Project Overview (DANIEL ALLEN)

As a member of TEAM 3, my primary objective was to ensure meteorological factors (e.g., wind speed, pressure, storm surge) were properly weighted against exposure-driven variables (e.g., population density, preparedness score). To create an optimized prediction model, the project involved extensive model testing, feature engineering, and bias mitigation strategies. Unfortunately, because of the failures of the model development, the model I developed failed miserably when I measured its validity to another dataset of three cyclones/typhoons.

PROCESS

1. **Data Preprocessing & Feature Engineering**
    - The dataset was cleaned and adjusted so all the numbers were on a similar scale. Using Standard Scale Gradient Boosting (GBM): Outperformed others in terms of predictive accuracy and generalization.
        - **Neural Networks**: Showed poor performance due to dataset limitations.
    - **Final Model:** GBM with hyperparameter tuning (n_estimators=200, learning_rate=0.03, max_depth=6).
2. **Addressing Bias & Overfitting**
    - **Stratified Cross-Validation** ensured balanced storm severity distribution across training and test sets.
    - **Feature Weighting** reduced Population Density's dominance, improving model fairness.
    - **Downsampling of high-population regions** was attempted, but the feature importance was not sufficiently shifted.
3. **Final Model Performance**
    - **Gradient Boosting (GBM) Achieved:**
        - **Mean Squared Error (MSE):** 3.4377
        - **R^2 Score:** 0.8331
    - GBM demonstrated the best stability and predictive power in cross-validation.

REFERENCES:

- Seaborn Heatmaps: https://seaborn.pydata.org/generated/seaborn.heatmap.html
- Scikit-Learn Linear Regression: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

- Seaborn Histplot: https://seaborn.pydata.org/generated/seaborn.histplot.html
- Mean Squared Error (Scikit-Learn): https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html
- Python Module Reloading: https://stackoverflow.com/questions/76009840/how-to-reload-a-module-in-python

FUTURE CONSIDERATIONS:

- Further refinement of pre-storm vulnerability metrics.
- Investigation of real-time hurricane severity data sources.
- Potential ensemble model approach (GBM + Random Forest) to enhance predictive stability.

This project highlights the challenges in hurricane fatality prediction and the importance of balancing meteorological and exposure-driven risk factors. The final model provides a strong foundation for improving hurricane preparedness and impact mitigation strategies.