KLE Technological University
Hubballi

A Course Project Report on

"Data-Driven Analysis On Modelling Earthquake Damage"

A Course Project Report Submitted in Partial Fulfillment of the Requirement for the Course of

Exploratory Data Analysis

in

4^{th} Semester of Computer Science and Engineering

by

Chandan Satwani	(02FE21BCS021)
Manasi Maridevarmath	(02FE21BCS045)
Shreya Dugani	(02FE21BCS080)
Veeresh Rathod	(02FE21BCS104)

Under the guidance of

Dr. Santosh Pattar

Assistant Professor,

Department of Computer Science and Engineering,

KLE Technological University's Dr. MSSCET, Belagavi.

KLE Technological University's

Dr. M. S. Sheshgiri College of Engineering and Technology,

Belagavi – 590 008.

July 2023

DECLARATION

We hereby declare that the matter embodied in this report entitled "**Data-Driven Analysis On Modelling Earthquake Damage**" submitted to KLE Technological University for the course completion of Exploratory Data Analysis (21ECSC210) in the 4^{th} Semester of Computer Science and Engineering is the result of the work done by us in the Department of Computer Science and Engineering, KLE Dr. M. S. Sheshgiri College of Engineering, Belagavi under the guidance of Dr. Santosh Pattar, Assistant Professor, Department of Computer Science and Engineering. We further declare that to the best of our knowledge and belief, the work reported here in doesn't form part of any other project on the basis of which a course or award was conferred on an earlier occasion on this by any other student(s), also the results of the work are not submitted for the award of any course, degree or diploma within this or in any other University or Institute. We hereby also confirm that all of the experimental work in this report has been done by us.

Belagavi – 590 008
Date :

Chandan Satwani (02FE21BCS021)	Shreya Duagni (02FE21BCS080)
Manasi Maridevarmath (02FE21BCS045)	Veeresh Rathod (02FE21BCS104)

CERTIFICATE

This is to certify that the project entitled "Data-Driven Analysis On Modelling Earthquake Damage" submitted to KLE Technological University's Dr. MSSCET, Belagavi for the partial fulfillment of the requirement for the course-Exploratory Data Analysis (21ECSC210) by Chandan Satwani (02FE21BCS021) , Manasi Maridevarmath (02FE21BCS045) , Shreya Dugani (02FE21BCS080) , Veeresh Rathod (02FE21BCS104) ,students in the Department of Computer Science and Engineering, KLE Technological University's Dr. MSSCET, Belagavi, is a bonafide record of the work carried out by them under my supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any other course completion.

Belagavi – 590 008
Date :

Dr. Santosh Pattar
(Course Teacher)

Prof. Priyanka Gavade
(Course Coordinator)

Dr. Rajashri Khanai
(Head of the Department)

Abstract

The Richter's predictor dataset grounded on the 2015 earthquake in the Gorkha region of Nepal appeared as a challenge on DRIVENDATA. In this analysis, we claw into a dataset obtained from the Central Bureau of Statistics and Kathmandu Living Labs. The dataset encompasses precious information collected through multitudinous checks conducted. The main idea of this analysis is to gain perceptive into the factors that contribute to the extent of earthquake damage. Through the combination of data analysis, sphere knowledge, and prophetic modeling, we aspire to develop accurate and dependable models that can help in prognosticating earthquake damage and eventually aid in disaster preparedness and response.

Through this design we aim to develop Richter's Predictor, a prophetic model for directly assessing earthquake damage. By using the dataset we seek to produce a robust and dependable model that considers various factors such as erecting structure, configuration, and position through a data-driven perceptive(i.e EDA) Results.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Earthquakes are natural marvels characterized by the unforeseen release of seismic energy in the Earth's crust, performing in seismic swells. These seismic events can beget significant damage to structures, infrastructure and pose trouble to mortal life.

Figure 1.1: A Structurally Damaged Building due to Earthquake

1.2 Problem Statement

The Richter's predictor dataset predicated on the 2015 earthquake in the Gorkha region of Nepal appeared as a challenge to DRIVENDATA. The earthquake bandied in this analysis is the one that struck the Gorkha region of Nepal in 2015. This earthquake passed on April 25, 2015, with a magnitude of 7.8 on the Richter scale.

