



UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE INGENIERÍA  
Año 2021 - 1er Cuatrimestre  
Trabajo Práctico 1

## Organización de Datos

 <https://github.com/rлуизaga/Tp1OrgaDeDatos>

Integrantes	Padrón	E-mail
Nicolás Menese	103581	nmenese@fi.uba.ar
Iván Leonardo Conde	102497	iconde@fi.uba.ar
Joseph Enrique Alvaro	103991	jalvaro@fi.uba.ar
Ricardo Luizaga	87528	rluizaga@fi.uba.ar

Observaciones:

<b>Introducción</b>	<b>3</b>
<b>Objetivos</b>	<b>3</b>
<b>Preparación de datos</b>	<b>3</b>
<b>Manejo de campos faltantes</b>	<b>5</b>
<b>Desarrollo</b>	<b>6</b>
1.0 Relaciones de grado de daño con materiales de la edificación	6
1.1 Con superestructura de adobe.	8
1.2 Con superestructura de barro y piedra.	8
1.3 Con superestructura de madera de construcción.	9
2.0 Relación de grado de daño con la zona donde se produjo el terremoto	10
3.0 Relación de grado de daño con antigüedad y estado de la edificación	10
4.0 Análisis de los tipos de usos de las edificaciones	12
5.0 Relación de cantidad de edificios y su grado de daño según su tipo de cimiento.	14
6.0 Análisis de la relación entre tipo de techo y grado de daño	14
7.0 Relación de la cantidad de pisos que tenía cada edificación y su grado de daño.	16
<b>Conclusiones</b>	<b>17</b>

# Introducción

En el presente trabajo se propone realizar un análisis exploratorio de los datos recolectados a través de encuestas realizadas por [Kathmandu Living Labs](#) y el [Central Bureau of Statistics](#) posterior al terremoto ‘**Gorkha**’ que golpeó Nepal y a ciudades cercanas en 2015.

El objetivo de este análisis es poder realizar un análisis exploratorio de los datos provistos con el objetivo de determinar características y variables importantes, descubrir insights interesantes, y analizar la estructura de los mismos.

## Objetivos

Con este análisis exploratorio esperamos:

- Entender las características y analizar la importancia de los datos propuestos.
- Encontrar relaciones entre las variables que nos ayuden a hacer un análisis exploratorio más profundo.
- Encontrar patrones de comportamientos útiles que nos ayuden a predecir el grado de daño en futuros terremotos.
- Descubrir Insights interesantes y poder sacar una conclusión exitosa de este análisis exploratorio.

## Preparación de datos

Al realizar el análisis introductorio a los distintos datos que fueron proporcionados, vimos que es de vital importancia realizar una preparación de los mismos, a fin de poder manejar con mayor facilidad el trabajo posterior.

Vistazo simple de los datasets

	building_id	geo_level_1_id	geo_level_2_id	geo_level_3_id	count_floors_pre_eq	age	area_percentage	height_percentage
0	802906	6	487	12198	2	30	6	5
1	28830	8	900	2812	2	10	8	7
2	94947	21	363	8973	2	10	5	5
3	590882	22	418	10694	2	10	6	5
4	201944	11	131	1488	3	30	8	9

	building_id	damage_grade
0	802906	3
1	28830	2
2	94947	3
3	590882	2
4	201944	3

Para concretar este análisis, el primer paso a realizar es hacer un **Merge** sobre 'Building\_id' entre ambos data frames con el fin de graficar las relaciones entre las características que contienen.

	building_id	damage_grade	geo_level_1_id	geo_level_2_id	geo_level_3_id	count_floors_pre_eq	age	z
0	802906	3	6	487	12198	2	30	
1	28830	2	8	900	2812	2	10	
2	94947	3	21	363	8973	2	10	
3	590882	2	22	418	10694	2	10	
4	201944	3	11	131	1488	3	30	
...	...	...	...	...	...	...	...	...
260596	688636	2	25	1335	1621	1	55	
260597	669485	3	17	715	2060	2	0	
260598	602512	3	17	51	8163	3	55	
260599	151409	2	26	39	1851	2	10	
260600	747594	3	21	9	9101	3	10	

## Manejo de campos faltantes

Data columns (total 40 columns):

#	Column	Non-Null Count		Dtype
0	building_id	260601	non-null	int64
1	geo_level_1_id	260601	non-null	int64
2	geo_level_2_id	260601	non-null	int64
3	geo_level_3_id	260601	non-null	int64
4	count_floors_pre_eq	260601	non-null	int64
5	age	260601	non-null	int64
6	area_percentage	260601	non-null	int64
7	height_percentage	260601	non-null	int64
8	land_surface_condition	260601	non-null	object
9	foundation_type	260601	non-null	object
10	roof_type	260601	non-null	object
11	ground_floor_type	260601	non-null	object
12	other_floor_type	260601	non-null	object
13	position	260601	non-null	object
14	plan_configuration	260601	non-null	object
15	has_superstructure_adobe_mud	260601	non-null	int64
16	has_superstructure_mud_mortar_stone	260601	non-null	int64
17	has_superstructure_stone_flag	260601	non-null	int64
18	has_superstructure_cement_mortar_stone	260601	non-null	int64
19	has_superstructure_mud_mortar_brick	260601	non-null	int64
20	has_superstructure_cement_mortar_brick	260601	non-null	int64
21	has_superstructure_timber	260601	non-null	int64
22	has_superstructure_bamboo	260601	non-null	int64
23	has_superstructure_rc_non_engineered	260601	non-null	int64
24	has_superstructure_rc_engineered	260601	non-null	int64
25	has_superstructure_other	260601	non-null	int64
26	legal_ownership_status	260601	non-null	object
27	count_families	260601	non-null	int64
28	has_secondary_use	260601	non-null	int64
29	has_secondary_use_agriculture	260601	non-null	int64
30	has_secondary_use_hotel	260601	non-null	int64
31	has_secondary_use_rental	260601	non-null	int64
32	has_secondary_use_institution	260601	non-null	int64
33	has_secondary_use_school	260601	non-null	int64
34	has_secondary_use_industry	260601	non-null	int64
35	has_secondary_use_health_post	260601	non-null	int64
36	has_secondary_use_gov_office	260601	non-null	int64
37	has_secondary_use_use_police	260601	non-null	int64
38	has_secondary_use_other	260601	non-null	int64
39	damage_grade	260601	non-null	int64

dtypes: int64(32), object(8)

memory usage: 81.5+ MB

Como se ve ejecutando el comando 'info' en el set de datos, se puede comprobar que no se encuentran valores nulos, por lo tanto, no tuvimos que realizar ninguna limpieza de datos en este aspecto en el raw data set.

