

# Eksploracyjna Analiza Danych Ramki Danych Tweetów Na Temat Koronawirusa

Igor Zawada Bartosz Zabkowski

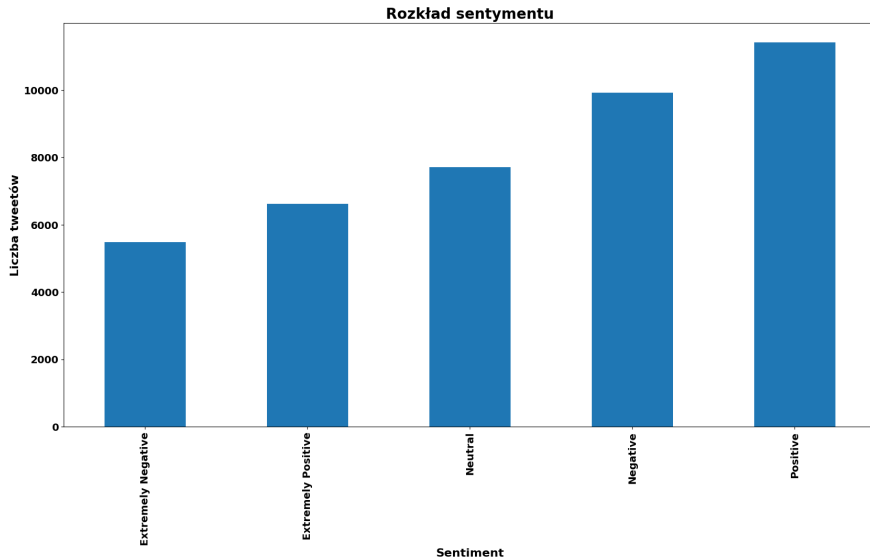
Wydział Matematyki i Nauk Informacyjnych Politechniki Warszawskiej

19 Marca 2025

# Poglad Ramki Danych

	UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment
0	3799	48751	London	16-03-2020	@MeNyrbie @Phil_Gahan @Chrisitv <a href="https://t.co/i...">https://t.co/i...</a>	Neutral
1	3800	48752	UK	16-03-2020	advice Talk to your neighbours family to excha...	Positive
2	3801	48753	Vagabonds	16-03-2020	Coronavirus Australia: Woolworths to give elde...	Positive
3	3802	48754	NaN	16-03-2020	My food stock is not the only one which is emp...	Positive
4	3803	48755	NaN	16-03-2020	Me, ready to go at supermarket during the #COV...	Extremely Negative

# Ustalenie Przewidywanej Zmiennej



# Location

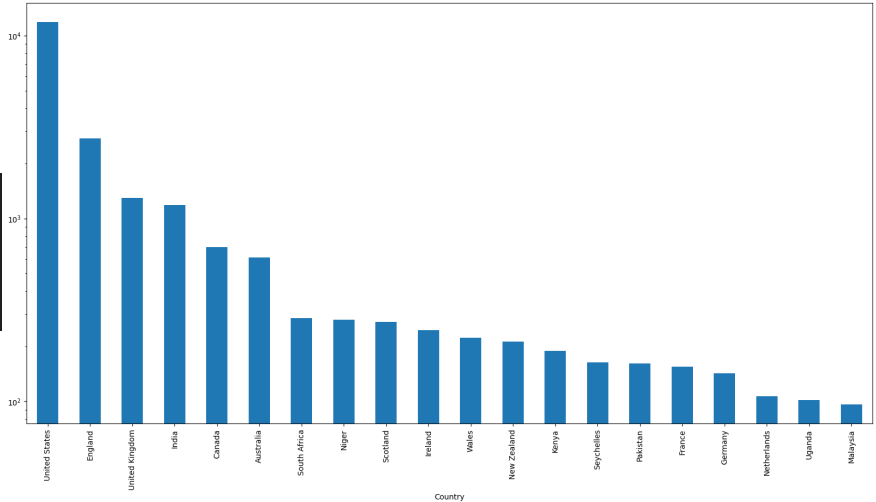
index	Location
20222	Espoo, Suomi
20223	New Jersey
20224	•
20225	1 Yonge St., Toronto, Canada
20226	#TorontoTheGood
20227	•
20228	•
20229	New Delhi, India
20230	Resist ??????
20231	Southeastern Pennsylvania
20232	Denver, CO
20233	•
20234	Green Bay, WI, United States
20235	Raymond, Alberta, Canada
20236	Tulsa, OK
20237	Kampala, Uganda
20238	•
20239	The World

