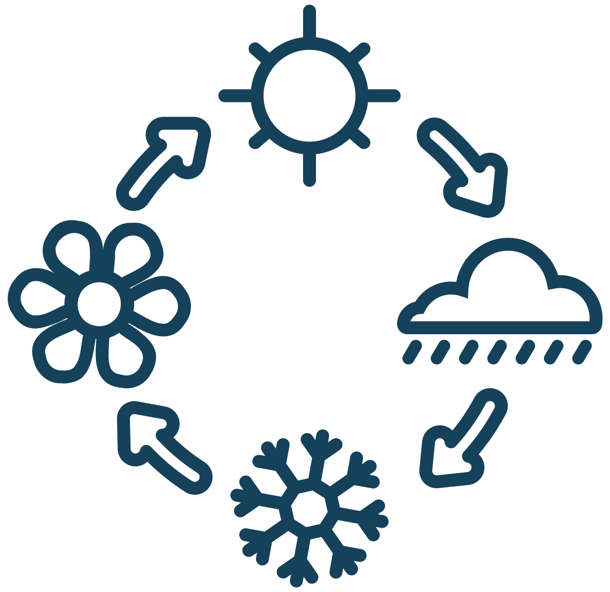
Dynamic Rhythms
Power Outages Forecasting

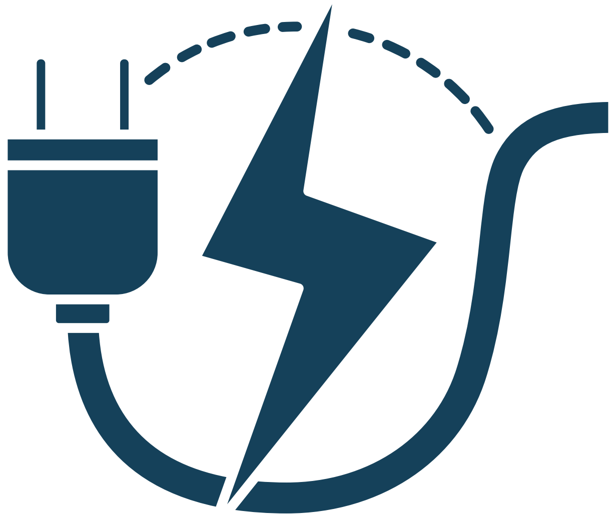
Authors:
Aleksandra Kwiatkowska
Małgorzata Mokwa
Bogumiła Okrojek

Introduction

Power outages caused by extreme weather events pose a significant risk to infrastructure, public safety, and economic stability. In the ***Dynamics Rhythms*** challenge, participants were invited to develop machine learning models capable of predicting such outages across the United States. Leveraging historical weather data and records of past power failures, the aim was to build a system that can reliably forecast the impact of severe weather on the electrical grid.

Datasets Description





Storm Events Dataset

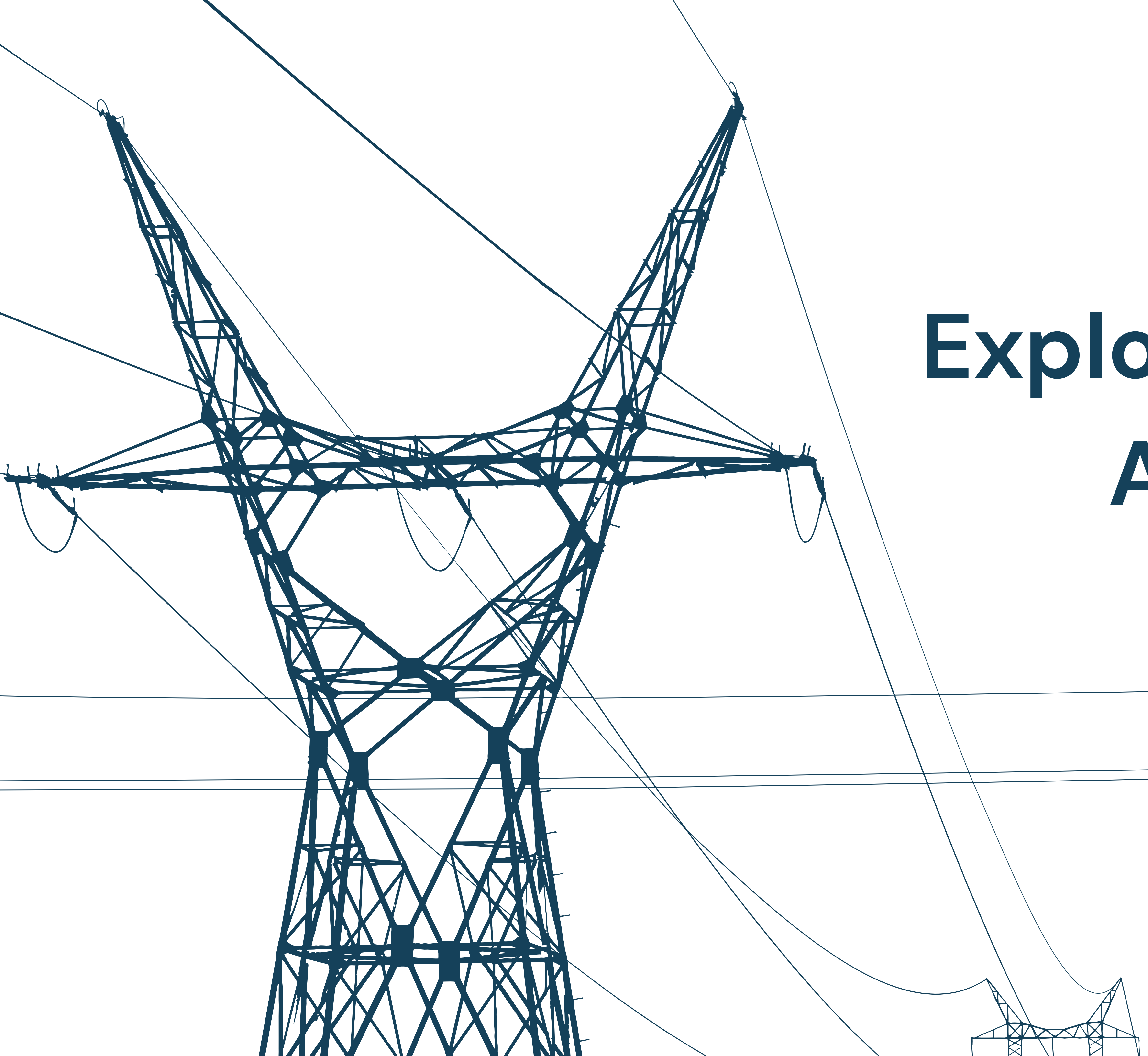
Collected from NOAA’s storm event database, this dataset includes severe weather events in the U.S. from 2014 to 2024. It offers information on event types, locations, and timing, serving as a key resource for linking storm activity to power outage incidents.

Power Outages Dataset

This dataset contains detailed records of power outages across U.S. counties from 2014 to 2023. It tracks the number of customers without power at 15-minute intervals, providing granular insight into the frequency, duration, and geographic distribution of outages

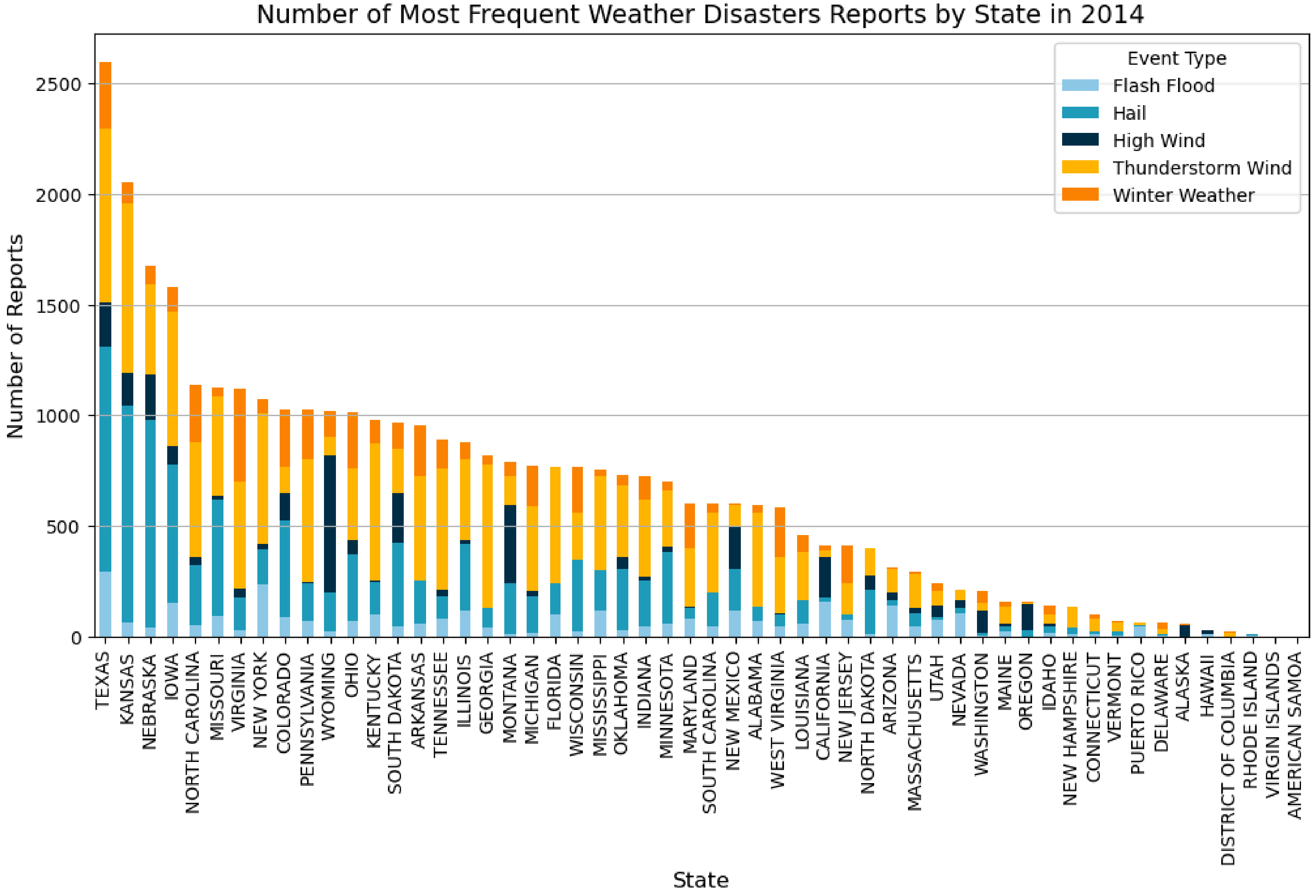
Challenge Approach: Step-by-Step

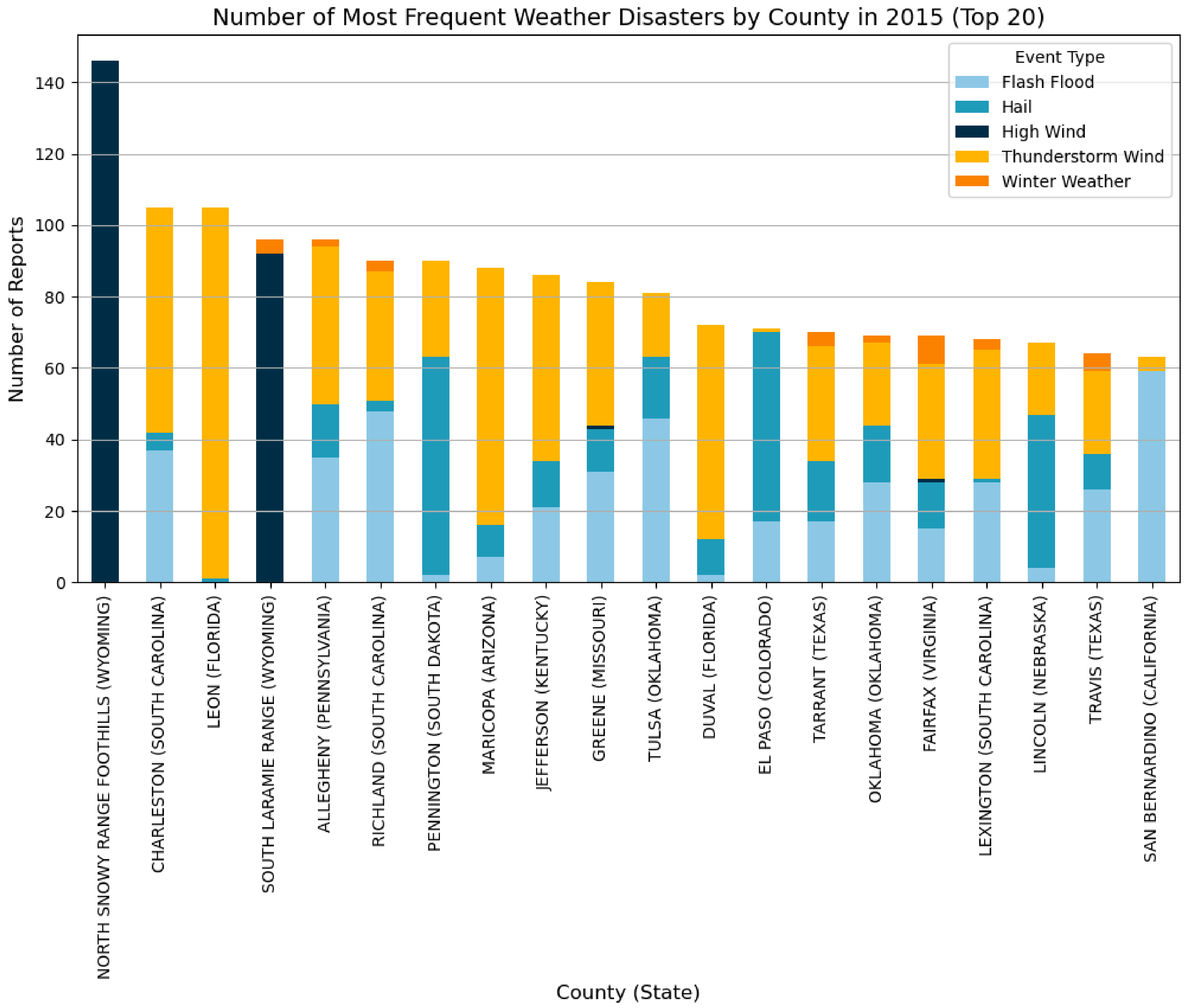
1. **Exploratory Data Analysis** – we analyzed the structure and distribution of the data to identify patterns, anomalies, and relationships that could inform modeling decisions.
2. **Feature Engineering** – we created and transformed features from raw data to better capture the relationship between weather conditions and power outages.
3. **Model Development** – we experimented with various machine learning models, tuning and validating them to predict power outages based on the engineered features.
4. **Interpretability & Insight** – we examined model outputs to extract insights and used interpretability tools to understand which factors most influenced predictions.



