SENTIMENT ANALYSIS AND BOT DETECTION: RUSSIA UKRAINE WAR

Twitter Sentiment Analysis Using BERT Model

0742137 Saurav Taneja
0728280 Riya Aggarwal

TRENT UNIVERSITY

Introduction

Sentiment Analysis is a crucial component of NLP, with need of constant research on new models. Twitter serves as a great resource for natural human language in a variety of languages from around the world. This research leverages BERT model for sentiment analysis to understand user sentiment trends on the ongoing Russia-Ukraine conflict using Twitter data. The study also clarifies how public perceptions of significant political leaders like Putin and Zelenskyy have changed.

Dataset

Twitter dataset based on the Russia- Ukraine conflict containing relevant hashtags from April 2022 to June 2023, obtained from a daily updated Kaggle repository.

18GB Twitter Dataset

450k Tweets Analyzed

Method

The study involves meticulous data pre-processing and cleaning on tweets. It is done by removing duplicates, and identifying bots based on the account's age and number of tweets per day.
The multilingual BERT model is then applied on the clean and the bots data to calculate respective average sentiment scores.
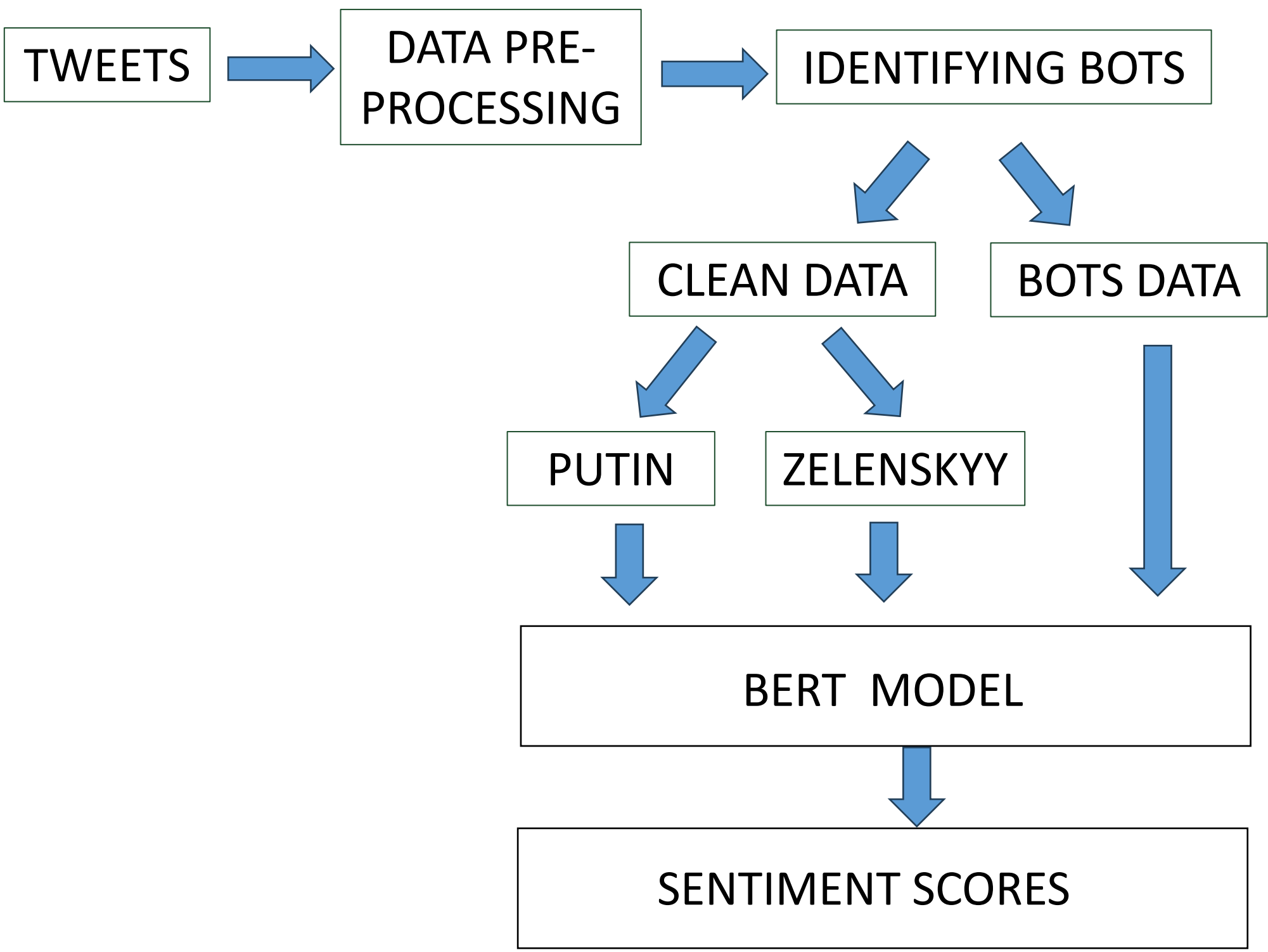
Data Analysis

