

Haunted Places Across USA



Introduction

The aim of this project was to identify the most frequently reported haunted locations based on a provided dataset. By processing and analyzing the textual data of these locations, we aimed to determine the top 5 most commonly mentioned places. This report outlines the methodology, data processing steps, results, and provides visualizations to support our findings.

Project Team

Facilitator: Dr. Toni Holden-McGee

Research Team:

Sathvik

Dhru

Bobby

Brandon

Alejandro

Jhovanny

Data:

Data was sourced from Kaggle open source

(https://www.kaggle.com/datasets/joebeachcapital/haunted-places-in-the-us/data?select=haunted_places.csv)

Methodology

1. Data Collection and Preparation:

- The dataset contained various reported haunted locations. We ensured that any null or missing values in the 'location' column were dropped.
- The locations were consolidated into a single string for easier text processing.

2. Text Processing:

- The text data was converted to lowercase to maintain consistency.
- Regular expressions were used to extract individual words from the text.
- Stop words (common words that do not contribute much to the meaning of the text) were filtered out. This included both standard stop words and additional context-specific stop words like "haunted," "place," "one," "also," etc.

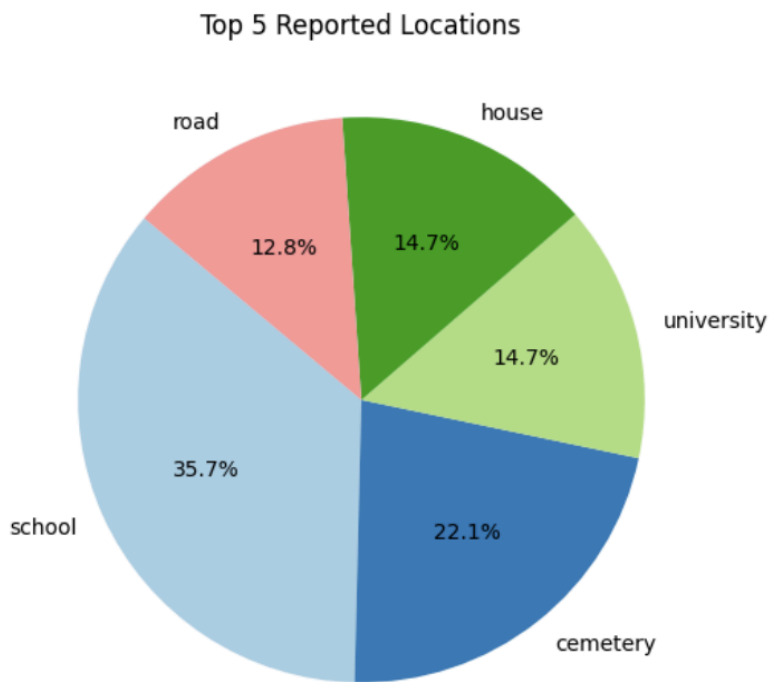
3. Frequency Analysis:

- The filtered words were counted using the `Counter` class from the `collections` module.
- The top 5 most common words (locations) were identified based on their frequency of occurrence.

