



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

[75.06/95.58] ORGANIZACIÓN DE DATOS

Curso 1: Argerich

1ER CUATRIMESTRE DE 2021

TRABAJO PRÁCTICO N° 1: Análisis exploratorio de datos

| GRUPO: Pandas en júpiter | |
|------------------------------|-----------|
| APELLIDO, Nombres | N° PADRÓN |
| MASTRICCHIO, Facundo Rodrigo | 100.874 |
| SEBELLIN, Camila Belén | 100.204 |
| PONCE, Antonella Belén | 100.635 |

Link al repositorio: <https://github.com/FacuMastri/tps-orga-de-datos>

Índice

| | |
|------------------------------------------------------------------------------------------------|-----------|
| Introducción | 2 |
| Información general sobre los datos | 3 |
| Manipulación de los datos | 4 |
| Truncamiento de los datos | 4 |
| Análisis por región geográfica | 5 |
| ¿Qué nivel de daño tiene cada zona? | 6 |
| Análisis por condición de la superficie terrestre | 7 |
| Análisis por tipo de cimientos | 8 |
| ¿Podemos emparejar algún tipo de cimiento con algún tipo de superestructura? | 8 |
| Análisis por tipo de planta baja | 11 |
| Análisis por tipo de estructura | 13 |
| ¿A mayor cantidad de superestructuras, menor es el daño recibido? | 16 |
| ¿Cuál fue el tipo de edificación que se vio más afectada? ¿Cuál fue la que menos fue afectada? | 17 |
| Comparación de morteros: cemento-ladrillo vs. barro-ladrillo | 19 |
| Análisis global sobre los distintos tipos de configuración de las viviendas | 21 |
| Análisis por orientación del edificio | 22 |
| Análisis del área y altura de las viviendas | 24 |
| Análisis por edad de la propiedad | 25 |
| ¿Los edificios con menor antigüedad sufrieron menos daño? | 25 |
| Analizando la altura y área para los edificios con 995 años | 28 |
| Análisis por cantidad de pisos de la vivienda | 29 |
| ¿Los edificios con mayor cantidad de pisos sufrieron un mayor daño? | 29 |
| ¿El plan de configuración para diseño sísmico ayudó a los edificios a recibir menos daño? | 30 |
| Análisis por tipo de construcción usado en sus otros pisos | 35 |
| Análisis por tipo de techo | 37 |
| Análisis de cantidad de familias por vivienda | 39 |
| Análisis del estado legal de la propiedad | 40 |
| Análisis por usos secundarios de los edificios | 41 |
| Conclusiones | 44 |
| Insights | 44 |

1. Introducción

Se propone analizar en el presente informe los datos obtenidos de www.drivendata.org, cuyo dataset está compuesto por encuestas realizadas por [Kathmandu Living Labs](#) y el [Central Bureau of Statistics](#). Particularmente, el dataset se enfoca en cómo eran las condiciones de una determinada vivienda y cuál fue su grado de daño luego del terremoto.

El objetivo principal de este informe es poder realizar un análisis exploratorio abarcativo donde, a medida que se exploran los datos, se vayan encontrando las preguntas como las respuestas a las incógnitas que fueron surgiendo. Se propone específicamente:

- Identificar patrones de viviendas que sufrieron un daño leve, medio y severo.
- Identificar cuáles atributos tuvieron más o menos incidencia a la hora de determinar el daño de una vivienda.
- Tratar de entender o darle un significado a los tipos ofuscados de algunos de los atributos.
- Identificar parcialmente qué zona geográfica de Nepal sufre una mayor cantidad de daño.

Finalmente, entre lo descubierto en el análisis exploratorio y los ítems marcados, se busca obtener un listado de insights aprendidos sobre los mismos y con ellos poder brindar posibles soluciones para reducir el nivel de daño recibido en las viviendas en Nepal en un contexto de terremoto.

2. Información general sobre los datos

Lo primero que debemos hacer es analizar la estructura general de los datos proporcionados, para comenzar a tener una idea de que es lo que se tiene y cómo debemos trabajar con ellos. Se observa que son 260.601 registros con 38 atributos.

Los atributos son:

- **building_id**: Identificador único de la edificación.
- **geo_level_1_id, geo_level_2_id, geo_level_3_id**: Región geográfica en la cual la edificación existe, desde la más general (level 1) a la más específica (level 3).
- **count_floors_pre_eq**: Número de pisos en la edificación antes del terremoto.
- **age**: Antigüedad de la edificación en años.
- **area_percentage**: Superficie normalizada ocupada por la edificación.
- **height_percentage**: Altura normalizada ocupada por la edificación.
- **land_surface_condition**: Condición de la superficie terrestre donde el edificio fue construido.
- **foundation_type**: Tipo de cimientos usados cuando se construyó la edificación.
- **roof_type**: Tipo de techo usado cuando se construyó la edificación.
- **ground_floor_type**: Tipo de construcción usado en la planta baja cuando se construyó la edificación.
- **other_floor_type**: Tipo de construcción usado en otros pisos cuando se construyó la edificación (exceptuando el techo).
- **position**: Orientación de la edificación.
- **plan_configuration**: Formato de la construcción de la edificación para diseño sísmicos.
- **has_superstructure_adobe_mud**: Indica si la edificación fue construida con adobe-barro.
- **has_superstructure_mud_mortar_stone**: Indica si la edificación fue construida con barro-piedra.
- **has_superstructure_stone_flag**: Indica si la edificación fue construida con piedra.
- **has_superstructure_cement_mortar_stone**: Indica si la edificación fue construida con cemento-piedra.
- **has_superstructure_mud_mortar_brick**: Indica si la edificación fue construida con barro-ladrillos.
- **has_superstructure_cement_mortar_brick**: Indica si la edificación fue construida con cemento-ladrillos.
- **has_superstructure_timber**: Indica si la edificación fue construida con Timber (madera específica para la construcción).
- **has_superstructure_bamboo**: Indica si la edificación fue construida con bambú (caña).
- **has_superstructure_rc_non_engineered**: Indica si la edificación fue construida con concreto reforzado no diseñado.
- **has_superstructure_rc_engineered**: Indica si la edificación fue construida con concreto reforzado diseñado.

- **has_superstructure_other:** Indica si la edificación fue construida con otro material.
- **legal_ownership_status:** Estado legal de la tierra donde la edificación fue construida.
- **count_families:** Número de familias que vivían en la edificación.
- **has_secondary_use:** Indica si la edificación era usada con un uso secundario.
- **has_secondary_use_X:** 10 atributos del mismo tipo, donde X puede ser agriculture, hotel, rental, institution, use_school, use_industry, use_health_post, use_gov_office, use_police, use_other.

2.1. Manipulación de los datos

En esta parte del análisis hacemos las configuraciones necesarias para trabajar mejor sobre el dataset. Las operaciones y transformaciones incluyen:

- **Conversión del tipo de datos:** Teniendo en cuenta que al cargar el set original no se infiere el tipo de dato de cada columna, se convierten los datos para tratarlos por su tipo original. La principal ventaja es el ahorro de memoria. Si bien en este caso puede que no sea tan necesario o el impacto no se vea reflejado, para datasets más grandes es bastante importante.
- **Manejo de nulos:** Afortunadamente el dataset no cuenta con algún nulo o NaN, permitiéndonos trabajar con mayor comodidad.
- **Data Mining:** Se extraen atributos importantes de los disponibles para trabajar de una mejor manera. Por ejemplo, las columnas de has_superstructure_X fueron agrupadas en una sola.

2.2. Truncamiento de los datos

Sabemos que el dataset proporcionado no representa el conjunto total de datos¹ de las viviendas que sufrieron algún tipo de daño en el contexto del terremoto del 2015 que afectó a Nepal. Por esta razón, las hipótesis que fueron elaboradas pueden no tener un fundamento claro si dichas hipótesis fueran comparadas con el dataset original. Sin embargo, revisando otras fuentes, creemos que el presente dataset representa con bastante fidelidad los resultados del terremoto de Gorkha.

¹ [Earthquake Data Portal](#)

3. Análisis por región geográfica

Analizaremos la información que nos brindan las columnas `geo_level_1_id`, `geo_level_2_id` y `geo_level_3_id`, intentando buscar si el terremoto afectó más a alguna zona en particular. Trabajando con los datos, podemos observar que hay 31 valores únicos de `geo_level_1_id`, que se corresponden con los 31 valores posibles del dataset. Se presentan 1.414 valores únicos de `geo_level_2_id`, sin embargo sabemos que puede tomar 1.428 valores posibles. Por lo tanto, para ciertos id's de esta zona no contamos con su información. Por su parte, de los 12.567 posibles valores para `geo_level_3_id`, tan sólo aparecen 11.595. Al igual que en el geo level 2, hay ciertos id's con los cuales no contamos con su información. Estas ubicaciones geográficas faltantes, tanto del nivel 2 como el 3, pueden ser debido a que no se contabilizaron las viviendas dañadas en esos lugares, que el dataset reducido con el que trabajamos no tenga algunos id's que sí tiene el dataset original o por alguna razón particular esas zonas no se vieron afectadas.

Cabe destacar que los id's 28, 19 y 29 del geo level 1 resultan ser las zonas menos damnificadas por el terremoto. Suponemos que esto se debe a que esas zonas se encuentran alejadas del epicentro del terremoto. Sin embargo, debido a la ofuscación de los datos, no podemos corroborarlo. A partir del id 2 se triplica la cantidad de viviendas destruidas. Los valores 6, 26, 10 y 17 de id correspondientes al geo level 1 representan a las zonas más afectadas por el terremoto.

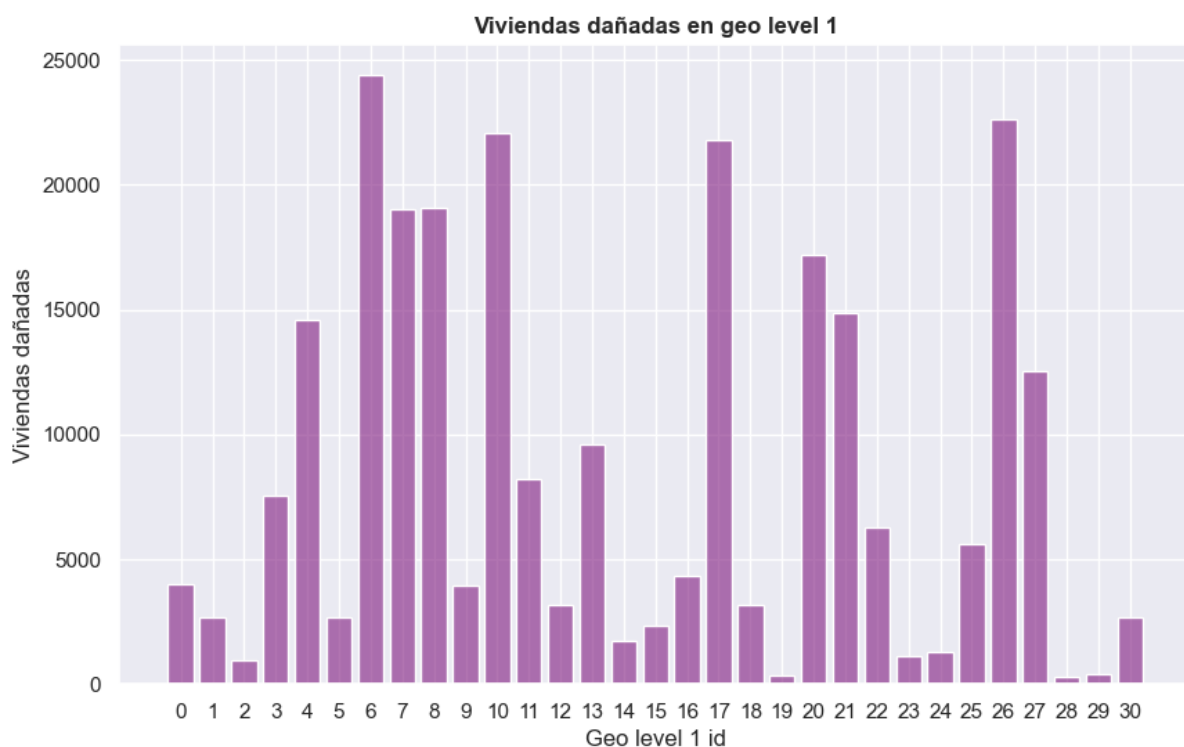


Figura 1: Viviendas dañadas en la zona 1

3.1. ¿Qué nivel de daño tiene cada zona?

