Modeling Earthquake Damage in Nepal

Jessely Enriquez

Gorkha Earthquake on April 25, 2015

About the earthquake:

- 7.8Mw (moment magnitude)
- Near Kathmandu
  - central city in Nepal

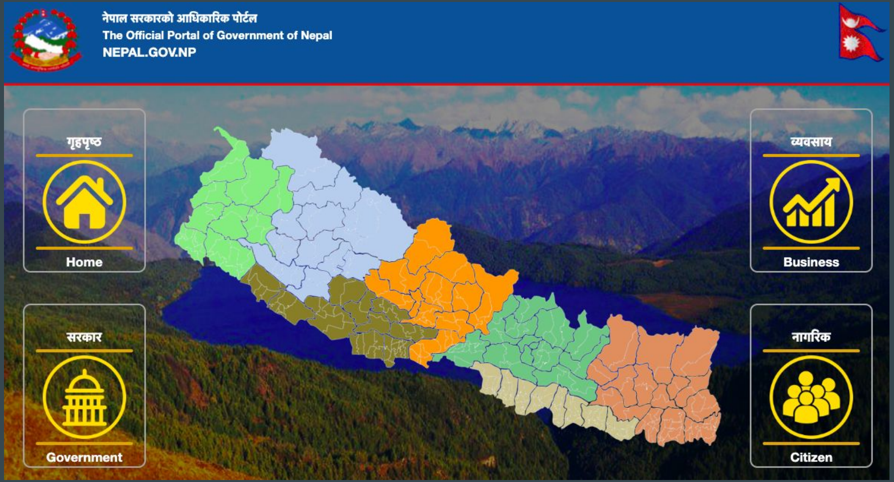
Impact:

- 9,000 lives lost
- 100,00 injuries

Massive Household Survey ~ 762,106 buildings, 11 districts, 77+ municipalities

- primary goal of to identify **beneficiaries** eligible for government assistance for housing reconstruction
- Assessed building damage in the earthquake-affected districts
- also collected census-level socio-economic information

