

Guerra da Ucrânia:

Estatísticas de Perdas do Exército Russo

Introdução/Justificativa

Atualmente a guerra entre Ucrânia e Rússia é o principal conflito mundial ativo. A guerra da Ucrânia foi iniciada pela Rússia no dia 24 de fevereiro de 2022, devido ao interesse da Ucrânia de se tornar um território independente da Rússia;

A partir de 24 de agosto de 1991, a Ucrânia foi internacionalmente reconhecida como um Estado Soberano, após o colapso da União Soviética. Contudo, existem alguns conflitos relacionados a certas regiões do país que fazem fronteira direta com a Rússia, e alguns outros territórios que possuem acesso direto ao mar. Desde 2013, a Ucrânia demonstrou interesse em se tornar um membro oficial da União Europeia. Caso a Ucrânia fosse efetivada como membro da União Europeia, o reconhecimento internacional da independência de toda a região seria ainda mais claro, mas essa efetivação não ocorreu;

A Rússia é historicamente reconhecida por oprimir seus vizinhos que fazem fronteiras diretas com o país. Um caso específico a ser citado, e que agravou as tensões entre Rússia e Ucrânia foi a anexação da Crimeia pela Rússia (uma região que até 2014, era reconhecida como parte da Ucrânia). Isso deixou a nação ucraniana em estado de alerta, devido a política de agressividade da Rússia com relação à anexação de territórios.

No dia 24 de fevereiro de 2022 a Rússia invadiu a Ucrânia, com o objetivo de tomar alguns territórios específicos do país para fortalecer o território Russo para frear o possível avanço da OTAN no território europeu, dentre outros motivos. Desde então, a Rússia tem direcionado seus esforços armamentistas para militarizar o território da Ucrânia e cumprir o objetivo de anexar territórios originalmente ucranianos à Rússia.

Objetivo

A objetivo desse relatório é avaliar o esforço deferido pela Rússia, de forma gráfica, em anexar partes do território Ucraniano ao seu país com baseado em um dataset de perdas do exército russo, composto por dados como número de soldados que vieram a óbito, prisioneiros de guerra, número de aviões, helicópteros, tanques de guerra enviados a Ucrânia, dentre outros indicadores e métricas que foram colhidos diretamente do Ministério de Defesa da Ucrânia.

Materiais e Metodologia

Para que seja possível identificar os principais indicadores relacionados à guerra, é necessário colher os dados desde o início da guerra da Ucrânia até o momento presente. Para isso, uma base de dados encontrada no Kaggle [1], alimentada por um cidadão Ucrâniano, baseada na base de dados do Ministério de Defesa da Ucrânia será utilizada.

O conjunto de dados utilizado descreve as Perdas de Equipamentos, o Número de Mortos, os Militares Feridos e os Prisioneiros de Guerra russos na Guerra entre Ucrânia e Rússia em

Os dados serão colhidos em formato CSV, e após o download dos dados, eles serão diretamente vinculados ao Google Collaborator, com o intuito de utilizar as ferramentas de armazenamento e processamento de dados disponíveis gratuitamente na plataforma. Além disso, o objetivo de realizar upload dos dados e do projeto no Google Collaborator é tornar o projeto aberto para a contribuição de outras pessoas em análises futuras.

As fontes de dados são tabulares e a maior parte das análises serão realizadas em Python, com as bibliotecas NumPy, Pandas, Matplotlib, Seaborn, e Statsmodels, dentre outras bibliotecas famosas contidas no Python, utilizando o Google Collaboratory para compilação em nuvem. Essas bibliotecas abrangem o conjunto de funcionalidades específicas para a realização de análise de dados, como: limpeza e tratamento de dados, criação e visualização de gráficos, manipulação estatística para criação de novas métricas baseadas nos dados fornecidos. Além disso, elas possuem fácil compreensão por usuários iniciantes e intermediários da linguagem python e possuem um bom nível de integração com outras linguagens de programação.

O formato de entrada dos dados deve ser o próprio arquivo em formato CSV, que possui os dados necessários para realizar as análises. Depois do upload dos dados, será realizada a limpeza de informações irrelevantes e de colunas que não serão usadas na análise de dados. Logo após esse processo, a análise exploratória dos dados será realizada, com o objetivo de identificar informações relevantes, entender a distribuição dos dados, identificar padrões, tendências e anomalias e utilizar visualizações gráficas para representar os dados de maneira reconhecível. Após a exploração dos dados, será realizado o processo de normalização, e em paralelo, a criação de novas variáveis relevantes para análise. Por fim, a última fase do projeto é a interpretação de resultados para a disponibilização pública deles.

