Pacific Northwest Earthquake Prediction and Analysis

Project 4 Group 4
Sofonias Abebe
Asres Dagnew
Jenn Andrews
Leo Rios

Objective

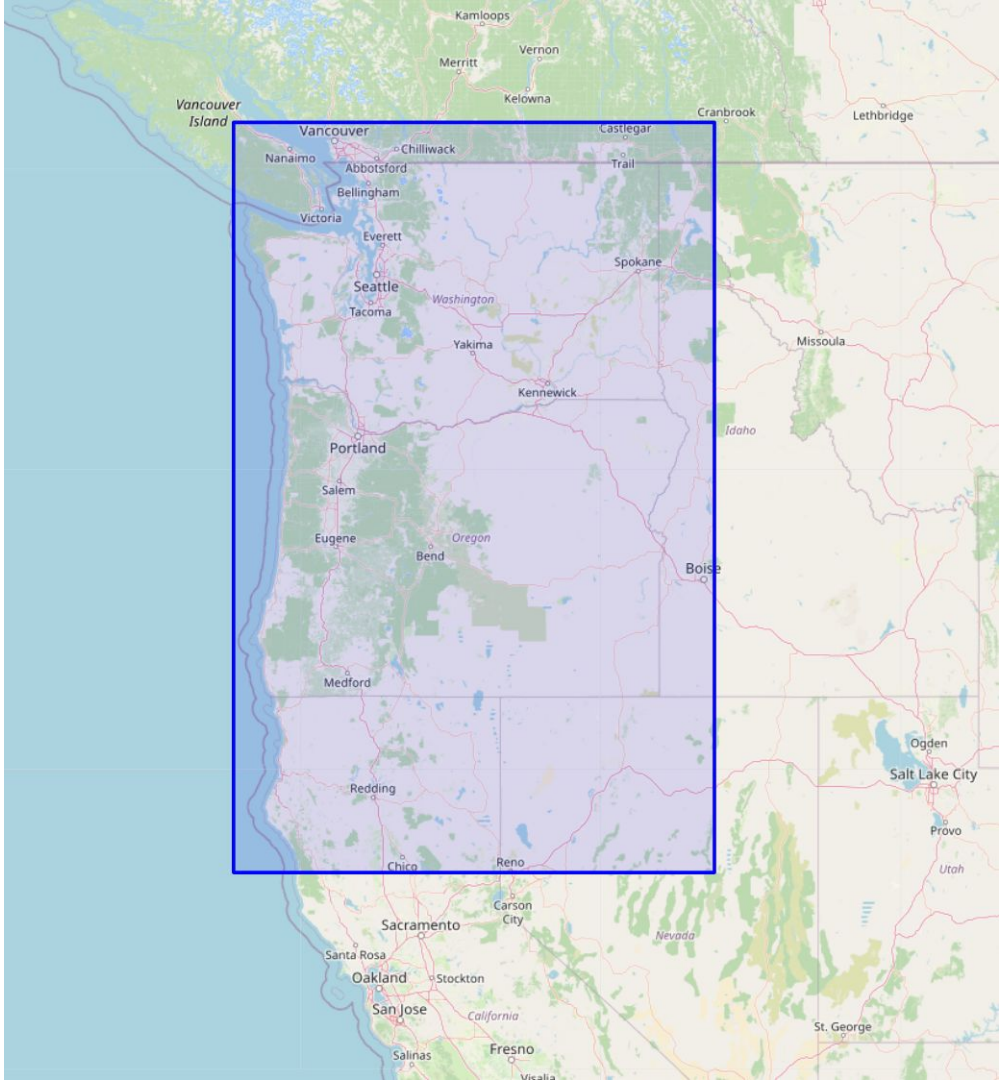
Analyze and predict earthquake patterns in the Pacific Northwest using historical data and machine learning.

This includes:

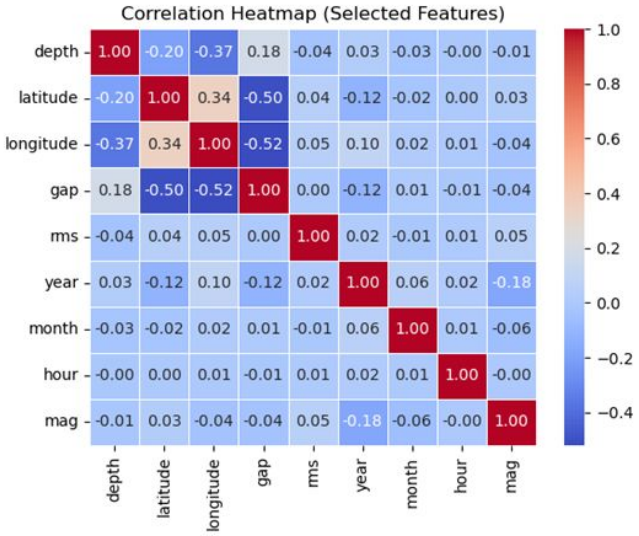
- ETL and database design
- Exploratory data analysis (EDA)
- ML modeling prediction
  - Occurrence
  - Magnitude
  - Occurrence + Magnitude prediction