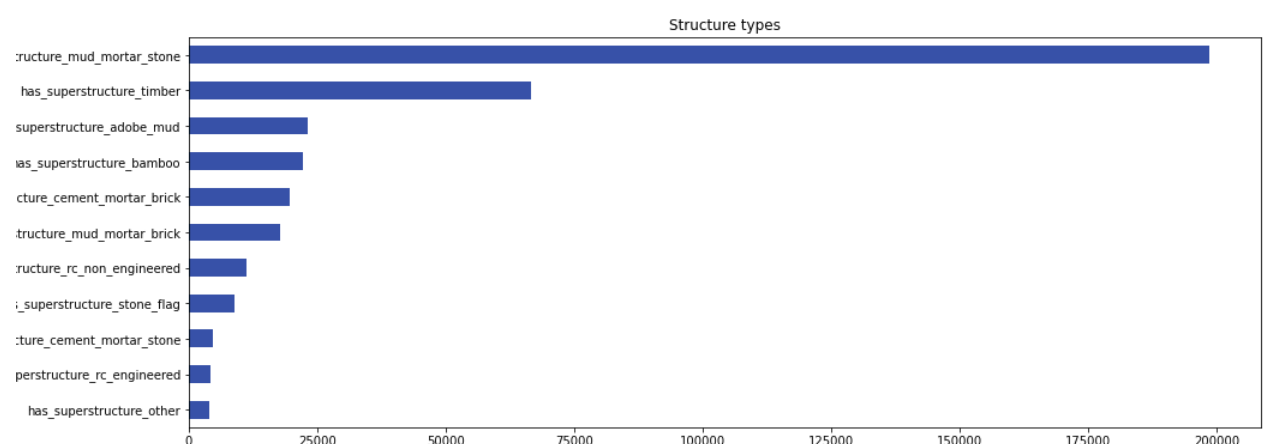
Como en todo set de datos -luego de realizar operaciones, es esperable encontrarnos con publicaciones que no se encuentren completas y que tengan datos ausentes, valores a los que nos referiremos como **valores nulos**. Frente a esta situación el criterio elegido en esta ocasión fue mantener todas las publicaciones que contengan valores nulos para no perder datos de manera innecesaria. Por ejemplo, si buscamos hallar una relación entre el "damage\_grade" y "count\_floors\_pre\_eq", no nos interesaría que el valor de "legal\_ownership\_status" sea nulo para dicha publicación.

## Desarrollo

### 1.0 Relaciones de grado de daño con materiales de la edificación

Con el fin de determinar los daños que pueden producir los futuros terremotos es necesario conocer los distintos tipos de estructuras y su comportamiento ante los diferentes movimientos de suelo para ello contamos con datos sobre el tipo de materiales utilizados en las edificaciones que recibieron daño en el terremoto. En base al análisis obtenido se podría mejorar los tipos de construcciones y hacer más estrictos los controles de calidad en obras de reciente construcción

Para empezar con nuestro análisis podemos empezar por ver cómo son las construcciones que recibieron daño para luego ver el tipo de daño que recibieron.



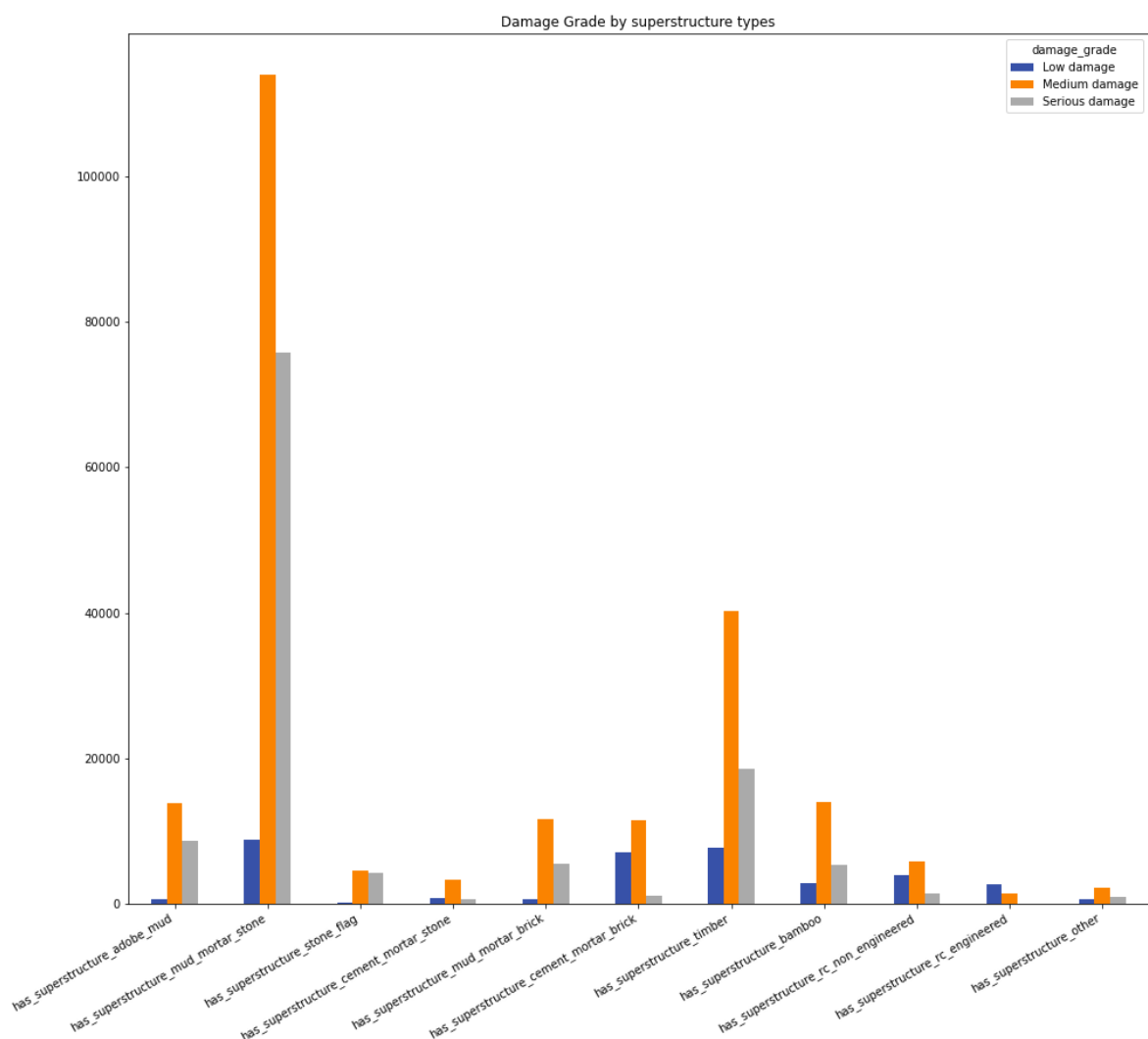
Observamos que la mayor parte de las estructuras que recibieron daño fue construida con barro-piedra seguida de las que fueron construidas con Timber

(madera específica para la construcción) y luego las que fueron construidas con adobe/barro.

