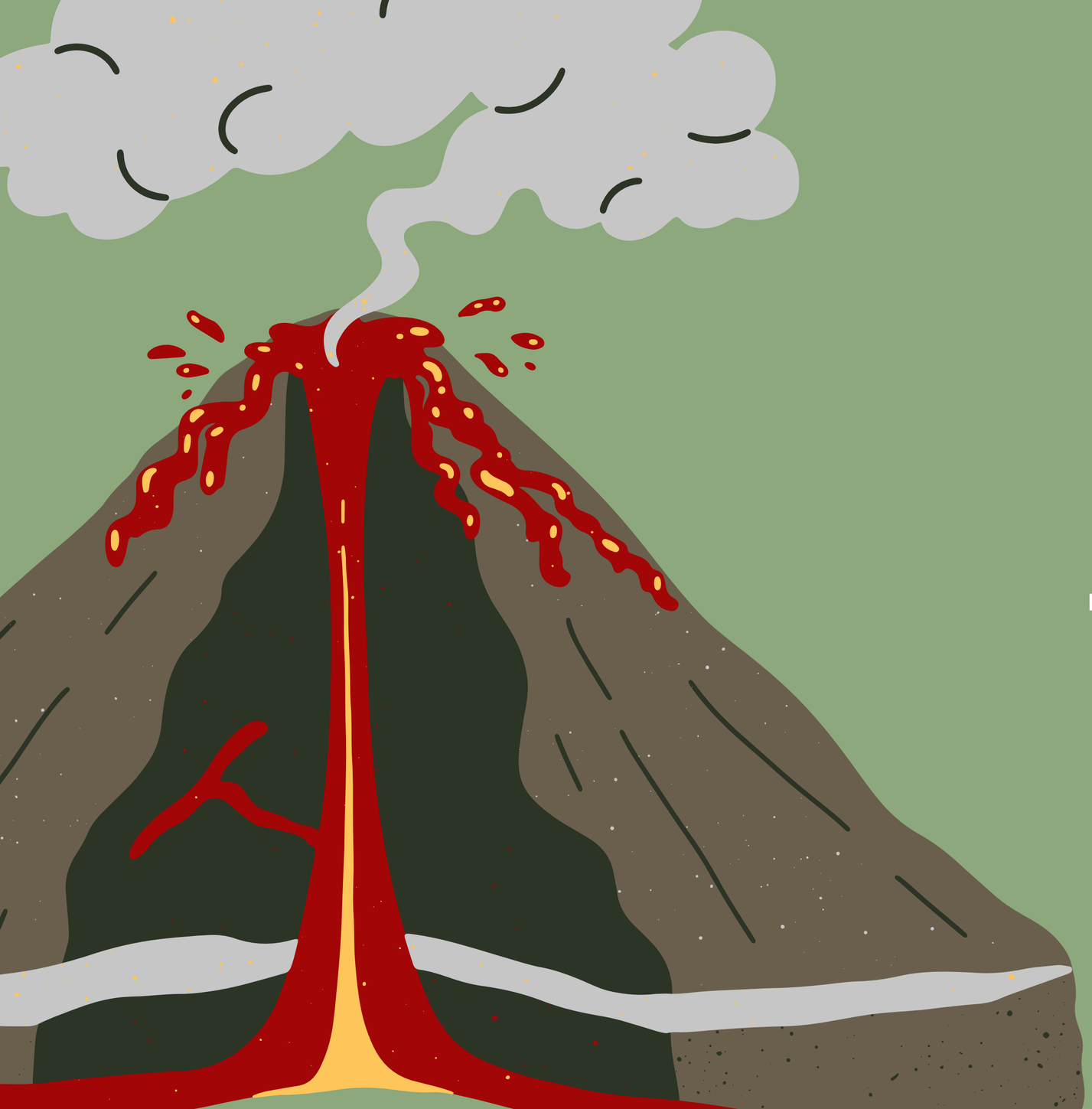


NLP With Disaster Tweets

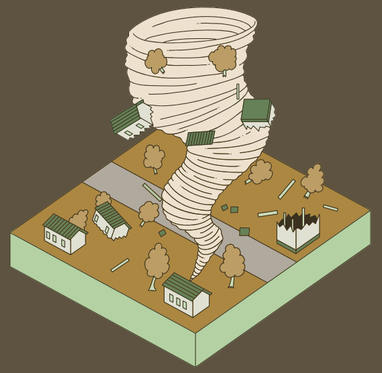
¿WHAT IS NLP?

