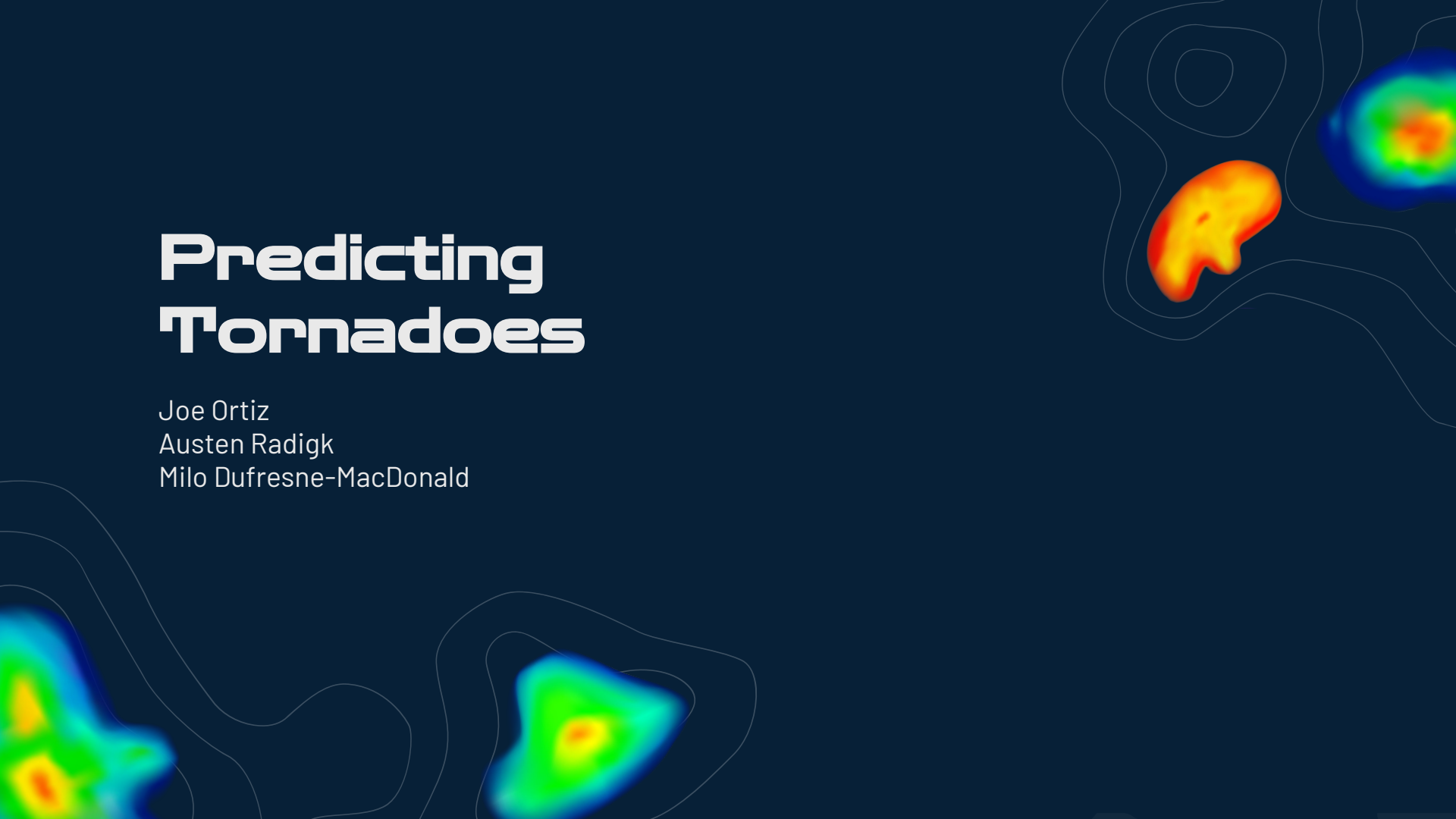


Predicting Tornadoes

Joe Ortiz

Austen Radigk

Milo Dufresne-MacDonald



Data Overview

68,868 tornadoes from 1950 - 2021



Temporal Data

Date
Time Zone
Time



Geographic Info

State
Starting Longitude/Latitude
Ending Longitude/Latitude
State FIPS Numbers
Total State(s) Affected



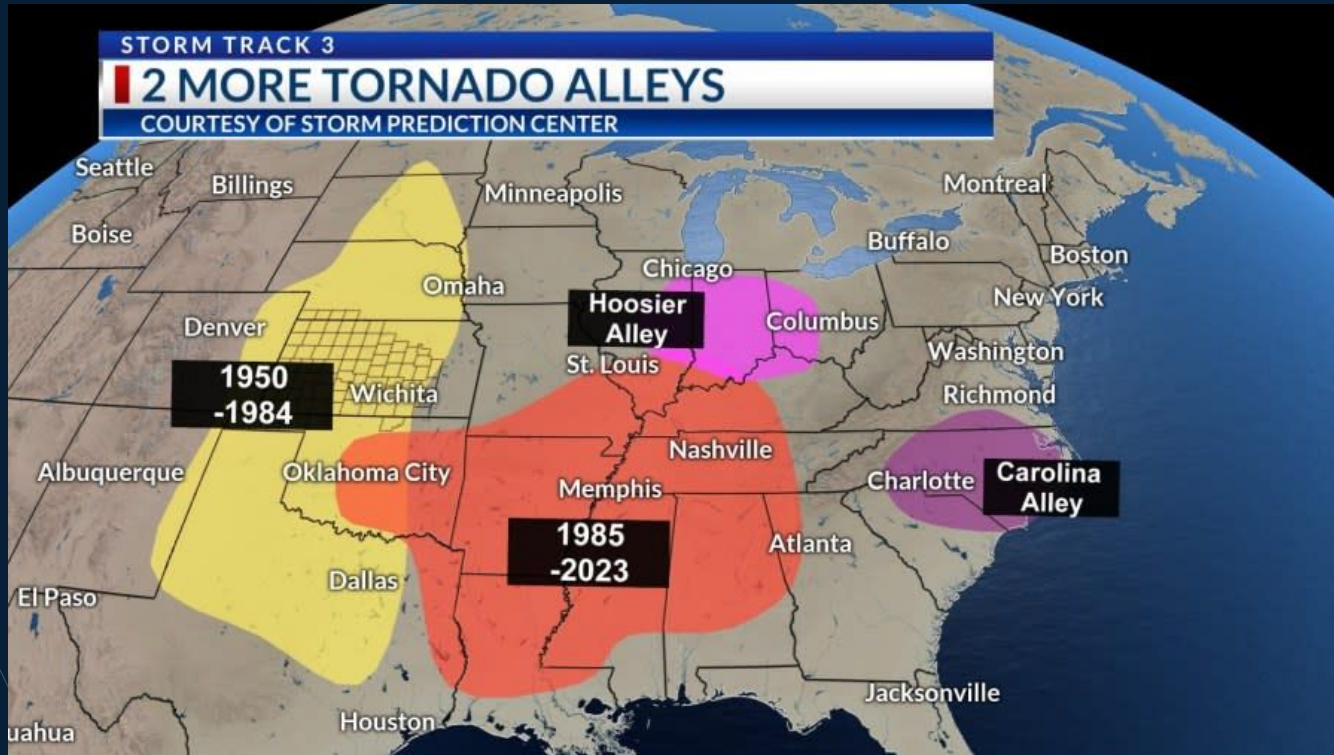
Tornado Characteristics

EF-Scale
Injuries
Fatalities
Estimated Property Loss
Estimated Crop Loss
Length (mi)
Width (yd)

Where do they occur?

