Ukraine Conflict Data Analysis

Max Bilyk

Table of contents

1 Introduction

The ongoing conflict in Ukraine has been marked by a complex interplay of military strategies, human losses, and societal responses. In an effort to comprehend the evolving dynamics of the situation, this analysis delves into various facets, ranging from equipment losses incurred by the Russian army to the human toll reflected in personnel casualties. Additionally, insights from sentiment analysis of social media and an interactive map displaying incidents resulting in civilian injury or casualties provide a comprehensive perspective on the multifaceted nature of the conflict. By examining trends in equipment losses, personnel casualties, and public sentiment, this study aims to shed light on the strategic implications, societal impact, and the evolving nature of the conflict in Ukraine. Through data-driven insights, we seek to gain a deeper understanding of the complexities inherent in this geopolitical struggle and explore potential avenues for informed decision-making and support for those affected by the conflict.

1.1 Imports

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.ticker import FuncFormatter
import re
import os
import glob
import gzip
from multiprocessing import Pool, cpu_count
from typing import Tuple
from tqdm import tqdm
from textblob import TextBlob
import plotly.express as px
```

2 How has Russian equipment and Personnel losses changed over time?

2.1 Equipment

The data processing for this research question included imputing null values to zero. and dropping a few columns that had null values for every instance.

	date	day	aircraft	helicopter	tank	APC	field artillery	MRL	military auto	fuel tank	d
0	2022-02-25	2	10	7	80	516	49	4	100.0	60.0	0
1	2022-02-26	3	27	26	146	706	49	4	130.0	60.0	2
2	2022-02-27	4	27	26	150	706	50	4	130.0	60.0	2

2.1.0.1 Examples of instances within the dataset.

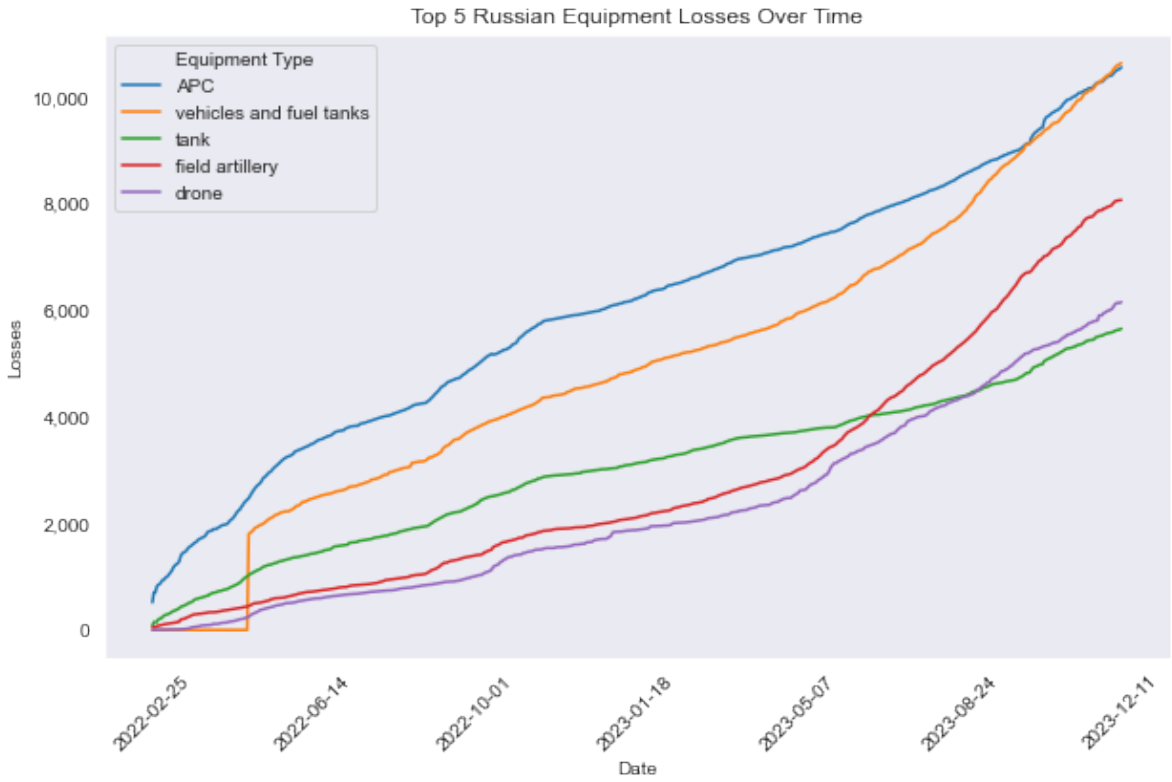

Figure 1: A line plot showing the losses in different Russian equipment over time

2.1.1 Figure Explanation

The above plot shows the losses of the top five most used equipment by the Russian army. Across all these equipments, Russia has been reporting increases in losses consitently since Feburary of 2022. The most used piece of equipment are APC's (Amphibous Personel Carriers) which are basically armored vehicles with a turret mainly used to transport troops in high danger areas.