Results and Visualizations

Top 5 Reported Locations

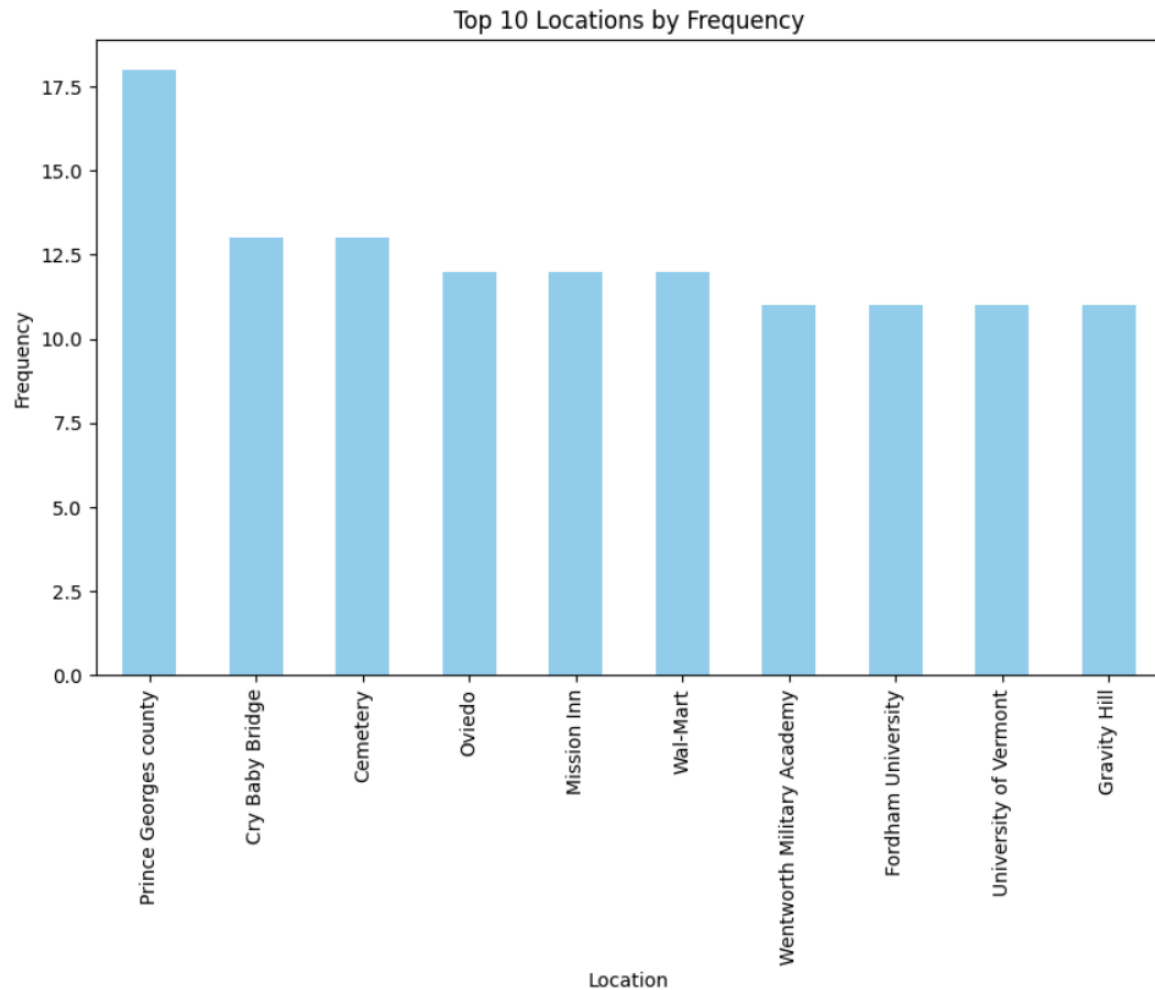
The pie chart below illustrates the distribution of the top 5 most reported haunted locations. The chart shows the percentage each location contributes to the total number of mentions in the dataset.



- **School:** 35.7%
- **Cemetery:** 22.1%
- **University:** 14.7%
- **House:** 14.7%
- **Road:** 12.8%