DATA PREPROCESSING → FEATURE ENGINEERING

DATA PRE-PROCESSING

1

Clean

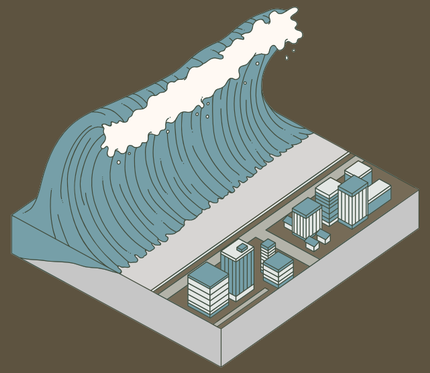


2

Tokenization

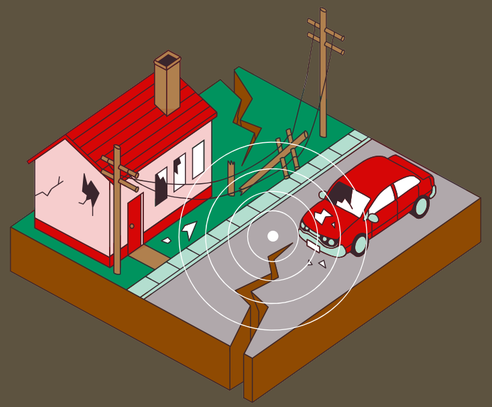
3

Stopwords



