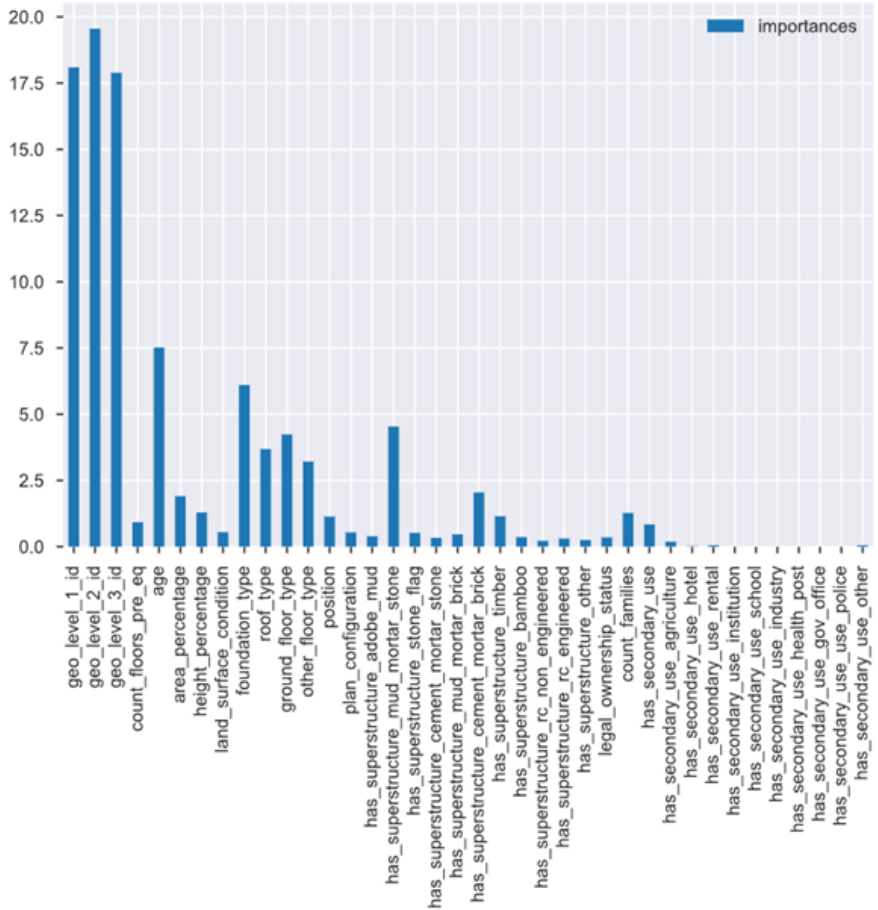
Automated Feature Engineering

The features geo_level_1_id, geo_level_2_id, and geo_level_3_id are poorly interpretable in the dataset. These features represent different areas (geographical regions) of Kathmandu with varying levels of granularity, ranging from broader to more specific.

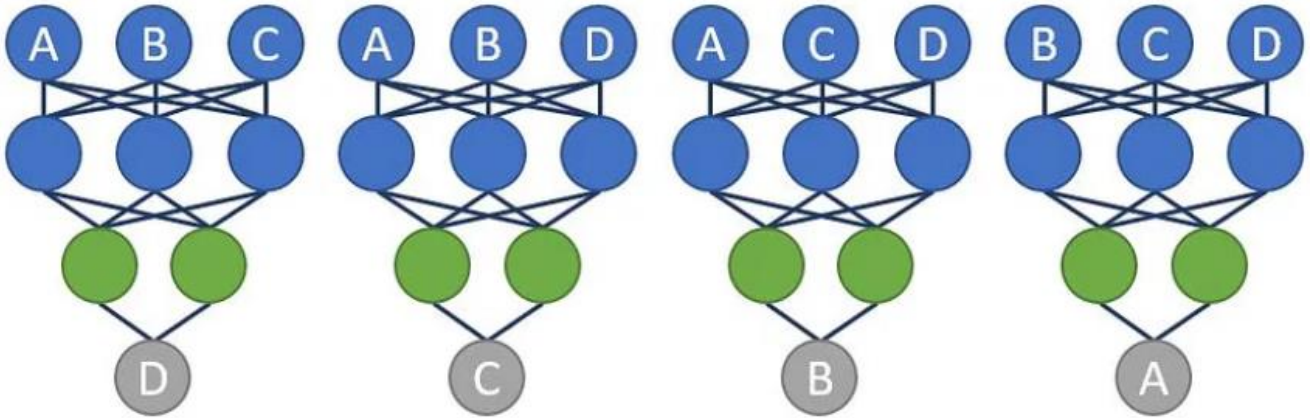
However, despite their lack of interpretability, these features contribute the most to the model's outcome.