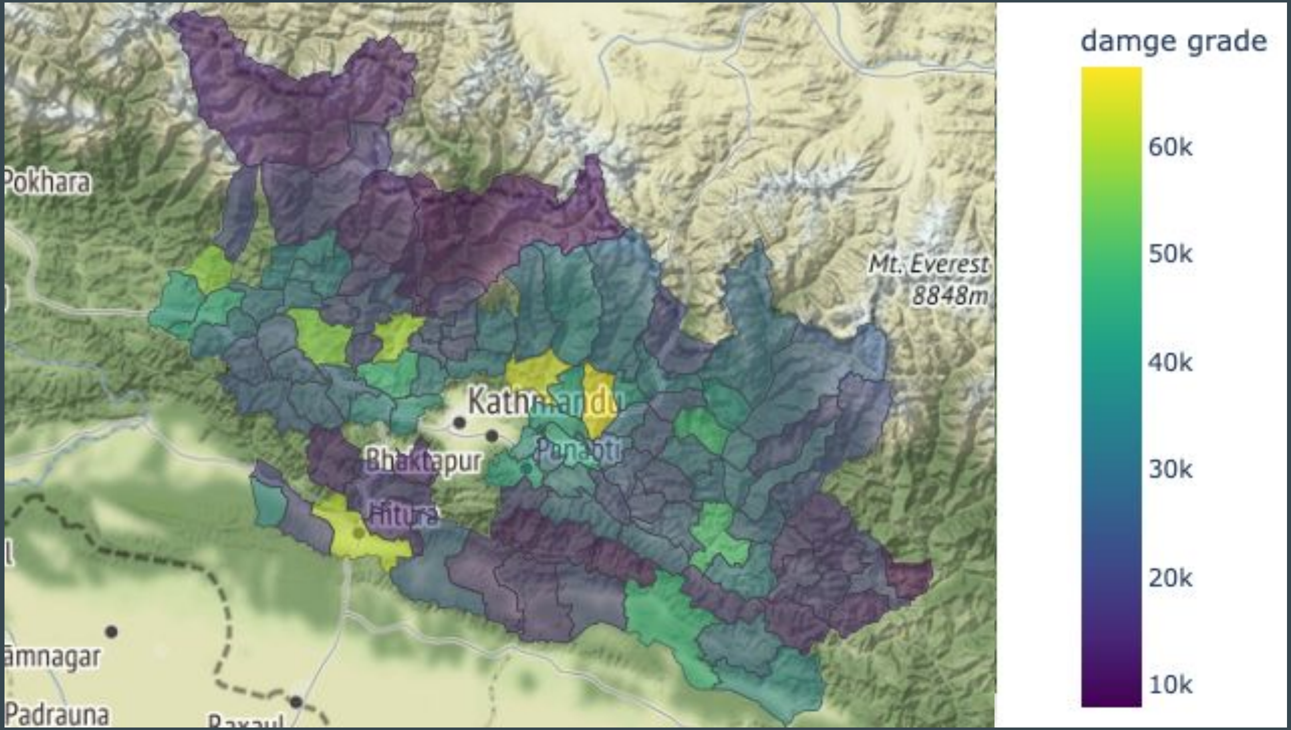
2015 NEPAL EARTHQUAKE OPEN DATA PORTAL



Problem Statement

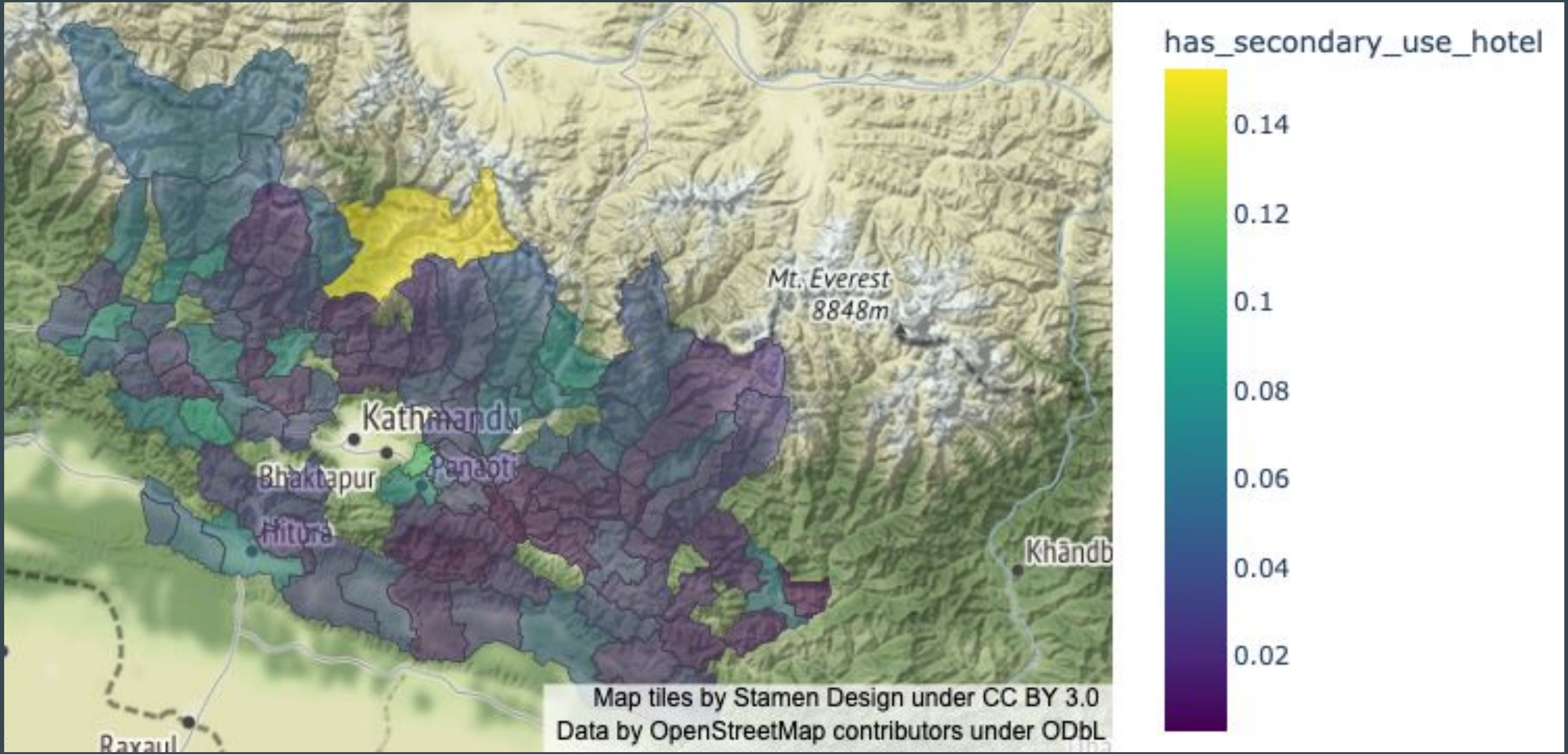
- In 2015 USAID trained locals to rebuild buildings
- This model then can be used
    - predict potential building damage to identify if those buildings created post earthquake are in need of more technical retrofitting at this time

