

Richter's Predictor: Modeling Earthquake Damage

PROBLEM STATEMENT

Our problem statement was published as an online challenge on driven data so as to develop Richter's predictor, a model which establishes relationships between various factors which have a direct influence on the damage incurred to buildings.

DRIVEN DATA

DATA DESCRIPTION

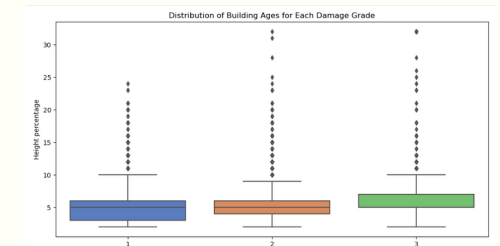
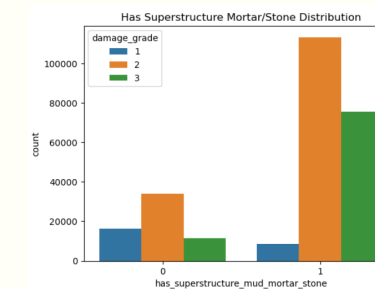
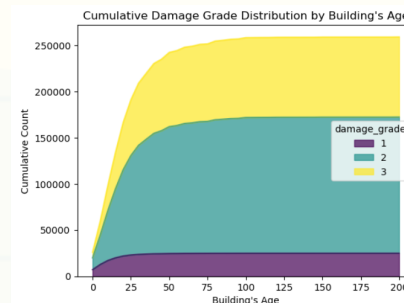
The dataset covers 39 parameters and 26061 sample spaces, including the building's age, number of floors, materials used in construction, height percentage and geographical location etc. The dataset is well-formatted and simple to use, and there are no missing values.

HYPOTHESIS

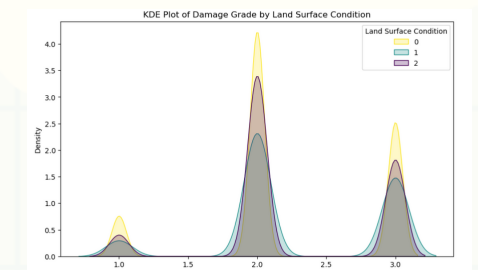
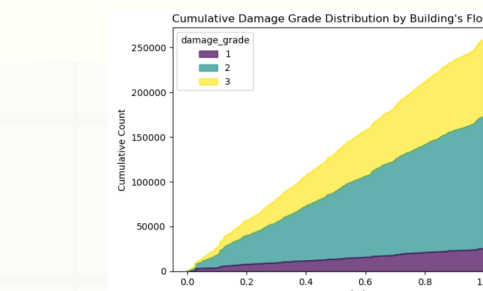
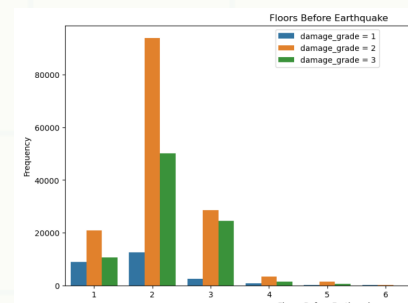
- Is the choice of construction materials a determining factor in the damage sustained by the building
- Which features within the dataset are crucial in determining the extent of damage
- Among the building's floors, which ones sustained the highest level of damage?
- In what ways did the geographical region contribute to the damage incurred by the building?
- What is the correlation between the number of family members residing in the building and the family size most significantly impacted?
- What parts of the building, in terms of its area footprint and height footprint, suffered the greatest impact from the earthquake?
- Does the age of a building play a role in determining the amount of damage it is likely to experience?
- Which ground types were most susceptible to causing significant damage to the buildings?
- To what extent did the building's foundation either alleviate or contribute to the damage incurred?
- Is the positioning of the building a contributing factor to the damage it inflicts upon itself?

EXPLORATORY DATA ANALYSIS

- The kernel trick graph demonstrates that as a building's age increases, so does the likelihood of destruction.
- Mud mortar stone constructions have higher sustainability and are less prone to withstand damage than other materials.
- Buildings with a higher percentage of height are more prone to experience earthquake damage because they are more susceptible to the pressures of the shaking.



- The building's second floor was the most seriously destroyed, followed by the third floor.



PRE PROCESSING TECHNIQUES

One-Hot Encoding :

In data preprocessing, one-hot encoding efficiently transformed categorical variables into numerical format. we used one-hot encoding to convert out categorical values such as land_surface_condition, roof_type, plan_configuration, legal_ownership_status etc.



Min-Max Normalization :

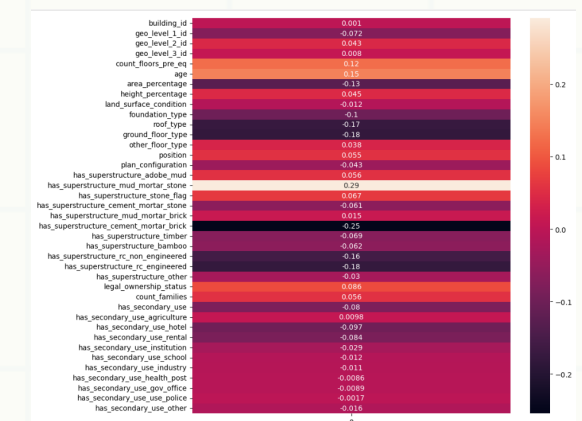
We used min-max normalization to normalize the geo_level_id_1, geo_level_id_2, and geo_level_id_3 columns. This guarantees that the scale and distribution of these columns are consistent.

$$v'_i = \frac{v_i - \min}{\max - \min} (\text{new_max} - \text{new_min}) + \text{new_min}$$

Outlier Management :

We looked for outliers in the columns age_of_building, area_percentage and height_percentage. We removed these outliers from the dataset by using a box plot to analyze them and then trimming them.

- A building's age, number of floors, materials used in superstructure construction, height percentage and geographical location are all important factors in influencing the level of damage it sustains.



TEAM A9

Chandan S -02FE21BCS021 Veeresh R- 02FE21BCS104
Manasi M - 02FE21BCS045 Shreya D- 02FE21BCS080