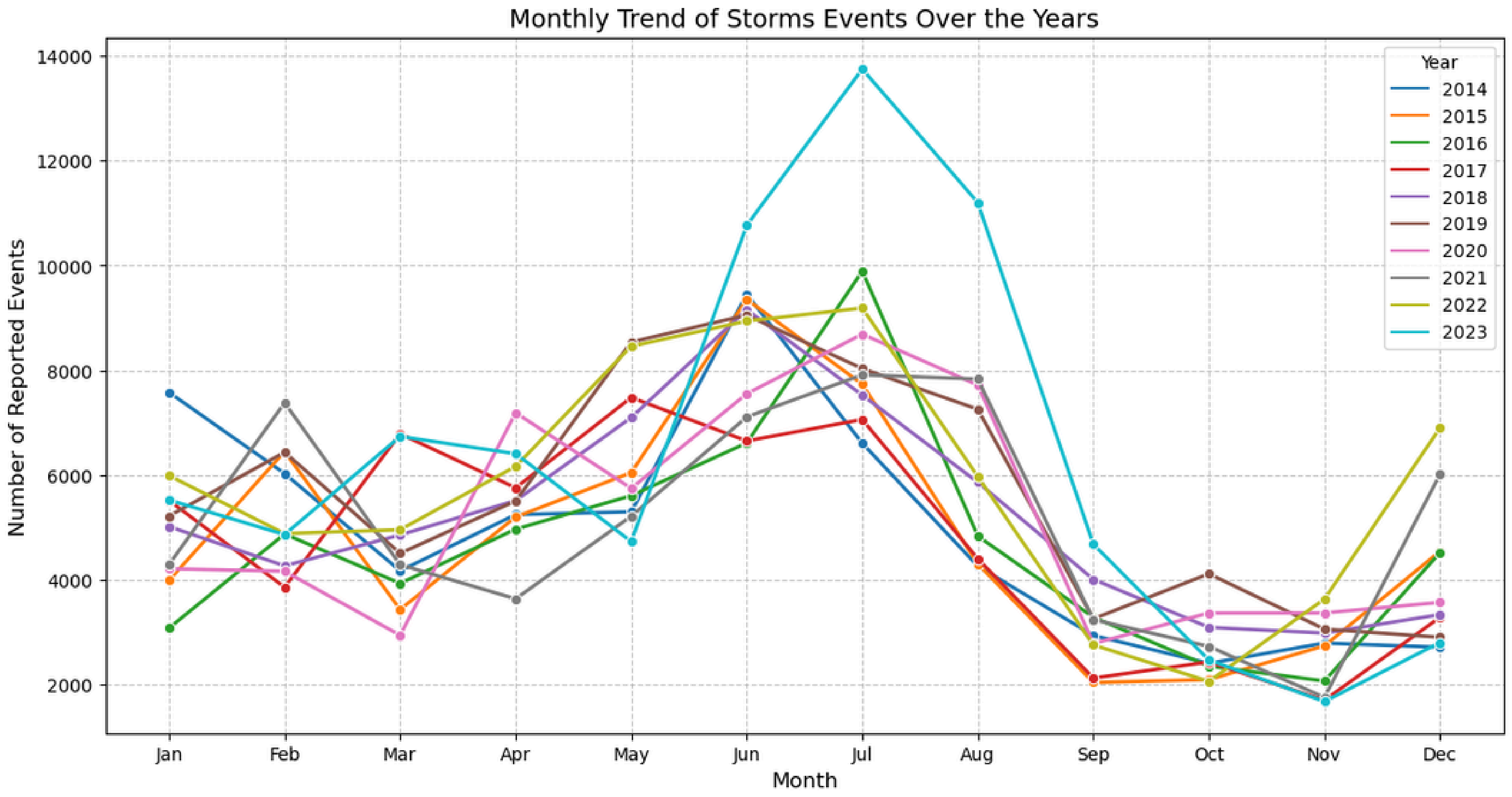
Exploratory Data Analysis

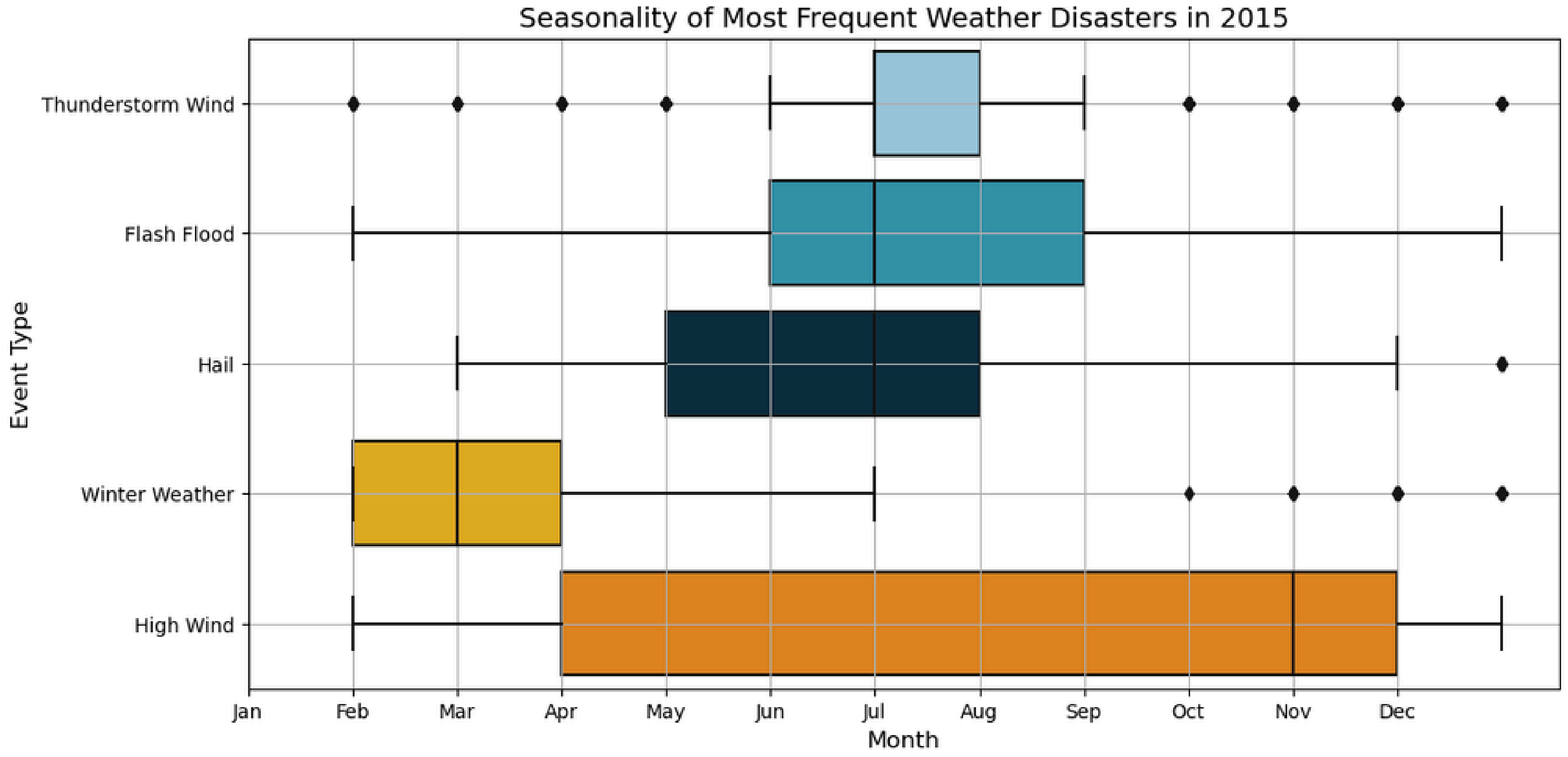
Storm events – location



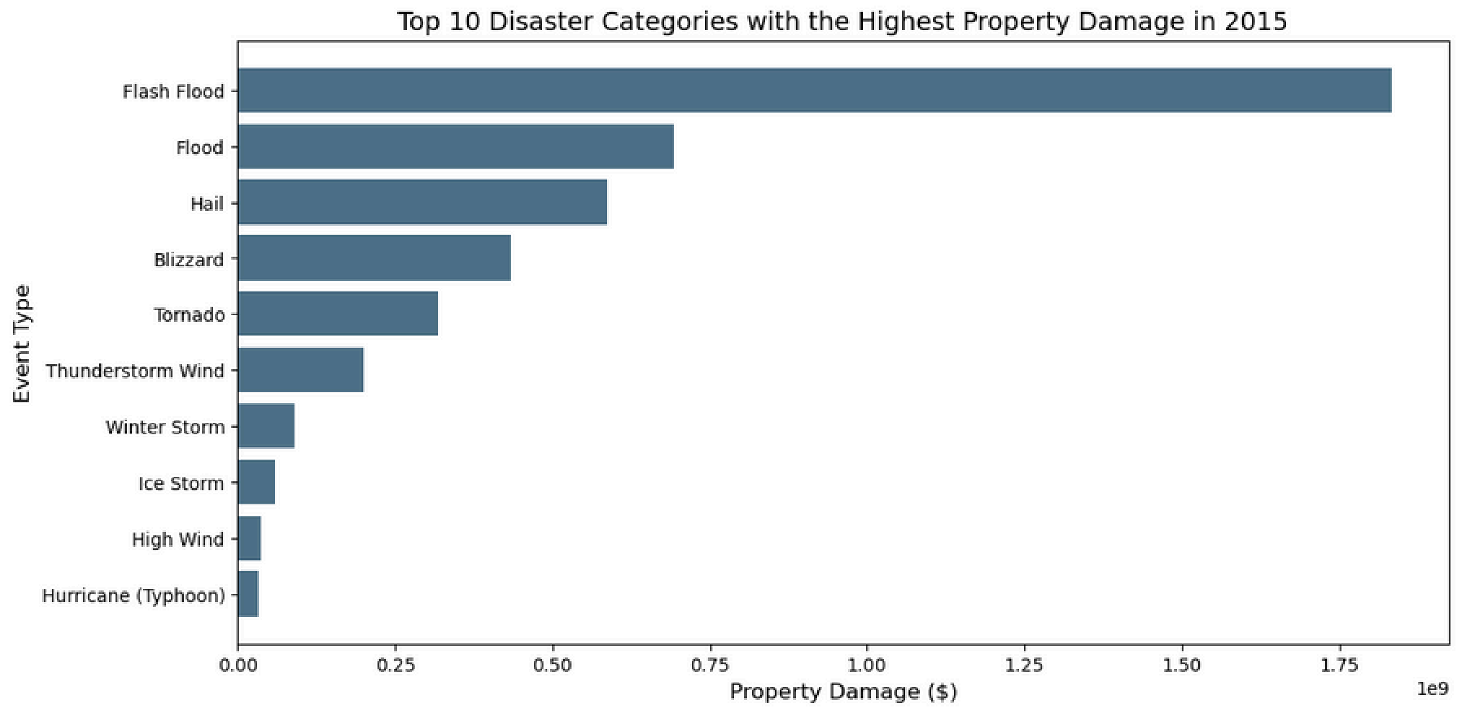


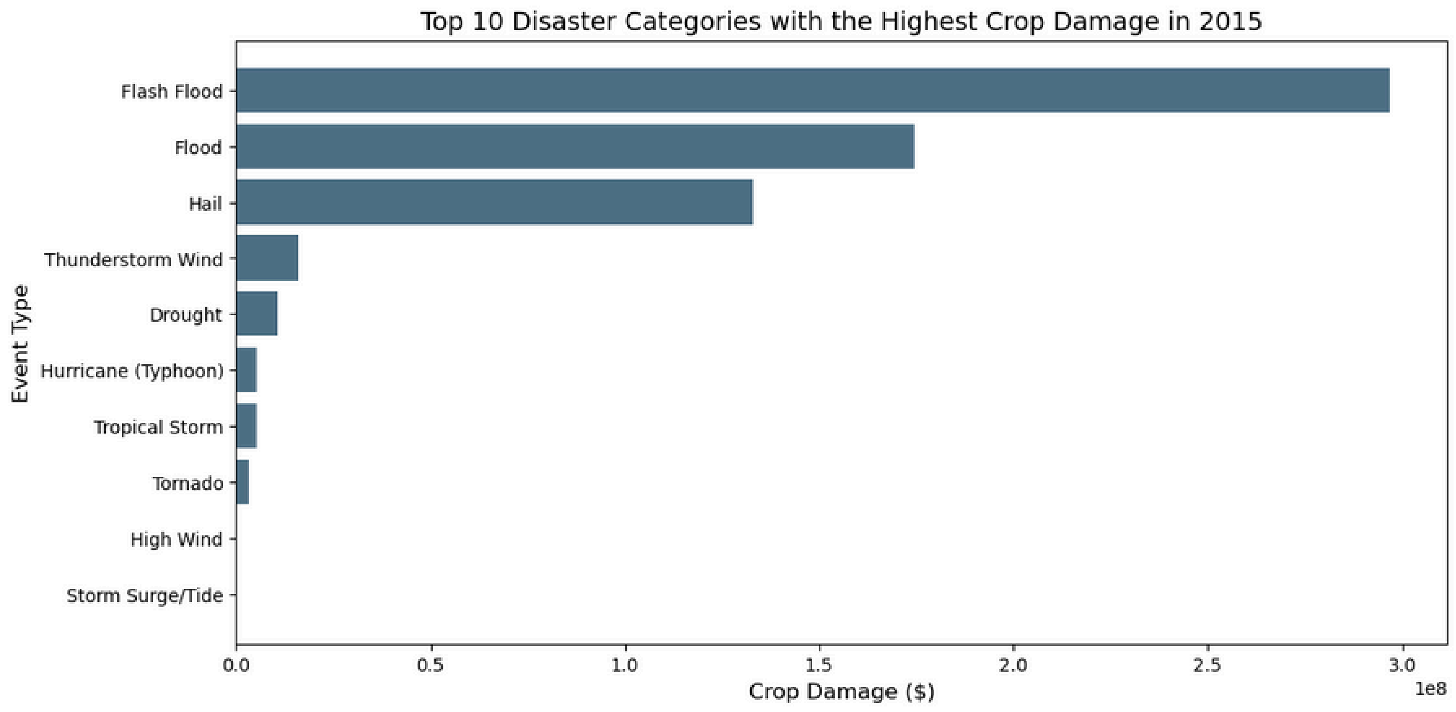
Storm events – seasonality

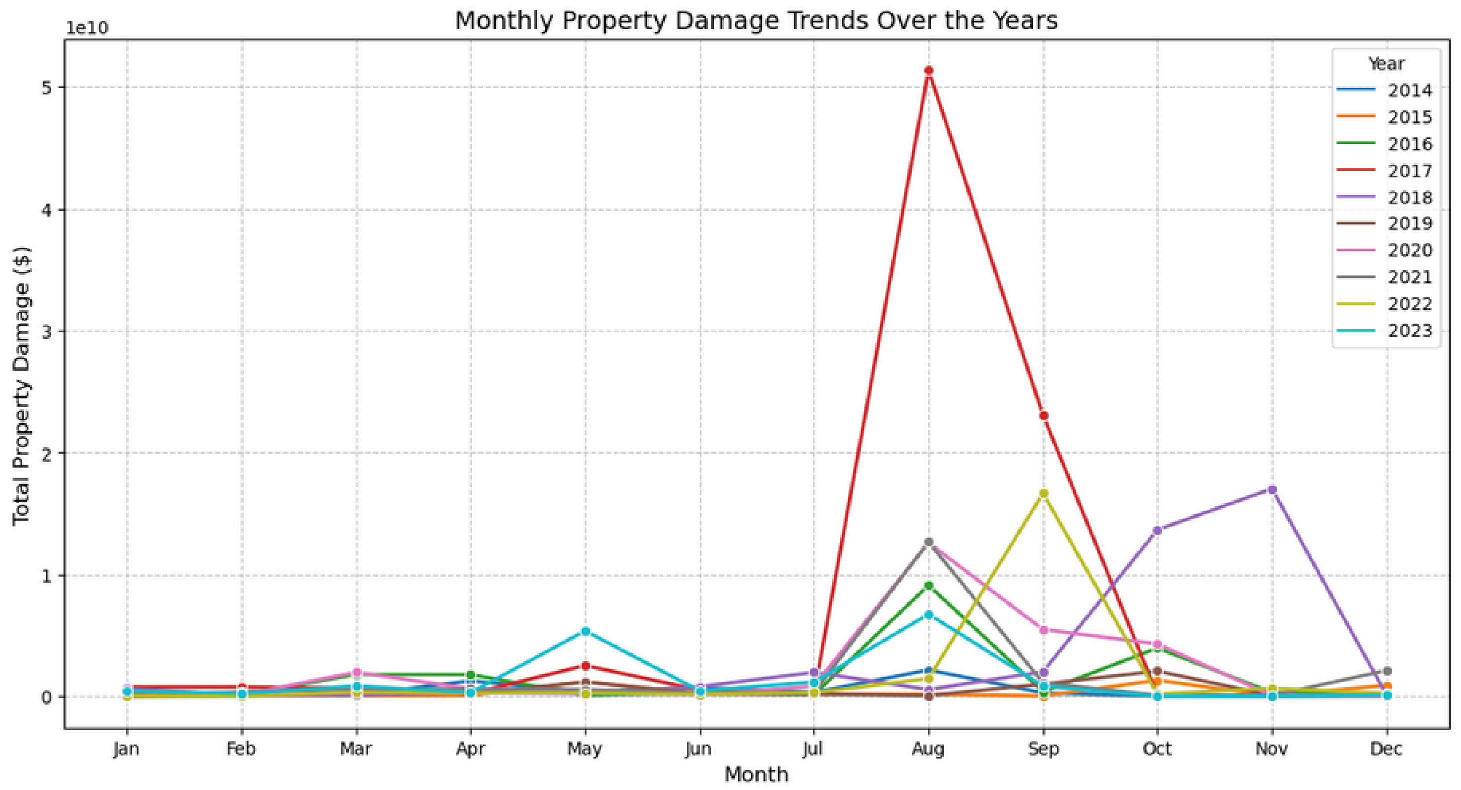


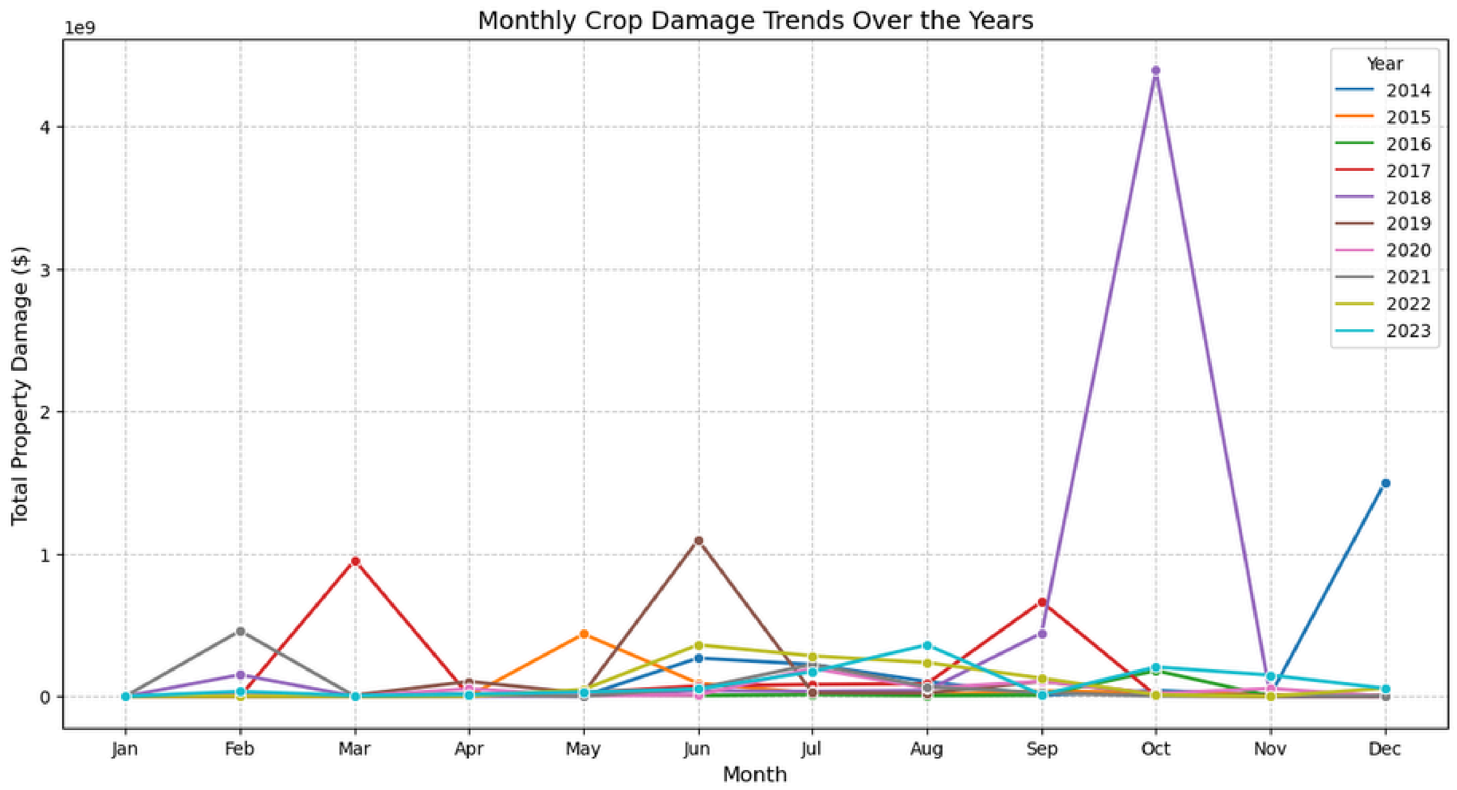


Storm events – damage analysis



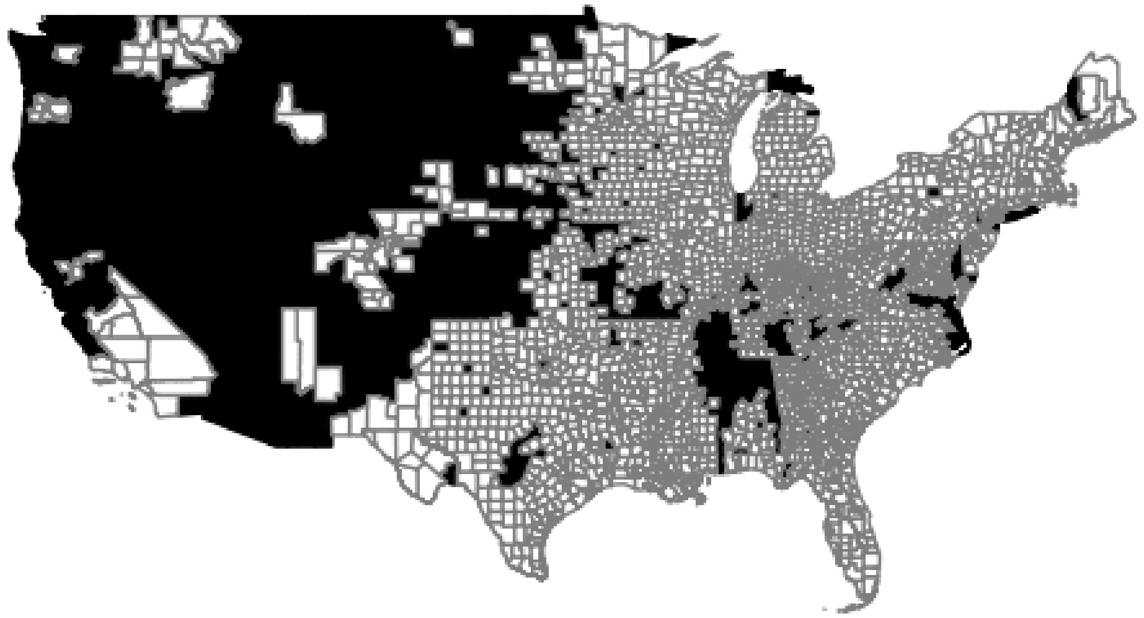




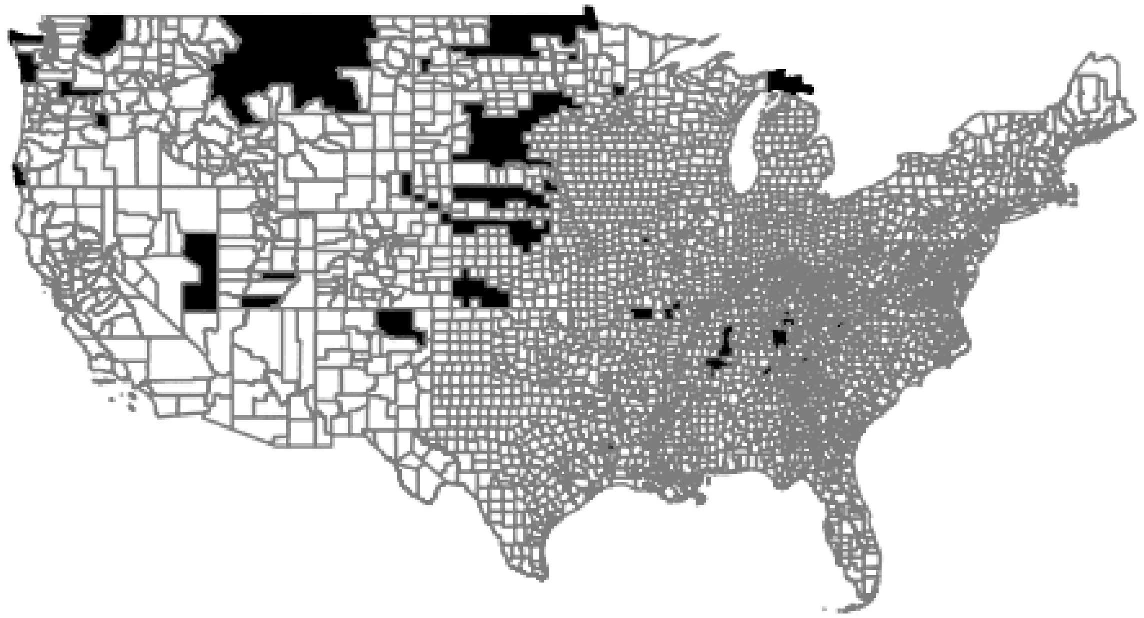


Power Outages – missing data

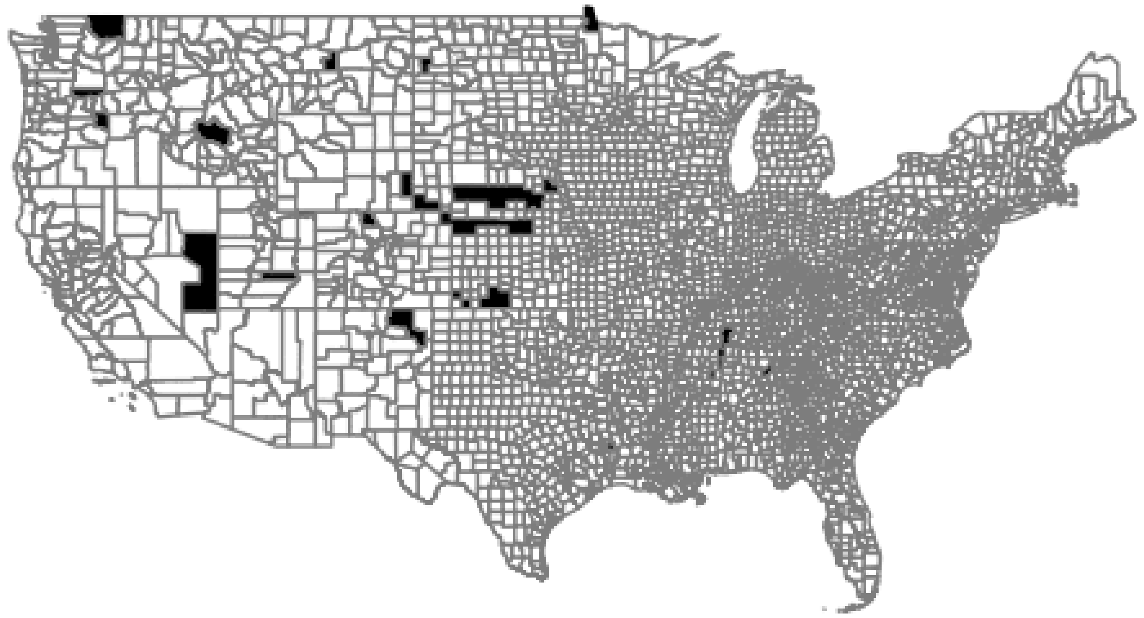
Missing counties 2014



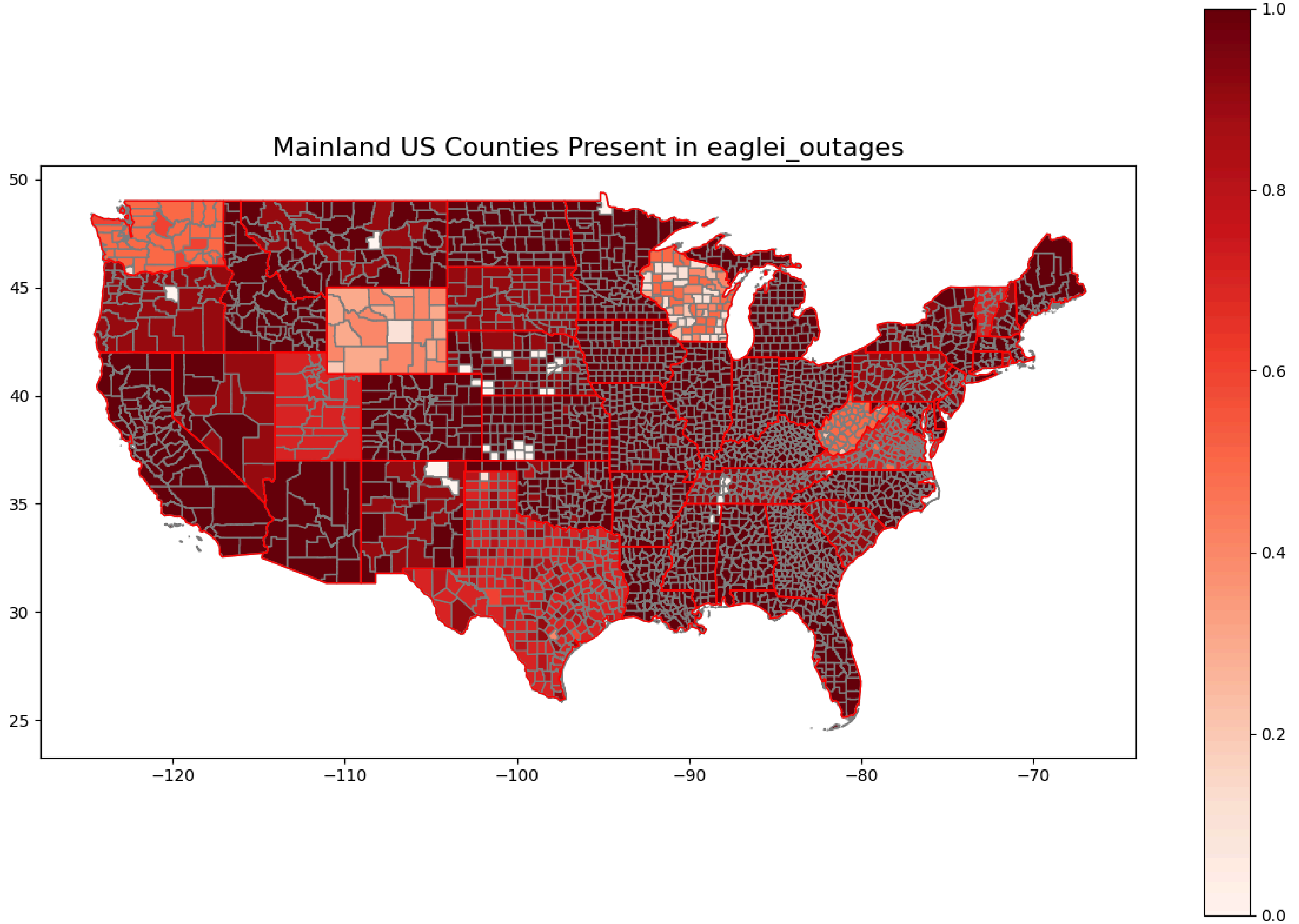
Missing counties 2018



Missing counties 2023



Power Outages – missing data

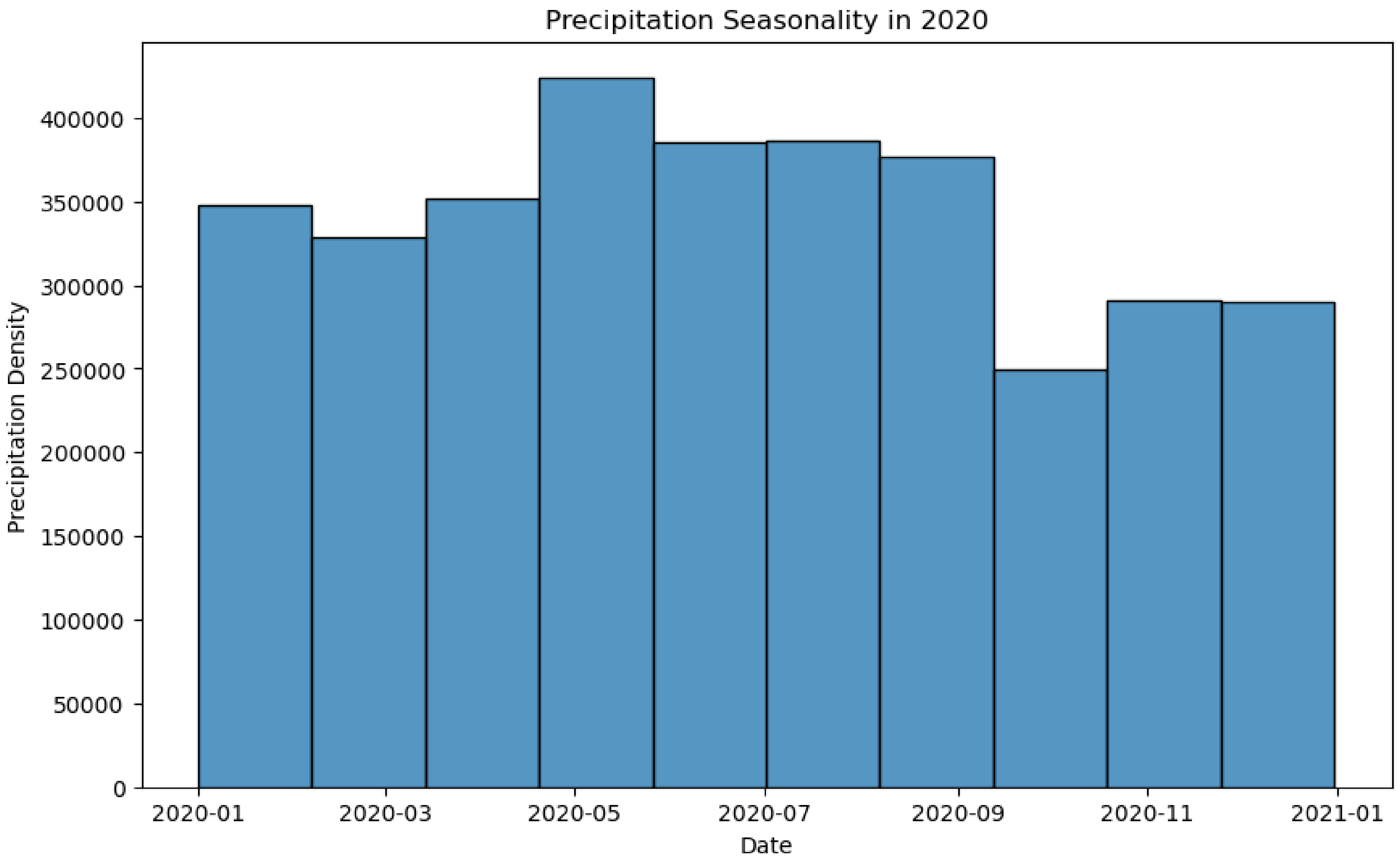


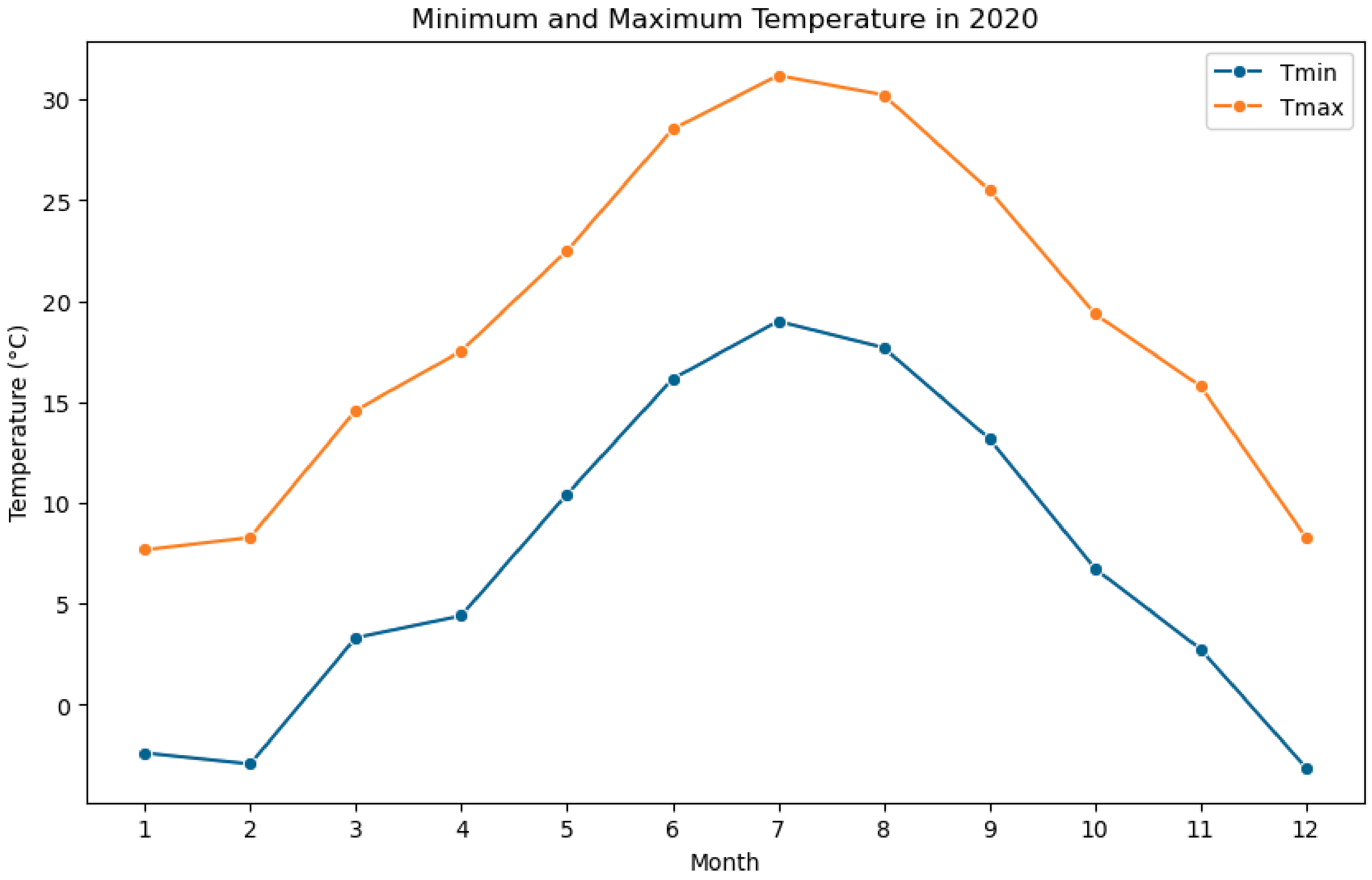
Weather dataset

