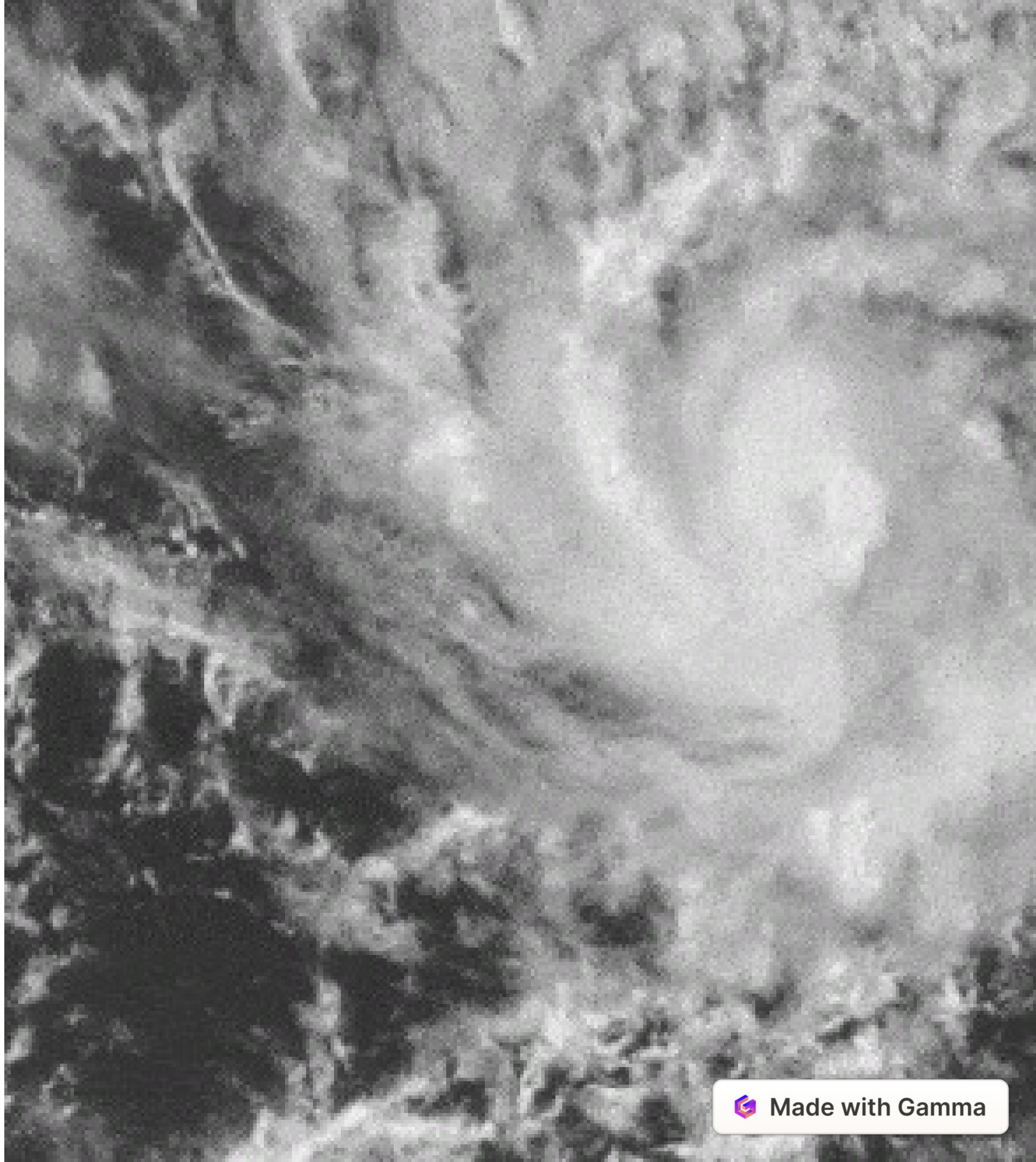
A Multi-Faceted Look at Storms:

Trends, Impacts, and Public Sentiments

Aleksandra Kutz
Daniel Allen
Anvita Iyer
Krishan Pandey
Hardeep Gumber



Made with Gamma

Key Topics: Hurricane Analysis and Impact

1 Project Overview

Data sources

2 Human Cost & Hurricane Trends

Financial loss and Fatalities (Krishan)

Analysis of Storm Trends (Anvita)

3 Storm path visualization

Hurricane Path and Precipitation (Hardeep)

4 Prediction

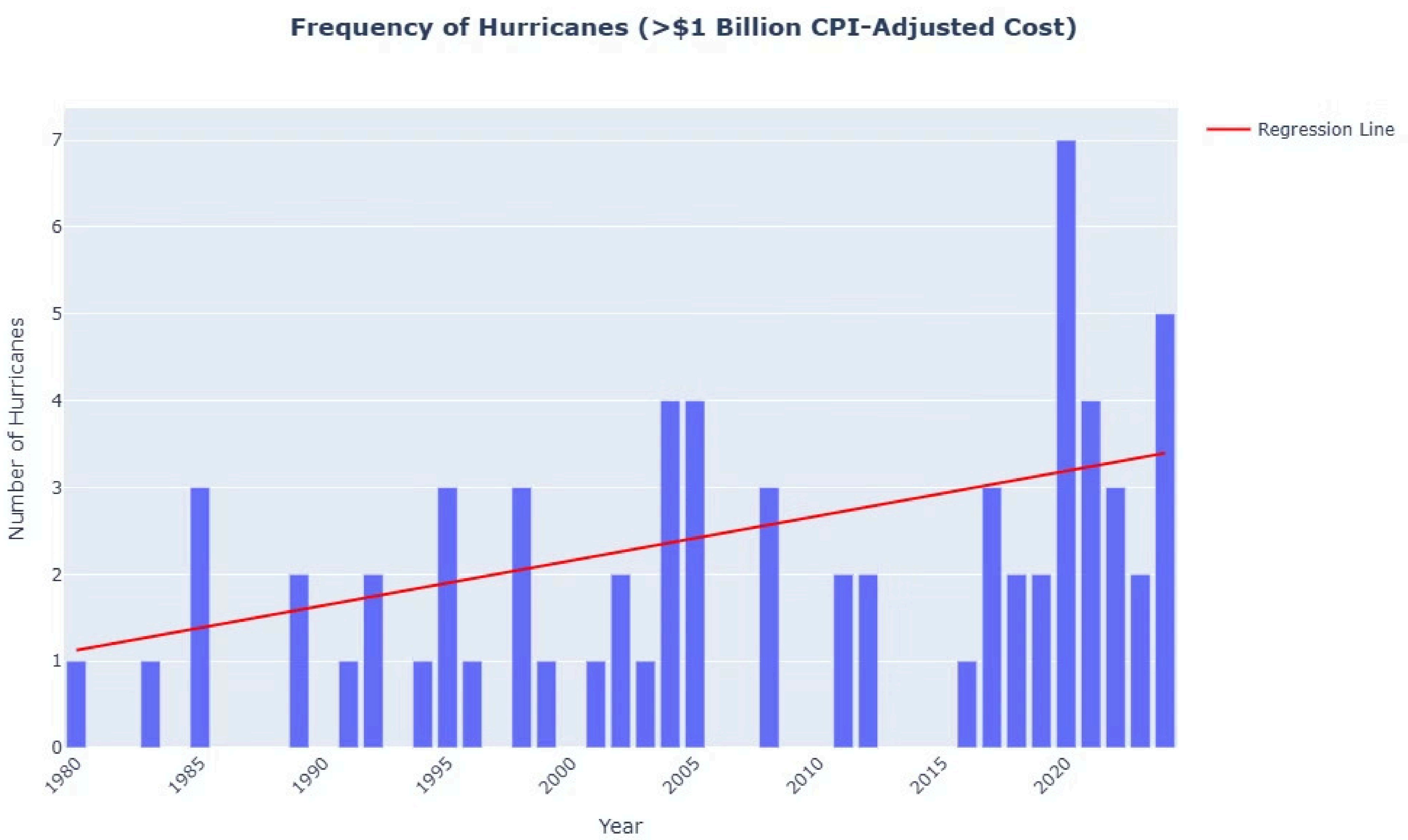
Prediction Models for Damages & Fatalities (Alex & Daniel)

5 Public Sentiment

Public Sentiments (Alex)

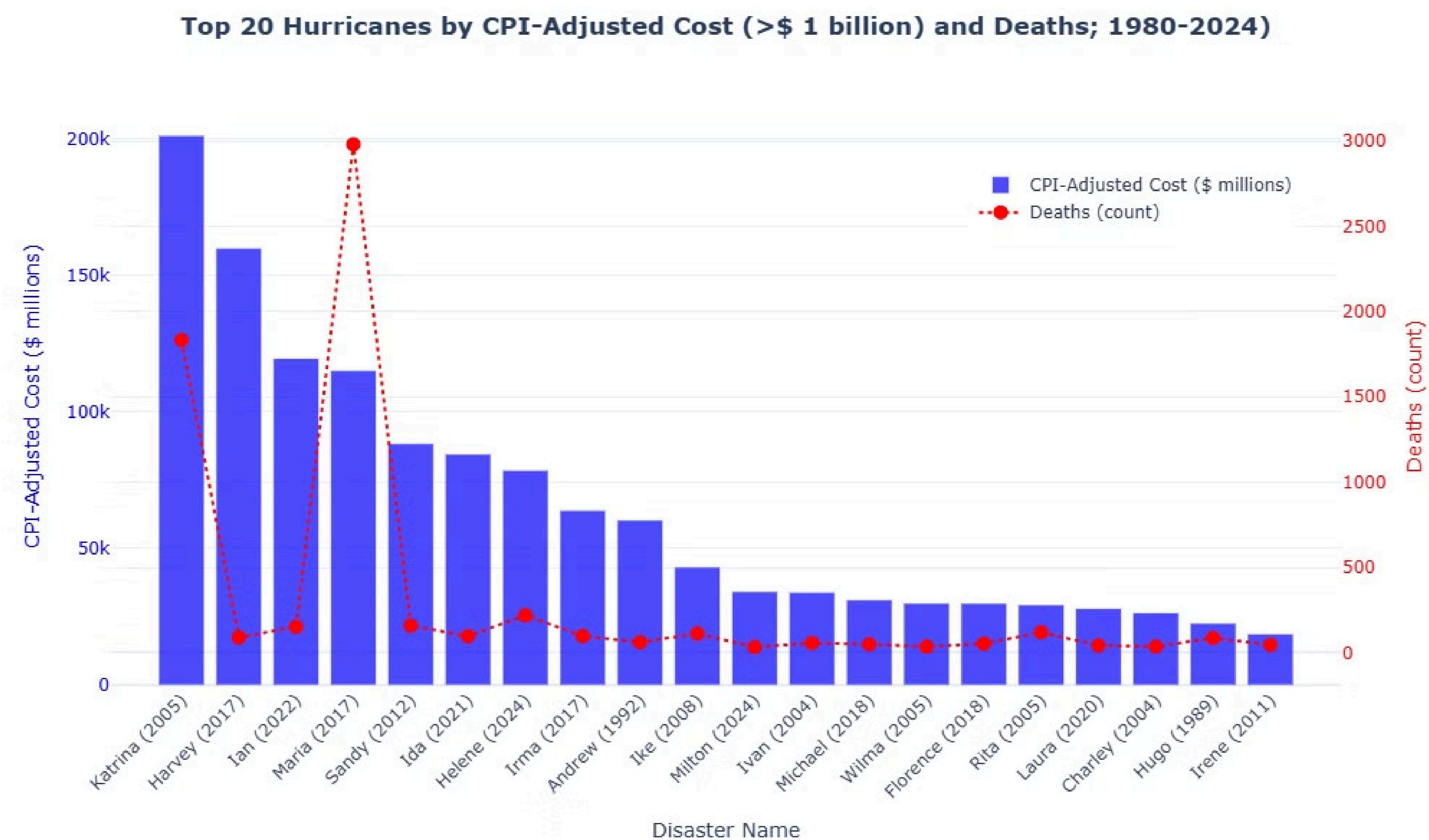
Made with Gamma

Frequency of Hurricanes (>$1 billion cost; 1980-2024)



Made with Gamma

Correlation of Top 20 >$1 Billion Dollar Hurricanes with Deaths (1980-2024)



- Even though Katrina (2005) has been the most expensive hurricane event, the most fatalities happened during hurricane Maria (2017) in Puerto Rico.
- It highlights that damages and fatalities are dependent on multiple factors, not just the hurricane severity.

