Modeling Earthquake Damage

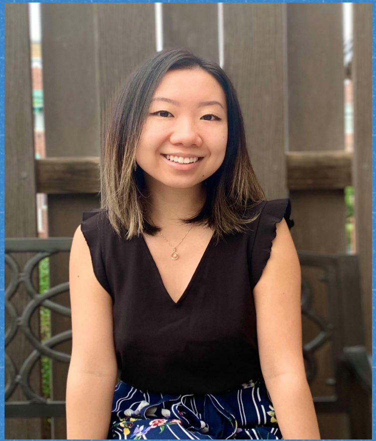
Chris Martin, Eric Cheng, Veronica Leong

Project 5
May 21, 2021

THE TEAM

Chris Martin
DATA SCIENTIST
LinkedIn

Eric Cheng
DATA SCIENTIST
LinkedIn



Victoria Leong
DATA SCIENTIST
LinkedIn

AGENDA

1. Background
2. Problem Statement
3. Data Cleaning
4. EDA
5. Modeling
6. Streamlit Demo
7. Conclusion / Recommendations
8. Limitations & Future Project Refinements

GORKHA EARTHQUAKE

Date: April 25, 2015

Location: 50 miles NW of Kathmandu

Intensity: 7.8 on the Richter scale

Deaths: 8,900

Injuries: 22,000

Homes affected:

