# NA CO NARZEKAJĄ MIESZKAŃCY MIAST?

*Tomasz Dróżdż* *Marcin Przymus* *Tomasz Pach*

Katedra Sztucznej Inteligencji

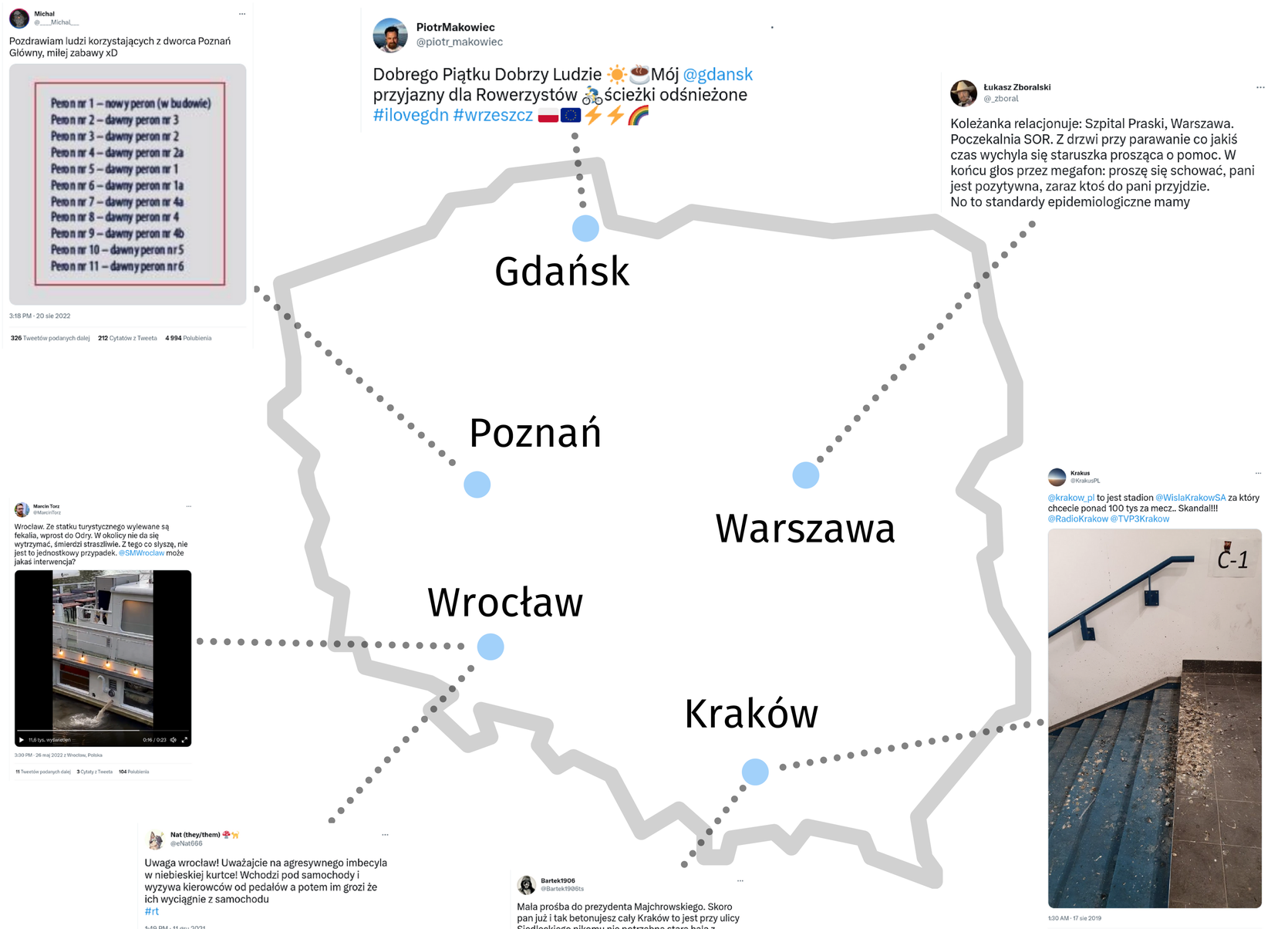
Politechnika Wrocławska

## Opis problemu

Każde miasto ma swoją specyfikę, atrakcje i problemy. Na temat większości z nich krążą również pewne stereotypy, które trudno zweryfikować. Postanowiliśmy sprawdzić w jaki sposób piszą użytkownicy Twittera piszą o największych polskich miastach.

## Dane

Korzystając ze scrappera *snscrape*, na podstawie słów kluczowych zebraliśmy posty i komentarze, które odnosiły się do dużych polskich miast: Warszawy, Wrocławia, Krakowa, Poznania i Gdańska. W sumie uzyskaliśmy 3,3 mln tweetów w języku polskim z okresu 2018 - 2022 o długości co najmniej 15 znaków.



