Analysis of the Ukraine vs Russia War Aid and Death Toll Prediction - White paper

I- Business Problem

The war in Ukraine has led to a significant humanitarian crisis, with many people in need of help. Governments and organizations are providing various types of aid, but it's tough to keep track of how much is being given and what impact it's having. The goal of this paper is to understand the types of aid being sent and predict the death toll from the conflict. By doing this, we can better support those affected by the war.

II- Background and History

Since the conflict began, the situation in Ukraine has drawn attention from around the world. Many countries have stepped in to provide aid, but the sheer amount of help can be overwhelming. Understanding the flow of resources and the human cost of the war is crucial for planning effective responses and assistance.

III- Data Explanation

To tackle this issue, we collected data from Kaggle and the Institute For The World Economy. Each source provided valuable information that helped us understand the humanitarian aid situation in Ukraine and the impact of the conflict.

Personnel Losses

This dataset includes details about military personnel losses:

- **ID**: Unique identifier for each entry

- **Date**: When the data was recorded
- **Daily Personnel Loss:** Number of personnel reported lost on that day
- **Total Personnel Loss:** Cumulative number of personnel lost
- **POW (Prisoners of War):** Number of captured personnel

In [6]:
```
# Personnel losses data
personnel_loss_data = pd.read_csv(personnel_data_link, sep=',')

# Previwing the dataset
personnel_loss_data.sample(n=5)
```

Out[6]:

	date	day	personnel	personnel*	POW
499	2023-05-12	443	197670	about	NaN
270	2023-12-27	672	355750	about	NaN
590	2023-02-10	352	135740	about	NaN
705	2022-10-18	237	65850	about	NaN
604	2023-01-27	338	124710	about	NaN

Importing the personnel losses data into a pandas data frame

Equipment losses

This dataset focuses on military equipment losses. It contains the following columns:

- **Date**: When the data was recorded
- **Aircraft**: Number of aircraft lost
- **Tanks**: Number of tanks lost
- **Drones, Naval Ships**, etc.: Counts for various types of military equipment

```
# Equipment losses data
equipment_loss_data = pd.read_csv(equipment_data_link, sep=',')

# Previewing the dataset
equipment_loss_data.sample(n=5)
```

	date	day	aircraft	helicopter	tank	APC	field artillery	MRL	military auto	fuel tank	drone	naval ship	anti-aircraft warfare	special equipment	mobile SRBM system	greatest losses direction	vehicles and fuel tanks	cruise missiles
301	2023-11-26	641	323	324	5513	10279	7874	907	NaN	NaN	5901	22	597	1113.0	NaN	NaN	10288.0	1565.0
931	2022-03-06	11	44	48	285	985	109	50	447.0	60.0	4	2	21	NaN	NaN	NaN	NaN	NaN
378	2023-09-10	564	315	316	4554	8755	5811	760	NaN	NaN	4593	19	509	872.0	NaN	NaN	8338.0	1455.0
400	2023-08-19	542	315	316	4340	8424	5212	714	NaN	NaN	4282	18	486	785.0	NaN	NaN	7665.0	1406.0
578	2023-02-22	364	299	287	3334	6569	2345	471	NaN	NaN	2026	18	243	226.0	NaN	NaN	5212.0	873.0

Importing equipment losses data into a pandas data frame

Aid Data

This dataset contains the types and amounts of aid provided:

- **ID**: Unique identifier for each aid entry
- **Countries**: Country providing the aid
- **Type of Aid**: Description of the aid (e.g., food, medical supplies)
- **Monetary Value**: Total value of the aid in monetary terms
- **Items Delivered**: Specific items included in the aid
- **Value Delivered**: Estimated worth of the items delivered
- **Date**: When the aid was sent

```
# Aid data
aid_data = pd.read_excel(aid_data_link, sheet_name=2)

# Checking the dataframe
aid_data.sample(n=5)
```