over 1,000,000

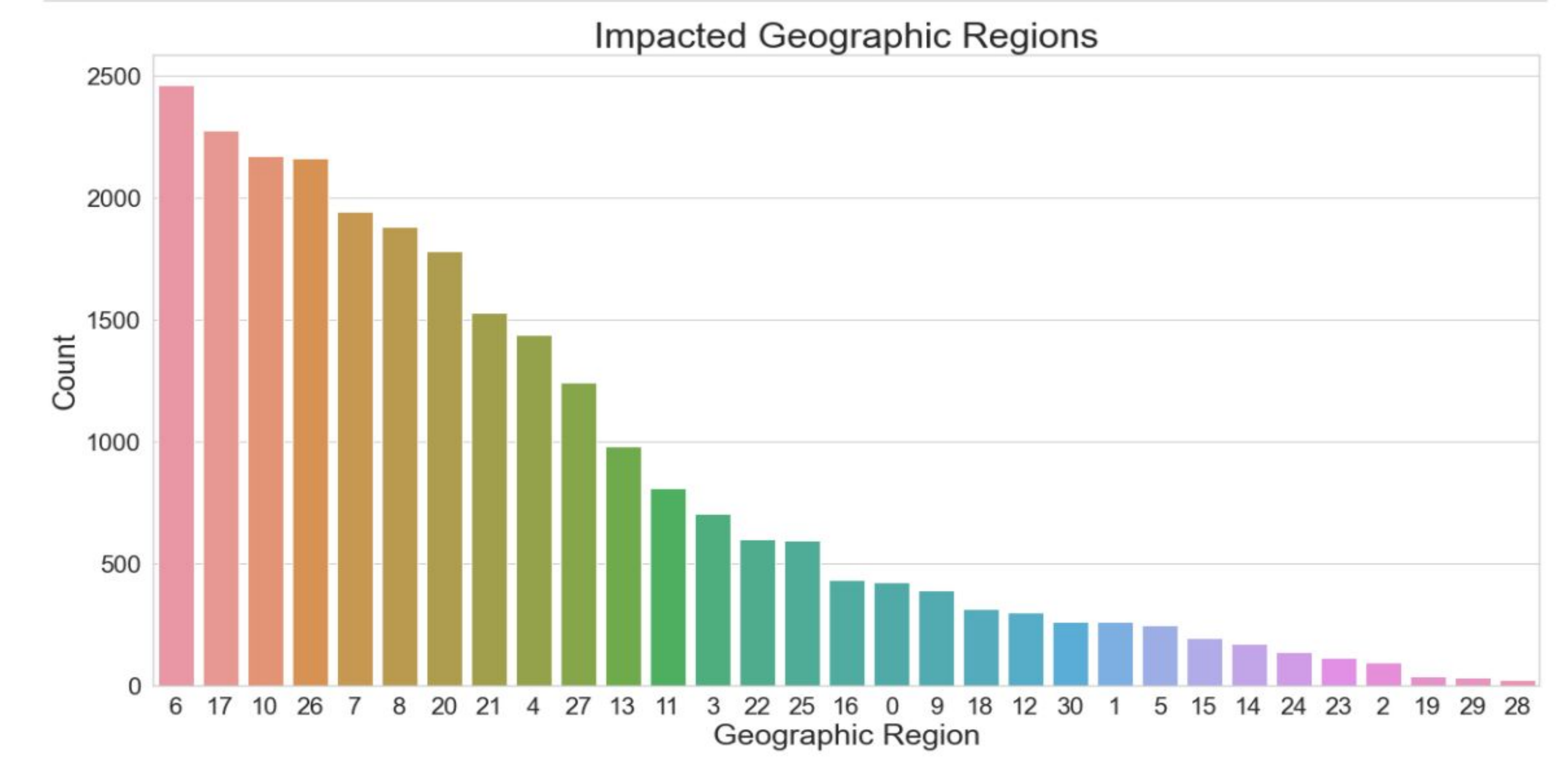
PROBLEM STATEMENT

Is it possible to predict the level of damage sustained by a building in the 2015 Gorkha earthquake in Nepal using classification modeling against the building’s features?

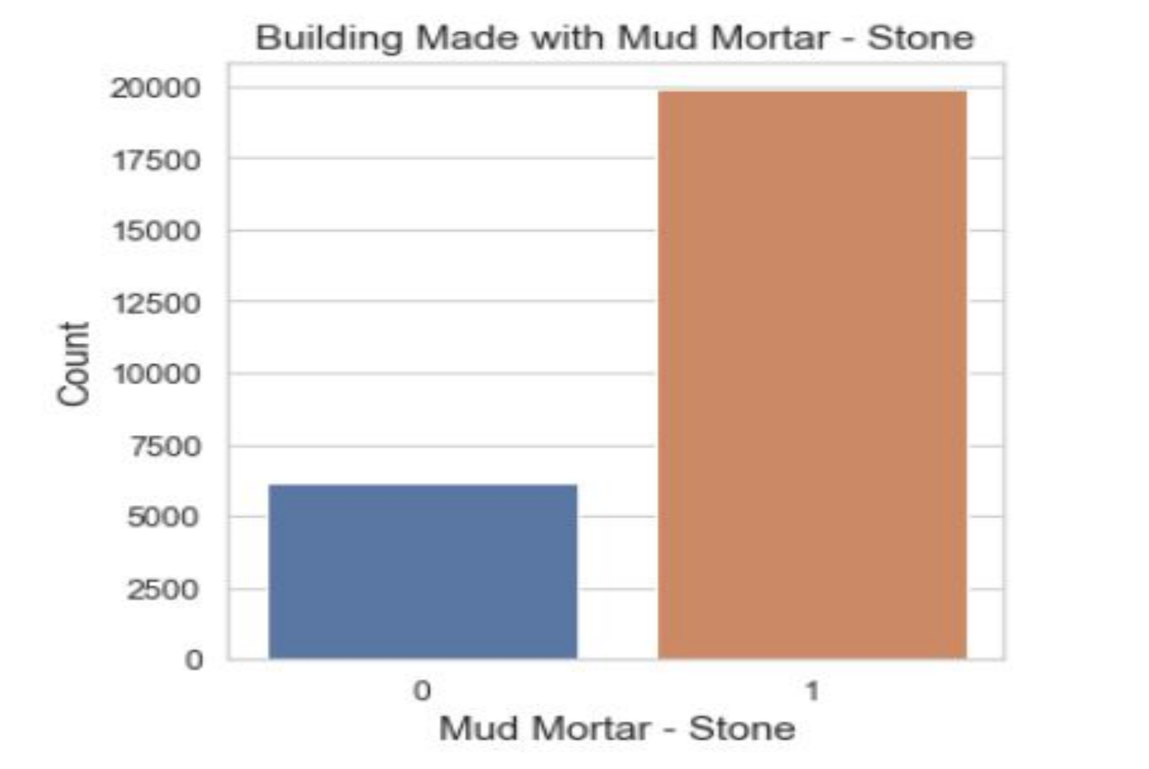
DATA COLLECTION & CLEANING

- Gathered data from drivendata.org
- Label Encoded features
- Checked for null values
- Utilized 10% of Dataset