Las construcciones que menos daño recibieron son las construidas con concreto reforzado diseñado y las construidas con cemento-piedra.

Se propone verificar si existe algún tipo de relación entre damage\_grade y el material con el que está construido la edificación.

A continuación se muestra un gráfico generalizado de la distribución de daños según el tipo de superestructura de la edificación.



A simple vista se puede observar que los edificios más comunes son los edificios cuya superestructura corresponde a 'barro y piedra' donde se puede observar que el daño de grado medio (2) es el más prominente seguido por el daño serio (3), siendo superior a más del 50% de los de grado medio; por último, se observa que muy

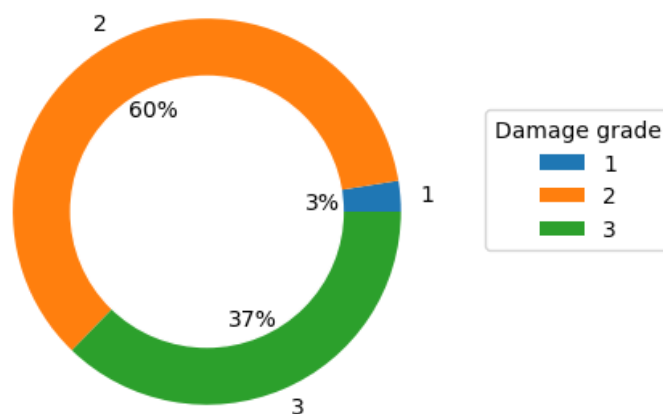
pocos edificios que contenían esta superestructura tuvieron un grado de daño bajo (1).

A continuación se mostrarán gráficos detallados con sus porcentajes de los 3 tipos de superestructura más utilizadas, los cuales se pueden apreciar en el gráfico del rótulo 1.0.

### 1.1 Con superestructura de adobe.

En la siguiente visualización se demuestra que las edificaciones construidas con superestructura de adobe tienden a tener un gran porcentaje de daño de grado 2 y 3 si son alcanzadas por un terremoto, siendo esto de esperarse ya que es un material muy frágil y no adecuado frente a terremotos.

Relacion de grado de daño con adobe

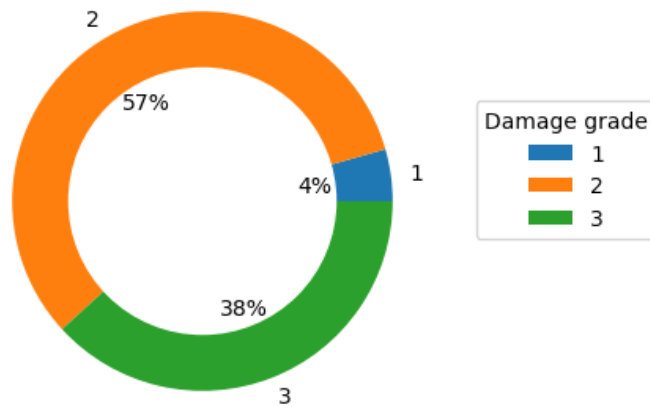


### 1.2 Con superestructura de barro y piedra.

En la siguiente visualización se ve con más detalle la relación en porcentajes de los grados de daño en los edificios que utilizan barro y piedra como superestructura. El resultado es prácticamente el mismo que el de los cimientos con adobe.



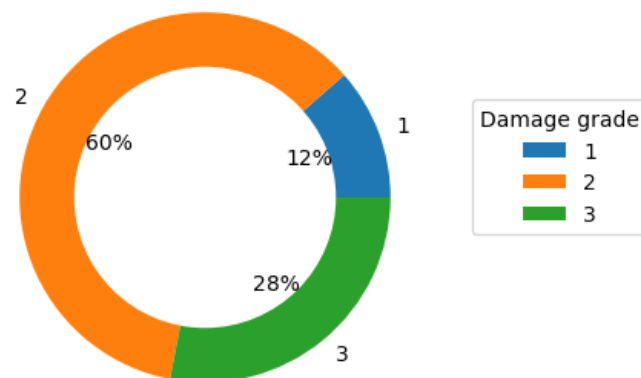
Relacion de grado de daño con barro - piedra



### 1.3 Con superestructura de madera de construcción.

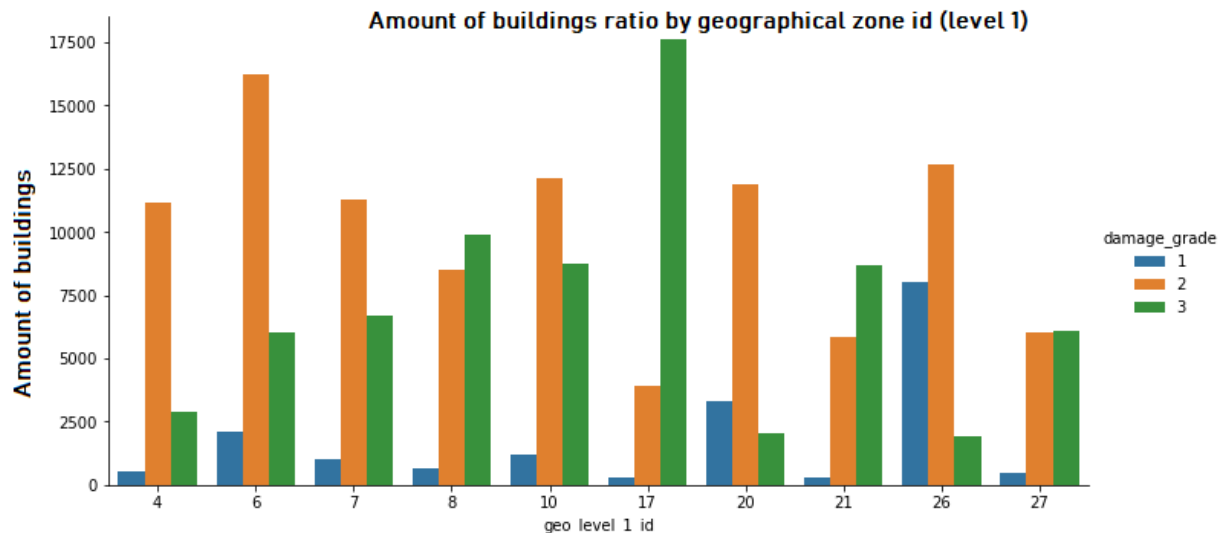
En la siguiente visualización se ve con más detalle la relación en porcentajes de los grados de daño en los edificios que utilizan madera de construcción como superestructura.