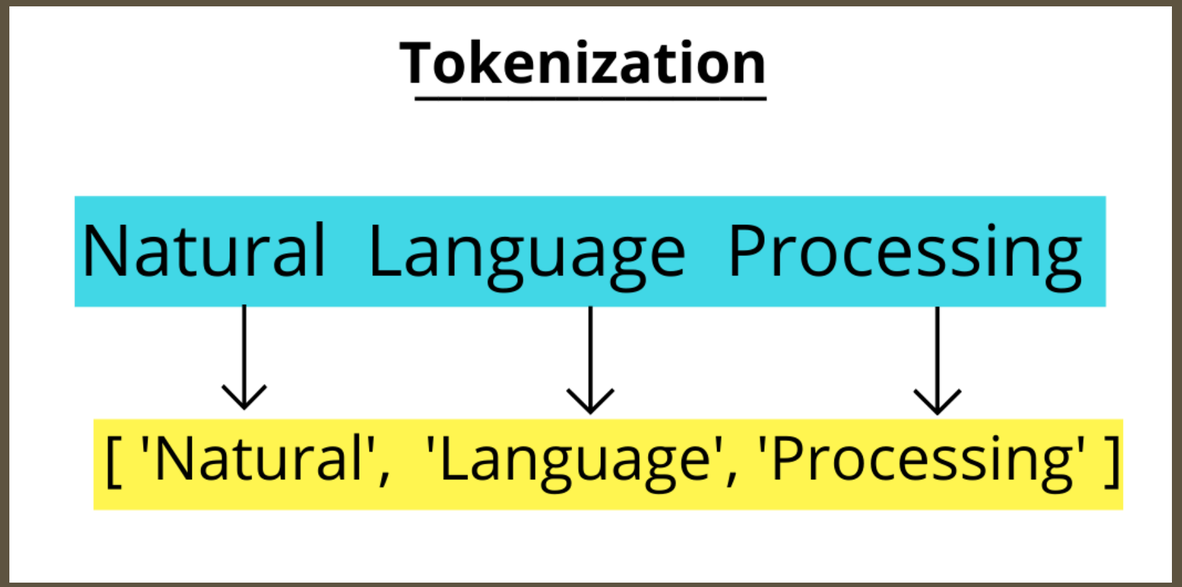
4

Lemmatization

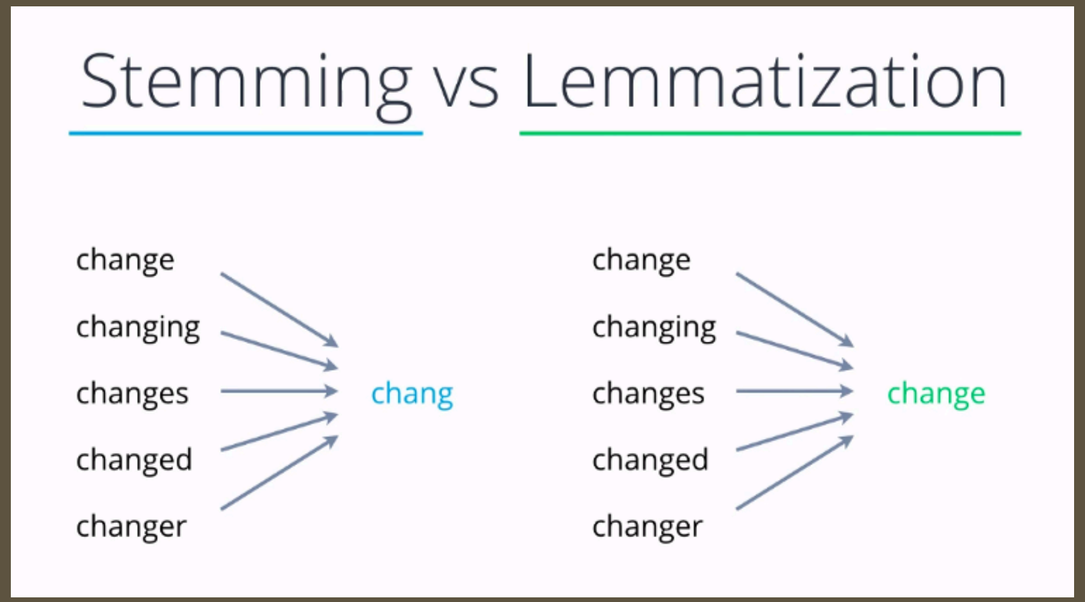


Exemples

Tokenization



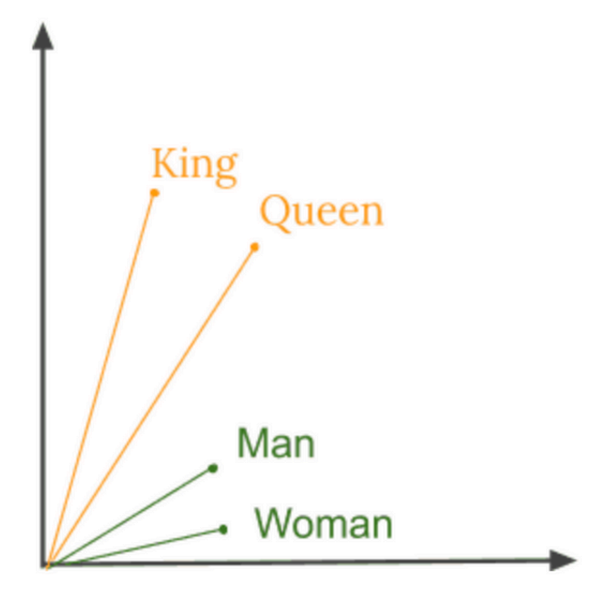
Lemmatization