# Location

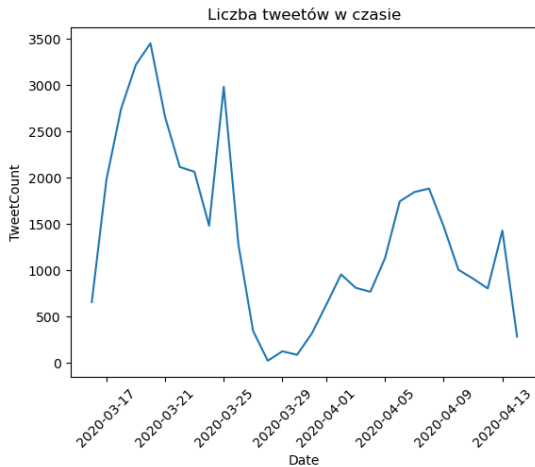
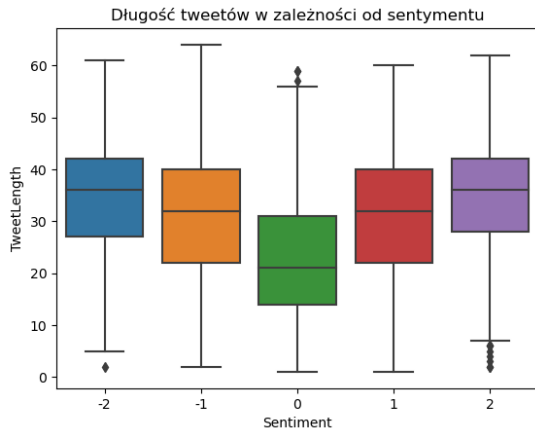
20222	Espoo, Suomi	•
20223	New Jersey	United States
20224	•	•
20225	1 Yonge St., Toronto, Canada	Canada
20226	#TorontoTheGood	•
20227	•	•
20228	•	•
20229	New Delhi, India	India
20230	Resist ?????	•
20231	Southeastern Pennsylvania	United States
20232	Denver, CO	United States
20233	•	•
20234	Green Bay, WI, United States	United States
20235	Raymond, Alberta, Canada	Canada
20236	Tulsa, OK	United States
20237	Kampala, Uganda	Uganda
20238	•	•
20239	The World	•

# Analiza Brakujących Wartości

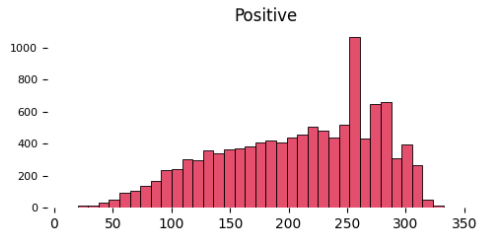
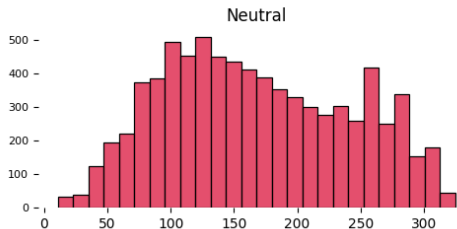
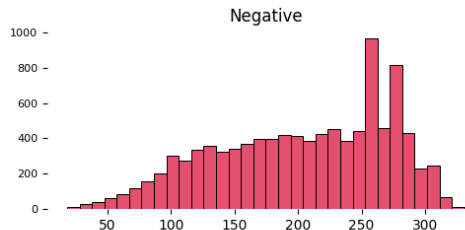
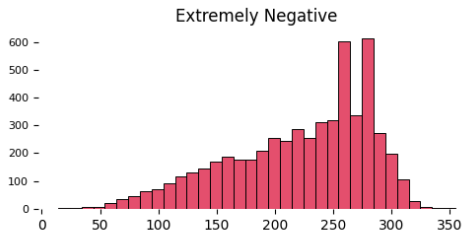
```
UserName      0  
ScreenName    0  
Location      8590  
TweetAt       0  
OriginalTweet 0  
Sentiment     0  
dtype: int64
```