Análise Exploratória

O objetivo da análise exploratória é entender a natureza dos dados é crucial para realizar análises concisas. Para isso, através do método `df.info()` foi possível visualizar que esse dataframe contém 200 colunas, 675 linhas, nas quais os dados são compostos por números inteiros e decimais. Também foi possível observar algumas colunas que continham dados denominados de Not a Number (NaN), ou seja, espaços vazios na tabela. Esses dados foram tratados postumamente.

```

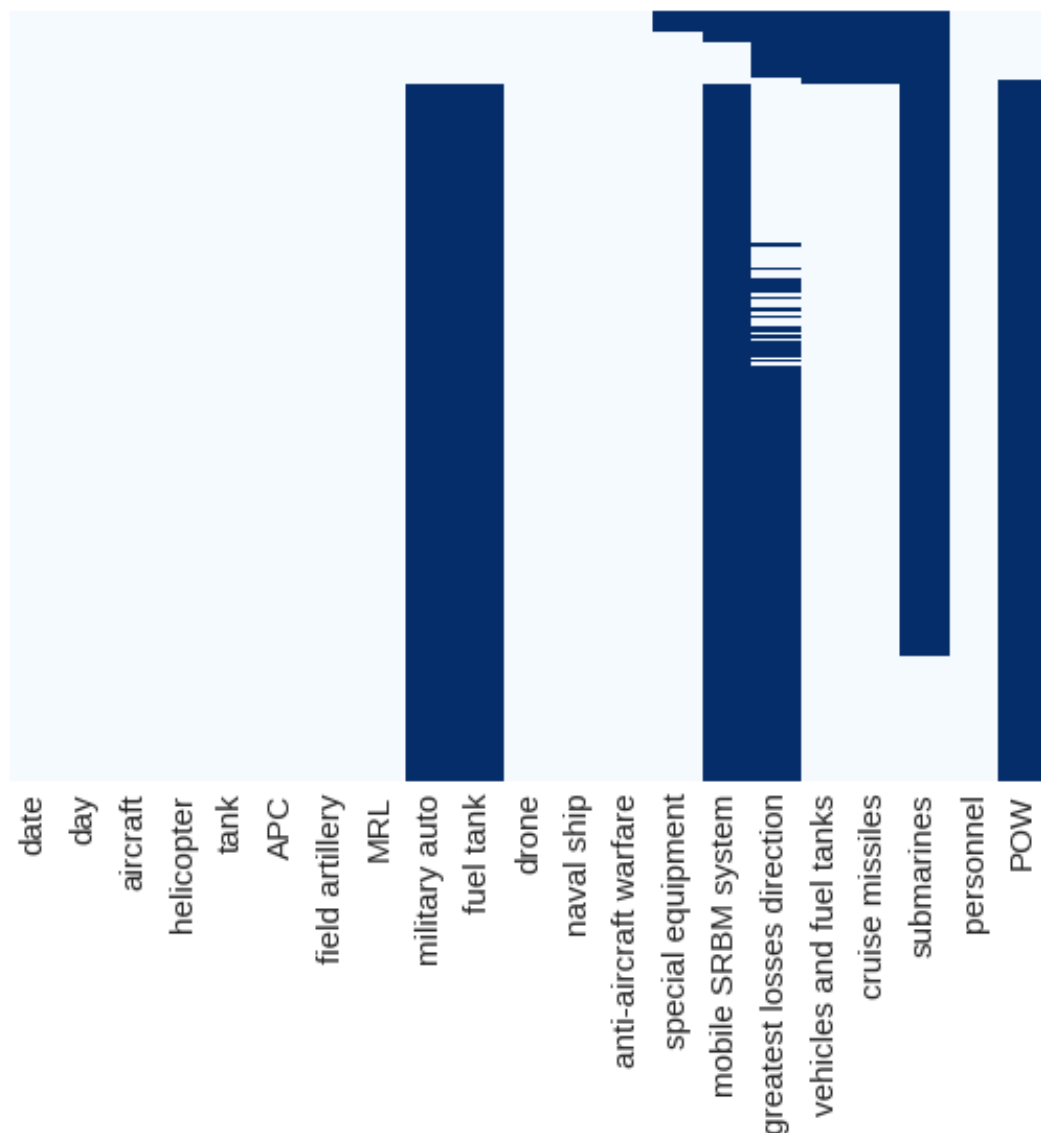
RangeIndex: 675 entries, 0 to 674
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   date                                675 non-null    object
1   day                                675 non-null    int64
2   aircraft                            675 non-null    int64
3   helicopter                          675 non-null    int64
4   tank                                675 non-null    int64
5   APC                                675 non-null    int64
6   field artillery                     675 non-null    int64
7   MRL                                675 non-null    int64
8   military auto                       65 non-null     float64
9   fuel tank                           65 non-null     float64
10  drone                               675 non-null    int64
11  naval ship                          675 non-null    int64
12  anti-aircraft warfare                675 non-null    int64
13  special equipment                   656 non-null    float64
14  mobile SRBM system                  36 non-null     float64
15  greatest losses direction            203 non-null    object
16  vehicles and fuel tanks              610 non-null    float64
17  cruise missiles                     610 non-null    float64
18  submarines                          109 non-null    float64
19  personnel                           675 non-null    int64
20  POW                                 62 non-null     float64
dtypes: float64(8), int64(11)

```

A partir dessas informações, foi possível observar as informações do dataset que precisam ser tratadas, em conjunto com as colunas de interesse para a execução da análise.

Após a visualização, foi necessário obter uma visualização gráfica das colunas NaN, para que fosse possível entender como os espaços vazios afetam a análise dos dados e o data frame no geral.

Abaixo, é possível ter uma visualização gráfica dos dados que não estavam preenchidos na tabela.