2.1.2 Trends

Analyzing trends of different equipments, we can see that since the beginning of 2022, APC's and Tanks are following a logorithmic growth, meaning they are growing but the rate of growth is not necessarily increasing. These types of trends usually lead to some kind of plateau. In contrast, Vehicles and fuel tanks, field artillery and drones all are seeing exponential growths. Meaning that the rate of losses for these equipments is increasing.

2.1.3 So what?

After analyzing this figure, we have multiple possible avenues of insight. Firstly, We can make the assumption that since the beginning of the war, Russia has increased their deployment of equipment into this war. After all, no equipment could see an increase in losses without a larger supply of equipment to lose. The second and more strategic insight, is that the Ukrainian Army has been progressivley getting better at destroying russian equipment. Just because the supply of equipment increases, doesn't necessarily mean that their has to be an increase in losses. For example, the Russian army could have potentially overwhelmed the Ukrainian army and forced surrender in battles, but based on the data, we can assume that this was not the case. All in all, this figure reports positive insight for the Ukrainians and negative for the Russians.

2.2 Personnel

The data processing for this section involved dropping columns that were not relevant and filling null values with zeros.

	date	day	personnel	POW
0	2022-02-25	2	2800	0.0
1	2022-02-26	3	4300	0.0
2	2022-02-27	4	4500	0.0

2.2.0.1 Examples of instances within the dataset.

2.2.1 Figure explanation

This figure shows the losses of Russian personell over time. The increase of personell lost appears to be following a slight exponential trend since Feburary 2022. This is very upsetting to see. After this graph printed in my notebook I was frozen for a few minutes. These numbers are so sad to see, all the families and friends that are losing someone in this war. Rest in Peace to all of those involved in this war of nonsense.

Figure 2: A line plot showing Russian Peronnel losses over time

2.2.2 Statistics