Relacion de grado de daño madera para construccion



## 2.0 Relación de grado de daño con la zona donde se produjo el terremoto

En este gráfico se detallan las 10 zonas con mayor cantidad de edificios para un análisis más preciso y profundo, ya que añadir todas las zonas implicaba tener datos menores que podrían afectar todo el gráfico.



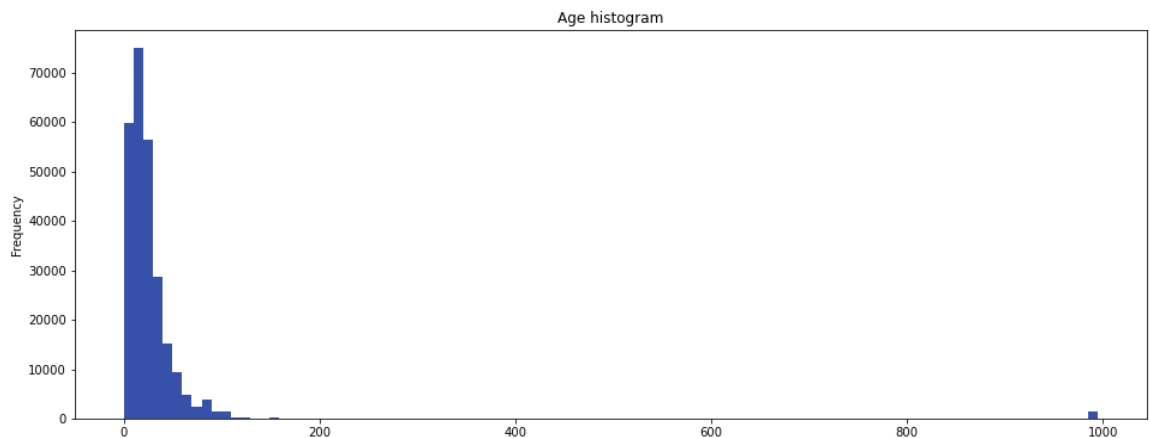
Se puede observar que la zona 17 es la que tiene mayor cantidad de edificios afectados por un daño de grado alto (3) y contiene una baja cantidad que sufrió un daño de grado medio(2) y bajo(1), en algunas zonas, el daño de grado medio y alto parece estar muy equilibrado (zonas 27, 21, 8), en casi todos predomina el daño de grado medio y por último en la zona 26 se ve que el daño de grado bajo es mayor que el daño de grado alto por bastante.

## 3.0 Relación de grado de daño con antigüedad y estado de la edificación

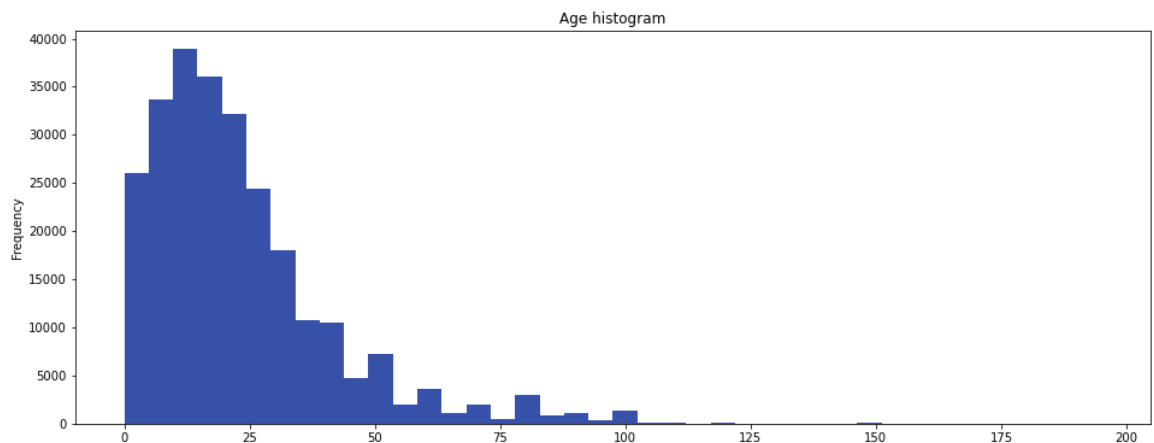
Veamos algunas estadísticas de la antigüedad de la edificación

count	260601
mean	26.535029
std	73.565937
min	0
25%	10
50%	15
75%	30
max	995

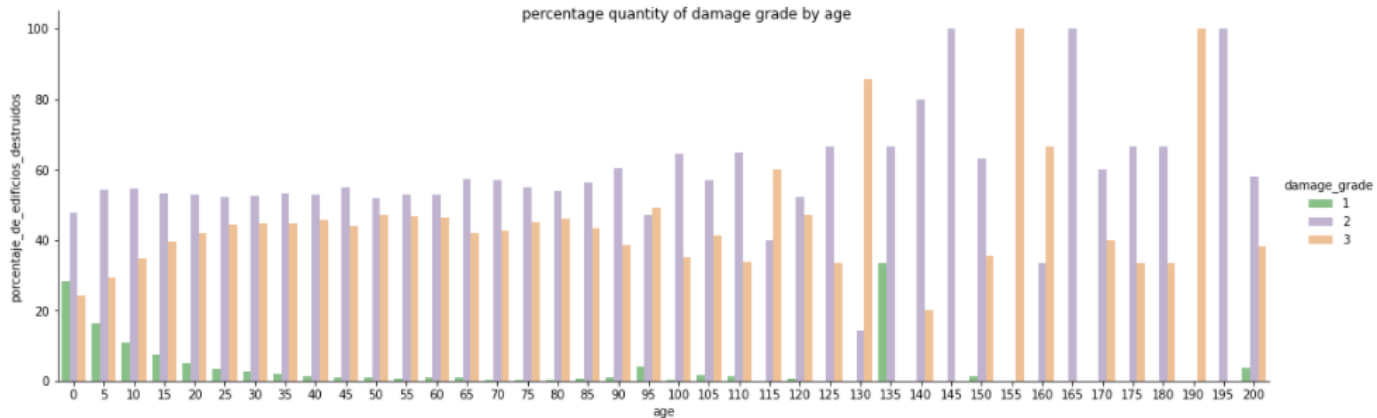
Podemos observar que el valor medio de la antigüedad es 26 años pero tenemos una desviación estándar muy alta, esto puede deberse a que tenemos valores muy dispersos ya que el máximo valor es 995. Para comprender un poco más los datos realizaremos un histograma con una ventana de 10 años



Podemos observar que los años se concentran en un rango de entre 0 y 200 años, además tenemos datos cercanos a 1000 años estos pueden ser edificaciones muy antiguas que sufrieron daños en el terremoto. Para nuestro análisis tomaremos el rango de 0 a 200 años ya que la cantidad de edificaciones fuera de este rango son mucho menores .