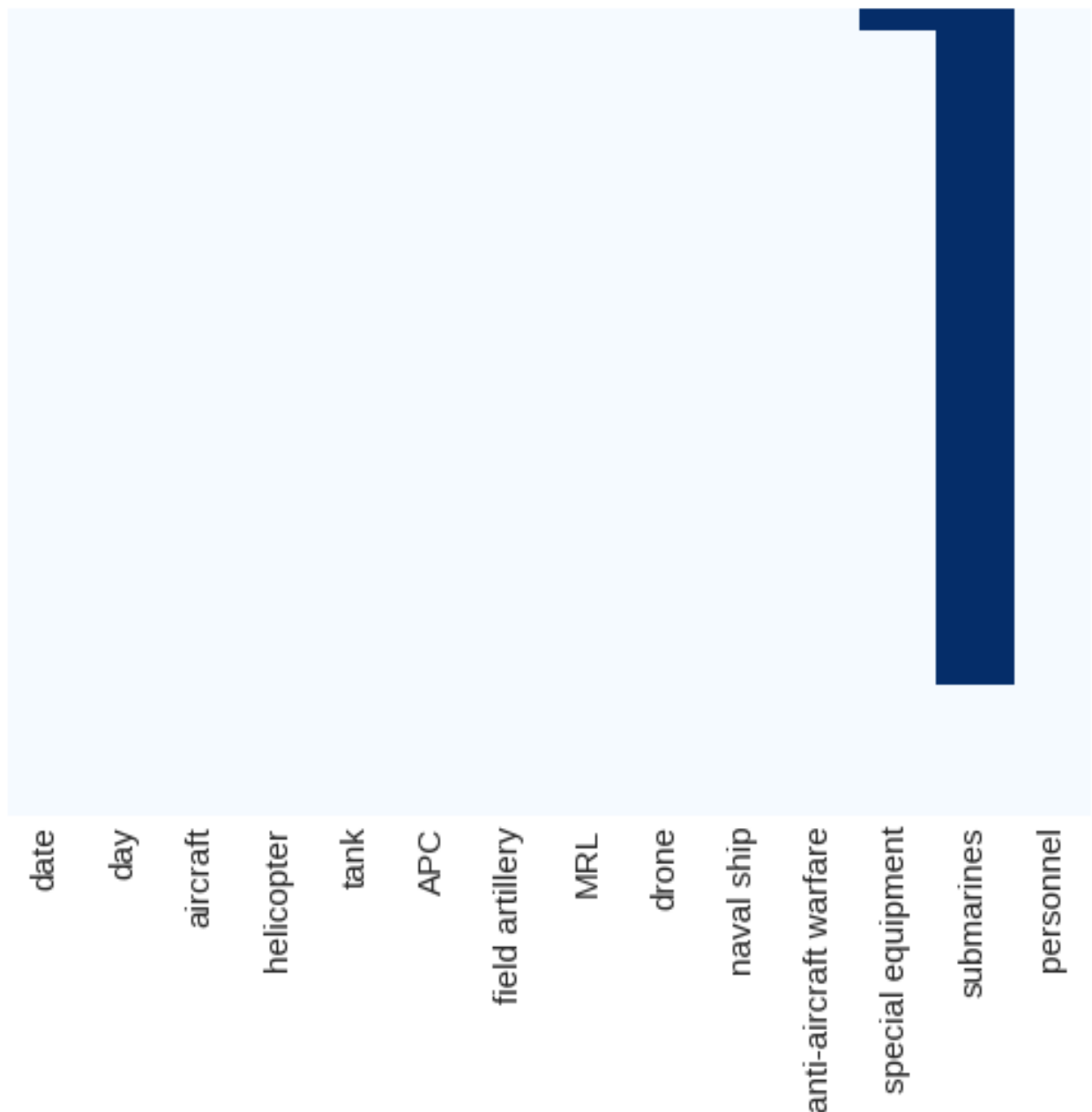


Limpeza dos Dados

Em posse do conhecimento das colunas que deveriam ser usadas para a análise dos dados, e dos dados que foram registrados como NaN, foi possível realizar a limpeza dos dados para que o dataset pudesse ser avaliado da forma mais correta possível sem ser acometido por desvio de dados.

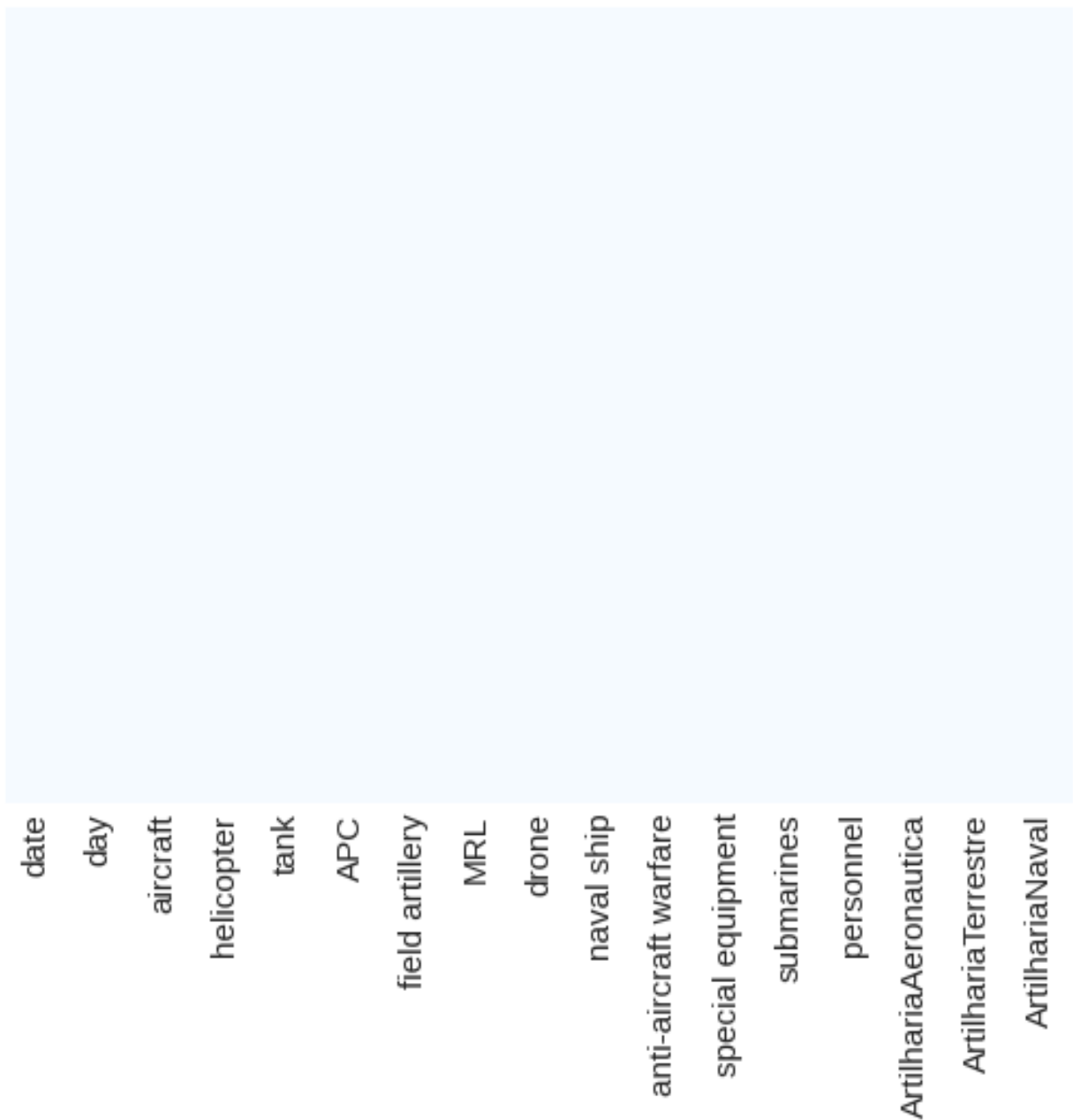
Originalmente, a tabela possuía 20 colunas, mas nesse estudo, nem todas seriam utilizadas no estudo. Dessa forma as colunas 'POW', 'fuel tank', 'military auto', 'mobile SRBM system', 'greatest losses direction', 'vehicles and fuel tanks', 'cruise missiles' foram excluídas do dataset.

Após esse processo, o resultado da visualização de dados obtida foi esse:



É possível observar que, mesmo após a exclusão das colunas que não seriam utilizadas, as colunas que seriam utilizadas (“special equipment” e “submarines”) continham dados nulos. Para tratar esse problema, “special equipment” e “submarines” foram tiveram seus campos NaN substituídos por 0, devido a ausência de dados na contagem dos itens registrados no dataset.