Nos concentramos en analizar sólo el `geo_level_id_1` puesto que tiene muchos menos valores posibles y, de acuerdo al significado de dicha columna, es la que nos brinda un panorama más general de lo acontecido a nivel geográfico.

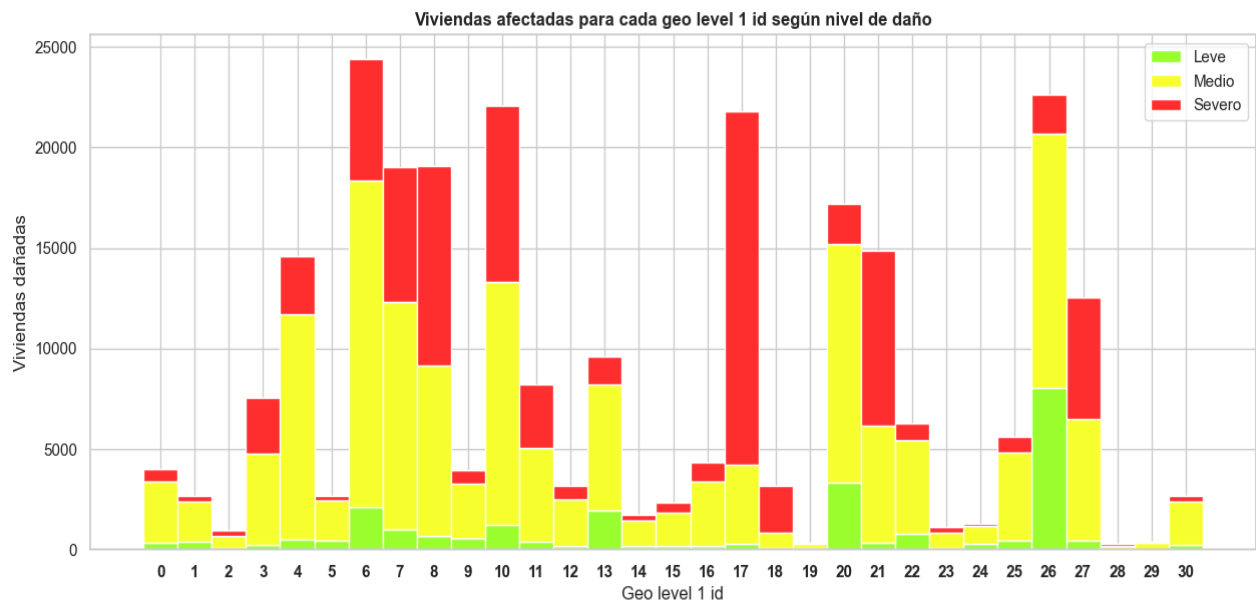


Figura 2: Viviendas dañadas de la zona 1 por nivel de daño

A partir de la figura 2 vemos que, dentro del geo level 1, el ID 17 se destaca notablemente de los demás ya que se lleva la mayor cantidad de viviendas afectadas con un nivel de daño severo.

Otro ID notable es el 6 que posee la mayor cantidad de viviendas damnificadas. Sin embargo, su gran mayoría son de daño medio. En contrapartida, remarcamos los id's 19, 28 y 29 por tener ínfimos daños en comparación al resto.

Si bien la zona 26 se encuentra entre las más afectadas, cabe señalar que hubo un gran porcentaje que sólo sufrió daños leves.

4. Análisis por condición de la superficie terrestre

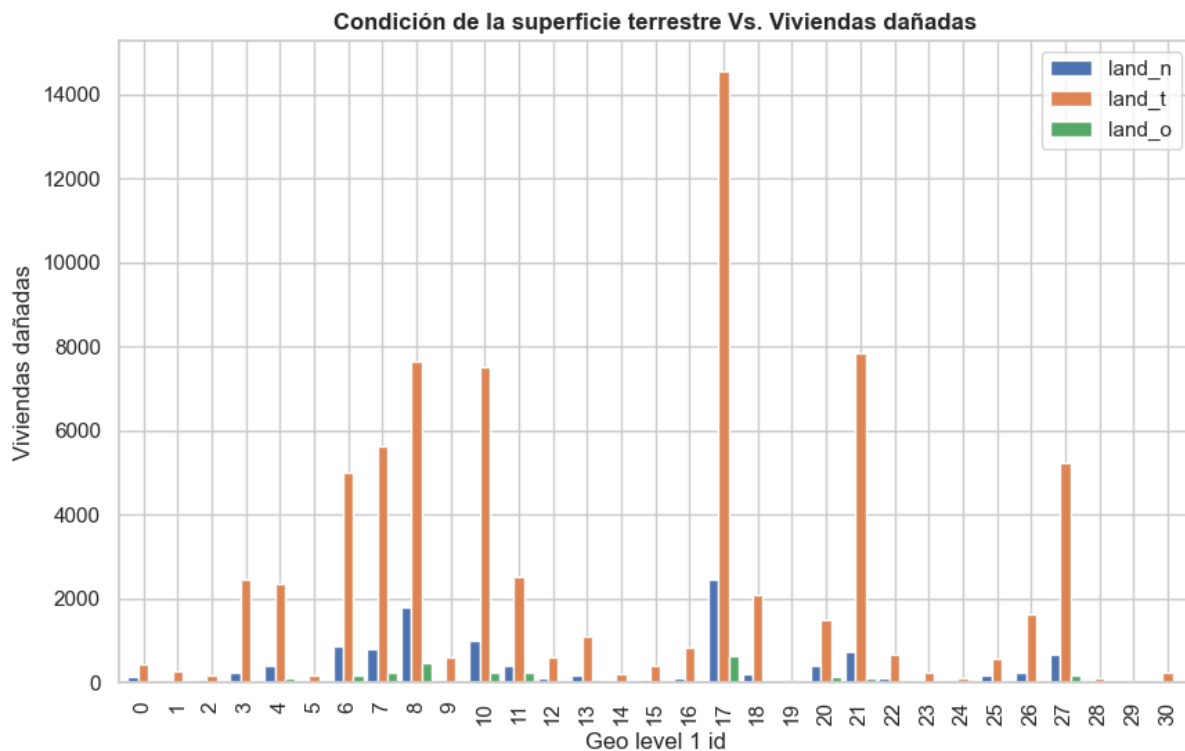


Figura 3: Condición de la superficie en la zona 1

A partir del gráfico anterior podemos visualizar que predomina la condición de superficie terrestre tipo “t”, por lo que suponemos que este tipo de superficie no es ideal para eventos sísmicos.

Además por el gráfico previo a este, sabíamos que la mayor cantidad de viviendas dañadas con nivel severo se encontraban en el ID 17, esto coincide con este gráfico ya que el máximo de viviendas dañadas de suelo “t” se encuentran en dicha zona.

5. Análisis por tipo de cimientos

5.1. ¿Podemos emparejar algún tipo de cimiento con algún tipo de superestructura?

