

Analýza strát ruskej techniky na Ukrajine

Šimon Fabuš

August 8, 2024

V tejto štatistickej práci chcem skúmať straty ruskej techniky a munície na Ukrajine od začiatku Ruskej invázie 24.2.2022. Zaujíma ma najmä závislosť medzi dňom v týždni a veľkosťou strát techniky. Z načrtnutých dát sa dajú ďalej pozorovať určité štatistické zaujímavosti.

Skúmané dáta pochádzaju z datasetu: https://www.kaggle.com/datasets/piterfm/2022-ukraine-russian-war?select=russia_losses_equipment.csv. Ktoré su aktualizované každý týždeň. V datasete sa nachádzajú aj záznamy o počte zranených, počtu obetí alebo POW nás budú zaujímať len straty techniky a munície.

Čistenie dát

Pretože som chcel skúmať hlavne straty techniky vyniechal som záznamy o strate munície a zameral sa na helikoptéry, tanky, drony, lode atď. Ajked' v strate munície je vidno podobný trend. Pretože ma zaujíma vzťah voči dňom v týždni vyniechal som pári posledných dní aby ich bol rovnaký počet. Jedná sa o 893 dní.

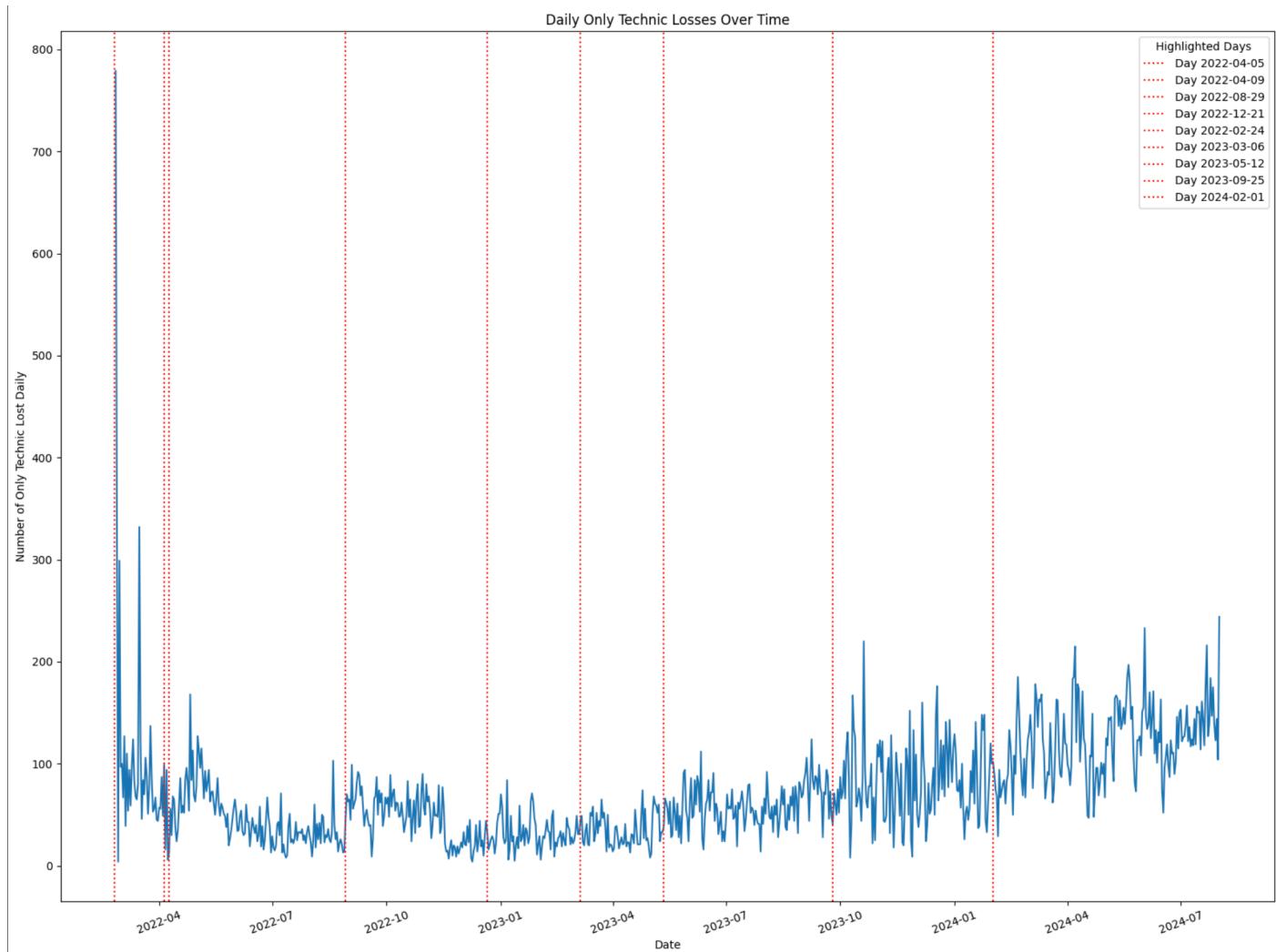
Príprava a ukážka dát

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from scipy.stats import ttest_1samp