1.2.1 Objectives

The objectives for the problem statement are:

1. To read the extent of structure damage caused by the 2015 Gorkha earthquake in Nepal. This thing can be met by creating a machine literacy model that can learn from data and estimate the extent of damage to a structure grounded on its position, construction accouterments, and other variables. The model's closeness, or how well it forecasts the factual extent of structure damage, will be assessed. The model can be used to understand the factors that contribute to earthquake damage and to establish better structure norms and construction practices.

2. Determine the factors that contribute to the extent of building damage, such as building location, construction materials, and building age. This goal can be met by evaluating data to find the factors that are most closely connected with the extent of building damage. This data can be used to direct resources and solutions to areas most vulnerable to seismic damage. It can also be utilized to create better building codes and building practices.

3. Create a model that can be used to forecast earthquake damage in other nations. This goal can be met by first training a machine learning model using data from the Nepal earthquake and then testing it on data from subsequent earthquakes. If the model can accurately anticipate the extent of damage to structures in other nations, it can be used to help decrease the impact of earthquakes on people and property in other regions of the world.

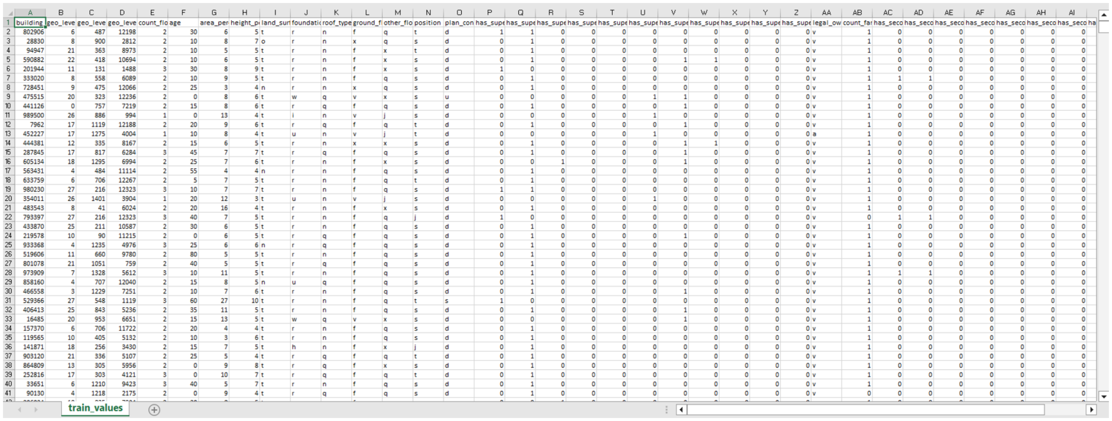
4. The fourth goal is to build a public database of earthquake damage data. This goal can be met by making the data from the Nepal earthquake available to the public. Researchers, engineers, and other stakeholders might utilize this database to increase our understanding of earthquakes and to help make buildings more resilient to future earthquakes.

Chapter 2

Knowing the Dataset

2.1 Dataset

- The data set consists of 39 features and 26061 sample spaces.
- Source URL: https://tinyurl.com/5cap334e