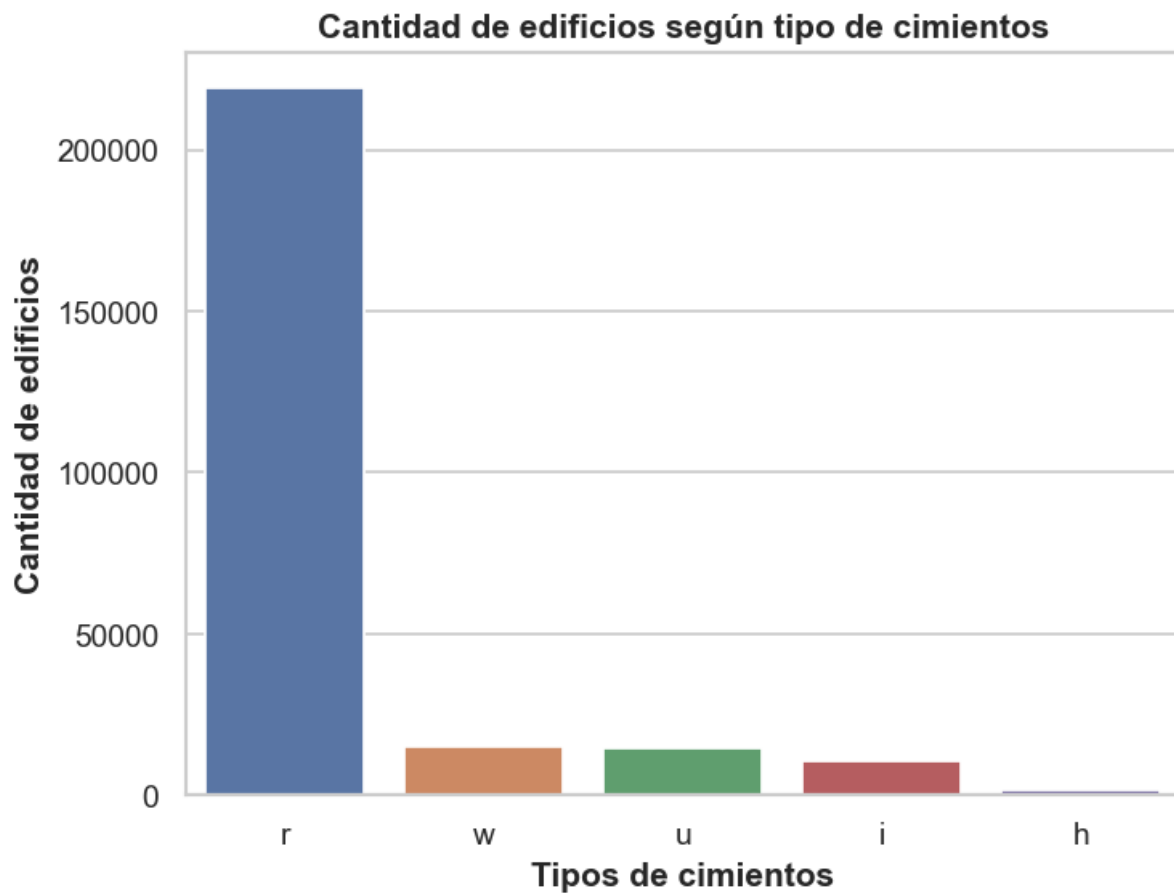


Figura 4: Cantidad de edificios según tipo de cimientos

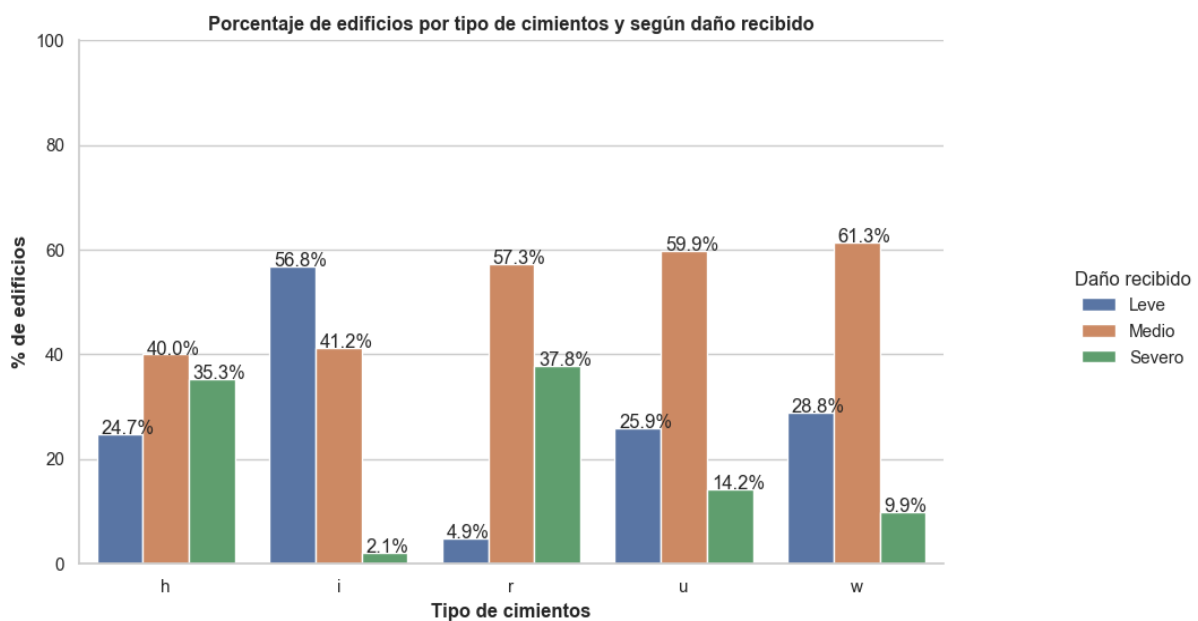


Figura 5: Porcentaje de edificios por tipo de cimientos y según daño recibido

Si comparamos los tipos “u” y “w” no notamos diferencias significativas. Llamen la atención los tipos “r” e “i”, el primero por tener la mayor cantidad de observaciones en el dataset como lo vimos en la Figura 4 y el menor porcentaje de viviendas con daño leve; y el segundo por tener el mayor porcentaje de viviendas con daño leve pasando el 50% y además teniendo el menor porcentaje de viviendas con daño severo. Creemos que el tipo “r” está relacionado con aquellas viviendas con barro-piedra y el tipo “i” con algún tipo de concreto reforzado o cemento-ladrillos que ya vimos que son las superestructuras más resistentes.

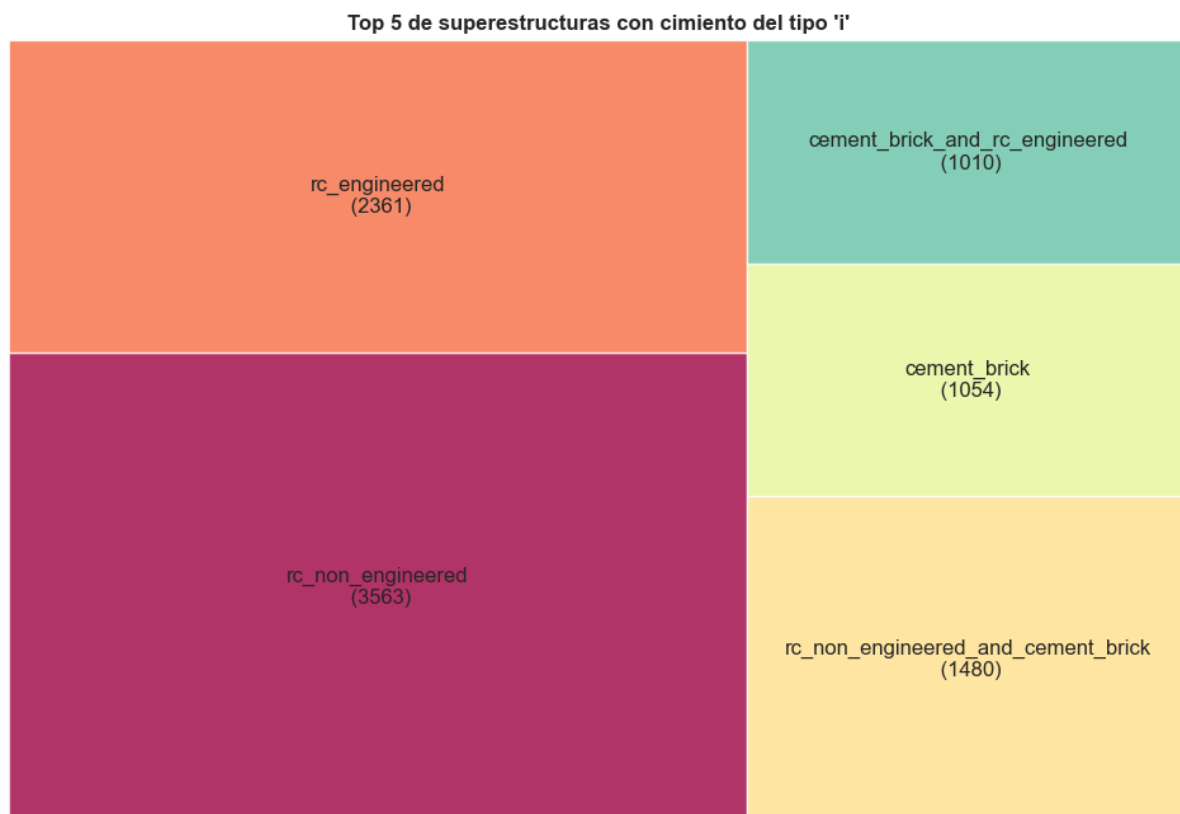


Figura 6: Top 5 de superestructuras con cimiento de tipo “i”

Efectivamente, el tipo de cimiento “i” lo podemos asociar a las viviendas que tengan las superestructuras de concreto reforzado diseñado/no diseñado y/o las que tengan cemento-ladrillo.

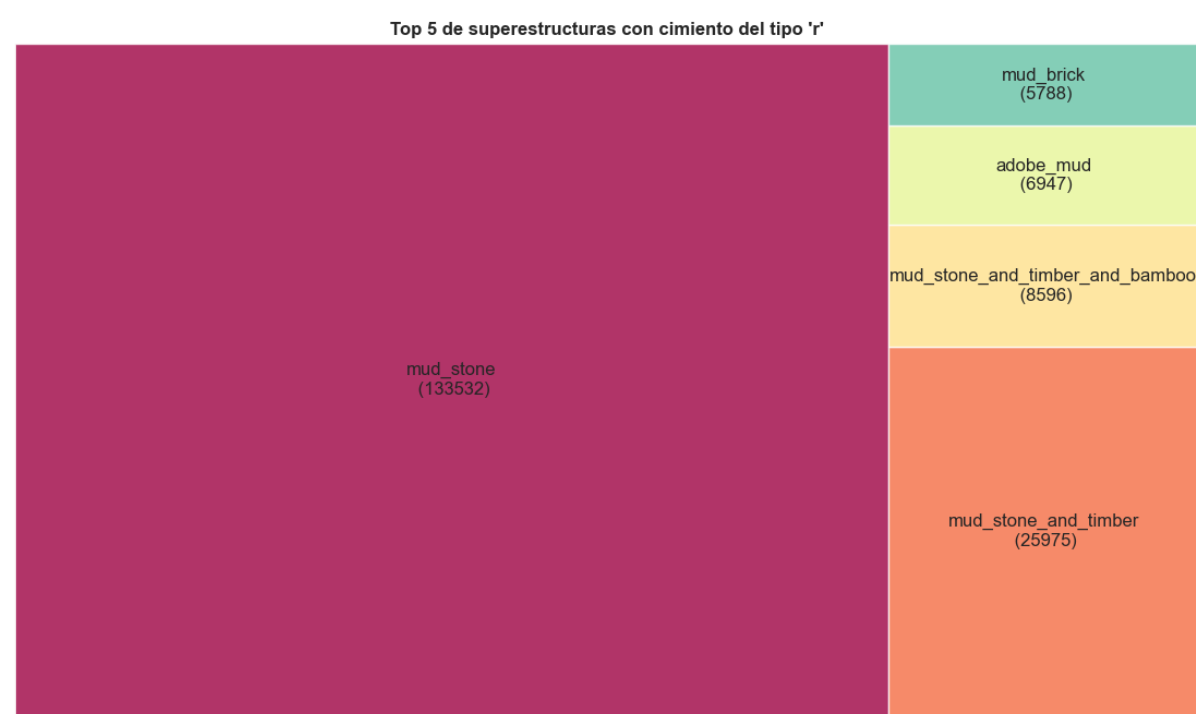


Figura 7: Top 5 de superestructuras con cimientto de tipo “r”

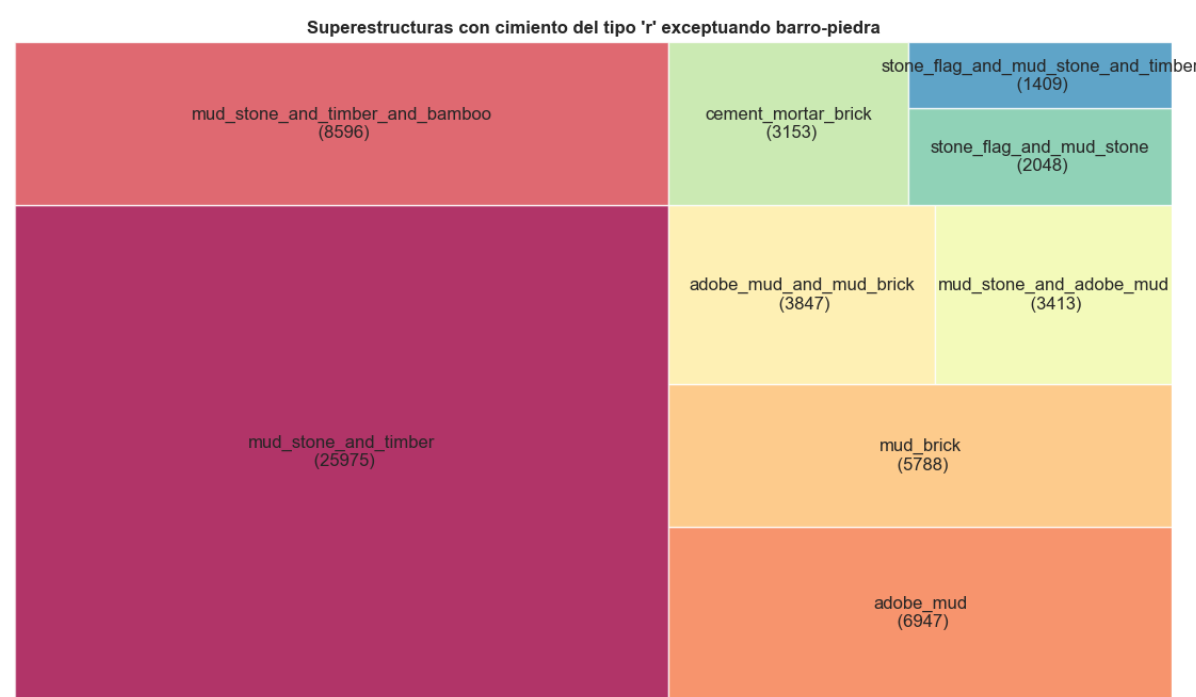


Figura 8: Superestructuras con cimientto de tipo “r” exceptuando barro-piedra

Soportando nuestra hipótesis, el tipo “r” de cimientto lo podemos asociar a aquellas superestructuras que tengan alguna combinación de barro-piedra, timber, bambú, barro-ladrillo o barro-adobe. Es decir, a aquella superestructura que haya usado al barro como elemento de unión.