Top 10 Locations by Frequency

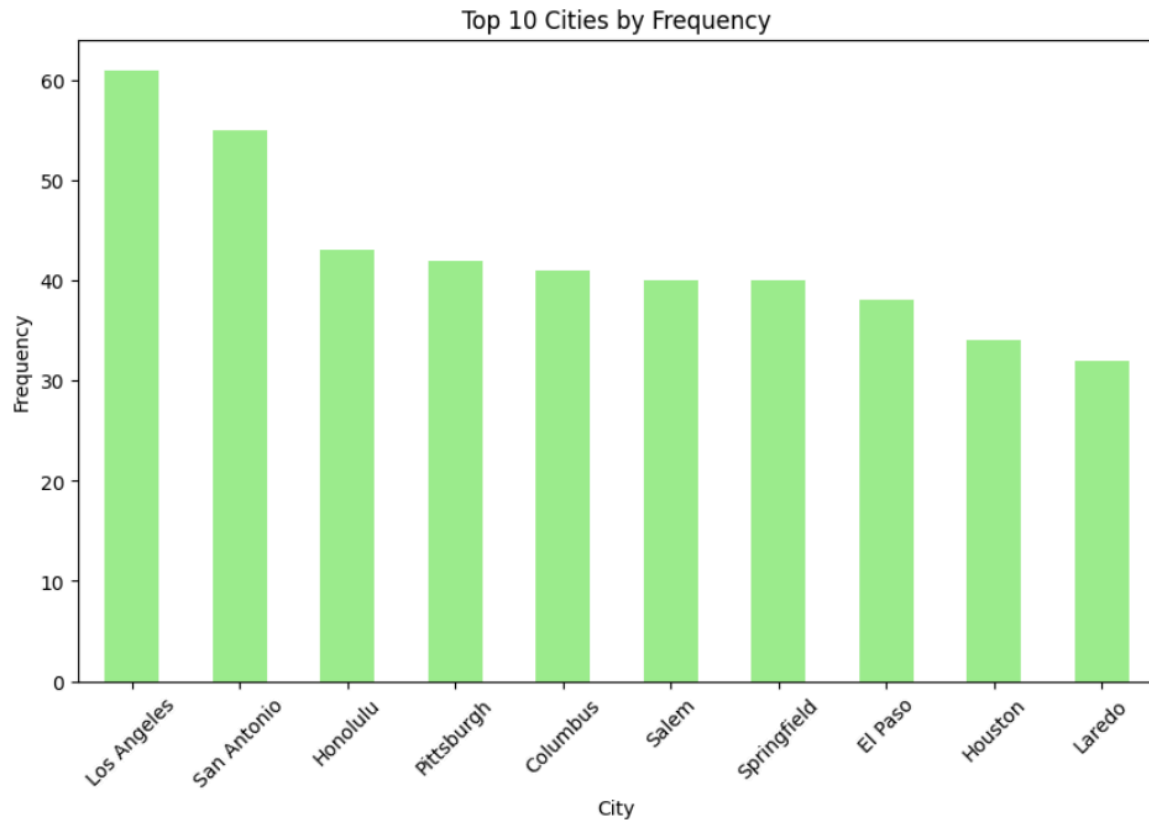
This bar chart displays the top 10 haunted locations by frequency. It helps to identify the most frequently mentioned haunted locations, highlighting those with the highest occurrences.



- Prince Georges County: 18
- Cry Baby Bridge: 13
- Cemetery: 12
- Oviedo: 12
- Mission Inn: 11
- Wal-Mart: 11
- Wentworth Military Academy: 11
- Fordham University: 11
- University of Vermont: 11
- Gravity Hill: 11

