Sweden
Oslo
Stockholm
Helsinki
Denmark
Copenhagen
Netherlands
Amsterdam
Latvia
Berlin
Germany
Moscow
Russia
Minsk
Belarus
Prague
Czech Republic
Slovakia
Vienna
Austria
Bratislava
Budapest
Hungary
Ukraine
Kyiv
Chisinau
Romania
Bucharest
Belgrade
Bosnia and Herzegovina
Sarajevo
Serbia
Bulgaria
Sofia
Skopje
Tirana
Albania
Kazakhsta

Russia & Ukraine Crisis Analysis Through Tweets

EDA Course Project: 4EDACP06

Team: E17

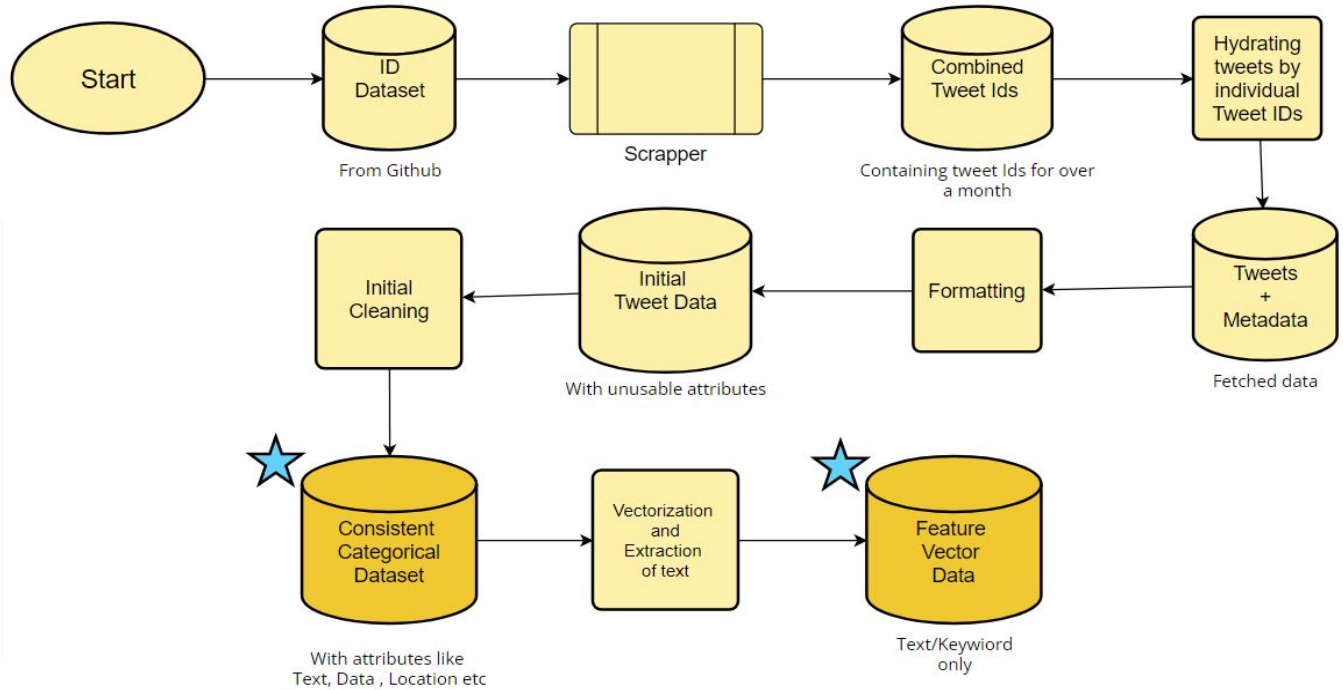
Team Members:
Prakash Kumar Singh (539)
Mohit Gumber (541)
Srishti Saurav (529)

Problem Statement

“Analysis of the current status of Ukraine affected by Russo-Ukraine War (from Twitter data)”