6. Análisis por tipo de planta baja

A continuación analizaremos los distintos tipos de suelos de las propiedades y su relación en el daño recibido.



Figura 9: Nivel de daño de las viviendas dañadas separadas por tipo de planta baja

Se puede apreciar que aquellas viviendas con tipo de planta baja “f” fueron claramente las más afectadas por el terremoto, recibiendo ampliamente el mayor daño de nivel medio y severo. Para analizar si también hay una influencia del tipo de suelo donde fueron construidas dichas propiedades realizaremos un análisis sobre estas viviendas en particular.

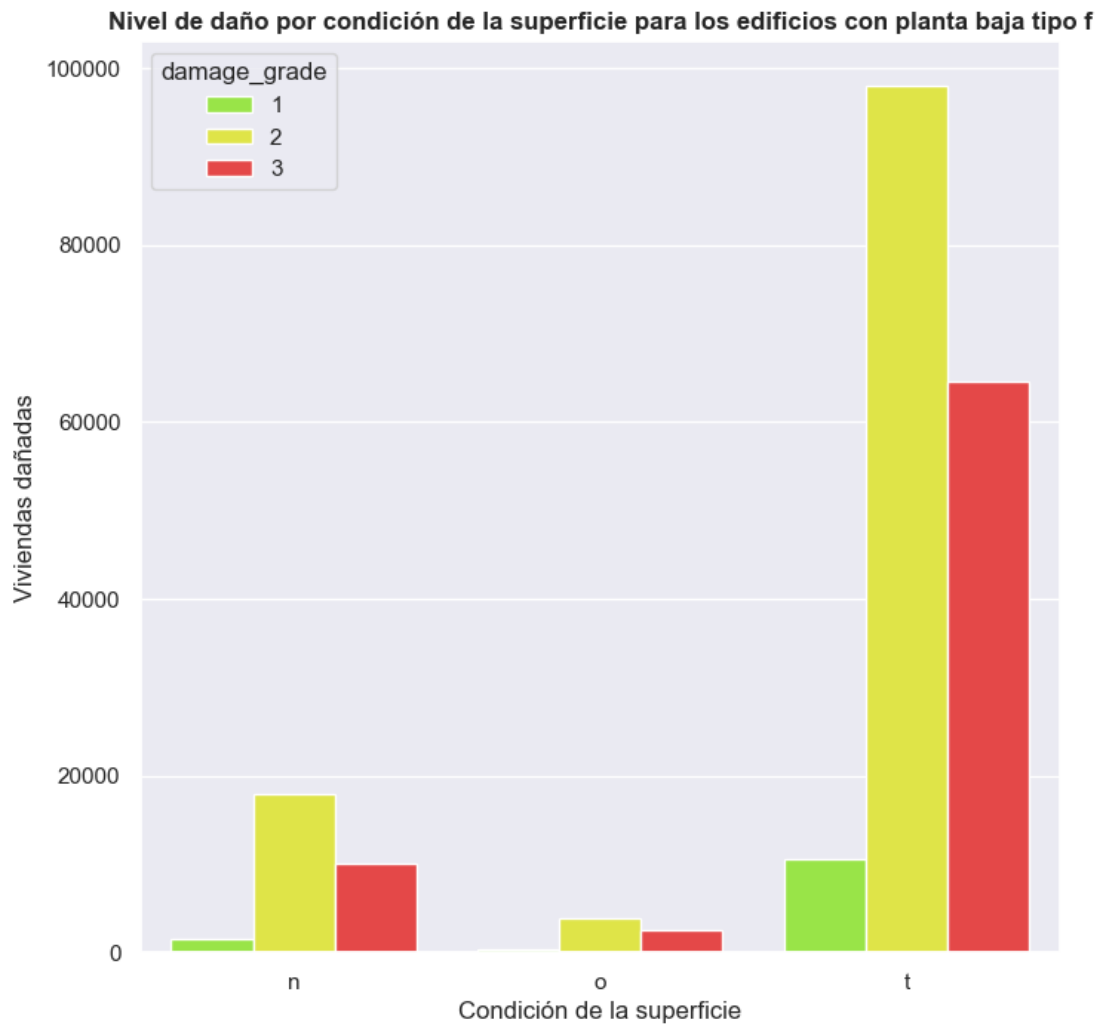


Figura 10: Nivel de daño por condición de la superficie de las viviendas con planta baja de tipo “f”

Con este segundo gráfico, podemos concluir que las viviendas más afectadas por el sismo, son aquellas en donde la propiedad fue construida en una superficie de condición “t” y con planta baja de tipo “f”.

7. Análisis por tipo de estructura

Comenzamos por un análisis cuantitativo, queremos ver con qué material fueron contruidos los edificios. Buscaremos analizar si dichos materiales tuvieron o no incidencia en el nivel de daño recibido.

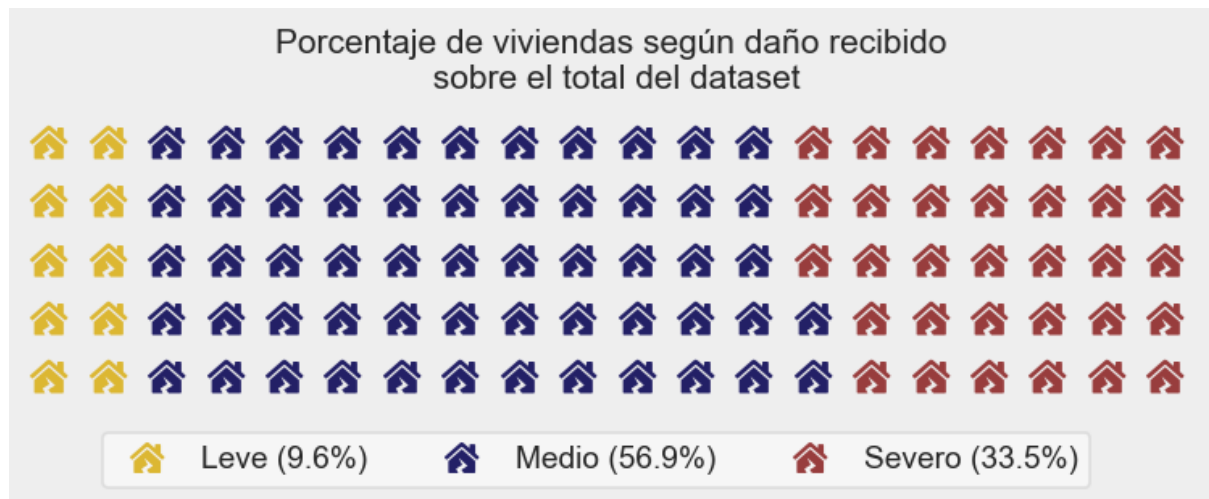


Figura 11: Porcentaje de viviendas según daño recibido

Vemos que alrededor del 57% de las viviendas recibieron un daño de nivel medio, mientras que alrededor del 33% recibieron un daño de nivel severo. Esto es importante para los análisis posteriores puesto que el dataset se encuentra medio sesgado a partir de estos valores; dándonos como resultado que el nivel de daño medio salga como el mayor en algunas comparaciones. Por lo tanto, nos centraremos en comparar tanto los niveles de daño leve y severo.

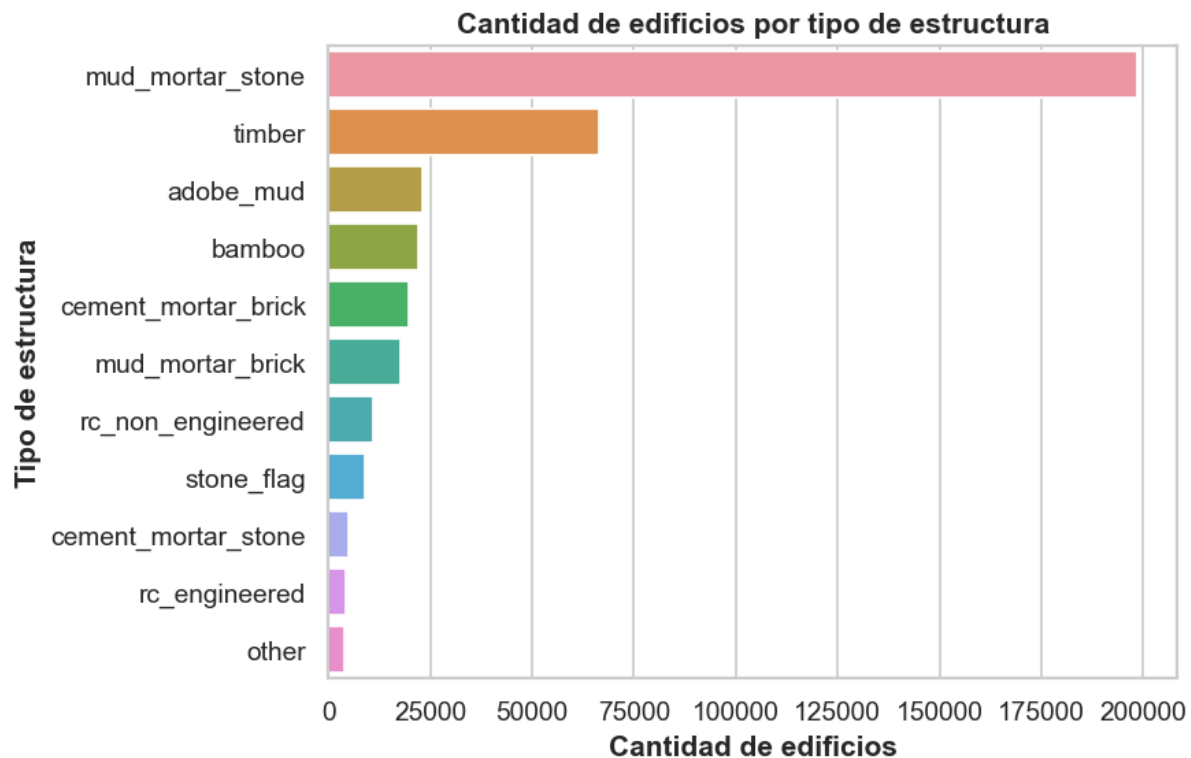


Figura 12: Cantidad de edificios por tipo de estructura

Podemos ver que la gran mayoría (76%) de las edificaciones totales utilizaron barro-piedra para su construcción, 25% fueron construidas con Timber (madera específica para construcción) seguido de un 9% con adobe-barro y bambú y un 8% con cemento-ladrillo. Estas categorías no son excluyentes, es decir, hay viviendas que fueron construidas con barro-piedra y por ejemplo se utilizó timber para soporte interno. Todos estos datos tienen concordancia con lo que se ve de Nepal a nivel nacional², fuera del contexto del terremoto. Entendemos que la gran cantidad de viviendas con barro-piedra, bambú o timber provienen de la población que se va mudando de zonas rurales a zonas urbanas, permitiéndoles construir un hogar en base a sus necesidades y al presupuesto disponible en el momento.