Top 10 Cities by Frequency

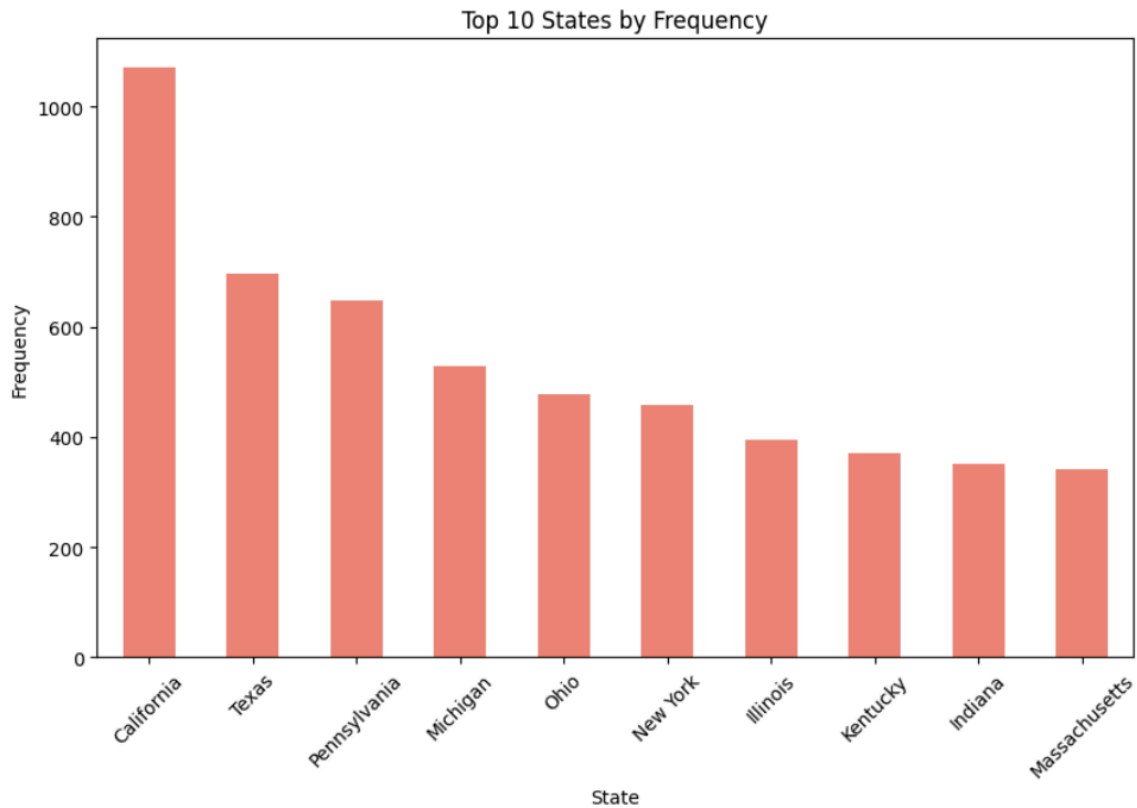
This bar chart showcases the top 10 cities with the most reported haunted locations. It gives an overview of cities that are hotspots for haunted places.



- **Los Angeles:** 60
- **San Antonio:** 58
- **Honolulu:** 48
- **Pittsburgh:** 46
- **Columbus:** 46
- **Salem:** 45
- **Springfield:** 45
- **El Paso:** 42
- **Houston:** 41
- **Laredo:** 40