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	building_	geo_leve	geo_leve	geo_leve	count_flo	age	area_per	height_p	land_surl	foundatic	roof_type	ground_fl	other_flo	position	plan_con	has_supe
2	802906	6	487	12198	2	30	6	5	t	r	n	f	q	t	d	1
3	28830	8	900	2812	2	10	8	7	o	r	n	x	q	s	d	0
4	94947	21	363	8973	2	10	5	5	t	r	n	f	x	t	d	0
5	590882	22	418	10694	2	10	6	5	t	r	n	f	x	s	d	0
6	201944	11	131	1488	3	30	8	9	t	r	n	f	x	s	d	1
7	333020	8	558	6089	2	10	9	5	t	r	n	f	q	s	d	0
8	728451	9	475	12066	2	25	3	4	n	r	n	x	q	s	d	0
9	475515	20	323	12236	2	0	8	6	t	w	q	v	x	s	u	0
10	441126	0	757	7219	2	15	8	6	t	r	q	f	q	s	d	0
11	989500	26	886	994	1	0	13	4	t	i	n	v	j	s	d	0
12	7962	17	1119	12188	2	20	9	6	t	r	q	f	q	t	d	0
13	452227	17	1275	4004	1	10	8	4	t	u	n	v	j	t	d	0
14	444381	12	335	8167	2	15	6	5	t	r	n	x	x	s	d	0
15	287845	17	817	6284	3	45	7	7	t	r	q	f	q	s	d	0
16	605134	18	1295	6994	2	25	7	6	t	r	n	f	x	s	d	0
17	563431	4	484	11114	2	55	4	4	n	r	n	f	q	s	d	0
18	633759	6	706	12267	2	5	7	5	t	r	n	f	q	t	d	0
19	980230	27	216	12323	3	10	7	7	t	r	n	f	q	s	d	1
20	354011	26	1401	3904	1	20	12	3	t	u	n	v	j	s	d	0
21	483543	8	41	6024	2	20	16	4	t	r	n	f	x	s	d	0
22	793397	27	216	12323	3	40	7	5	t	r	n	f	q	j	d	1
23	433870	25	211	10587	2	30	6	5	t	r	n	f	q	s	d	0
24	219578	10	90	11215	2	0	6	5	t	r	q	f	q	s	d	0
25	933368	4	1235	4976	3	25	6	6	n	r	q	f	q	s	d	0
26	519606	11	660	9780	2	80	5	5	t	r	n	f	q	s	d	0
27	801078	21	1051	759	2	40	5	5	t	r	q	f	q	s	d	0
28	973909	7	1328	5612	3	10	11	5	t	r	n	f	q	s	d	0
29	858160	4	707	12040	2	15	8	5	n	u	q	f	q	s	d	0
30	466558	3	1229	7251	2	10	7	6	t	r	n	f	q	s	d	0
31	529366	27	548	1119	3	60	27	10	t	r	n	f	q	t	s	1
32	406413	25	843	5236	2	35	11	5	t	r	n	f	q	s	d	0
33	16485	20	953	6651	2	15	13	5	t	w	q	v	x	s	d	0
34	157370	6	706	11722	2	20	4	4	t	r	n	f	q	s	d	0
35	119565	10	405	5132	2	10	3	6	t	r	n	f	q	s	d	0
36	141871	18	256	3430	2	15	7	5	t	h	n	f	x	j	d	0
37	903120	21	336	5107	2	25	5	4	t	r	q	f	q	t	d	0
38	864809	13	305	5956	2	0	9	8	t	r	q	f	x	s	d	0
39	252816	17	303	4121	3	0	10	7	t	r	q	f	q	t	d	0
40	33651	6	1210	9423	3	40	5	7	t	r	n	f	q	s	d	0
41	90130	4	1218	2175	2	0	9	4	t	r	q	f	q	s	d	0

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
1	has_supe	has_supe	has_supe	has_supe	has_supe	has_supe	has_supe	has_supe	has_supe	has_supe	legal_ow	count_far	has_seco	has_seco	has_seco	has_seco	has_seco	has_seco	has_seco	ha
2	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
4	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
5	1	0	0	0	0	1	1	0	0	0	v	1	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
7	1	0	0	0	0	0	0	0	0	0	v	1	1	1	0	0	0	0	0	
8	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
9	0	0	0	0	1	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
10	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
11	0	0	0	0	1	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
12	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
13	0	0	0	0	1	0	0	0	0	0	a	1	0	0	0	0	0	0	0	
14	1	0	0	0	0	1	1	0	0	0	v	1	0	0	0	0	0	0	0	
15	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
16	0	1	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
17	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
18	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
19	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
20	0	0	0	0	1	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
21	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	v	0	1	1	0	0	0	0	0	
23	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
24	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
25	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
26	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
27	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
28	1	0	0	0	0	0	0	0	0	0	v	1	1	1	0	0	0	0	0	
29	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
30	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
31	0	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
32	1	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
33	0	0	0	0	0	1	0	0	0	0	v	1	0	0	0	0	0	0	0	
34	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
35	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
36	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
37	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
38	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
39	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
40	1	0	0	0	0	0	0	0	0	0	v	1	0	0	0	0	0	0	0	
41	1	0	0	0	0	0	0	0	0	0	v	0	0	0	0	0	0	0	0	