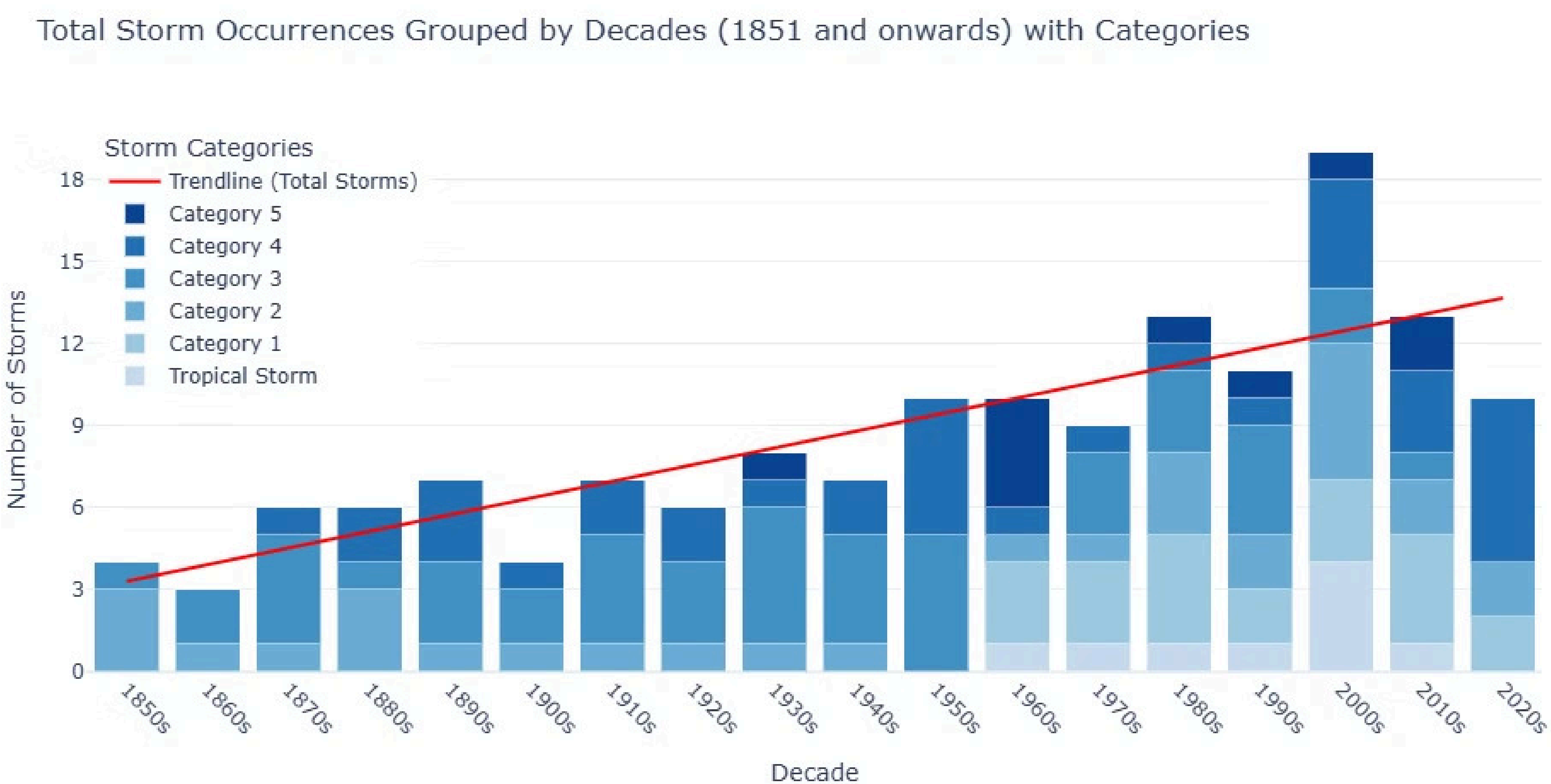
Made with Gamma

Hurricane Categories - Saffir-Simpson Hurricane Wind Scale

Category	Max sustained wind	Examples
1	74-95 mph (64-82 knots)	Humberto (2007)
2	96-110 mph (83-95 knots)	Zeta, Sally (2020)
3	111-129 mph (96-112 knots)	Ian (2022)
4	130-156 mph (113-136 knots)	Harvey, Maria (2017)
5	>= 157 mph (>= 137 knots)	Katrina (2005), Andrew (1992)