Tornado Alley

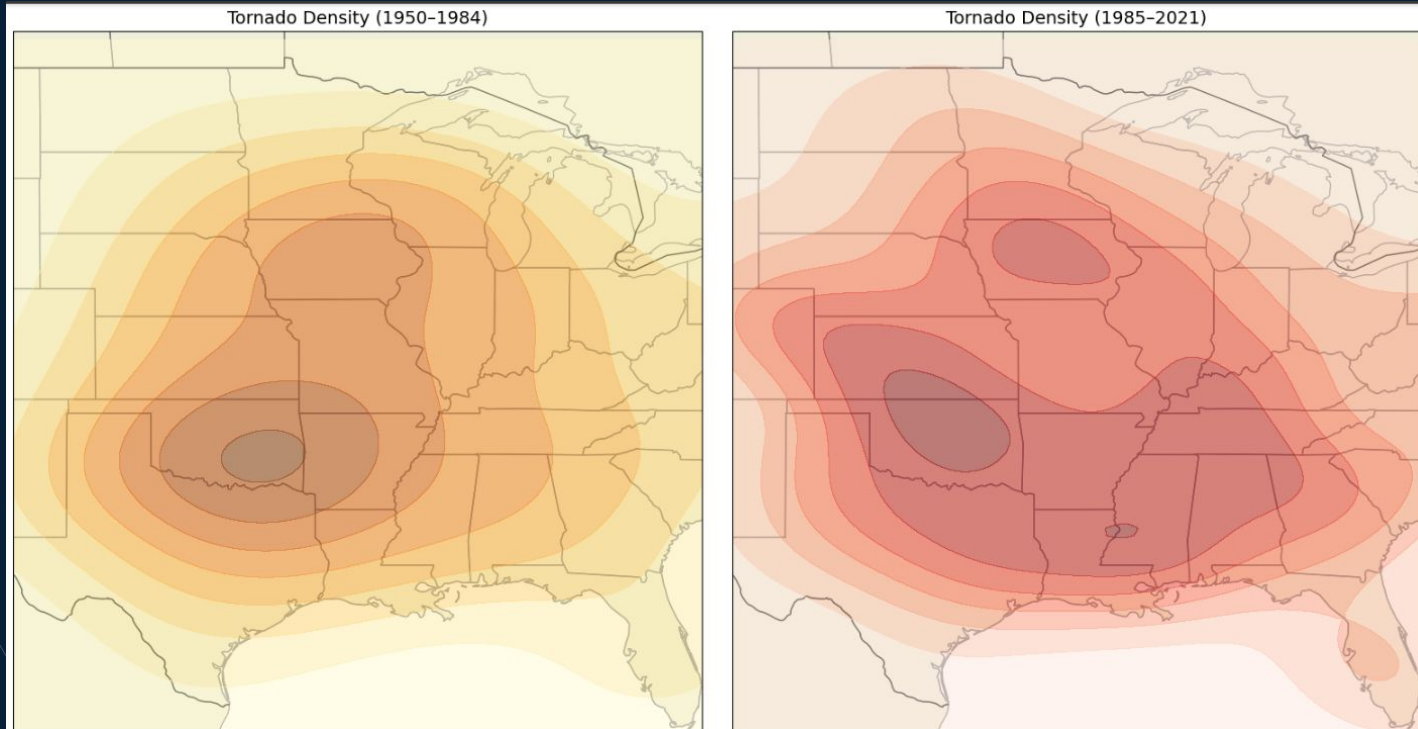
Loosely defined geographical location of the Central United States where tornadoes are the most frequent.



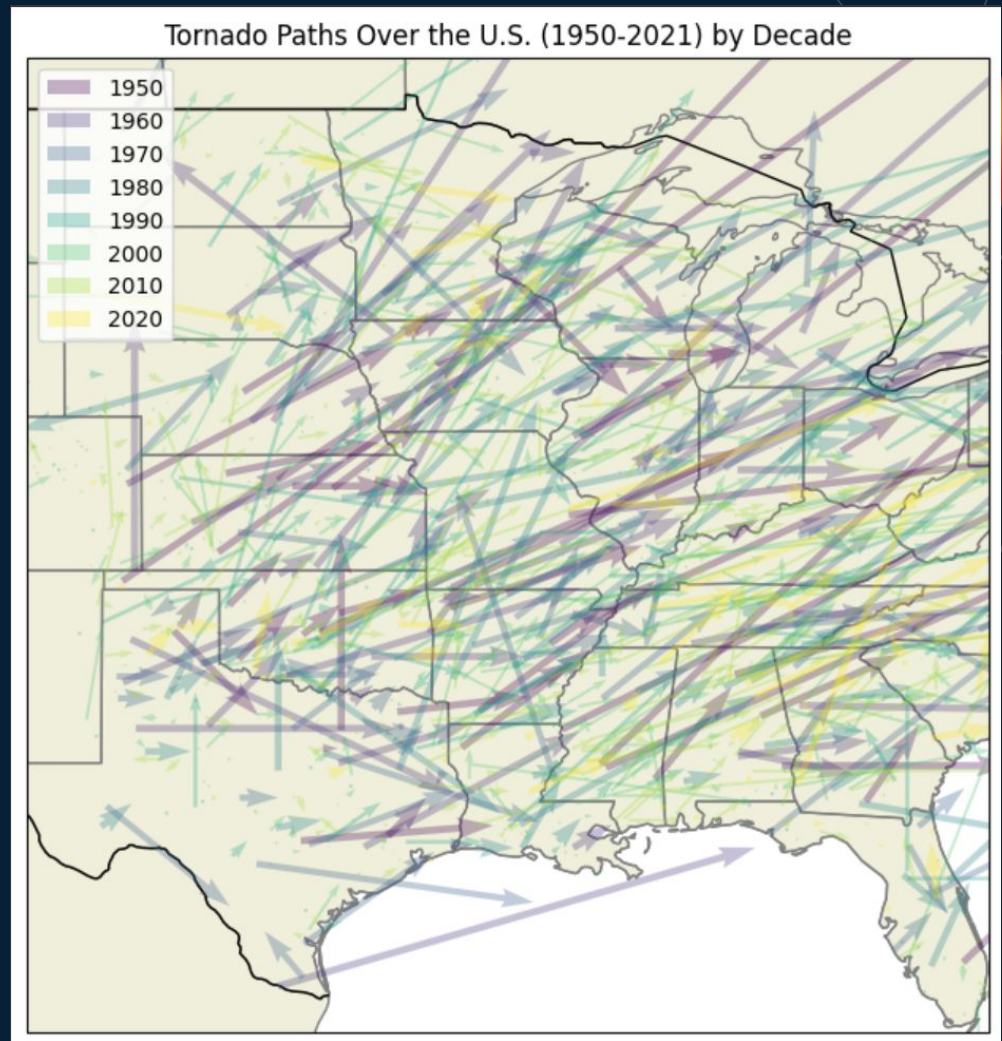
Where do they occur?