train_values

Figure 2.1: Snapshot of our dataset

2.2 Features of the Dataset

Table 2.1: Feature Descriptions

Sl.No	Feature name	Description
1	Geo level ID	geographic region in which building exists from largest to most specific sub-region
2	No of floors	number of floors in the building before the earthquake.
3	Age	age of the building in years.
4	Area Percentage	normalized area of the building footprint
5	Height Percentage	normalized height of the building footprint
6	Land Surface Condition	surface condition of the land where the building was built
7	FoundationType	type of foundation used while building.
8	RoofType	type of roof used while building
9	GroundfloorType	type of the ground floor
10	Other floorType	type of constructions used higher than the ground floors (except for the roof).
11	Postion	Postion of the building
12	Plan configuration	building plan configuration
13	Has superstructure adobe/mud	flag variable that indicates if the superstructure was made of Adobe/Mud
14	Has superstructure mud mortars tone	flag variable that indicates if the superstructure was made of Mud Mortar-Stone
15	Has superstructure stone flag	flag variable that indicates if the superstructure was made of Stone.
16	Has superstructure cement mortar stone	flag variable that indicates if the superstructure was made of Cement Mortar Stone.
17	Has superstructure mud mortar brick	flag variable that indicates if the superstructure was made of Mud Mortar Brick

Continued on next page

Table 2.1 – continued from previous page

Sl.No	Feature name	Description
18	Has superstructure cement mortar brick	flag variable that indicates if the superstructure was made of Cement Mortar Brick.
19	Has superstructure timber	flag variable that indicates if the superstructure was made of Timber.
20	Has superstructure bamboo	flag variable that indicates if the superstructure was made of Bamboo.
21	Has superstructure rc non-engineered	flag variable that indicates if the superstructure was made of non-engineered reinforced concrete.
22	Has superstructure rc engineered	flag variable that indicates if the superstructure was made of engineered reinforced concrete.
23	Has superstructure other	flag variable that indicates if the superstructure was made of any other material.
24	Legal ownership status	legal ownership status of the land where the building was built
25	Count families	number of families that live in the building.
26	Has secondary use	flag variable that indicates if the building was used for any secondary purpose.
27	Has secondary use agriculture	flag variable that indicates if the building was used for agricultural purposes.
28	Has secondary use hotel	flag variable that indicates if the building was used as a hotel.
29	Has secondary use rental	flag variable that indicates if the building was used for rental purposes.
30	Has secondary use institution	flag variable that indicates if the building was used as a location of any institution.

Continued on next page

Table 2.1 – continued from previous page

Sl.No	Feature name	Description
31	Has secondary use school	flag variable that indicates if the building was used as a school.
32	Has secondary use industry	flag variable that indicates if the building was used for industrial purposes.
33	Has secondary use health post	flag variable that indicates if the building was used as a health post.
34	Has secondary use gov office	flag variable that indicates if the building was used as a government office.
35	Has secondary use police	flag variable that indicates if the building was used as a police station
36	Has secondary use other	flag variable that indicates if the building was secondarily used for other purposes.

Table 2.2: **Details of the features in a dataset**

Sl.No	Feature name	Data Type	Distinct Values	Missing Values
1	geo level 1 id	Numerical	30	nil
2	geo level 2 id	Numerical	1427	Nil
3	geo level 3 id	Numerical	12567	Nil
4	count floors pre eq	9	Nil	
5	Age	Numerical	995	Nil
6	Area percentage	Numerical	100	Nil
7	Height percentage	Numerical	32	Nil
8	Land surface condition	Categorical	3	Nil
9	Foundation type	Categorical	5	Nil
10	Roof type	Categorical	3	Nil
11	Ground floor type	Categorical	5	Nil
12	Other floor type	Categorical	4	Nil
13	Position	Categorical	4	Nil
14	Plan configuration	Categorical	10	Nil
15	Has superstructure adobe mud	Binary	2	Nil
16	Has superstructure mud mortar stone	Binary	2	Nil
17	Has superstructure stone flag	Binary	2	Nil

Continued on next page

Table 2.2 – continued from previous page

Sl.No	Feature Name	Data Types	Distinct Values	Missing Values
18	Has superstructure cement mortar stone	Binary	2	Nil
19	Has superstructure mud mortar brick	Binary	2	Nil
20	Has superstructure cement mortar brick	Binary	2	Nil
21	Has superstructure timber	Binary	2	Nil
22	Has superstructure bamboo	Binary	2	Nil
23	has superstructure rc non engineered	Binary	2	Nil
24	has superstructure rc engineered	Binary	2	Nil
25	has superstructure other	Binary	2	Nil
26	Legal ownership status	Categorical	4	Nil
27	Count families	Numerical	9	Nil
28	has secondary use	Binary	2	Nil