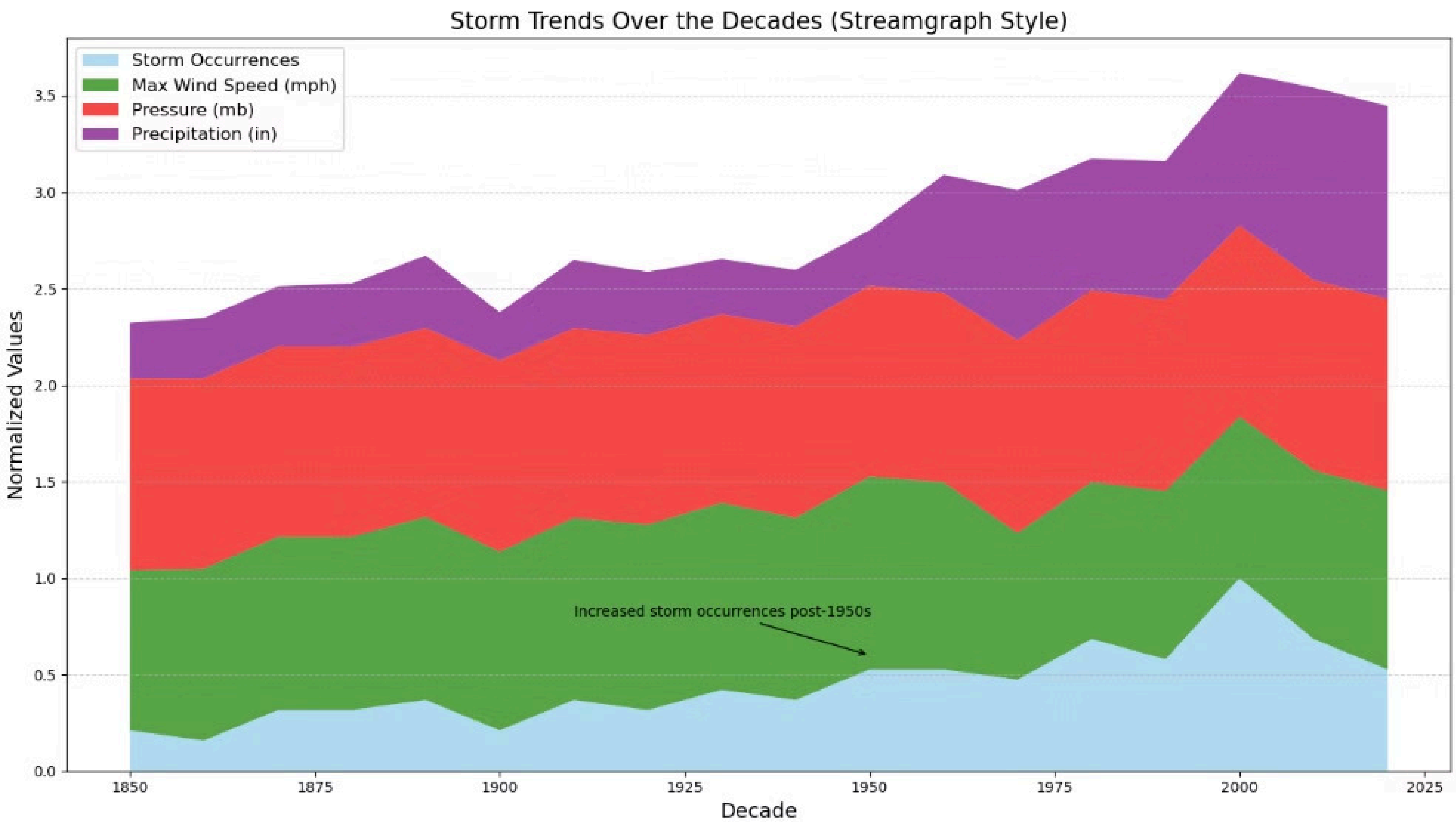
Made with Gamma

Storm Occurrences over the Decades



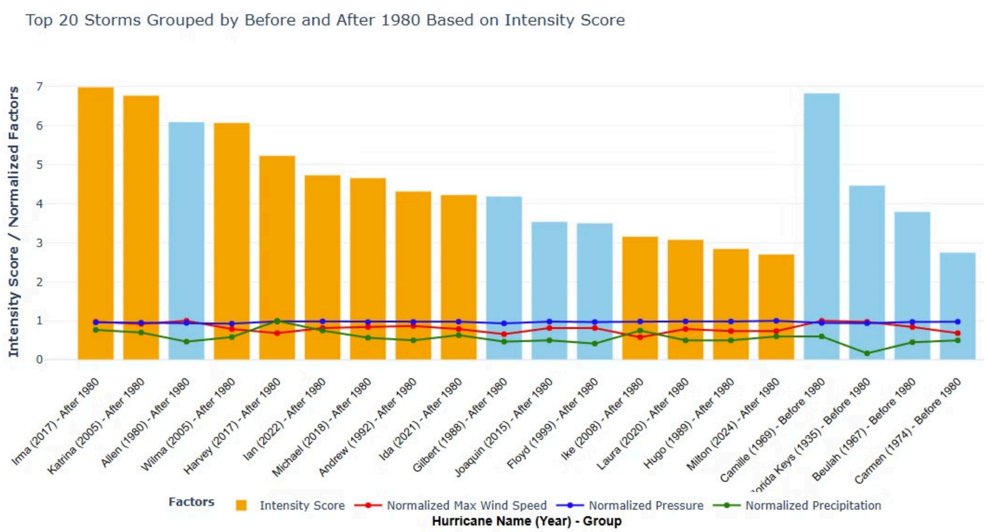
Made with Gamma

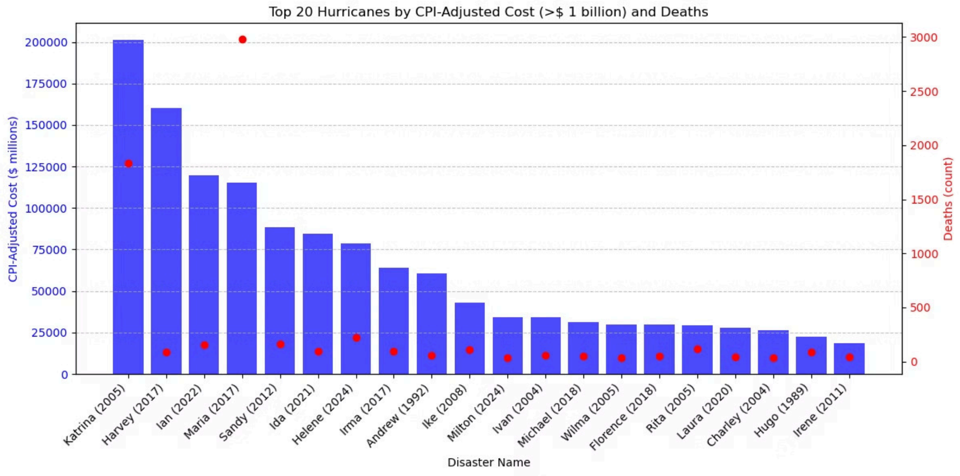
Storm Characteristics over the Decades



Made with Gamma

Top 20 Most Intense Storms

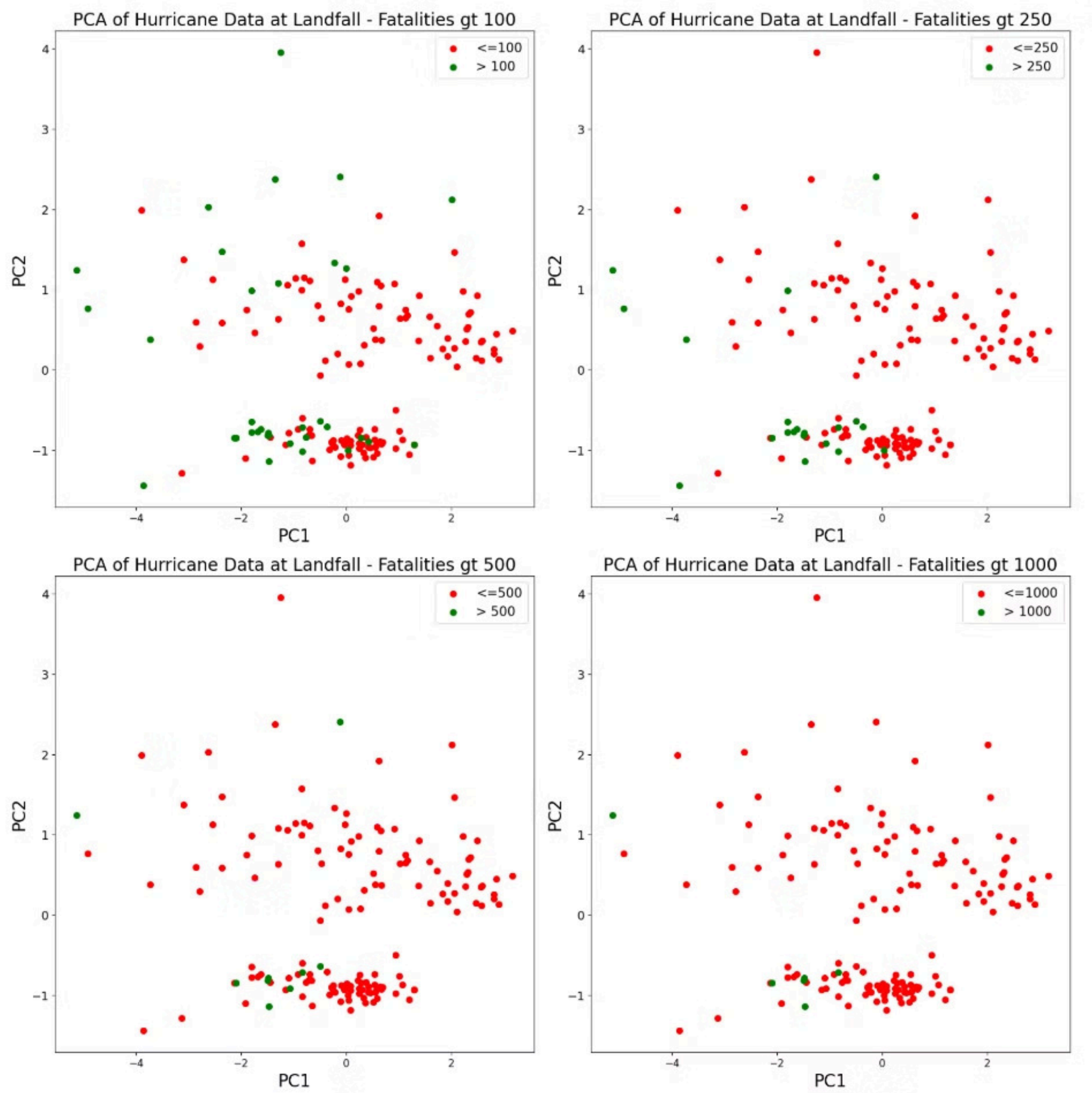




Made with Gamma

PCA: Fatalities from Hurricane Parameters

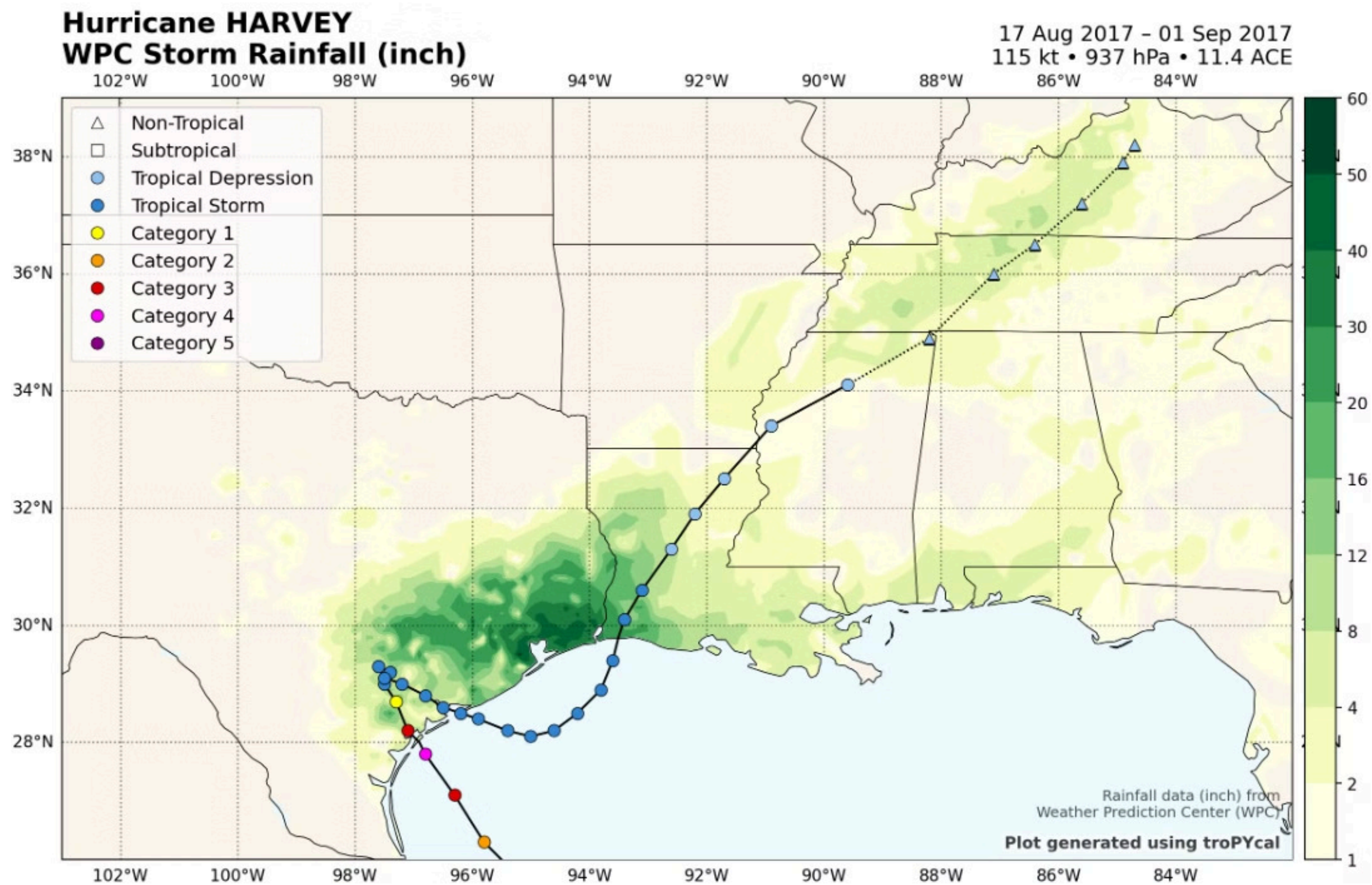
Max wind speed, pressure, storm surge and precipitation



Made with Gamma

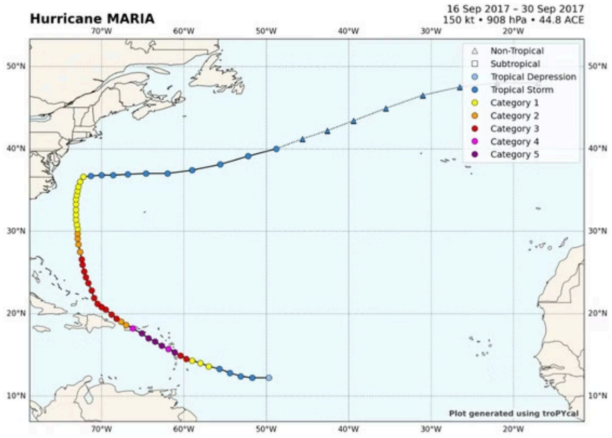
Hurricane Harvey's Path and Precipitation

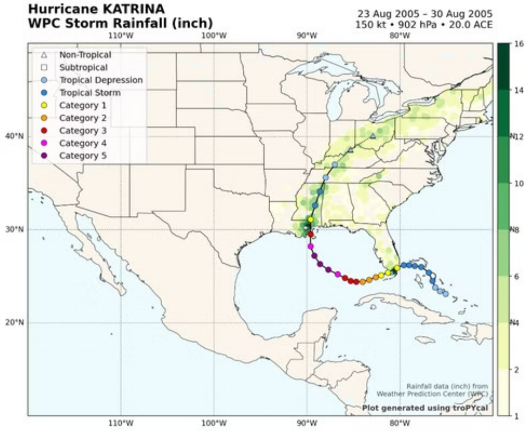
- **Landfall**: Rockport, Texas
- **Category**: 4
- **Rainfall**: up to 60 inches
- **Storm surge**: 9 feet
- **Fatalities**: 175
- **Economic damage**: $125 billion
- **Infrastructure**: Over 300,000 buildings

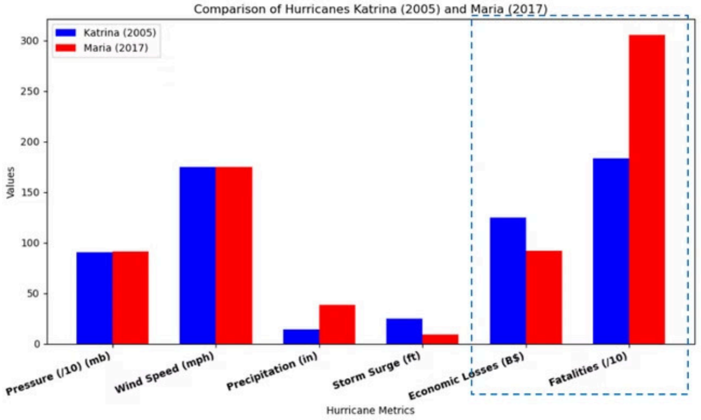


Made with Gamma

Island (Maria-4) vs. Mainland (Katrina -5) Hurricane Analysis







The number of fatalities is influenced by the landfall location and the effectiveness of resource management.

Made with Gamma

Sentiment - What is that and Why does it matter?

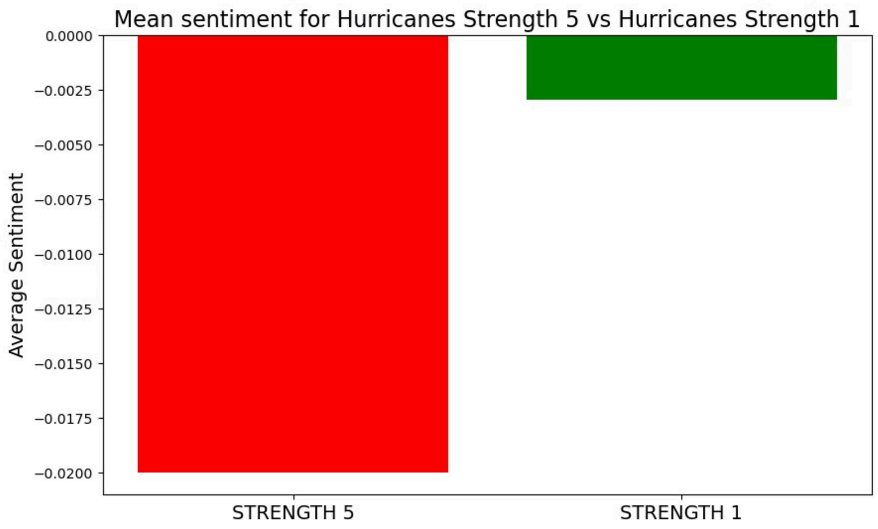
Sentiment is the feeling or mood shown in a piece of text.

It can be positive, neutral or negative.

Sentiment analysis helps us see how people react to an event. It can show whether the public had a positive or negative reaction.