Após essa substituição, foi possível manipular os dados para que depois a análise gráfica fosse implementada da forma mais correta possível. No gráfico abaixo, podemos notar que após o tratamento de dados, os NaN foram eliminados.



Análise Tabular

O objetivo da análise tabular é extrair as principais informações e métricas do dataset. Para isso, realizei a execução do método `df.describe()`. Esse método é útil porque permite a sumarização, a identificação de padrões e a comparação de categorias. É eficaz para visualizar relações e discrepâncias nos dados. Além disso, o uso desse método teve como objetivo obter métricas estatísticas descritivas, como média, desvio padrão e quartis, oferecendo uma visão abrangente das características centrais, variabilidade e identificação de anomalias nos dados. Abaixo, é possível conferir as principais métricas do dataset utilizado.

	count	mean	std	min	25%	50%	75%	max
day	675.0	339.000000	195.000000	2.0	170.5	339.0	507.5	676.0
aircraft	675.0	267.152593	64.210259	10.0	233.0	293.0	315.0	329.0
helicopter	675.0	252.845926	68.189581	7.0	193.0	284.0	310.0	324.0
tank	675.0	3087.740741	1461.455039	80.0	1852.5	3189.0	4104.5	5977.0
APC	675.0	6171.054815	2537.646733	516.0	4111.5	6344.0	8022.5	11070.0
field artillery	675.0	2956.068148	2379.435588	49.0	976.5	2188.0	4472.0	8464.0
MRL	675.0	474.506667	244.134421	4.0	261.0	453.0	680.5	943.0
drone	675.0	2369.122963	1852.747238	0.0	778.5	1947.0	3813.5	6591.0
naval ship	675.0	16.063704	4.288253	2.0	15.0	18.0	18.0	23.0
anti-aircraft warfare	675.0	278.162963	175.143341	0.0	136.0	221.0	425.0	623.0
special equipment	675.0	378.847407	369.440742	0.0	90.5	199.0	665.0	1268.0
submarines	675.0	0.161481	0.368247	0.0	0.0	0.0	0.0	1.0
personnel	675.0	144856.727407	104888.396656	2800.0	43300.0	125510.0	237430.0	359230.0

Principais métricas descritivas do dataset.

A partir dessa visualização, foram separadas algumas colunas específicas para possibilitar a avaliação da artilharia de guerra usada, usando a diferenciação entre artilharia naval, terrestre e aeronáutica.

Dessa forma, as colunas de interesse relacionadas às 3 respectivas categorias foram organizadas como exemplificado no seguinte código:

```
ukraine_war_cleaned_df['ArtilhariaAeronautica'] =
ukraine_war_cleaned_df[['aircraft', 'helicopter',
'drone']].sum(axis=1)
```

O mesmo processo foi repetido para artilharia naval e terrestre, e também para as armas específicas utilizadas na guerra, pela Rússia, utilizando o código:

```
colunas Equipamento Guerra = [ 'aircraft', 'helicopter',
'tank', 'field artillery', 'MRL', 'drone', 'naval ship',
'anti-aircraft warfare', 'special equipment', 'submarines']
```

Dessa forma, foi possível obter a compreensão dos dados de interesse agrupados por categorias, como consta abaixo:

	count	mean	std	min	25%	\
ArtilhariaAeronautica	675.0	2889.121481	1959.994739	17.0	1204.5	
ArtilhariaTerrestre	675.0	13346.380741	7057.541488	649.0	7428.0	
ArtilhariaNaval	675.0	16.225185	4.485349	2.0	15.0	
	50%	75%	max			
ArtilhariaAeronautica	2524.0	4438.5	7244.0			
ArtilhariaTerrestre	12594.0	18369.5	28345.0			
ArtilhariaNaval	18.0	18.0	24.0			

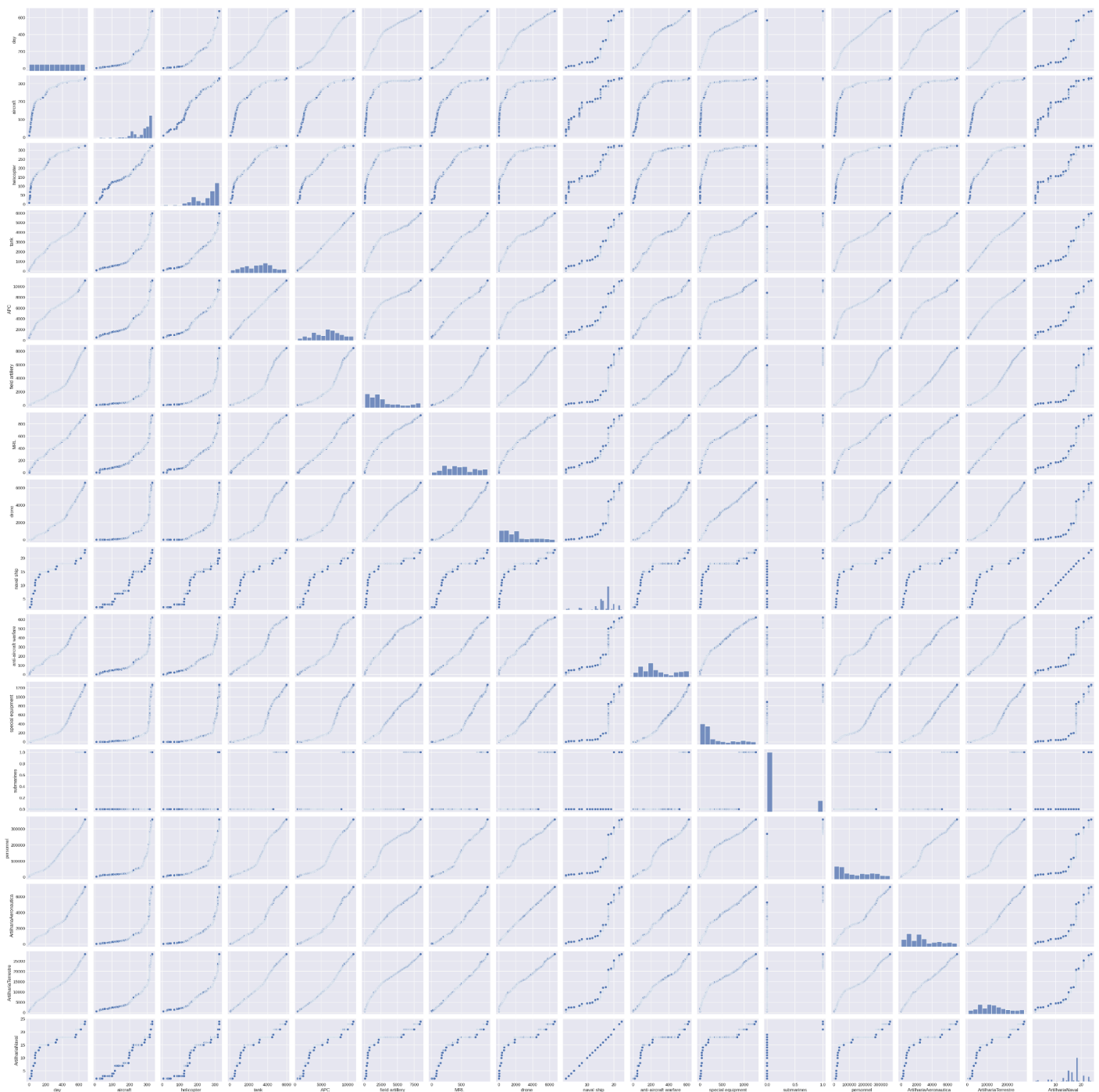
Tabela Referente a artilharia aeronáutica, naval e terrestre, utilizada pela Rússia.

