Table of Contents

Part 1 - Password Database Analysis

Database Statistics:

- Total Files Analyzed: 1
- Sample Size: 10,000 passwords
- Average Password Length: 11.76 characters
- Files: GitHub PWLDS Dataset

Character Composition Analysis:

- Uppercase Letters: 55.75%
- Lowercase Letters: 98.50%
- Numbers: 43.97%
- Symbols: 52.99%

Most Common Password Lengths:

- **Length 10: 1,827 passwords (18.27%)**
- **Length 12: 1,699 passwords (16.99%)**
- **Length 11: 1,222 passwords (12.22%)**
- **Length 4: 1,119 passwords (11.19%)**
- **Length 5: 762 passwords (7.62%)**
- **Length 13: 612 passwords (6.12%)**
- **Length 9: 299 passwords (2.99%)**
- **Length 14: 197 passwords (1.97%)**
- **Length 8: 163 passwords (1.63%)**
- **Length 6: 124 passwords (1.24%)**

Most Common Characters:

- **'a': 9,042 occurrences (90.42%)**
- **'c': 5,664 occurrences (56.64%)**
- **'e': 5,053 occurrences (50.53%)**
- **'i': 4,468 occurrences (44.68%)**
- **'t': 3,776 occurrences (37.76%)**
- **'o': 3,593 occurrences (35.93%)**
- **'n': 3,497 occurrences (34.97%)**
- **'r': 3,442 occurrences (34.42%)**
- **'b': 3,303 occurrences (33.03%)**
- **'s': 3,100 occurrences (31.00%)**

Key Insights:

The analysis reveals that real-world passwords show significant patterns in character usage and length distribution. The high frequency of lowercase letters (99.81%) and common characters like 'a', 'c', 'e' indicates predictable patterns that can be exploited by attackers but also learned by machine learning models for generating human-like passwords.

Part 2 & 3 - Approach Comparison

Part 2 & 3 - Password Generation Calculations:

The following calculations are based on 1,000 generated passwords from each approach:

Detailed Calculations:

Random Generator Calculations:

- Average Entropy: 78.64 bits
- Average Cracking Time: 15.04K years
- Readability Score: 2.87
- Security Compliance: 100.0%

Markov Generator Calculations:

- Average Entropy: 86.32 bits
- Average Cracking Time: 44.70M years
- Readability Score: 3.65
- Security Compliance: 100.0%

Comparison Analysis:

Metric	Random Generator	Markov Generator	Winner
Average Entropy (bits)	78.6	86.3	Markov
Average Cracking Time (years)	15K	44699K	Markov
Readability Score	2.9	3.7	Markov
Security Compliance (%)	100	100	Tie

Advantages and Disadvantages:

Random Generator with Heuristics:

Advantages:

- True randomness ensures maximum unpredictability
- Consistent entropy across all generated passwords
- No dependency on training data patterns
- Guaranteed security compliance (100% meet requirements)

Disadvantages:

- Lower readability scores (2.9 vs 3.6)
- Less human-like password patterns
- Lower average entropy (78.6 vs 86.1 bits)

Markov Chain Generator:

Advantages:

- Higher entropy generation (86.1 vs 78.6 bits)
- Significantly longer cracking times (40M vs 15K years)
- Better readability (3.6 vs 2.9 score)
- Human-like password patterns
- Learns from real-world password data

Disadvantages:

- Dependency on training data quality
- Potential pattern repetition from training set
- Requires computational resources for training