df = pd.read_csv('archive/russia_losses_equipment.csv')
df['date'] = pd.to_datetime(df['date'])
df.set_index('date', inplace=True)
print(df)
```

	day	aircraft	helicopter	tank	APC	field artillery	MRL	\	
date									
2024-08-01	890	363	326	8399	16203	16119	1132		
2024-07-31	889	363	326	8390	16161	16086	1131		
2024-07-30	888	363	326	8381	16141	16056	1130		
2024-07-29	887	363	326	8356	16103	16010	1129		
2024-07-28	886	363	326	8344	16095	15936	1127		
...		
2022-03-01	6	29	29	198	846	77	24		
2022-02-28	5	29	29	150	816	74	21		
2022-02-27	4	27	26	150	706	50	4		
2022-02-26	3	27	26	146	706	49	4		
2022-02-25	2	10	7	80	516	49	4		
		military	auto	fuel	tank	drone	naval	ship	\
date									
2024-08-01		NaN	NaN	12968		28			
2024-07-31		NaN	NaN	12853		28			
2024-07-30		NaN	NaN	12832		28			
2024-07-29		NaN	NaN	12805		28			
2024-07-28		NaN	NaN	12766		28			
...			
2022-03-01		305.0	60.0	3		2			
2022-02-28		291.0	60.0	3		2			
2022-02-27		130.0	60.0	2		2			
...									
2022-02-26		NaN	NaN						
2022-02-25		NaN	NaN						

Miesta, na ktorých je záznam NaN som nastavil na 0. Následne si vizualizujem graf denných strát techniky a vyznačím si niektoré zaujímavé udalosti. Je vidno že veľa týchto udalostí nasledoval nárast strát ruskej techniky.



Z grafu je vidno, že najvyššie straty boli na začiatku invázie, v nasledujúcich mesiacoch klesali a celkom výrazne kolísali. Od cca apríla 2023 začali straty znova stúpať. Kolisavosť je podľa môjho názoru zapríčinená pripravovaním väčších útokov a taktiež obliehaním miest. Vplyv mali tiež momenty kedy sa začala využívať darovaná technika Ukrajinskej armáde.

Udalosti vyznačené v grafe

2022-02-24 - Obliehanie Mariupolu

2022-04-05 - Dodanie prvých tankov od ČR. <https://www.reuters.com/world/europe/czech-republic-sends-tanks-ukraine-czech-tv-reports-2022-04-05/>