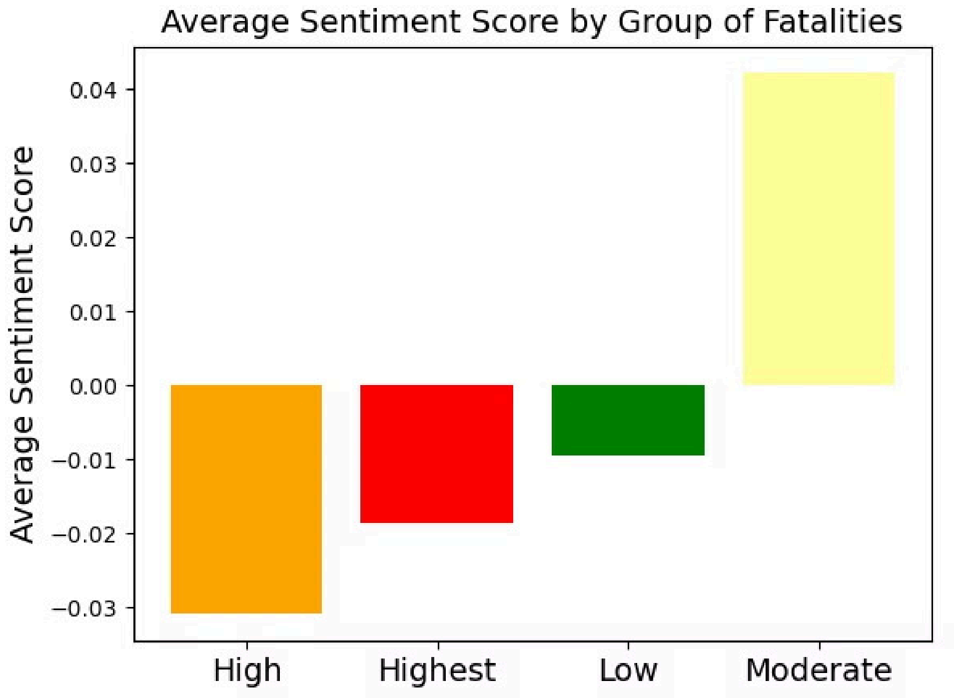
NLTK (Natural Language Toolkit), Word Clouds,

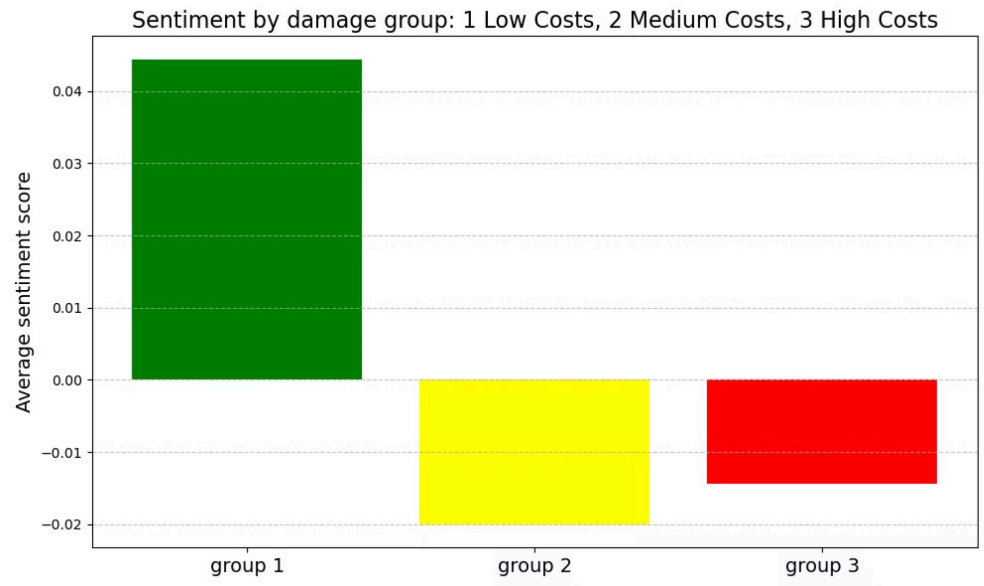
VADER (Valence Aware Dictionary and sEntiment Reasoner)



Made with Gamma

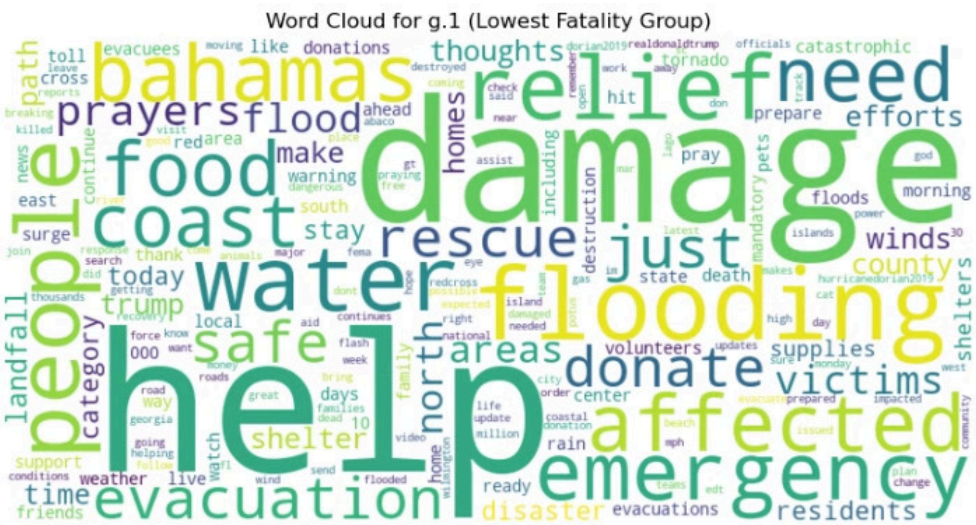
Average Sentiment Distribution Across Hurricanes Grouped by Fatalities and Cost of Damage

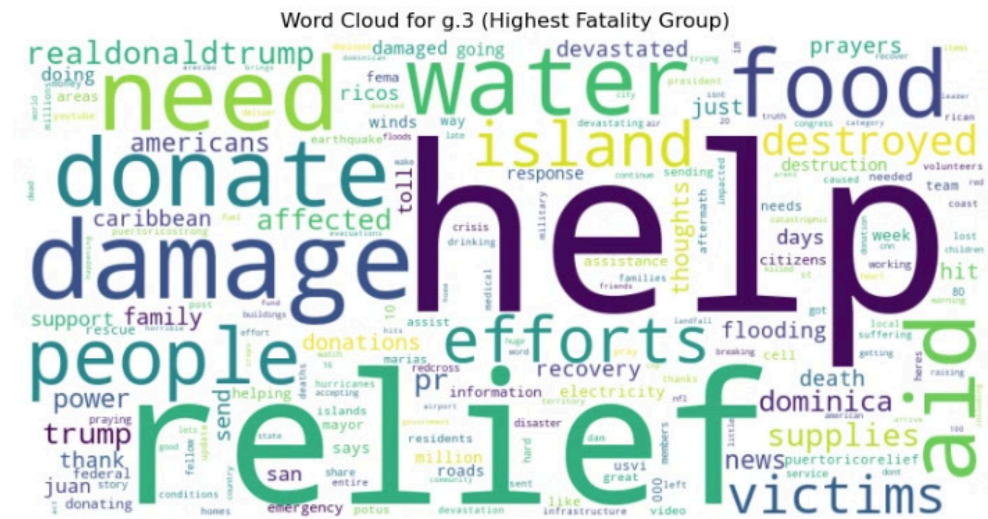




Made with Gamma

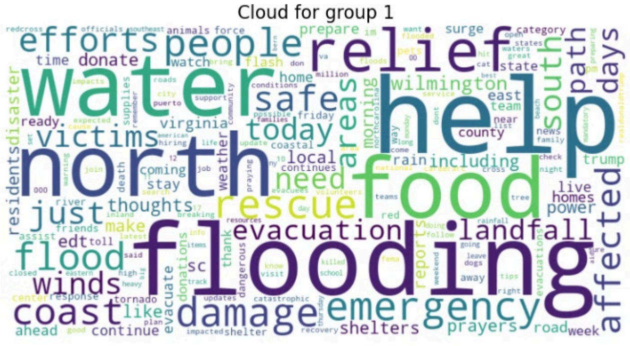
Average Sentiment Distribution Across Hurricanes Grouped by Fatalities





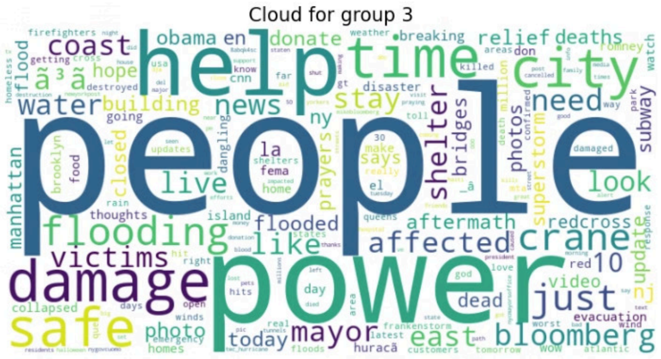
Made with Gamma

Word Cloud for Hurricanes Grouped by Damage Costs



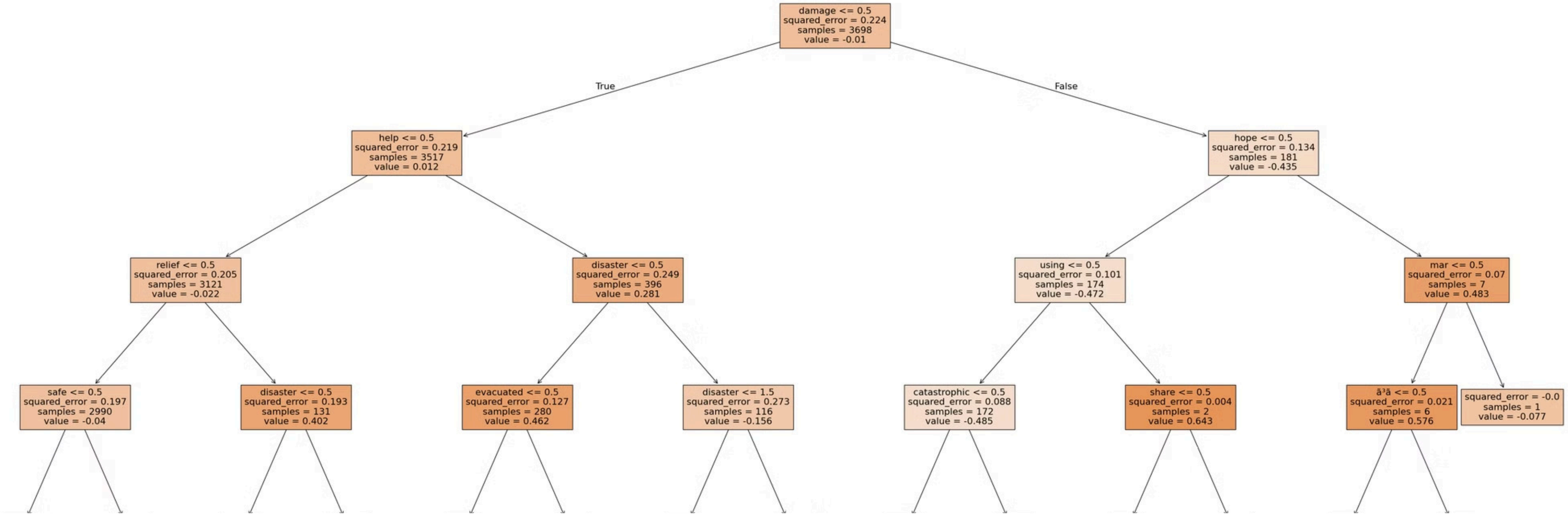
Low costs

Medium costs



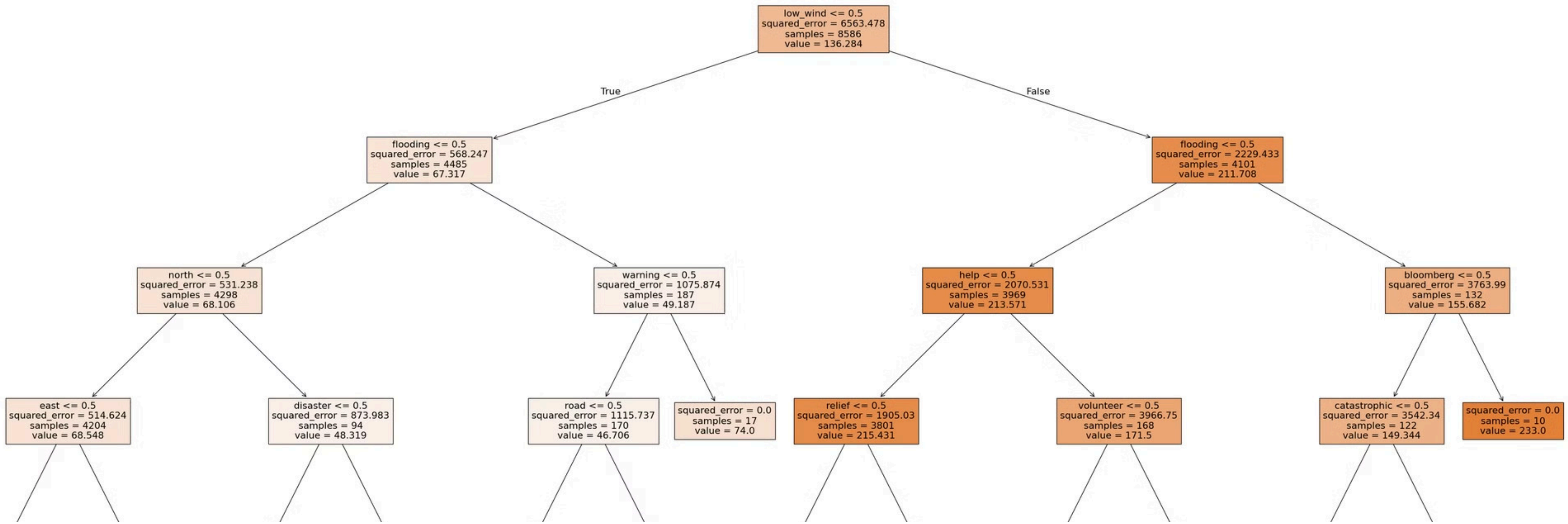
High costs

Made with Gamma



Regression Tree Based on Hurricane-Related Tweets Predicting Sentiment Value

Made with Gamma



How do Wind Parameters influence the Regression Tree's Prediction of Fatalities?