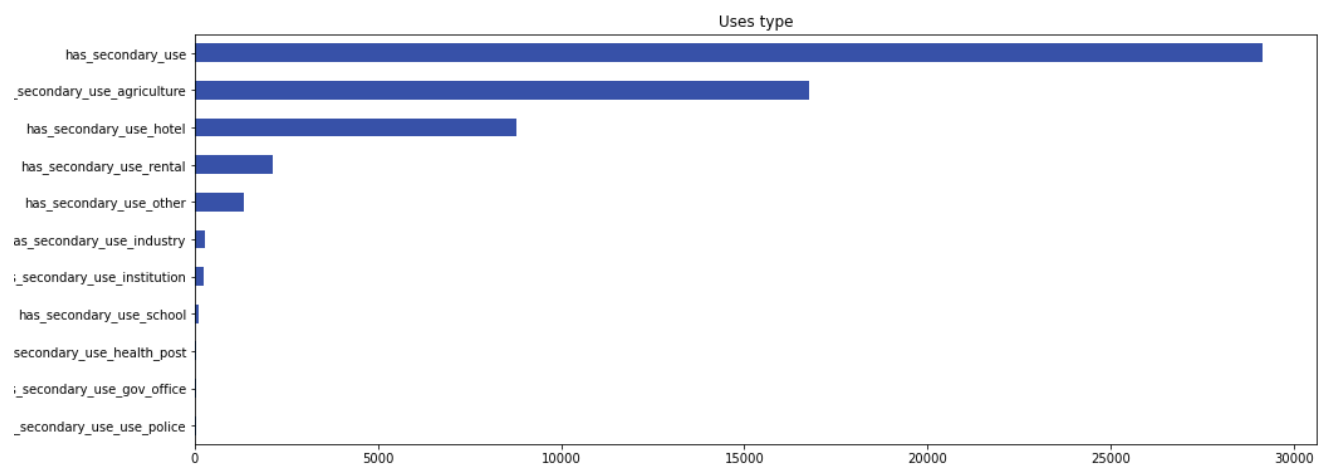
El siguiente gráfico explora la relación que existe entre la edad y el porcentaje de los edificios junto con su grado de daño, para este análisis, decidimos centrarnos en un zona en particular en lugar de incluir todas. (elegimos la zona 6)



Se puede observar que mientras más edad tenga un edificio menor será el porcentaje de grado 'menor' (1) de las edificaciones y en cambio crece el porcentaje de edificios que sufrieron un daño de grado 'medio' (2), en el gráfico también se observa que a partir de los 120~ años (aproximadamente) debido a que la cantidad de estructuras con esta edad empieza a ser muy rara, los datos quizás no tiene mucho sentido y son muy volátiles, como se podría observar en los 135 años donde se nota un grado de daño 'bajo' mucho mayor que en edades tempranas.

## 4.0 Análisis de los tipos de usos de las edificaciones

Para empezar con nuestro análisis podemos empezar por ver qué tipo de usos tenían las edificaciones que recibieron daño para luego ver el tipo de daño que recibieron.



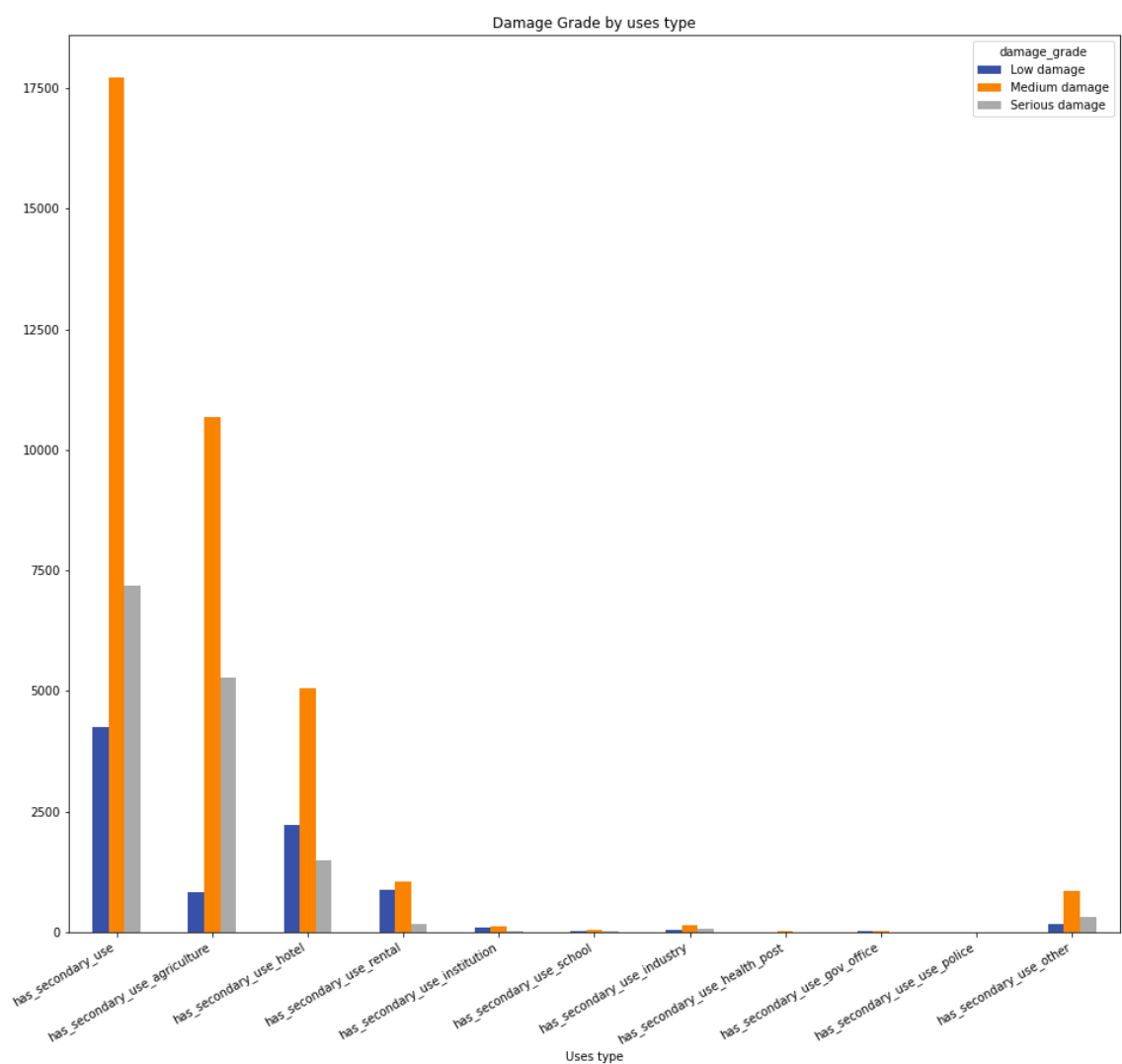
Podemos ver que los edificios que más registros de daño poseen son los que tienen un uso secundario seguido por las que tenían un propósito de agricultura y hoteles.