Automated Feature Engineering

Idea:
Augment the dataset with features that somehow reflect the dependencies between these "encoded" geographical regions.

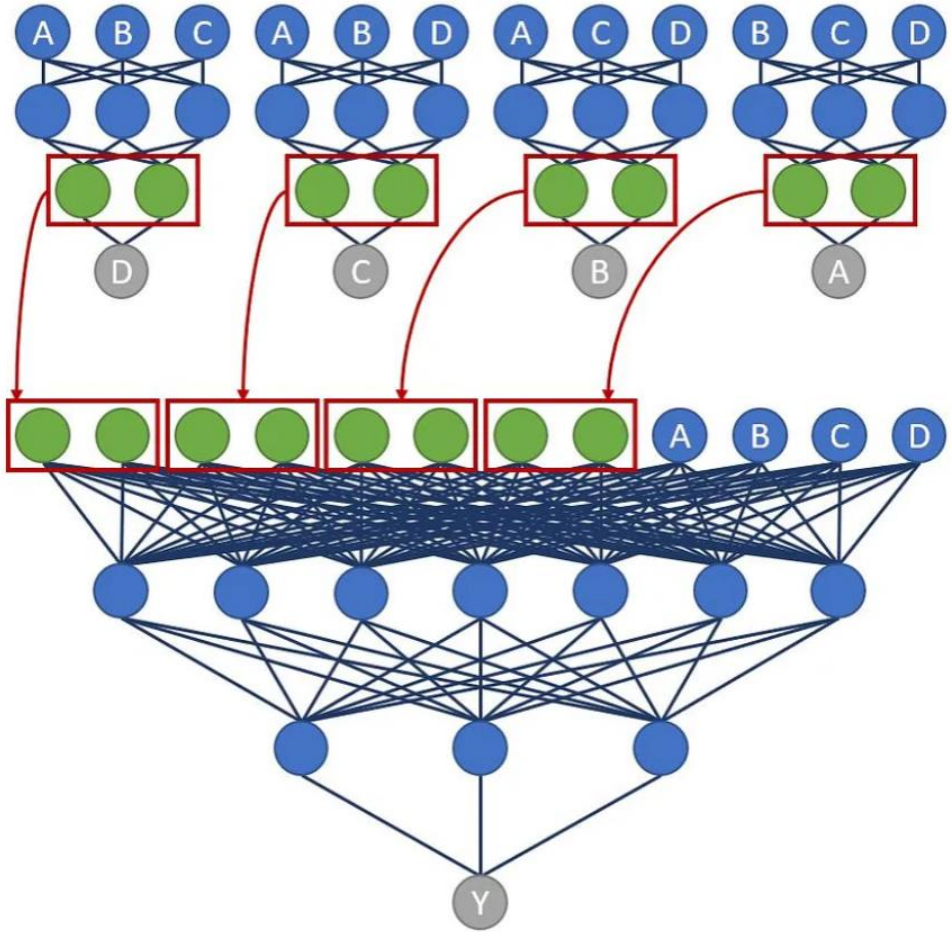
Neural networks excel at creating various feature representations of an object, providing more accurate descriptions of dependencies in the data



Automated Feature Engineering

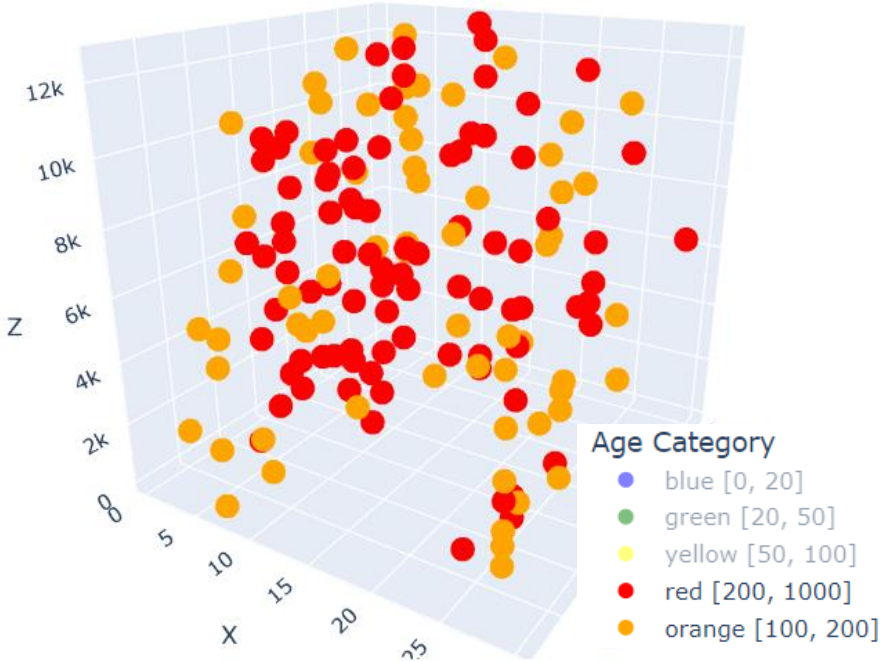
The last layer contains our new features that describe the dependencies between the input features passed to the neural network, including a selective feature from the dataset

