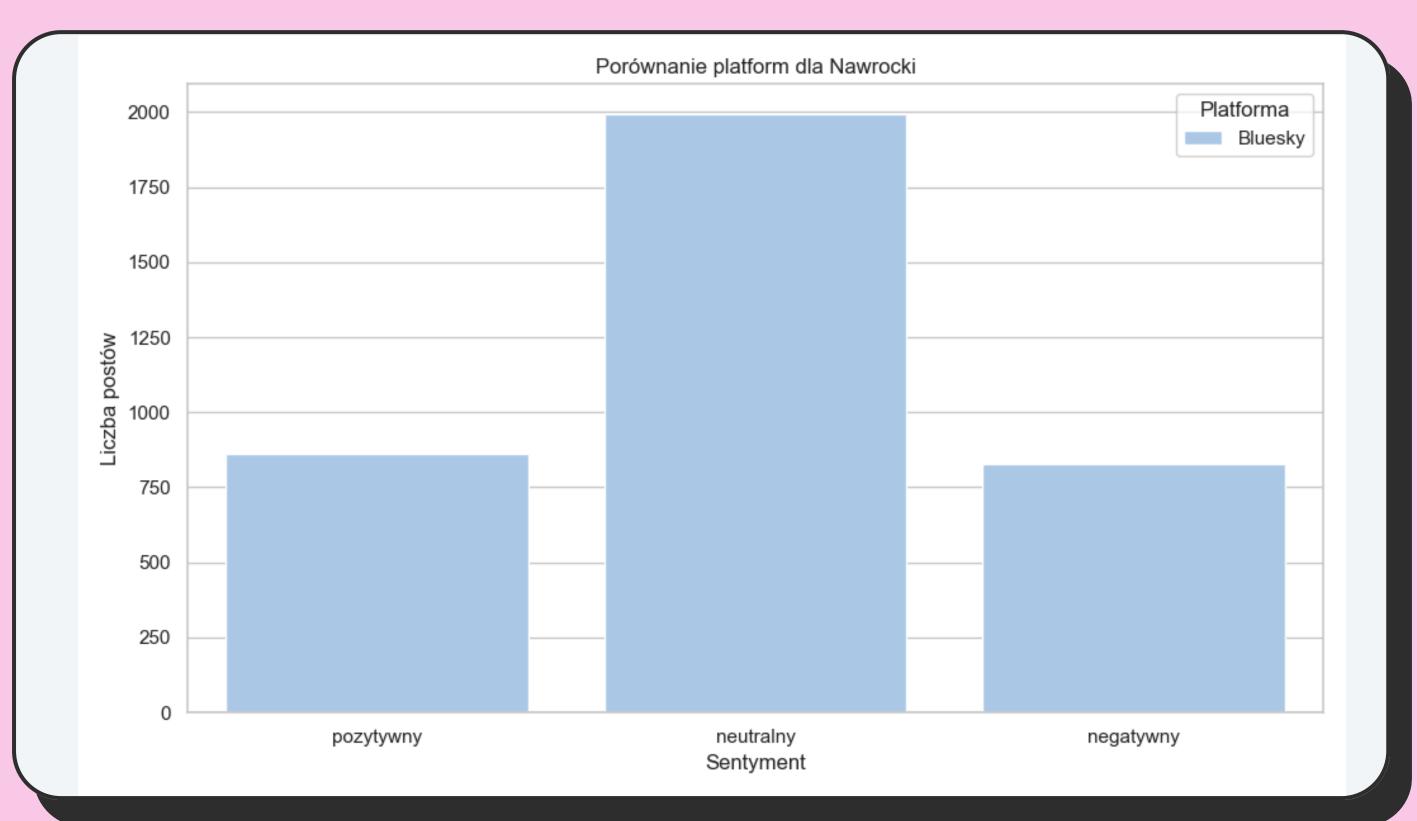
```
=== NAWROCKI_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
             12
pozytywny
            120
neutralny
             47
negatywny
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.242
Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.007
Negatywny (neg): 0.025
Neutralny (neu): 0.967
Zagregowany wynik (compound): -1.000
```



```
=== NAWROCKI (BLUESKY) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
              860
pozytywny
neutralny
            1993
              826
negatywny
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): 0.008
Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.067
Negatywny (neg): 0.065
Neutralny (neu): 0.868
Zagregowany wynik (compound): 1.000
```