Como observación, recién al final de la década de 1970 se empezó a construir viviendas con concreto reforzado, y por esa razón las pocas que hay construidas son en su mayoría de concreto reforzado no diseñado.

² [Common structural and construction deficiencies of Nepalese buildings](#)

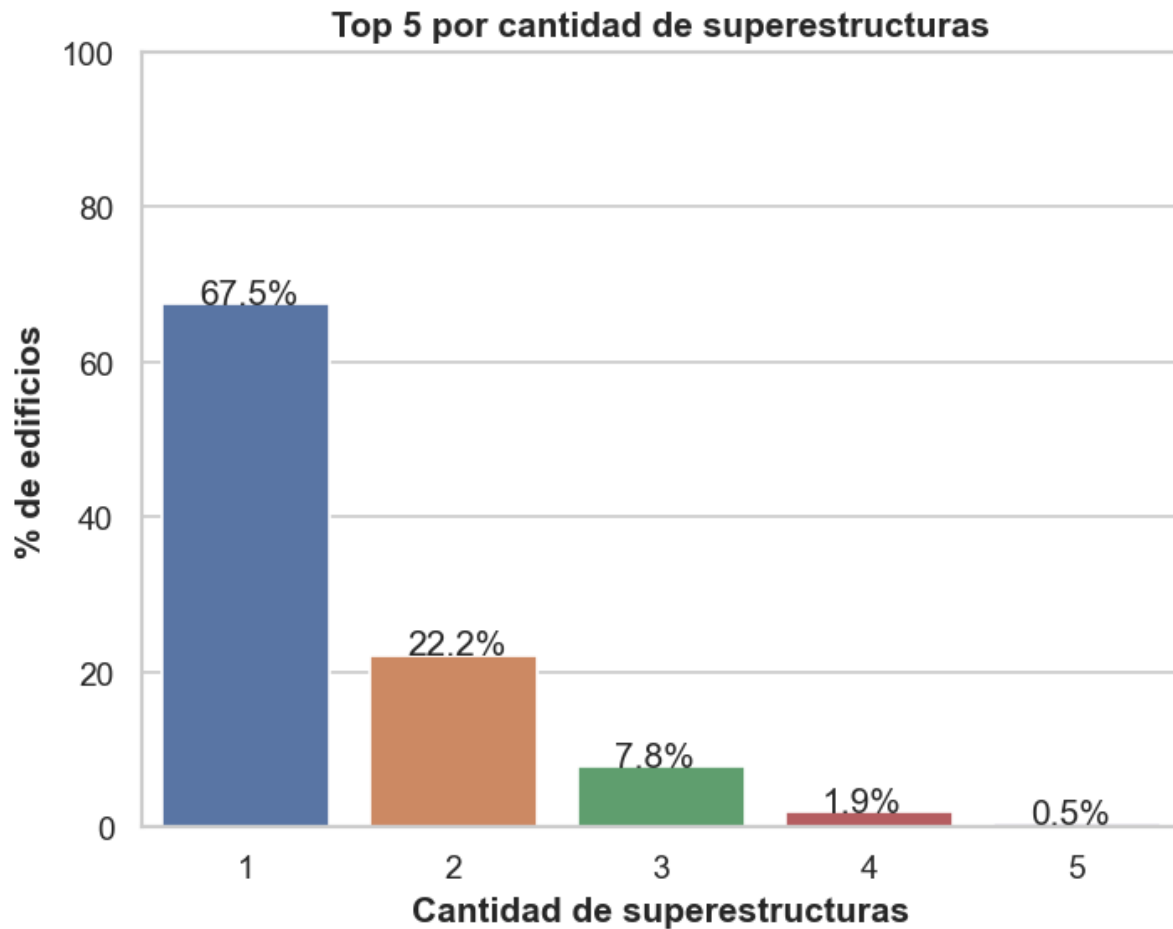


Figura 13: Top 5 por cantidad de superestructuras

La gran mayoría de las viviendas del dataset poseen sólo una única superestructura, seguidas de un porcentaje que no debemos despreciar del 22% que poseen dos superestructuras. Esto nos plantea las preguntas de si a mayor número de superestructuras menor fue el daño recibido.

7.1. ¿A mayor cantidad de superestructuras, menor es el daño recibido?

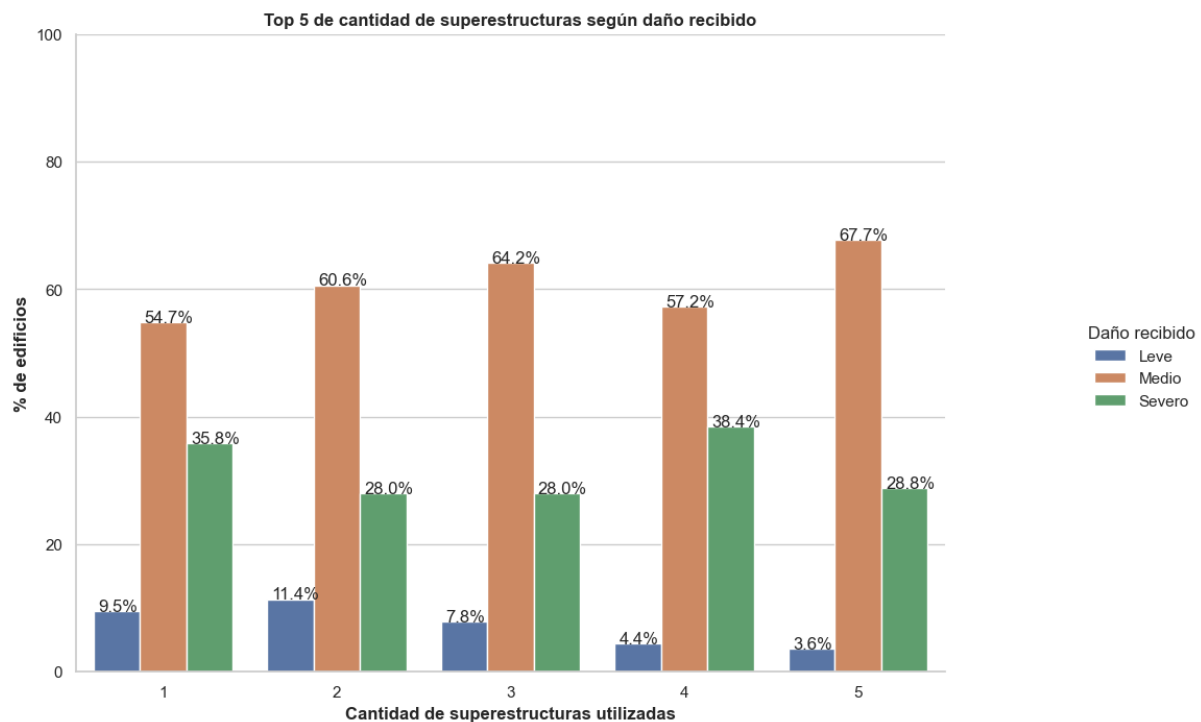


Figura 14: Top 5 de cantidad de superestructuras según daño recibido

Podemos apreciar que a medida que aumenta la cantidad de superestructuras utilizadas, aumenta parcialmente el porcentaje de edificios que sufrieron un daño medio. Esto se encuentra principalmente afectado por la cantidad de viviendas que sufrieron un grado de daño de nivel 2, que representan la mayoría del dataset. Si nos centramos únicamente en los niveles 1-leve y 3-severo, tampoco podemos sacar mucha información ya que en algunos casos éstos aumentan o disminuyen para aumentar el porcentaje de viviendas que recibieron un daño medio.

En general, no podemos afirmar que a mayor cantidad de superestructuras utilizadas, menor es el daño recibido. Por lo tanto, vamos a realizar un análisis un poco más profundo centrándonos en los tipos de superestructuras utilizados en las viviendas.

7.2. ¿Cuál fue el tipo de edificación que se vio más afectada? ¿Cuál fue la que menos fue afectada?

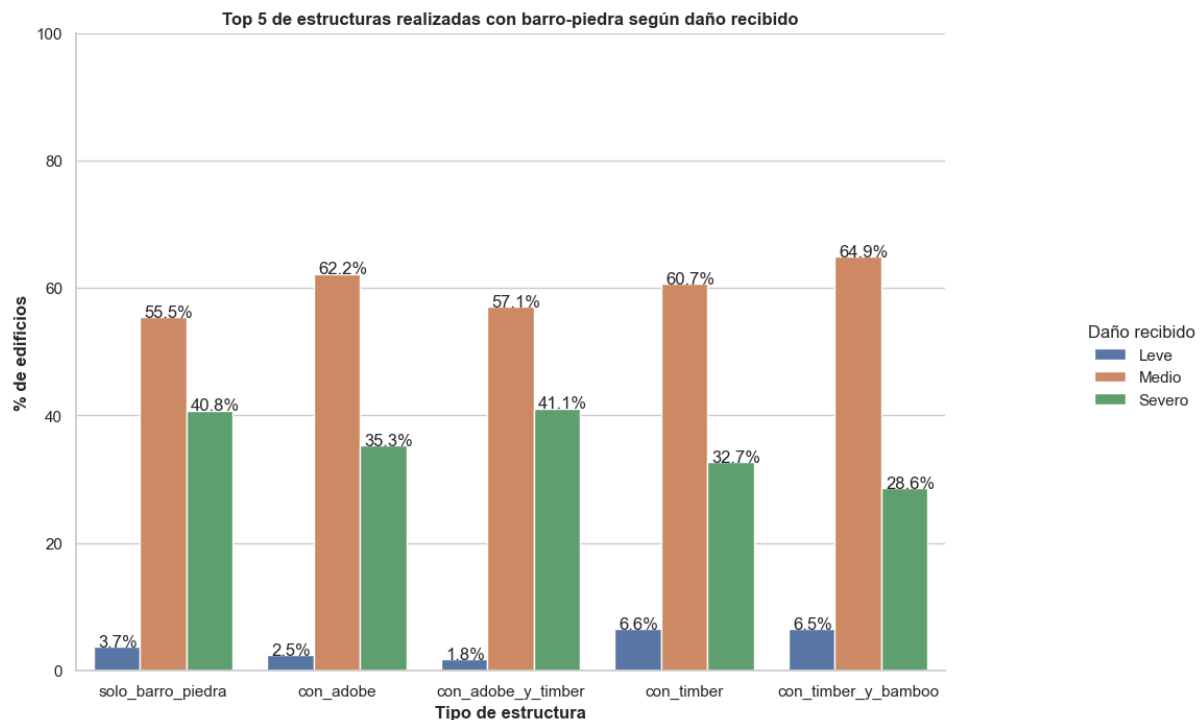


Figura 15: Top 5 de estructuras realizadas con barro-piedra según daño recibido

Dentro del grupo de construcciones fabricadas con barro-piedra, podemos ver que aquellas estructuras construidas con únicamente barro-piedra y aquellas con barro-piedra-adobe-timber fueron las más afectadas pero no por mucho margen. Por otro lado, las construcciones que fueron fabricadas con timber recibieron un poco menos de daño. Esto tiene sentido ya que el timber es un material que suele ser resistente a daños sísmicos si se usa como soporte interno³. Sin embargo, el caso de barro-piedra-adobe-timber nos puede indicar que el nivel de intensidad de los terremotos que afectaron a ese tipo de estructuras fueron lo suficientemente fuertes como para dañarlos, pese a tener una mayor cantidad de estructuras diversas.