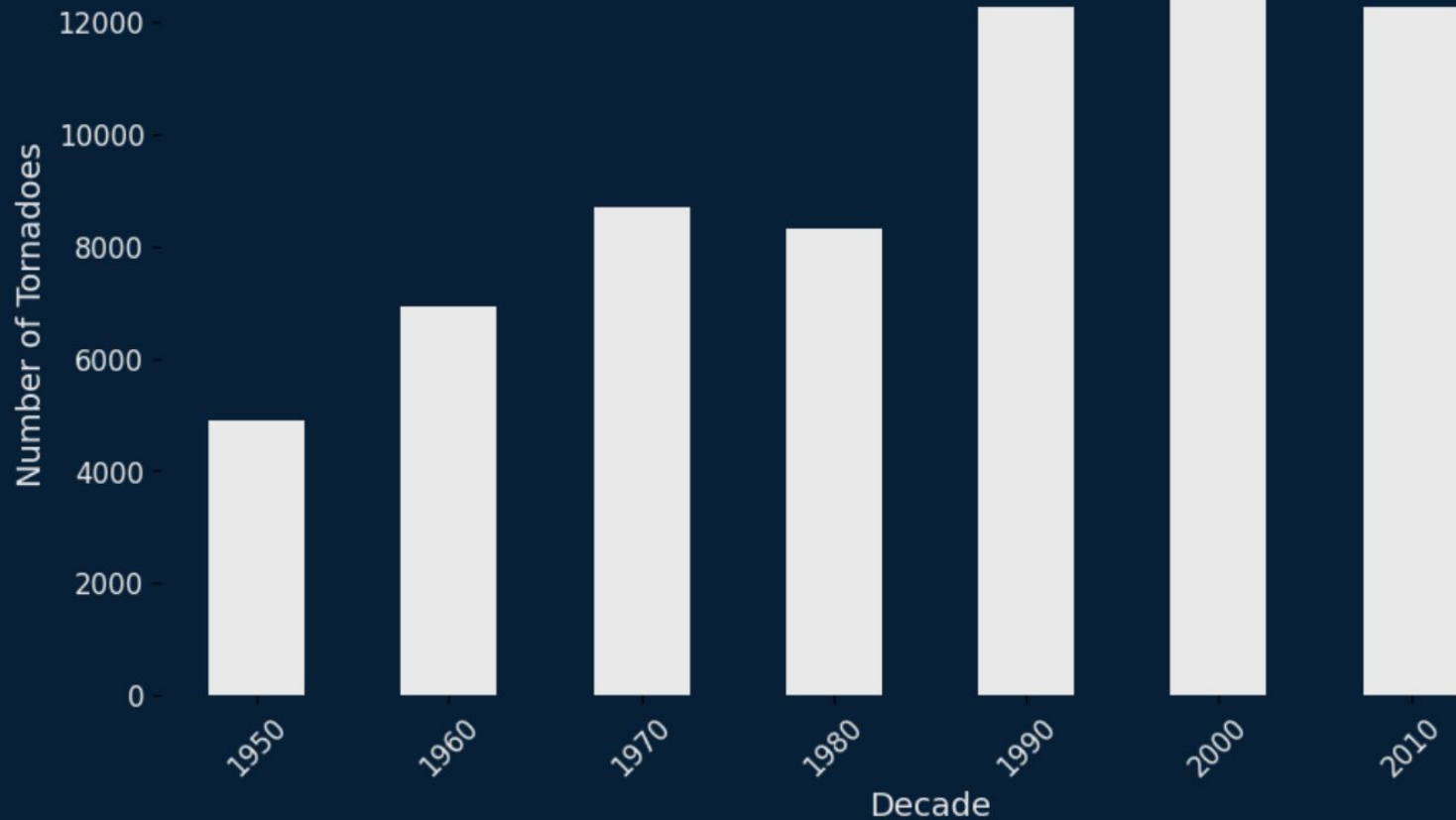
Tornado Alley

Loosely defined geographical location of the Central United States where tornadoes are the most frequent.



Tornado Paths 1950 - 2021





How are tornadoes measured?

The **Enhanced Fujita Scale (EF Scale)** is based on:

- Estimated Wind Speeds
- Related Damage

Tornado damage is compared to Damage Indicators and Degrees of Damage to better estimate the range of wind speeds the tornado likely produced.



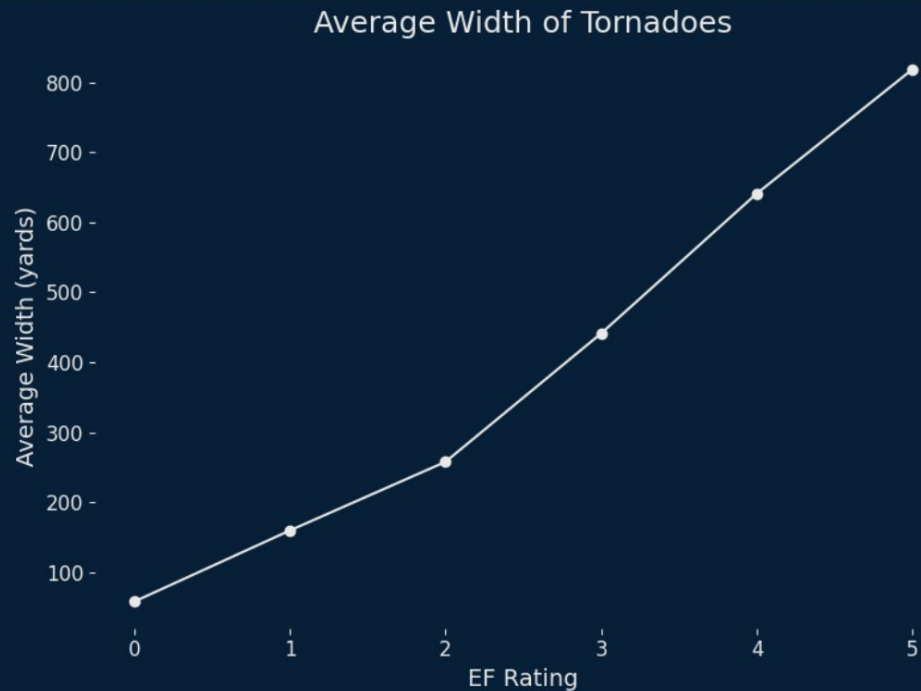
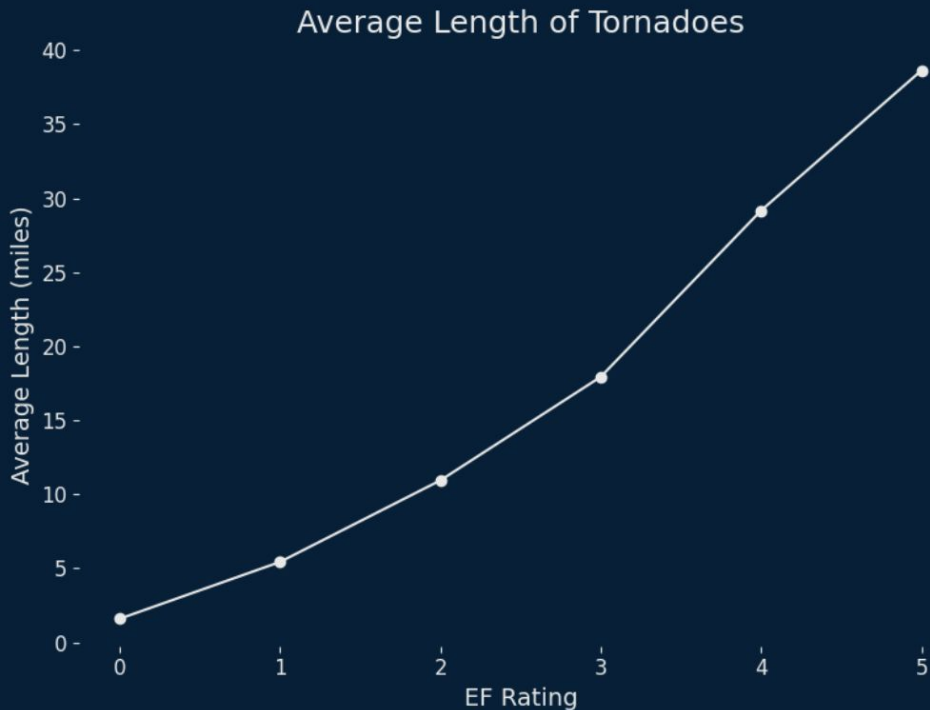
Source:

<https://www.weather.gov/eun/efscale>

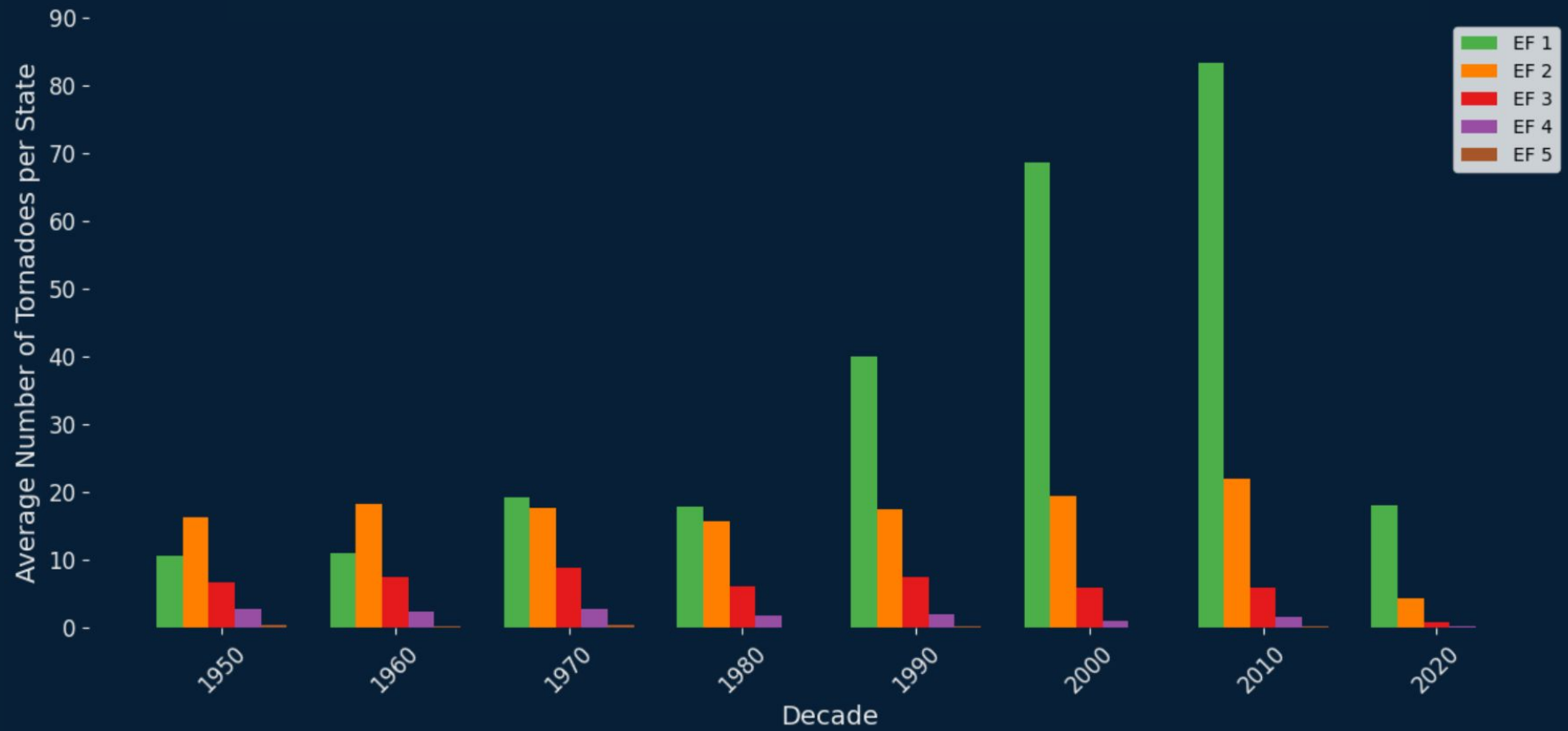
Tornado Metrics

Average Length - 5.56 mi

Average Width - 150.65 yd



EF-Ratings by



Modeling Objectives

Focusing on EF1 - EF4

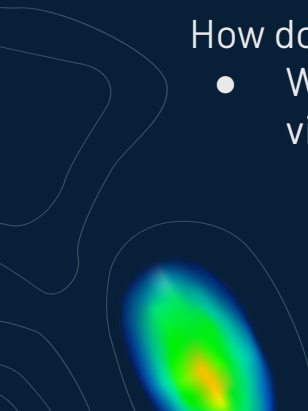
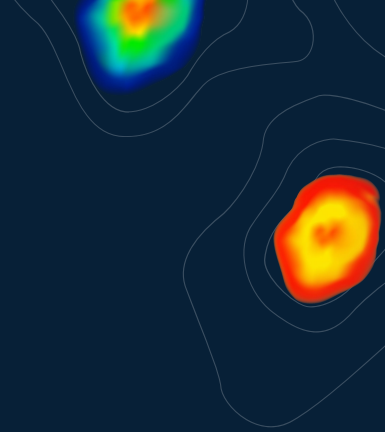
- EF5 → Unused due to improvements in technology.
- EF0 → Difficult to detect.