Wykresy



```
=== WYBORY_POLSKA (REDDIT) ===
Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 5
neutralny 42
negatywny 26
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.330

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.006
```

Negatywny (neg): 0.027

Neutralny (neu): 0.968

Zagregowany wynik (compound): -1.000

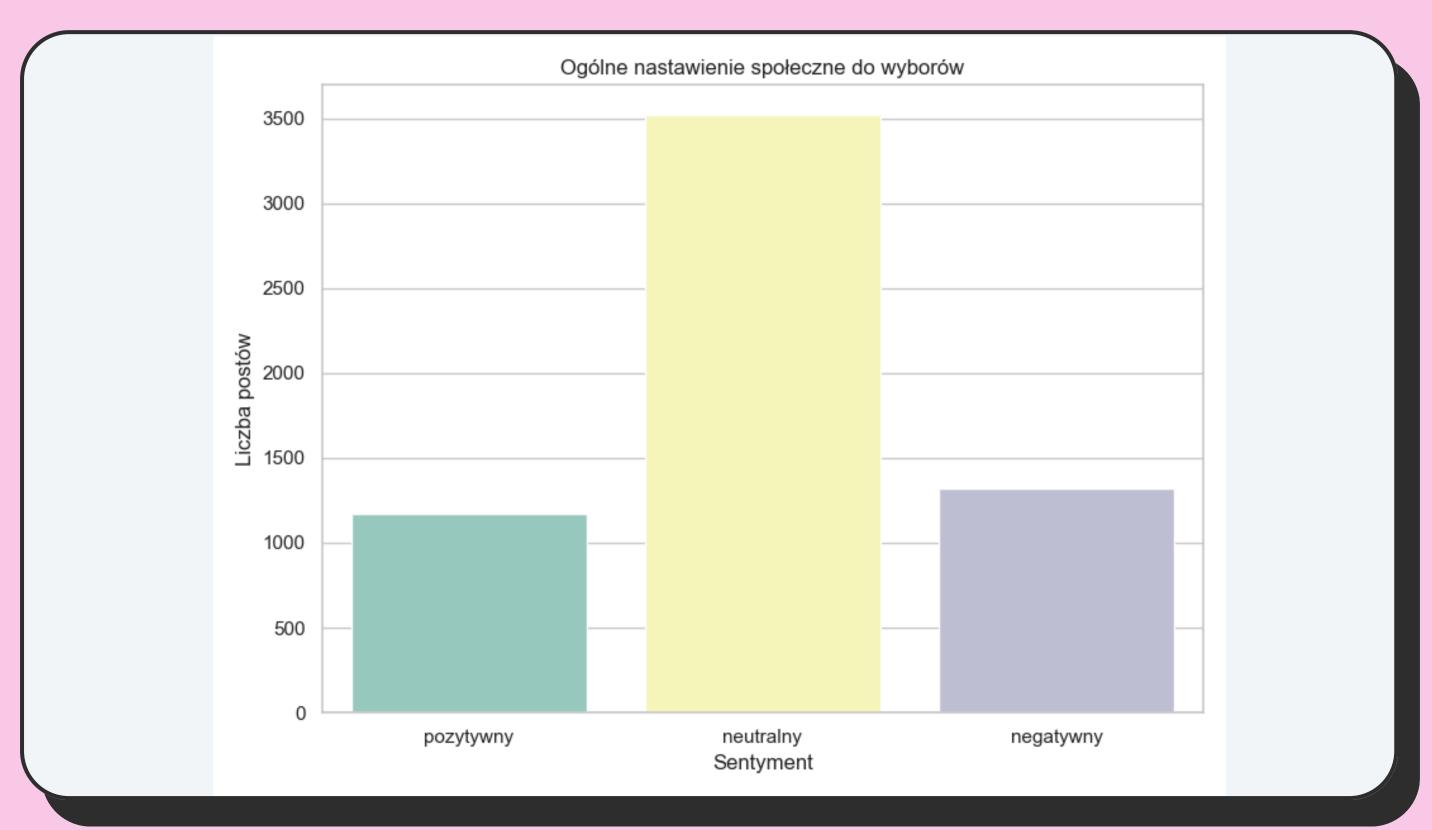


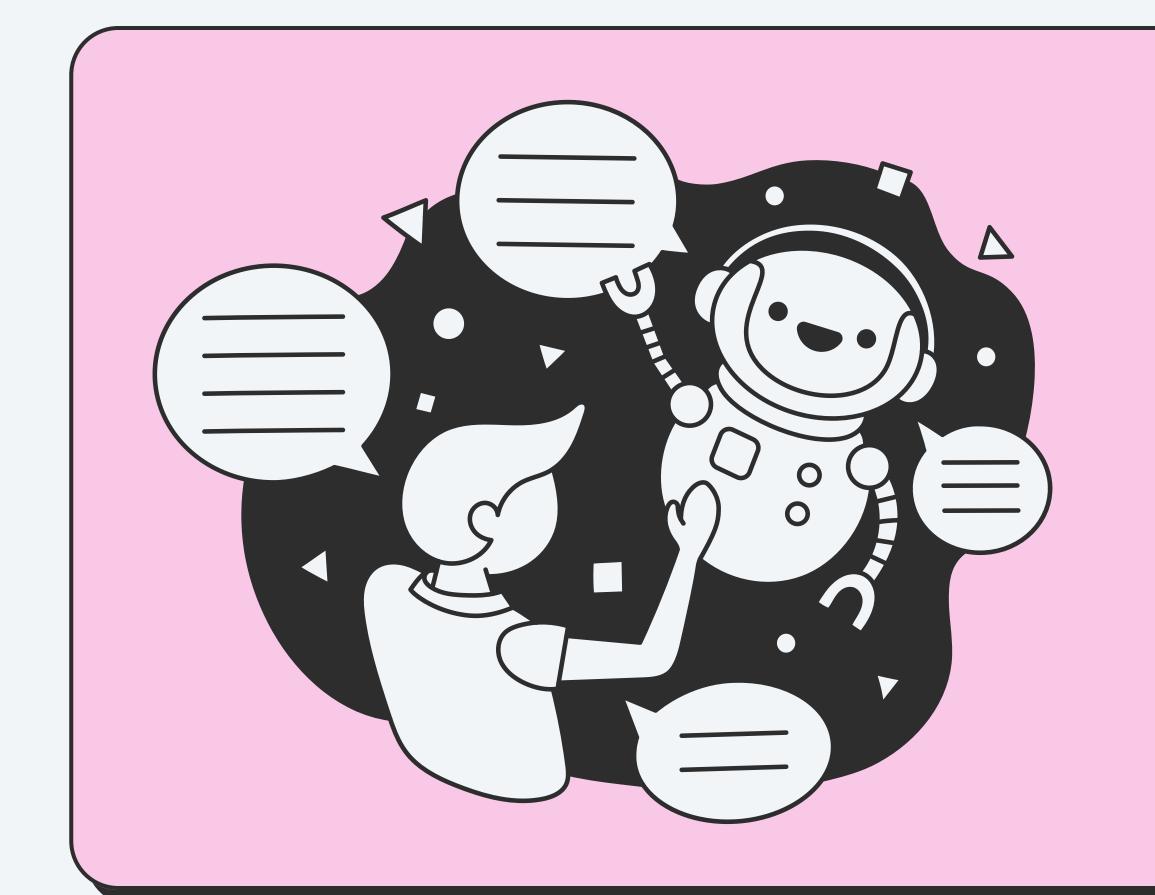
```
=== SONDAŻ_POLSKA (REDDIT) ===

Liczba postów wg klasyfikacji sentymentu (analiza każdego posta osobno):
sentiment
pozytywny 0
neutralny 12
negatywny 2
Name: count, dtype: int64
Średni compound_score (średnia z każdego posta): -0.359

Zagregowane wyniki sentymentu (cały tekst jako jedna wypowiedź):
Pozytywny (pos): 0.000
Negatywny (neg): 0.054
Neutralny (neu): 0.946
Zagregowany wynik (compound): -0.929
```

Wykresy





Dziękuję za

uwagę!!