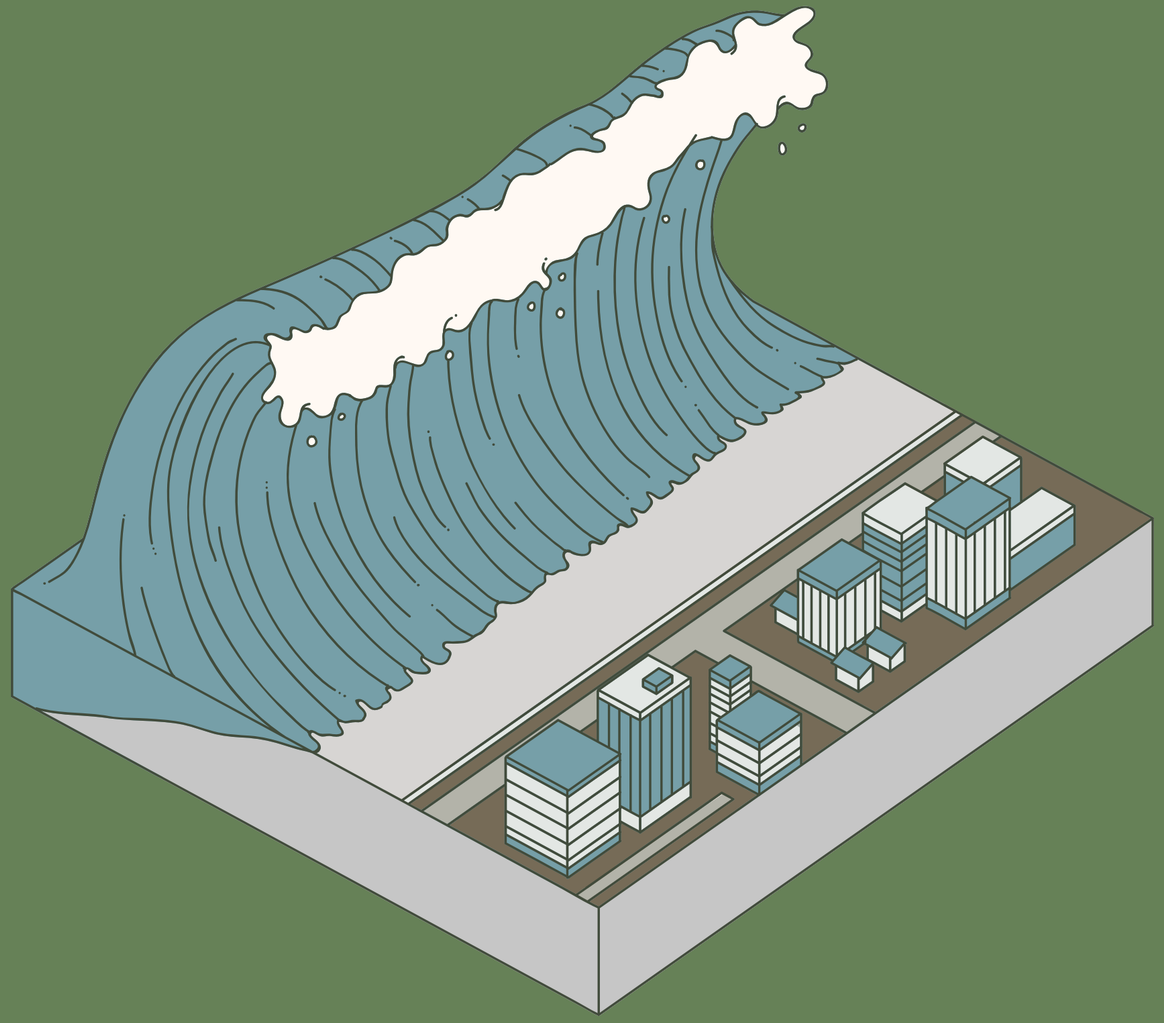
FEATURE ENGINEERING

FEATURE ENGINEERING

Word Embedding



Sentiment Analysis



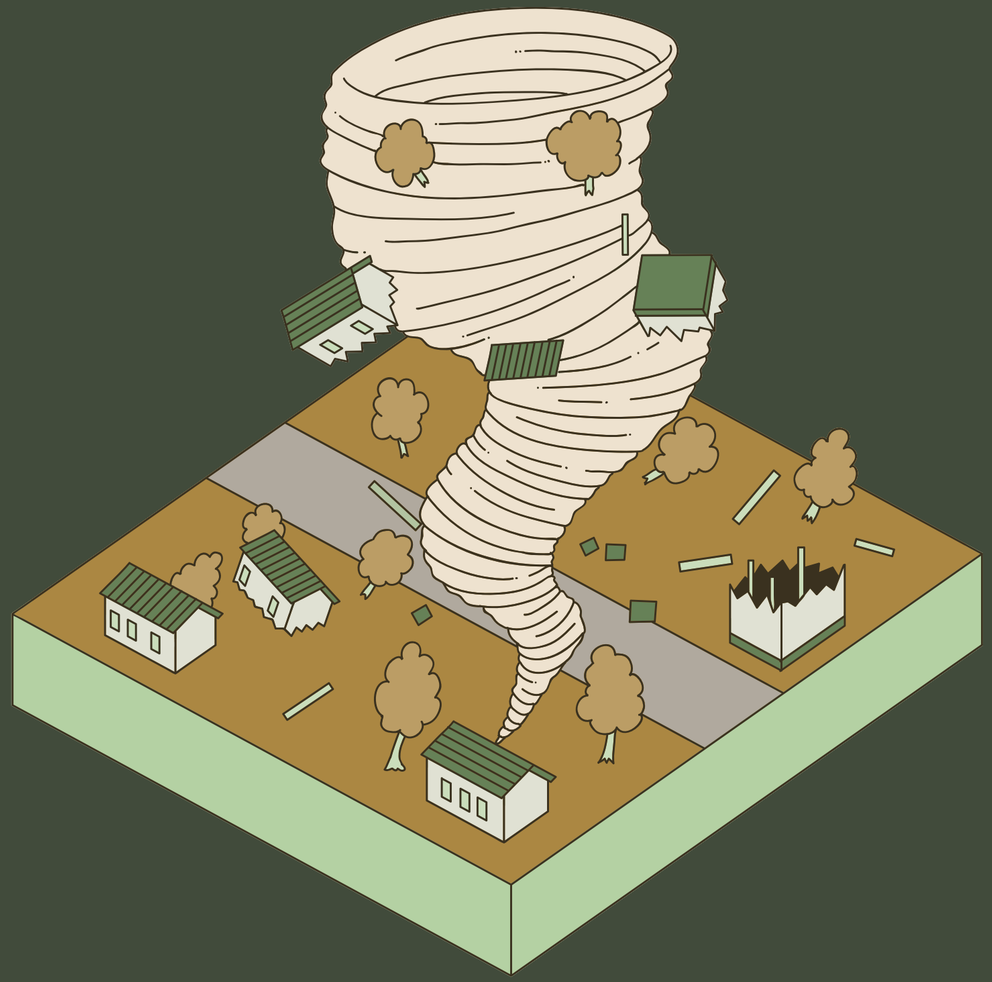
Classification Report

Prediction