```
Russian Personnel loss totals are: 339850
Russian Prisoner of War totals are: 496.0
```

3 Sentiment Analysis on Tweets about the Ukraine War around the World

The pre-processing part of this twitter data was from a public notebook linked here

It consisted of going through all of the files containing the tweets mined from twitter using a twitter API, appending them together in one giant file, filtering them from January 2023 - June 2023 because of computational limitations of my computer and only grabbing tweets that said the word Kyiv in either English, Russian or Ukrainian.

	userid	username	acctdesc	loca
58360	2852858199	interdamazonia	O olhar dos Internacionalistas Amazônidas para...	Belé
58361	1575330860181880832	eldoggo75710233	#NAFO fellephant armed with a fophana fophaser...	#N
58362	805486618077315072	christoalnz	Natural Health Therapist, Global Warming belie...	Auc

3.0.0.1 Examples of instances within the dataset.

3.0.1 Figure Explanation

The figure above is a line plot showing the change in sentiment score of tweets containing the word "Kyiv", the capital of Ukraine. The figure has three seperate lines that represent three different languages from the tweet which are English, Russian and Ukrainian. One thing to note is that a decent number of Ukrainian civilians living on the eastern front actually speak Russian.

3.0.2 Analyzing Spikes and Dips

The first big spike in this figure is a Russian tweet sentiment spike in january 2023.

Here are some events that happened in Kyiv in January 2023: - New Year's Day: Explosions in Kyiv on New Year's Day kill four people and wound many more - January 2, 2023: Russian drones attack infrastructure in Kyiv, damaging energy facilities and causing power outages - January 23, 2023: Russia launches missiles and drones at Kyiv, killing at least 11 people - January 26, 2023: Russia launches 55 missiles at Kyiv, including air and sea-based missiles

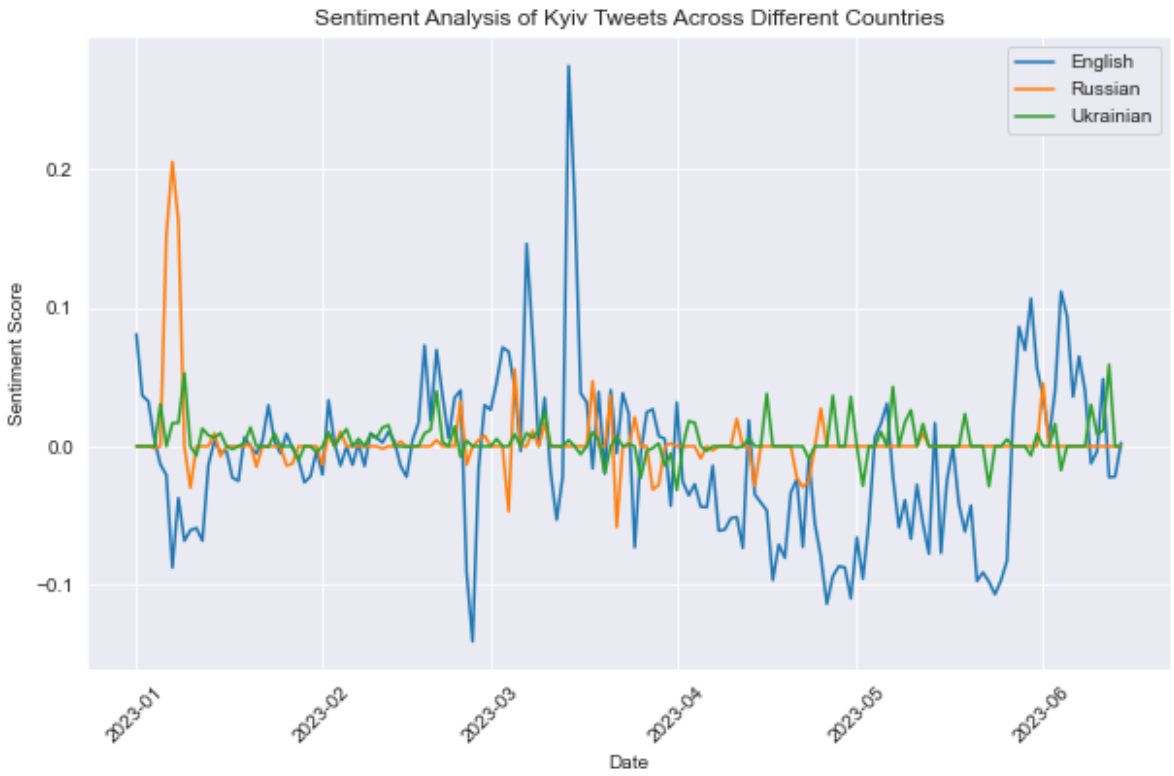

Figure 3: A line plot showing change in twitter sentiment in tweets containing word 'Kyiv' in multiple conuntries.

- January 7, 2023: Russia launches attacks across Ukraine, killing at least two civilians, in violation of a self-declared ceasefire

Russia put a lot of pressure on Ukraine in Janurary and the Russian sentiment over these attacks seemed to be positive.

The first big dip in this figure is in english tweets in late feburary. This time period marked the one year anniversary of when the war started. It is likely that this reminded people of how long this war has been raging on and brought sorrow feelings to english speakers a like that had to watch as so many people die.

The next big spike in the data is in March 2023, once again from the english speakers. Around this time, Russia launched a country wide barage of missle and drone strikes against multiple cities in Ukraine. This was an inhumane attack by the russians that led to so many civilian deaths and destruction of infrastructure and this attack really shifted a tone amongst the world in regards to supporting Ukraine. It is likely that the increase is sentiment score in this time period is from words of encouragement and support towards Kyiv, one of the cities affected by this attack.

After this period their are no major spikes or dips, but their is a steady decrease in sentiment among english until June. This is likely due to the dislike of how America was giving so much money to Ukraine during this time. With all the issues going on the U.S., citizens likely felt like their taxes should be spent more internally rather than internationally.

3.0.3 Statistics