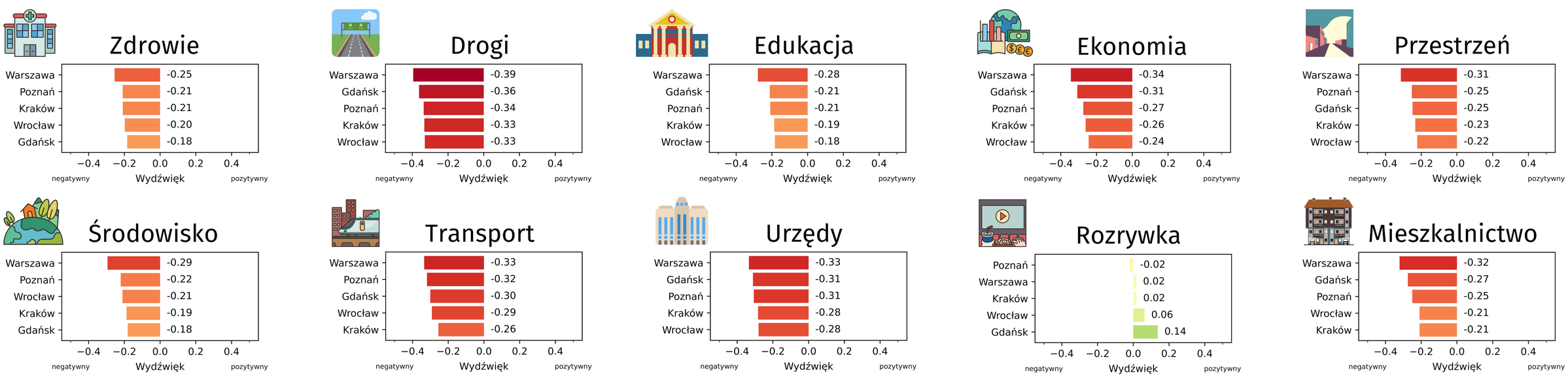

## Metodologia

Ze zbioru danych usunięto tweety wspominające o kilku miastach jednocześnie oraz ręcznie odfiltrowano firmowe konta z największą liczbą tweetów, takie jak stacje telewizyjne i programy informacyjne.

Za pomocą modelu DeBERTA przystosowanego do zadania NLI (Natural Language Inference) określiliśmy przynależność każdego tweeta do jednego z 10 wybranych przez nas tematów. Tweety, które nie zostały w ten sposób przypisane do chociaż jednego z tematów zostały odrzucone. Ostatecznie do dalszej analizy wykorzystano 862 tys. tweetów z pierwotnego zbioru.

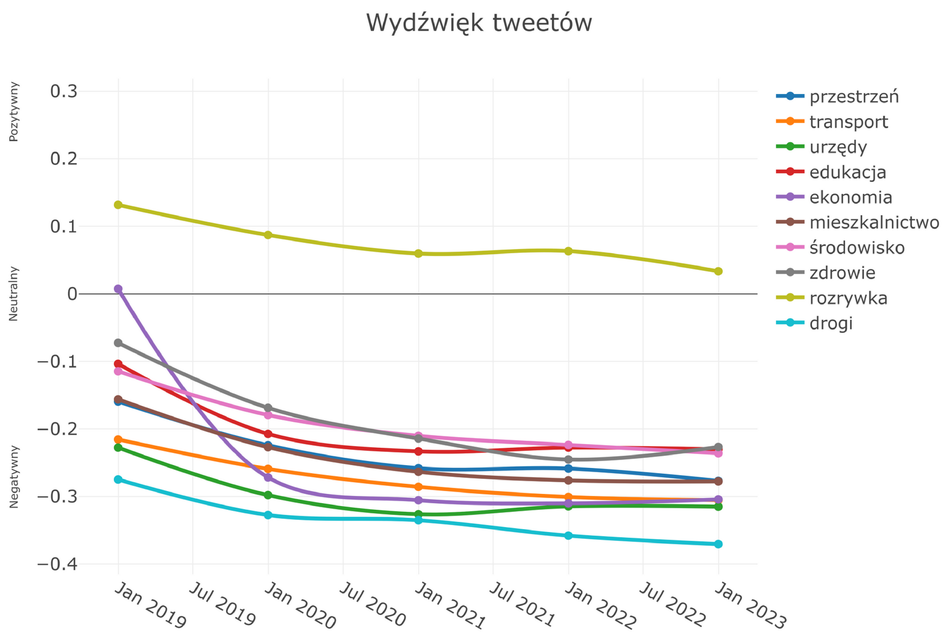
W ramach ostatniego etapu procesu przetwarzania danych określony został wydźwięk wypowiedzi za pomocą modelu opartego na HerBERTcie z biblioteki *sentimentPL*. Każdemu tweetowi przypisana została wartość od -1 (wypowiedź o wydźwięku skrajnie negatywnym) do 1 (wypowiedź o wydźwięku skrajnie pozytywnym).


## Średni wydźwięk tweetów z roku 2022




## Jak nastroje zmieniały się w czasie?

Wydźwięk wypowiedzi spada dla wszystkich kategorii, niezależnie od miasta. Analiza pokazuje, że nastroje w internecie są coraz gorsze. Wyraźne pogorszenie następuje do 2021 roku. Najbardziej widoczny spadek jest w kategorii ekonomia. W 2021 trend spadkowy ustał i sentyment we wszystkich kategoriach pozostaje na podobnym poziomie. Występują niewielkie wzrosty lub spadki, jednak ich wielkość nie wskazuje na to, by obecne wartości mogły się wyraźnie zmienić w najbliższej przyszłości.




## Ciekawostki

- Najniższy średni wydźwięk wśród tweetów dotyczących Wrocławia miały te napisane z urządzeń z systemem Android.
- Średni wydźwięk wypowiedzi dotyczących Wrocławia użytkowników zweryfikowanych jest bliski 0, a za ujemny sentyment wypowiedzi odpowiadają niezweryfikowani użytkownicy.
- Tweety o negatywnym wydźwięku dotyczące Wrocławia są tak samo często lajkowane jak pozostałe.
- Z czasem rośnie liczba użytkowników Twittera wypowiadających się w danych tematach, ale spada średnia liczba tweetów na użytkownika.