Data Gathering

Data Gathering Flow

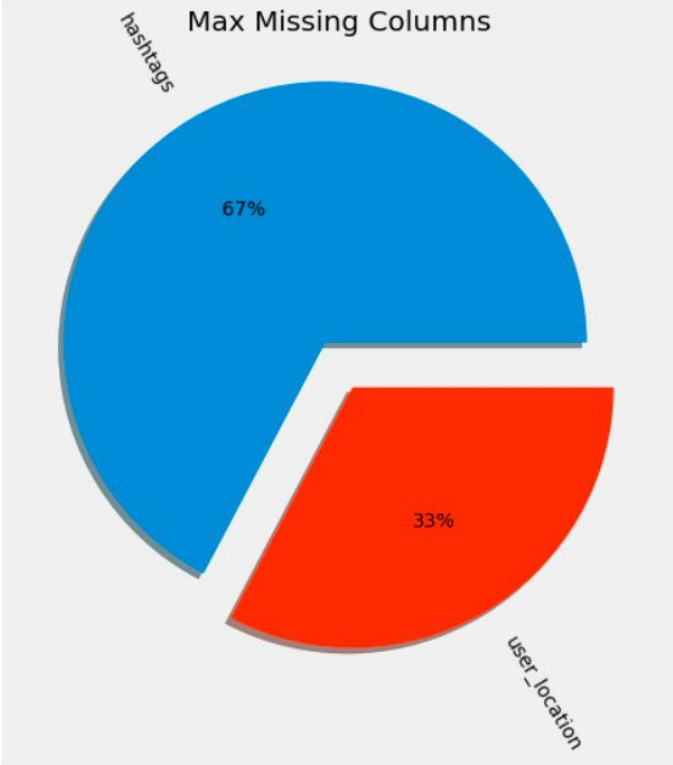


Data Understanding

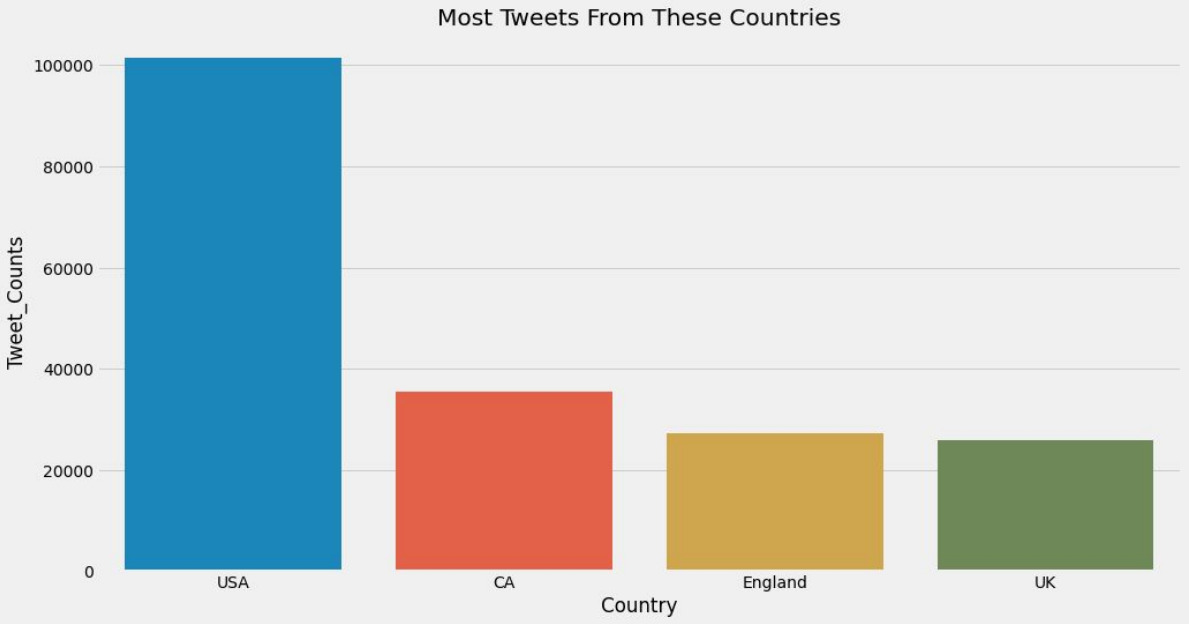
Obtained Dataset

Our dataset obtained via Twitter API consists of data related to tweets stored in a .csv format with the attributes:

```
 #   Column           Non-Null Count    Dtype
---  ------           --------------    -----
 0   created_at       1136882 non-null  datetime64[ns]
 1   hashtags         173892 non-null   object
 2   id               1136882 non-null  int64
 3   lang             1136882 non-null  object
 4   retweet_count    1136882 non-null  int64
 5   source           1136882 non-null  object
 6   text             1136882 non-null  object
 7   user_created_at  1136882 non-null  datetime64[ns]
 8   user_id          1136882 non-null  int64
 9   user_location    668836 non-null   object
 10  user_verified    1136882 non-null  bool
dtypes: bool(1), datetime64[ns](2), int64(3), object(5)
memory usage: 96.5+ MB
```



Majority of the People didn’t use any hashtags Or share their location While tweeting about the crisis still being vocal about their thoughts on the matter making twitter a source of vast range of emotion and information



As evident from the plot, there is a large proportion of the Americans being vocal about the speculation about the crisis, involvement in which is deterministic for diplomatic relations between the US and Russia

Data Preprocessing

President Biden
@POTUS
United States government official
From the beginning of this crisis, I have been clear and consistent: The United States is prepared no matter what happens.
We are ready for diplomacy to improve stability and security in Europe as a whole.
And we are ready to respond decisively if Russia attacks Ukraine.
2:27 AM · Feb 16, 2022 · The White House
5,553 Retweets
787 Quote Tweets
32.3K Likes
text
Created_at