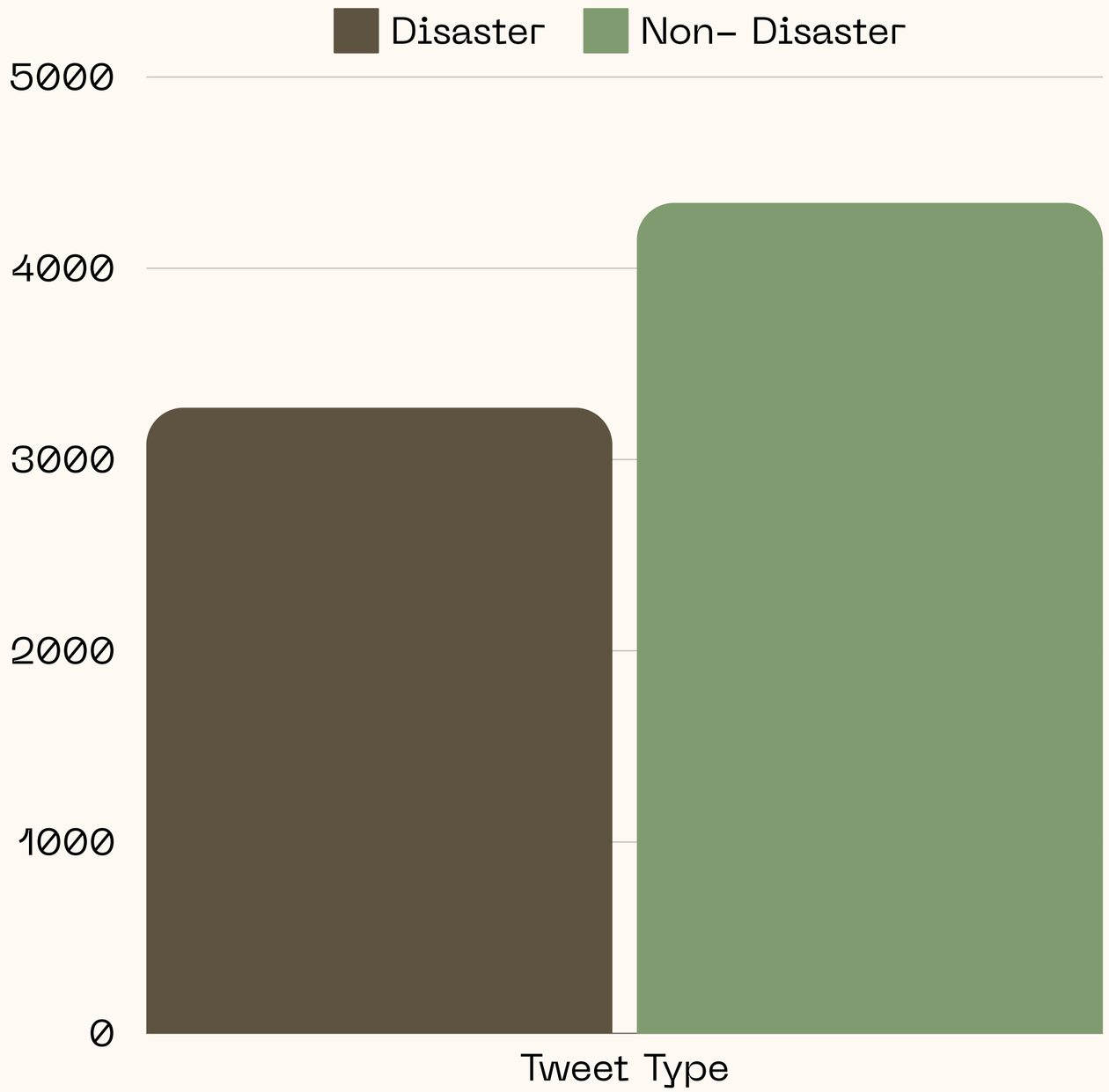
	precision	recall	f1-score	support
0	0.60	0.82	0.69	4342
1	0.53	0.27	0.35	3271
accuracy			0.58	7613
macro avg	0.56	0.54	0.52	7613
weighted avg	0.57	0.58	0.55	7613

