Passwords Around the World: Key Insights from Worldwide Datasets

By: Teiba Al-Hami — Tech for Jobs – Technical Development and Job Matching Fellowship

Highlights

- Global Prevalence of Weak and Predictable Passwords
- Widespread Lack of Complexity in Passwords
- Significant Security Vulnerabilities in High-Value Sectors
- Short Passwords Are Extremely Susceptible to Brute-Force Attacks

Background

Passwords are the first line of defense in securing online accounts, yet many users create weak, predictable, or reused passwords that are vulnerable to brute-force, dictionary attacks and more. This vulnerability contributes to data breaches, costing businesses billions annually in lost revenue, regulatory penalties, and damaged customer trust.

Data

This analysis leverages three datasets: the RockYou Password Dataset (sample) from the RockYou breach dataset (2009), commonly included in Kali Linux as rockyou.txt, contains nearly one million leaked passwords. The Top 200 Passwords by Country (2020) and the Wealthiest Companies with the Weakest Passwords both are from NordPass. The analysis identifies common password vulnerabilities and highlights security gaps across different regions and industries.

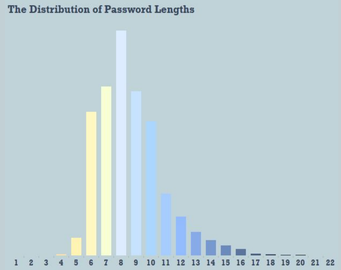
In Data Curation process involved sourcing and preparing multiple datasets for analysis. Each dataset was verified for completeness, consistency, and uniqueness. Specific attention was given to identifying passwords starting with characters like - and =, which could be misinterpreted by Excel, ensuring they were properly handled. Data was cleaned, with formatting issues addressed and missing values handled, while maintaining data integrity. No transformations were made to preserve the authenticity of the original datasets.

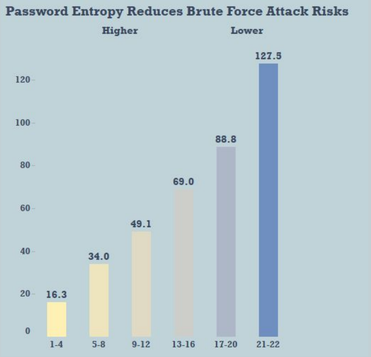
Tools

For Exploratory Data Analysis (EDA), Python was the primary tool, enabling the analysis of password length distributions, the identification of the most common passwords, and the exploration of correlations between password length and time to crack. For visualizations and further interactive analysis, Tableau was used to create an engaging dashboard, highlighting key patterns and trends across different datasets, such as password vulnerabilities and regional trends.

Methodology

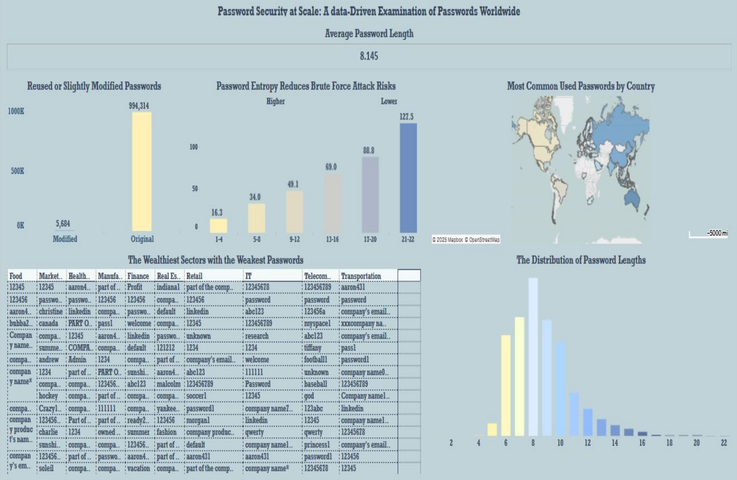
Key analyses in this project focus on password length distribution, the most commonly used passwords, and how password strength impacts security. One of the most critical findings is the relationship between password length and time to crack, demonstrating that shorter passwords are significantly more vulnerable to brute-force attacks





Additionally, a bar chart illustrates how increasing password entropy significantly reduces vulnerability to brute-force attacks. The visualization clearly shows that most passwords in the dataset fall into low-entropy ranges, making them highly susceptible to being cracked. Only a small portion of passwords reach higher entropy levels, indicating a much lower risk and highlighting the sharp contrast between commonly used weak passwords and those that offer stronger protection.

Dashboard



You can access the dashboard by clicking here.

Overall Vulnerabilities: A cross the board, passwords are excessively short, lack complexity, and follow predictable patterns (such as 123456). The majority of passwords are vulnerable to brute-force attacks and more. By addressing these weaknesses, organizations can protect themselves from costly breaches, help build best practices for password security, and contribute to a more secure digital environment globally.

Limitations

The analysis relies on a RockYou sample, limiting its global representation. The study focuses on plaintext passwords, but real-world security concerns often involve hashed or encrypted passwords. Some passwords in real-world breaches may be partially leaked, salted, or hashed, making them harder to analyze.