Tweet Text Preprocessing

We perform the necessary text preprocessing steps on the Tweet text:

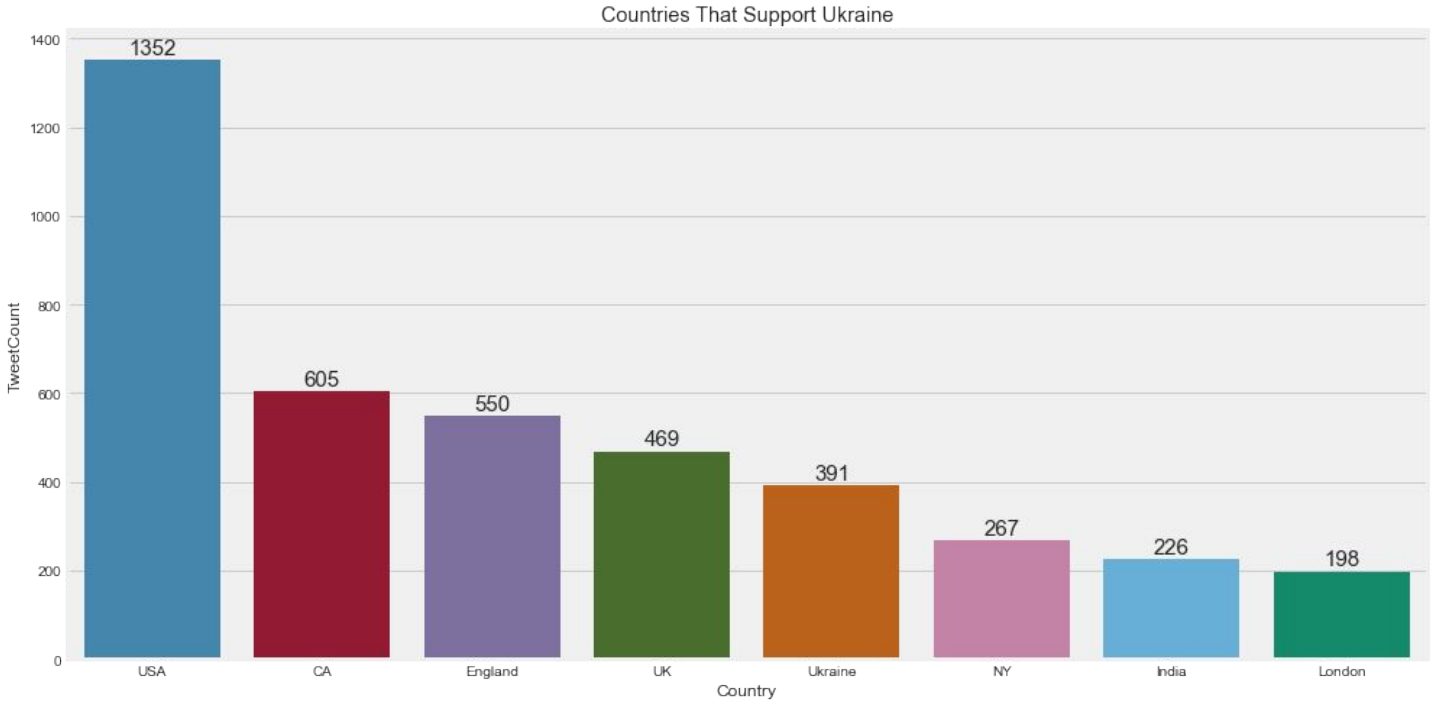
1. Removal of links to websites, such as news sites, youtube, etc.
2. Removal of punctuation and special characters.
3. Removal of stop words
4. Performing of stemming.
5. Performing lemmatization.
6. Creation of n-grams of chosen topics in the form of keywords.
7. Creation of corpus that is to be analyzed.