	ID	Countries	Announcement Date	Type of Aid General	Type of Aid Specific	Explanation	Original Currency	Type of donation	Monetary Value as Given by Source	Items	...	Unnamed: 82	Unnamed: 83	Unn
696	EEM16	Estonia	2023-12-14 00:00:00	Military	Weapons and equipment	According to the Minister of Defence Hanno Pev...	EUR	Allocation	80000000	rounds of small arms ammunition	...	NaN	NaN	
633	EEM1	Estonia	2022-02-18 00:00:00	Military	Weapons	Estonia donated Javelin anti-tank missiles to ...	USD	Allocation	Not given	Javelin anti-tank missile	...	NaN	NaN	
3669	EUF4	EU (Commission and Council)	2022-11-22 00:00:00	Financial	Loan	MFA: 5 EUR billion allocated funds	EUR	Allocation	2500000000	.	...	NaN	NaN	
2053	NOM3	Norway	2022-04-20 00:00:00	Military	Weapons	Norway has announced 100 units of the Mistral ...	USD	Allocation	Not given	Mistral Air Defence Missile	...	NaN	NaN	
1527	ISH5	Iceland	2022-05-05 00:00:00	Humanitarian	Assistance	According to Source of Aid 1, the Prime Minist...	ISK	Allocation	295000000	.	...	NaN	NaN	

Importing the aid data into a pandas data frame

1- Data Cleaning Process

Cleaning the data was a crucial step in our analysis. Here's how we handled it:

1. **Removing Duplicates**: We checked for duplicate entries in both datasets and removed them to ensure accuracy.

2. **Handling Missing Values**: We found some entries with missing information. For these:

- If the missing data was minimal, we filled the numerical values with 0
- In cases where too much data was missing, we decided to remove those entries altogether to maintain the dataset's integrity

```
# We will remove the column personnel* from the personnel_loss_data
personnel_loss_data.drop('personnel*', axis=1, inplace=True)
```

Dealing with missing values The equipment_loss_data has missing values. Because those values are numerical, we will replace them with 0

```
# Replacing missing values with 0
equipment_loss_data = equipment_loss_data.fillna(0)
personnel_loss_data = personnel_loss_data.fillna(0)
```

```
# Removing the columns that are missing Annoucement Date
aid_data = aid_data.dropna(subset='Announcement Date')

```


```
# Replacing missing values with 0
equipment_loss_data = equipment_loss_data.fillna(0)
personnel_loss_data = personnel_loss_data.fillna(0)
```


Handling missing data

3. **Data Type Conversion**: We made sure that all columns had the correct data types. For example, dates were converted to a date format, and monetary values were set as numerical data.

```
# Let's convert 'Announcement Date' to string and create a new column 'date' using a lambda function
aid_data = aid_data.assign(date=lambda df_: df_['Announcement Date'].astype(str))

# Let's filter and check if the column contains valid dates in the format YYYY-MM-DD
aid_data = aid_data.loc[lambda df_: df_['date'].str.contains(r'^\d{4}-\d{2}-\d{2}', regex=True)]

# Now let's convert the column to datetime format
aid_data = aid_data.assign(date=lambda df_: pd.to_datetime(df_['date']))

# Let's drop the origiinal Announcement Date columns
aid_data = aid_data.drop('Announcement Date', axis='columns')

# The new dataset
aid_data.head(5)
```

```
# Converting the date column type to datetime
personnel_loss_data['date'] = pd.to_datetime(personnel_loss_data['date'])
equipment_loss_data['date'] = pd.to_datetime(equipment_loss_data['date'])
```

```python
# Let's convert the monetary values to Numeric
aid_data['Monetary Value as Given by Source'] = pd.to_numeric(aid_data['Monetary Value as Given by Source'],
                                                              errors='coerce')

# Dropping the missing values
aid_data = aid_data.dropna(subset=['Monetary Value as Given by Source'])

```

Data type conversion

2- Transformations Applied

To make the data more useful for analysis, we applied some transformations:

- **Aggregating Data**: We grouped data by date to calculate daily totals for personnel and equipment losses, as well as aid delivered. This helped us see trends over time.
- **Creating New Columns**: We added new columns for analysis, like calculating the ratio of aid delivered to personnel losses, which can provide insights into the effectiveness of aid.