Made with Gamma

Our Models to Predict Hurricanes

Cost of Damage:

82% Accuracy to predict 1 from 3 classes.

We used RandomForestClassifier as a base model, and improved it by using XGBClassifier with RandomizedSearchCV.

["CPI-Adjusted Cost_class"]

Number of Fatalities:

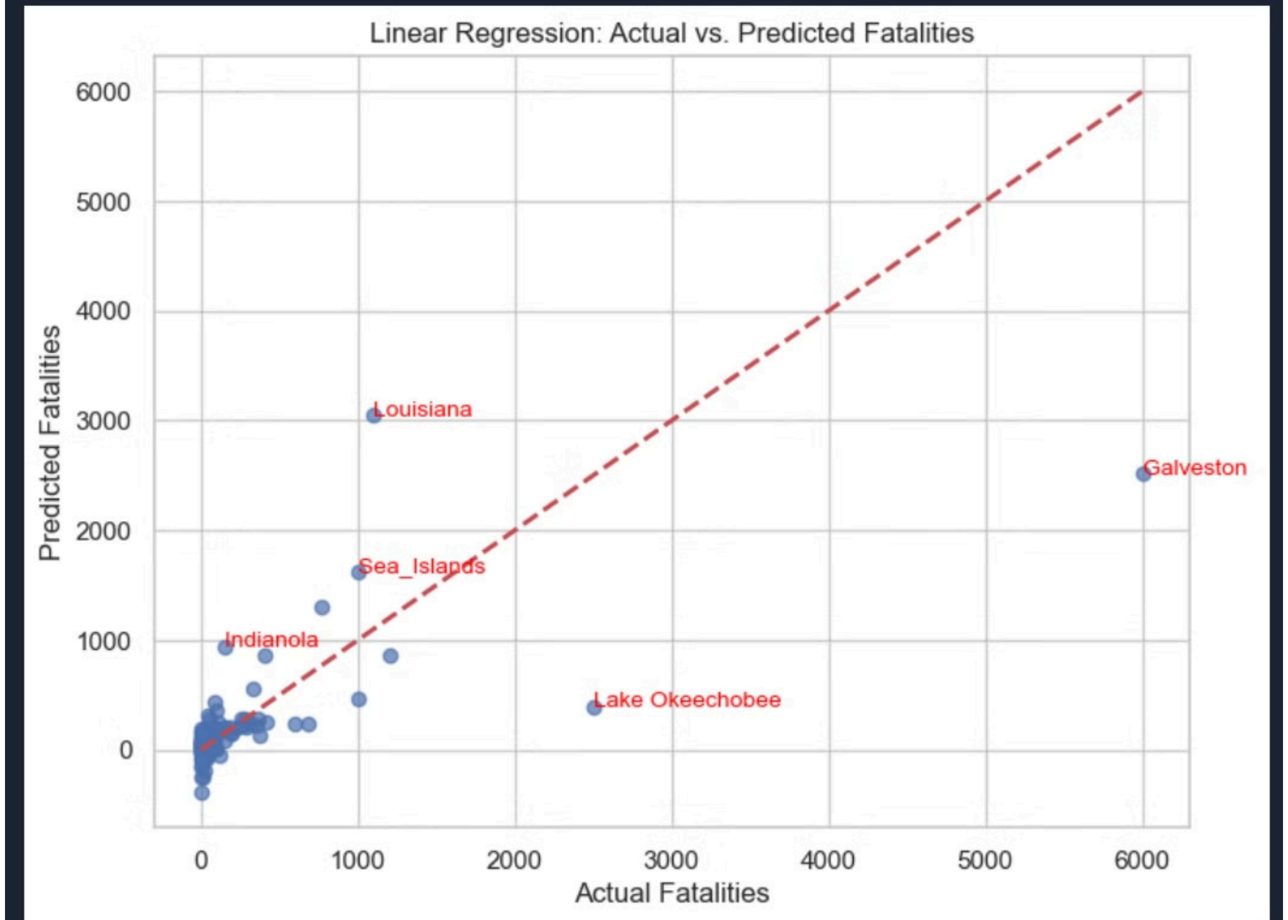
63% Accuracy to predict 1 from 4 classes.

(96% Accuracy if it's a binary class)

We used RandomForestClassifier as a base model and improved it with XGBClassifier.

["Fatal_class"]

Made with Gamma

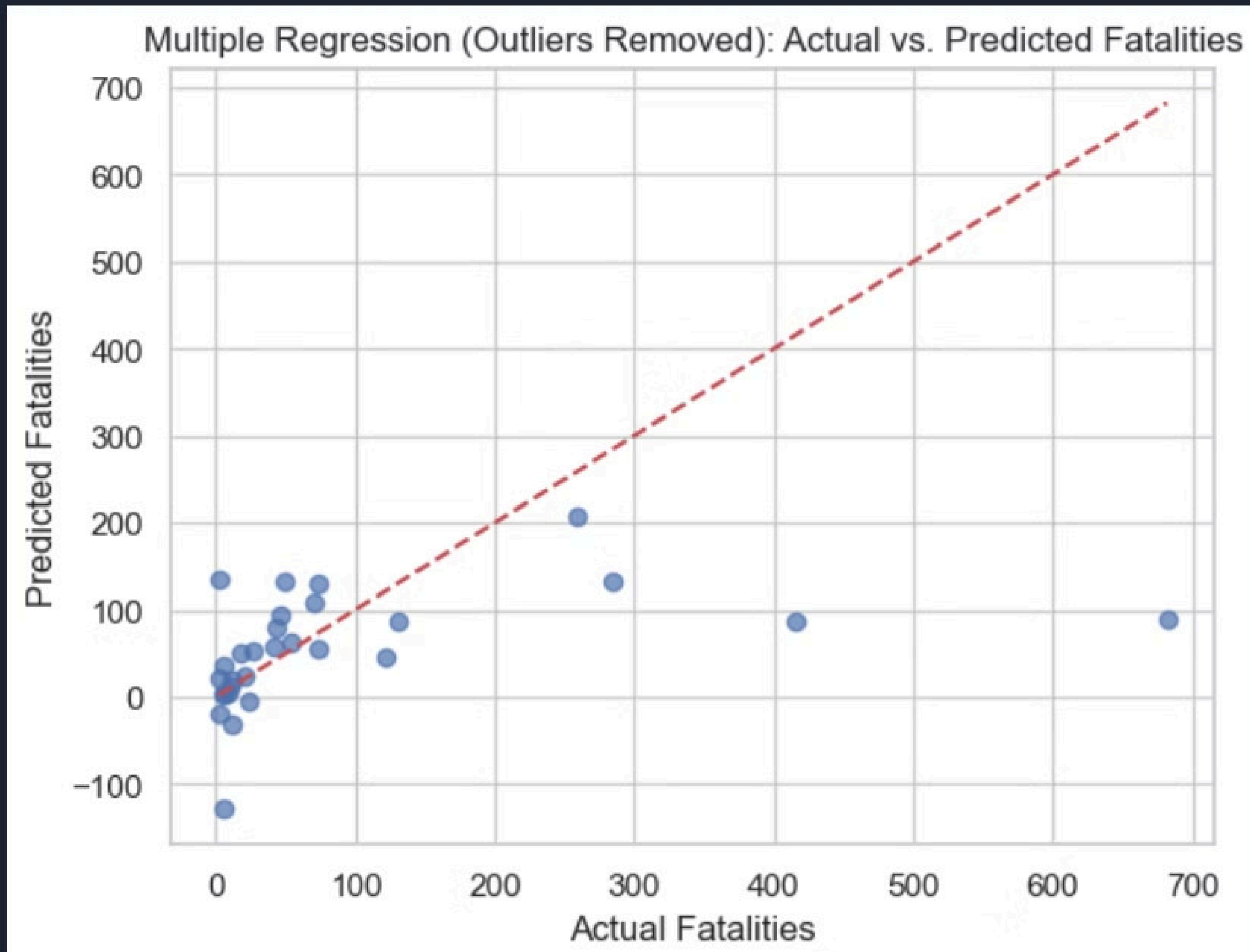



```
Identified Outliers:
           Hurricanes  Fatalities
124         Galveston        6000
110   Lake Okeechobee        2500
131         Louisiana        1100
133         Indianola         150
130       Sea_Islands        1000
```

- **Outliers Identified:** Major hurricane events such as **Galveston (6000 fatalities)**, **Lake Okeechobee (2500 fatalities)**, and **Louisiana (1100 fatalities)** deviate significantly from the model's predictions.
- **Prediction vs. Actual Trends:** The **majority of points cluster near the regression line**, indicating reasonable predictive performance, but extreme fatality events are **underestimated** by the model.
- **Systematic Underprediction:** The model struggles with high-fatality hurricanes, suggesting **nonlinear relationships or missing key predictive features** in the dataset

Made with Gamma

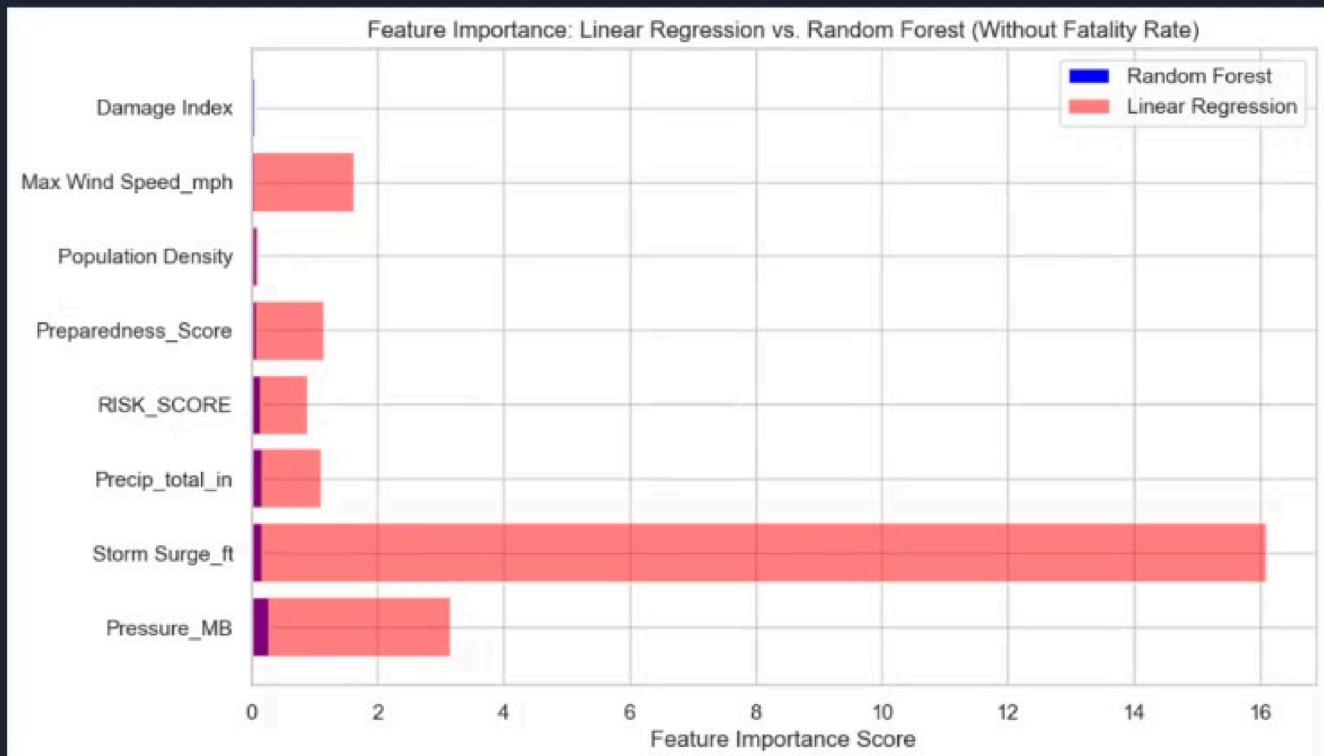
```
Mean Squared Error: 18970.57988204361
R²: 0.1281518636778829
```



- **The model struggles to predict fatalities accurately.** The **R^2 score is only 0.13**, meaning it explains just **13% of the changes in fatalities**—which is very low.
- **Most predictions are close to zero.** The model often **underestimates fatalities**, especially for larger storms, and even predicts **negative fatalities**, which is not realistic.
- **Removing outliers didn't fix the problem.** Even after **removing extreme storms**, the model still doesn't follow the trend well, suggesting **it may be missing key hurricane impact factors**.