2022-04-09 - Dodanie protivdušných systémov obrany od SR. <https://www.mosr.sk/51353-en/s-300-nahradil-jeden-z-najmodernejsich-systemov-protivzdusnej-obrany/>

2022-07-03- Obliehanie Bakhmutu

2022-08-29 - Zahájenie ukrajinskej protiofenzívy. <https://www.politico.eu/article/ukraine-liberates-kherson-army-war-defeat-russia-dnipro/>

2022-12-21 - USA poskytla Ukrajine protivdušný systém obrany Patriot. <https://www.reuters.com/world/europe/what-is-patriot-missile-defense-system-2022-12-21/>

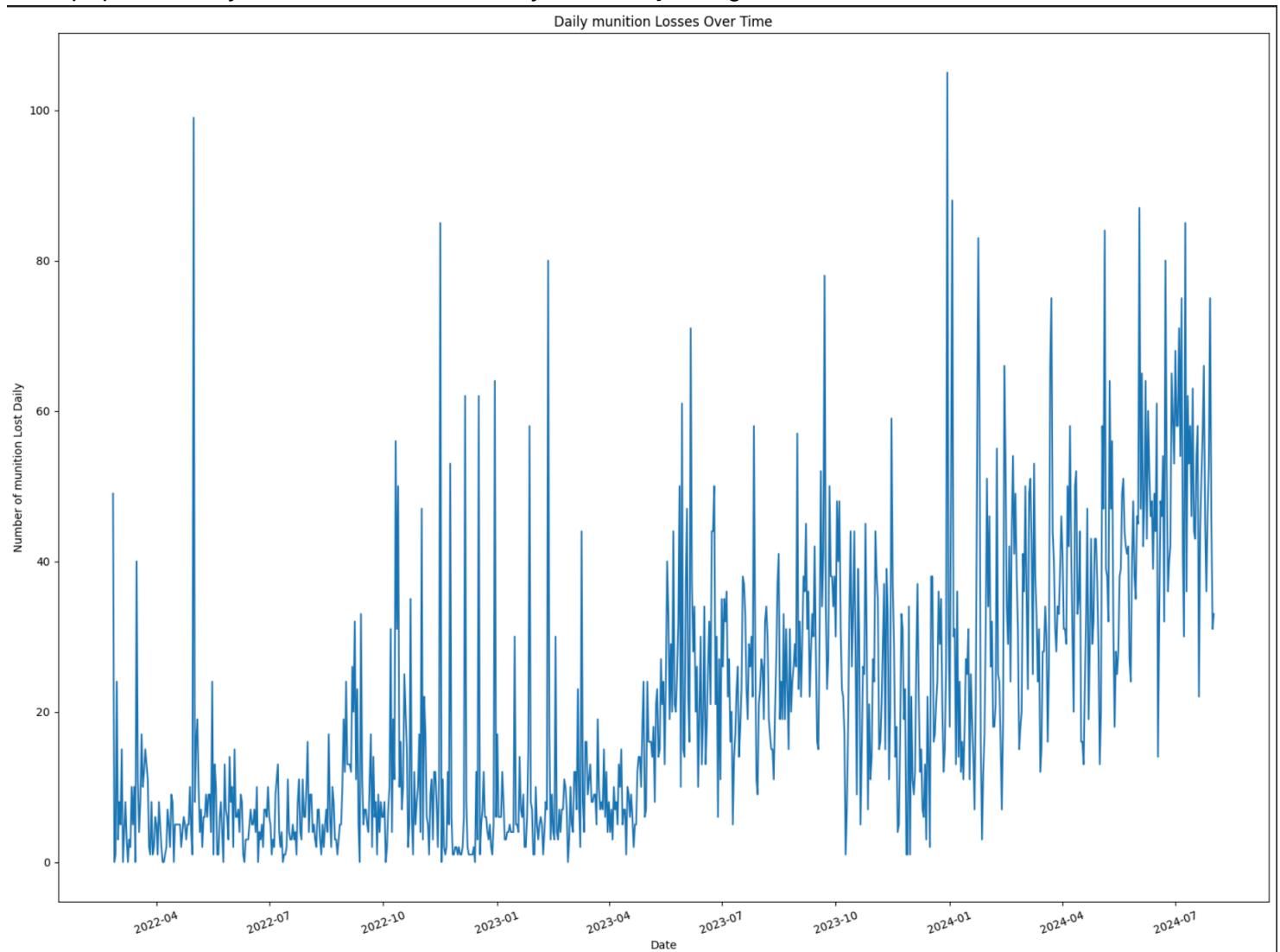
2023-03-06 - Najnáročnejšie boje v Bakhmute.(7. marca sa ukrajinská armáda stiahla z Bakhmutu).

2023-05-12 - Doručenie stromshadow air-lunched cruise-missile od UK

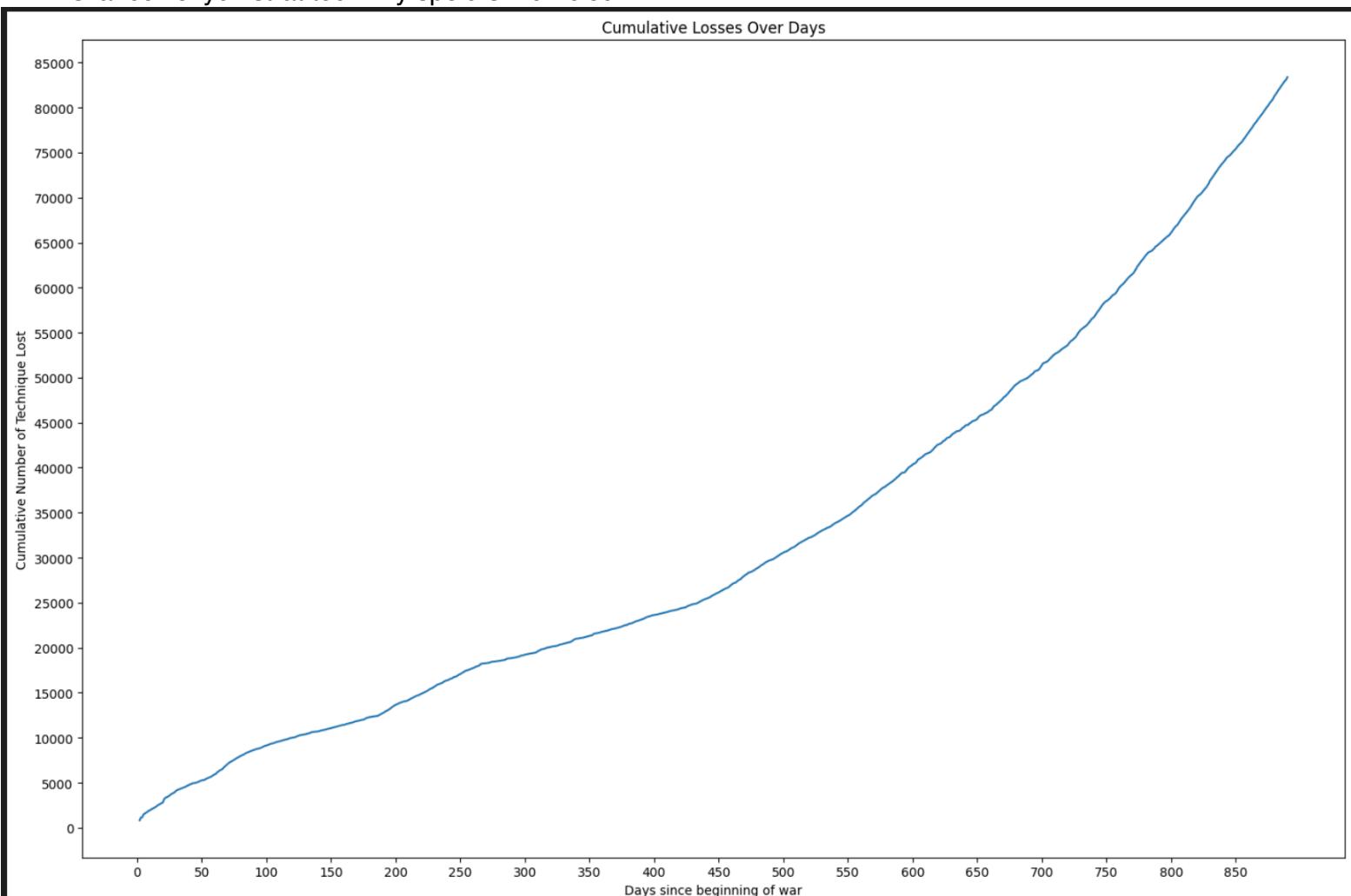
2023-09-25 - Ukrajine boli dodané prvé tanky M1 Abrams. <https://www.nytimes.com/2023/09/25/world/europe/us-abrams-tanks-ukraine.html>