Model	RMSE	Cross Entropy	Error
Auto-feature	0.4314893	0.55944	37.2%
MLP - emb	0.4315610	0.55960	37.6%
XGBoost	0.4316558	0.55982	37.6%

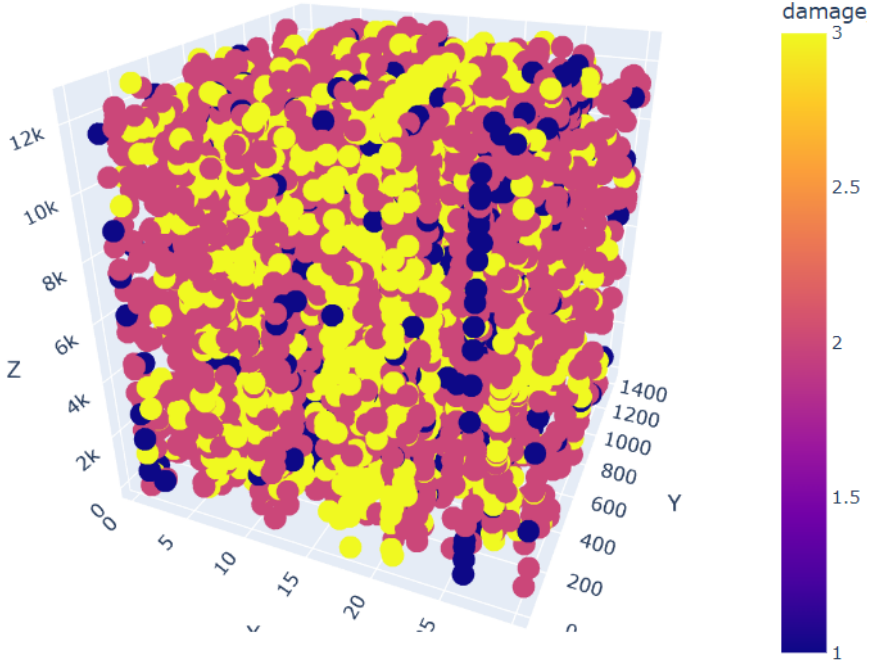


More about geo_level_id

Age Distribution of Buildings by geo_level_id

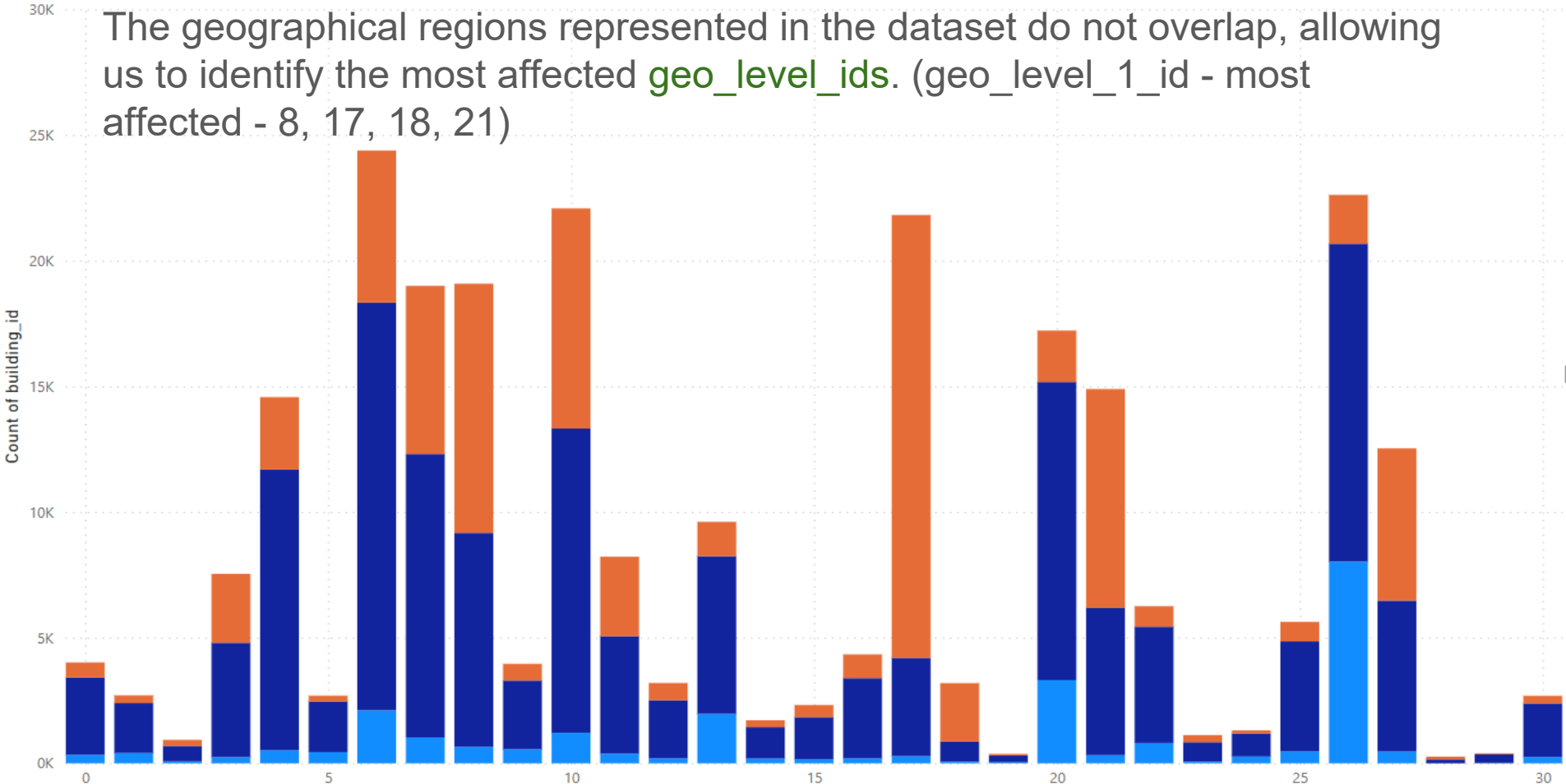


Distribution of Buildings by Destruction Severity across geo_level_id



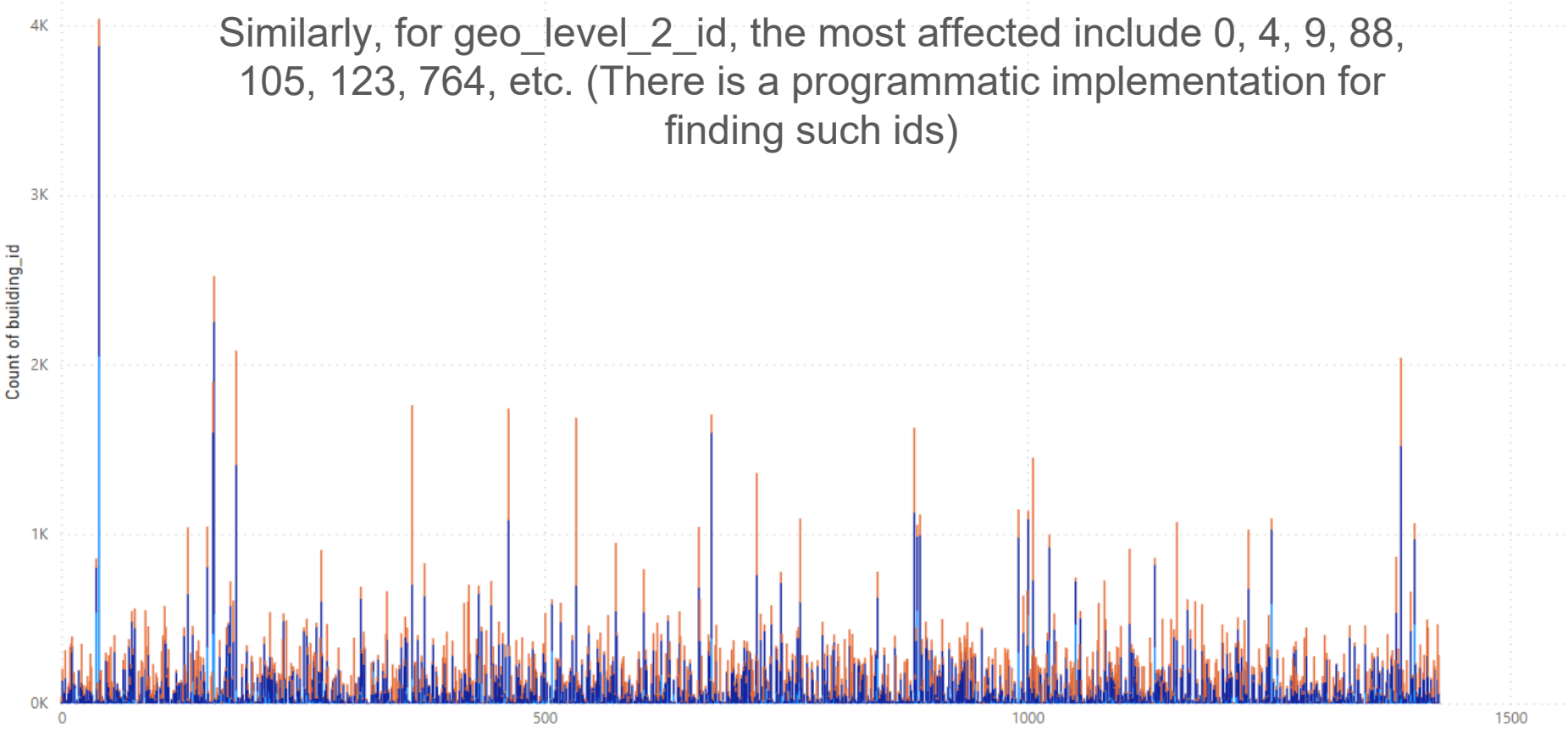
More about geo_level_id

The geographical regions represented in the dataset do not overlap, allowing us to identify the most affected geo_level_ids. (geo_level_1_id - most affected - 8, 17, 18, 21)



More about geo_level_id

Similarly, for geo_level_2_id, the most affected include 0, 4, 9, 88, 105, 123, 764, etc. (There is a programmatic implementation for finding such ids)



More about geo_level_id

Contrary to the fact that specific building locations cannot be extracted from these features, we are still able to make presumptive identifications of certain existing areas within the city in the dataset. Additionally, we can make assumptions about destructibility.