The data analysis is done using BERT model to obtain sentiment scores of tweets. BERT stands for Bidirectional Encoder Representations from Transformers.

Sentiment score varies from -1 to 1, where negative score signifies negative sentiment, and vice-versa.

Using the BERT model, we analysed-
- Sentiment score on clean data
- Sentiment score on bots
- Sentiment score for Putin and Zelenskyy
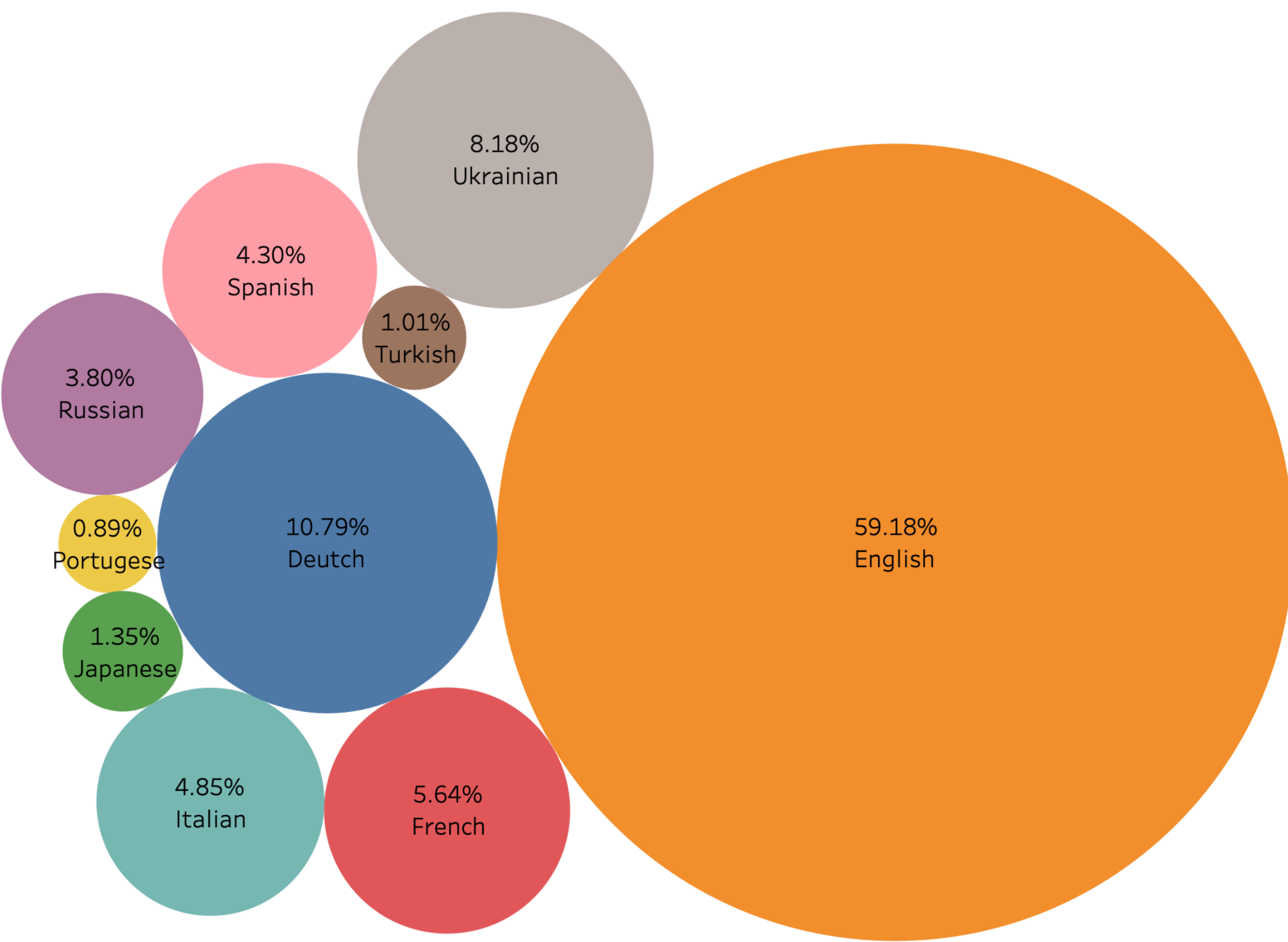- Countries with tweets about the war
- Major languages and keywords