# Poglądowe Wykresy

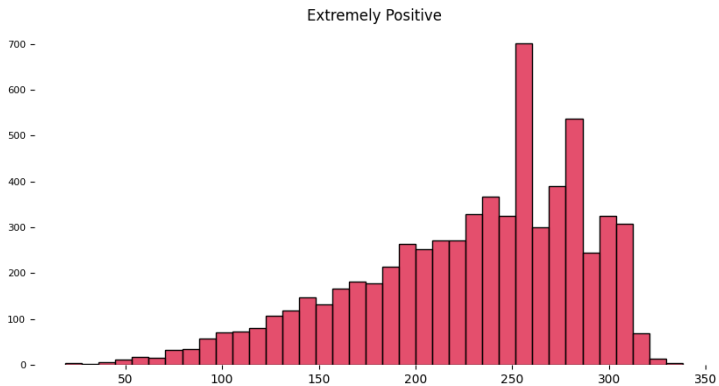


# Długość Tweetów





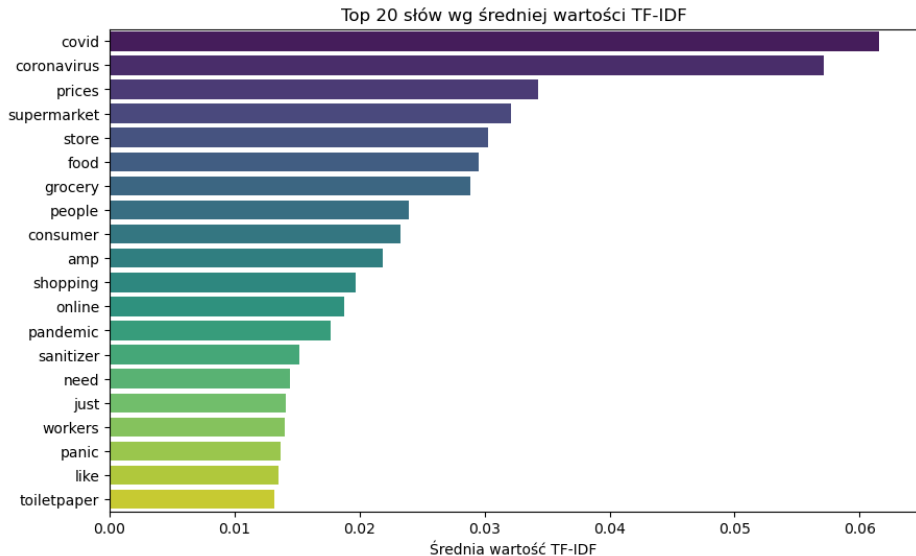
# Długość Tweetów



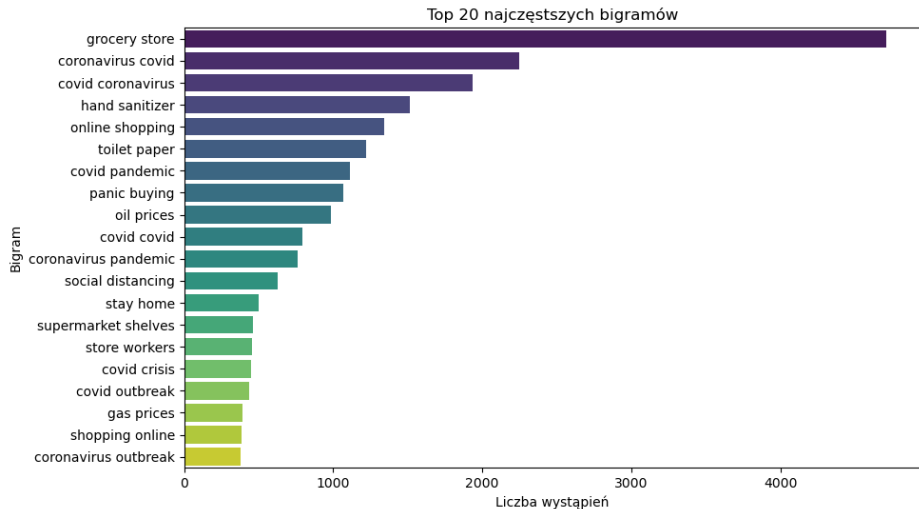
**Kształt macierzy TF-IDF:** (41157, 57082)