```
Average sentiment score of English tweets: -0.00866654949977672
Average sentiment score of Ukrainian tweets: 0.003815393431503507
Average sentiment score of Russian tweets: 0.0030602195795483927
```

4 Civilians

The data processing in this section included concatenating two dataframes together, creating two new columns based on a string value in another column and filtering by only instances classified as weapon systems that are known.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2552 entries, 0 to 2551
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   id            2552 non-null   object
```

```
 1   date          2552 non-null   object
 2   latitude      2552 non-null   float64
 3   longitude     2552 non-null   float64
 4   location      1366 non-null   object
 5   description   2552 non-null   object
 6   sources       2500 non-null   object
 7   associations  2539 non-null   object
dtypes: float64(2), object(6)
memory usage: 159.6+ KB
```

	id	date	latitude	longitude	location	description
0	CIV0001	02/24/2022	49.850050	36.659031	Chuhuiv, south of Kharkiv	Apartment block hit. Crate
1	CIV0002	02/24/2022	48.748564	30.218515	Uman	Civilians hit by what appea
2	CIV0003	02/24/2022	50.470055	30.527381	Kyiv	Explosion in central Kyiv, n

4.0.0.1 Examples of instances within the dataset.

```
Weapon System            2433
Type of area affected     106
Name: weapons2, dtype: int64
```

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): application/vnd.plotly.v1+json, text/html
```

4.0.1 Figure Explanation

The figure above is an interactive map that plots incidents that resulted in civilian injury or casualty from the Ukraine conflict, and each point is colored by the weapons that were responsible for the incident.

4.0.2 Why?

The purpose of this figure is to provide insight on where civilians are being affected, and what is the means of descruction related to the incident. The insights provided can help to defend civilians against certain types of weapons by allocated appropriate equipment and resources. For example, Air strikes are much more common in the Eastern areas of Ukraine, so anti-air defense can be moved from the Western regions and into these Eastern regions. With 14 different type of weapons found in this dataset, the same type of insight can be repeated for a plethora of data-driven decicisions to help the civialians affected in this conflict.

4.0.3 Further Insight

Cluster Munitions - Which are usually a ground-launched explosive, has hotspots towards Eastern Ukraine. Defensive equipment could be distirbuted to these regions more heavily. It is very sad to see this type of weapon commonly used in this conflict, as they are notoriously unreliable and unpredictable, meaning that they are very dangerous for noncombatants anywhere near the area of their explosion.

Cruise Missles - These missles have an even distribution all over Ukraine, so their is no one area where defense equipment should be distrubuted more heavily.

Ballistic Missles - These missles are included in civilian casualty in the middle and eastern regions of Ukraine. So defensive equipment should be distributed to these areas.

4.0.4 Statistics

```
Cluster munitions         201
Cruise missile            109
HE rocket artillery        65
Small arms                 52
Loitering munition         36
Ballistic missile          33
Anti-air missile           31
Air strike                 30
Incendiary munitions       28
Vehicle mounted weapon     23
Name: weapons1, dtype: int64
```

4.0.4.1 Different types of Weapons and their frequencies from the dataset.

5 Conclusion

In conclusion, the comprehensive analysis of various aspects related to the conflict in Ukraine offers valuable insights into the evolving dynamics of the situation. The examination of losses incurred by the Russian army in terms of equipment reveals notable trends. The data indicates consistent increases in losses across the top five most used pieces of equipment since February 2022. Specifically, Amphibious Personnel Carriers (APCs), characterized by their armored vehicles with turrets, emerge as the most frequently used equipment.

A detailed analysis of equipment trends further distinguishes between logarithmic growth observed in APCs and tanks, suggesting a potential plateau, and exponential growth in vehicles,

fuel tanks, field artillery, and drones, implying an accelerating rate of losses. The interpretation of these trends leads to two key insights. Firstly, the increased losses may be attributed to a larger deployment of equipment by Russia since the beginning of the conflict. Secondly, the Ukrainian Army appears to be progressively improving its ability to neutralize Russian equipment, suggesting a strategic advantage for Ukraine.

Moving beyond equipment losses, the examination of personnel losses by the Russian army reveals a distressing linear trend since February 2022. The magnitude of human casualties evokes a profound emotional response, highlighting the human toll and tragedy associated with armed conflicts. The sentiment analysis of tweets containing the word "Kyiv" provides a unique perspective on public sentiment across different languages. Notable spikes and dips in sentiment coincide with significant events, such as attacks on Kyiv, the anniversary of the war, and global responses to Russian actions.

The interactive map plotting incidents resulting in civilian injury or casualties, colored by responsible weapons, serves as a crucial tool for understanding the geographical distribution of the conflict's impact. Insights gained from the map, such as the prevalence of cluster munitions in Eastern Ukraine and the widespread use of cruise missiles, can inform strategic decisions to better defend civilians. The analysis underscores the importance of data-driven decision-making to allocate appropriate resources and equipment for the protection of civilians in conflict zones.

In summary, the multi-faceted analysis of equipment losses, personnel casualties, sentiment trends, and incident mapping offers a nuanced understanding of the ongoing conflict in Ukraine. These insights contribute to a broader comprehension of the strategic, human, and societal dimensions of the conflict, highlighting the challenges and opportunities for all stakeholders involved.

6 References

Equipment and Personnel Datasets

Tweet Dataset

Civilian Dataset

https://stackoverflow.com/questions/63644131/how-to-use-geopandas-to-plot-latitude-and-longitude-on-a-more-detailed-map-with

https://www.cnbc.com/2023/02/24/russia-ukraine-live-updates.html

https://seaborn.pydata.org/generated/seaborn.lineplot.html

https://www.w3schools.com/python/python_lambda.asp

https://plotly.github.io/plotly.py-docs/generated/plotly.express.choropleth.html

ChatGPT

Google's Search Generative Experience