Target Variable

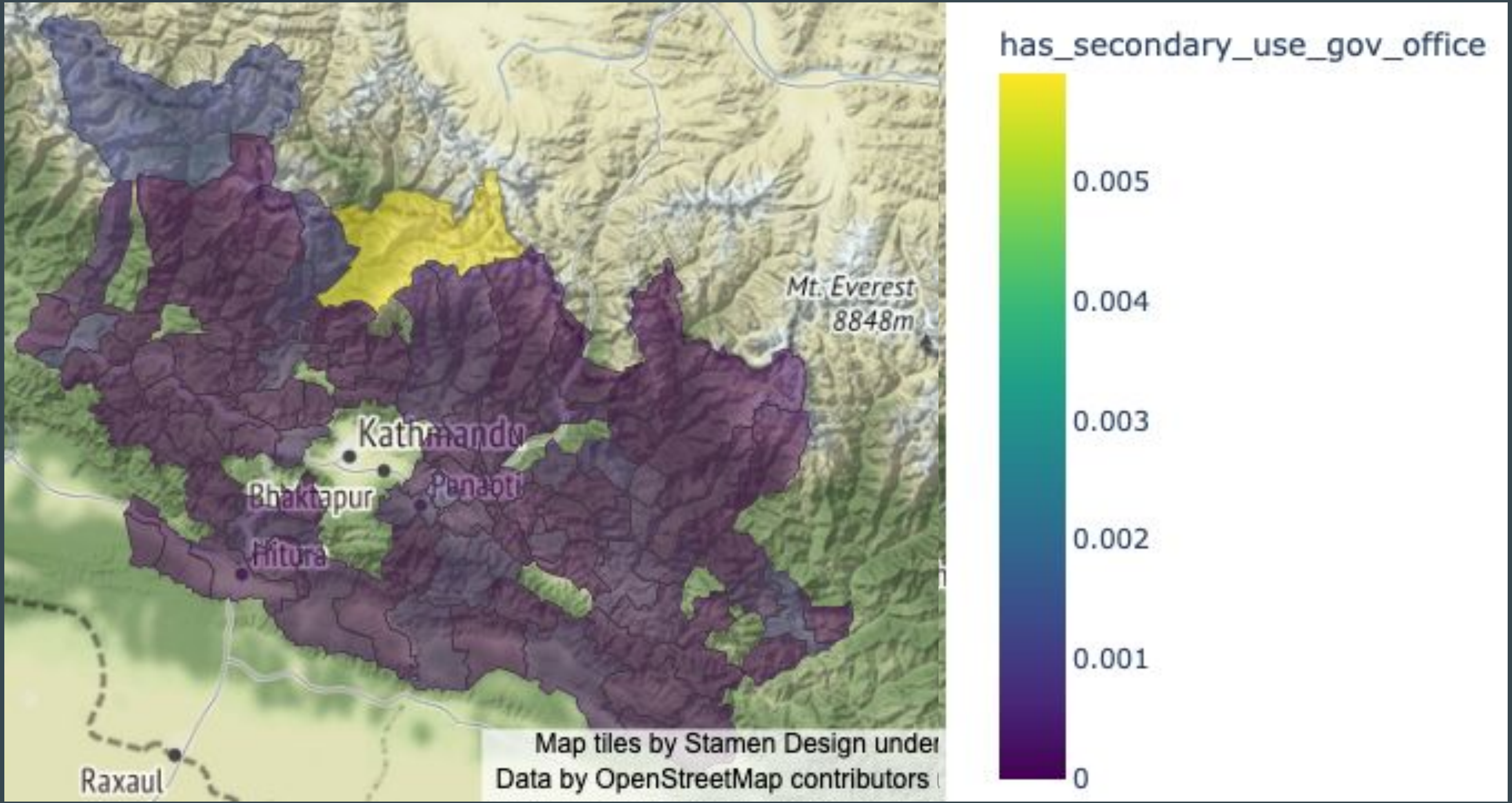


Damage Grade	Percent of Data
5 Total Collapse	36.1 %
4	24.1 %
3	17.9 %
3	17.9 %
2	11.4 %
1 Hairline cracks	10.3 %