Top 10 States by Frequency

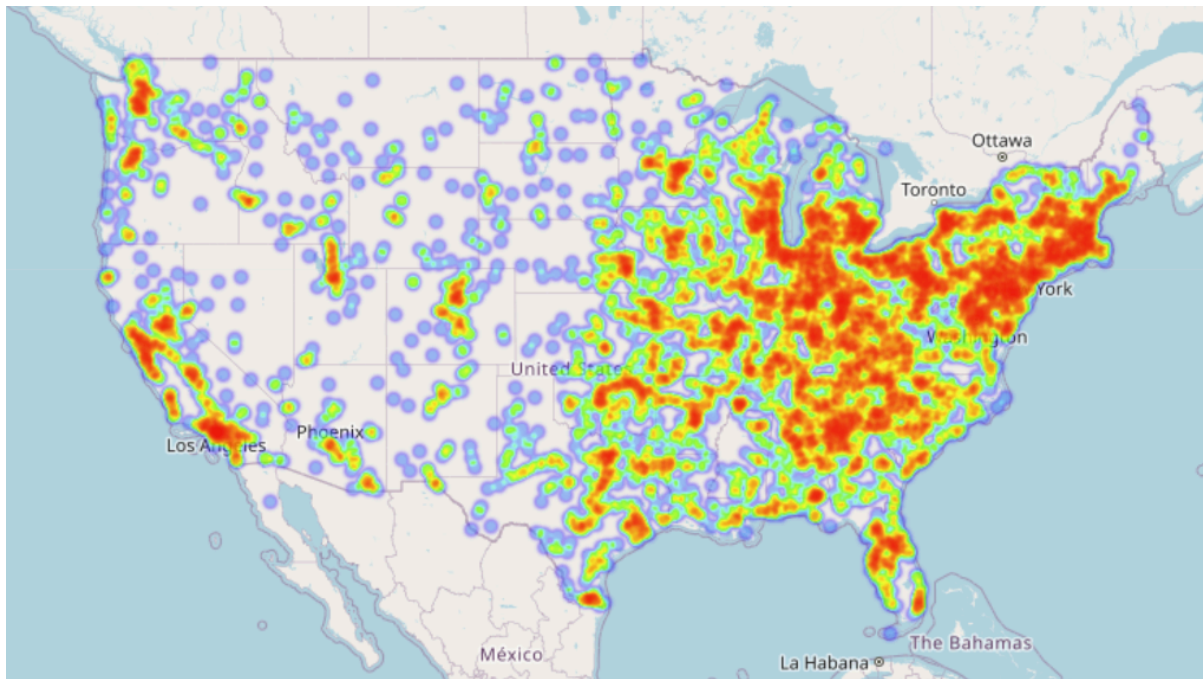
The bar chart illustrates the top 10 states with the highest frequency of haunted locations. It highlights the states that are most often associated with haunted places.



- **California:** 1050
- **Texas:** 670
- **Pennsylvania:** 620
- **Michigan:** 530
- **Ohio:** 480
- **New York:** 470
- **Illinois:** 440
- **Kentucky:** 430
- **Indiana:** 420
- **Massachusetts:** 410