President Biden

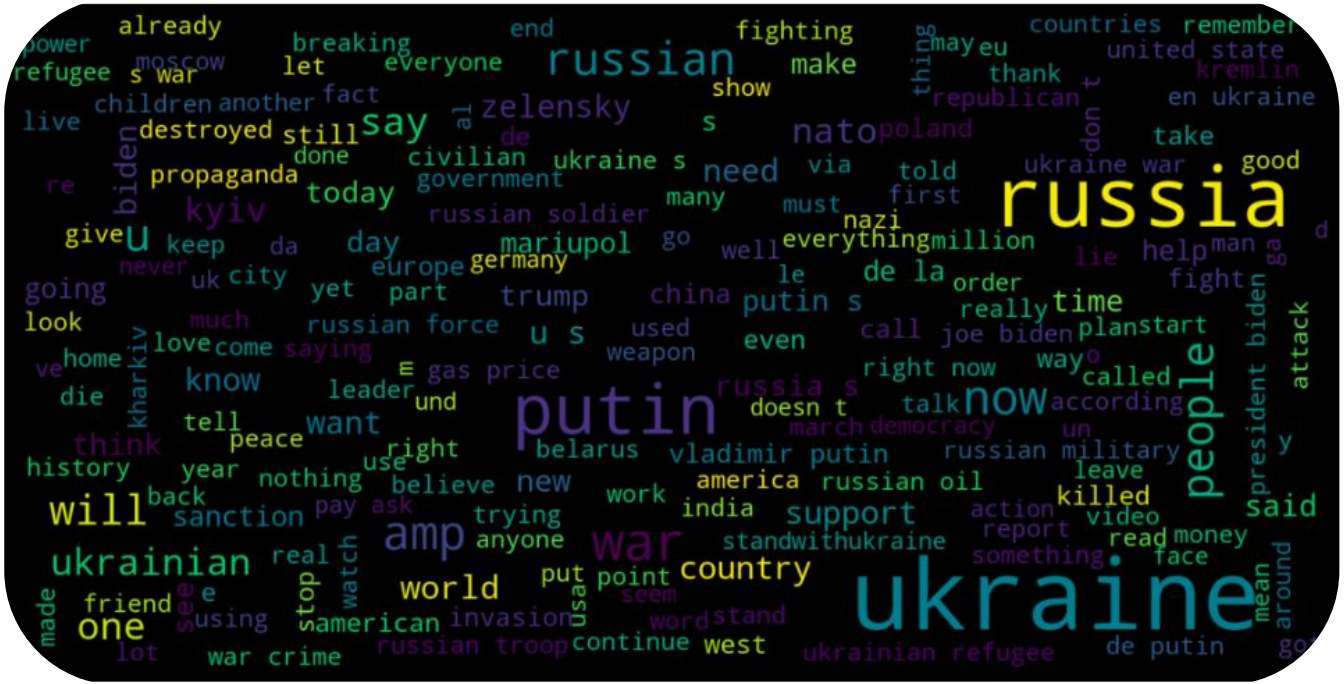
@POTUS

beginning crisis, clear consistent The United States prepared no matter. ready diplomacy improve stability security Europe whole respond decisively when Russia attacks Ukraine