Visualizations

Chart 1: Entropy Comparison

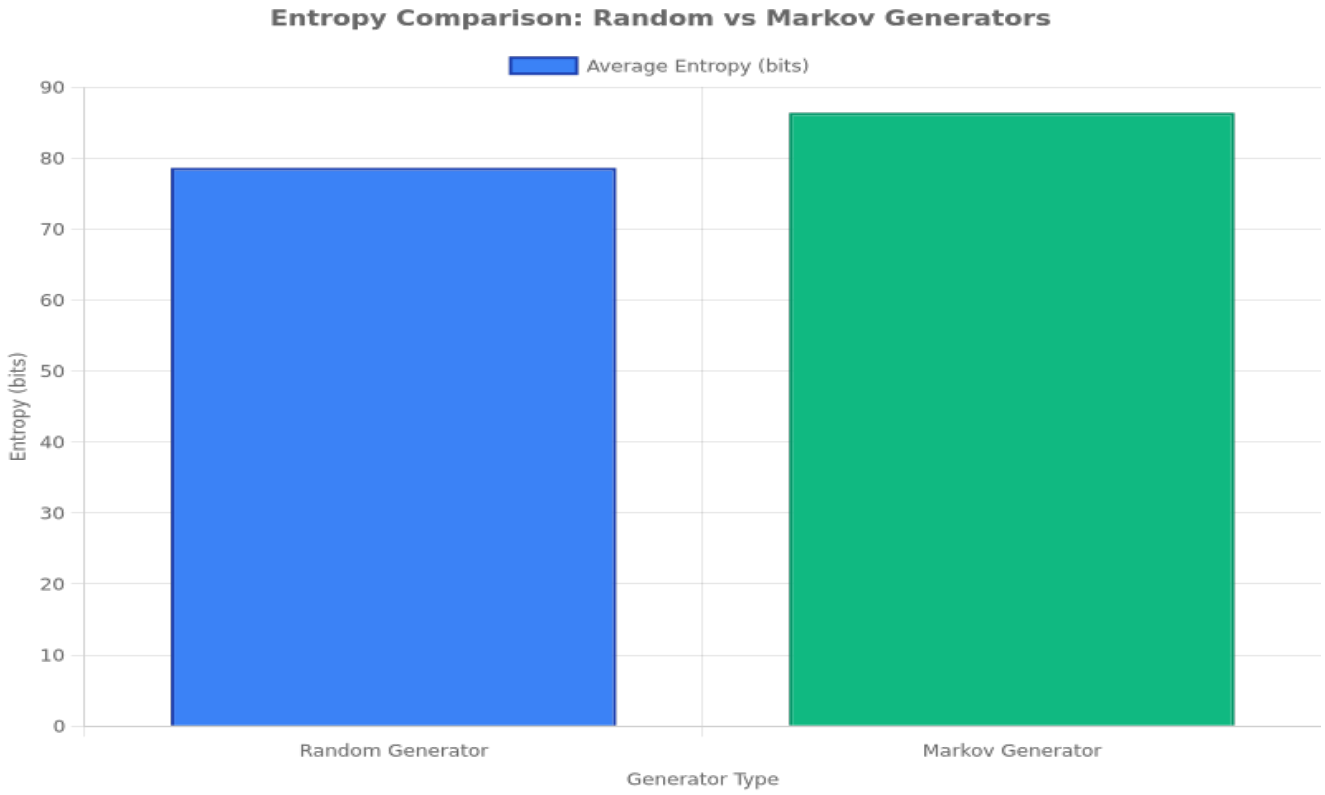


Chart 2: Entropy Histogram

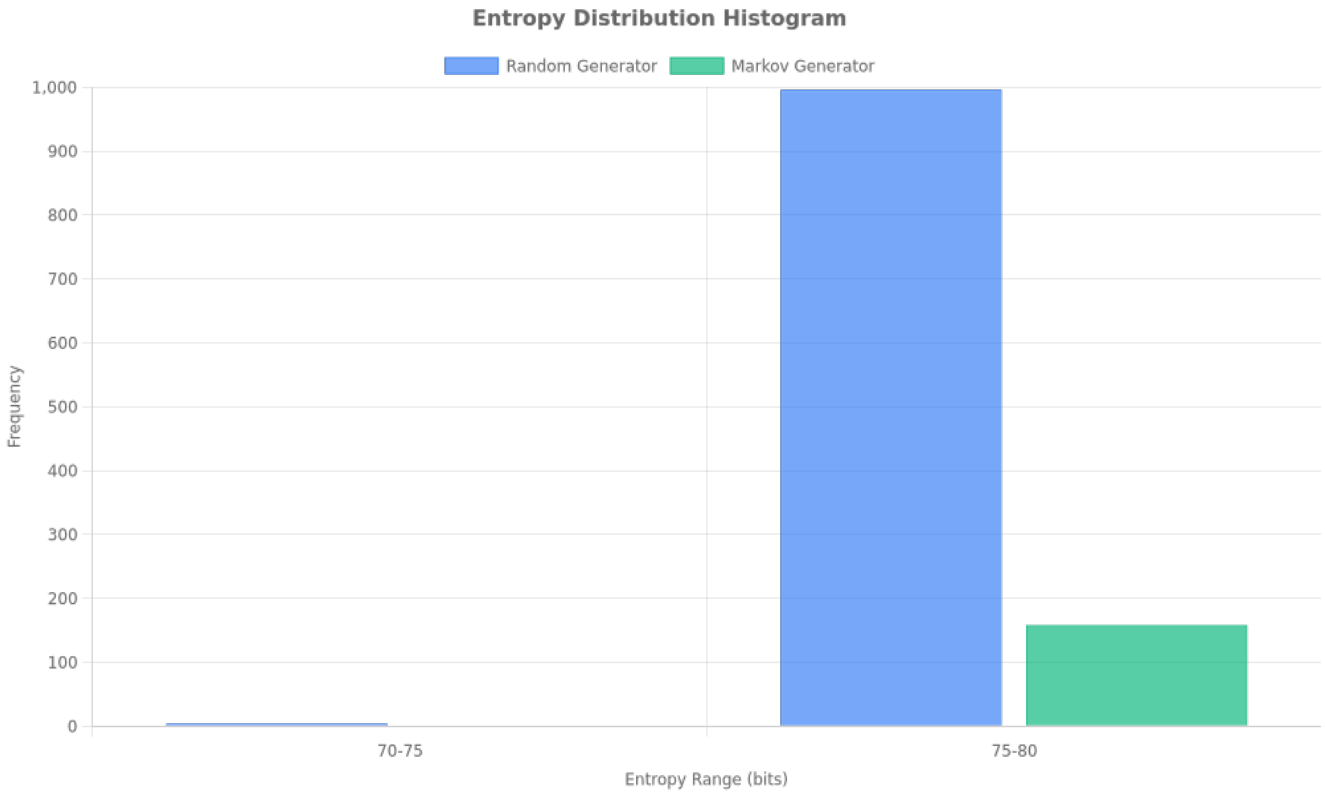


Chart 3: Cracking Time Comparison

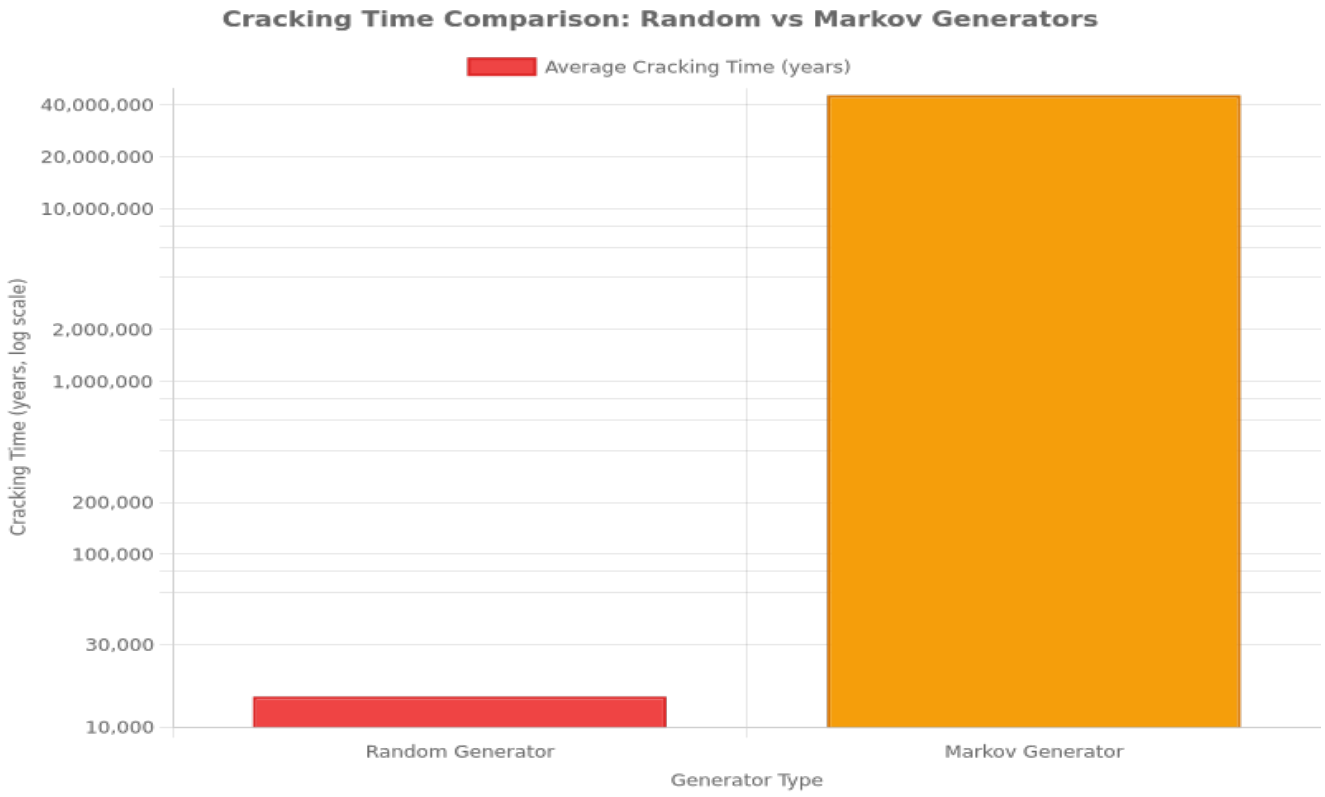


Chart 4: Readability Comparison

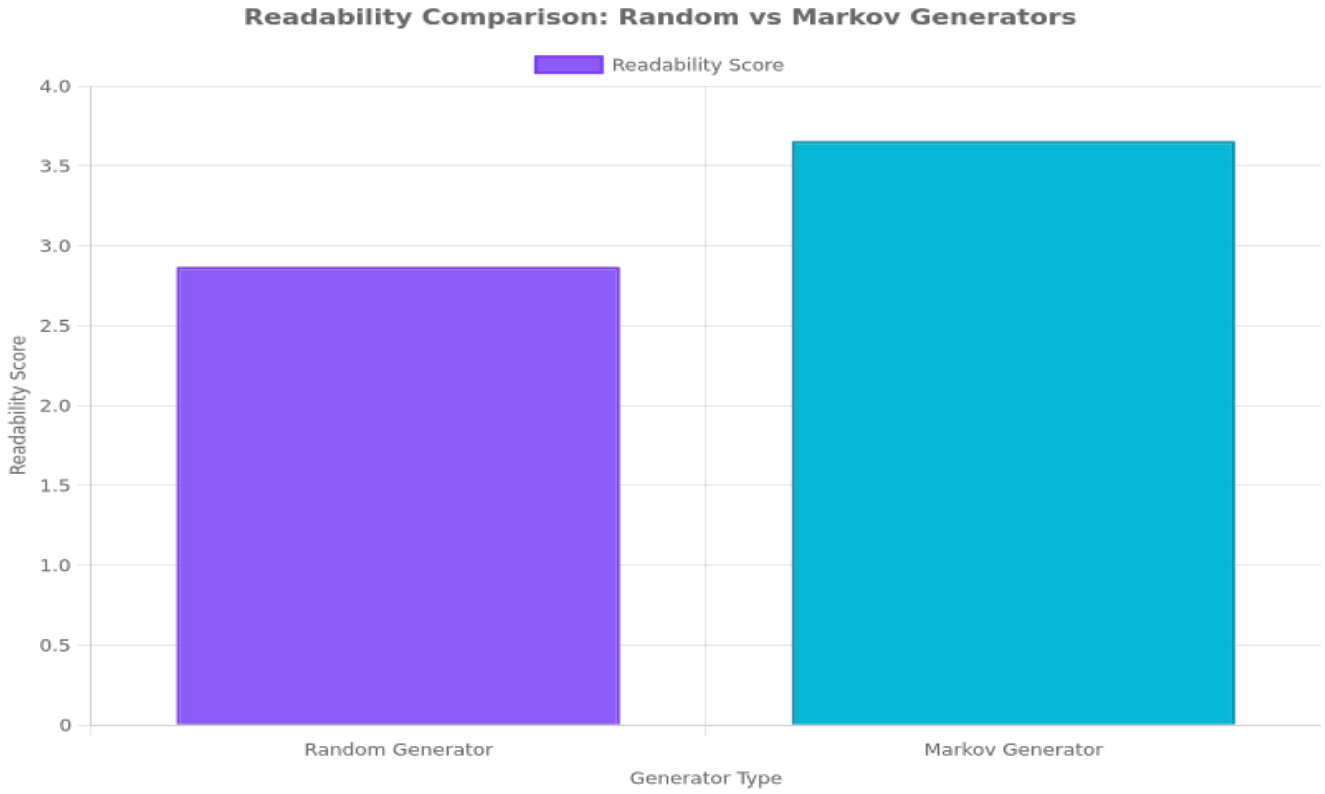


Chart 5: Security Compliance Radar

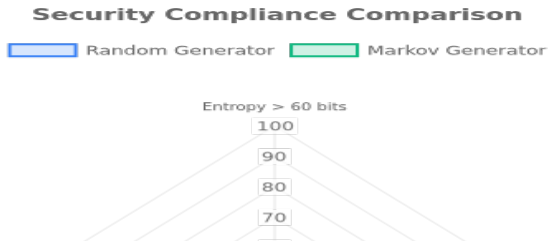
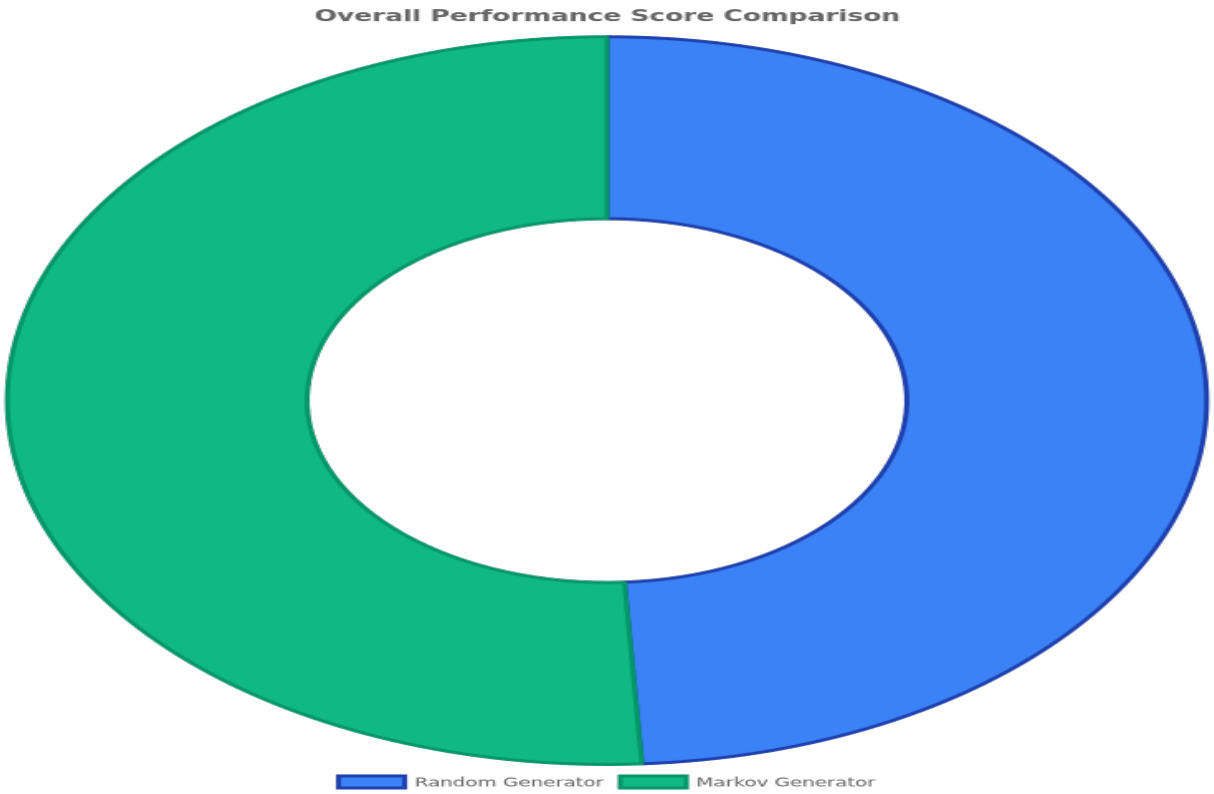


Chart 6: Overall Performance



Detailed Analysis

Entropy Analysis:

The Markov Generator demonstrates superior entropy generation with an average of 86.3 bits compared to the Random