2024-02-01 - EU sa odhodla na darovaní 54 miliárd dolárov pre Ukrajinu. https://www.eeas.europa.eu/delegations/united-states-america/eu-assistance-ukraine-us-dollars_en?s=253

V prípade denných strát munície a riadených striel vyzerá graf obdobne:

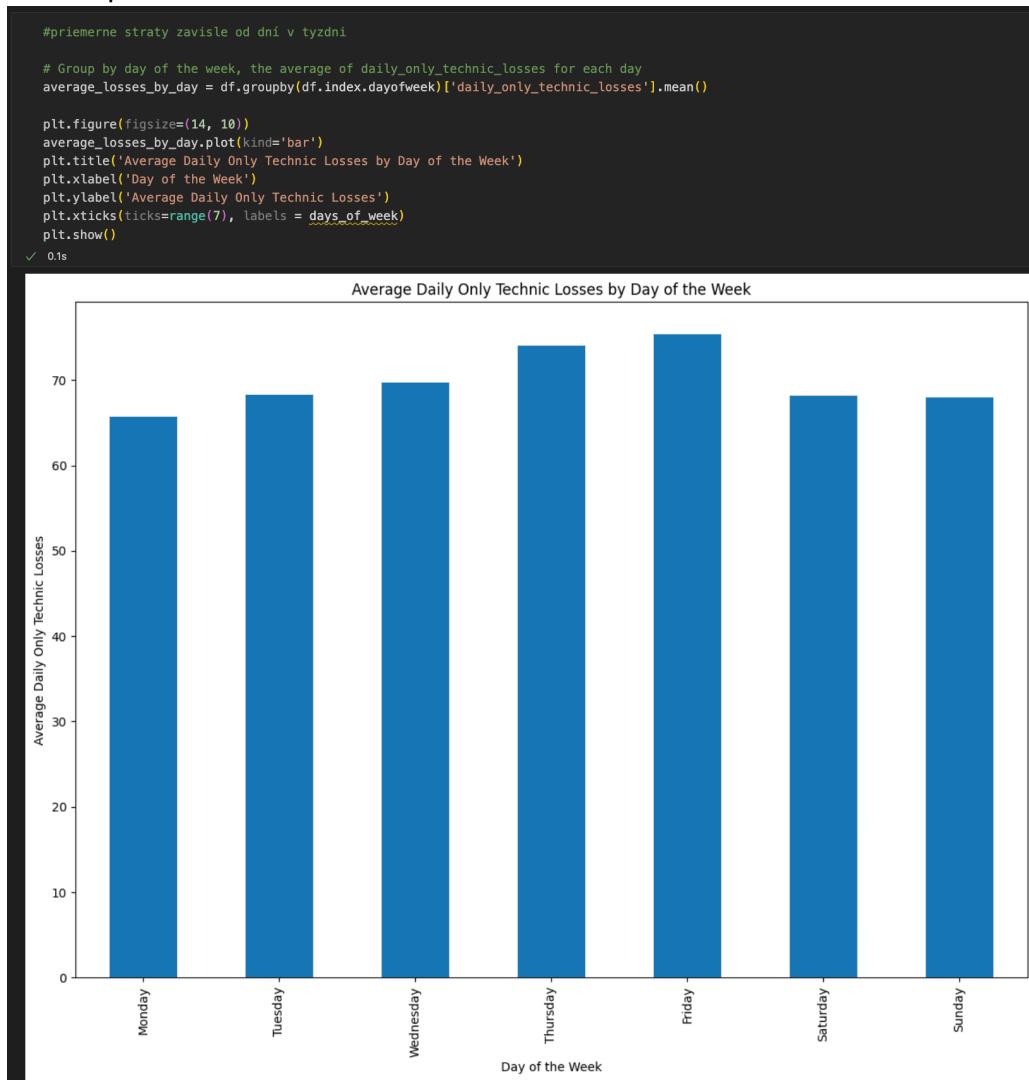


Graf celkových strát techniky spolu s muníciou:



Analýza

Moja nulová hypotéza, ktorú budem skúmať H_0 : Veľkosť strát techniky nie je závislá od dňa v týždni. Čo v našom prípade znamená, že strata techniky z konkrétneho dňa sa nebude výrazne lísiť od očakávanej priemernej straty. Z grafu priemerných strát v dni týždňa je však vidieť že najviac sa v priemere stratí v piatok a vo štvrtok.



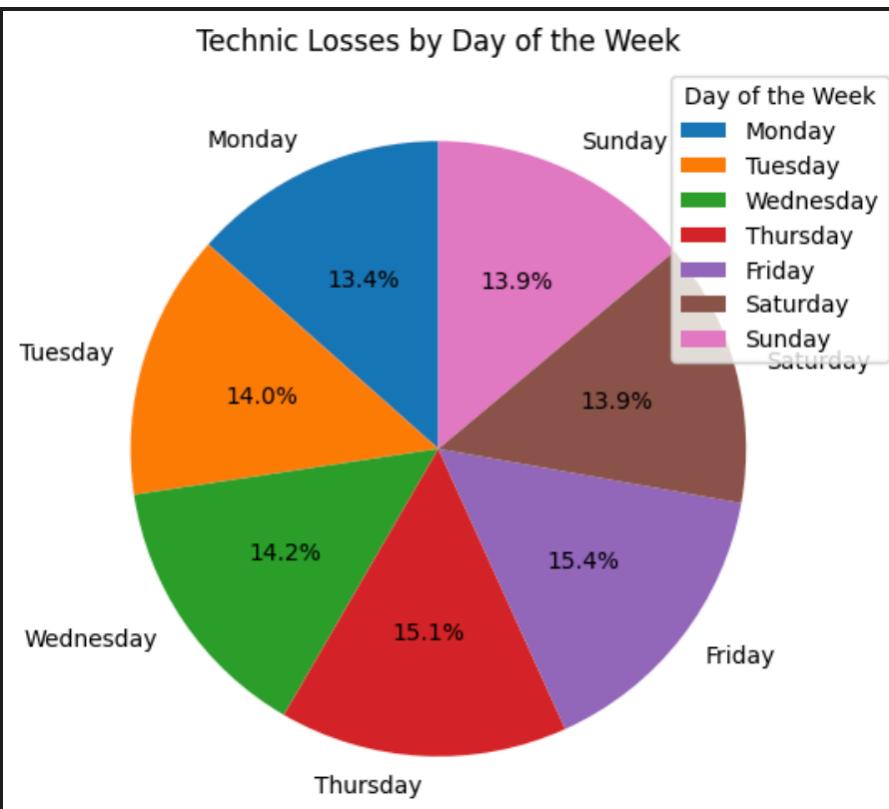
Znázornenie distribúcie strát techniky na základe dňa v týždni koláčovým grafom:

```
df['week_number'] = df.index.to_period('W').week
day_counts = df['day_of_week'].value_counts().sort_index()

# Print the counts
print("uistenie ze vsetkych dni mame rovnako")
print(day_counts)

# Nacrt kolacovo grafu:
# Calculate total losses for each day of the week
total_losses_by_day = df.groupby('day_of_week')['daily_only_technic_losses'].sum()
|_
plt.figure(figsize=(10, 6))
plt.pie(total_losses_by_day, labels=days_of_week, autopct='%1.1f%%', startangle=90)
plt.title('Technic Losses by Day of the Week')
plt.legend(title='Day of the Week', loc='upper right', bbox_to_anchor=(1.1, 1))
plt.show()
✓ 0.0s

uistenie ze vsetkych dni mame rovnako
day_of_week
0    127
1    127
2    127
3    127
4    127
5    127
6    127
Name: count, dtype: int64
```

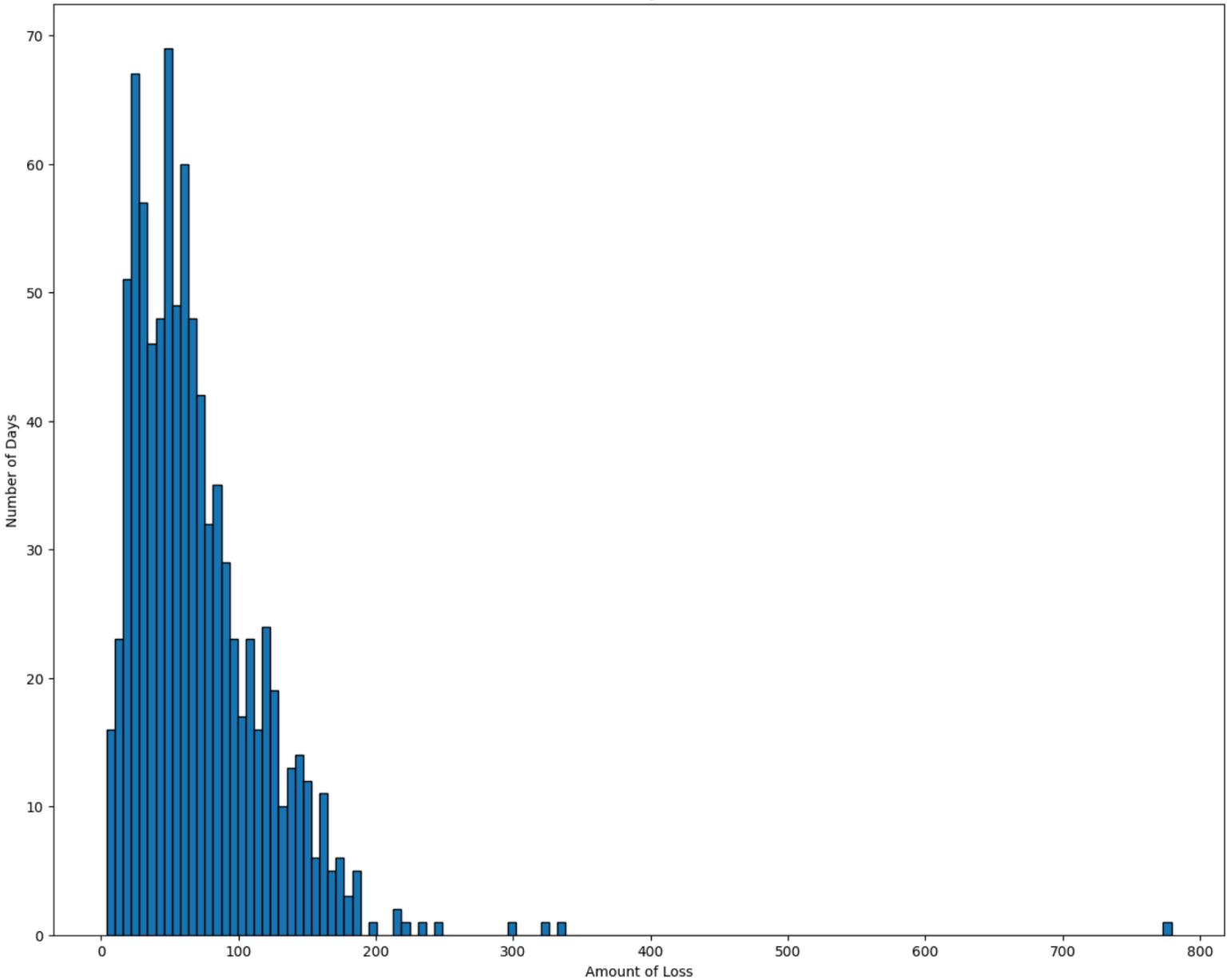


Načrtnem si ešte histogram denných strát, kde vidím že najčastejšie vyskytujúce sa hodnoty sú niekde okolo 70