En general, el uso de barro-piedra como combinación principal de la vivienda no es sería lo aconsejado para este tipo de episodios ya que presentan una pobre performance a la hora de reducir los daños catalogados como severos.

³ [Earthquake resistant timber panelised structures with resilient connections](#)

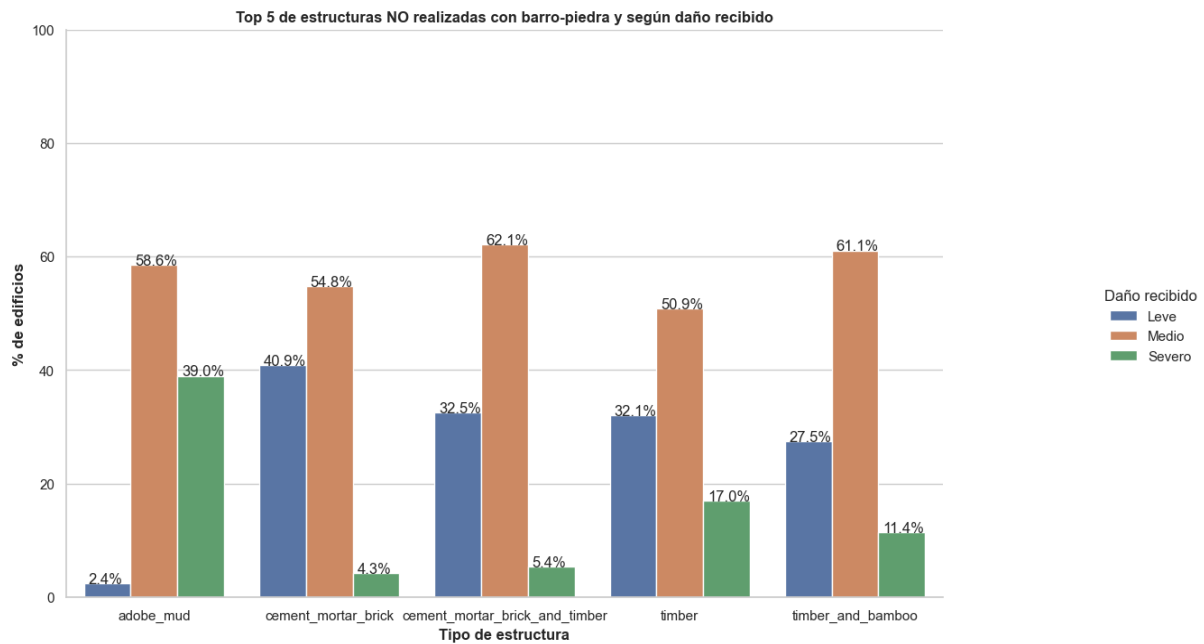


Figura 16: Top 5 de estructuras no realizadas con barro-piedra y según daño recibido

Si dejamos afuera a las viviendas de barro-piedra, vemos una vez más que el timber es uno de los materiales importantes a la hora de reducir los daños recibidos. Si bien es un tipo de madera, ésta es resistente a daños sísmicos, pero también puede suceder que los edificios construidos con este material se encuentren en una zona que no fue el epicentro del terremoto. Al mismo tiempo, se observa que el cemento-ladrillo es muchísimo mejor que barro-piedra, siendo este último más propicio a recibir peores daños. Esto puede deberse a que, independientemente de la diferencia entre materiales ladrillo-piedra, el barro es una pésima sustancia para usarla como mortero a la hora de unir materiales. Por otro lado, se ve que adobe-barro tiene un comportamiento similar a barro-piedra en cuanto a los 3 niveles de daños comparados.

Es de suma importancia aclarar que, en función a la cantidad de datos registrados, se dejó de lado en esta comparación a 3 tipos de edificaciones que, a priori, serían las viviendas más y menos resistentes a daños sísmicos. Los 3 tipos mencionados se compararán a continuación.

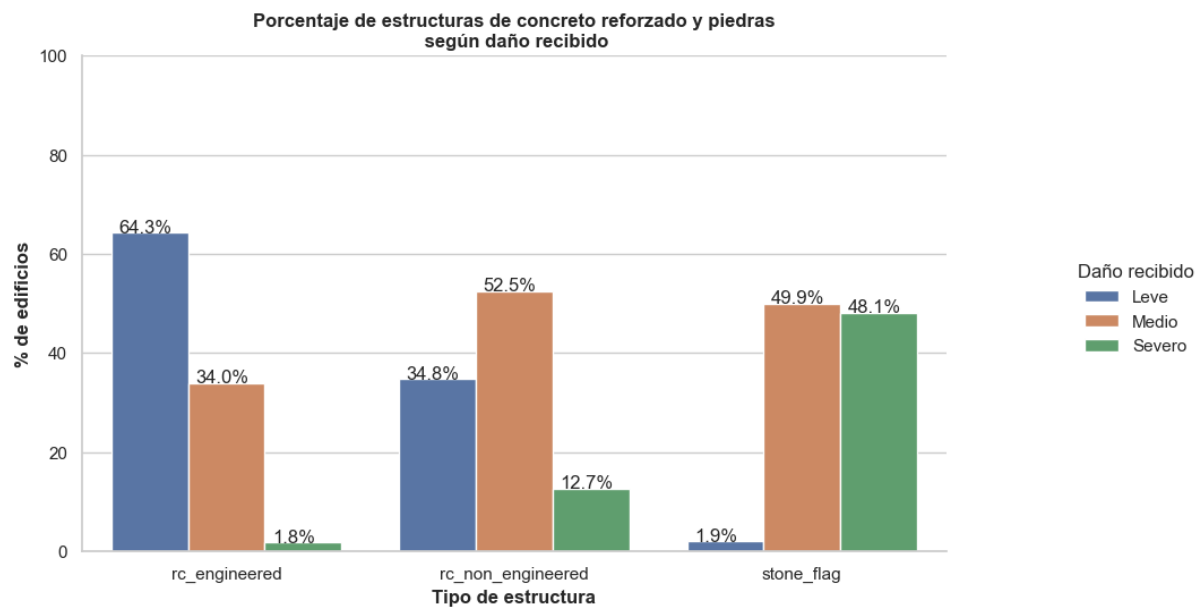


Figura 17: Porcentaje de estructuras de concreto reforzado y piedras según daño recibido

Es evidente que dentro de este pequeño grupo, en concordancia con nuestra hipótesis, aquellas edificaciones con concreto reforzado diseñado son las que menos daños sufrieron mientras que aquellas fabricadas con solamente piedra fueron las que más daño recibieron. Esto es un claro indicio que un diseño estructural adecuado es la clave para reducir algún riesgo de daño proveniente de un terremoto.

7.3. Comparación de morteros: cemento-ladrillo vs. barro-ladrillo

Ya vimos en los gráficos anteriores que cualquier combinación que presente a barro-piedra como sus principales materiales va a tener un rendimiento muy pobre a la hora de mitigar daños. Ahora bien, esta hipótesis puede deberse a que tanto el barro o la piedra sean materiales pocos propicios. Por lo tanto, vamos a comparar los casos que poseen cemento-ladrillo contra aquellos que tienen barro-ladrillo ya que ambas categorías tienen una cantidad de observaciones muy similar.

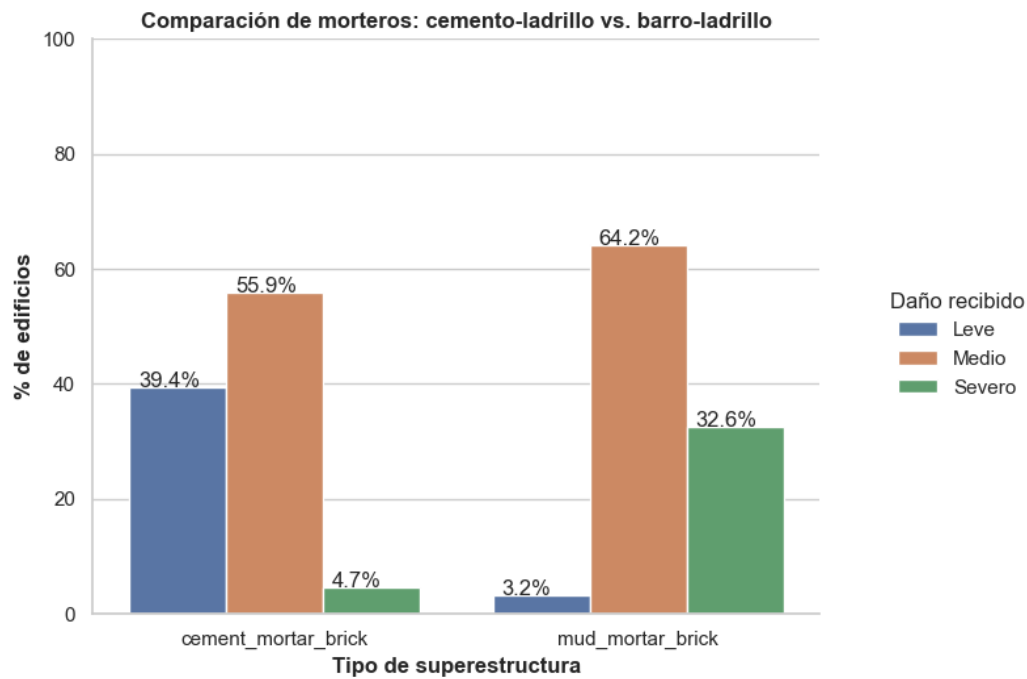


Figura 18: Comparación de morteros

Los resultados se encuentran a la vista, para un mismo material la variante más importante es el tipo de mortero utilizado reduciendo en alrededor de un 28% la cantidad de edificios que recibieron daño severo, y disminuyendo alrededor de un 8% la cantidad de viviendas que recibieron un daño medio; aumentando así un 36% los edificios que recibieron daño leve. Este dato es muy importante teniendo en cuenta que la gran mayoría de viviendas de este dataset (y también a nivel nacional) presentan algún tipo de combinación que incluyen barro-piedra.