Continued on next page

Table 2.2 – continued from previous page

Sl.No	Feature Name	Data Types	Distinct Values	Missing Values
29	Has secondary use agriculture	Binary	2	Nil
30	Has secondary use hotel	Binary	2	Nil
31	Has secondary use rental	Binary	2	Nil
32	Has secondary use institution	Binary	2	Nil
33	Has secondary use school	Binary	2	Nil
34	Has secondary use industry	Binary	2	Nil
35	Has secondary use health post	Binary	2	Nil
36	Has secondary use gov office	Binary	2	Nil
37	Has secondary use police	Binary	2	Nil
38	Has secondary use other	Binary	2	Nil

2.3 Observations

List your observations from the dataset here.

1. How are the features? All categorical? Mix?

   - We have mixed features in our dataset.

2. Are there any missing values? If yes, are they large or small?

   -There are no missing values in our data set.

3. What is the range of data items? How are they distributed?

   - Data items in our data set are categorical and numerical

4. Are there any outliers?

   -Yes, there are outliers in our data set such as the age of the building, etc,

5. Are any of the features skewed?

   - Yes, there are skewed features in our dataset

6. Does any of the features require normalization, or scaling?

   -Yes some features require normalization and standardization such as land surface condition, foundation type, etc.

7. Overall what are the characteristics of your dataset?

   -In our dataset, there is more categorical data in that we have nominal and ordinal and

2.4 Statistical Data Analysis

Making conclusions about a population based on a sample of data is the process of using statistics. A dataset is considered clean if there are no null values in it. Even so, noise can exist in a clean dataset and can be observed using a histogram. A histogram is a diagram that shows how data is distributed. A bar map illustrating the frequency of each value in a dataset is displayed. By scanning for outliners and distributing gaps, histograms can be used to simulate the noise in a dataset. Histograms can be used in the context of earthquake damage modeling to:

- Visualize the damage grade distribution for structures 1, 2, and 3.
- Visualize how the number of floors in buildings changed before and after the earthquake.
- Visualize the age distribution of structures.

Chapter 3

Implement Framework

The implementation framework for EDA on earthquake damage data can be divided into the following steps:

Data collection: The first step is to collect data on earthquake damage. This data has come from sources such as government reports and surveys. Data cleaning: Once the data has been collected, it is important to clean it to remove any errors or inconsistencies. This includes checking for missing values, outliers, and duplicate records. The data should be cleaned in a way that preserves the integrity of the data and that allows for accurate analysis.

Data exploration: The next step is to explore the data to understand its distribution and to identify patterns and trends. This can be done using a variety of statistical and visualization techniques. Some of the statistical techniques that can be used include:

Univariate analysis: This involves analyzing the distribution of each variable in the data set. This can be done using descriptive statistics, such as the mean, median, and standard deviation. Bivariate analysis: This involves analyzing the relationship between two variables. This can be done using scatter plots, correlation coefficients, and regression models.

Some of the visualization techniques that can be used include:

- Histograms: These can be used to visualize the distribution of a single variable.

 Chapter 3 Implement Framework

- Box plots: These can be used to visualize the distribution of a single variable and identify outliers.

- Heatmaps: These can be used to visualize the relationship between multiple variables.

Hypothesis generation: The final step is to generate hypotheses about how different factors are associated with earthquake damage. These hypotheses can then be tested with more sophisticated statistical methods

Chapter 4

Data Pre-processing

Data pre-processing is a process of preparing data for further analysis.

We have performed the following data preprocessing steps on the dataset:

1. Data cleaning
2. Data normalization
3. Data encoding
4. Data visualization

1. Cleaning of data

   We checked the dataset for missing values and discovered that there were none. We also looked for outliers in the columns' age of the 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.

2. Normalization of data

   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.