Słowo	Wartość
covid	0.061562
coronavirus	0.057175
prices	0.034287
supermarket	0.032125
store	0.030281
food	0.029506
grocery	0.028875
people	0.023961
consumer	0.023298

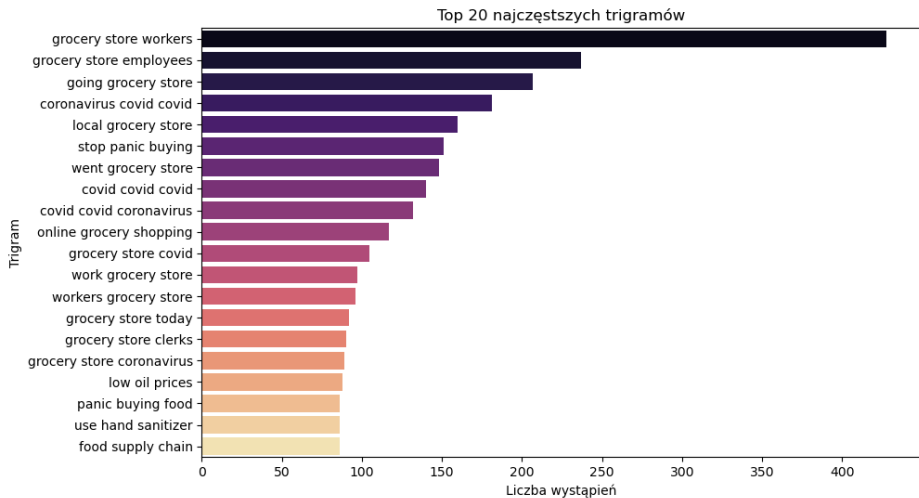
# Wizualizacja TF-IDF



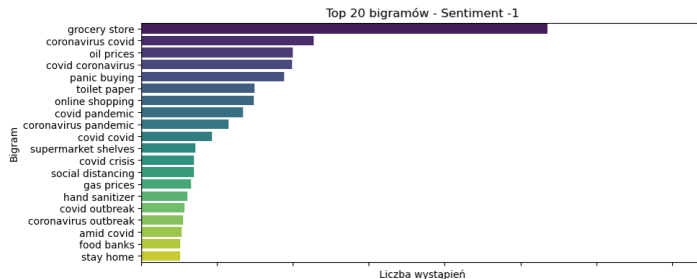
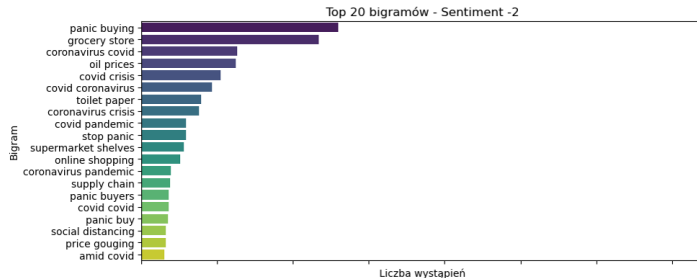
# Bigramy i Trigramy