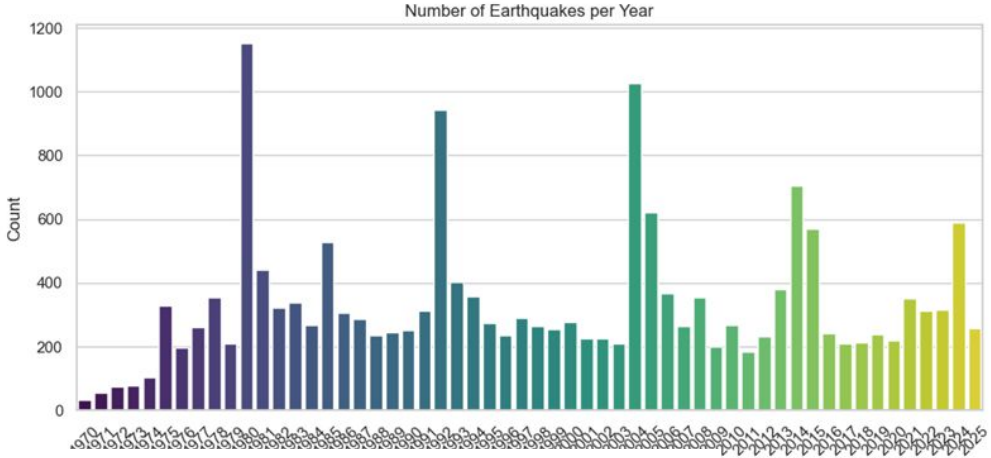
Data Sources and ETL

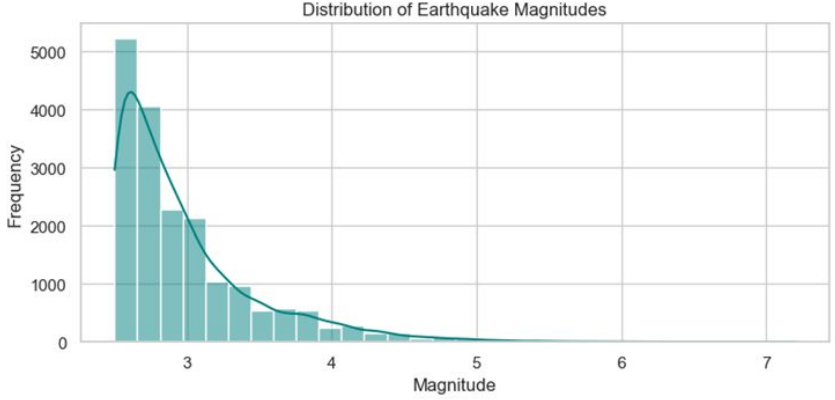
- Seismic activity dataset was obtained from the USGS Earthquake Hazards Program Catalog
- Filtered measurements for:
  - Magnitude >2.5
  - Jan 1970 – May 2025
  - Latitude: 39.5 to 49.5
  - Longitude: -125 to -116
- Exported to CSV and transformed with Pandas
- Created a PostgreSQL database with schema



Data Exploration (EDA)