Results

300k Bots Detected   **150k** Clean Tweets

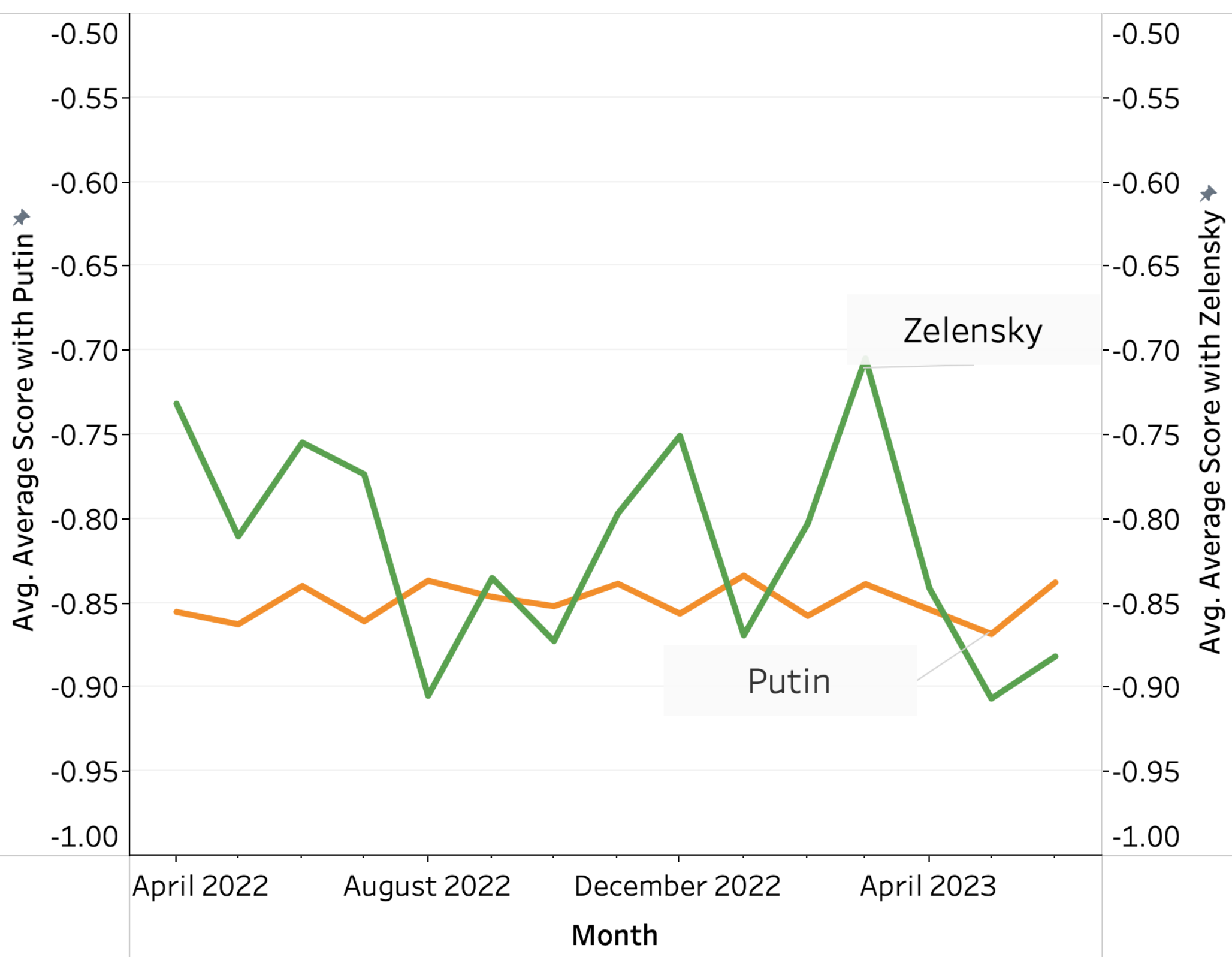|	Positive	Negative
Bots	5.15%	44.84%
Clean	6.15%	43.84%