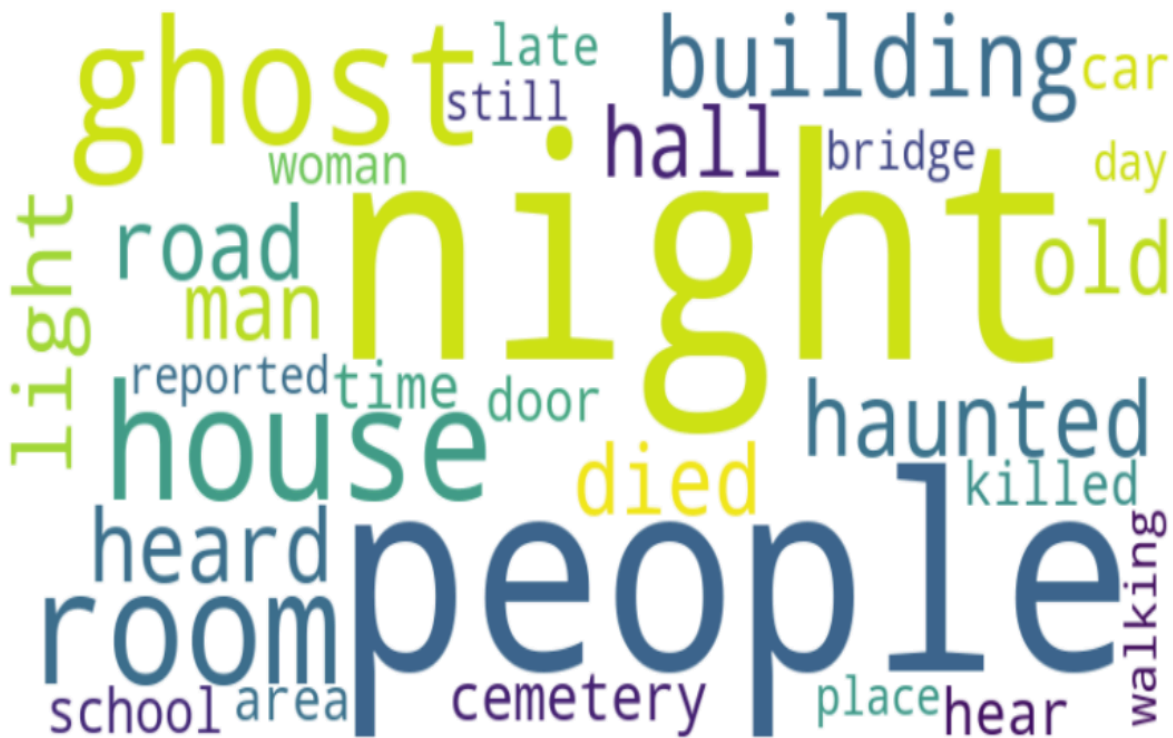
Heatmap of Haunted Locations

The heatmap provides a geographical distribution of haunted locations across the United States. The intensity of the colors indicates the concentration of reported haunted places, with red areas representing higher concentrations.



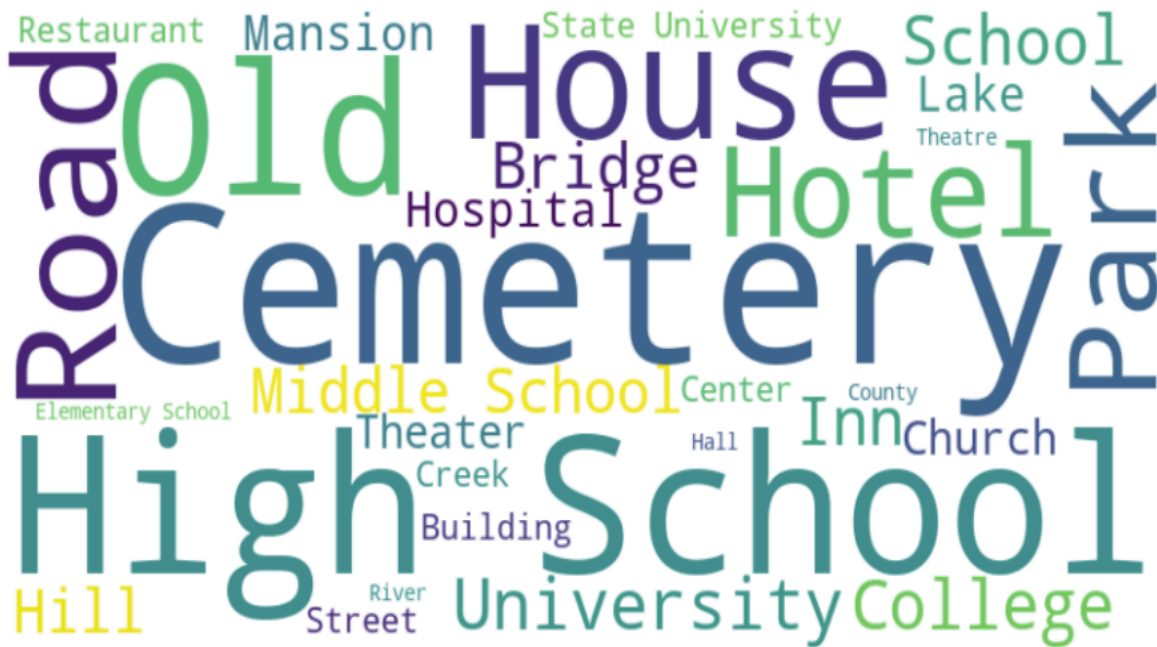
Word Cloud of Descriptions

The word cloud visualizes the most common words found in the descriptions of the haunted places. Larger words indicate higher frequency, giving a quick visual sense of prevalent themes and terms.