3. Visualization of data

   To visualize the distribution of data in the land surface condition, foundation type, roof type, ground floor type, other floor type, position distribution,

 Chapter 4 Data Pre-processing

plan configuration, has superstructure made of mud mortar, brick or cement, and rc engineered or non-engineered concrete columns, we used violin and kernel density estimation plots. This aided us in identifying any potential issues with the data.

4. Encoding data

   We used one-shot encoding to encode the categorical columns. This creates a separate binary column for each classification value. This enables us to incorporate these columns into a machine-learning model.

 15

Chapter 5

Exploratory Data Analysis

5.1 Hypothesis on the Problem Statement

1. In what ways did the geographical region contribute to the damage incurred by the building?
2. Among the building's floors, which ones sustained the highest level of damage?
3. Does the age of a building play a role in determining the amount of damage it is likely to experience?
4. What parts of the building, in terms of its area footprint and height footprint, suffered the greatest impact from the earthquake?
5. Which ground types were most susceptible to causing significant damage to the buildings?
6. To what extent did the building's foundation either alleviate or contribute to the damage incurred?
7. Was the roof of the building a primary factor in the extent of the damage?
8. Is there a correlation between the type of ground floor and the flooring materials used in other levels of the building?
9. Is the positioning of the building a contributing factor to the damage it inflicts upon itself?

10. Is the choice of construction materials a determining factor in the damage sustained by the building?

11. What is the correlation between the number of family members residing in the building and the family size most significantly impacted?

12. Which features within the dataset are crucial in determining the extent of the damage?

5.2 Analysis

Our hypothesis is as mentioned below and we have analyzed them using various graphs such as kernel trick graphs etc.

1. Among the building floors, which one sustained the highest damage?

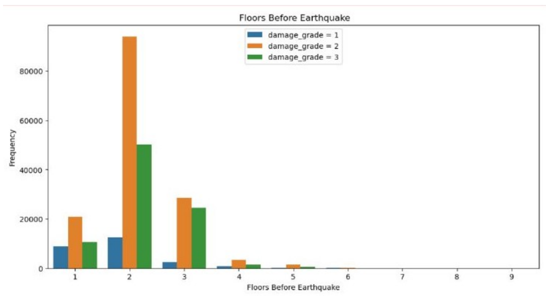


FIGURE 5.1: Floors Before Earthquake

During an earthquake, the second story of a building is most likely to experience damage, followed by the third floor. This conclusion is corroborated by the plots, which demonstrate a clear positive link between a building's floor and the amount of damage sustained. This means that the second, followed by the third of a building are more prone than the other floors to experience damage. When calculating the potential of damage during an earthquake, the floor of a building is an important component to consider.

2. Does the age of a building play a role in determining the amount of damage it is likely to experience?

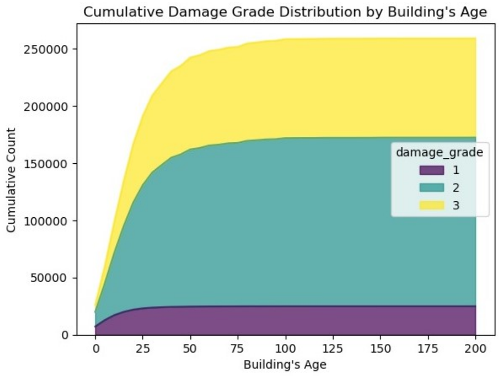


FIGURE 5.2: Damage Grade by Building Age

This conclusion is reinforced by the kernel trick graph, which shows a clear positive link between the age of a building and the amount of damage it is expected to sustain. This indicates that as a building's age increases, so does the likelihood of it being damaged.

This study supports the theory that the age of a building influences the amount of damage it is likely to sustain. This is because older buildings are more likely to have structural flaws and other vulnerabilities that render them vulnerable to harm.

3. What parts of the building, in terms of its area footprint and height footprint, suffered the greatest impact from the earthquake?

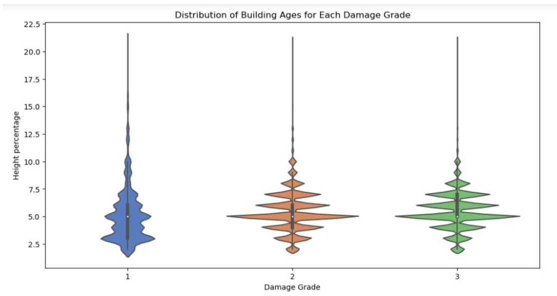


FIGURE 5.3: Building Ages for each Damage Grade

Buildings with a higher height percentage are more likely to sustain damage from earthquakes, as they are more susceptible to the forces of shaking. This conclusion is backed by the plots, which demonstrate a clear positive link between a building's height percentage and the amount of damage it takes.