DATA VISUALIZATIONS

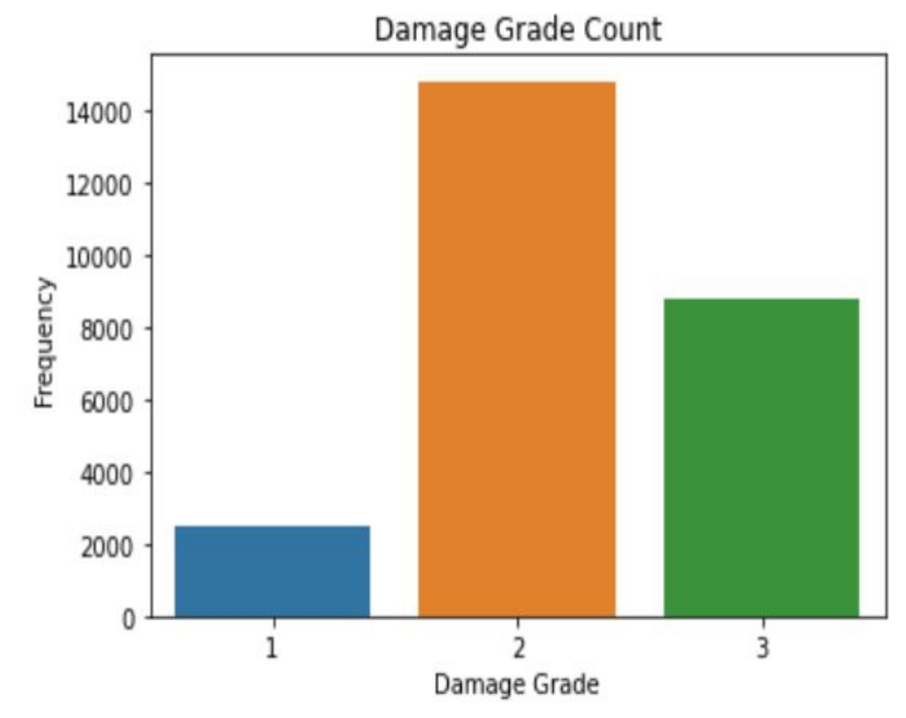


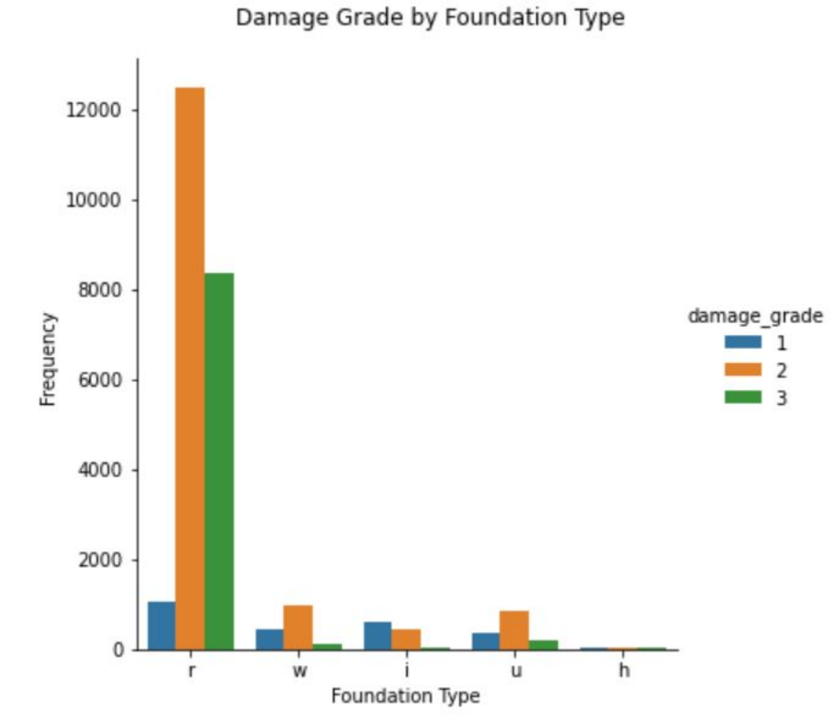
DATA VISUALIZATIONS (CONT’D)