	aircraft	helicopter	tank	field artillery	MRL	drone	naval ship	\
0	10	7	80	49	4	0	2	
1	27	26	146	49	4	2	2	
2	27	26	150	50	4	2	2	
3	29	29	150	74	21	3	2	
4	29	29	198	77	24	3	2	
..	
670	329	324	5913	8376	934	6471	23	
671	329	324	5940	8391	935	6503	23	
672	329	324	5953	8417	938	6539	23	
673	329	324	5969	8434	939	6554	23	
674	329	324	5977	8464	943	6591	23	
	anti-aircraft warfare	special equipment	submarines					
0	0	0.0	0.0					
1	0	0.0	0.0					
2	0	0.0	0.0					
3	5	0.0	0.0					
4	7	0.0	0.0					
..					
670	616	1240.0	1.0					
671	617	1247.0	1.0					
672	620	1254.0	1.0					
673	621	1260.0	1.0					
674	623	1268.0	1.0					

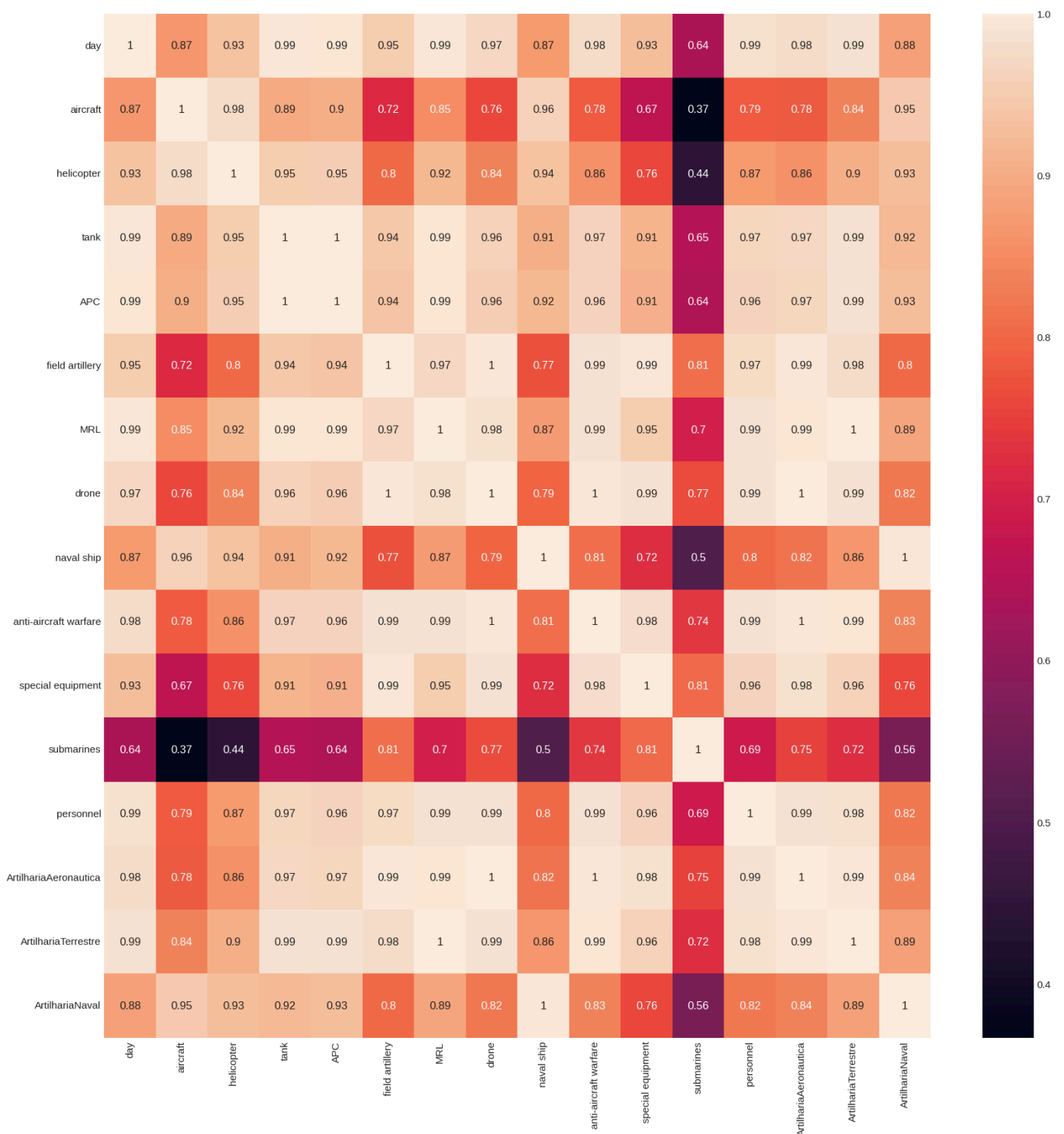
Tabela referente ao número específico de armistício utilizado pela Rússia.