```
plt.figure(figsize=(15, 12))
plt.hist(df['daily_only_technic_losses'], bins=130, edgecolor='black')
plt.title('Distribution of Daily Technic Losses')
plt.xlabel('Amount of Loss')
plt.ylabel('Number of Days')
plt.show()
# priponima nejake deformovane normalne rozdelenie
# data su relativne normalne rozdelene
```

✓ 0.1s

Distribution of Daily Technic Losses



Ako posledné si zobrazím ešte distribúciu strát techniky pre jednotlivé dni týždňa

```
# Create a subplot for each day of the week with a 2x4 grid
fig, axs = plt.subplots(nrows=2, ncols=4, figsize=(20, 10))
axs = axs.ravel() # Flatten the array

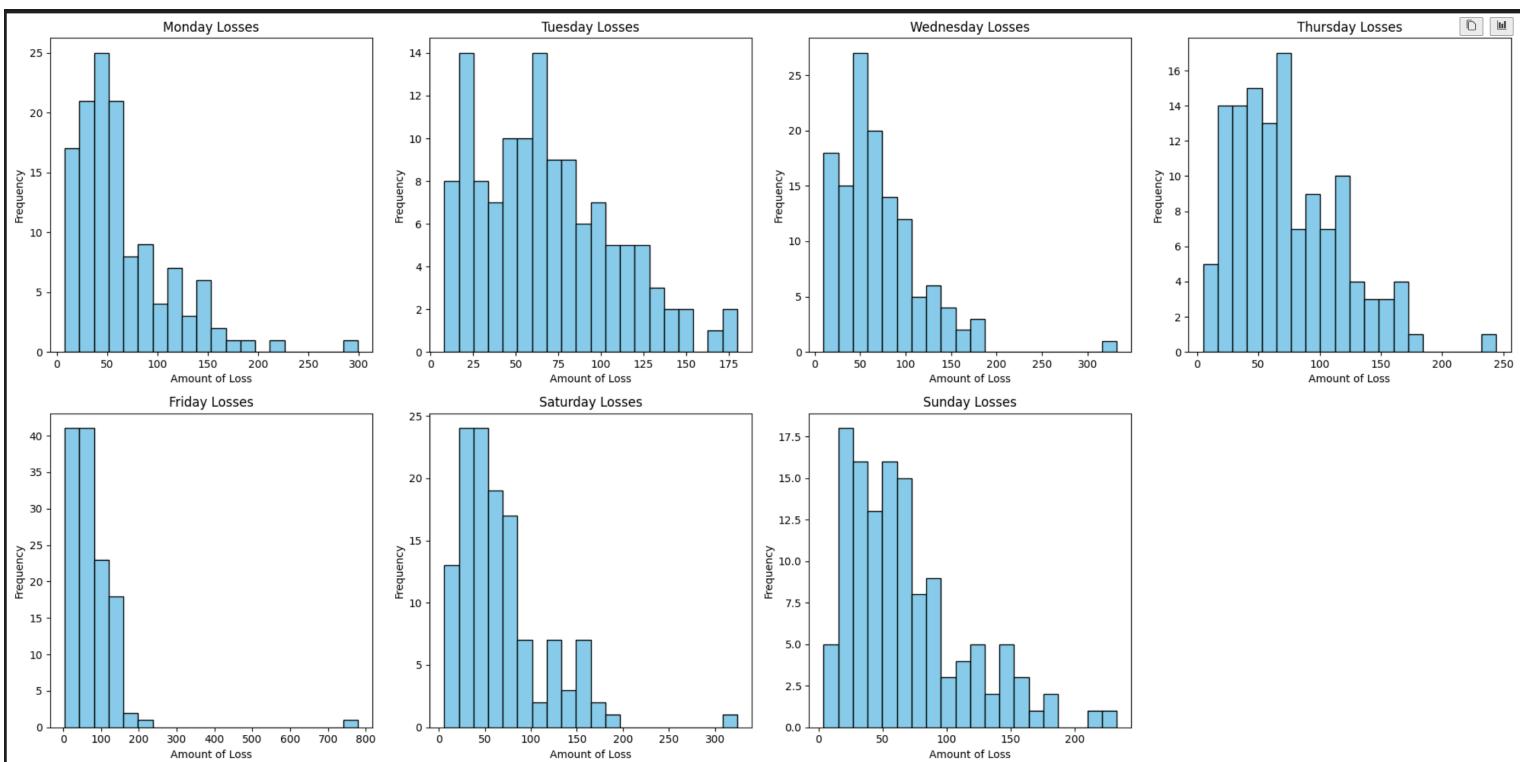
# limits for axis
x_min, x_max = 0, 1000
y_min, y_max = 0, 500