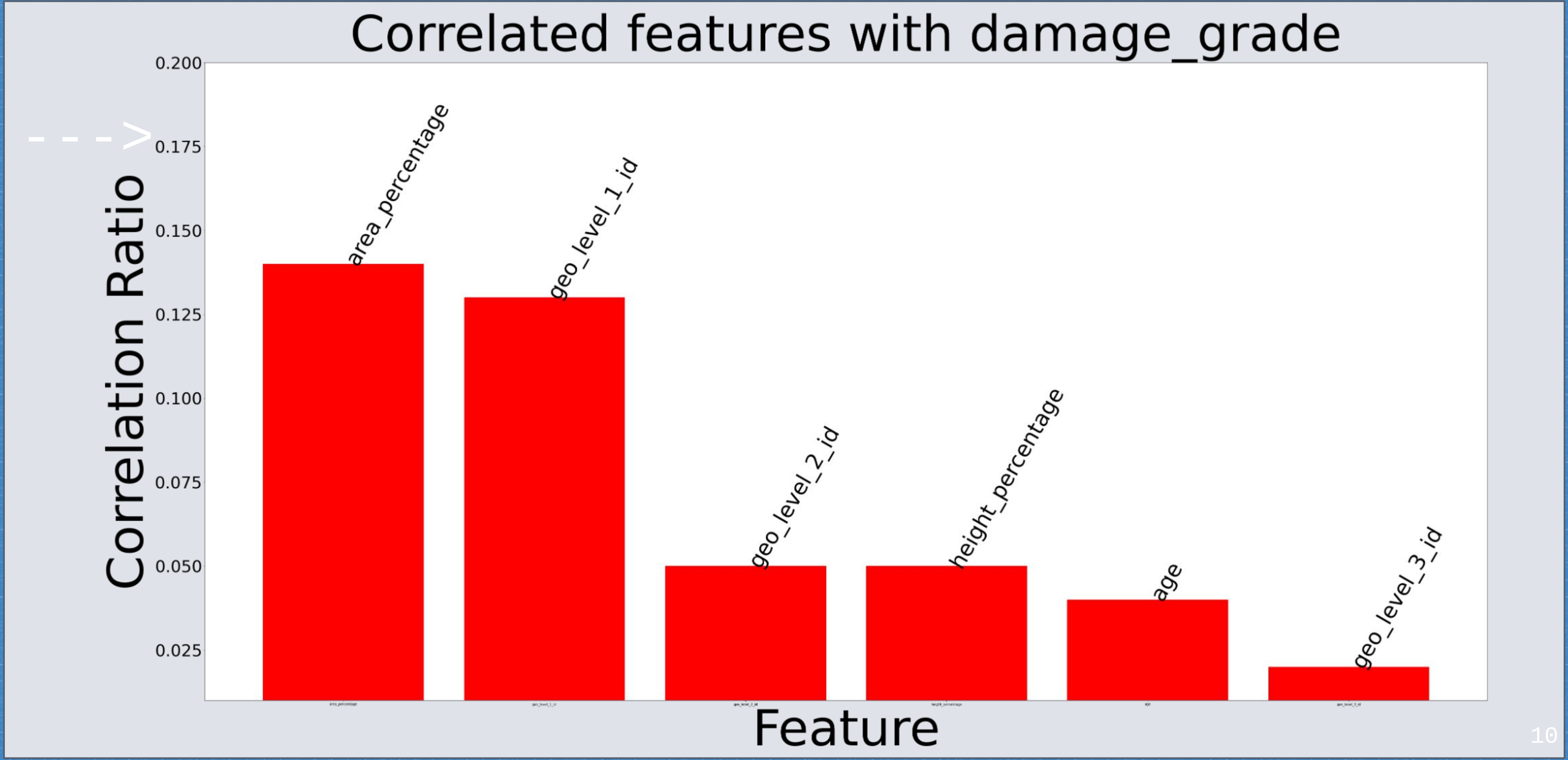
Sample size = 26,000
Mud = 20,000
Remainder = 6,000