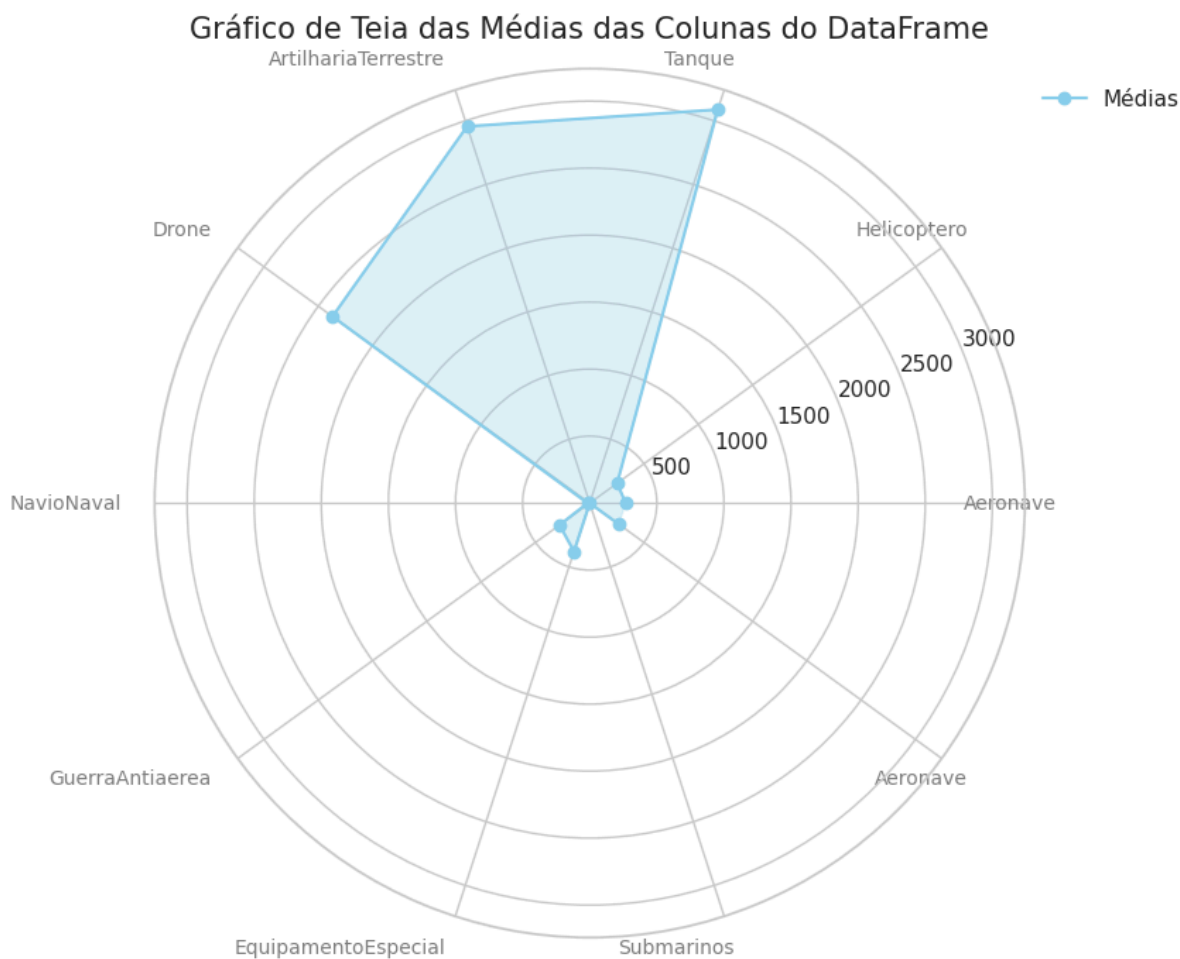
Análise gráfica dos dados

O objetivo da análise gráfica é a culminância do projeto para observar as principais variações dos dados ao longo do tempo. Inicialmente, utilizei o `sns.pairplot()` para criar gráficos de dispersão bivariados com o objetivo de visualizar as relações e o comportamento das variáveis no dataset para ter uma visualização clara dos dados.

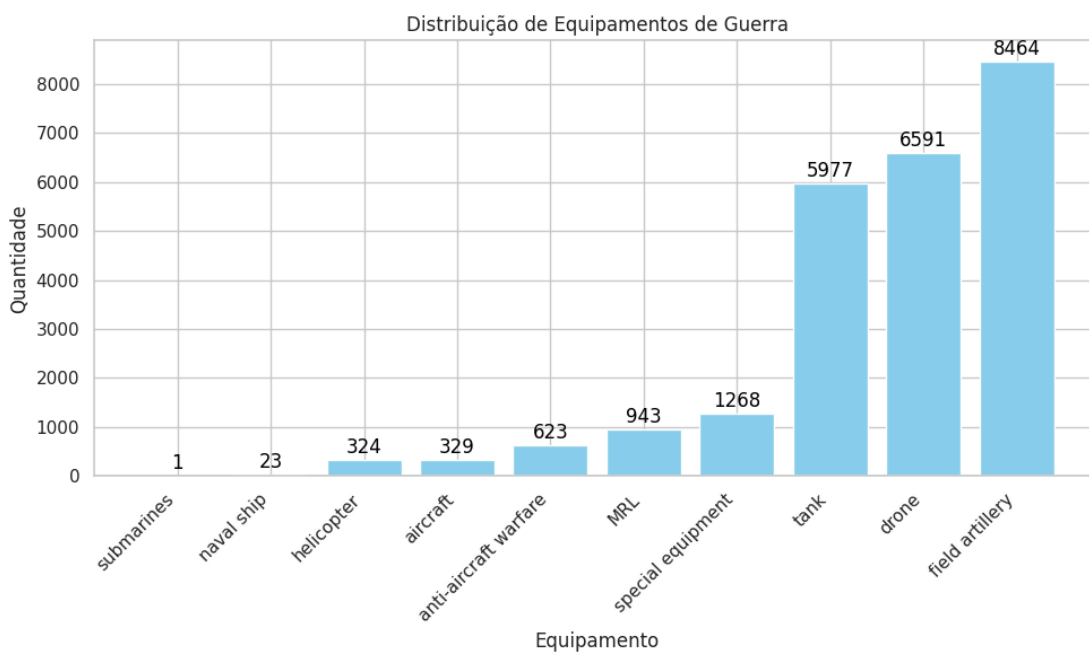


Após obter a variação desses dados, criei uma matriz de correlação paraa entender quais métricas se relacionavam entre si ao longo do tempo, para levantar possíveis hipóteses, como: O número de drones utilizados aumentou de forma proporcional ao número de aeronaves utilizadas? Quantas pessoas foram mortas ao longo do tempo? O número de pessoas mortas ao longo do tempo foi proporcional ao investimento em artilharia? Qual foi o investimento em artilharia aeronáutica, naval, e terrestre? Quais foram as armas de guerra mais utilizadas?