The goal was to find some significant, well-known building, such as the Kasthamandap Temple

Basing our findings on the material - wood, and considering the age of the buildings, along with the fact that the temple was previously used as a hotel, we have tentatively identified this temple (and the Durbar Square itself) in our dataset.

About the Temple

It is believed that the Kasthamandap Temple was constructed from a single Sal tree trunk, known for its enhanced strength and resistance to environmental factors.

This three-tiered pagoda was formerly used as a shelter for merchants and travelers journeying from Tibet to India. Such small inns in the Kathmandu Valley were not uncommon.

About the materials (has_superstructure_…)

№1

Reinforced concrete buildings tend to collapse primarily due to the soft lower floor, which lacks rigid walls (usually the commercial floor, with 90% reinforcement).

- Roller shutters are also common (some shops use a curtain instead of a door).
- Upper floors are typically made of brick masonry, which is somewhat less stable.
- In multi-story buildings, there is usually a ground-level parking floor with few internal walls, leading to less stability and increased vulnerability.

Important features in the dataset - **ground_floor_type, height_percentage, other_floor_type**

About the materials (has_superstructure_…)

№2

The proximity of buildings really strongly affects the destructibility

Important features in dataset - **count_floors_pre_eq**, **position**

№3

Heavy roof contributes to additional load (tile coverings).It is especially important for weak houses made, for example, of some kind of clay and mud

Important features in dataset - **roof_type**

About the matetials (has_superstructure_…)

№4

Buildings with a random roof are quite vulnerable. Buildings that had not only stone, but also wood are more stable

Important features in dataset - **has_superstructure_is_…**

№5

It is necessary to pay attention to the height of the building and the material. For example, three-storey clay (adobe) are more vulnerable