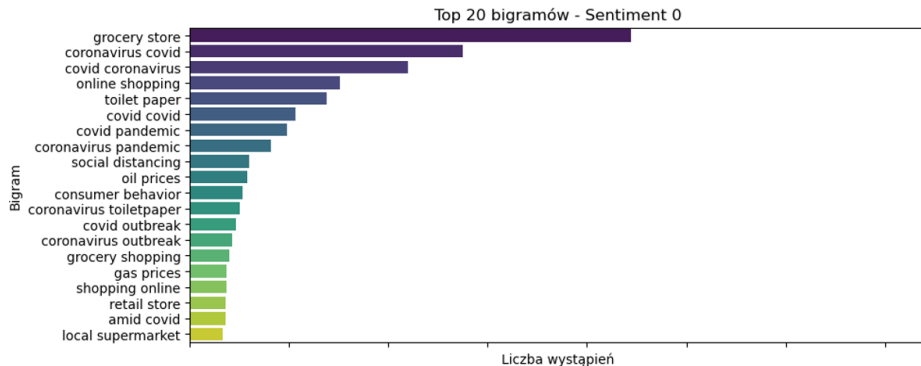
# Bigramy i Trigramy



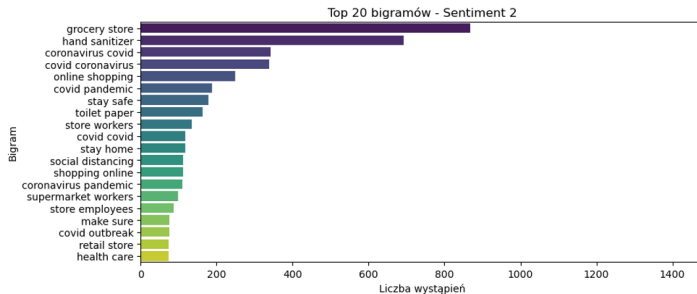
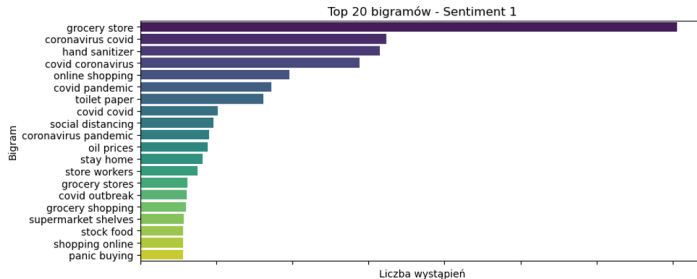
# Bigramy a Sentyment