- Average Sentiment for clean tweets is -0.724
- Average Sentiment for bots is -0.787

- Average Sentiment for Putin is -0.845
- Average Sentiment for Zelenskyy is -0.819

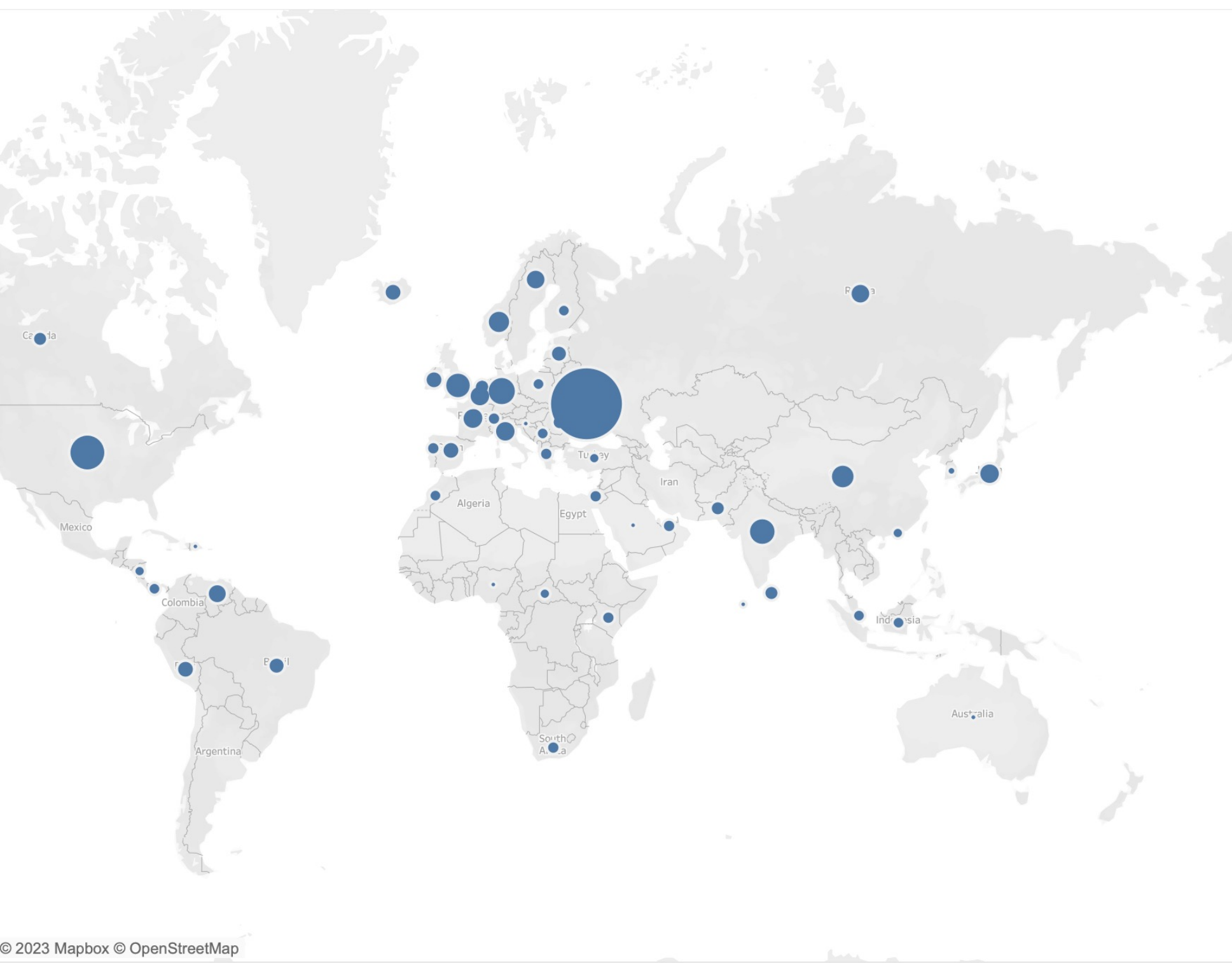### Percent of Tweets by Language

The maximum tweets were in English, followed by Dutch and Ukrainian

Sentiment Score for Putin and Zelensky

The average sentiment for Zelenskyy shows major fluctuations while sentiment for Putin remains almost the same throughout.
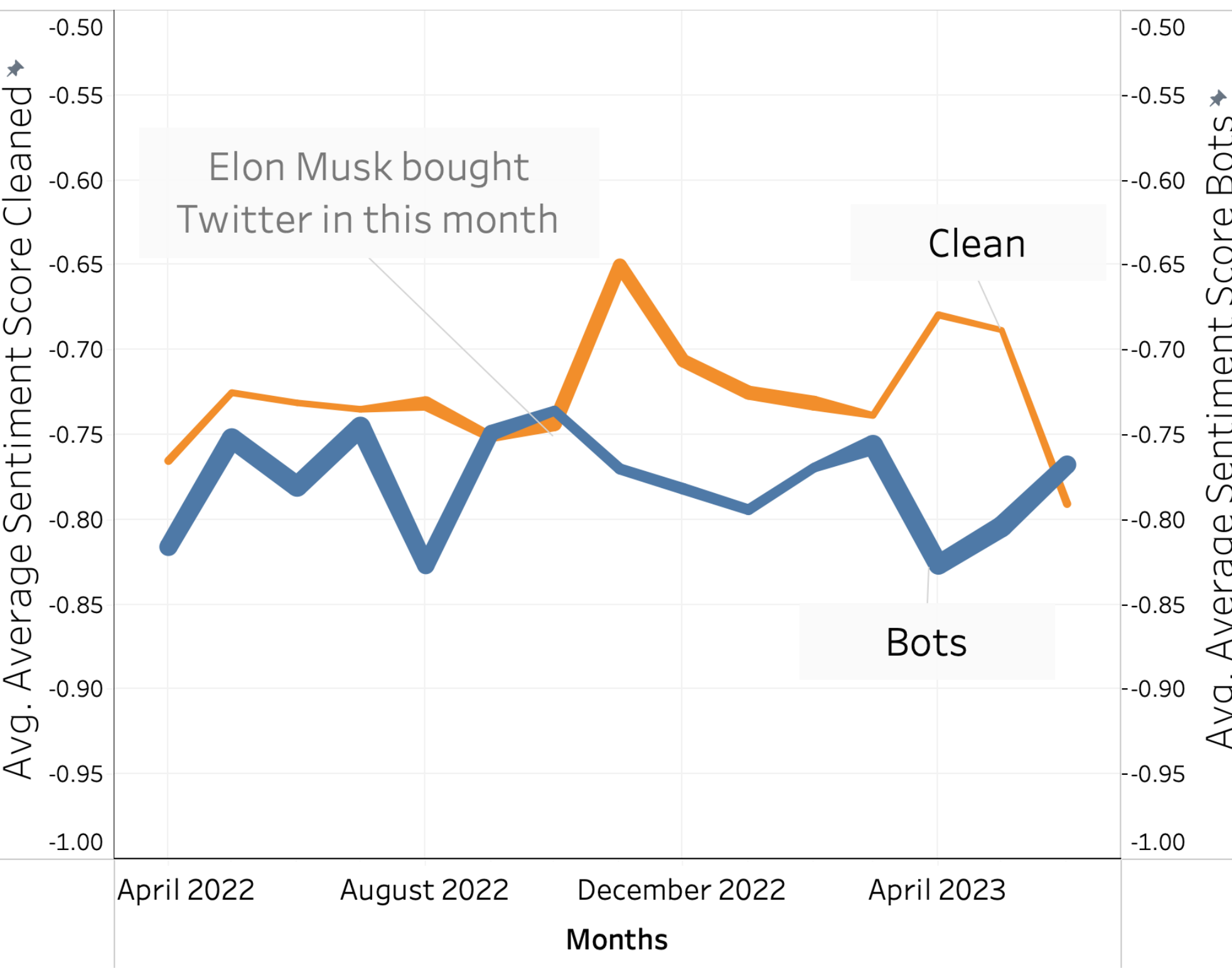
No. of Tweets by Country

It shows the maximum number of tweets originate from Ukraine, followed by U.S.A., western Europe, India and China

Word Cloud for the most used words