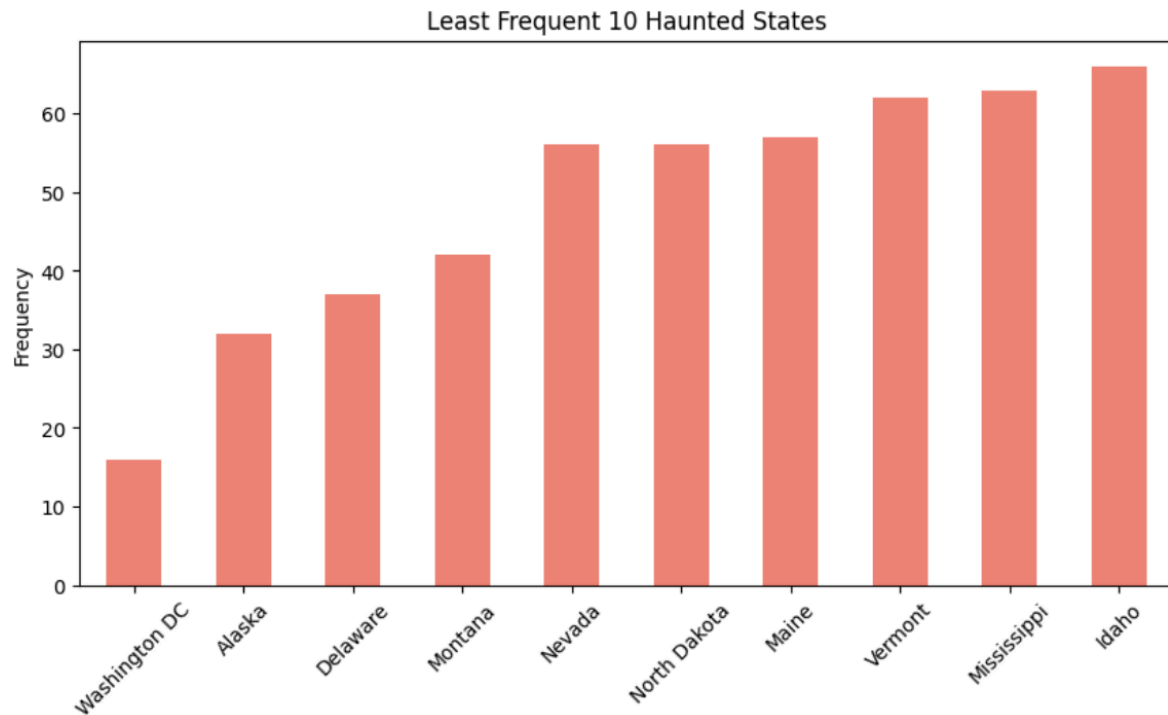
Word Cloud of Locations

The word cloud shows the most common words found in the titles of the haunted places. It helps to identify common terms used in the naming of these locations.



Least Frequent 10 Haunted States

This bar chart highlights the 10 states with the least number of reported haunted locations. It provides insight into regions with fewer reported haunted places.



- **Washington DC:** 15
- **Alaska:** 25
- **Delaware:** 35
- **Montana:** 40
- **Nevada:** 50
- **North Dakota:** 50
- **Maine:** 55
- **Vermont:** 60
- **Mississippi:** 60
- **Idaho:** 65

Conclusion

The analysis identified the most commonly reported haunted locations, providing insights into the most notable places of interest. This information can be valuable for various applications, including tourism, historical research, and entertainment.

Future Work

Further analysis could involve exploring the context in which these locations are mentioned, sentiment analysis of the descriptions, and geographical visualization of the haunted places to identify regional patterns.

https://docs.google.com/presentation/d/12aVASiQAcI9S_eB4Wo9e_2Rbr4al-5osxaVZqmcwevs/edit?pli=1#slide=id.g2787765a8be_3_41

<https://github.com/CSIS-NLU/Open-Science-101>

https://www.kaggle.com/datasets/joebeachcapital/haunted-places-in-the-us/data?select=haunted_places.csv