report

September 12, 2023

1 Laboratory Work 1 Report

1.1 Using Data Exploration and Visualization with Python

1.1.1 Objective

To gain basic skills in using Python for data exploration and visualization, focusing on analyzing the armored personnel carrier (APC) losses during a specified period.

1.1.2 Methodology

In this laboratory work, we utilized the data from the 'russia_losses_equipment.csv' file available on Kaggle. The primary libraries employed for data analysis and visualization were NumPy and Matplotlib. The analysis focused on extracting, manipulating, and visualizing data about APC losses.

1.1.3 Implementation

Task 1: Setting up the Environment "'python import csv import numpy as np import matplotlib.pyplot as plt

Task 2: Reading Data Here we defined a function to read a specific column from a CSV file and convert it into a numpy array.

[16]:
```python
def read_column_from_csv(filename, column_name):
    with open(filename, 'r') as file:
        csv_reader = csv.reader(file)
        data = list(csv_reader)
        column_index = data[0].index(column_name)

        return np.array([row[column_index] for row in data[1:]])

# Testing the function
apc_data = read_column_from_csv('russia_losses_equipment.csv', 'APC')
print("The first 5 data points for APC losses are:", apc_data[:5])  #␣
 ↪Displaying the first 5 data points
```

```
The first 5 data points for APC losses are: ['516' '706' '706' '816' '846']
```

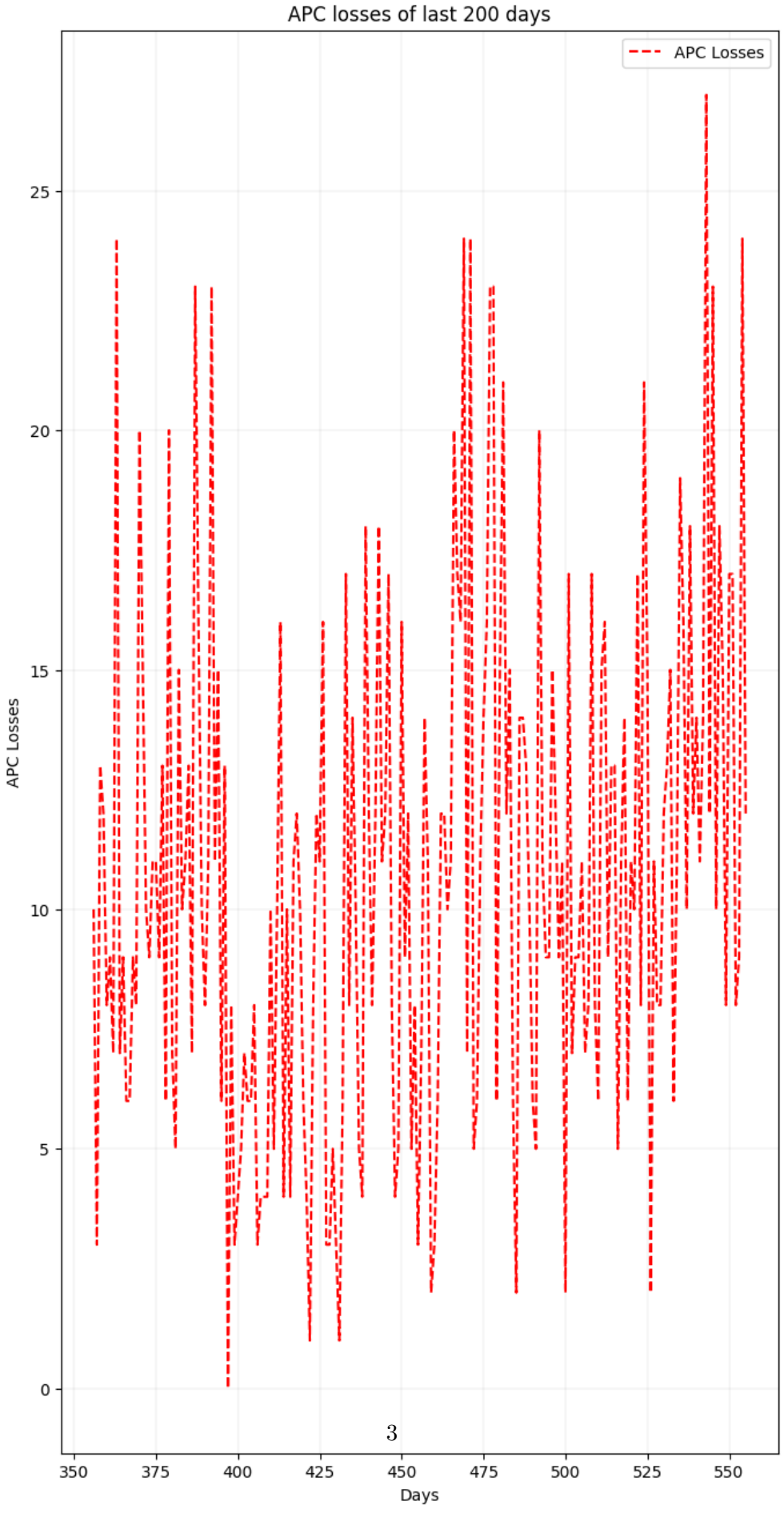
Task 3: Data Exploration In this task, we calculate the daily losses, the five greatest daily losses, the total summer losses, and the mean losses between the 100th and 500th days.