The organizers emphasized—both in the official rules and in the provided starter notebook—the importance of identifying and utilizing meteorological data related to weather conditions. Our team conducted research to find a comprehensive, publicly available, and free dataset that covers all the required years and accurately represents weather conditions across the United States.

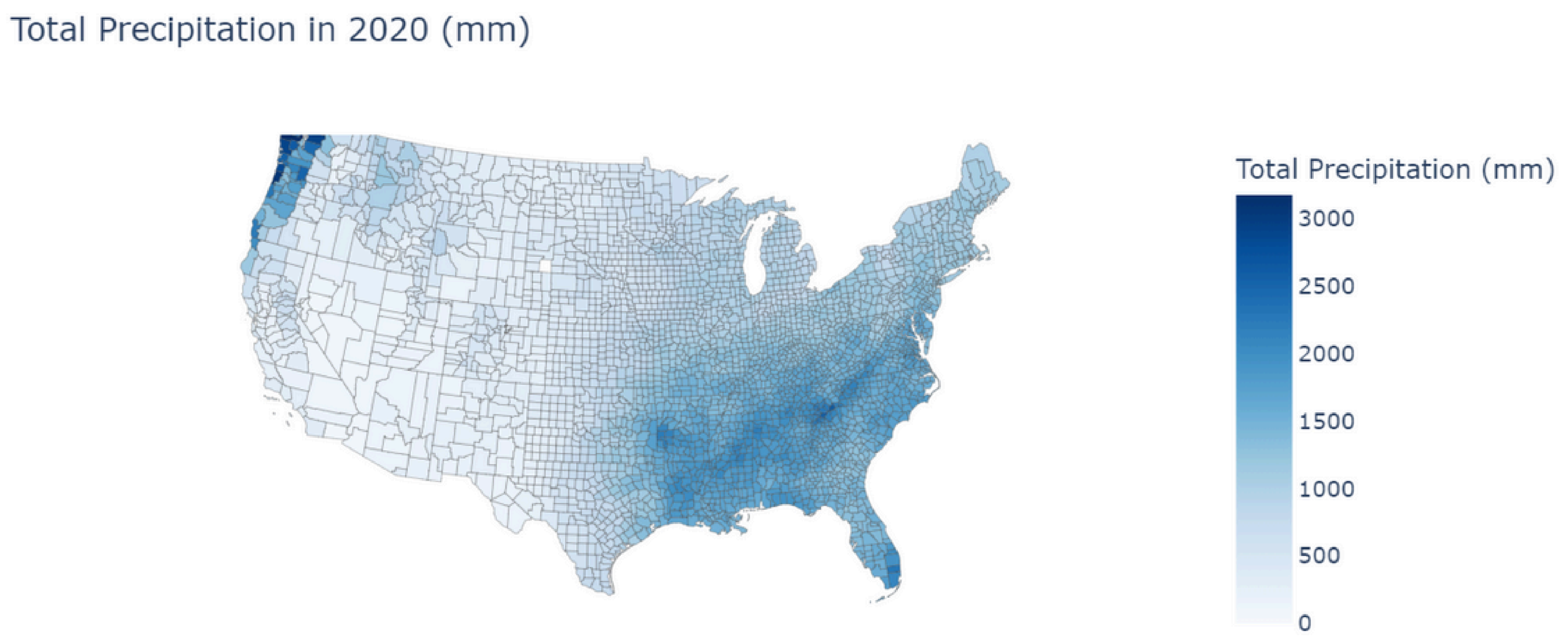
Source: The dataset is sourced from Aaron Smith's website. It provides daily weather observations across various U.S. counties, making it suitable for time-series analysis and integration with other datasets aggregated at the daily level.

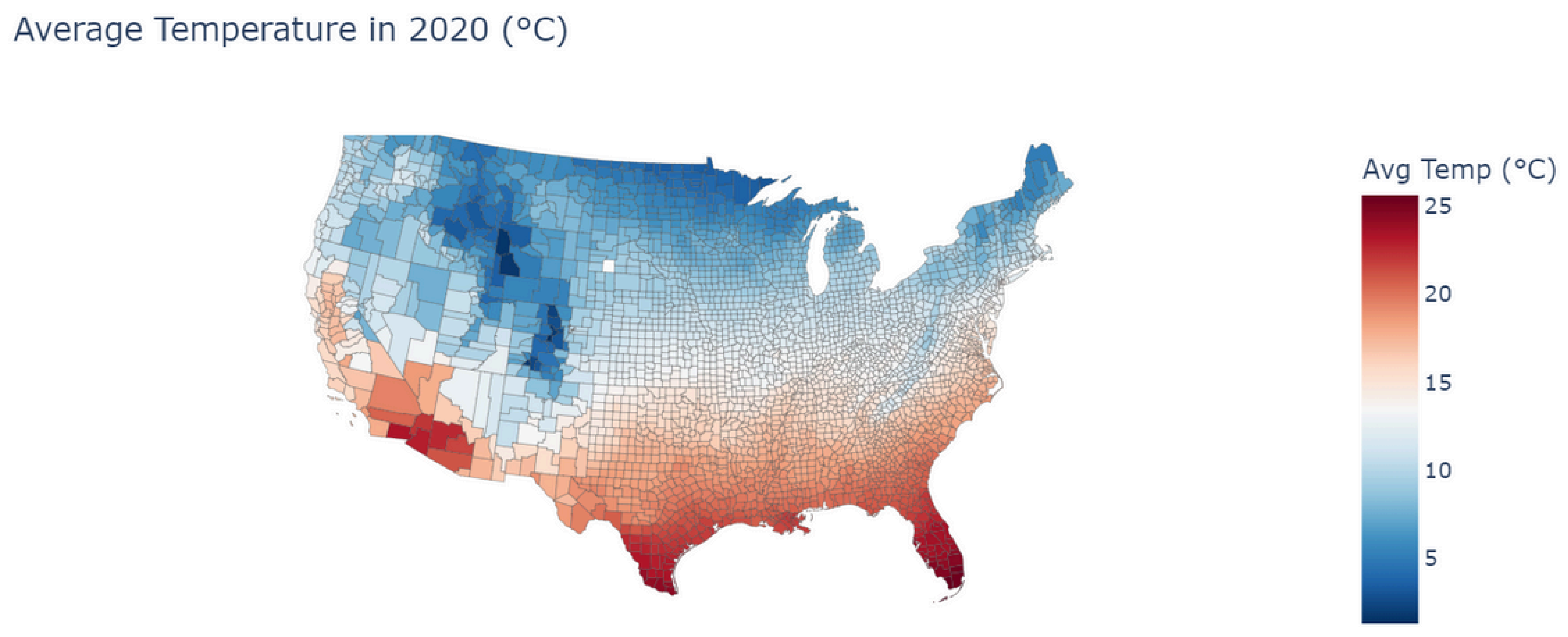
Weather dataset – seasonality





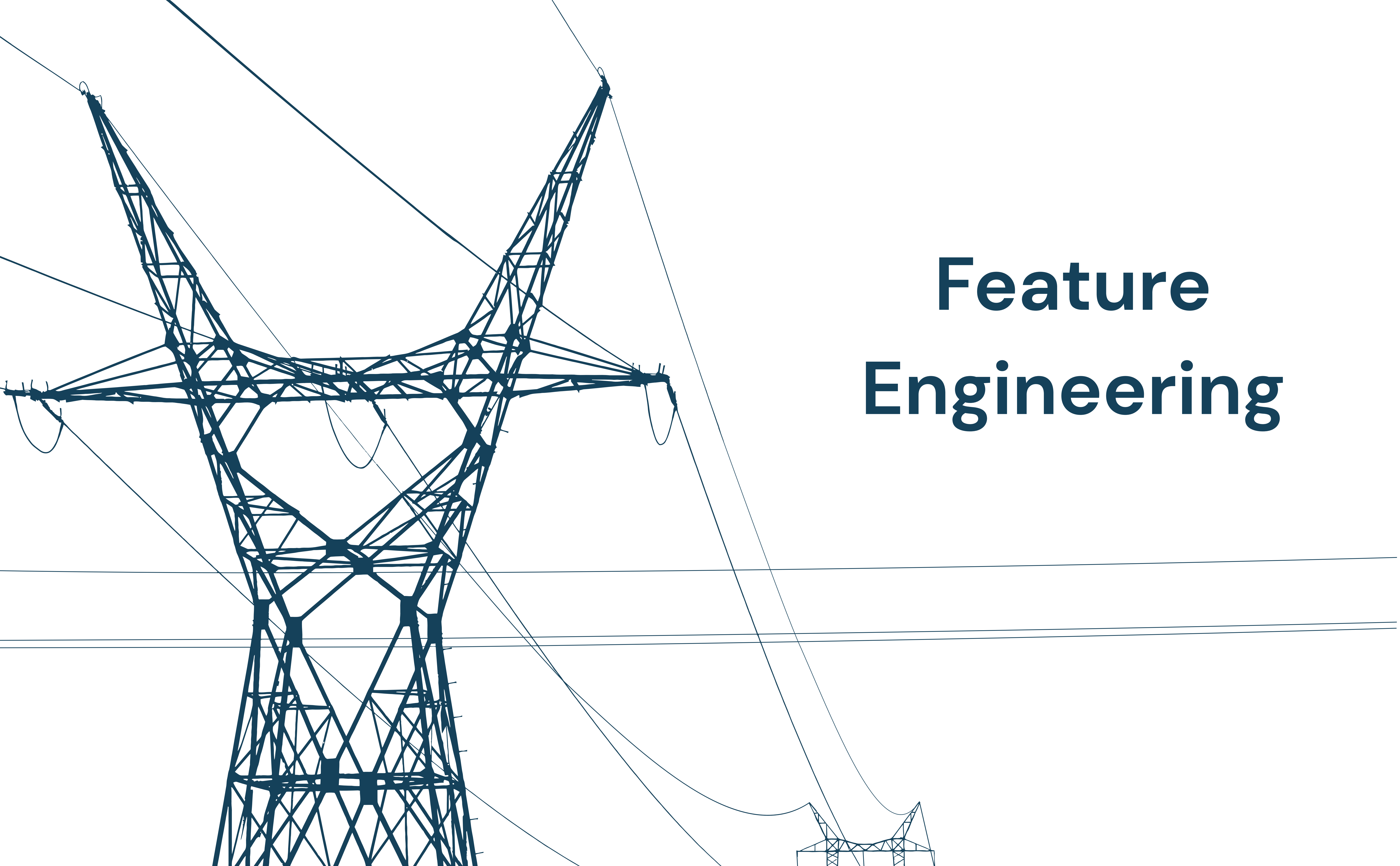
Weather dataset – location





EDA – Key Findings and Considerations

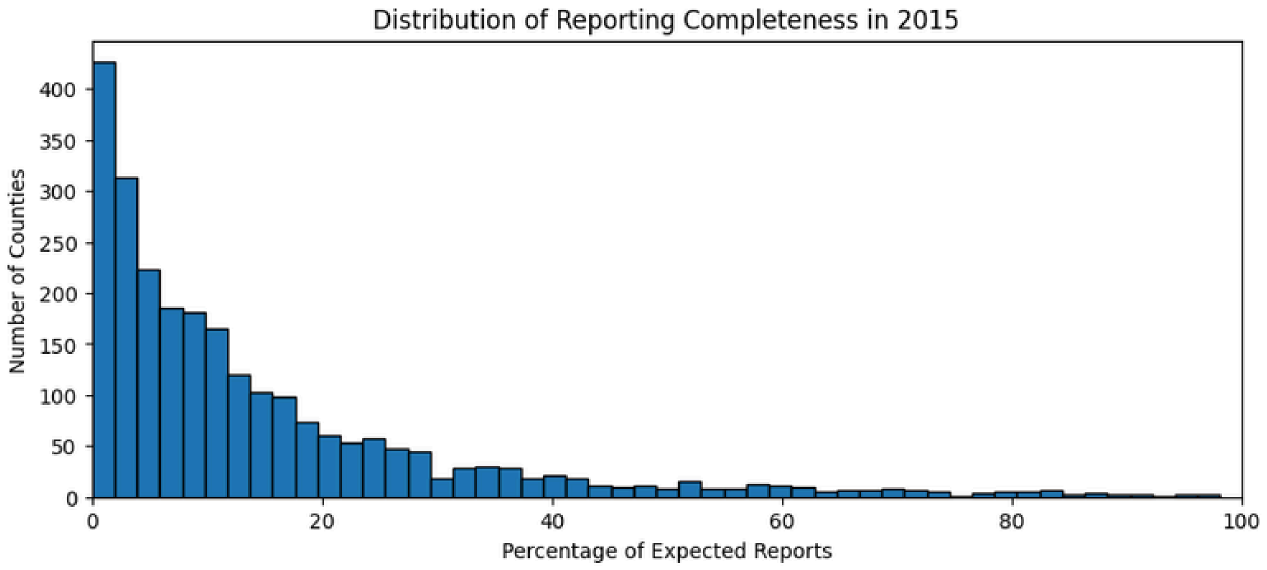
- **Location as a Critical Factor –** Geographic location plays a significant role in modeling outcomes. It is advisable to engineer additional location-based features to enhance predictive performance.
- **Seasonality of Weather Events –** Many extreme weather events follow seasonal patterns. Identifying and incorporating features that capture this seasonality can improve the model's ability to generalize.
- **Data Gaps and Incompleteness –** One of the main challenges lies in filling missing data. The power outage dataset contains substantial gaps, including counties that do not report outages and inconsistent reporting frequencies.
- **Rarity of Catastrophic Events –** Severe weather events are relatively rare, and not all of them result in power outages. This class imbalance complicates modeling and requires careful handling during training.