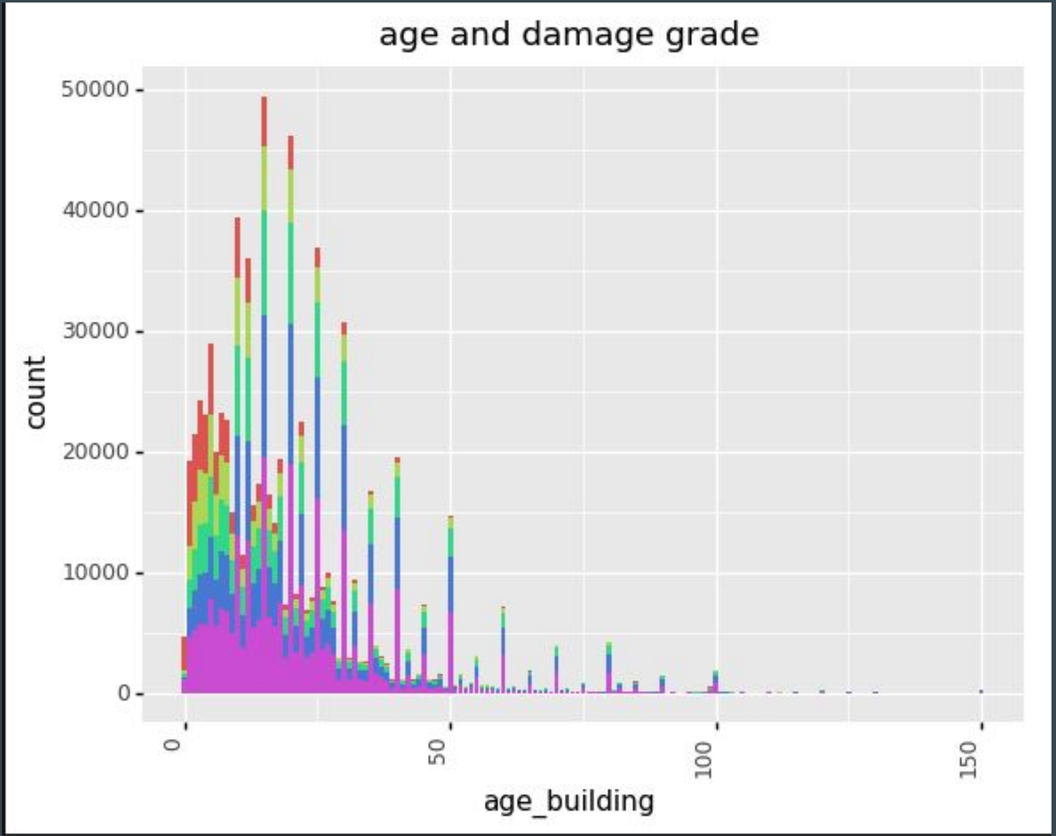
Looking at Densities of Secondary Building Usage



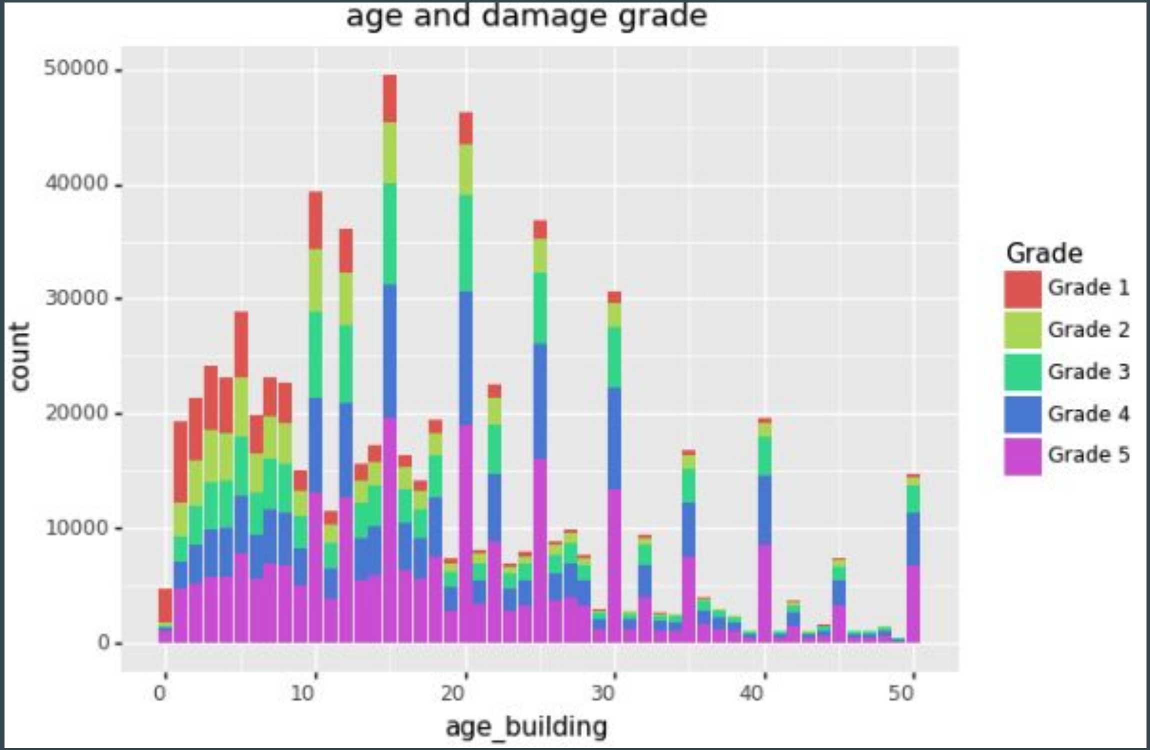
Looking at Densities of Secondary Building Usage Cont'd