[17]:
```
daily_losses = np.diff(apc_data.astype(int), prepend=0)
print("The first 5 data points for APC losses are:", apc_data[:5])  #␣
 ↪Displaying the first 5 daily losses

greatest_losses = np.sort(daily_losses)[-5:]
print("The 5 days with the greatest APC losses are:", greatest_losses)

# Extracting and converting date data
dates = read_column_from_csv('russia_losses_equipment.csv', 'date')
dates_as_datetime = np.array([np.datetime64(date) for date in dates])

# Calculating summer losses
start_date = np.datetime64('2023-06-01')
end_date = np.datetime64('2023-08-31')
summer_losses = daily_losses[(dates_as_datetime >= start_date) &␣
 ↪(dates_as_datetime <= end_date)].sum()
print("The total APC losses for the summer of 2023 are:", summer_losses)

# Calculating mean losses between day 100 and 500
mean_losses = daily_losses[100:500].mean()
print("The mean APC losses between day 100 and day 500 are:", mean_losses)
```

```
The first 5 data points for APC losses are: ['516' '706' '706' '816' '846']
The 5 days with the greatest APC losses are: [ 60  96 110 190 516]
The total APC losses for the summer of 2023 are: 1132
The mean APC losses between day 100 and day 500 are: 11.4625
```

Task 4: Visualization In this task, we visualize the APC losses for the last 200 days.

[14]:
```
last_200_losses = daily_losses[-200:]
days = np.arange(len(daily_losses) - 200, len(daily_losses))

plt.figure(figsize=(8, 16), dpi=100)
plt.plot(days, last_200_losses, linestyle='--', color='red', label='APC Losses')
plt.title("APC losses of last 200 days")
plt.xlabel("Days")
plt.ylabel("APC Losses")
plt.grid(color='gray', linestyle='-', linewidth=0.1)
plt.legend()
plt.savefig("apc_losses_last_200_days.png", dpi=100)
plt.show()
```


APC losses of last 200 days
APC Losses
APC Losses
Days
0
5
10
15
20
25
350
375
400
425
450
475
500
525
550

1.1.4 Conclusion

In this laboratory work, we successfully performed data exploration and visualization to analyze APC losses data. We calculated daily losses, identified the days with the highest losses, and analyzed the data for specific periods, providing insights into the patterns and trends in APC losses over time.

1.1.5 References

1. Data source: Kaggle Dataset
2. NumPy Documentation
3. Matplotlib Documentation