IV- Methods

1. Modeling

We used the ARIMA model to predict the number of losses based on historical data.

Predicting Future Death Tolls Using ARIMA

```
# Checking for stationarity
# Performing Dickey-Fuller test to check for stationarity
result = adfuller(losses_data['daily_personnel_loss'].dropna())
print('ADF Statistic:', result[0])
print('p-value:', result[1])
```

```
ADF Statistic: -1.1034198219169915
p-value: 0.7137061945528895
```

The p-value is 0.71 this means that the time series data for daily personnel losses is not stationary.

Let's transform the time series data to make it stationary before using ARIMA.

```
# First differencing - subtracting the previous value from the current one
losses_data['diff_personnel_loss'] = losses_data['daily_personnel_loss'].diff()

# Drop any resulting NaN values from the differencing
losses_data.dropna(subset=['diff_personnel_loss'], inplace=True)

# View the first few rows of the new differenced column
losses_data[['daily_personnel_loss', 'diff_personnel_loss']].head()

```

```
# Perform Dickey-Fuller test on the differenced data to check for stationarity
result_diff = adfuller(losses_data['diff_personnel_loss'])

print('ADF Statistic after differencing:', result_diff[0])
print('p-value after differencing:', result_diff[1])
```

```
ADF Statistic after differencing: -12.797325151155027
p-value after differencing: 6.871473982995574e-24
```

Using Arima to predict future death tolls

2. Evaluation

We evaluated our model using Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

```
# Calculating the evaluation metrics
mae = mean_absolute_error(predicted_data['Actual'], predicted_data['Predicted'])
mse = mean_squared_error(predicted_data['Actual'], predicted_data['Predicted'])
rmse = mse ** 0.5  # Root Mean Squared Error

# Printing the metrics
print(f'Mean Absolute Error (MAE): {mae}')
print(f'Mean Squared Error (MSE): {mse}')
print(f'Root Mean Squared Error (RMSE): {rmse}')

```

Evaluating the model

V- Analysis

1- Type of aid

```
# Print the data
aid_counts
```

```
Out[10]: Type of Aid Specific
Weapons And Equipment                                              1686
Assistance                                                          390
Weapons                                                             312
Grant                                                               250
Equipment                                                           206
Budgetary Appropriations                                            120
Military Equipment                                                  103
Equipment And Assistance                                             85
Funding For Weapon Acquisition Program                               82
Loan                                                                 65
Guarantee                                                            40
Reconstruction                                                       29
Material Assistance                                                  17
Weapons And Military Equipment                                       16
Assistance And Equipment                                             15
Weapons, Equipment And Assistance                                     9
European Peace Facility                                               9
Weapons And Assistance                                                8
Training                                                              8
Funding For Weapon Acquisition Program/Weapons And Equipment          2
Equipment And Medicines                                               2
Training And Equipment                                                2
Funding For Weapons Acquisition                                       1
Swap-Line                                                             1
Protective Equipment And Medical Supplies                             1
Medicines                                                             1
Recovery                                                              1
Consolidation Entry                                                   1
Name: count, dtype: int64
```

Type of aid

The data shows that most of the aid provided in the conflict is focused on military support. The largest category is “Weapons and Equipment” with 1,686 instances, followed by Assistance (390) and Weapons (312). This shows that providing weapons and equipment is a priority. There are also financial contributions like grants, loans, and funding for weapon programs, but these are smaller compared to the direct military aid.

Overall, the data suggests that while military support is the main focus, there is a mix of aid types, ranging from financial to medical and recovery support.

2- Air equipment loss

The air equipment loss data shows that drones make up the biggest share, with 89.1% of all losses, likely because they are used more often or are easier to lose compared to other air equipment. In contrast, aircraft and helicopters each account for only about 5-6% of the losses, which suggests they are used more carefully or are harder to take down. This highlights the key role drones play in the conflict, but also how vulnerable they are.