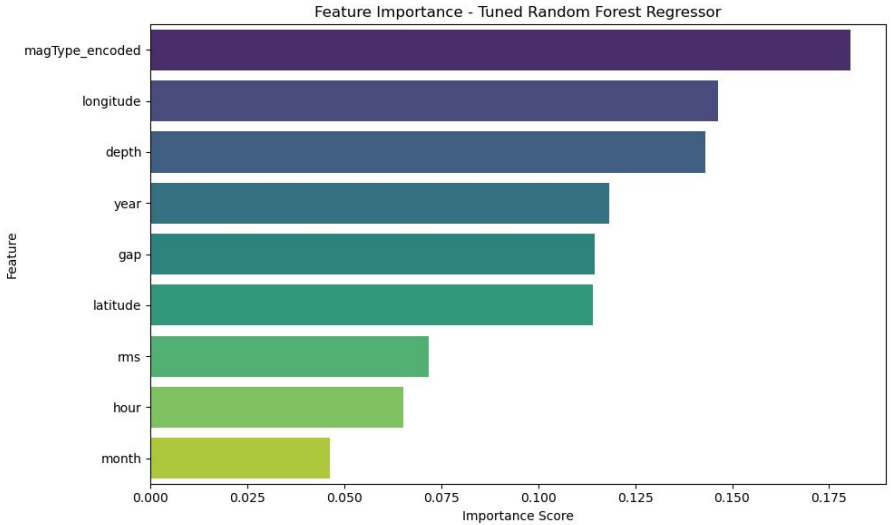


Modeling

1. Initial Magnitude model - Sofonias
    a. Without binning
    b. Results
2. Occurrence prediction - Asres
3. Magnitude prediction - Asres
    a. Refinement with binning
4. Magnitude + Occurrence prediction tab in dashboard - Sofonias/Asres?

Magnitude Prediction: Initial Model

- Initial training using Linear Regression and Random Forest using Magnitude as the target feature
  - Linear Regression
    - MAE: 0.3599
    - R^2: 0.0273
  - Random Forest Regressor
    - MAE: 0.3253
    - R^2: 0.1902
- Feature Expansion and Grid Search
  - Encode ‘magType’ as categorical
  - Hyperparameter tuning



Occurrence Classification Model

- Two-Pipeline Modeling
  - Pipeline 1: Classification Pipeline
  - Goal: Predict whether an earthquake will occur in a given time/location bin
- Applies a Random Forest Classifier to predict whether an earthquake will occur in a specific region and time frame. Includes data balancing using SMOTE and feature selection.

Magnitude Prediction: Model Refinement

- Pipeline 2: Regression Pipeline
  - Goal: If a quake is predicted to occur, estimate its magnitude
- Similar objective as the initial magnitude model, with feature refinements
  - Created temporal (monthly) and spatial (0.5°) bins
  - Add seasonal features

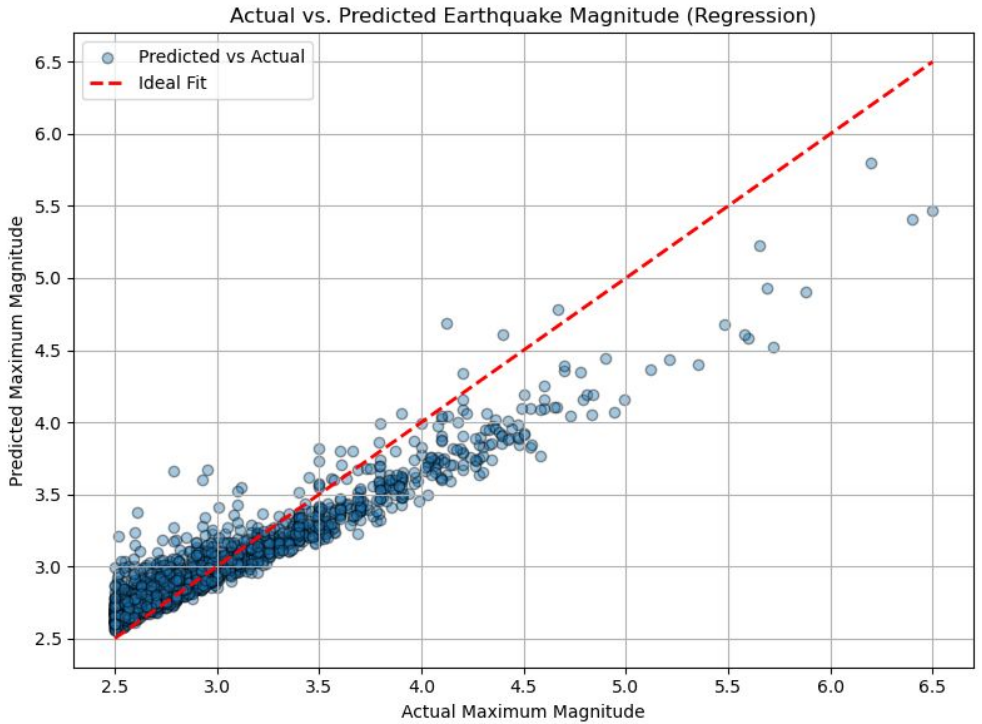
Model Results

Performance Summary:

- RMSE: 0.198
- R^2 Score: 0.852

Regression Performance Interpretation:

- Points near the line: good predictions.
- Points above the line: model overestimated magnitude.
- Points below the line: model underestimated magnitude.
- Scattered points far from the line: model uncertainty or noisy data.



Summary, Challenges/Limitations, Further Research

- Our two-pipeline model predicted the accuracy of an earthquake occurring at a specific time and location of 96% and an R^2 of that occurrence magnitude of 85.2%

HOWEVER...

Questions?