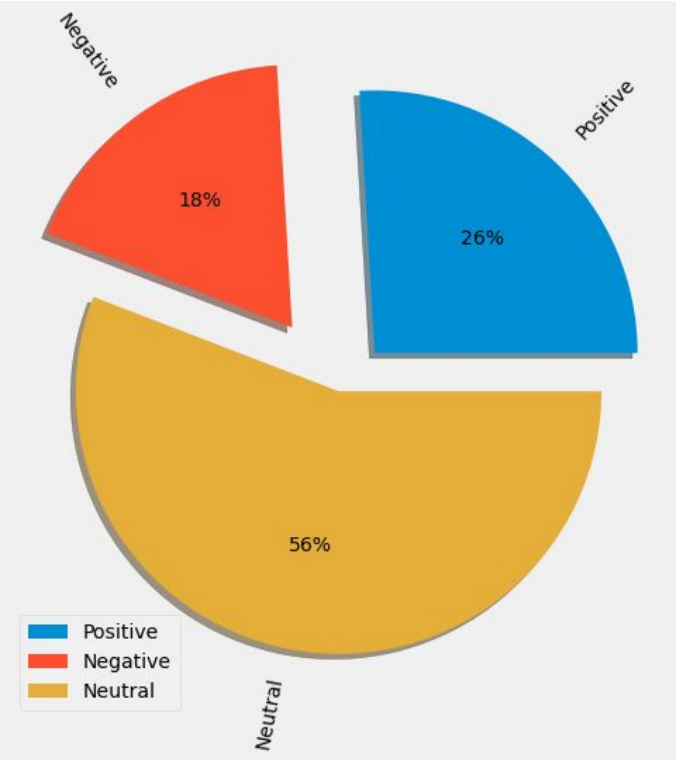
Data Visualization



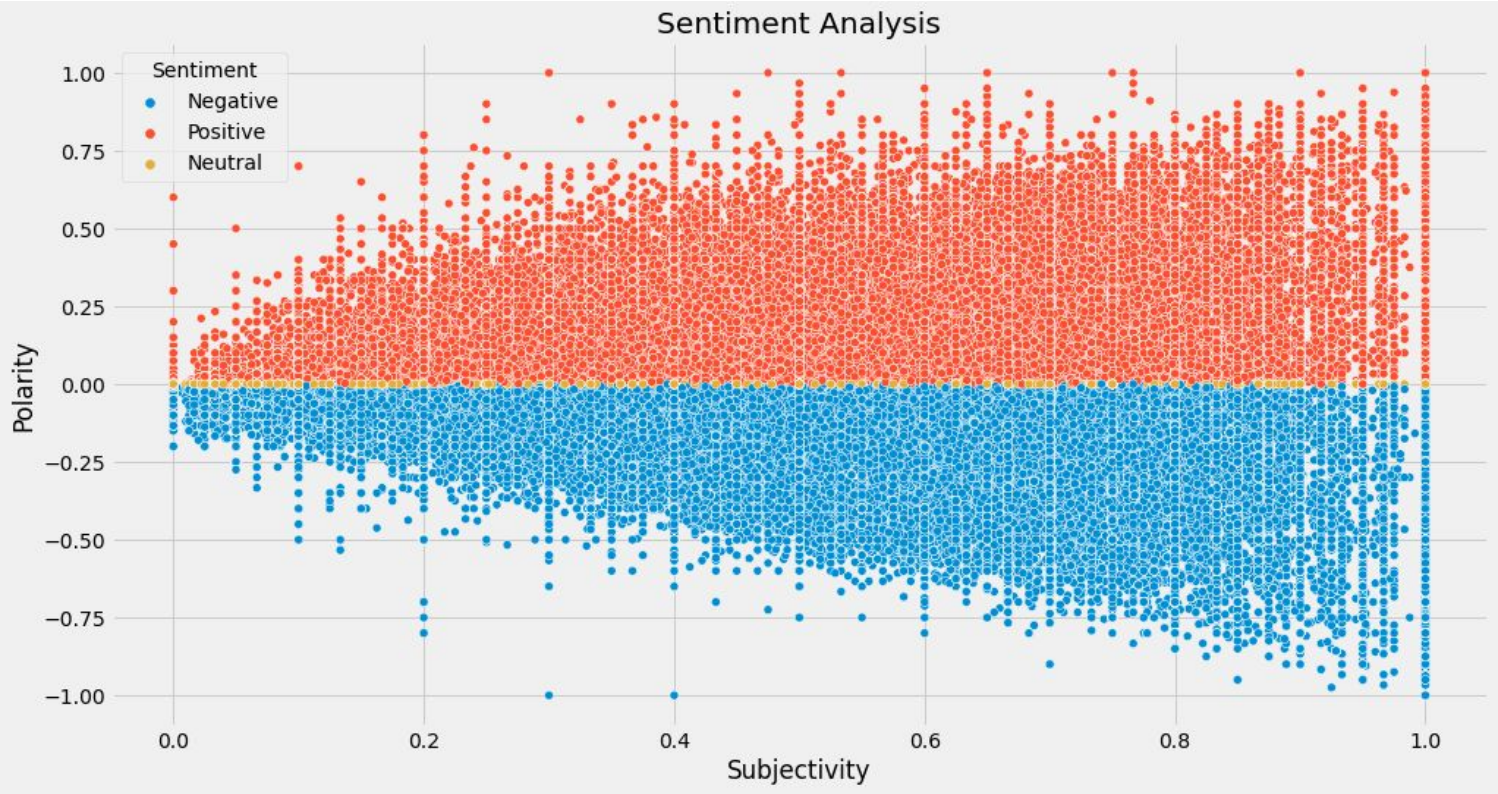
As evident from the plot, there is a large proportion of the Americans being vocal about the speculation about the crisis, involvement in which is deterministic for diplomatic relations between the US and Russia



Word cloud showing the prominent words in the tweets.



Besides the neutral sentiment, that includes discussions of the crisis, its impact on the globe, etc., there is a large positive sentiment of tweets in our data.


Sentiment Analysis
Sentiment
Negative
Positive
Neutral
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Polarity
0.0
0.2
0.4
0.6
0.8
1.0
Subjectivity

Approach To Analysis & Model Building I

Approach To Analysis & Model Building I

Determine how prevalent the topic of Ukraine's socio-economic conditions is in the tweets made in context of the Russo-Ukrainian crisis.

Relevancy & Overlap Of Topics

For making appropriate predictions, we are using the LDA (Latent Dirichlet Allocation) algorithm which is an unsupervised machine-learning model that takes documents as input and finds topics as output. The model also says in what percentage each document talks about each topic.