Made with Gamma

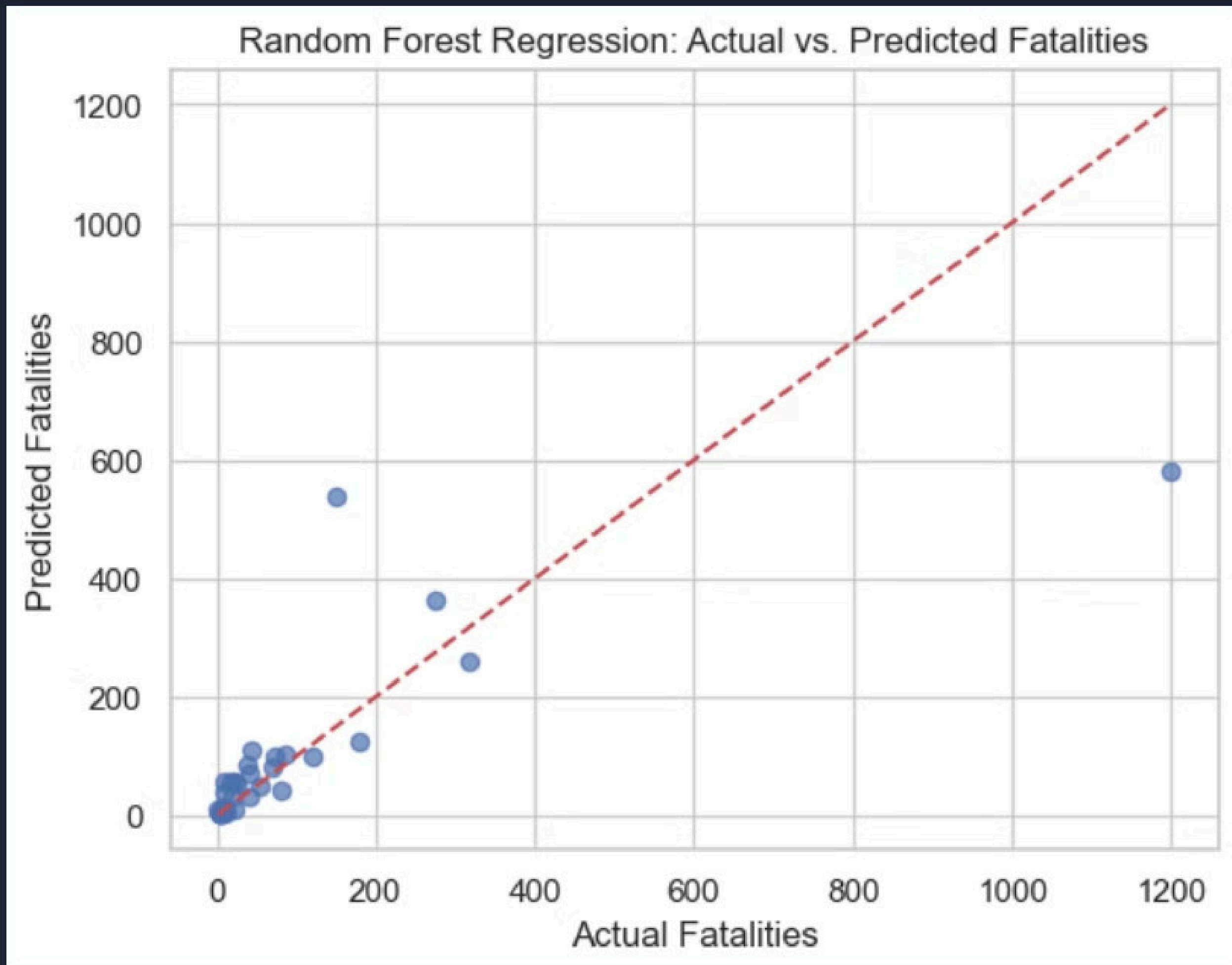
```
Feature Importance Comparison (Without Fatality Rate):
              Feature  Linear Regression  Random Forest
1         Pressure_MB       3.144191e+00       0.266418
2       Storm Surge_ft      1.608365e+01       0.171550
3      Precip_total_in      1.100919e+00       0.151907
4          RISK_SCORE       8.866127e-01       0.140400
7   Preparedness_Score      1.141094e+00       0.079736
5   Population Density      8.855626e-02       0.067983
0   Max Wind Speed_mph      1.621071e+00       0.061143
6        Damage Index       7.487204e-07       0.060863
```



- **Linear Regression puts too much focus on individual weather factors.** It thinks **Storm Surge and Pressure** are the most important, but this might be **over-exaggerated** because the model assumes straight-line relationships.
- **Random Forest spreads out the importance more evenly.** It **balances storm-related features like wind speed** and **doesn't let one factor dominate**, which helps it better understand complex storms.
- **Different models see hurricane impacts in different ways.** Linear models **focus on a few strong factors**, while Random Forest **considers how multiple factors work together**, making it **better at predicting real-world hurricane deaths**.

Made with Gamma

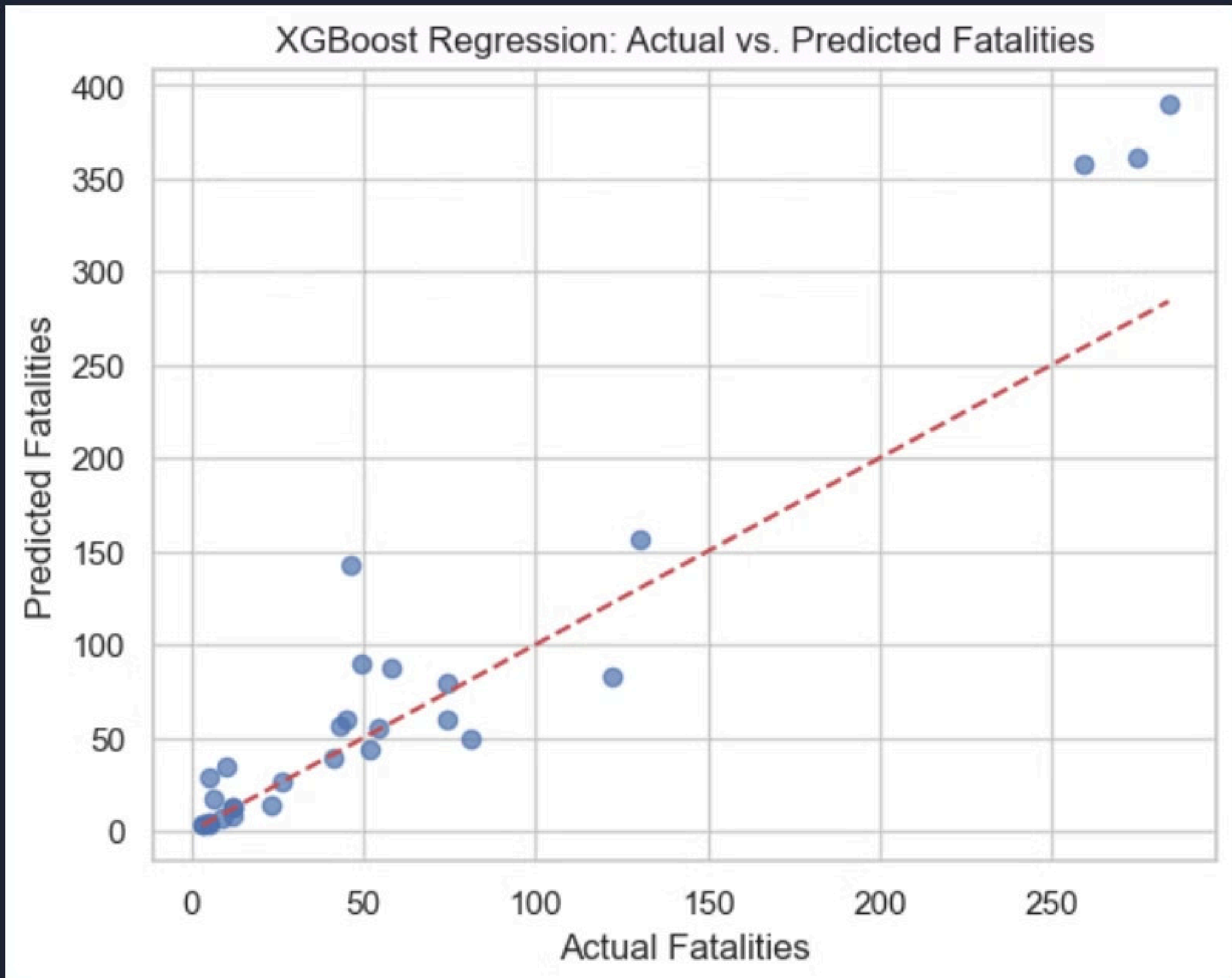
```
Mean Squared Error: 18813.750976666666
R²: 0.6064191518576794
```



- **It is unrealistic to use fatality rate to predict fatality rate. It maybe good for historical measures but not predicators .**
- **Exposure Factors May Have Influenced the Results Too Much.** If features like **population density or preparedness score** were based on **data collected after the storm**, then the model isn't truly predicting fatalities—it's just repeating patterns it already saw.
- **Test Data May Have Been Too Similar to Training Data.** If some of the same hurricanes were in both training and testing datasets, the model **wasn't really predicting new hurricanes—it was memorizing old ones**, leading to **higher R^2 than expected**

Made with Gamma

```
XGBoost Mean Squared Error: 1566.420654296875
XGBoost R²: 0.7472000122070312
```



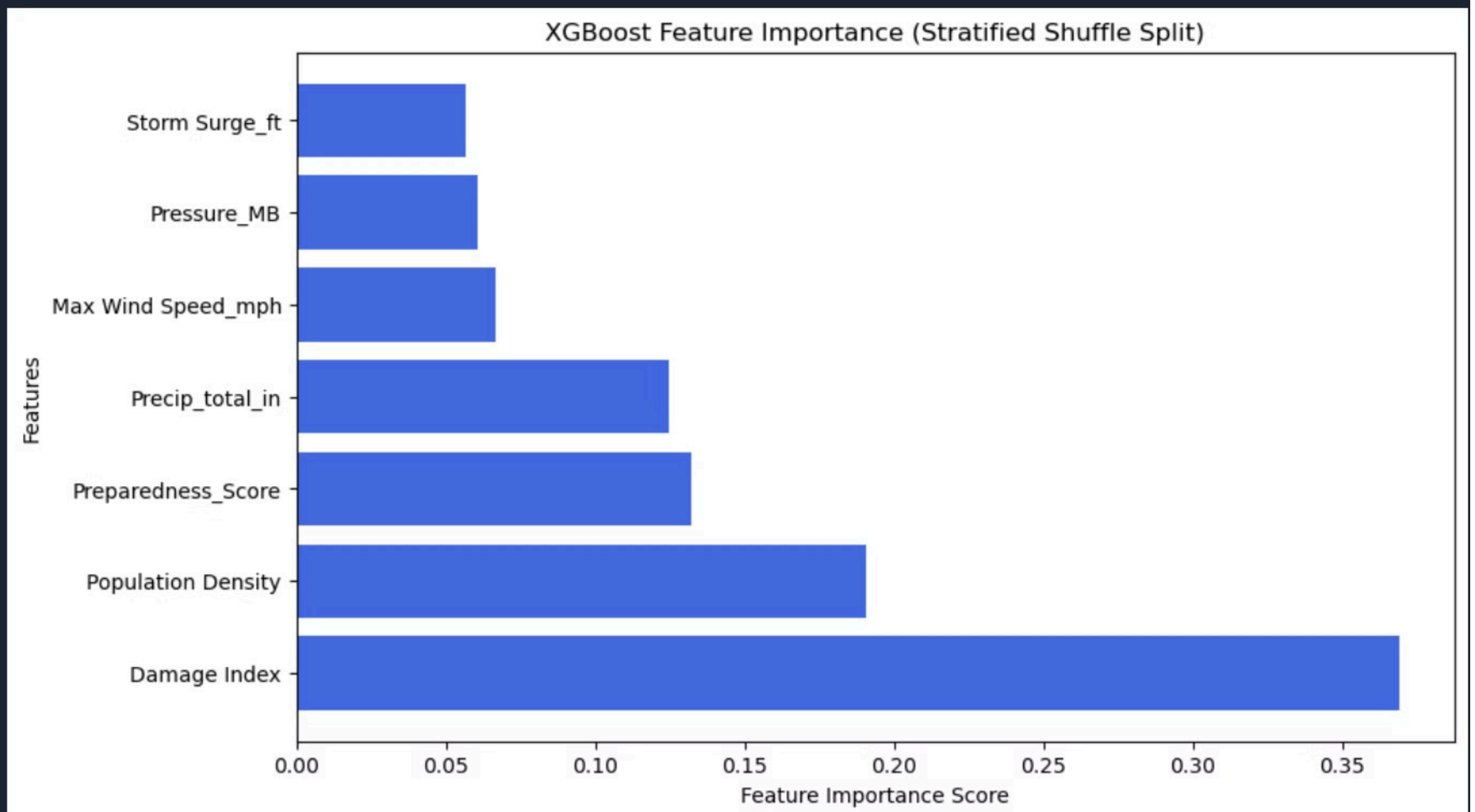
- **Model might be using future data.** If it learned from **post-storm information**, it would **predict fatalities too well**, making it look more accurate than it really is.