# Bigramy a Sentyment

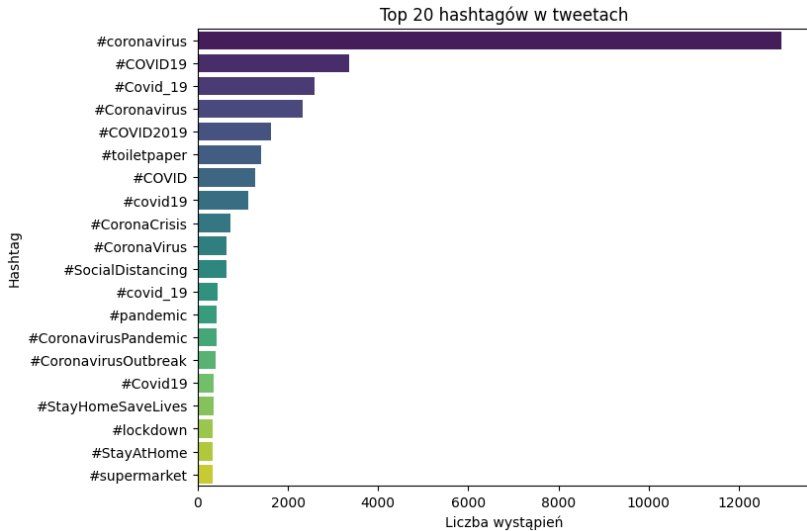


# Bigramy a Sentyment

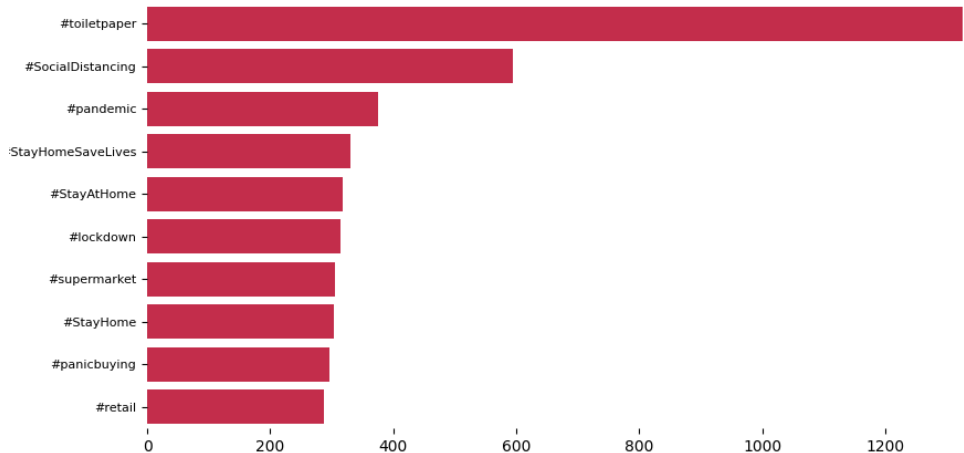




# Analiza Hashtagów

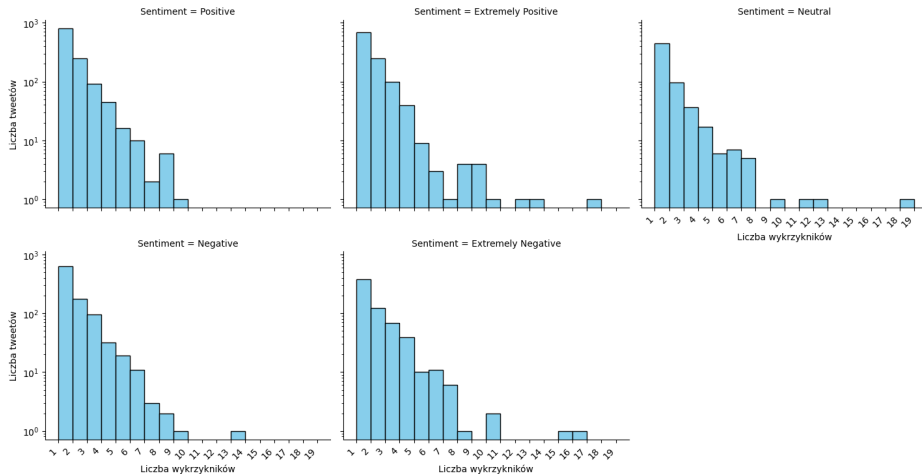


# Analiza Hashtagów

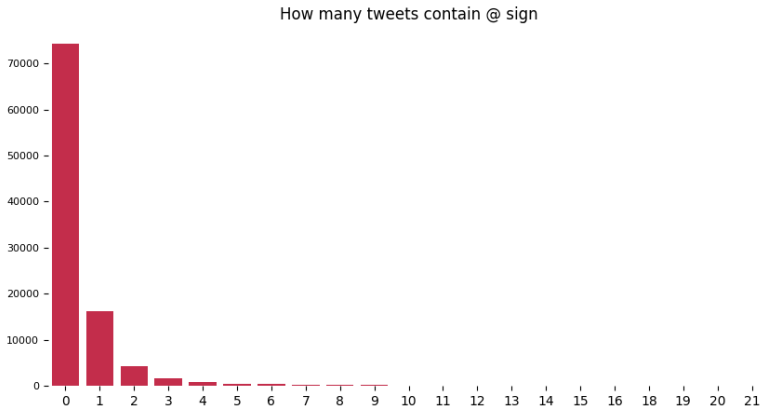


# Znaki specjalne - wykrzyknik

Histogramy liczby wykrzykników w zależności od sentymentu  
(tylko tweety z wykrzyknikami)



# Analiza liczby oznaczeń



**Dziekujemy Za Uwage**