A topic is represented as a weighted list of words:

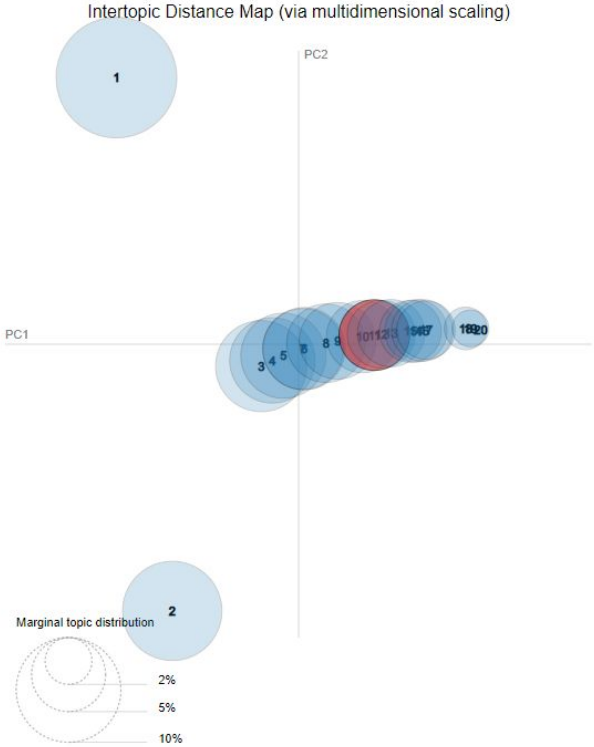
Ukraine * 0.9 | Russia * 0.65 | peace * 0.19 |...

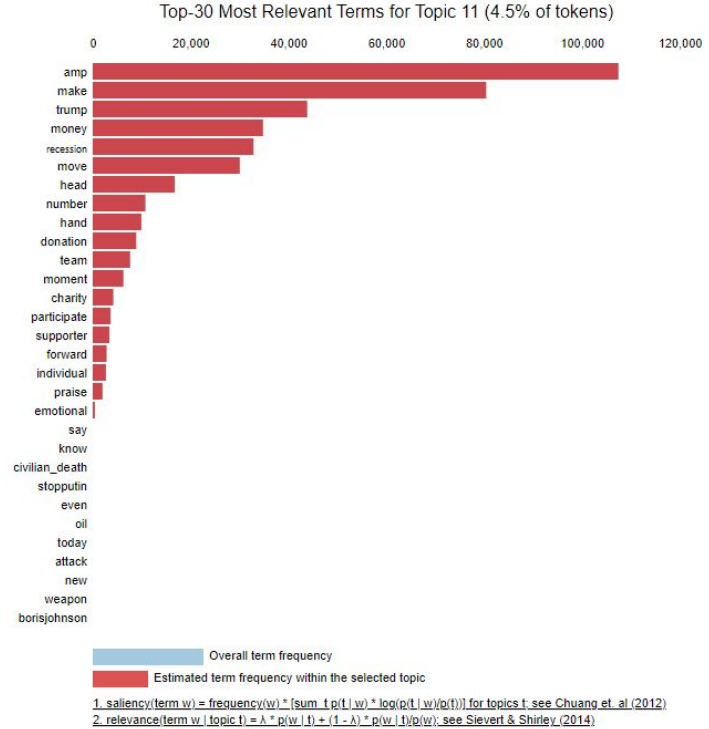
Significance Of LDA Algorithm

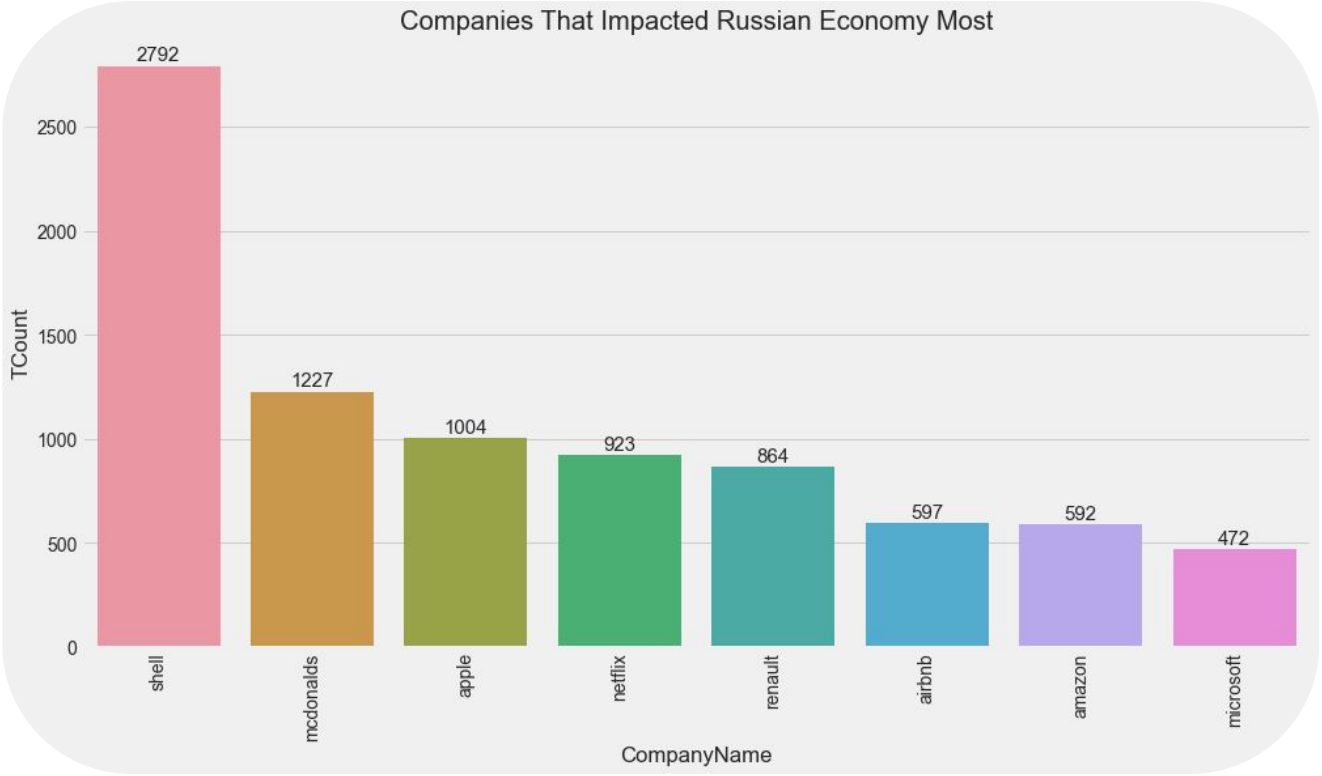
From our considered tweet text, it was difficult to derive a problem/topic to focus on, so we used LDA to see the relevancy of topics, as in understand what topics are more talked about.