Exploration

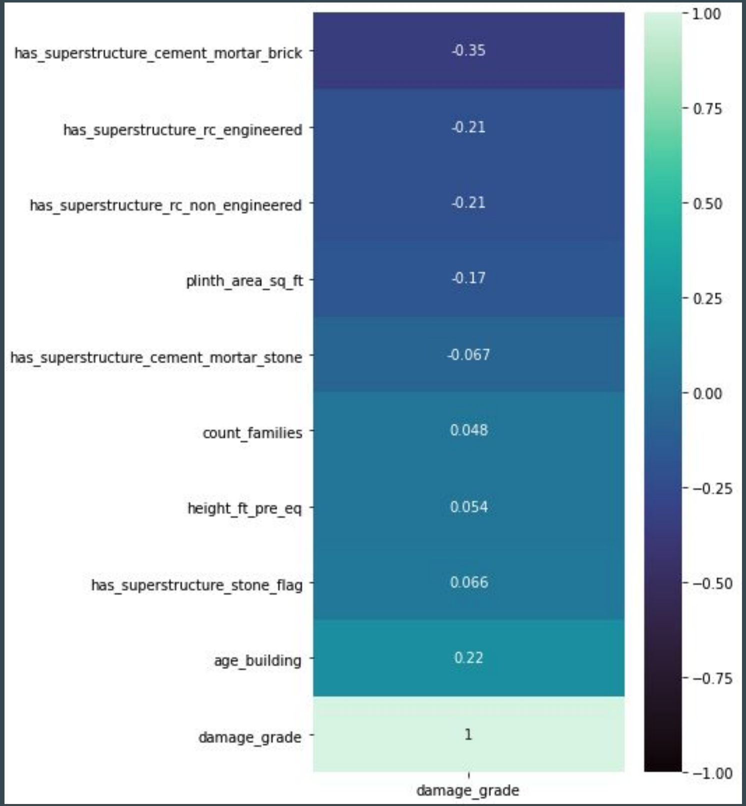


Zooming in ...



age and damage grade
50000
40000
30000
20000
10000
0
count
0
10
20
30
40
50
age_building
Grade
Grade 1
Grade 2
Grade 3
Grade 4
Grade 5

Linear Relationships ...?