The word cloud shows the commonly discussed terms on Twitter throughout the war timeline

Sentiment Score Timeline (Multiple Languages)

Sentiment Score for Bots and clean text changes over time with major peak in sentiment of clean Tweets in October 2022

Key Conclusions

- It is observed that the proportion of positive and negative tweets by bots and actual users were almost the same, but the bots post more extreme tweets making the average sentiment for bots more negative

- There is a big dip in the proportion of bots in October 2022 and a rise in the actual users and their sentiment

- The number of tweets from Russia is considerably less than Ukraine and U.S.A. given the major role of Russia in the conflict.

- The sentiment for Zelenskyy fluctuates more often with time as compared to Putin

References

[1] "(🌅Sunset) 🇺🇦 Ukraine Conflict Twitter Dataset," *www.kaggle.com*. https://www.kaggle.com/datasets/bwandowando/Ukraine-russian-crisis-twitter-dataset- 1-2-m-rows?resource=download&select=20230505_to_20230508_UkraineCombinedTweetsDe duped.csv.gzip (accessed Jul. 20, 2023).

[2] E. Treyger, *Russian Disinformation Efforts On Social Media*. S.L.: Rand Corporation, 2022.

[3] M. Pota, M. Ventura, R. Catelli, and M. Esposito, "An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian," *Sensors*, vol. 21, no. 1, p. 133, Dec. 2020, doi: https://doi.org/10.3390/s21010133.