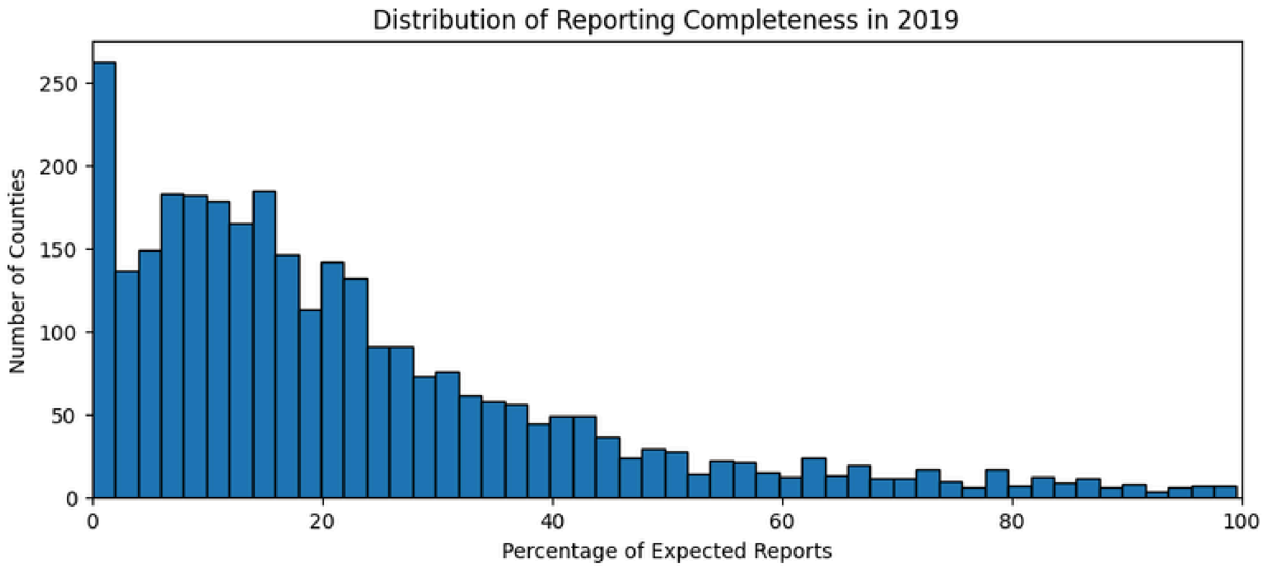


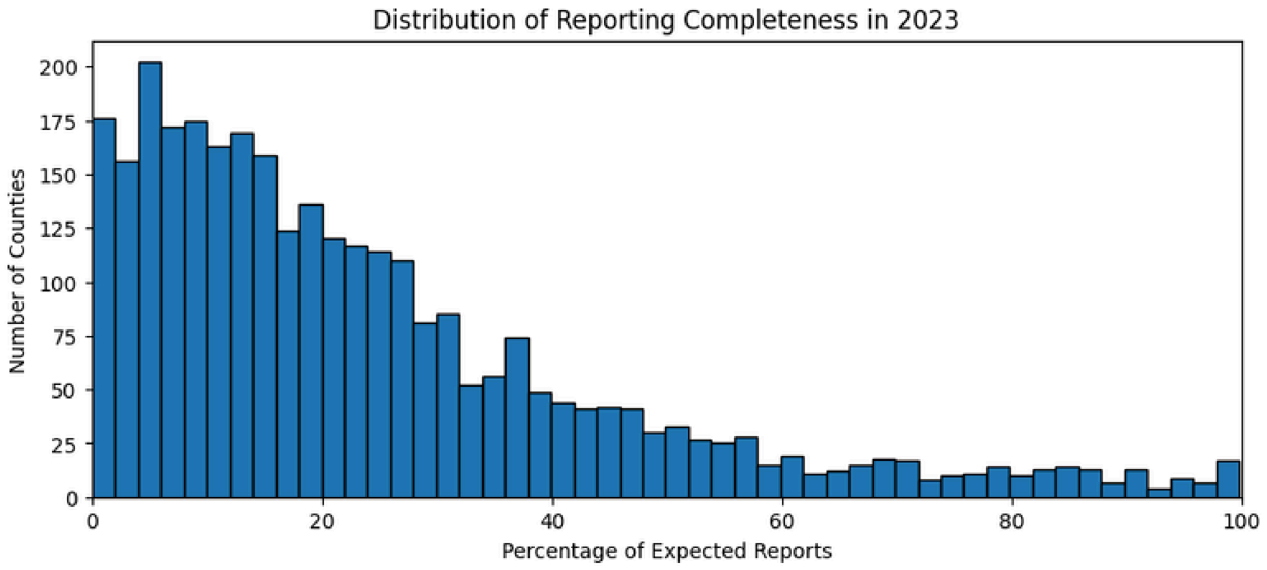
Feature Engineering

Power Outages – Missing data handling

We examined the completeness of reporting in counties, assuming that each county should report at least once every **15 minutes**.

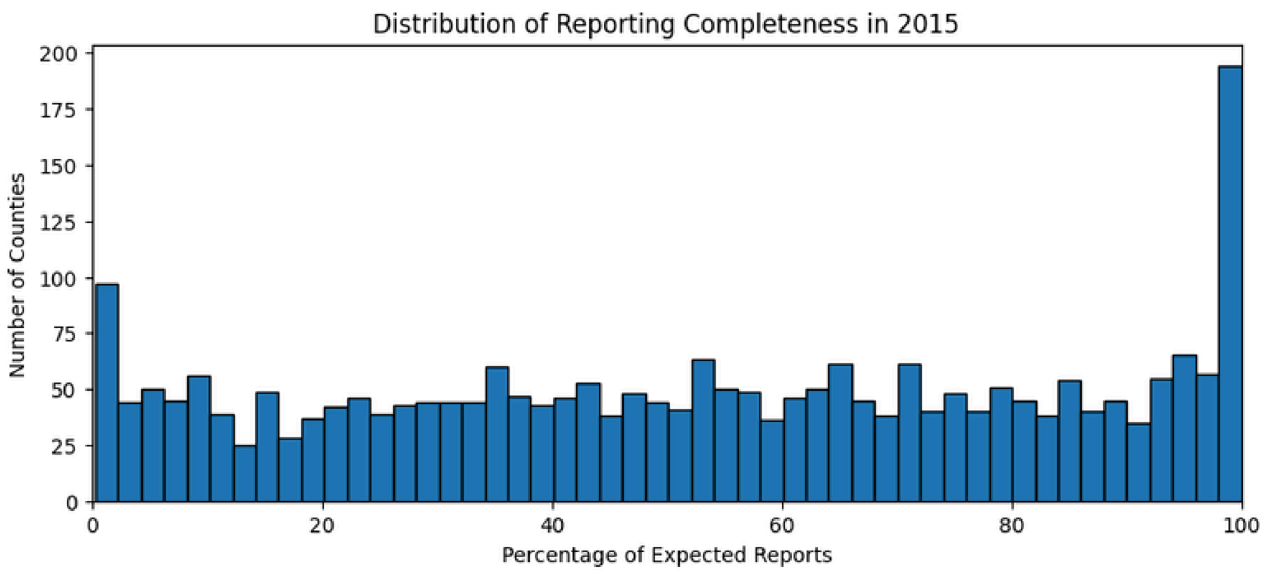


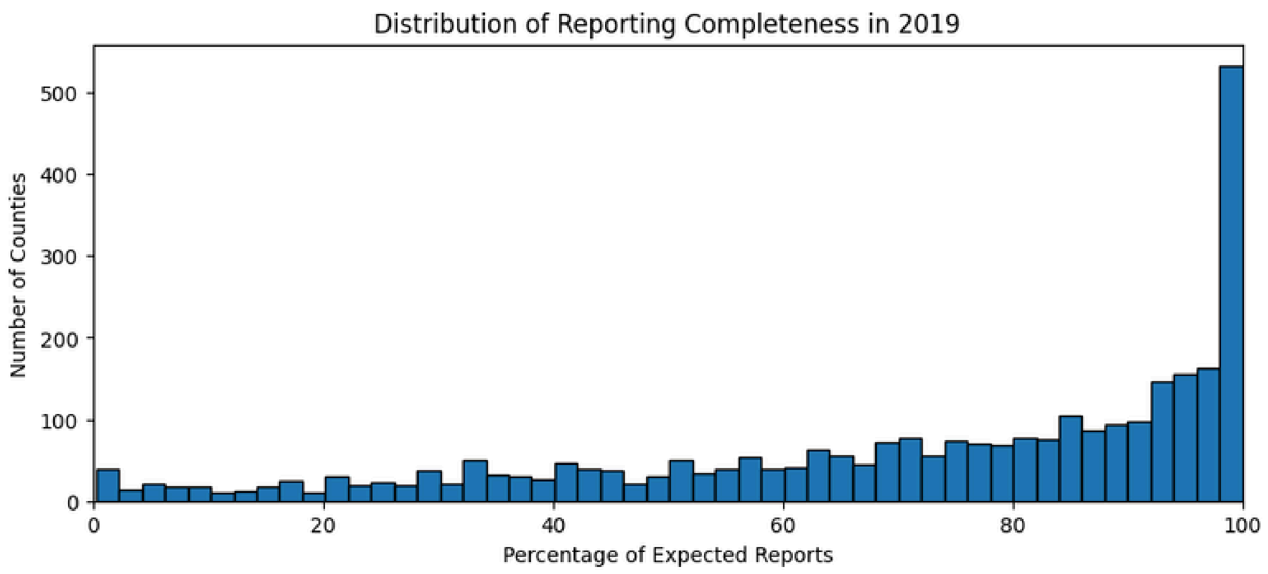


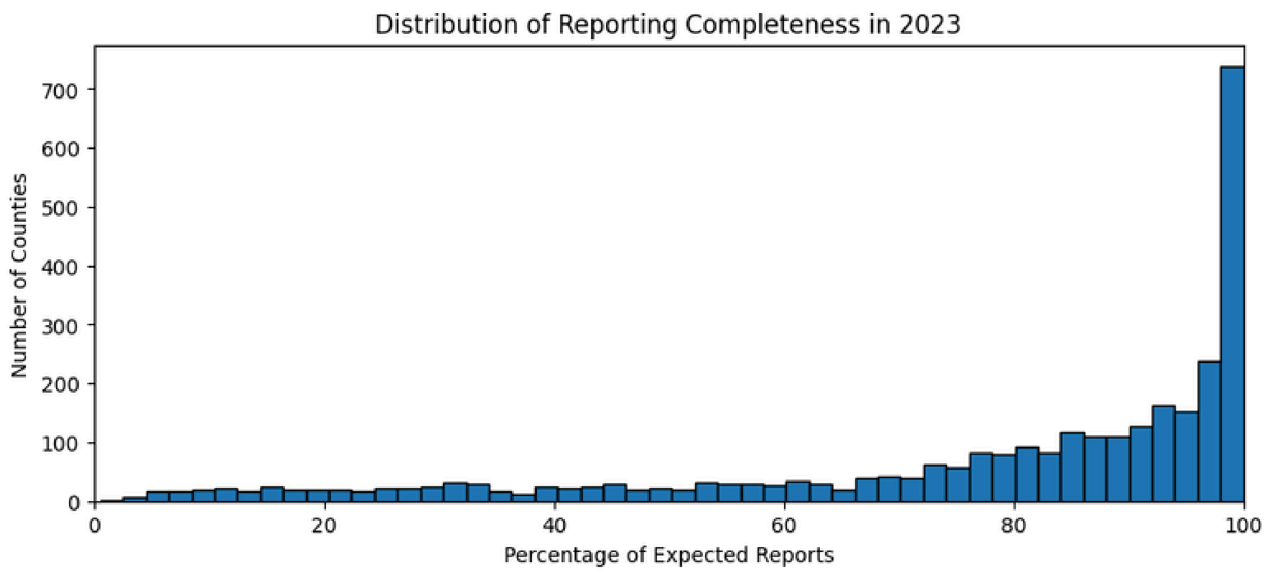


Power Outages – Missing data handling

Later we aggregated data and examined the completeness of reporting in counties, assuming that each county should report at least once every **day**.