DATA VISUALIZATIONS (CONTINUED)





FEATURE CORRELATIONS



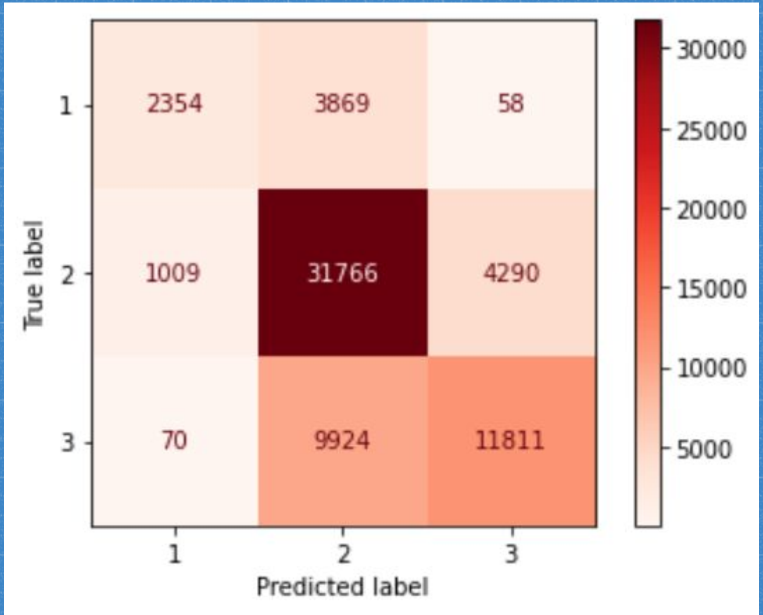
CLASSIFICATION MODELS TESTED

MODEL NAME	ACCURACY SCORE
Logistic Regression	0.583
KNearestNeighbors	0.571
Random Forest Classifier	**0.689**
Extra Trees Classifier	0.678
Decision Tree Classifier	0.665
Bagging Classifier	0.668
AdaBoost Classifier	0.660
Neural Network	0.568
Baseline Accuracy Model	***0.567***

BEST MODEL

- **Random Forest Classifier**
  70% Accuracy Score (All Data)

- Label Encoding
- Standard Scaler
- GridSearchCV -- Best Params
  - Max_depth = 11
  - Max_features = 35



BEST MODEL (CONT'D)

Feature Importances
Geo Level ID 1
Geo Level ID 2
Geo Level ID 3
Superstructure Mud Mortar Stone
Age

Permutation Feature Importances
Foundation Type
Age
Roof Type
Count Families
Superstructure Mud Mortar Stone

Streamlit Demo

CONCLUSIONS

- Neural Networks, KNN models did not perform well
- Logistic Regression, Boosting, Tree models did perform better
- Random Forest Classifier best classified the earthquake damage grade based off the features modeled
- Most important features: foundation type, building age, roof type, family count, mud mortar/stone superstructure

RECOMMENDATIONS

- Emphasize proper roof/foundation type for new construction
- Consider building superstructure
- Reinforce older buildings
- Reinforce multi-family housing buildings
- Focus efforts on heavily-impacted geographic areas

