# Overview

Based on aspects of building location and construction, your goal is to predict the level of damage to buildings caused by the 2015 Gorkha earthquake in Nepal.

The data was collected through surveys by [Kathmandu Living Labs](http://www.kathmandulivinglabs.org/) and the [Central Bureau of Statistics](https://cbs.gov.np/), which works under the National Planning Commission Secretariat of Nepal. This survey is one of the largest post-disaster datasets ever collected, containing valuable information on earthquake impacts, household conditions, and socio-economic-demographic statistics.

# Problem description

We're trying to predict the [ordinal variable](https://www.ma.utexas.edu/users/mks/statmistakes/ordinal.html) damage\_grade, which represents a level of damage to the building that was hit by the earthquake. There are 3 grades of the damage:

* 1 represents low damage
* 2 represents a medium amount of damage
* 3 represents almost complete destruction

## Features

The dataset mainly consists of information on the buildings' structure and their legal ownership. Each row in the dataset represents a specific building in the region that was hit by Gorkha earthquake.

There are 39 columns in this dataset, where the building\_id column is a unique and random identifier. The remaining 38 features are described in the section below. Categorical variables have been obfuscated random lowercase ascii characters. The appearance of the same character in distinct columns does not imply the same original value.

## Description

* geo\_level\_1\_id, geo\_level\_2\_id, geo\_level\_3\_id (type: int): geographic region in which building exists, from largest (level 1) to most specific sub-region (level 3). Possible values: level 1: 0-30, level 2: 0-1427, level 3: 0-12567.
* count\_floors\_pre\_eq (type: int): number of floors in the building before the earthquake.
* age (type: int): age of the building in years.
* area\_percentage (type: int): normalized area of the building footprint.
* height\_percentage (type: int): normalized height of the building footprint.
* land\_surface\_condition (type: categorical): surface condition of the land where the building was built. Possible values: n, o, t.
* foundation\_type (type: categorical): type of foundation used while building. Possible values: h, i, r, u, w.
* roof\_type (type: categorical): type of roof used while building. Possible values: n, q, x.
* ground\_floor\_type (type: categorical): type of the ground floor. Possible values: f, m, v, x, z.
* other\_floor\_type (type: categorical): type of constructions used in higher than the ground floors (except of roof). Possible values: j, q, s, x.
* position (type: categorical): position of the building. Possible values: j, o, s, t.
* plan\_configuration (type: categorical): building plan configuration. Possible values: a, c, d, f, m, n, o, q, s, u.
* has\_superstructure\_adobe\_mud (type: binary): flag variable that indicates if the superstructure was made of Adobe/Mud.
* has\_superstructure\_mud\_mortar\_stone (type: binary): flag variable that indicates if the superstructure was made of Mud Mortar - Stone.
* has\_superstructure\_stone\_flag (type: binary): flag variable that indicates if the superstructure was made of Stone.
* has\_superstructure\_cement\_mortar\_stone (type: binary): flag variable that indicates if the superstructure was made of Cement Mortar - Stone.
* has\_superstructure\_mud\_mortar\_brick (type: binary): flag variable that indicates if the superstructure was made of Mud Mortar - Brick.
* has\_superstructure\_cement\_mortar\_brick (type: binary): flag variable that indicates if the superstructure was made of Cement Mortar - Brick.
* has\_superstructure\_timber (type: binary): flag variable that indicates if the superstructure was made of Timber.
* has\_superstructure\_bamboo (type: binary): flag variable that indicates if the superstructure was made of Bamboo.
* has\_superstructure\_rc\_non\_engineered (type: binary): flag variable that indicates if the superstructure was made of non-engineered reinforced concrete.
* has\_superstructure\_rc\_engineered (type: binary): flag variable that indicates if the superstructure was made of engineered reinforced concrete.
* has\_superstructure\_other (type: binary): flag variable that indicates if the superstructure was made of any other material.
* legal\_ownership\_status (type: categorical): legal ownership status of the land where building was built. Possible values: a, r, v, w.
* count\_families (type: int): number of families that live in the building.
* has\_secondary\_use (type: binary): flag variable that indicates if the building was used for any secondary purpose.
* has\_secondary\_use\_agriculture (type: binary): flag variable that indicates if the building was used for agricultural purposes.
* has\_secondary\_use\_hotel (type: binary): flag variable that indicates if the building was used as a hotel.
* has\_secondary\_use\_rental (type: binary): flag variable that indicates if the building was used for rental purposes.
* has\_secondary\_use\_institution (type: binary): flag variable that indicates if the building was used as a location of any institution.
* has\_secondary\_use\_school (type: binary): flag variable that indicates if the building was used as a school.
* has\_secondary\_use\_industry (type: binary): flag variable that indicates if the building was used for industrial purposes.
* has\_secondary\_use\_health\_post (type: binary): flag variable that indicates if the building was used as a health post.
* has\_secondary\_use\_gov\_office (type: binary): flag variable that indicates if the building was used fas a government office.
* has\_secondary\_use\_use\_police (type: binary): flag variable that indicates if the building was used as a police station.
* has\_secondary\_use\_other (type: binary): flag variable that indicates if the building was secondarily used for other purposes.

### Feature data example

Here's an example of one of the rows in the dataset so that you can see the kinds of values you might expect in the dataset. Some are numeric, some are categorical.

| field | value |
| --- | --- |
| geo\_level\_1\_id | 8 |
| geo\_level\_2\_id | 396 |
| geo\_level\_3\_id | 1108 |
| count\_floors\_pre\_eq | 2 |
| age | 15 |
| area\_percentage | 4 |
| height\_percentage | 7 |
| land\_surface\_condition | t |
| foundation\_type | r |
| roof\_type | n |
| ground\_floor\_type | v |
| other\_floor\_type | q |
| position | s |
| plan\_configuration | d |
| has\_superstructure\_adobe\_mud | 1 |
| has\_superstructure\_mud\_mortar\_stone | 1 |
| has\_superstructure\_stone\_flag | 0 |
| has\_superstructure\_cement\_mortar\_stone | 0 |
| has\_superstructure\_mud\_mortar\_brick | 0 |
| has\_superstructure\_cement\_mortar\_brick | 1 |
| has\_superstructure\_timber | 0 |
| has\_superstructure\_bamboo | 0 |
| has\_superstructure\_rc\_non\_engineered | 0 |
| has\_superstructure\_rc\_engineered | 0 |
| has\_superstructure\_other | 1 |
| legal\_ownership\_status | v |
| count\_families | 1 |
| has\_secondary\_use | 0 |
| has\_secondary\_use\_agriculture | 0 |
| has\_secondary\_use\_hotel | 0 |
| has\_secondary\_use\_rental | 0 |
| has\_secondary\_use\_institution | 0 |
| has\_secondary\_use\_school | 0 |
| has\_secondary\_use\_industry | 0 |
| has\_secondary\_use\_health\_post | 0 |
| has\_secondary\_use\_gov\_office | 0 |
| has\_secondary\_use\_use\_police | 0 |
| has\_secondary\_use\_other | 0 |