This suggests that structures with a higher percentage of height are more likely to experience earthquake damage than ones with a lower percentage of height.

This study supports the concept that the areas of a building with the highest height percentage are the ones most vulnerable to earthquake effects. This is because the taller a structure is, the more it will shake and bend during an earthquake. This swinging and flexing can damage the structure as well as the contents of the building.

4. Is the choice of construction materials a determining factor in the damage sustained by the building?

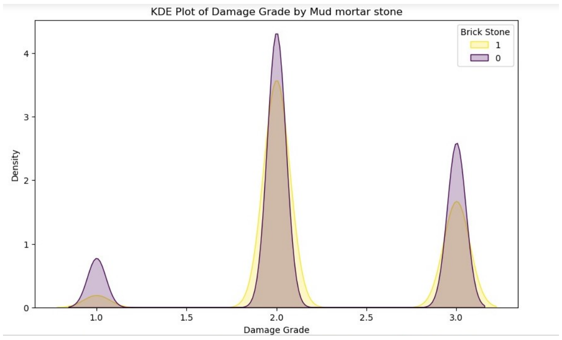


FIGURE 5.4: Damage Grade by Mud Motor Stone

Mud mortar stone constructions have higher sustainability and are less prone to withstand damage than other materials.

This conclusion is verified by the kernel density estimation plot, which demonstrates a clear positive association between the use of mud mortar stone and construction sustainability. This means that mud mortar stone buildings are less prone to incur damage than other types of buildings. This data supports the concept that the materials used in construction play a role in the damage incurred by a building. This is because mud mortar stone is a natural and sustainable material that is extremely resistant to damage.

5. Which features within the dataset are crucial in determining the extent of the damage?

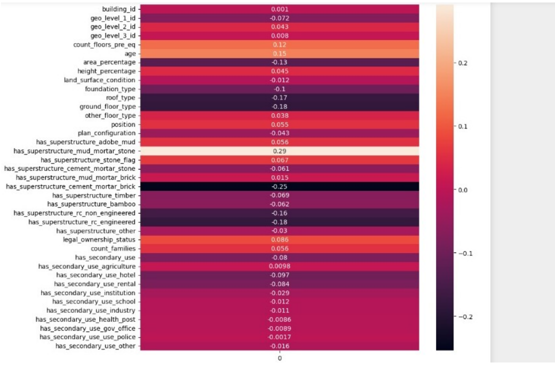


FIGURE 5.5: HeatMap

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.

This conclusion is corroborated by the heatmap, which demonstrates a clear positive link between these traits and the level of damage sustained by a building. This indicates that buildings with these elements are more vulnerable to damage than buildings without them. Older buildings, that have more floors, are composed of less durable materials, are higher, or are located in natural disaster-prone areas, for example, are all more likely to experience damage

6. Was the proof of the building a primary factor in the extent of the damage

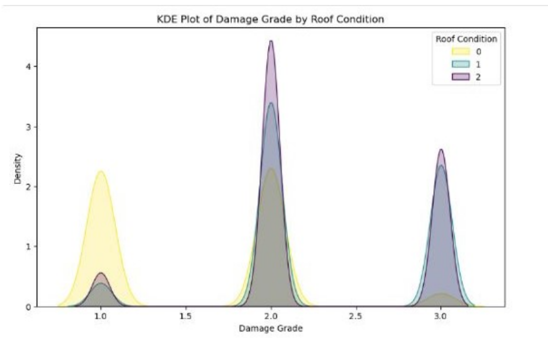


FIGURE 5.6: Damage Grade by Roof Condition

The level of damage sustained by a building during an earthquake is primarily determined by its roof.

This conclusion is reinforced by the kernel density estimate graph, which shows that a certain roof type labeled with ID 0 has incurred the most damage. This means that buildings with particular types of roofs are more prone to experience damage than structures with other types of roofs. The level of damage sustained by a building during an earthquake is primarily determined by its roof. This is because the roof is the most vulnerable section of a building to earthquake pressures. The roof is also the most vulnerable to harm from falling debris.