Important features in dataset - **has_superstructure_is_…, count_floor**

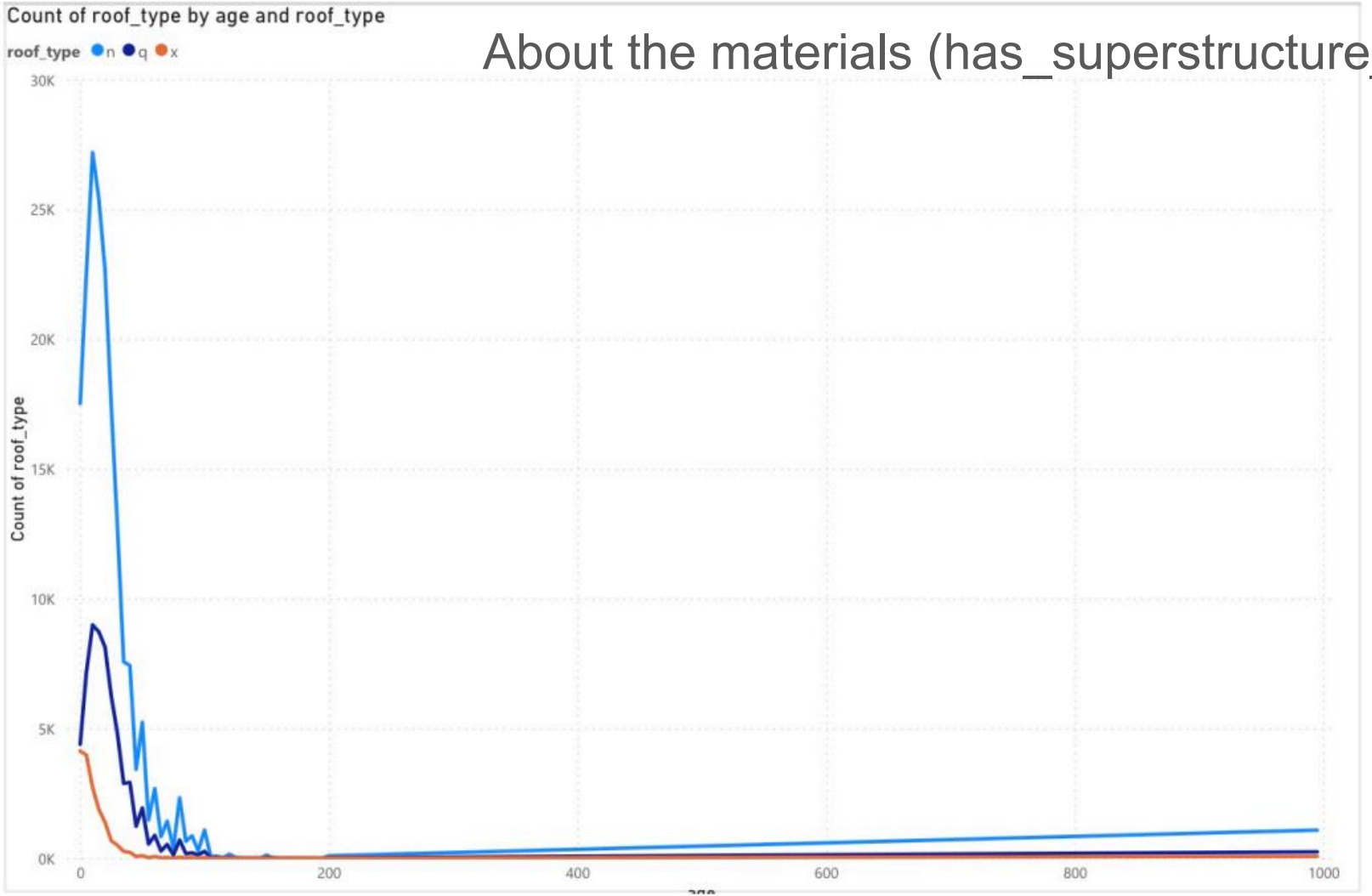
About the materials (has_superstructure_…)