Heuristic

	precision	recall	f1-score	support
0	0.63	0.74	0.68	4342
1	0.55	0.42	0.48	3271
accuracy			0.60	7613
macro avg	0.59	0.58	0.58	7613
weighted avg	0.59	0.60	0.59	7613

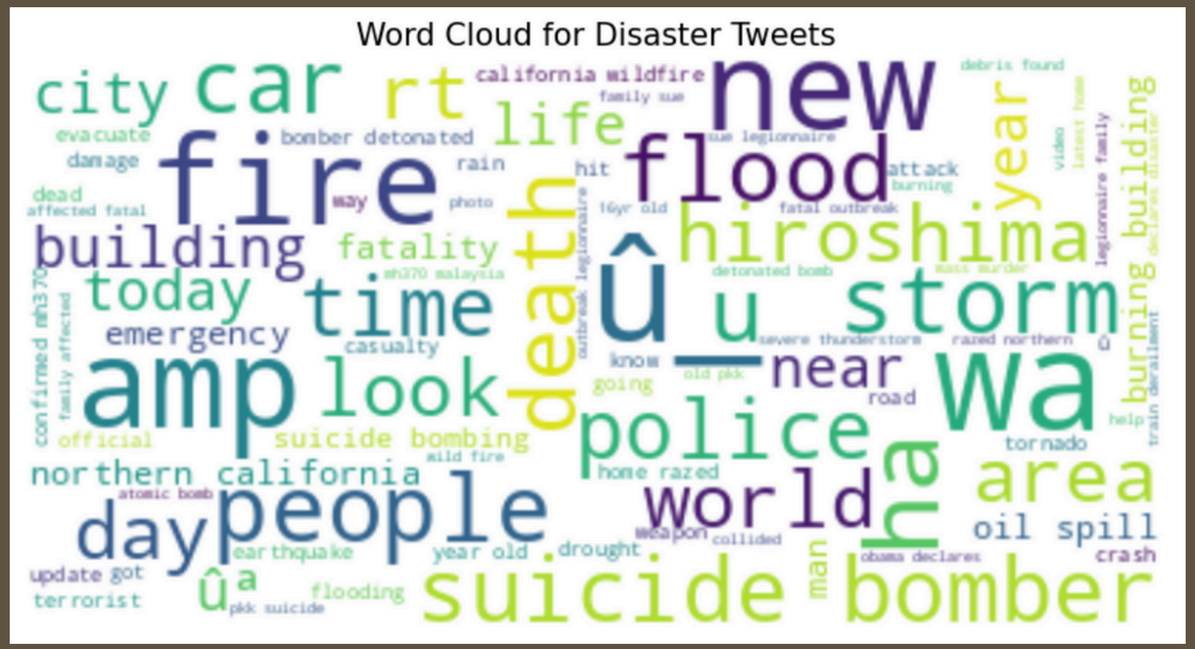


VISUALIZATION

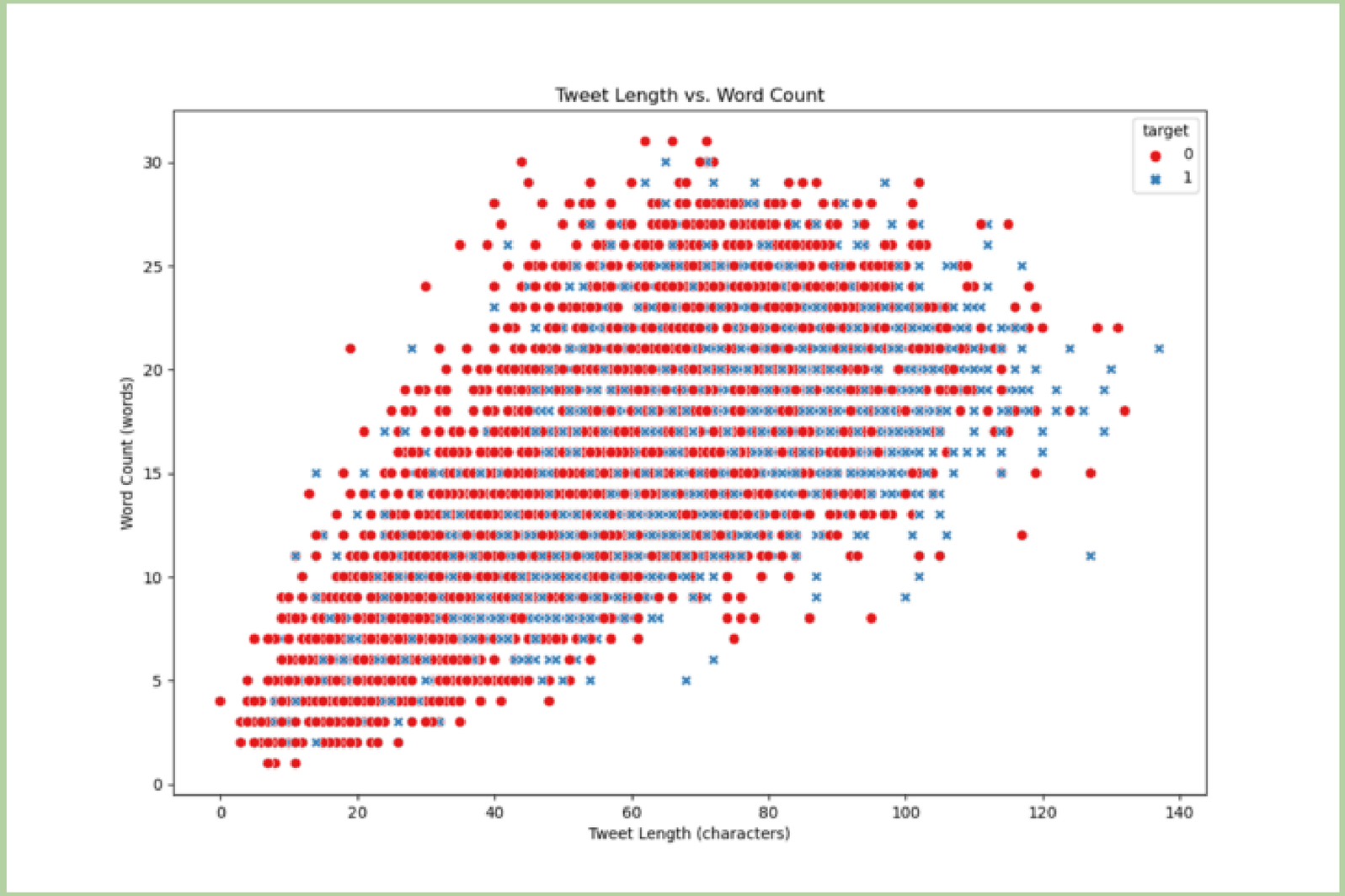


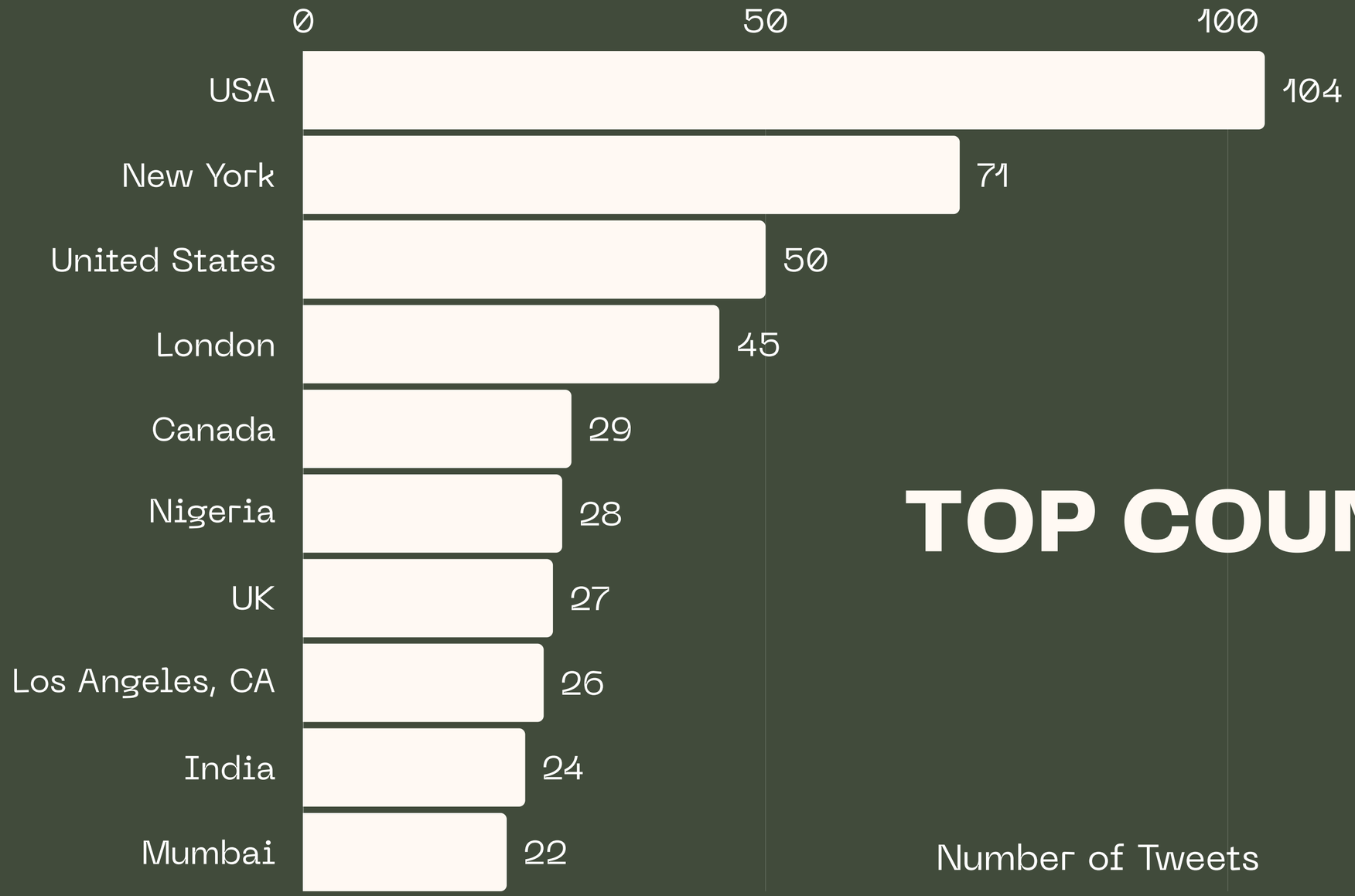
TWEETS DISTRIBUTION

MOST REPEATED WORDS ACCORDING TO DISASTERS OR NON-DISASTERS



CORRELATION BETWEEN TWEET LENGTH VS WORDCOUNT





TOP COUNTRIES

Thank u!