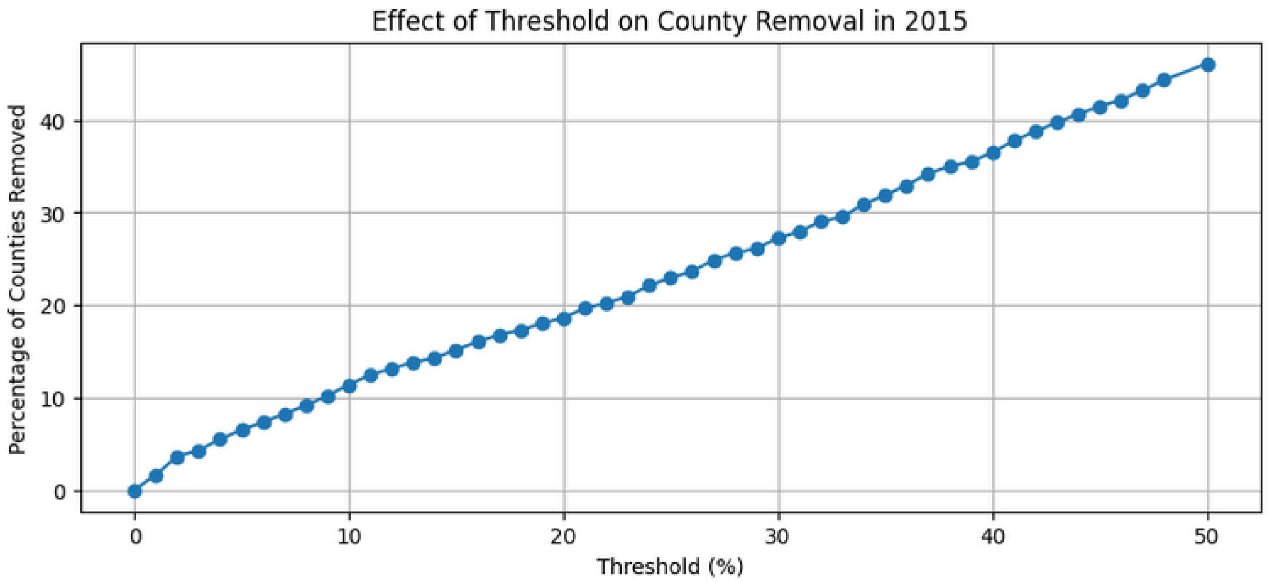


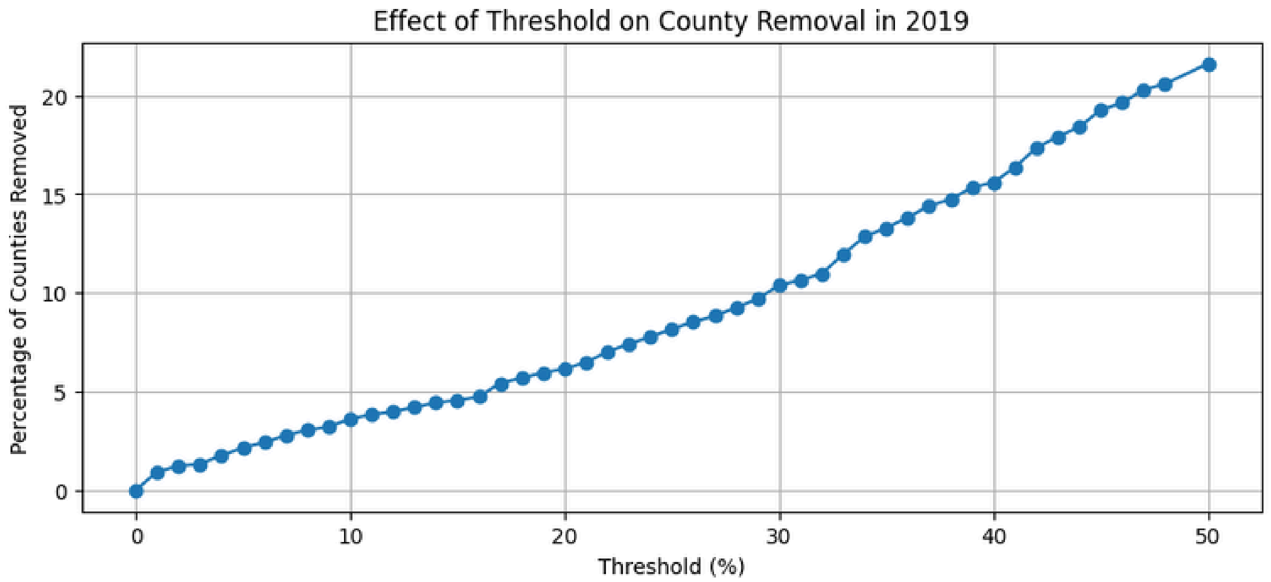


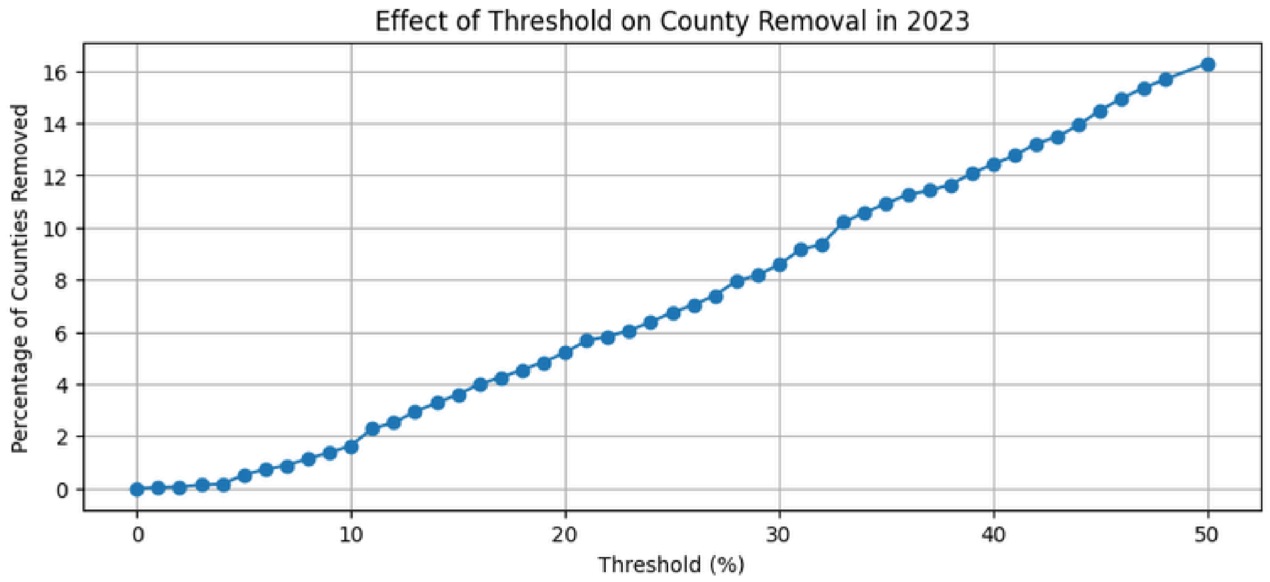


Power Outages – Missing data handling

After performing appropriate daily aggregation, we decided to remove records from counties with the lowest reporting frequency. Prior to this, we assessed how this removal would affect the overall data volume. The analysis showed that excluding the weakest-reporting counties did not result in significant information loss. The final threshold for inclusion was set at **30%.**







Power Outages – Missing data handling

We will use new dataset that provides detailed population data by county in the United States. This dataset includes up-to-date information gathered from authoritative sources such as the U.S. Census Bureau and the Bureau of Labor Statistics.

This dataset is based on the freely available data from SimpleMaps U.S. Counties Dataset.

Additionally, the dataset includes latitude and longitude coordinates for each county, which can serve as valuable features in modeling spatial patterns and regional effects related to power outages.

Power Outages – Missing data handling

CustomersOut

This column represents the raw number of customers without power, as provided in the original dataset. Missing values were filled with 0, based on the assumption that no outages occurred when data is absent.

CustomersOutEstimate

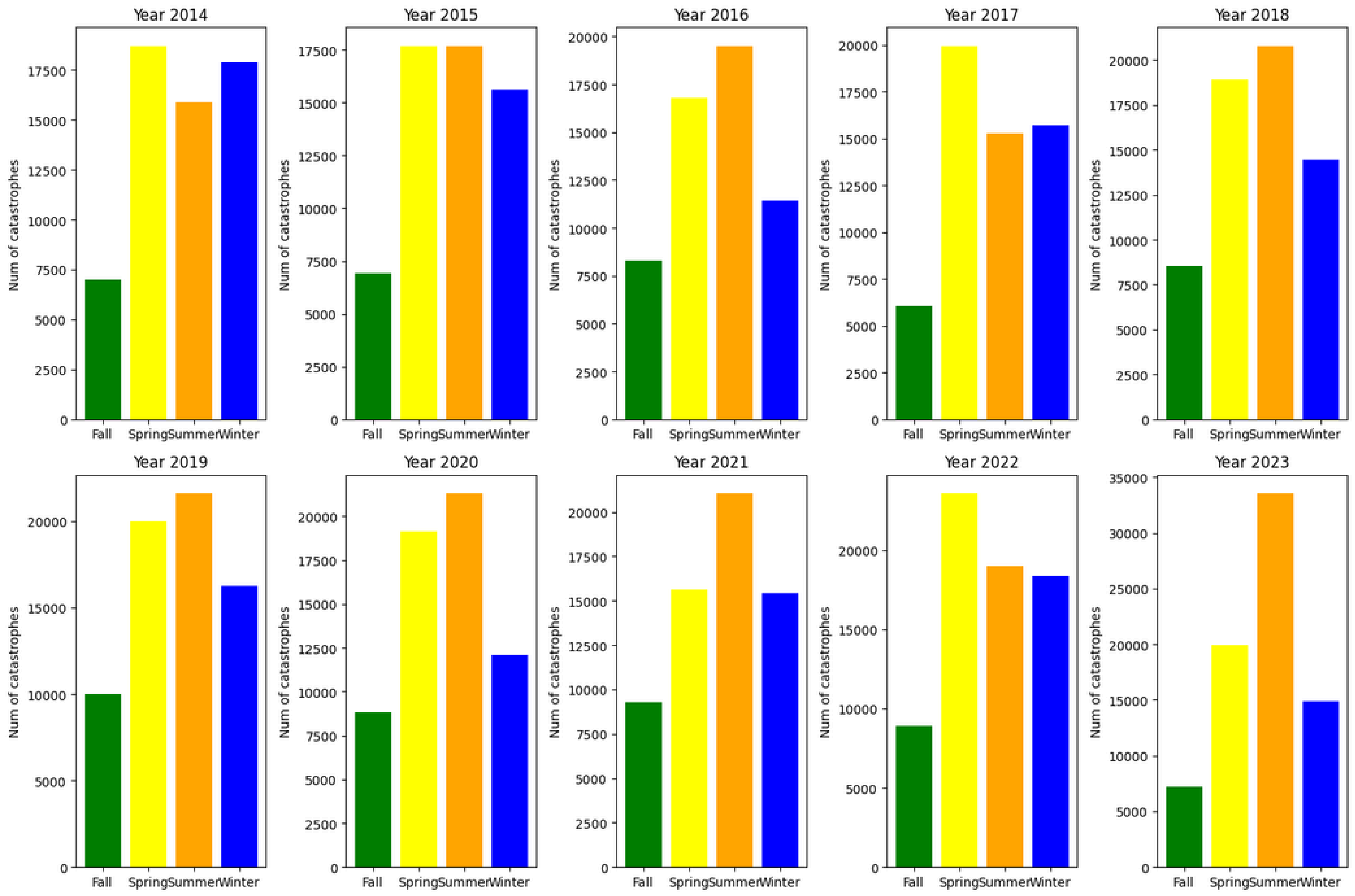
In this column, missing values are estimated using a proxy approach: we calculated the average percentage of customers without power in neighboring counties and multiplied it by the population of each target county. This method aims to approximate likely outages where direct data is unavailable.

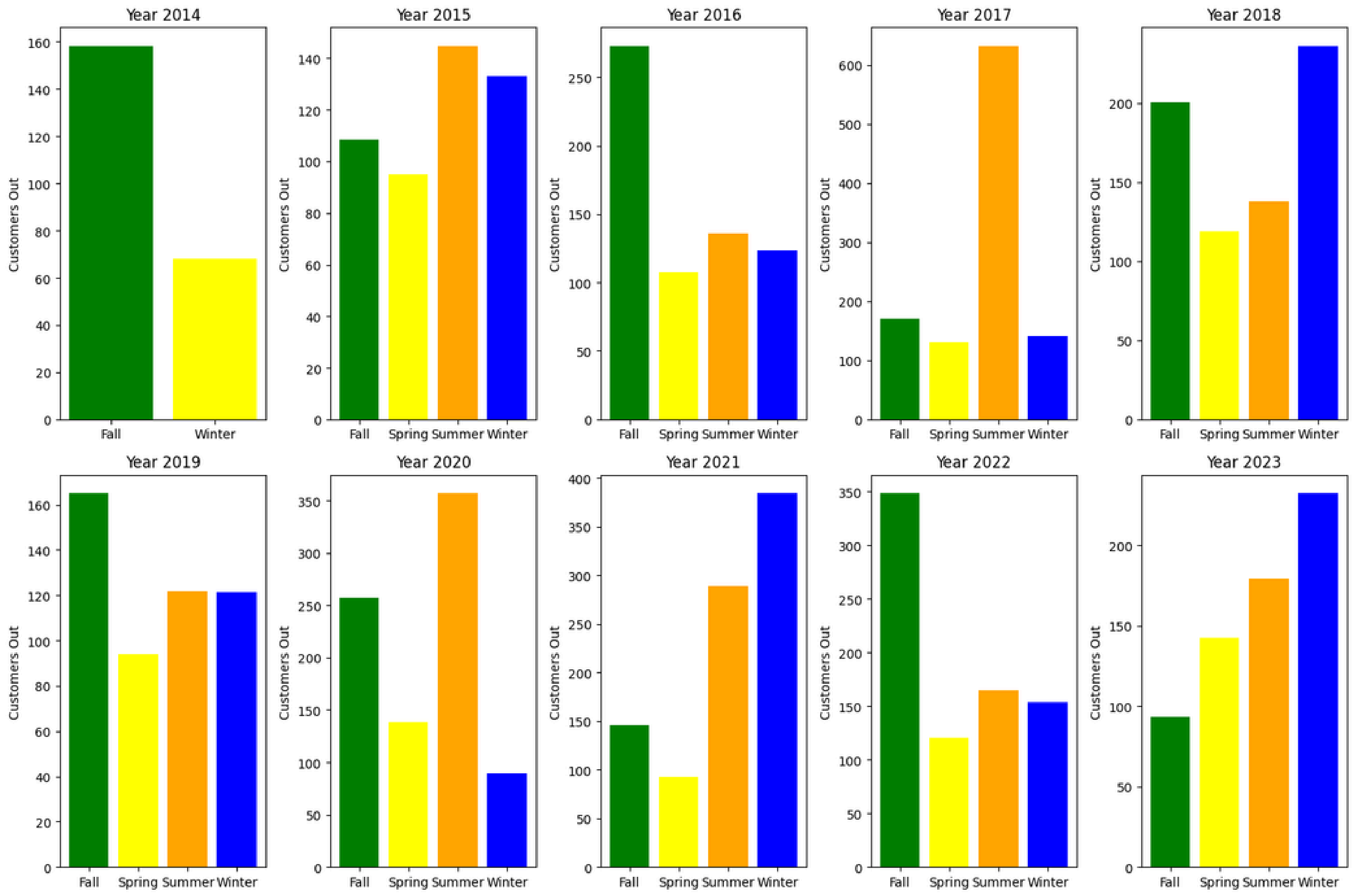
PercentCustomersOut

This column reflects the percentage of the county's population without electricity. It is calculated by dividing the number of customers without power by the total population, providing a normalized measure of outage impact across counties of varying sizes.

New feature – reflecting season

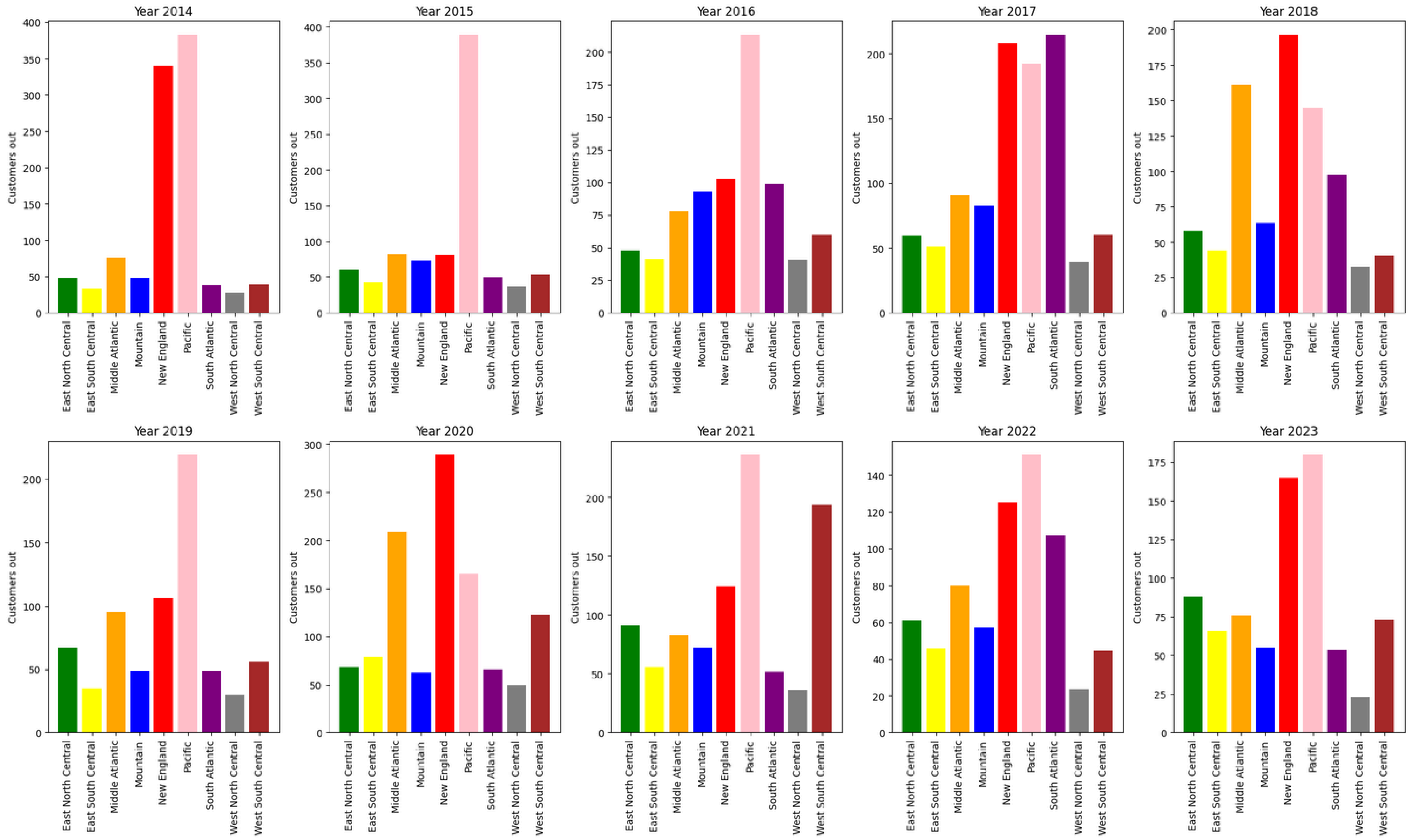
Based on insights from our exploratory data analysis, we identified seasonality as an important factor. Therefore, we decided to add an additional feature indicating the season of the year to better capture temporal patterns in the data.