8. Análisis global sobre los distintos tipos de configuración de las viviendas

En esta parte del informe, haremos un análisis general sobre las categorías que consideramos merecen un análisis general: la condición de la superficie terrestre, los tipos de cimiento, el tipo de piso de la planta baja y el plan de configuración para diseño sísmico.

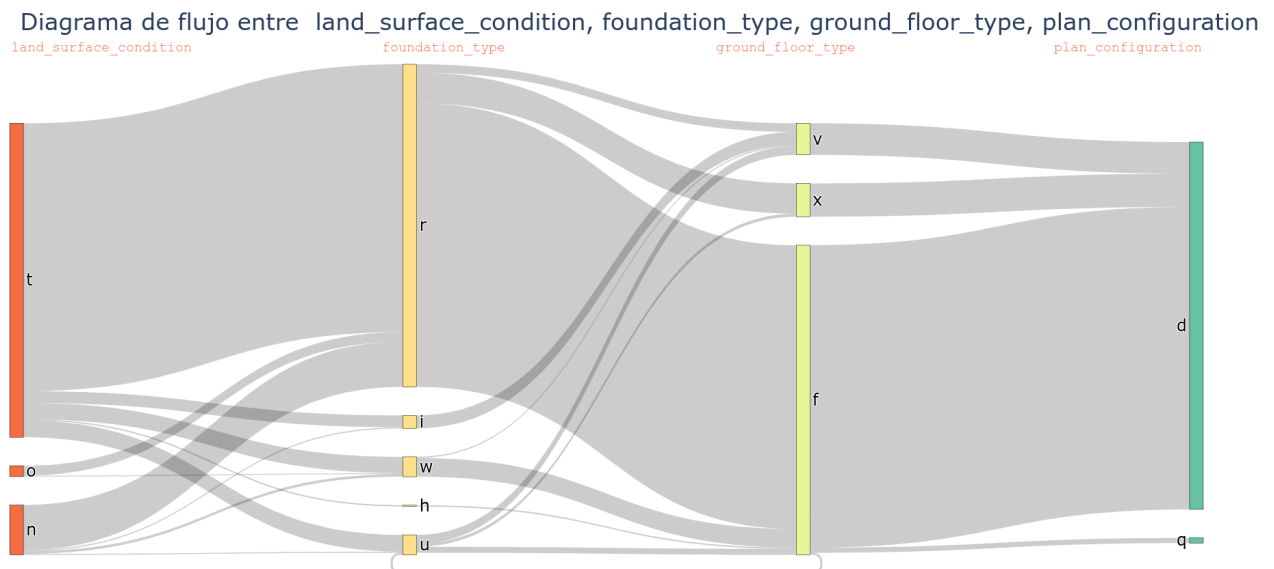


Figura 19: Diagrama de flujo entre los distintos tipos de configuración de las viviendas

La figura 19 ilustra de una manera muy clara las distintas variantes que se encontraron a lo largo de este análisis. El orden que se aprecia de izquierda a derecha es el siguiente:

- La primera columna (naranja) representa los tres tipos con mayor cantidad de observaciones de configuración de la superficie terrestre: “n”, “o” y “t”.
- La segunda columna (amarilla) representa los cinco tipos de cimientos posibles: “r”, “i”, “w”, “h” y “u”.
- La tercera columna (verde claro) representa los tres tipos con mayor cantidad de observaciones de piso para planta baja: “v”, “x” y “f”.
- La cuarta columna (verde oscuro) representa los dos tipos con mayor cantidad de observaciones de configuración para diseño sísmico: “d” y “q”.

Lo interesante de esta figura, es que permite ver rápidamente cuál fue la **combinación más vista**, siendo esta la que contiene las características menos resistentes a la hora de enfrentar un sismo. Esta es:

- ❖ *Configuración de superficie terrestre: t*
- ❖ *Tipo de cimiento: r*
- ❖ *Tipo de piso para planta baja: f*
- ❖ *Tipo de configuración para diseño sísmico: d*

9. Análisis por orientación del edificio

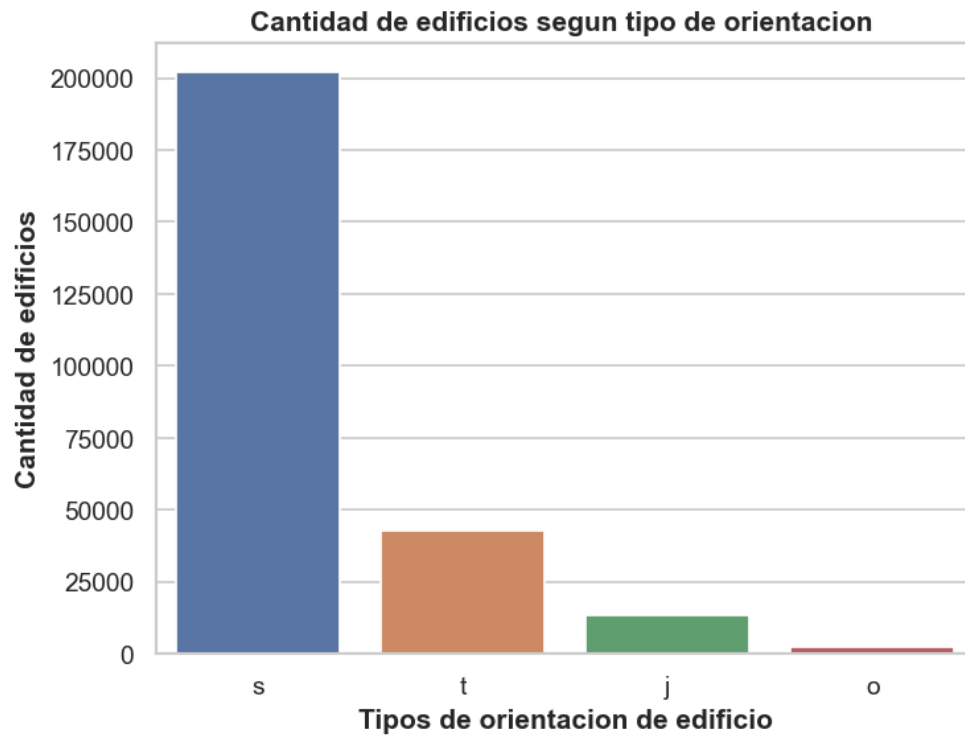


Figura 20: Cantidad de edificios según tipo de orientación

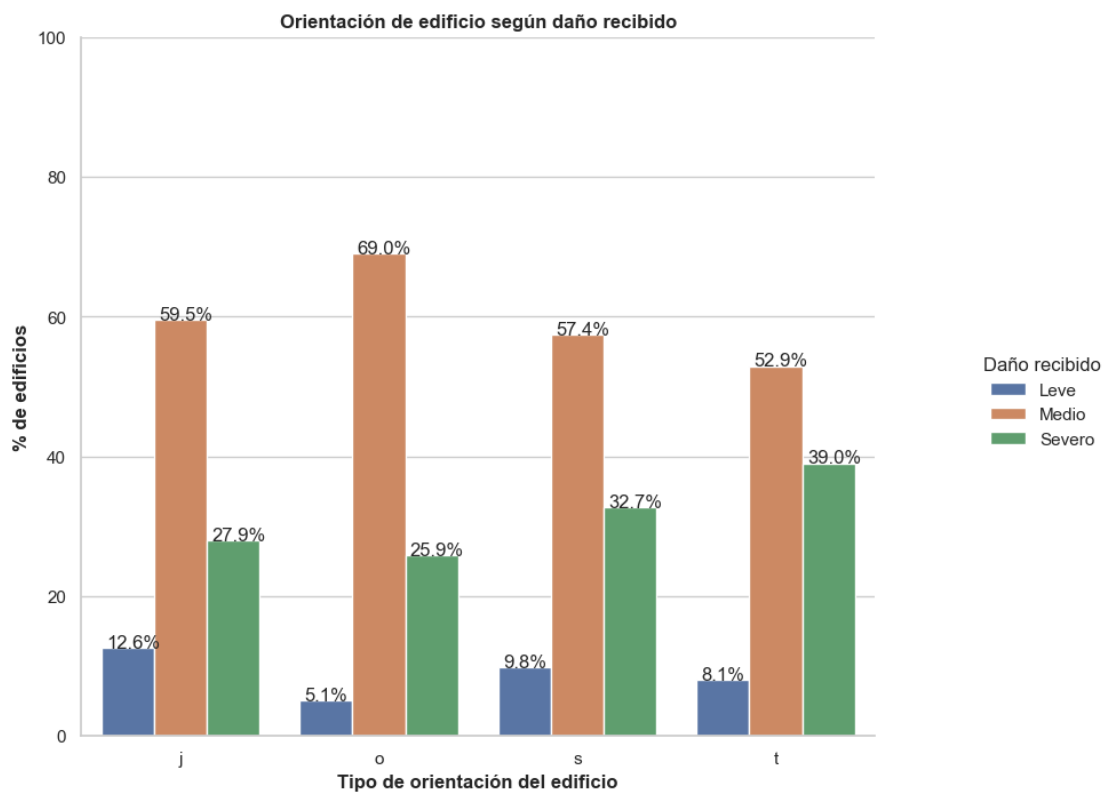


Figura 21: Porcentaje de edificios dañados según su orientación

A priori, el tipo de orientación que mejor actuó ante daños sísmicos es el tipo “j” ya que es el que tiene el mayor porcentaje de viviendas con daño leve y es el segundo con un porcentaje menor de edificios con un daño severo. Creemos que el tipo “j” es aquel que indica que un edificio tiene una orientación lo más recta posible. Por otro lado, el tipo “t” es una de las peores orientaciones con el mayor porcentaje de edificios que recibieron un daño severo. Este tipo de orientación debe estar relacionada con que el edificio presenta una inclinación bastante marcada, produciendo que al mínimo movimiento del suelo se maximicen los daños recibidos.

En la figura 20 vimos que el tipo “s” es el que tiene mayor representación en el dataset y puede ser que este tipo represente algún tipo de inclinación pero no tan acentuada como el tipo “t”. La hipótesis se basa en que Nepal tiene una topografía compleja con poco terreno llano generando que las viviendas se construyan sobre lugares pocos propicios en condiciones normales; siendo mucho peor si le agregamos que es una zona con alta probabilidad de terremotos.

De todas maneras, no vemos ningún cambio significativo entre los distintos tipos de esta categoría, por lo que tendría poca relevancia a la hora de determinar el nivel de daño recibido de una vivienda.

10. Análisis del área y altura de las viviendas

En esta sección analizaremos la incidencia de la altura y el área de cada vivienda en el daño sufrido.