Los usos que menos registros poseen son las estaciones de policía, oficinas del gobierno y puestos de salud.

A continuación se muestra un gráfico generalizado de la distribución de daños según el uso de la edificación.

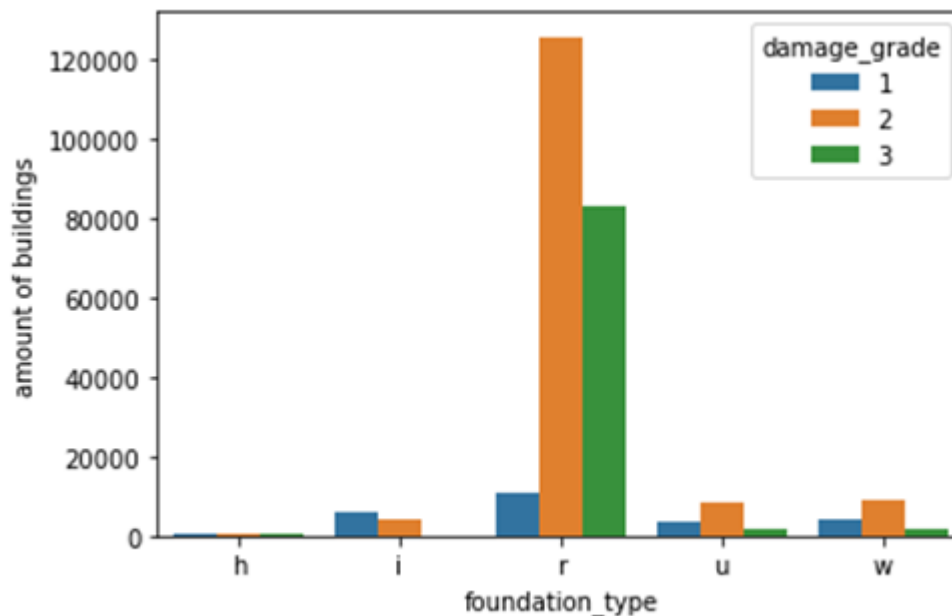
Podemos observar que las edificaciones con uso secundario recibieron mayormente daño medio los edificios con uso de agricultura recibieron mayormente daño de grado medio y alto, los hoteles recibieron daño bajo medio.

Viendo que las edificaciones de agricultura debemos tener poner más énfasis en proteger dichas estructuras



## 5.0 Relación de cantidad de edificios y su grado de daño según su tipo de cimiento.

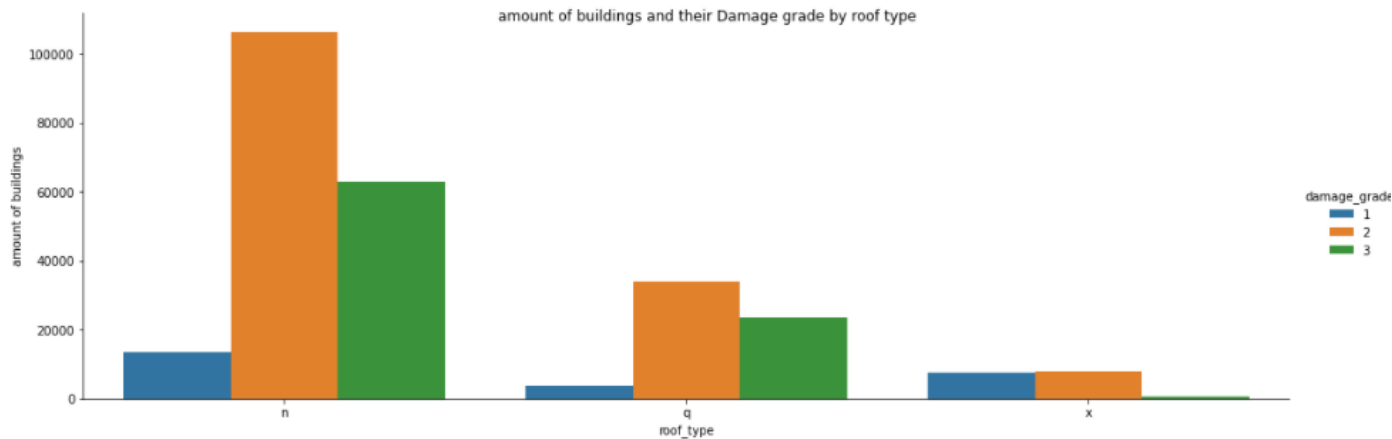
En el siguiente gráfico se puede apreciar la cantidad de edificios y el grado de daño que estos recibieron durante el terremoto según el tipo de cimientos que estos poseían. Se puede apreciar que el tipo de fundación o cimiento más propenso a recibir daños durante un terremoto de estas magnitudes es el denominado “r” siendo el resto mucho más resistentes a este tipo de desastre.



Por otro lado los de tipo “h” fueron los que menos sufrieron la catástrofe, siendo casi inexistentes frente al resto. “i”, “u” y “w” se mantienen parejos entre ellos, siendo dentro de todo baja cantidad comparado a los de tipo “r”.

## 6.0 Análisis de la relación entre tipo de techo y grado de daño

La siguiente visualización muestra la cantidad de edificaciones que tienen un tipo de techo ‘n’, ‘q’ ó ‘x’ separado según el grado de daño que recibieron estos.