- **Predictions are too close to actual values.** The **high R^2 (0.75)** suggests the model may have **seen similar data during training**, meaning it's **memorizing instead of truly predicting.**

- **It underestimates extreme events.** The model **struggles with higher fatality storms**, meaning it **might be missing key disaster impact factors.**

Made with Gamma

```
Feature Importance Analysis:
               Feature  Importance
5        Damage Index    0.368847
4  Population Density    0.190847
6  Preparedness_Score    0.132103
3       Precip_total_in    0.124479
0  Max Wind Speed_mph    0.066500
1         Pressure_MB    0.060499
2       Storm Surge_ft    0.056726
```

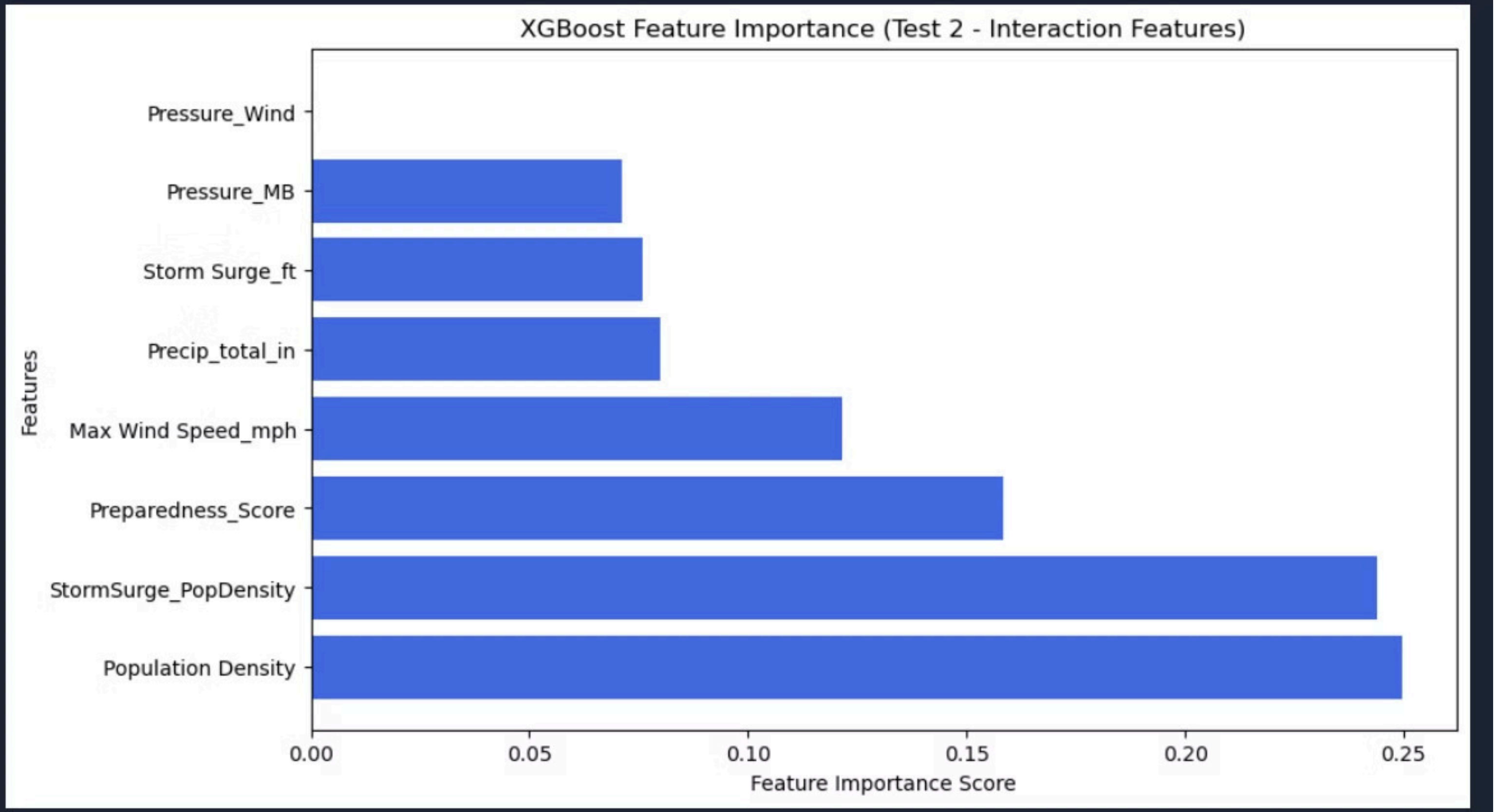


- **Damage Index and Population Density dominate,** meaning the model focuses more on exposure than storm severity.
- **Wind Speed, Pressure, and Storm Surge are weak,** suggesting the model isn't prioritizing meteorological factors.
- **Possible data leakage** if **Damage Index** includes post-storm data, making predictions misleading.

Made with Gamma

```
TEST 2 - Interaction Features (Without Damage Index)
Mean Squared Error: 5.3099
R² Score: 0.7422

Feature Importance Analysis (Test 2):
                Feature  Importance
4     Population Density    0.249506
6   StormSurge_PopDensity    0.243775
5     Preparedness_Score    0.158501
0     Max Wind Speed_mph    0.121563
3        Precip_total_in    0.079852
2         Storm Surge_ft    0.075789
1            Pressure_MB    0.071014
7          Pressure_Wind    0.000000
```

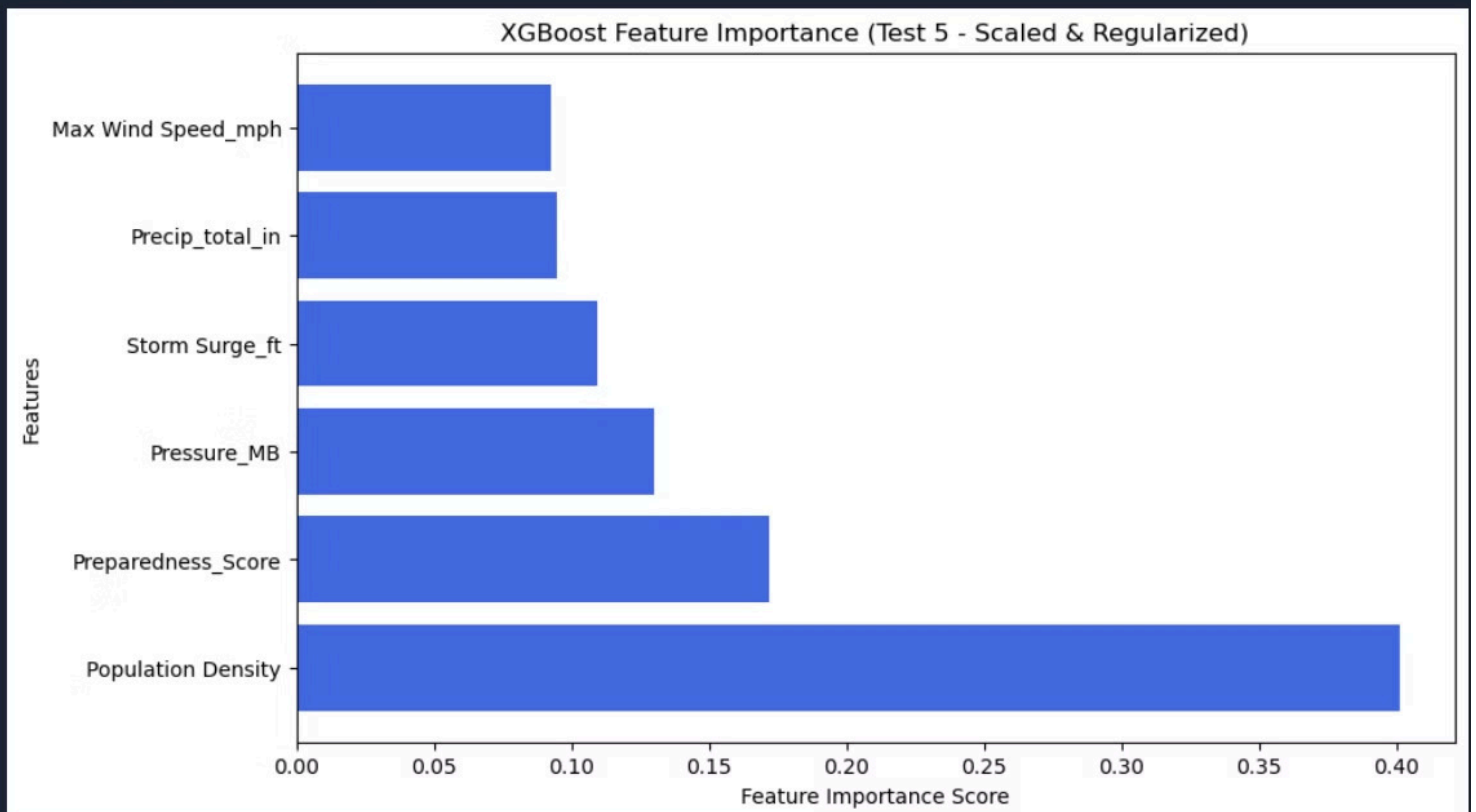


- **Removing Damage Index shifted importance to Population Density** and **Storm Surge interactions**, but exposure still dominates.
- **Meteorological factors like Wind Speed and Pressure improved slightly,** but not enough to drive predictions.
- **Pressure_Wind had no impact,** showing some interaction features didn't help the model.