By performing LDA, we derived that “economy” was a highly significant topic, so we decide to focus on the same. From our domain understanding we know that we need to try to forecast the situation of the Ukrainian economy. This can be derived from the user sentiment with regards to it, which may or may not align with actual events.

LDA Plot


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
Marginal topic distribution
2%
5%
10%


Top-30 Most Relevant Terms for Topic 11 (4.5% of tokens)
0
20,000
40,000
60,000
80,000
100,000
120,000
amp
make
trump
money
recession
move
head
number
hand
donation
team
moment
charity
participate
supporter
forward
individual
praise
emotional
say
know
civilian_death
stopputin
even
oil
today
attack
new
weapon
borisjohnson
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)



Companies with a significant impact on the Russian economy

Question, Analysis & Inference

Question 1

“What is the general sentiment regarding the situation and further developments that can take place in the Ukrainian economy gravely affected by the war?”

Here, we are looking into the analysis of a multitude of factors that directly or indirectly impact the Ukrainian as well as global economy.

Approach To Analysis & Model Building II

Approach To Analysis II

Predicting the future public sentiment regarding developments in the Ukrainian economy

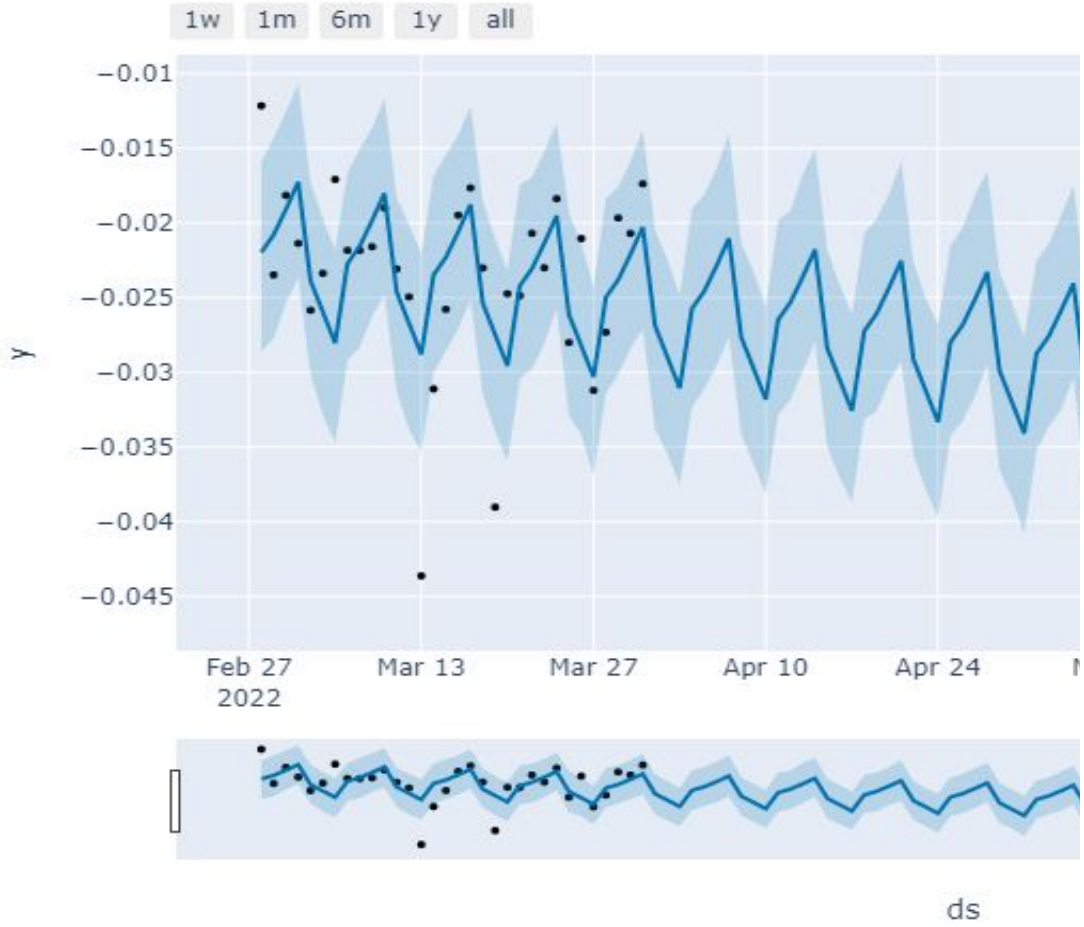
Predicting / Forecasting Future Sentiment