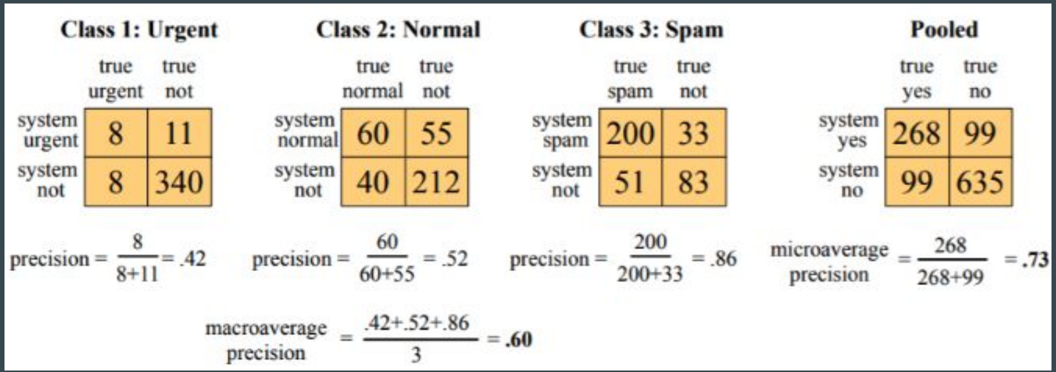


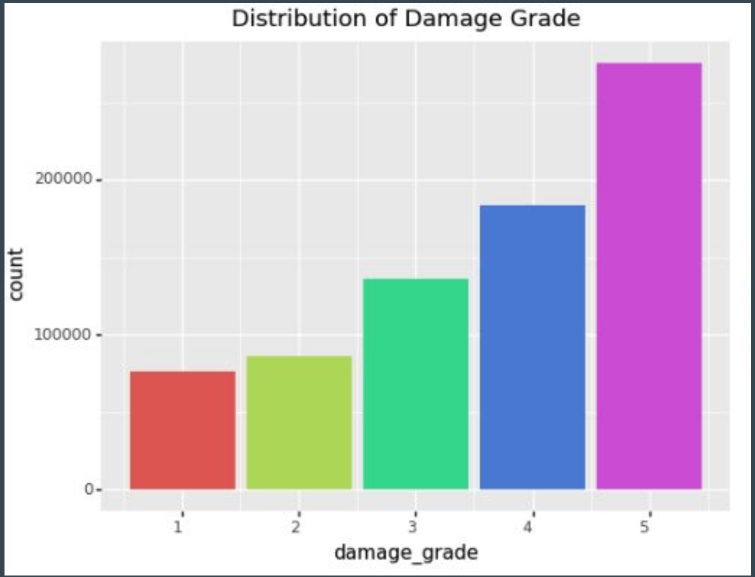
PSA: Correlation is not causation

Model Evaluation Metric

F1 Score:

- Macro
- Micro





Modeling Overview

- Logistic Regression
- Random Forest
- XGBoost

Best Model:

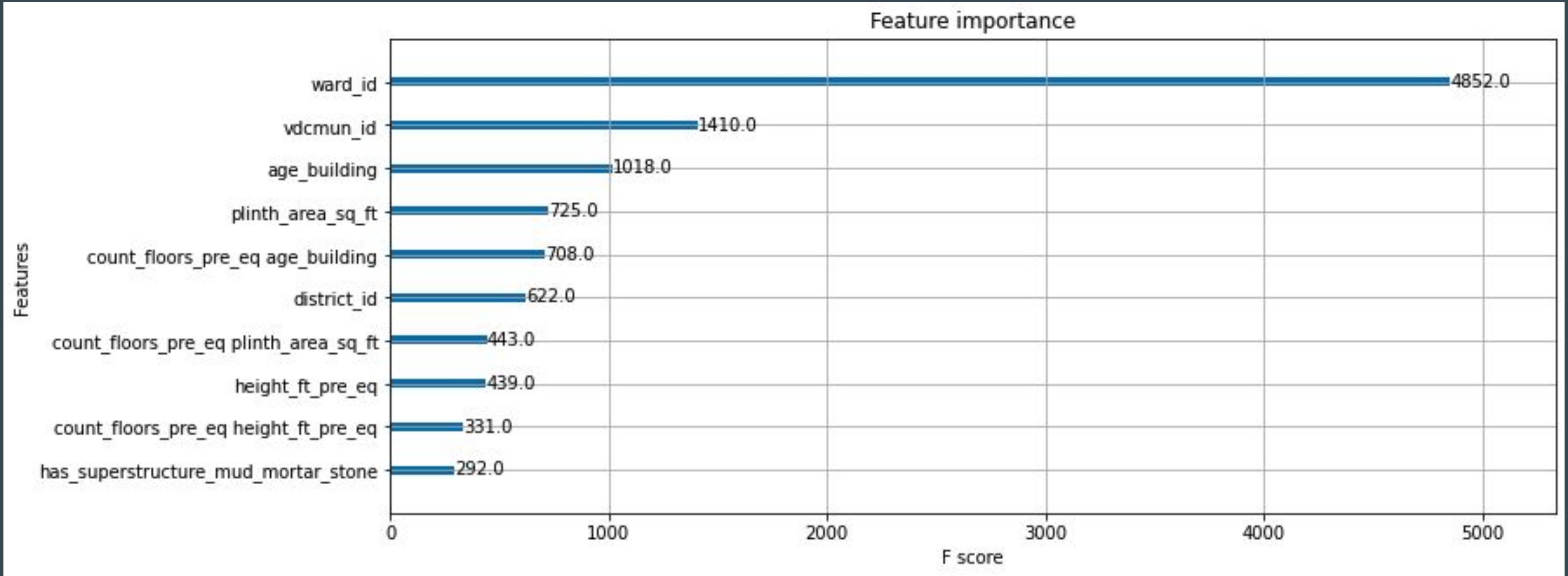
- Logistic Regression (multi_class='multinomial', C=0.1, solver='lbfgs')

Feature Engineering Attempt:

- Feature Interactions (100 to 200)

Feature Interactions

- Models with Polynomial Features performed similarly (the interactions were making top 10)



Important Features (Log Reg)