3- Aid monetary values by country

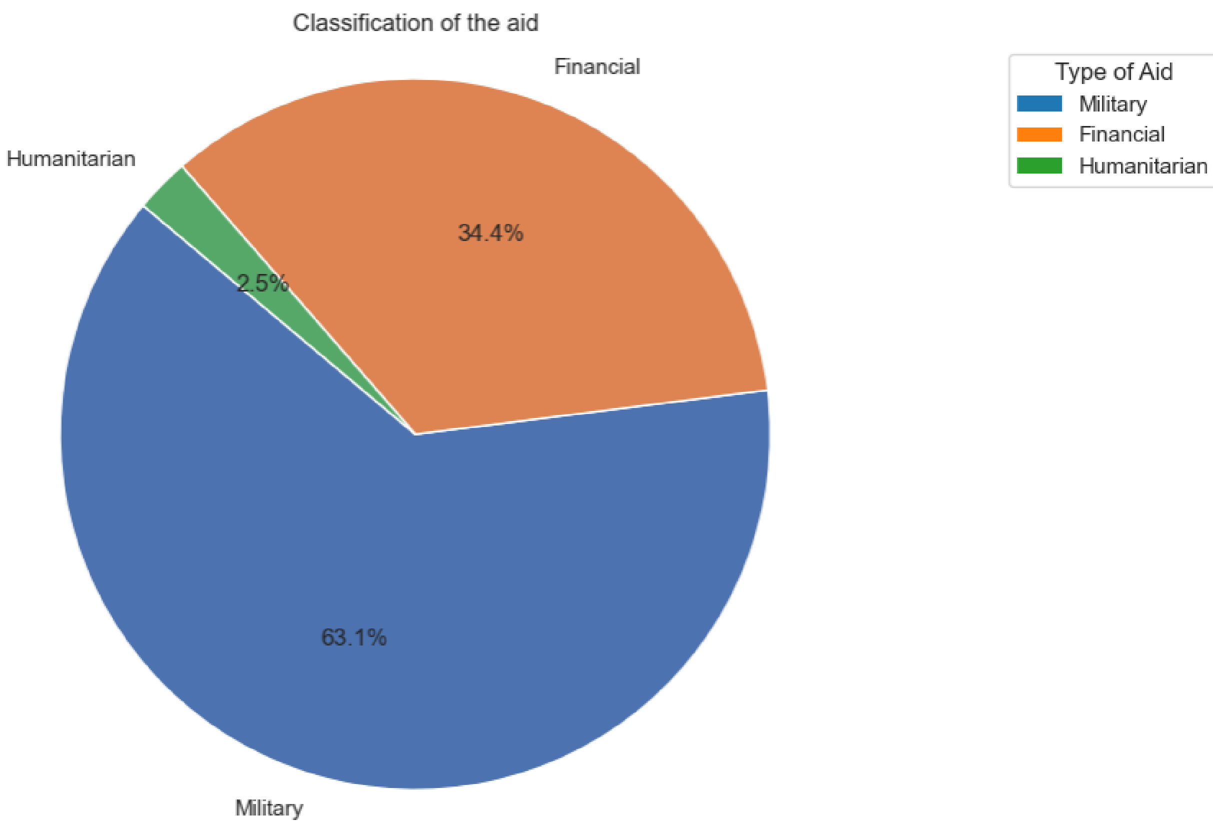
```
# Summarizing by country
country_summary = aid_data.groupby('Countries')['Monetary Value as Given by Source'].sum().reset_index()

# Sorting the summary by monetary value in descending order
country_summary = country_summary.sort_values(by='Monetary Value as Given by Source', ascending=False)

print(country_summary)
```

```
                    Countries  Monetary Value as Given by Source
40              United States                       8.716467e+11
34                South Korea                       5.030364e+11
36                     Sweden                       2.270970e+11
9                     Denmark                       2.021996e+11
10  EU (Commission and Council)                     1.115543e+11
28                     Norway                       9.271663e+10
21                      Japan                       3.659419e+10
16                    Germany                       2.266663e+10
39             United Kingdom                       1.859045e+10
4                      Canada                       1.695493e+10
26                Netherlands                       1.261615e+10
8              Czech Republic                       1.072731e+10
2                     Belgium                       6.339534e+09
13    European Peace Facility                       5.600000e+09
15                     France                       5.383800e+09
```

Aid monetary values by country

This data shows the value of the aid (including financial aid) different countries have contributed, with the United States giving the most at around 871 billion. South Korea and Sweden also gave large amounts, over 500 billion and 227 billion, respectively. Denmark, the EU, Norway, and Japan contributed billions as well. Other countries like Belgium, France, and Iceland also made smaller but still significant contributions, showing that many nations are helping.