New feature – reflecting location

We also introduced a new feature to represent the geographic location of each county, as our analysis confirmed that location plays a significant role in the occurrence of power outages.



Feature Engineering – Key Findings and Considerations

- **Missing Data Handling –** One of the main challenges was dealing with incomplete data, as highlighted by the organizers. We applied multiple imputation strategies to enable the use of various features during modeling.
- **New Feature Creation –** We added variables representing geographic location and seasonality to better reflect the spatial and temporal patterns observed in the data.
- **Final Dataset Construction –** We built a comprehensive dataset for modeling, including features related to weather conditions, location, time, and catastrophic events.

Model Development & Interpretability

XGBRegressor

Key model features:

- Target Value: CustomersOut
- Removing event-related features with more than 99.8% zero values
- Keeping only rows with valid flag
- Two approaches:
  - Global Model: One model trained on data from all states combined
  - Local Approach: Separate models trained individually for each state

Results from global model:

PERFORMANCE

RMSE:	2124
MAE:	170
R2:	0.1

XGBRegressor

Due to the low performance of the global model, it was decided to train separate models for each state. Below is a comparison of the results for individual states achieved using both the global and local models.

The tables presents four states:

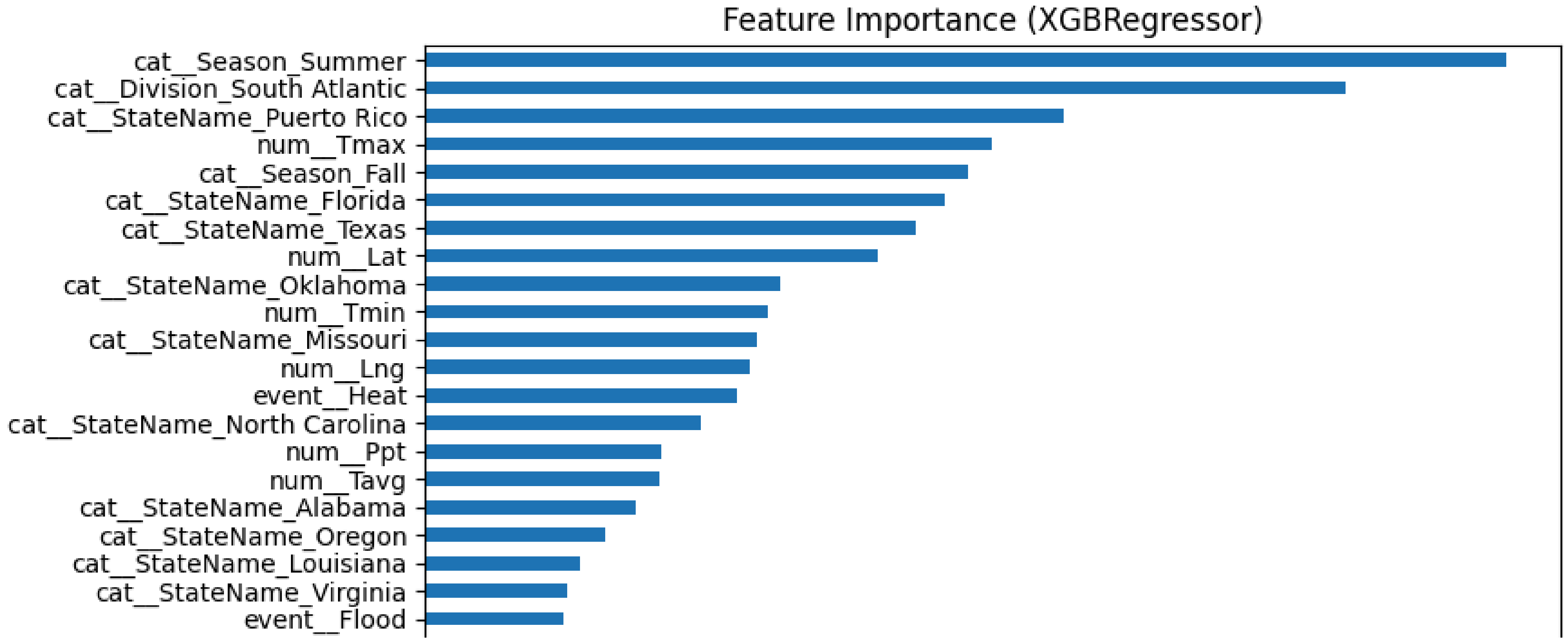
- **The two best-performing states**
- **The two worst-performing states based on the global model's results.**

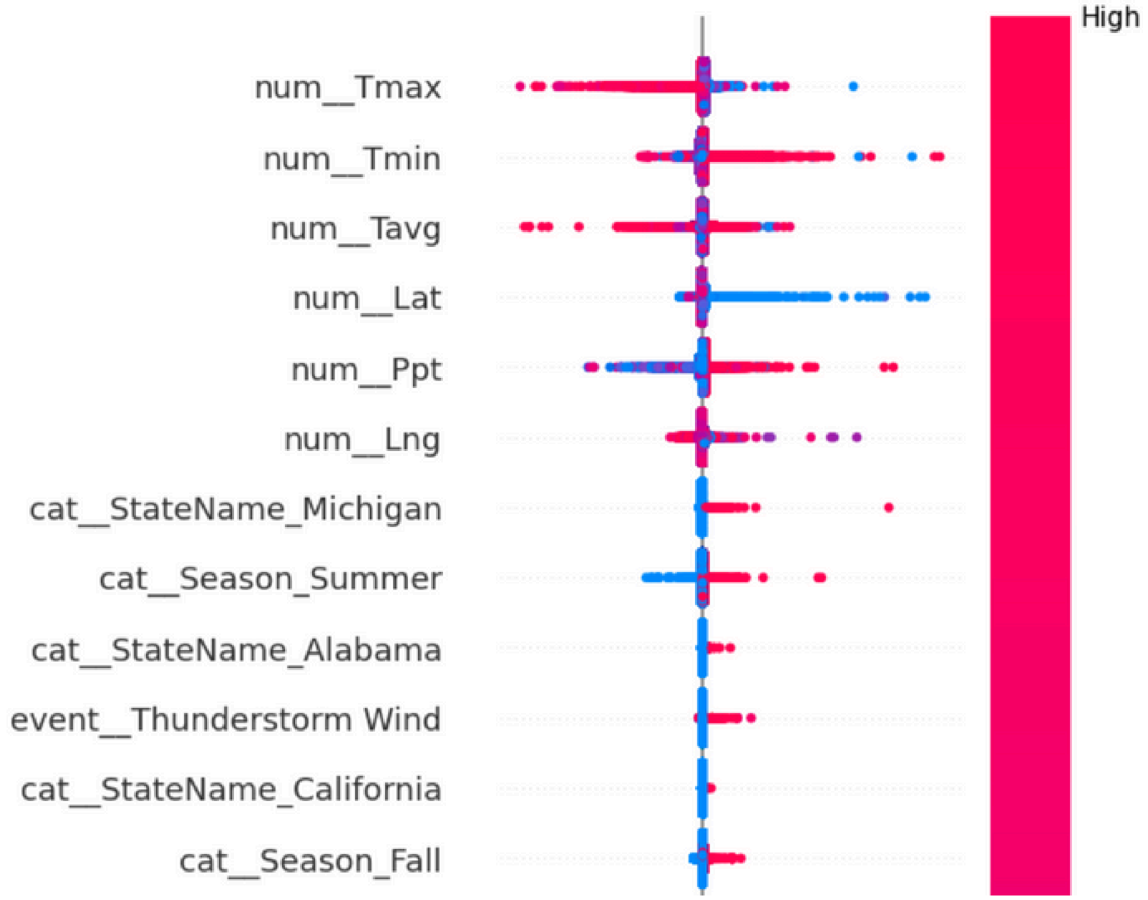
GLOBAL MODEL	MAE	R2	RMSE
Wyoming	44	-0.17	112
South Dakota	63	-0.13	198
Florida	402	0.49	4708
Puerto Rico	2839	-0.14	13321

LOCAL MODELS	MAE	R2	RMSE
Wyoming	32	-0.09	136
South Dakota	39	-0.26	128
Florida	531	-0.02	6038
Puerto Rico	3074	-0.01	13819

XGBRegressor

Due to the high complexity and noise in the data, the model struggled to accurately estimate target values. While state-specific models improved performance in certain regions, high-error areas like Puerto Rico and Florida continued to show significant prediction challenges





MLPRegressor

A basic MLPRegressor model was developed as an initial benchmark. However, the MLP struggled with high prediction errors and showed weak predictive performance, with an R^2 score of approximately 0.07. In comparison, the XGBRegressor demonstrated better generalization capabilities, achieving consistently higher R^2 scores across different states. These results indicate that tree-based models, such as XGBoost, are better suited for this particular prediction task.

Results from global model:

PERFORMANCE

RMSE:	2159
MAE:	174
R2:	0.07

Sarimax

Key model features:

- County-Level Analysis: The dataset is grouped by county to capture local patterns and characteristics more effectively
- Target Variable – PercentCustomersOut: This metric accounts for differences in county sizes
- Prediction Focus: The model is trained to predict the final year of available data, allowing evaluation on the most recent patterns and trends

PERFORMANCE

MEAN MAE:	0.45
MEAN RMSE:	1.38

Sarimax

Analysis by county – lowest MAE and RMSE:

COUNTY	MAE	RMSE
Swain	0.00	0.00
Haywood	0.00	0.00
Macon	0.00	0.00

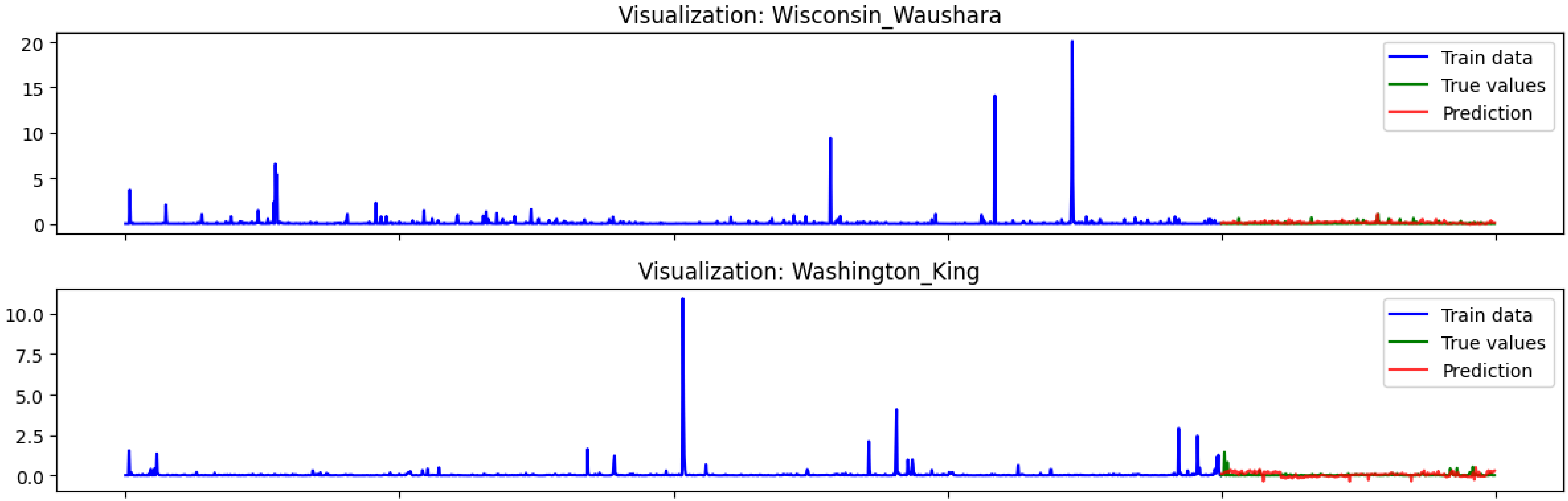
Analysis by county – highest MAE and RMSE:

COUNTY	MAE	RMSE
Waynesboro	37.5	41.42
Henderson	33	37.86
Presidio	7.8	18.75

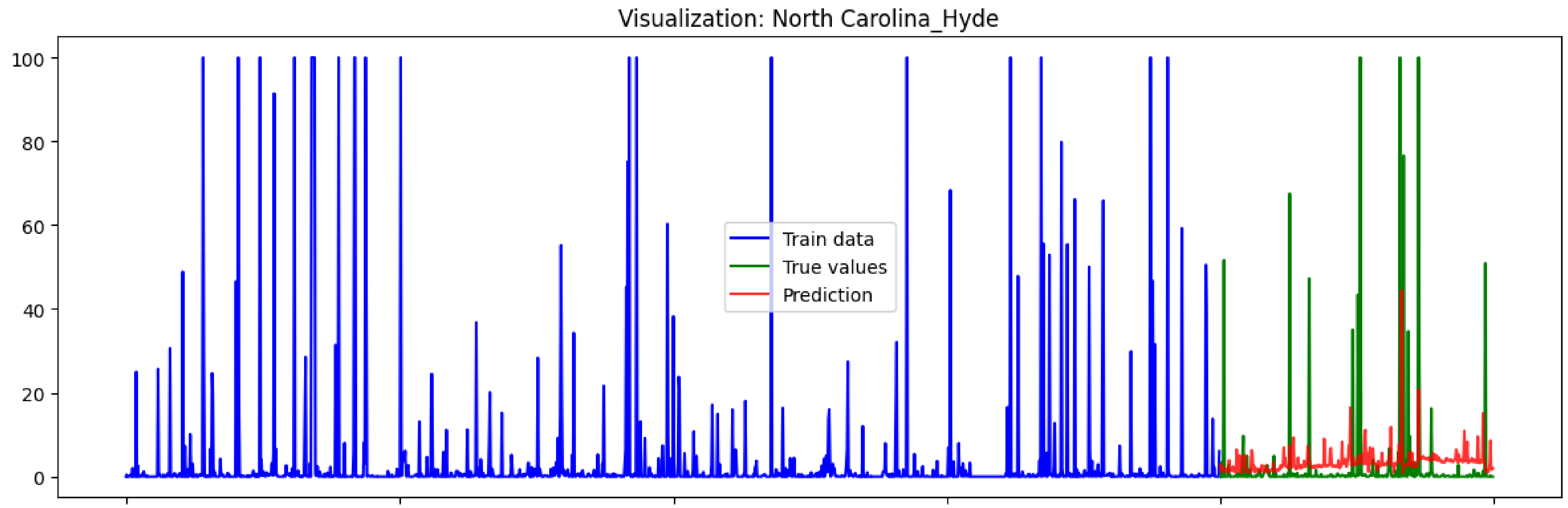
Sarimax

Counties with very low error metrics typically correspond to areas where no major power outages occurred in the training data. In such cases, the models learned to predict zero outages consistently, which happened to be correct. However, this does not necessarily indicate strong model performance—rather, it reflects a lack of variability in the target data. These artificially low errors can give a false sense of accuracy, as the model has not truly learned to distinguish between outage and non-outage scenarios. Therefore, caution is needed when interpreting results from such counties.