Coefficients	Word
-0.589	vdcmun_id
-0.553	ward_id
-0.358	has_superstructure_mud_mortar_stone
-0.16	has_secondary_use_agriculture
-0.12	count_floors_pre_eq

Coefficients	Word
1.0514	district_id
0.19	roof_type_bamboo_timber_light_roof
0.189	roof_type_rcc_rb_rbc
0.175	has_secondary_use
0.13	ground_floor_type_rc

Model Evaluation

- F1 Score Micro/Macro & Accuracy for reference
  - Hardest class to classify was buildings with damage grade 3

Metric	Logistic Regression	XGBoost	Random Forest
Accuracy	66%	58%	56%
Micro	66%	58%	56%
Macro	64%	54%	52%

Conclusion

- Functional model to utilize for the damage predictions to identify those in need of more technical retrofitting
    - Out performs baseline by about 30%

Future Work:

- Could utilize the geo coordinates to gather geospatial characteristics for each location
- Build a functional tool that is user friendly for locals to use in building eval

Questions ?

Thank you!

Streamlit multipage notes

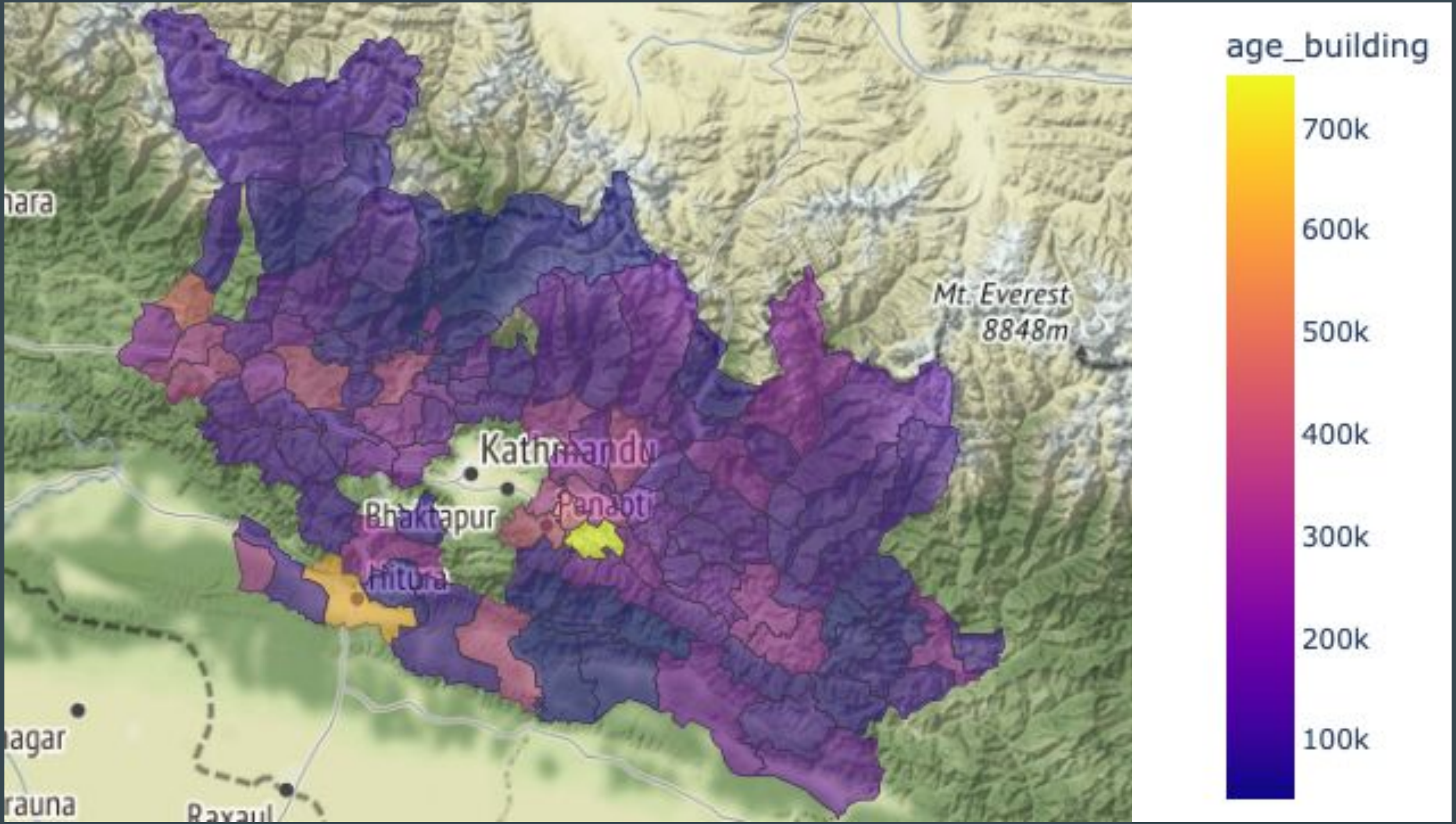
https://towardsdatascience.com/creating-multipage-applications-using-streamlit-efficiently-b58a58134030

https://towardsdatascience.com/a-multi-page-interactive-dashboard-with-streamlit-and-plotly-c3182443871a