LIMITATIONS & FUTURE PROJECT REFINEMENTS

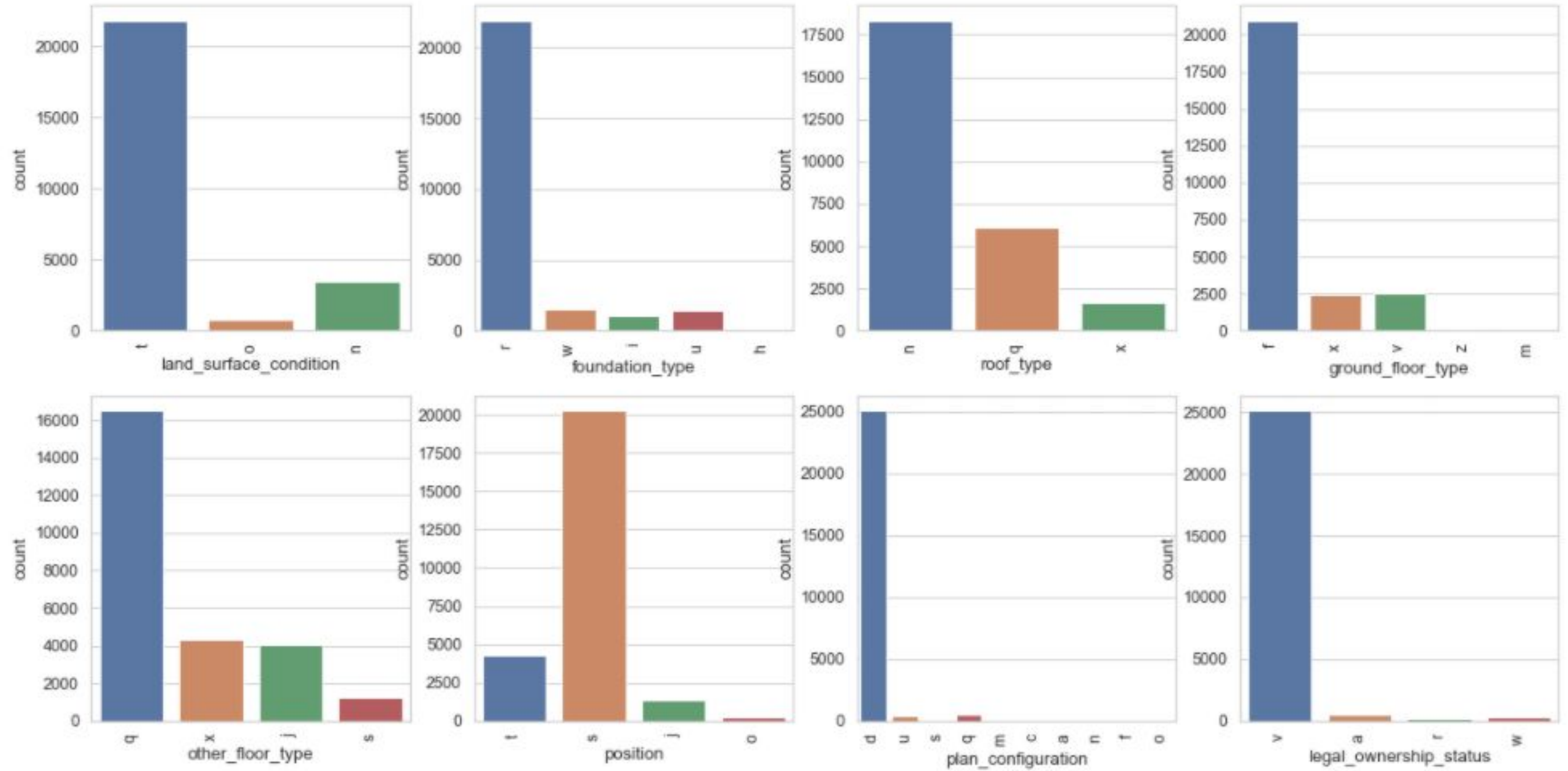
- Data Dictionary was limited in description
- Integrate imbalanced learning strategies
  - SMOTE, ADASYN, RandomOverSampler
- Integrate Boosting across the other tested models
  - AdaBoost, GradientBoost, HistGradientBoost

RESOURCES

- Project and Data source
  - Richtor's Predictor: Modeling Earthquake Damage
  - https://www.drivendata.org/competitions/57/nepal-earthquake/page/134/
- GA Instructional Team
  - Jeff Hale
  - Jacob Koehler
  - Eric Bayless

THANK YOU!
Questions?

Appendix


land_surface_condition
foundation_type
roof_type
ground_floor_type
other_floor_type
position
plan_configuration
legal_ownership_status
count