#print(max(df['daily_only_technic_losses']) + 1)

for i, day in enumerate(days_of_week):
    day_losses = df[df['day_of_week'] == i]['daily_only_technic_losses']
    # Plot histogram for the day
    axs[i].hist(day_losses, bins=20, edgecolor='black', color='skyblue') # Adjust bins as needed
    axs[i].set_title(f'{day} Losses')
    axs[i].set_xlabel('Amount of Loss')
    axs[i].set_ylabel('Frequency')

# removing empty frame
if len(days_of_week) < len(axs):
    axs[-1].remove() # This removes the last subplot if it's not used
plt.tight_layout()
plt.show()

# data su relativne normalne rozdelene
✓ 0.6s
```



Z histogramov vidno že v piatok, kde priemerné straty boli najvyššie sa často vyskytujú hodnoty medzi 0 a 150 raz sa ale stalo že v piatok boli straty skoro 800, preto priemer vyšiel vyšší. Preto si myslím že hypotéza H0 by sa mohla potvrdiť. Pre nás prípad nieje predpoklad approximácie velkosti strát normálnym rozdelením potrebný.

Pri testovaní chceme vlastne zistiť či sa straty z konkrétnych dní výrazne líšia od priemernej straty. t.j. čísla

$$\mu_d = \frac{\sum_{s \in d} \text{veľkosť strát v deň } s}{|d|},$$

kde d su všetky dni z ktorých máme záznam o stratách. Túto odlišnosť a jej signifikanciu určím pomocou jednovýberového t-testu. Použijem na to funkciu `ttest_1samp` z knižnice `scipy.stats` a ako parametre jej predám denné straty techniky a priemernú dennú stratu techniky. Za hladinu významnosti si zvolím typicky $\alpha = 0.05$.

```
# T-test
# spocita mi_r a urci t-hodnoty a p-hodnoty
# one sample t-test
# two-tailed alternative='two-sided' is the default value, checking for a two-sided alternative hypothesis
average_daily_loss = df['daily_only_technic_losses'].mean()
print(average_daily_loss)

for day in days_of_week:
    day_losses = df[df['day_of_week'] == days_of_week.index(day)]['daily_only_technic_losses']
    # Perform one-sample t-test against the overall mean loss
    t_statistic, p_value = ttest_1samp(day_losses, average_daily_loss)
    print(f"{day}: T-statistic = {t_statistic}, P-value = {p_value}")

] 0.0s
69.90213723284589
Monday: T-statistic = -0.9795458315490803, P-value = 0.3291874017438906
Tuesday: T-statistic = -0.46440215832118237, P-value = 0.6431610759445054
Wednesday: T-statistic = -0.04969671531621782, P-value = 0.9604427670342814
Thursday: T-statistic = 1.06016390717508, P-value = 0.2910980459223831
Friday: T-statistic = 0.8075698689149147, P-value = 0.4208601655627199
Saturday: T-statistic = -0.4087804514031083, P-value = 0.6833947309862864
Sunday: T-statistic = -0.45816042840044785, P-value = 0.647626376379667
```

Ako vidieť z výsledkov, p-hodnoty pre všetky dni sú rádovo vyššie ako 0.05 a preto túto hypotézu nemôžeme zamietnúť.

Záver, Diskusia

Pri overovaní hypotézy o závislosti dňa v týždni sme dospeli k záveru že môžeme tvrdiť: veľkosť strát techniky nie je štatisticky významne závislá od dňa v týždni. A pretože sú p-hodnoty pre všetky dni výrazne nad hladinou významnosti 0.05 podporuje to našu hypotézu H₀. Závislosti voči dňom sú relatívne slabé a bolo by zaujímavé skúmať tieto straty voči hodiny dňa, kde by sa mohli ukázať zaujímavé pozorovania, taktiež by sa dala spraviť aj analýza lokácií v ktorých sú tieto straty najvyššie.

Záver, Diskusia

Pri overovaní hypotézy o závislosti od dní v týždni sme dospeli k záveru že môžeme tvrdiť, veľkosť strát techniky nie je štatisticky významne závislá od dňa v týždni. A pretože sú p-hodnoty pre všetky dni výrazne nad hladinou významnosti 0.05 podporuje to našu hypotézu. Závislosti voči dňom sú relatívne slabé a bolo by zaujímavé skúmať tieto straty od času dňa kde by sa mohli ukázať zaujímavé pozorovania, taktiež by sa dala spraviť aj analýza lokácií v ktorých sú tieto straty navyššie.