What characteristics can be used to predict tornado severity?

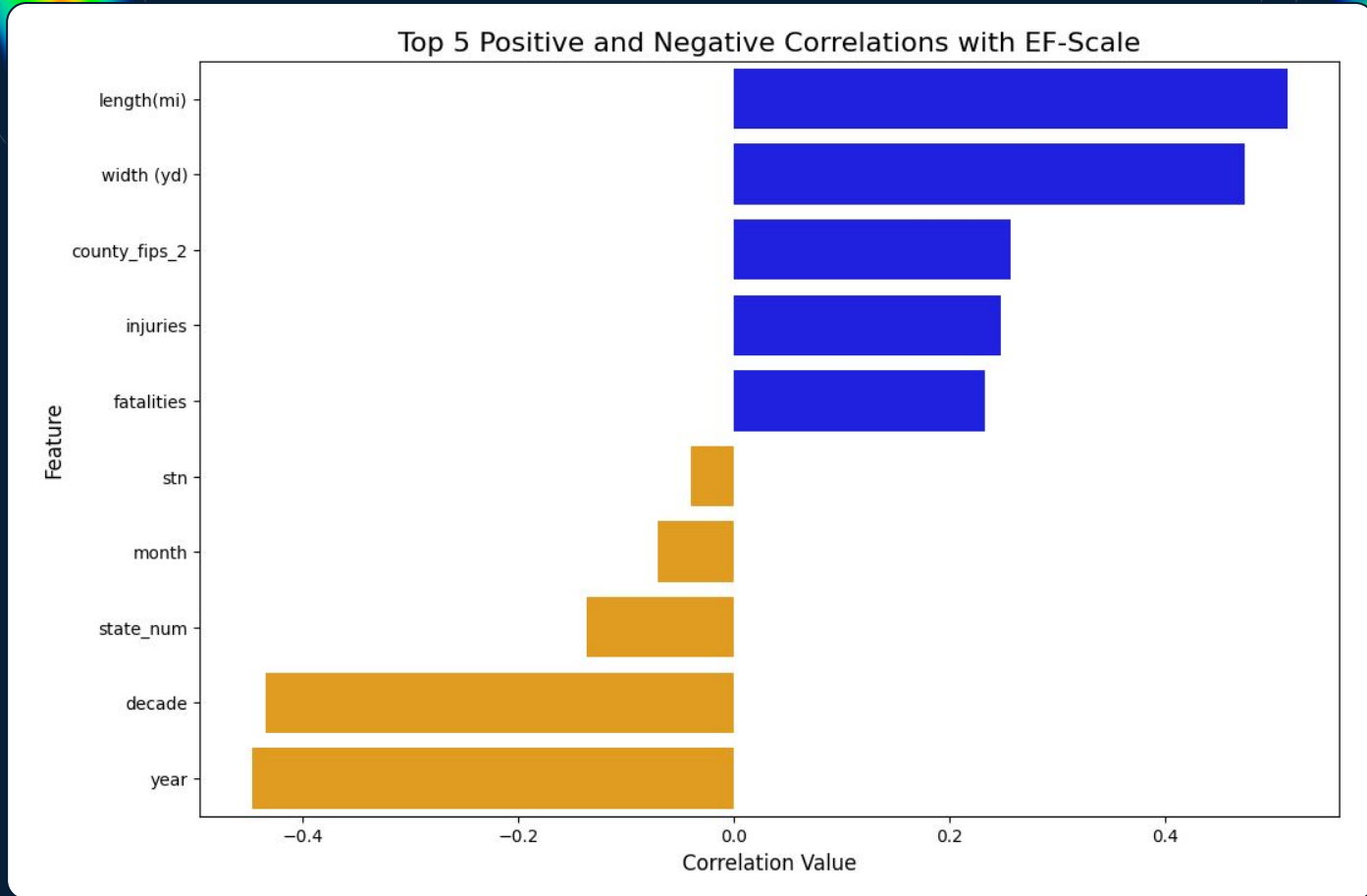
- What are the features of tornadoes that cause damage?
- What features make a tornado violent?

How do tornado characteristics differ between EF-ratings?

- What features are different between the majority of all damaging tornadoes and violent tornadoes?



Correlations

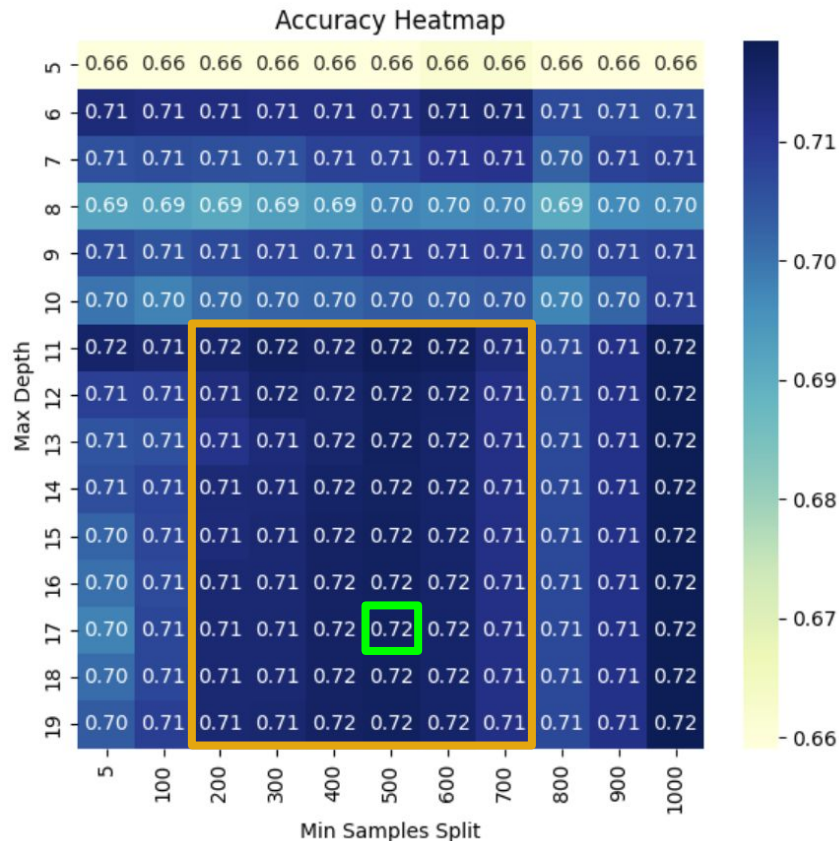


The background is a dark blue gradient. It features several white contour lines that form irregular, wavy shapes. Overlaid on these are three heatmaps. One heatmap in the top left shows a bright yellow-green spot. Another in the top right shows a bright yellow-orange spot. A third, larger heatmap in the middle right shows a bright green spot with a yellow center.