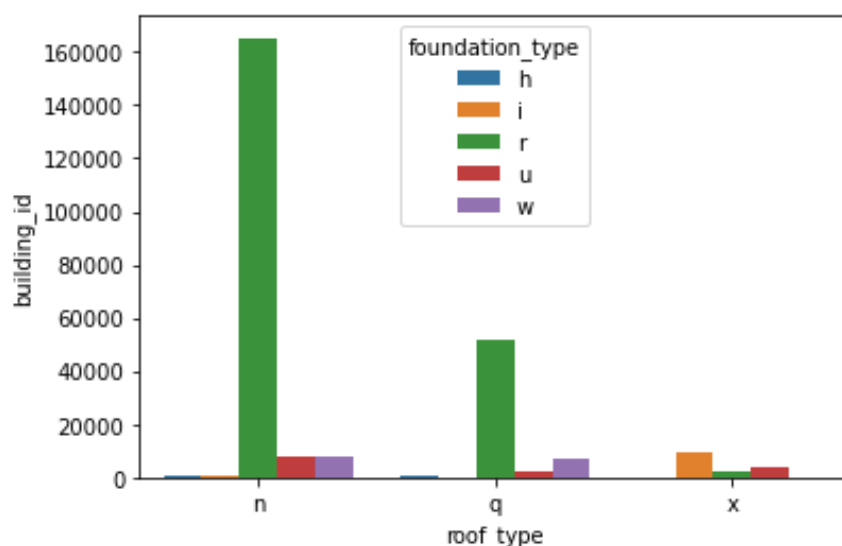


Se observa que el tipo de techo más común es el del tipo 'n', seguido por el 'q' y por último el 'x'; siempre se observa que el grado de daño más común es el grado de daño medio seguido por el daño de grado alto, excepto por el tipo de techo 'x' donde el daño de grado 'bajo' y 'medio' parece tener la misma cantidad de casos.

Esto podría sugerir que el uso de techo de tipo 'x' es un factor muy importante a tomar en cuenta a la hora de construir nuevos edificios con el objetivo de que haya la menor cantidad de edificios con un grado de daño alto.

Esta relación tipo de 'techo'-'daño de grado' parece ser muy fuerte por lo que vamos a hacer otro análisis más de este atributo (tipo de techo).

## 6.1 Relación de la cantidad de edificios con un cierto tipo de techo y su respectivo tipo de fundación

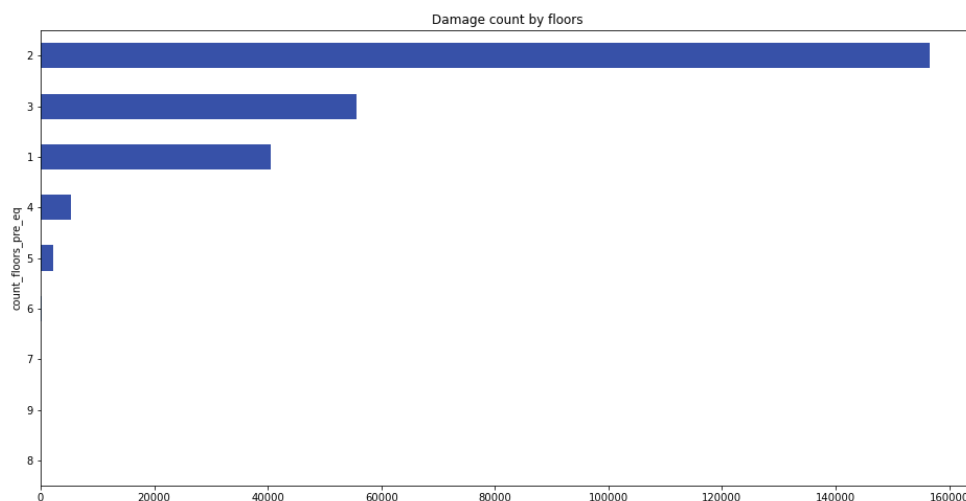


Se observa que los tipos de techo 'n' y 'q' tiene un alto índice de tipo de fundación de tipo 'r' seguido por 'u' y 'w', mientras que el tipo de fundación predominante en el tipo de techo 'x' es el tipo de fundación 'i'.

De nuevo, esto añadiría a la sugerencia que de usar un tipo de techo 'x', el tipo de fundación quizás debería ser 'i', también que la fundación de tipo 'r' no es la más recomendada.

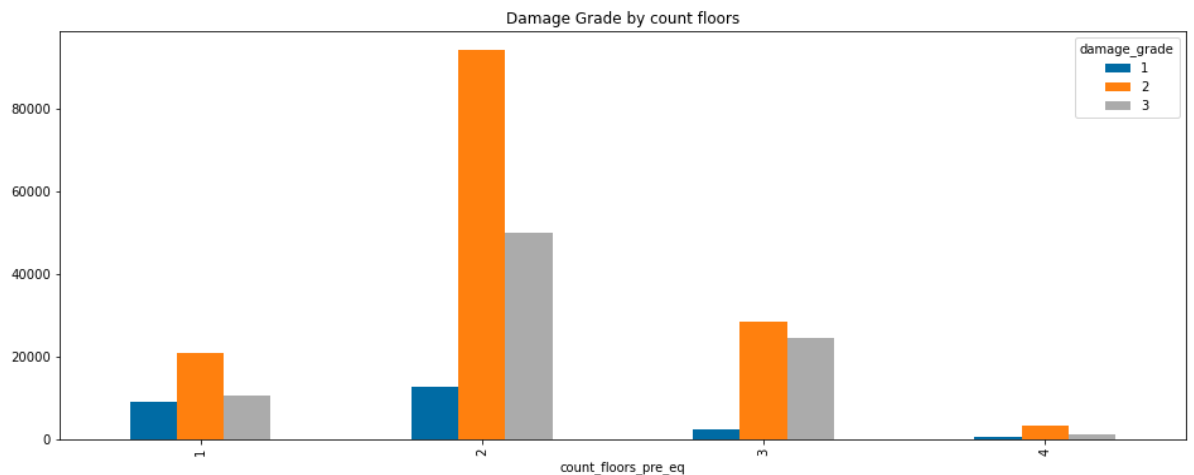
## 7.0 Relación de la cantidad de pisos que tenía cada edificación y su grado de daño.

En la siguiente visualización se puede observar que la mayoría de las edificaciones contenían 2 pisos



Se denota una gran cantidad de edificaciones con daño de grado 2 los cuales representan el 60% del total de registros algo llamativo, en este caso, mientras que en las casas con 1 piso el daño de grado 3 es equivalente al grado 1 y el daño de grado 2 también es menor, por el contrario en el 3 sucede lo opuesto, el daño de grado 3 está a la par con el daño de grado 2 en estas edificaciones lo que supone que mientras más pisos hay, mayor el daño de grado 3 y menor el daño de grado 1, en el caso de las edificaciones con 4 pisos, la información que se encuentra en el dataset es muy baja como para sacar una conclusión concreta, pero se podría suponer que sigue el mismo patrón que las anteriores.