№6

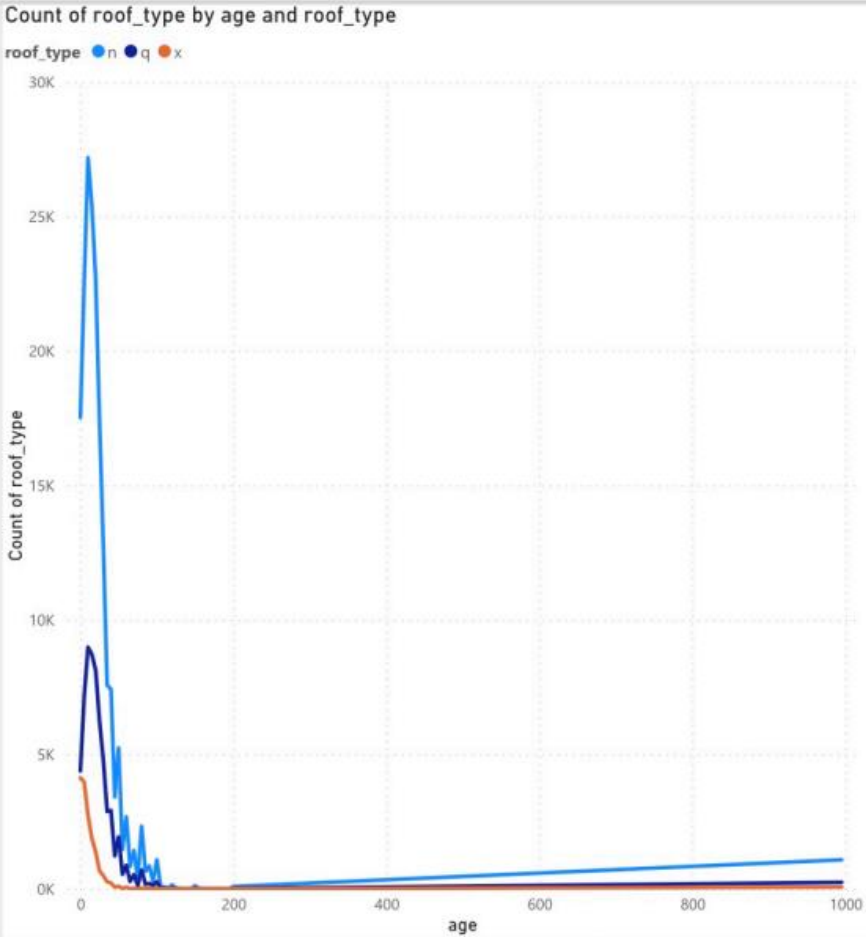
For unreinforced buildings, the main causes of complete collapse include the accumulation of significant loads (roof, the building, and furnishings), age, and a lack of reinforcements that could help the building withstand horizontal seismic waves. The earthquake shakes not only vertically but also horizontally.

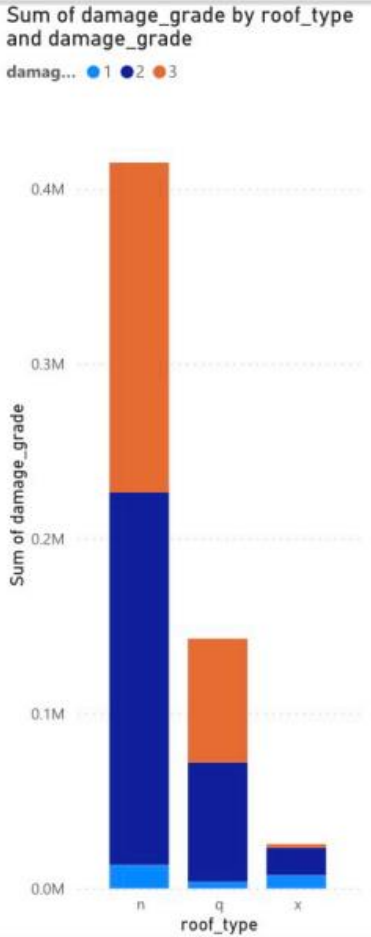
Important features in dataset - **age**, **has_superstructure_is_…**

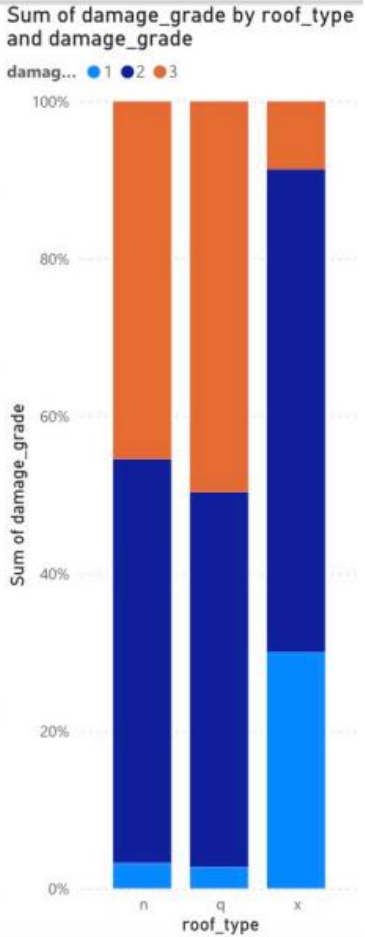
About the materials (has_superstructure_…)