Predicting Non-Violent Tornadoes

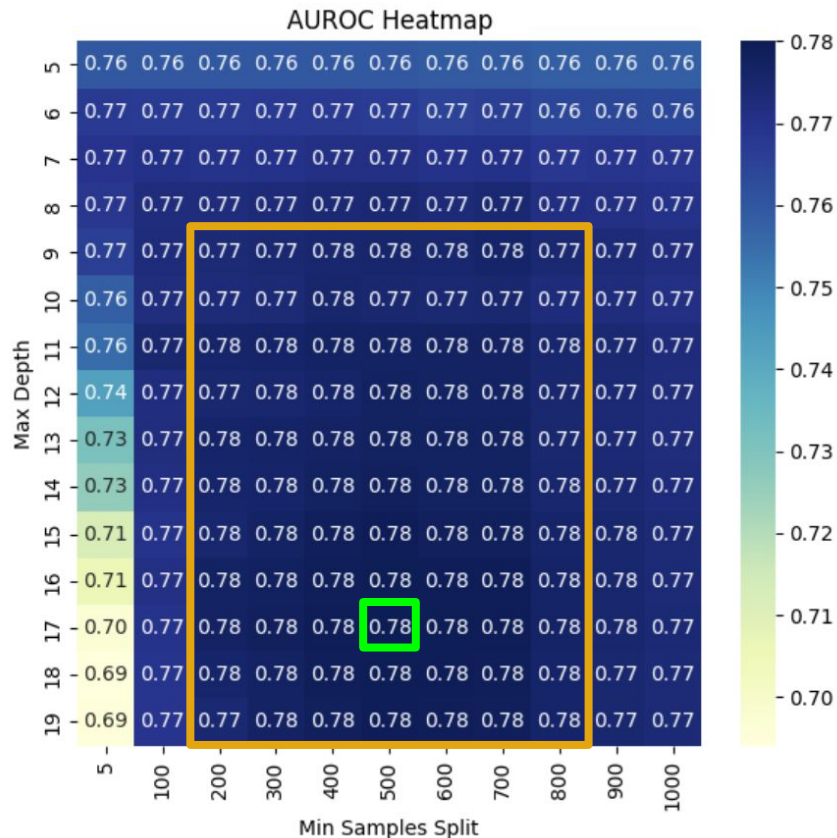
EF1 and EF2 Ratings

Decision Tree Accuracy



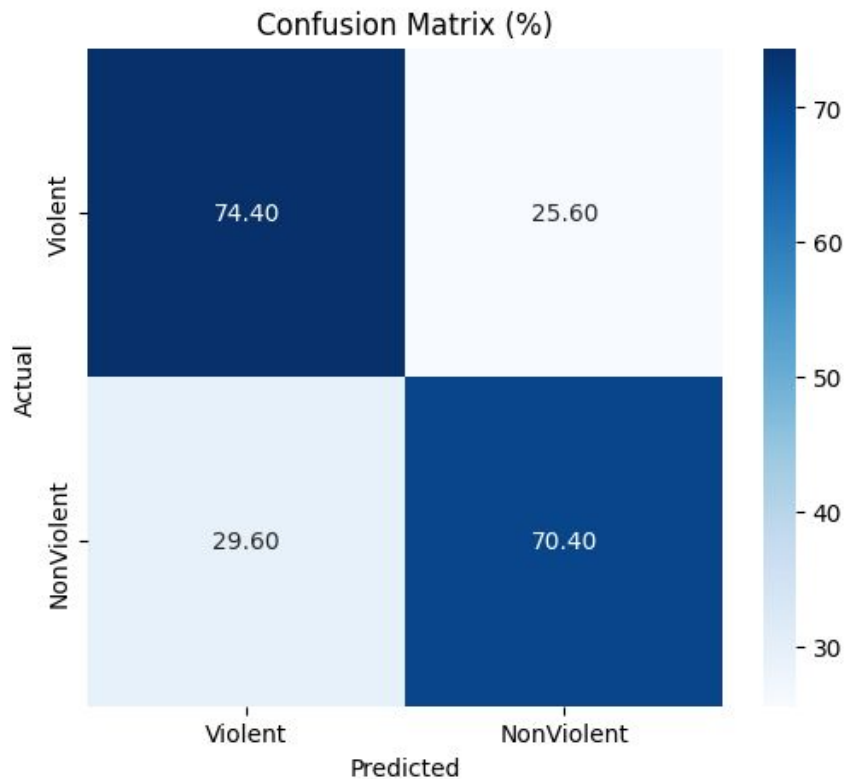
- **Predicting Non-Violent Ranged**
 - EF1 - EF2
- **Best Hyperparameters**
 - Max Depth: **17**
 - Min Samples Split: **500**
- **Best Metric**
 - Accuracy: **0.72**

Decision Tree Auroc



- **Predicting Non-Violent Ranged**
 - EF1 - EF2
- **Best Hyperparameters**
 - Max Depth: **17**
 - Min Samples Split: **500**
- **Best Metric**
 - AUROC: **0.78**