Here, we are making use of the Prophet Model which is an open source library published by Meta that is based on decomposable (trend+seasonality+holidays) models.

It provides us with the ability to make time series predictions with good accuracy using simple intuitive parameters and has support for including impact of custom seasonality and holidays.

Significance Of Prophet Algorithm

We decided to use Prophet algorithm as it converts the data into stationary form and it has clearer visualization compared to other similar models/methods.



Prediction plot showing the growing negative concern regarding economy.

Inference

From the previous plot, we can infer that there will be a growing negative sentiment regarding the Ukrainian economy, which may imply that the Ukrainian economy is on a way to shrinking, making it very difficult for the country to be brought back up to the pre-war Ukrainian standard of living.

Weekly & Daily Trend

Inference

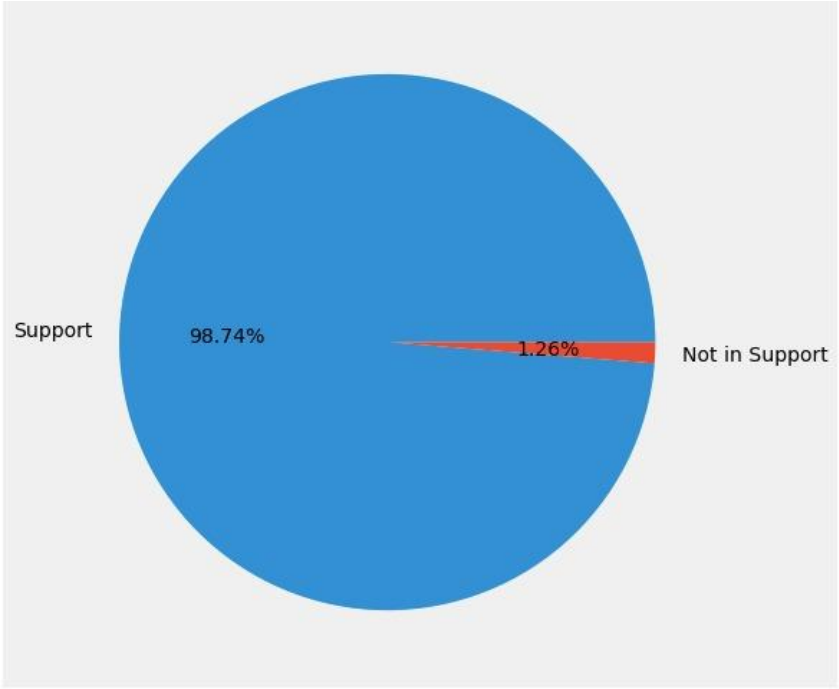
From the previous plot, we can infer that on a weekly & daily basis, the concern for the Ukrainian economy fluctuates, but peaks every week. This is potentially due to news reports that are generated in context of the crisis on a daily & weekly basis.

Question 2

“Predicting what is the general stance regarding Ukraine in the Russo-Ukrainian crisis?”

Here we are trying to segregate the data into two groups mainly, “In Support” and “Not in Support”

Segregated Data



Naive Bayes Algorithms

Gaussian Naive Bayes
Accuracy: 35%
Precision Score: 15%

Multinomial Naive Bayes
Accuracy: 91%
Precision Score: 81%

Bernoulli Naive Bayes
Accuracy: 99%
Precision Score: 95%

Conclusions

From our analysis, we have concluded that there is a growing concern regarding the state of Ukraine, under the Russo-Ukrainian crisis and a general support for Ukraine against Russian aggression that needs to end.

Thank you