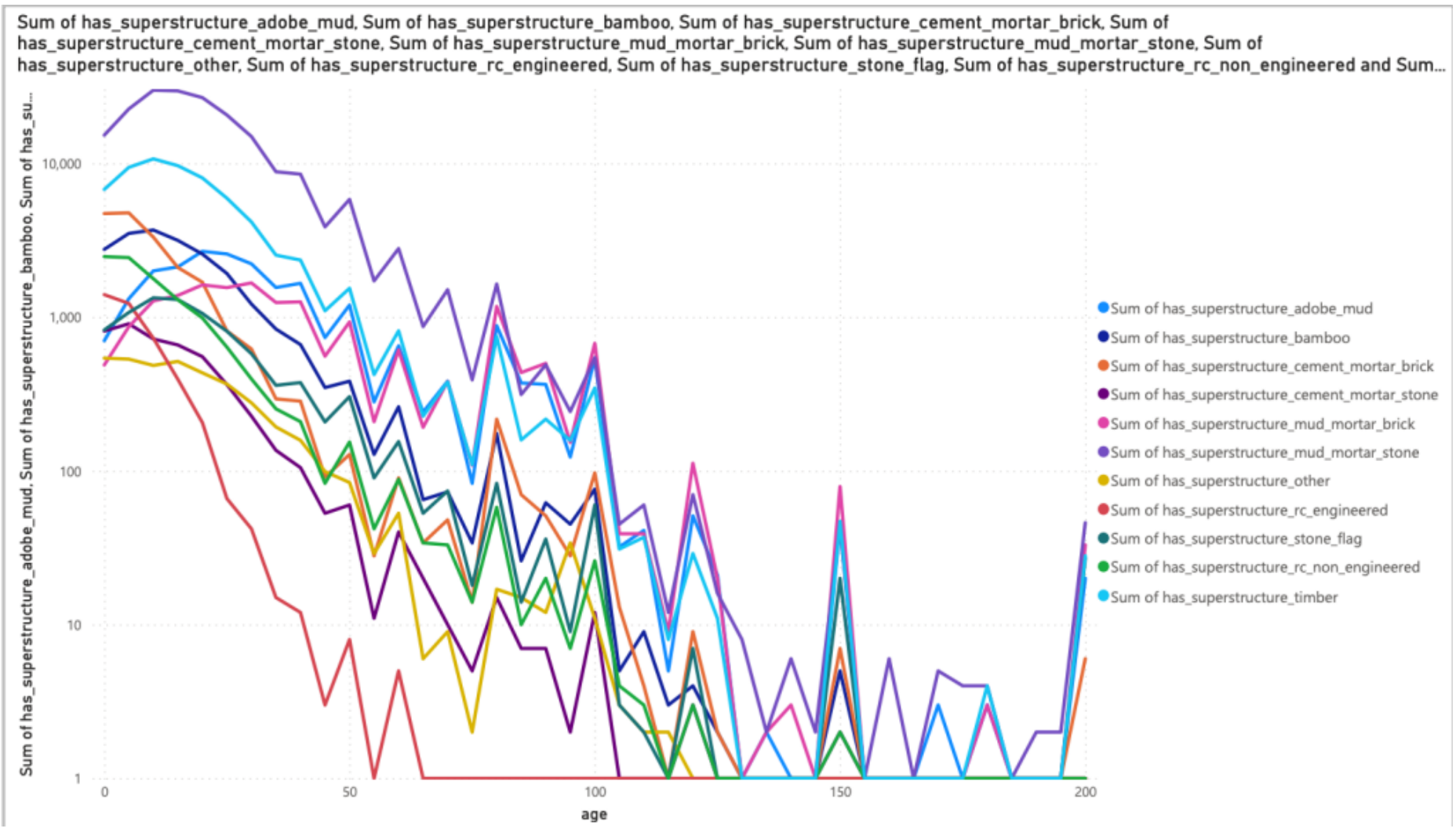


About the materials (has_superstructure_…)








Sum of has_superstructure_adobe_mud, Sum of has_superstructure_bamboo, Sum of has_superstructure_cement_mortar_brick, Sum of has_superstructure_cement_mortar_stone, Sum of has_superstructure_mud_mortar_brick, Sum of has_superstructure_mud_mortar_stone, Sum of has_superstructure_other, Sum of has_superstructure_rc_engineered, Sum of has_superstructure_stone_flag, Sum of has_superstructure_rc_non_engineered and Sum...
Sum of has_superstructure_adobe_mud, Sum of has_superstructure_bamboo, Sum of has_superstructure_bamboo, Sum of has_su...
10,000
1,000
100
10
1
0
50
100
150
200
age
Sum of has_superstructure_adobe_mud
Sum of has_superstructure_bamboo
Sum of has_superstructure_cement_mortar_brick
Sum of has_superstructure_cement_mortar_stone
Sum of has_superstructure_mud_mortar_brick
Sum of has_superstructure_mud_mortar_stone
Sum of has_superstructure_other
Sum of has_superstructure_rc_engineered
Sum of has_superstructure_stone_flag
Sum of has_superstructure_rc_non_engineered
Sum of has_superstructure_timber