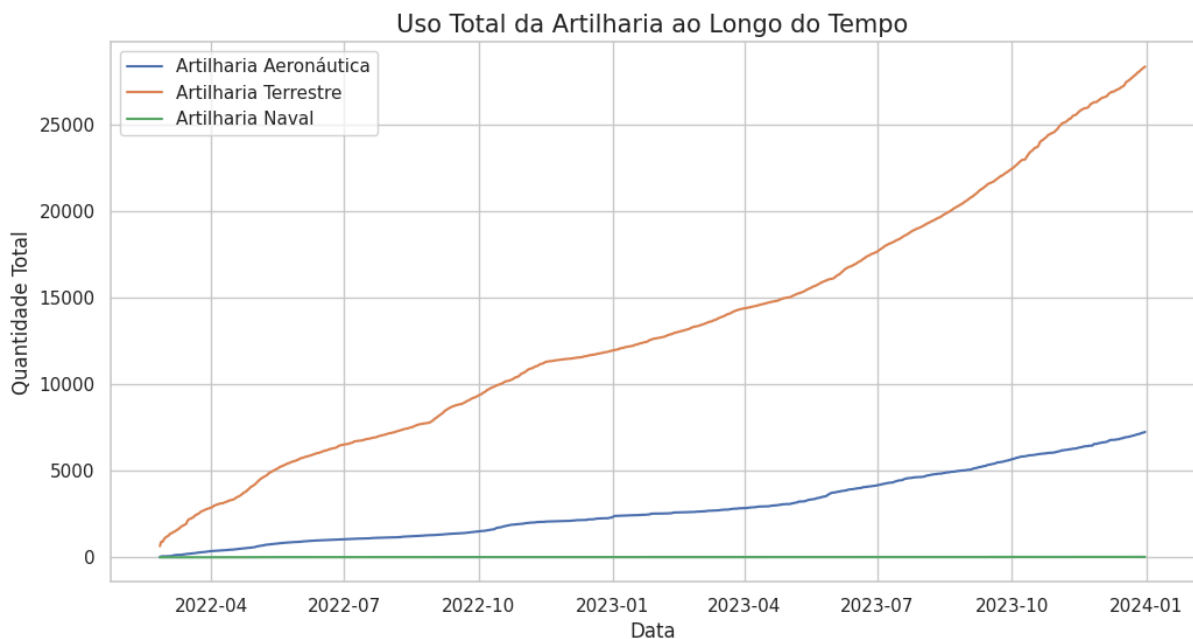




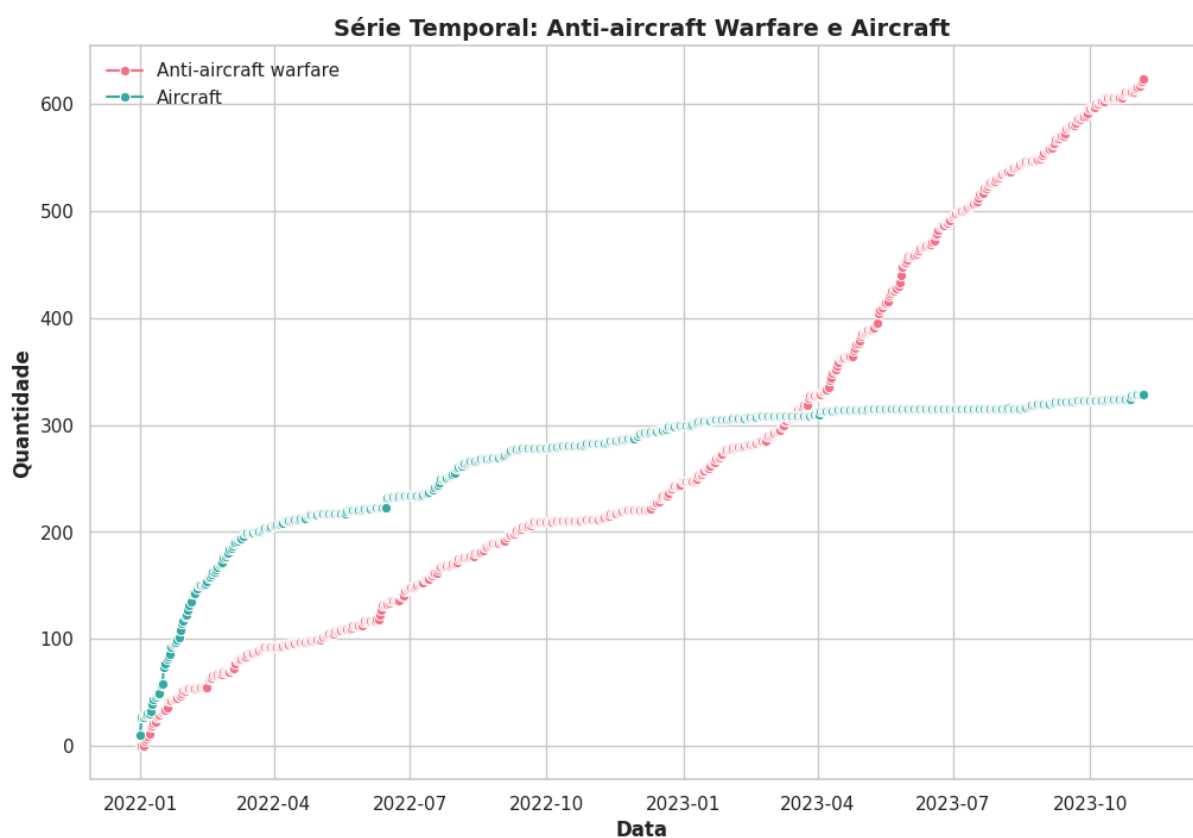
O gráfico exibe que a média do número de artilharia terrestre utilizado é superior à média do uso de artilharia naval e aeronáutica. Em paralelo, o número médio de tanques é proporcional ao número de artilharia terrestre utilizada.