Made with Gamma

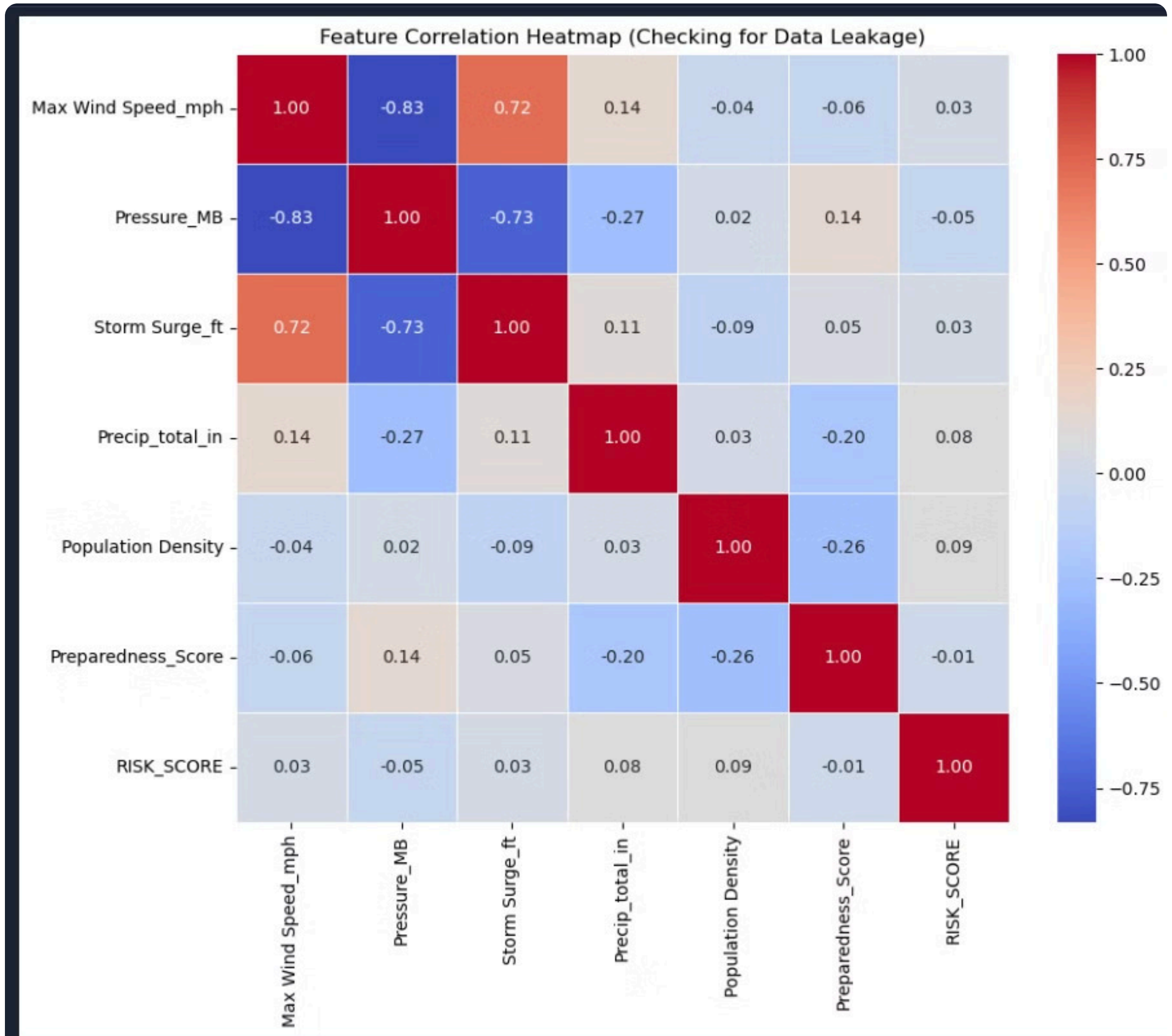
```
TEST 5 – Feature Scaling & Regularization
Mean Squared Error: 6.6361
R² Score: 0.6778

Feature Importance Analysis (Test 5):
              Feature  Importance
4  Population Density    0.401053
5  Preparedness_Score    0.171887
1         Pressure_MB    0.130088
2      Storm Surge_ft    0.109360
3     Precip_total_in    0.094717
0  Max Wind Speed_mph    0.092896
```



- **Scaling and regularization didn't fix exposure bias.** Population Density still dominates.
- **Storm Surge, Pressure, and Wind Speed gained importance,** but they're still weaker than exposure factors.
- **The model is slightly more balanced,** but more adjustments are needed to focus on storm severity.

Made with Gamma



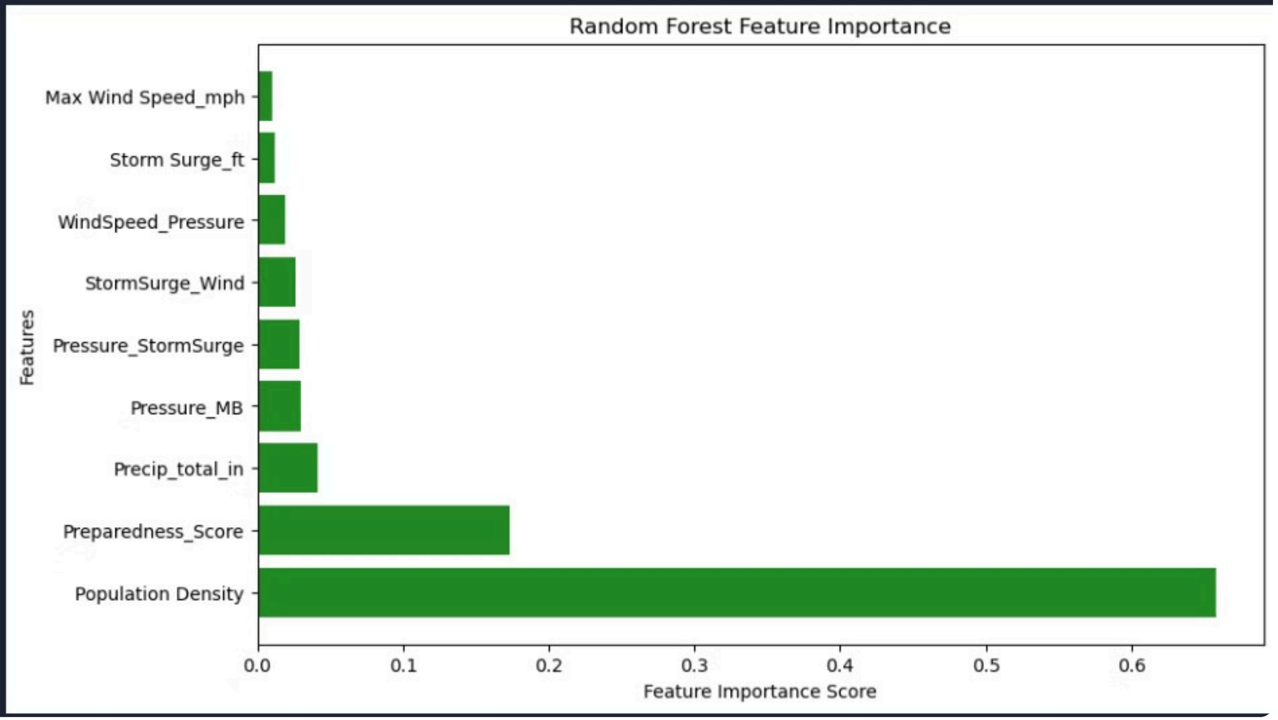

```
Highly Correlated Features with RISK_SCORE (Potential Leakage Check):
RISK_SCORE            1.000000
Population Density    0.089715
Precip_total_in       0.081500
Pressure_MB           0.051896
Max Wind Speed_mph    0.033038
Storm Surge_ft        0.027929
Preparedness_Score    0.005651
Name: RISK_SCORE, dtype: float64
```

- **No extreme data leakage found.** Population Density has a weak correlation (0.09) with Risk Score.
- **Meteorological features (Wind Speed, Storm Surge, Pressure) are weakly linked** to Risk Score, meaning storm severity isn't driving predictions.
- **The model still relies on exposure features,** but there's no direct sign of it using post-storm data incorrectly.

Made with Gamma

```
RANDOM FOREST MODEL - Quick Test
Mean Squared Error: 5.5260
R² Score: 0.7317

Feature Importance Analysis (Random Forest):
              Feature  Importance
4  Population Density    0.657638
5   Preparedness_Score   0.173675
3       Precip_total_in   0.041480
1           Pressure_MB   0.030176
8  Pressure_StormSurge   0.029188
7       StormSurge_Wind   0.026176
6    WindSpeed_Pressure   0.019529
2        Storm Surge_ft   0.011907
0    Max Wind Speed_mph   0.010232
```

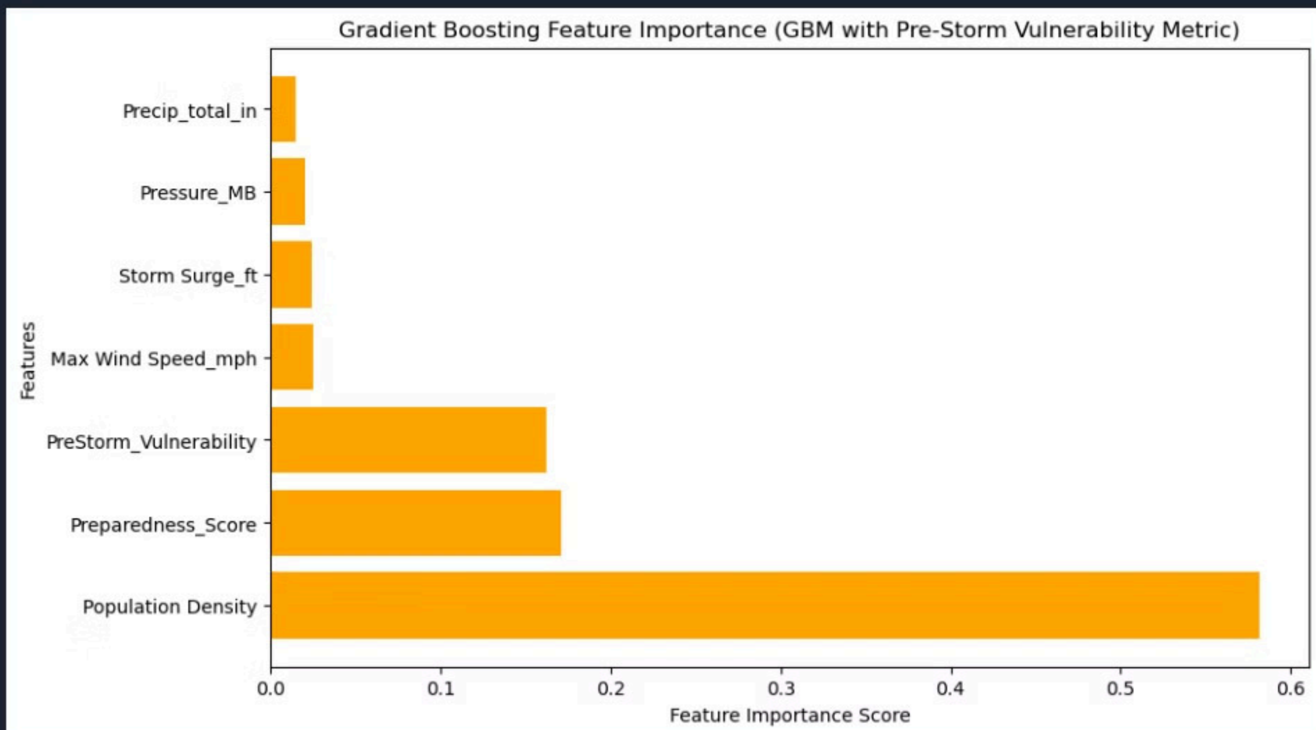


- **Population Density dominates predictions,** meaning the model relies more on where people live than storm severity.
- **Meteorological features (Wind Speed, Storm Surge, Pressure) have low importance,** showing the model isn't prioritizing storm strength.
- **Preparedness Score is the second most important,** suggesting the model factors in how well communities are ready but still undervalues storm intensity.

Made with Gamma

```
TEST 14 - GBM with Engineered Pre-Storm Vulnerability Metric
Mean Squared Error: 3.4377
R² Score: 0.8331

Feature Importance Analysis (GBM with Pre-Storm Metric):
                 Feature  Importance
4       Population Density    0.581991
6       Preparedness_Score    0.170967
5   PreStorm_Vulnerability    0.162209
0       Max Wind Speed_mph    0.025286
2           Storm Surge_ft    0.024336
1              Pressure_MB    0.020339
3          Precip_total_in    0.014872
```



- **Population Density is still the most important,** meaning exposure remains the biggest factor in predicting fatalities.
- **Pre-Storm Vulnerability helped but didn't shift focus to storm severity,** showing the model still favor
- **Meteorological factors (Wind Speed, Storm Surge, Pressure) remain weak,** meaning storm intensity isn't driving predictions enough.

- **The model failed badly,** with an **R^2 of -1.33**, meaning it made worse predictions than just guessing.
- **Loss decreased over time,** showing the model learned patterns, but they **weren't useful for predicting fatalities.**
- **Neural networks did not work well for this dataset, confirming that simpler models like GBM were the better choice.**

Made with Gamma

Key Takeaways

Hurricane Trends

Hurricane occurrence and intensity have increased in recent decades.

Feature Engineering

Ratio-based features (e.g., wind/pressure) were introduced to enhance model accuracy.

Data Handling

Handling missing data (dropping vs. median fill) significantly affects prediction outcomes.

Financial Loss Correlation

Financial losses are closely linked to deaths, storm surge, precipitation, and pressure.

Fatality Correlation

Fatalities are most strongly correlated with storm surge, wind speed, and pressure.

Made with Gamma

Thank you Professor, Andrew and the instructional team!

Your guidance has been our **calm before the storm**, helping us **weather challenges** with confidence. We appreciate your **steady leadership** throughout this journey!

Made with Gamma