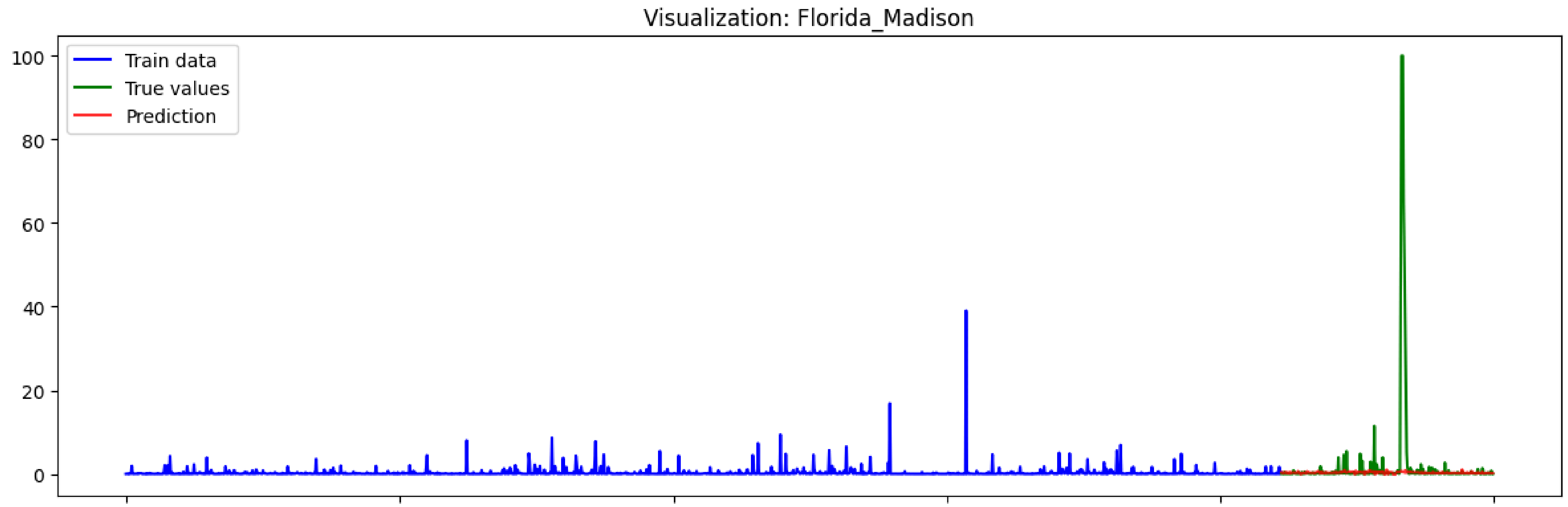
Sarimax


Visualization: Wisconsin_Waushara
20
15
10
5
0
Train data
True values
Prediction
Visualization: Washington_King
10.0
7.5
5.0
2.5
0.0
Train data
True values
Prediction

Sarimax



Visualization: North Carolina_Hyde
100
80
60
40
20
0
Train data
True values
Prediction



Visualization: Florida_Madison
100
80
60
40
20
0
Train data
True values
Prediction

Sarimax

What Could Have Influenced the Differences in Modeling?

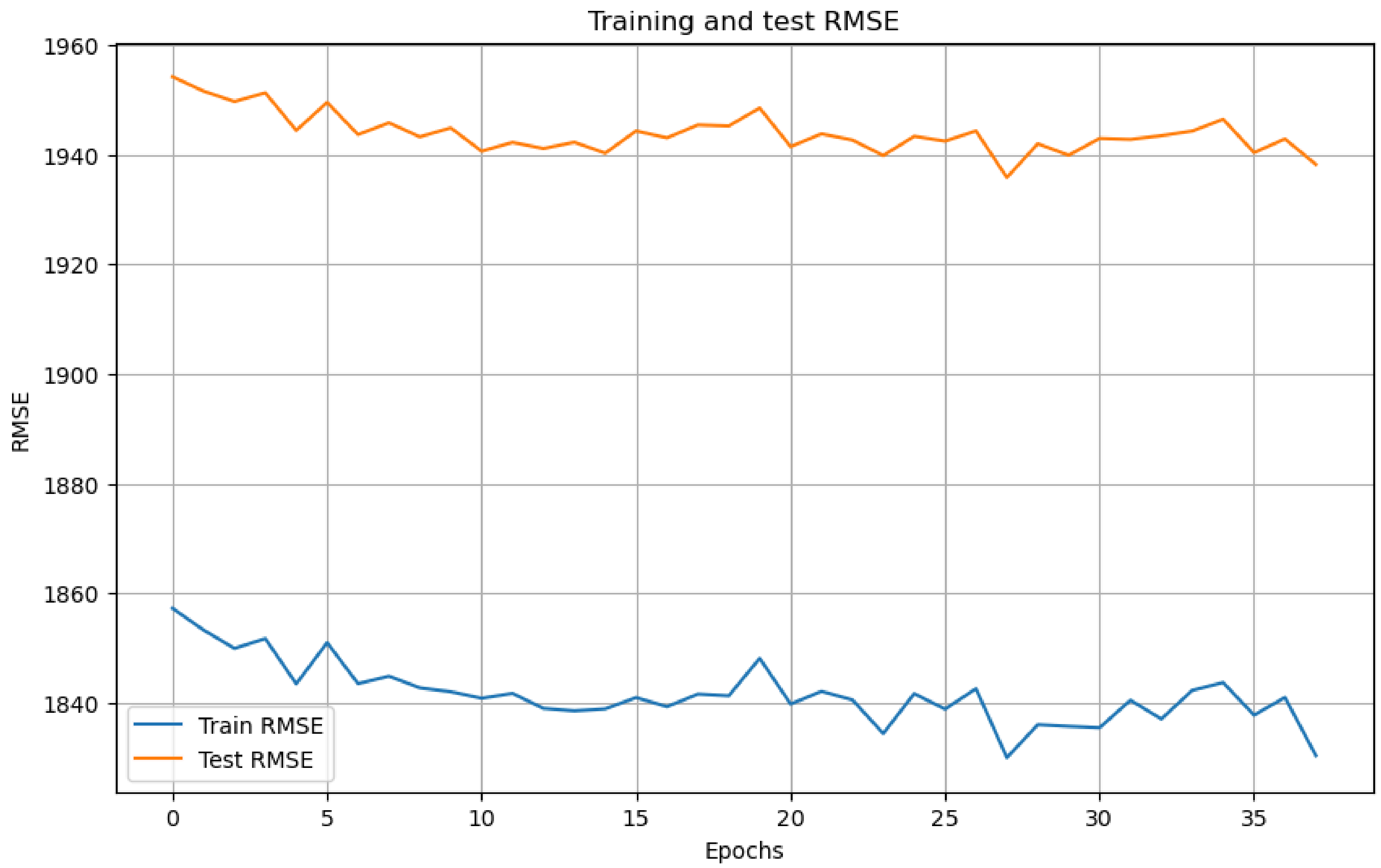
Data Variability Across Different Counties:
Counties with frequent large power outages may cause the model to have higher forecasting errors (higher RMSE), as SARIMAX models struggle to predict large, unpredictable events. On the other hand, counties with rare outages might result in more stable model predictions, leading to lower errors (lower RMSE).
Scale and Frequency of Events:
In regions with frequent extreme weather events (such as tornadoes and storms), SARIMAX models may face difficulties in forecasting such rare but significant occurrences, leading to higher RMSE values. In contrast, in less dynamic regions, the model might be more stable and accurate in its predictions.
Conclusions: The Model Depends on the County
Models are not universal and should be tailored to the specific data characteristics of each county. This suggests that the success of the model largely depends on the regional characteristics, such as the frequency and severity of events, which can significantly affect the quality and accuracy of forecasts.

TabNet – global model

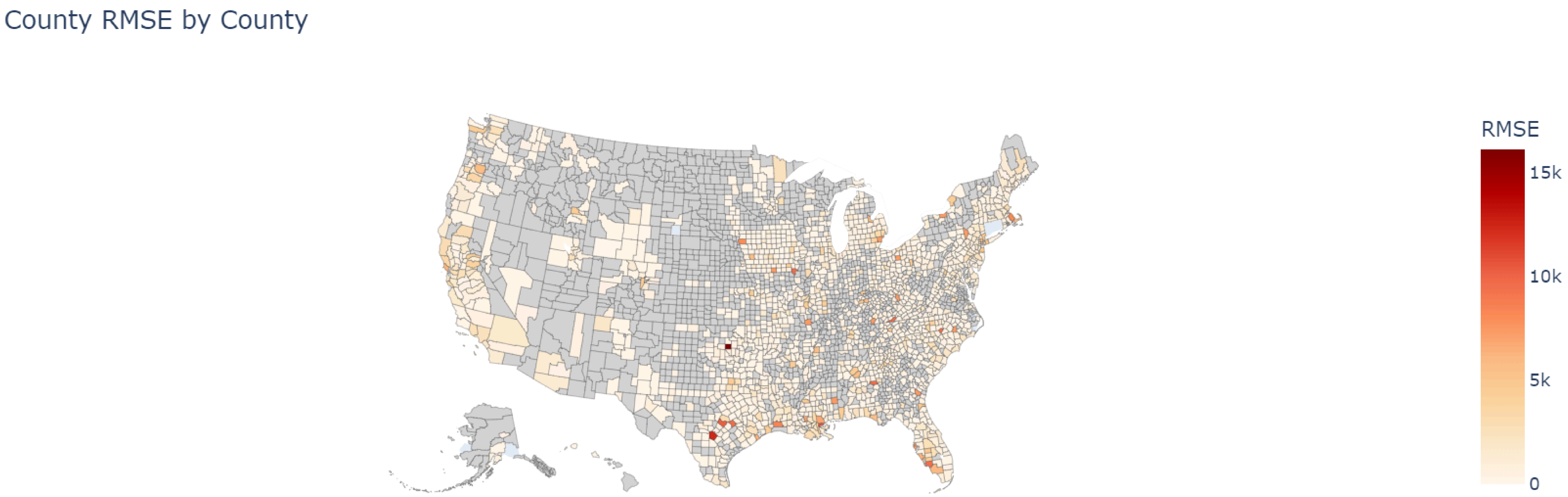
We can observe that the RMSE decreases during training on both the training and test sets. However, the results are not satisfactory.

PERFORMANCE (TEST)

RMSE:	1938
R2:	0.02

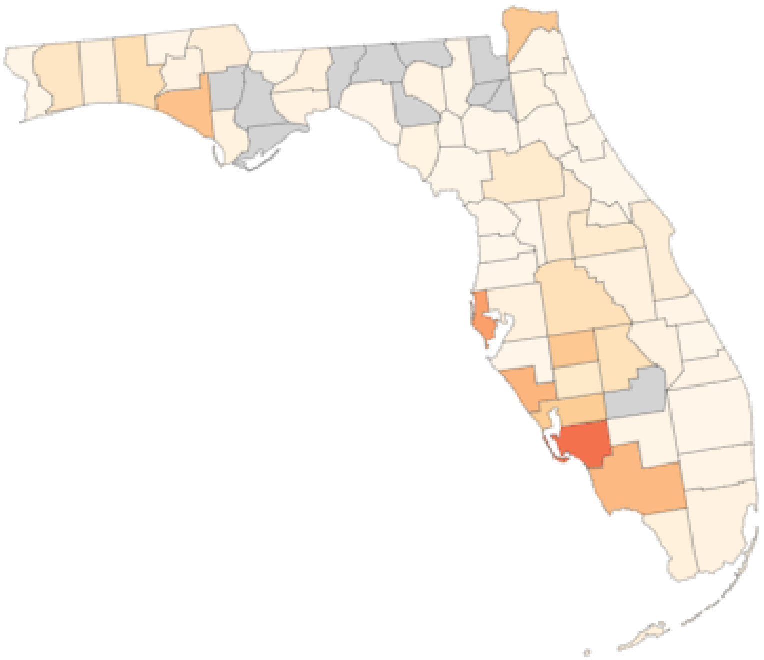


TabNet – global model

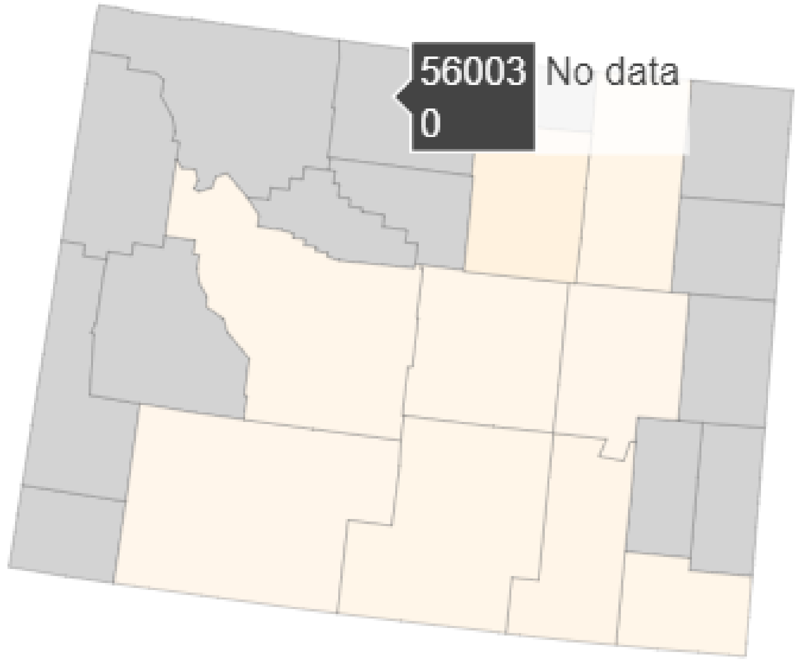


Tabnet by States - global model

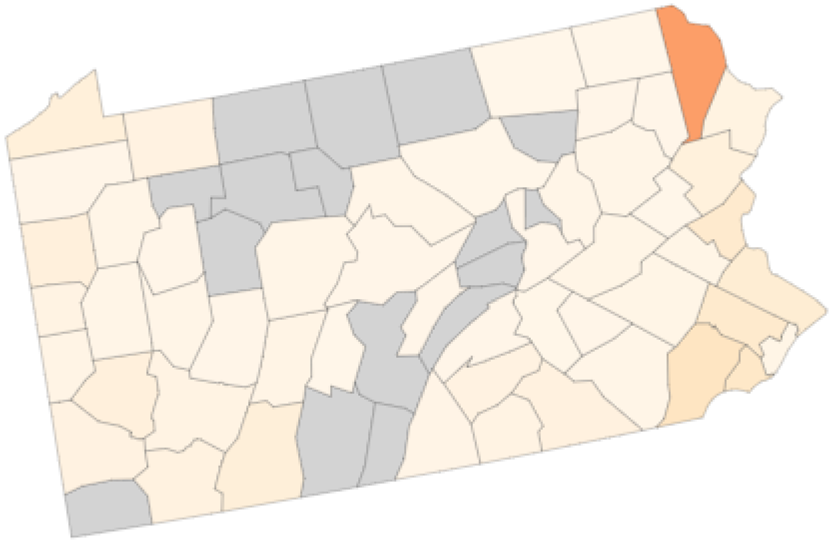
Florida



Wyoming



Pennsylvania



Tabnet by States – global model

Florida

PERFORMANCE

Mean RMSE:	1499
Max RMSE:	9462
Min RMSE:	113

Florida shows the weakest results, likely due to high population density, geographical diversity, and extreme weather events not well captured by the model

Wyoming

PERFORMANCE

Mean RMSE:	208
Max RMSE:	699
Min RMSE:	68

The model performs best in Wyoming, likely due to its small, uniform population and simpler climate conditions, despite a high rate of excluded counties

Pennsylvania

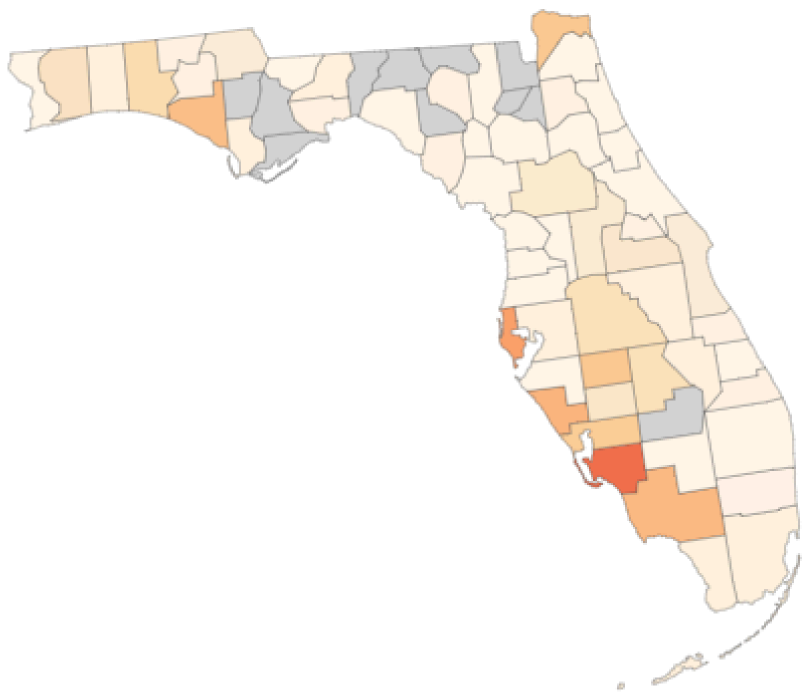
PERFORMANCE

Mean RMSE:	676
Max RMSE:	7332
Min RMSE:	167

Moderate performance with higher RMSE suggests the model struggled with Pennsylvania’s larger population and more complex, regionally varied data

Tabnet – Global model vs model for State

Global model



Model for Florida

Tabnet - Summary

In the interpretability phase, we aimed to demonstrate that modeling data across the entire United States is an inherently complex task with limited practical applicability. We achieved significantly better results when the model was trained and tailored for specific regions—such as individual states.

For example, even in the case of Florida, which was one of the poorest-performing states in the general model, training a dedicated model specifically for Florida led to a substantial improvement in performance.

THANK YOU