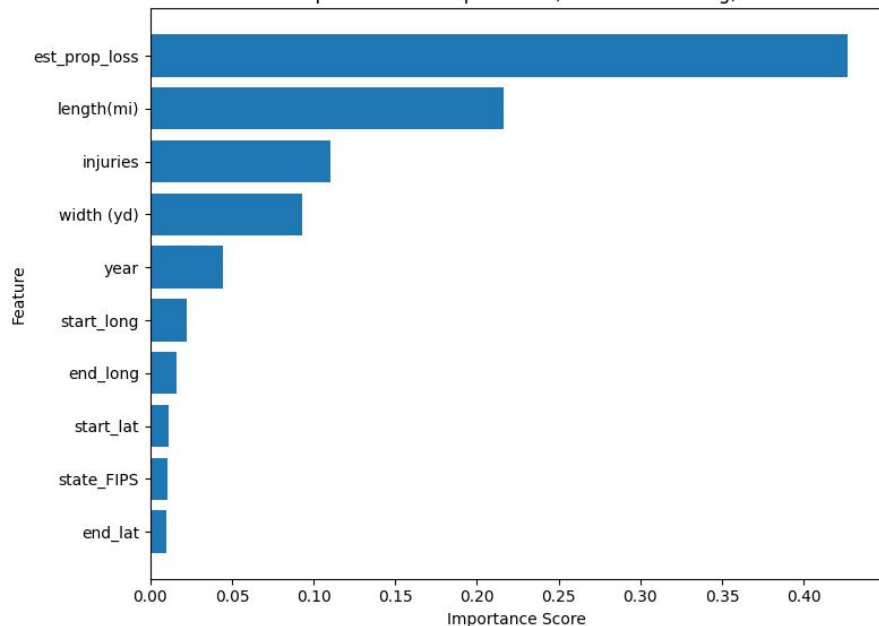
Gradient Boosting



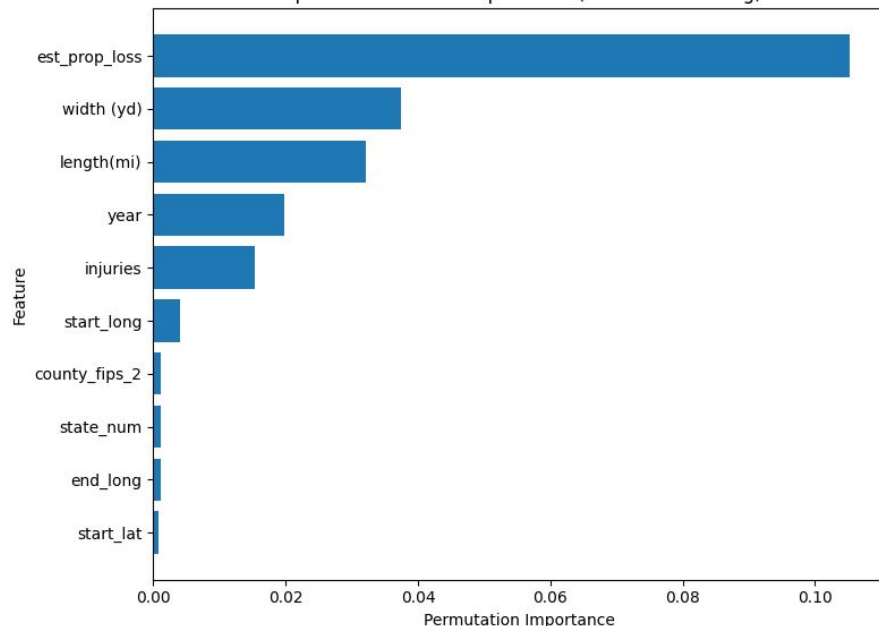
- **Predicting Non-Violent Ranged**
 - EF1 - EF2
- **Hyperparameters Used**
 - Max Depth: **17**
 - Min Samples Split: **500**
- **Metrics**
 - Accuracy: **0.73**
 - AUROC: **0.80**

Gradient Boosting Features

Top 10 Feature Importance (Gradient Boosting)



Top 10 Permutation Importance (Gradient Boosting)



- **Permutation**

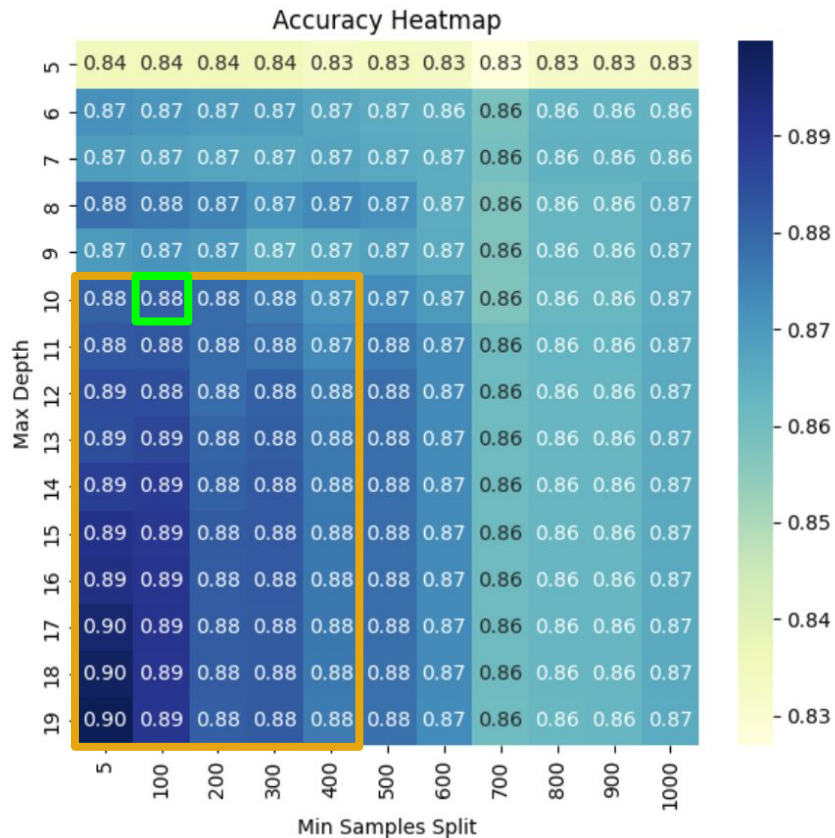
- Measures how much the model's performance degrades when feature values are randomly shuffled (permuted).

The background is a dark blue gradient. It features several white contour lines that form irregular, wavy shapes. Overlaid on these are three heatmaps. One heatmap in the top left shows a bright yellow and orange core surrounded by green and blue. Another in the top right shows a bright yellow and orange core surrounded by green and blue. A third, larger heatmap in the bottom right shows a bright yellow and orange core surrounded by green and blue.