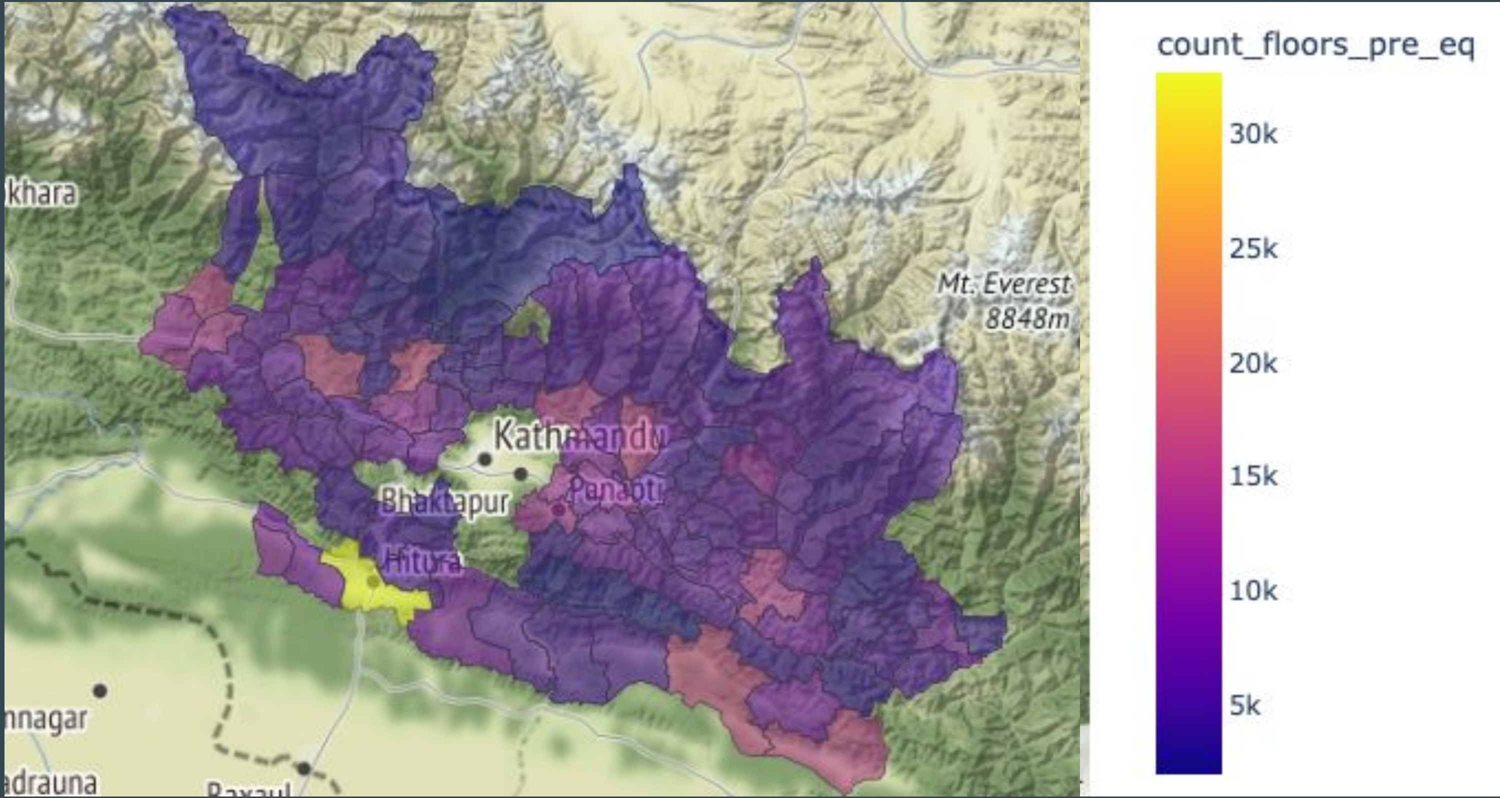
Data resource:

https://observablehq.com/collection/@arkoblog/opendataportal

Distribution of Building Age



Distribution of Count of Floors Pre - Earthquake



Distribution of Count of Floors Post - Earthquake