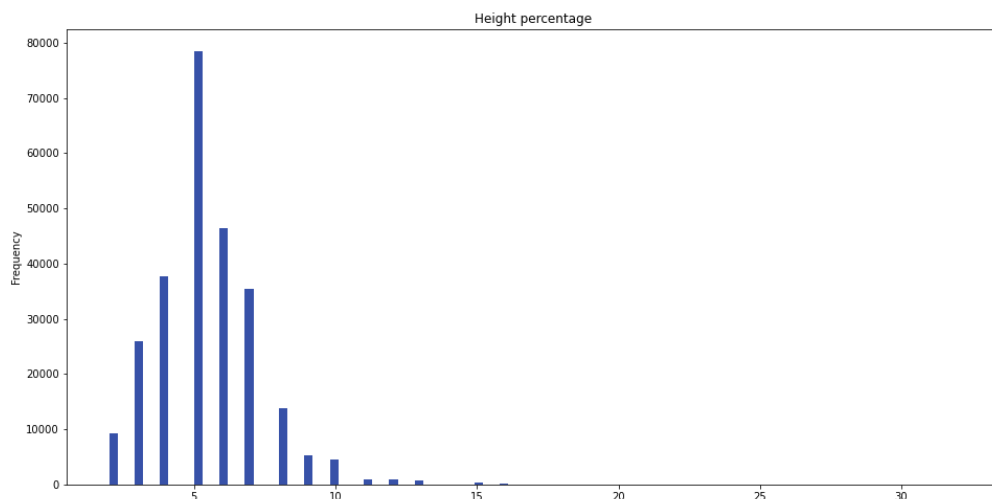
## 8.0 Análisis de la altura normalizada de los edificios

Las ondas sísmicas afectan a las estructuras en función de diferentes características

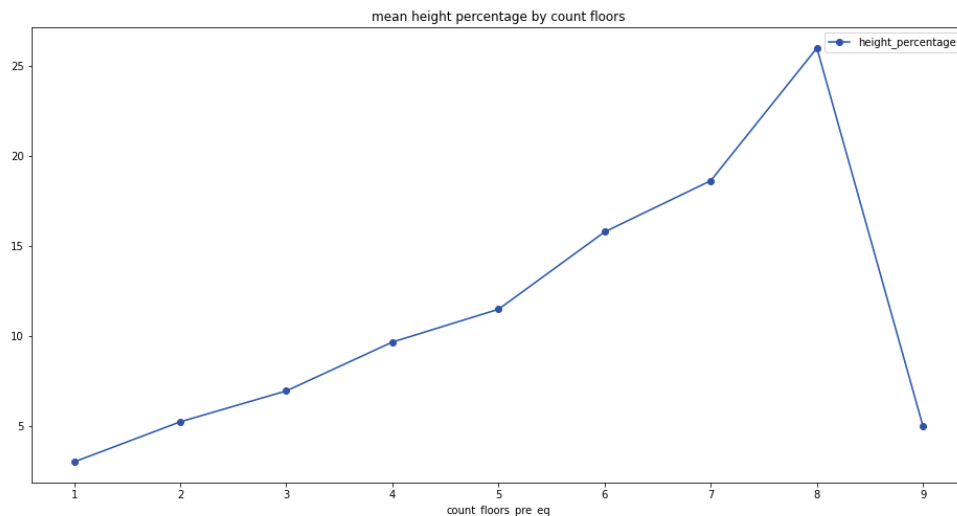
- Ondas de baja frecuencia: hacen que los edificios altos entren en resonancia, afectando en menor grado a los bajos.
- Ondas de Alta frecuencia: afectan sobre todo a las casas bajas y en menor grado a las grandes estructuras

Por lo cual un análisis del daño en función de la altura nos daría una idea del tipo de onda sísmica que tuvo lugar en el terremoto de Gorkha.

Para lo cual empezamos analizando cómo se distribuye la altura normalizada de los edificios dañados



podemos observar claramente que los edificios que recibieron daño están concentradas entre 2 y 7. Para tener una idea de estos valores se hace un análisis de la altura promedio por la cantidad de pisos que posee el edificio



podemos ver que a medida que aumenta la cantidad de pisos aumenta la altura lo cual es razonable por lo cual el valor para edificios de nueve pisos resulta extraño. Revisando el set de datos encontramos que solo tenemos un solo edificio con nueve pisos por lo cual puede ser un valor mal computado.

En análisis anteriores llegamos a la conclusión de que la mayor cantidad de edificios dañados eran los de uno dos y tres pisos lo cual se corresponde con edificios de baja altura.

Podemos llegar a la conclusión de que el tipo de sismo que afectó a la zona fue una onda sísmica de alta frecuencia.

## Conclusiones

Durante el transcurso del trabajo práctico, tuvimos la oportunidad de implementar diversas técnicas de parseo de datos para extraer información relevante, también pudimos observar que para buscar características e irregularidades fue necesario realizar preguntas interesantes e ir resolviendolas mediante análisis y creando visualizaciones.

En lo que respecta a extraer información relevante del set de datos, prestamos especial atención a encontrar puntos en común entre el material con el cual fue construido la edificación y su respectivo grado de daño que podría llegar a darnos una idea del daño que provocaría un terremoto de esta magnitud para una edificación con características similares a la trabajas en el data set.

Al dividir y visualizar los datos según criterios de tipo de material, encontramos comportamientos esperados con edificaciones que tenían materiales de construcción muy precarios ya que predominaban un daño de grado de tipo 2 y 3.

Lo mismo sucedió al realizar el análisis con respecto a la edad del edificio pudimos observar que a mayor antigüedad del edificio el daño causado sobre ellos era superior a edificios con estructuras más nuevas, aunque tuvimos que limitar los datos ya que nos encontramos con datos de edificios de más antigüedad que no tenían un daño tan fuerte como el esperado según el análisis.

Un análisis de la altura de los edificios dañados nos permitió conocer el tipo de sismo que afectó a la zona. Se pudo concluir que fue una onda sísmica de alta frecuencia.

El dataset nos proporciona información de las edificaciones que se encontraban en el epicentro, aunque hubiese sido interesante conocer el radio del terremoto para poder sacar conclusiones más exactas con respecto al daño ocurrido de una edificación que estaba en el centro a otro que estaba casi en el límite del epicentro del terremoto.