Predicting Violent Tornadoes

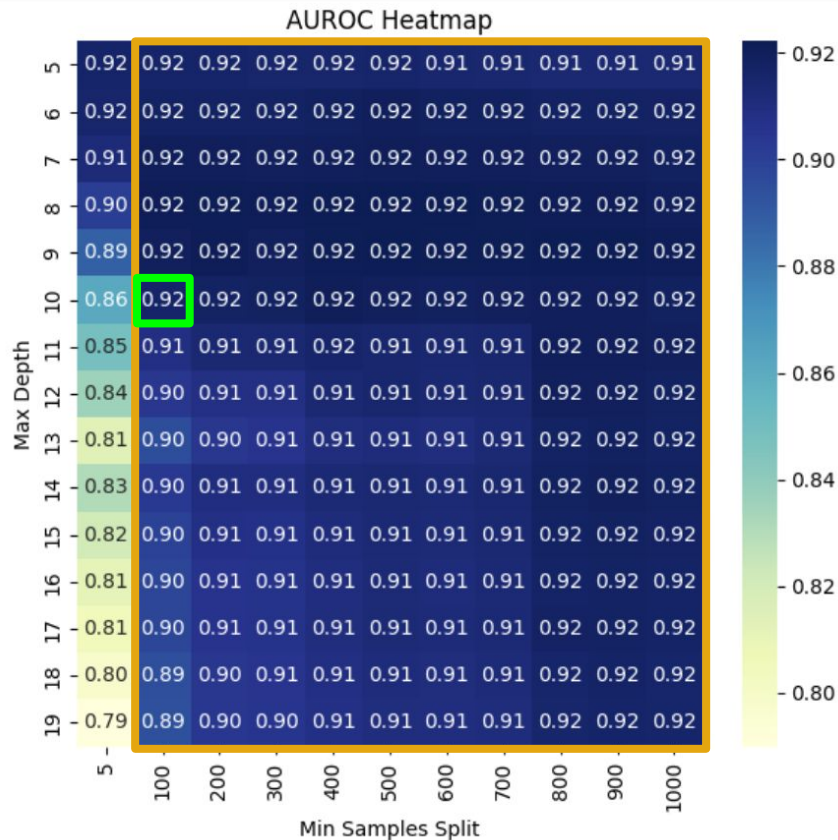
EF3 and **EF4** Ratings

Decision Tree Accuracy



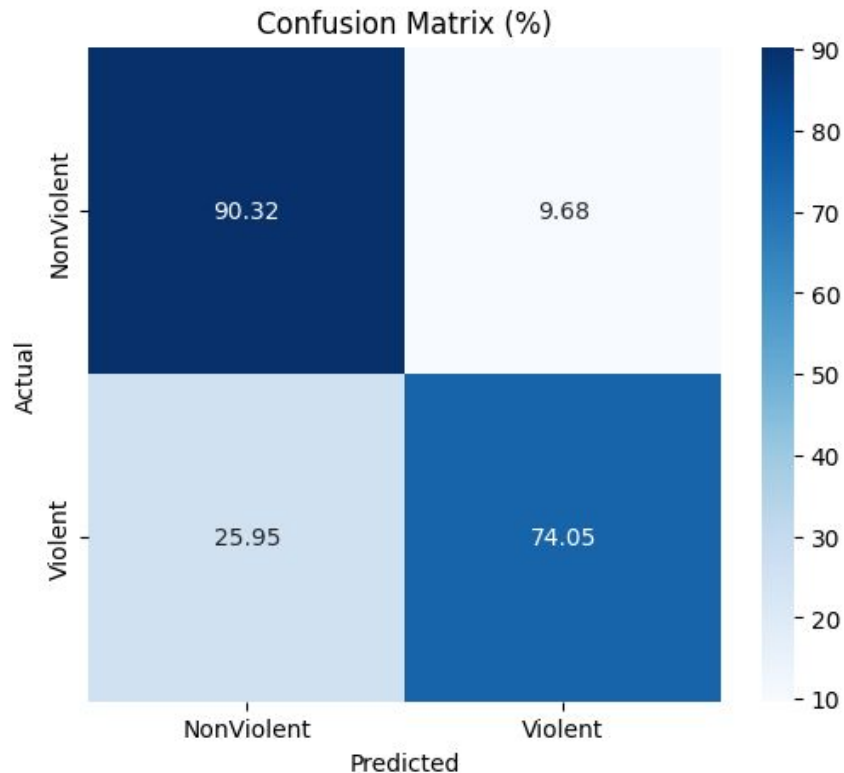
- **Predicting Violent Ranged**
 - EF3 - EF4
- **Best Hyperparameters**
 - Max Depth: **10**
 - Min Samples Split: **100**
- **Best Metric**
 - Accuracy: **0.88**

Decision Tree AUROC



- **Predicting Violent Ranged**
 - EF3 - EF4
- **Best Hyperparameters**
 - Max Depth: **10**
 - Min Samples Split: **100**
- **Best Metric**
 - AUROC: **0.92**

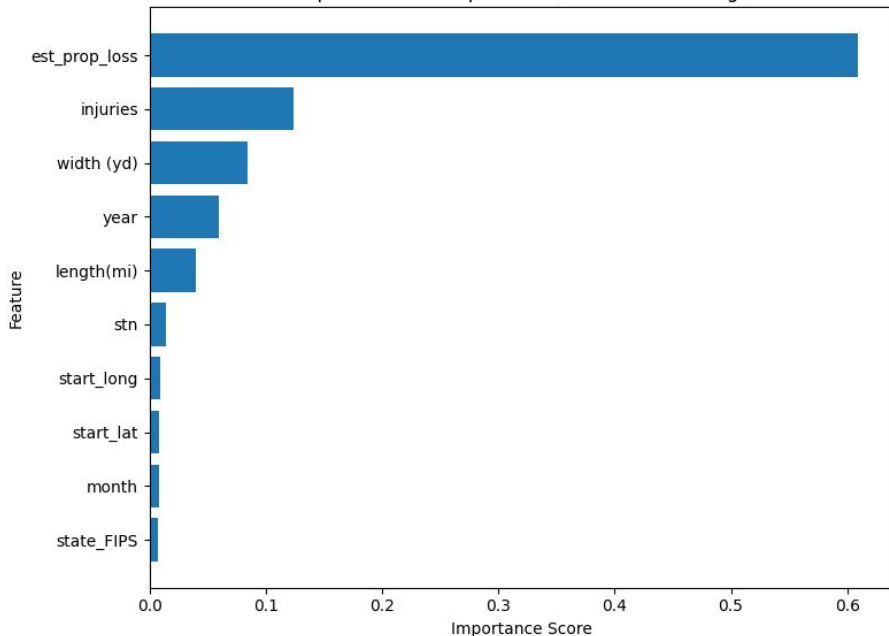
Gradient Boosting



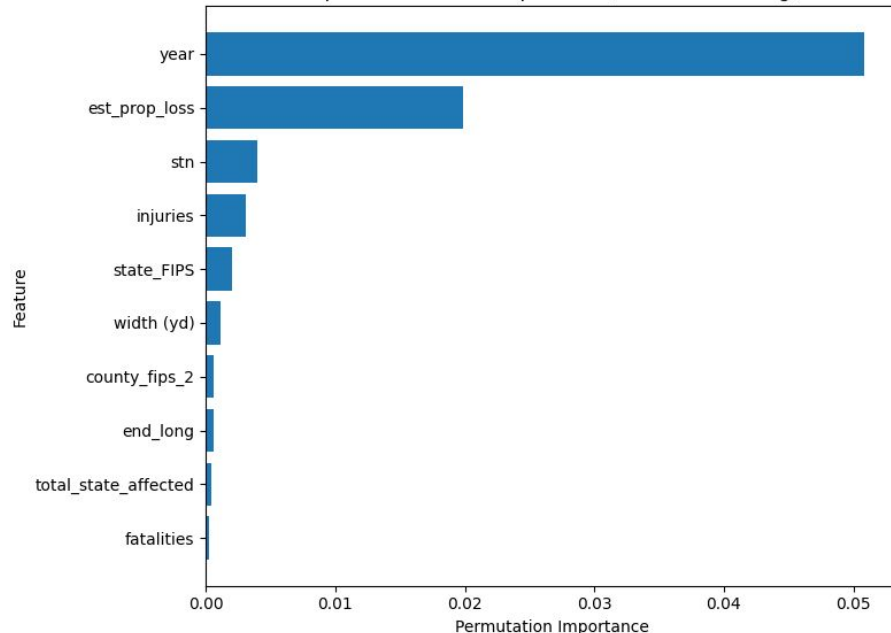
- **Predicting Violent Ranged**
 - EF3 - EF4
- **Hyperparameters Used**
 - Max Depth: **10**
 - Min Samples Split: **100**
- **Metrics**
 - Accuracy: **0.89**
 - AUROC: **0.93**

Gradient Boosting Features

Top 10 Feature Importance (Gradient Boosting)



Top 10 Permutation Importance (Gradient Boosting)



- **Mismatch?**

- Features are what the models uses to split the data and reduce impurity
- Permutations show what's important to the model

Modeling Takeaways

Predicting EF1 - EF2 Tornadoes (in order)

- Property Loss
- Length (mi) / Width (yd)
- Year
- Injuries

Predicting EF3 - EF4 Tornadoes (in order)

- Year
 - Potentially Climate / Technology Related
- Property Loss
- State Number / State FIP
 - Which State the Tornado Took Place In
- Injuries

Models are predicting tornadoes with 73% - 89% accuracy