7. To what extent did the building's foundation either alleviate or contribute to the damage incurred?

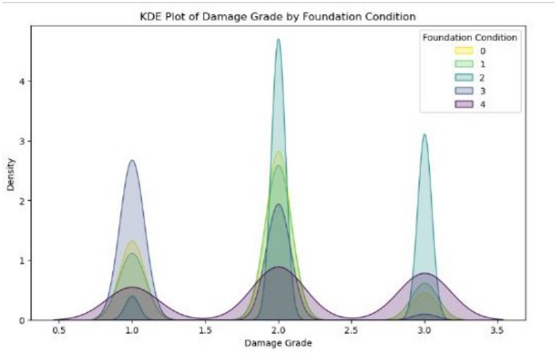


FIGURE 5.7: Damage Grade by Foundation Condition

The foundation of a building can have a considerable impact on the amount of damage sustained during an earthquake.

This conclusion is backed by the kernel density estimation figure, which shows that a specific foundation type of a building designated with ID 2 sustained the most damage. This means that buildings with particular types of foundations are more prone to experience damage than those with other types of foundations. Because the foundation of a building can have a considerable impact on the level of damage sustained after an earthquake.

This is because the foundation is what connects the building to the earth. If the foundation is not solid enough, it may collapse as a result of the power of an earthquake.

Chapter 6

Results and Outcomes

Results and Outcomes of Exploratory Data Analysis on Earthquake Damage Dataset

1. The age of a building is an important element in evaluating its sensitivity to seismic damage. Older buildings are more susceptible to being harmed since they are frequently composed of inferior materials and have weaker foundations.

2. The number of floors in a building also influences its sensitivity to seismic damage. Buildings with several stories are more likely to collapse since they are taller and must hold greater weight.

3. Height percentage: A building's height percentage is the ratio of its height to its width. Buildings having a higher percentage of height are more likely to collapse since they are more exposed.

4. Superstructure: The type of superstructure used in a building also affects its susceptibility to earthquake damage. Buildings with superstructures made of mud mortar stone are more sustainable than buildings with other types of superstructures.

5. Geographical location: The geographical location of a building also plays a role in determining its susceptibility to earthquake damage. Buildings located in areas with high seismic activity are more likely to be damaged than buildings located in areas with low seismic activity

Conclusions

Our problem statement was published as an online challenge on Kaggle, and the dataset we got was derived from surveys conducted by Kathmandu Living Labs and the Central Bureau of Statistics. We accepted this challenge because we hoped to use it to develop Richter's Predictor, a prediction method for precisely assessing earthquake damage. Using the information, we hope to create a robust and trustworthy model that considers various factors such as building configuration, construction, and location, among others, to prevent future disasters.

To solve this problem statement we performed exploratory data analysis and it assisted us in identifying the elements most likely to influence the extent of damage. This was accomplished by running several statistical studies on the dataset, including: Univariate analysis which entails examining the distribution of each variable in the dataset.

Bivariate analysis entails examining the relationship between two variables. EDA aided us in understanding the links between the various elements. This was accomplished by visualizing the data and searching for trends. For example, we discovered a large negative link between a building's age and the extent of damage it experienced. This indicates that older structures were more likely to sustain damage from the earthquake.

It also assisted us in removing outliers from the dataset. Outliers are data points that deviate greatly from the average. They have the potential to influence the outcomes of statistical research. We were able to generate a more accurate picture of the link between the different components by deleting outliers. It also assisted us in determining the most essential variables. This was accomplished by

 Chapter 6 Results and Outcomes

employing a range of statistical techniques, including: Feature Selection entailed determining which characteristics were most predictive of the extent of damage.

Overall, the use of EDA was critical in resolving the issue statement on modeling earthquake damage. It aided us in identifying the elements most likely to influence the extent of damage, understanding the correlations between the various factors, and removing outliers from the dataset. This enabled us to construct a more precise understanding of the problem and a more effective model.

Based on the study's findings, the following recommendations are made:

- Older structures should be retrofitted to make them more earthquake resistant.
- New structures should be built with earthquake-resistant materials and procedures.
- The conclusions of this study should be incorporated into building codes.
- Educational activities should be developed to enhance knowledge of earthquake safety.