4- Type of aid



Type of aid

Most of the aid given is military support, making up 63.1% of the total. Financial aid is next at 34.4%, showing that money is also a big part of the help provided. Humanitarian aid is the smallest at just 2.5%, meaning there is less focus on direct help for people's immediate needs. Overall, the data shows that military aid is the top priority.

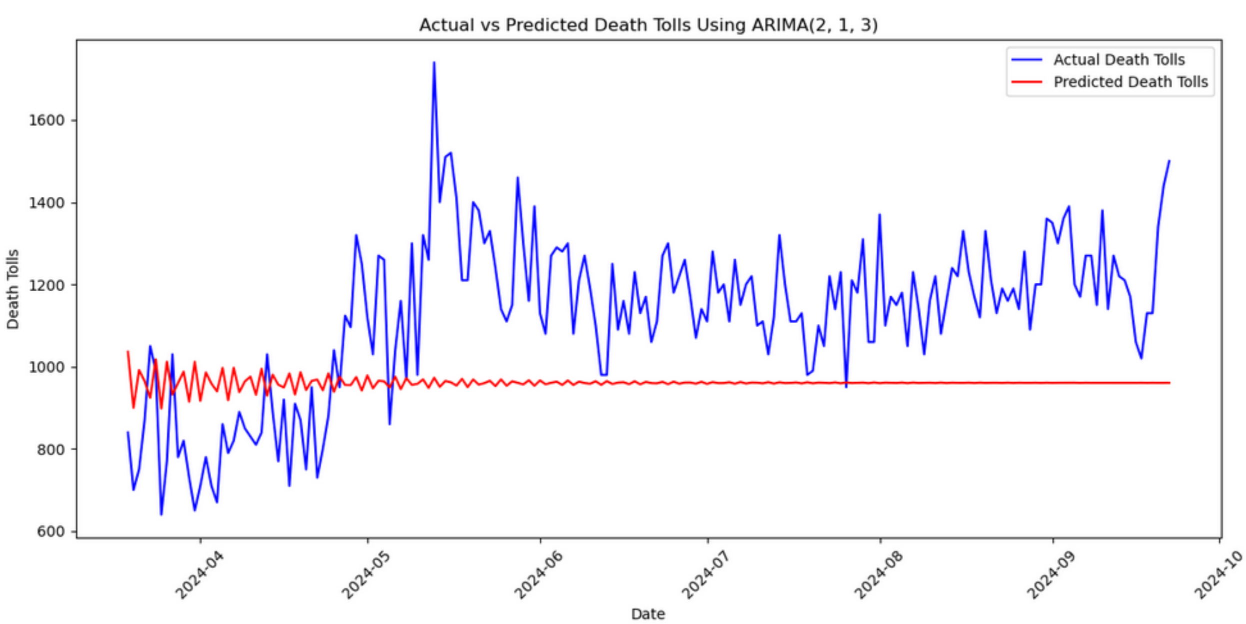
5- Future deaths predictions

```
#Predicting future losses for September 23 to September 29
index_future_dates = pd.date_range(start='2024-09-23', end='2024-09-29')
print(index_future_dates)
predictions = model.predict(start=len(losses_data),end=len(losses_data)+6, typ='levels').rename('ARIMA predictions')
predictions.index = index_future_dates
print(predictions)
```

```
DatetimeIndex(['2024-09-23', '2024-09-24', '2024-09-25', '2024-09-26',
               '2024-09-27', '2024-09-28', '2024-09-29'],
              dtype='datetime64[ns]', freq='D')
2024-09-23    960.405133
2024-09-24    960.715383
2024-09-25    960.424161
2024-09-26    960.624084
2024-09-27    960.560139
2024-09-28    960.480549
2024-09-29    960.674542
Freq: D, Name: ARIMA predictions, dtype: float64
```

According to the model, the war will cause about 960 personnel losses every day from September 23 to 29.