O gráfico exibe que o número absoluto de drones utilizados supera o de tanques, e é apenas menor do que o número de artilharia terrestre.

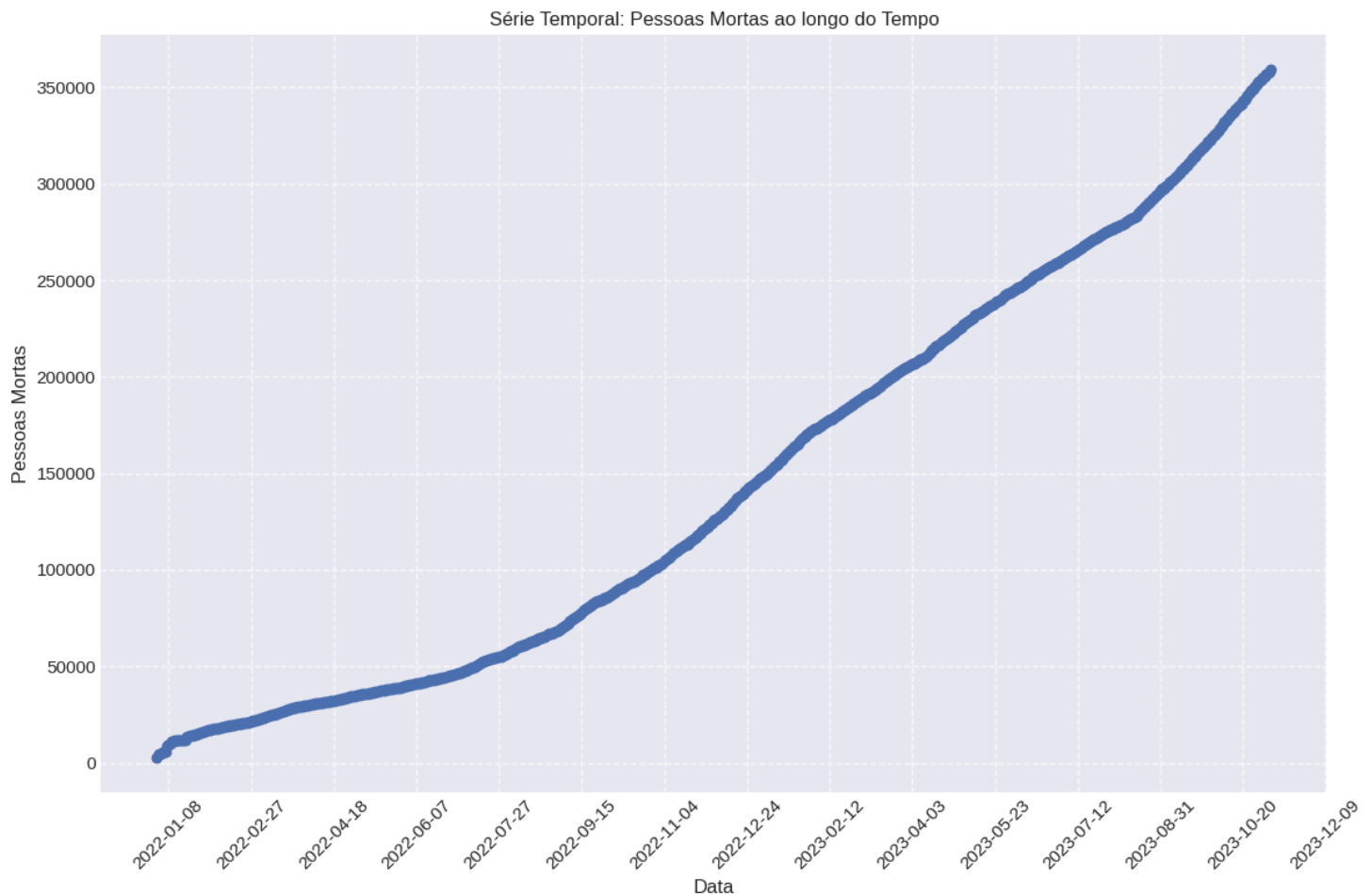


O gráfico exibe que o investimento em artilharia terrestre foi maior do que o investimento em artilharia naval e aeronáutica, ao longo do tempo.



O gráfico exibe a evolução ao longo do tempo do número de aeronaves e do uso de defesa antiaérea pela Rússia. Observa-se uma estabilização no uso de aeronaves,

indicando uma relativa constância nesse aspecto. No entanto, destaca-se um aumento significativo no emprego de equipamentos de defesa antiaérea



O gráfico exhibe que o número de mortes aumentou ao longo do tempo.

Conclusões

1. O maior esforço armamentista Russo está relacionado à artilharia terrestre, tendo em vista que o objetivo do país é anexar territórios da Ucrânia, por razões geopolíticas.
2. O número absoluto de drones utilizados foi de 6591. O número absoluto de tanques foi de 5977. Isso indica que os drones estão sendo utilizados diretamente como sistema de apoio para a guerra em solo.
3. Por mais que ao longo do tempo o uso de aeronaves tenha se estabilizado, o número de sistemas de defesa-anti-aérea cresceu ao longo do tempo, o que indica.
4. Mais de 350.000 soldados russos foram mortos, entre os períodos de 02/2022 e 12/2023.

Apêndice

Acrônimos:

POW: Prisioneiro de Guerra,
MRL: Lançador Múltiplo de Foguetes,
APC: Veículo Blindado de Transporte de Pessoal,
SRBM: Míssil Balístico de Curto Alcance,
UAV: Veículo Aéreo Não Tripulado,
RPA: Veículo Pilotado Remotamente.

Histórico do Conjunto de Dados:

2022-03-02: Conjunto de dados criado (após 7 dias de guerra).

Referências

1. Petro Ivaniuk. (2022). <i>2022 Ukraine Russia War</i> [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/1967621>
2. “Google Sheets: Online Spreadsheet Editor.” n.d. Google. Accessed Jan 18, 2024. <https://docs.google.com/spreadsheets/u/0/?tgif=d>.
3. Ariel Bandeira. (2024). <i>Guerra da Ucrânia: Estatísticas de Perdas do Exército Russo</i>. https://colab.research.google.com/drive/1yoDXCwB_MfsYc0icgN8xcEDaGA9ho1bz?usp=sharing