**Helicopter Prison Escape Analysis**

This project analyzes historical helicopter prison escapes by collecting, cleaning, and visualizing data to uncover patterns and trends. The notebook combines **web scraping**, **data wrangling**, and **exploratory data analysis (EDA)** techniques to provide insights into when, where, and how these escapes occurred.

**Data Collection and Preparation**

The dataset is scraped from the **Wikipedia page "List of helicopter prison escapes"** using requestsandBeautifulSoup.  
Key steps include:

* **Scraping:** Extracting tabular data (date, prison name, country, escapees, outcome).
* **Cleaning:**
  + Removing duplicates and empty rows.
  + Standardizing date formats using datetime.
  + Normalizing country and prison names to handle inconsistencies.
  + Cleaning text using unicodedata and re to remove special characters.
* **Structuring:** Converting the cleaned data into a **pandas DataFrame** for easy analysis and visualization.

This stage ensures that the dataset is reliable enough for pattern discovery, while also acknowledging potential biases such as missing incidents or incomplete records.

**Exploratory Data Analysis**

The analysis focuses on discovering patterns in helicopter prison escapes:

* **Temporal Analysis:**
  + Number of escapes per year and per decade to identify peaks and declines.
  + Distribution across months and weekdays to see if certain times are more common.
* **Geographical Analysis:**
  + Counting incidents per country to find high-risk regions.
  + Comparing regions or specific prisons with repeat incidents.
* **Notable Escapees:**
  + Identifying individuals with multiple attempts (e.g., Vassilis Paleokostas, Pascal Payet).
  + Highlighting their repeated methods and escape success rates.
* **Success vs. Failure:**
  + Assessing how many attempts were successful and any visible improvement in security over time.

Visualizations (using matplotlib) include bar charts and time-series plots to clearly display these findings.

**Findings**

The notebook reveals fascinating insights:

* **Rare but Spectacular:** Helicopter escapes are relatively uncommon but show clustered activity in certain decades (e.g., 1980s–2000s).
* **Geographical Hotspots:** A handful of countries account for most of the incidents, suggesting systemic security weaknesses in those regions.
* **Repeat Offenders:** Some escapees attempted multiple escapes, often successfully, showing persistence and planning sophistication.
* **Timing Patterns:** Some correlation between escape attempts and specific times of year, hinting at operational opportunities (weather, visibility, etc.).

**Conclusion**

This project highlights the **importance of prison security evolution** and how data can reveal systemic weaknesses.  
The insights could help policymakers strengthen prison protocols and anticipate future risks.

By combining **data scraping, cleaning, and visualization**, this notebook provides a strong foundation for further research into criminal behavior and institutional vulnerabilities.

**Note:** The results should be interpreted cautiously due to possible underreporting or missing data. Further enrichment with official prison records could improve reliability.