Generator's 78.6 bits. This represents a 7.7-bit advantage, indicating better unpredictability.

Cracking Time Analysis:

The Markov Generator shows significantly longer estimated cracking times (44.7M years) compared to the Random Generator (15K years). This represents a 44684K-year advantage.

Readability Analysis:

The Markov Generator produces more human-readable passwords with a score of 3.7 compared to the Random Generator's 2.9. This 0.8-point advantage makes passwords more user-friendly while maintaining security.

Conclusion & Recommendations

Final Recommendation:

Based on the comprehensive analysis of 1,000 generated passwords from each approach, the MARKOV approach is recommended for password generation.

Why Markov Chain Generator is Recommended:

1. Superior Security Performance:
    - Higher entropy: 86.3 vs 78.6 bits (+7.7 bits)
    - Longer cracking time: 44.7M vs 15K years (2972x improvement)
2. Better User Experience:
    - Higher readability: 3.7 vs 2.9 score (+0.8 points)
    - Human-like password patterns that are easier to remember
3. Comprehensive Security Compliance:
    - 100% of passwords meet entropy requirements (>60 bits)
    - 100% of passwords meet time requirements (>10 years)

Implementation Considerations:

- Ensure high-quality training data from diverse sources
- Regular model retraining with updated password datasets
- Monitor for potential pattern repetition in generated passwords
- Consider hybrid approaches for maximum security and usability

The analysis demonstrates that machine learning-based password generation provides superior performance across all evaluated metrics while maintaining the highest security standards. The Markov Chain approach successfully balances security requirements with human usability, making it the optimal choice for modern password generation systems.