6- Evaluating the Arima model



Actual vs Predictions

```
Evaluation Metrics
```

```python
# Calculating the evaluation metrics
mae = mean_absolute_error(predicted_data['Actual'], predicted_data['Predicted'])
mse = mean_squared_error(predicted_data['Actual'], predicted_data['Predicted'])
rmse = mse ** 0.5  # Root Mean Squared Error

# Printing the metrics
print(f'Mean Absolute Error (MAE): {mae}')
print(f'Mean Squared Error (MSE): {mse}')
print(f'Root Mean Squared Error (RMSE): {rmse}')

```

```
Mean Absolute Error (MAE): 221.03703494220113
Mean Squared Error (MSE): 63508.332885840406
Root Mean Squared Error (RMSE): 252.00859684907658
```

Metrics

The Mean Absolute Error (MAE) is about 221, which means the predictions are typically off by around 221 units from the real values. The Mean Squared Error (MSE) is around 63508 and the Root Mean Squared Error (RMSE) is about 252. These numbers suggest that the model makes fairly good predictions, but there are some noticeable errors.

VI- Conclusion

The analysis of the aid and its impact on the ongoing conflict in Ukraine shows the complexity of the situation. We found that military aid makes up the largest portion of support, followed by financial assistance and humanitarian aid, which remains relatively small. Our predictive model, based on ARIMA, shows the estimated future losses, although it also reveals some areas for improvement, as indicated by the error metrics.

These findings can help guide policymakers in making decisions about the types of aid that may be most effective moving forward. Ultimately, understanding the different forms of aid and their implications is crucial for addressing the challenges faced in this conflict.

VII-Assumptions

We assume that the data we used is accurate and reliable. We also assume that the patterns we see in the past will continue in the future. Finally, we think that the methods we used to analyze the data are suitable for our goals.

VIII- Limitations

One limitation of our study is that we can only make predictions based on the data we have. If there are important events or changes in the situation that are not reflected in our data, our predictions may not be correct. For example, if the aid provided changes (due to changes in policies, recession, or economic factors), public opinion, international relations, or political turmoil), these factors can impact our predictions. Additionally, our models might not capture all the complexities of the real world.

IX- Challenges

We faced several challenges during this project. One major challenge was dealing with missing data, which made it hard to get a complete picture. Another challenge was figuring out the best ARIMA parameters for the model.

X- Future Uses/Additional Applications

In the future, our findings could help governments and organizations make better decisions about aid and support. Other researchers could also build on our work to explore different aspects of the situation or apply similar methods to other conflicts.

XI- Recommendations

We recommend that governments and organizations focus their efforts where they are most needed. It is important to keep updating the data and the models to ensure that they reflect the current situation accurately.

XII-Ethical Assessment

In our project, we considered the ethical implications of our work. We made sure to handle data responsibly and be sensitive to the people affected by the conflict. It's important that our findings are used to support, not harm, those in need.

References

Hernandez, J. C. (2024). Senate's aid package includes support for Ukraine, Israel, Taiwan. . *The New York Times*.

Kiel, I. (2024, 04 10). Retrieved from Institute For The World Economy: https://www.ifw-kiel.de/publications/ukraine-support-tracker-data-20758/

Peter, B. (2024). Is $60 billion enough to save Ukraine? *The New York Times*.

Wirtschaftsforschung., I. I. (2024). *Ukraine support tracker data.* nternationales Institut für Wirtschaftsforschung.

Kaggle Datasets:

- russia_losses_personnel.csv: https://www.kaggle.com/datasets/2022-ukraine-russian-war/russia_losses_personnel
- russia_losses_equipment.csv: https://www.kaggle.com/datasets/2022-ukraine-russian-war/russia_losses_equipment