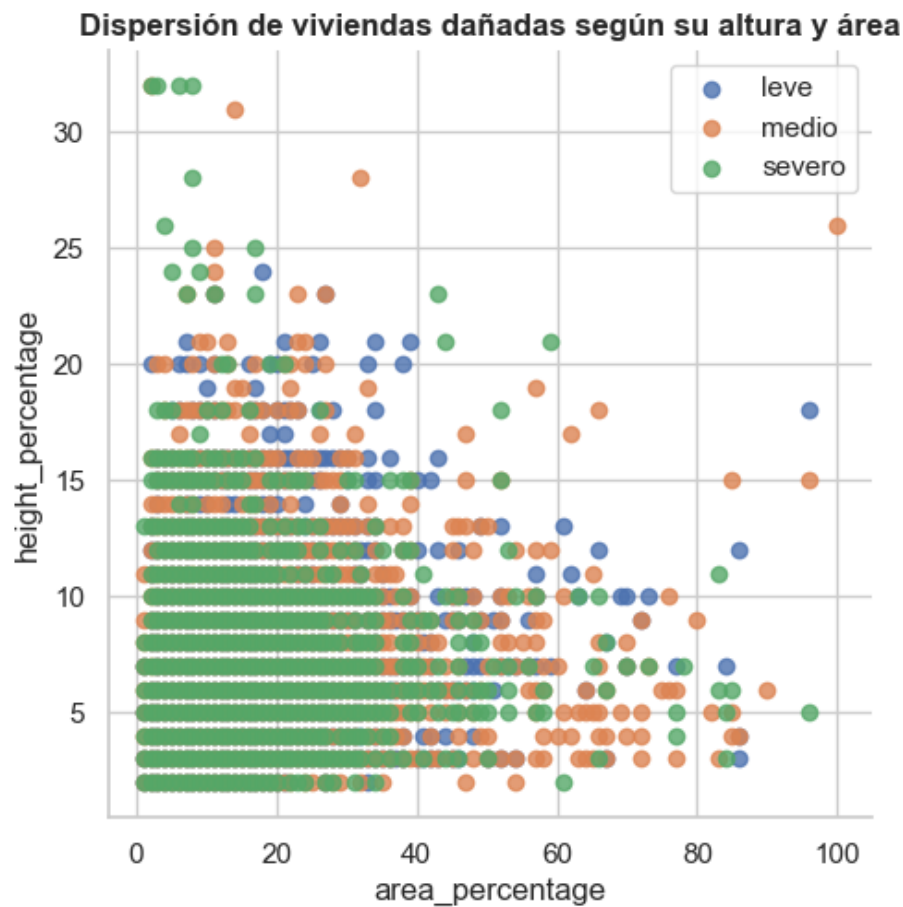


Figura 22: Dispersión de viviendas dañadas según su altura y área

En el gráfico de dispersión podemos observar que la mayor cantidad de viviendas destruidas se encuentran entre los valores 0 - 40 de área y 0 - 17 de altura. No podemos asegurar que el área y/o la altura hayan aumentado el nivel de daño sufrido por la vivienda ya que para propiedades de igual valor de área y altura hay distintos niveles de daño.

11. Análisis por edad de la propiedad

11.1. ¿Los edificios con menor antigüedad sufrieron menos daño?

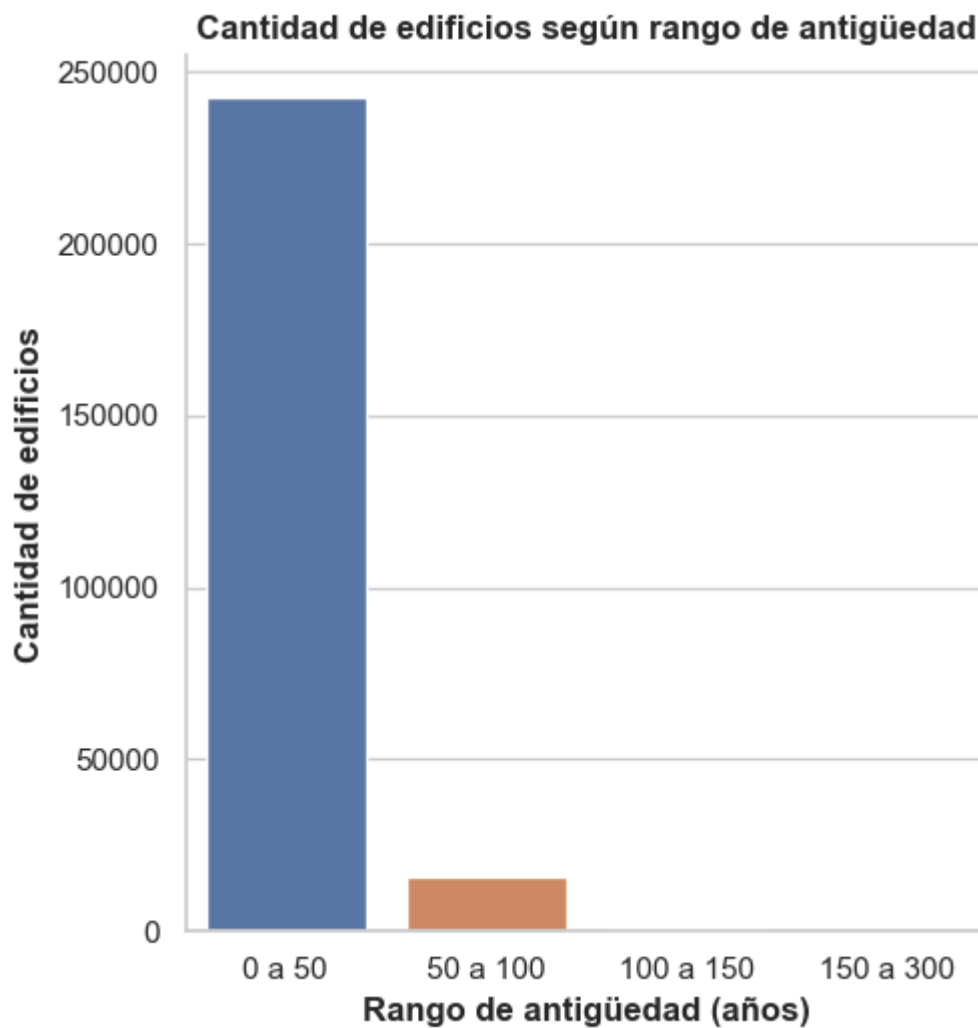


Figura 23: Cantidad de edificios según rango de antigüedad

Casi la totalidad de los edificios del dataset están concentrados en el rango de 0 a 50 años. Nos quedaremos con este rango para análisis posteriores.

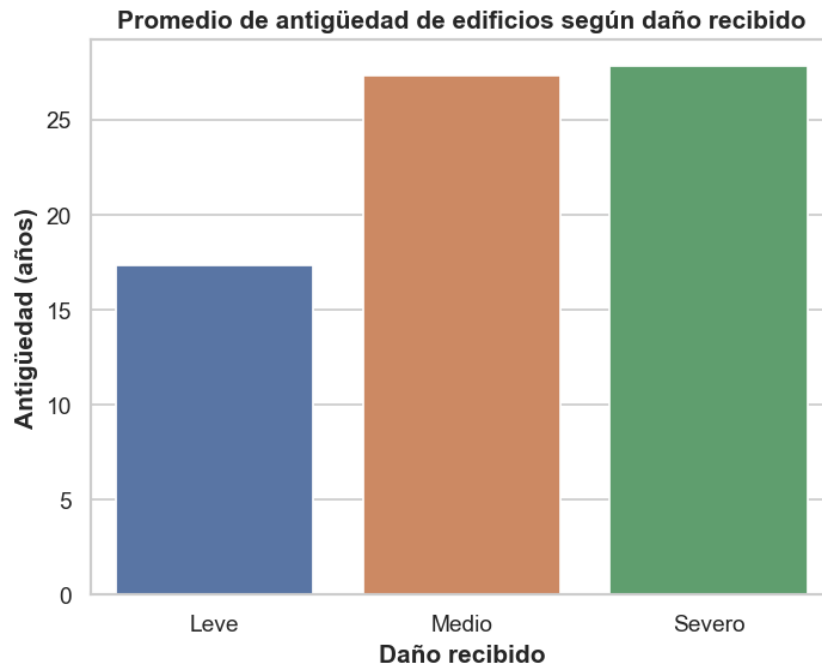


Figura 24: Promedio de antigüedad de edificios según daño recibido

Si consideramos el promedio de años de antigüedad de los edificios, vemos que a priori a menor antigüedad de los edificios menor es el daño que recibieron. Sin embargo, el nivel de daño medio y severo tienen un promedio de edad muy similar, razón por la cual vamos a profundizar sobre estas cuestiones.

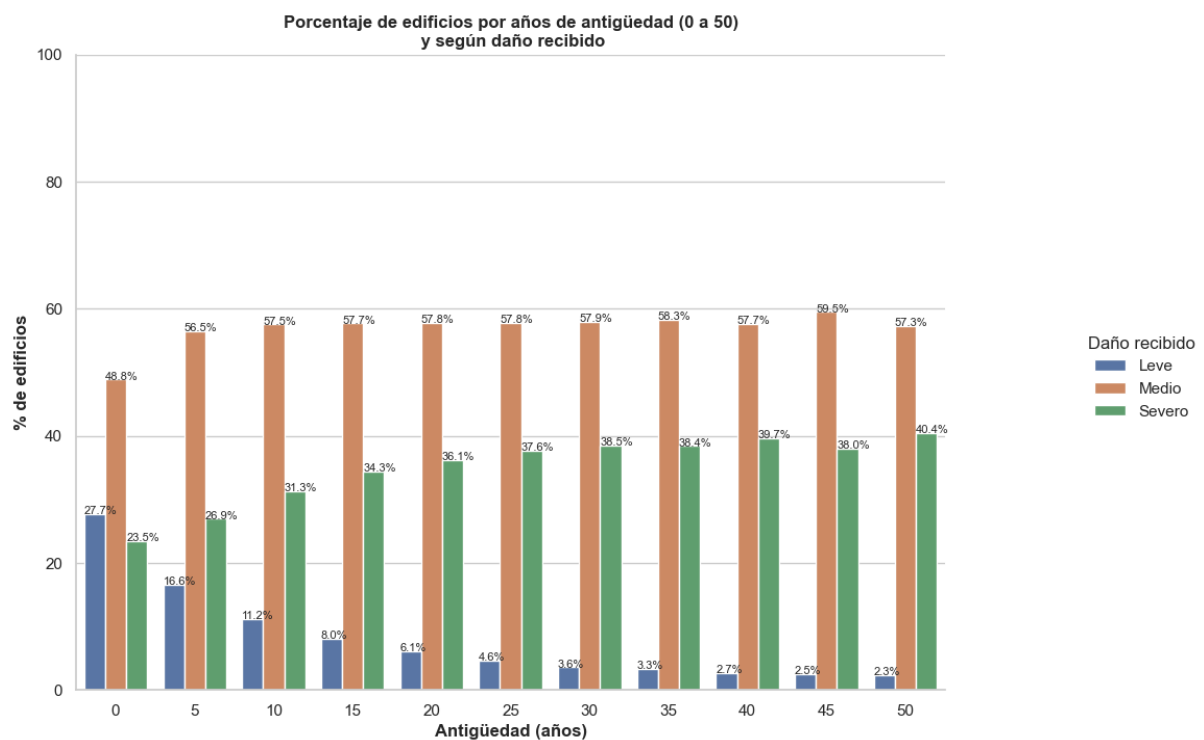


Figura 25: Porcentaje de edificios por años de antigüedad (0 a 50) y según daño recibido

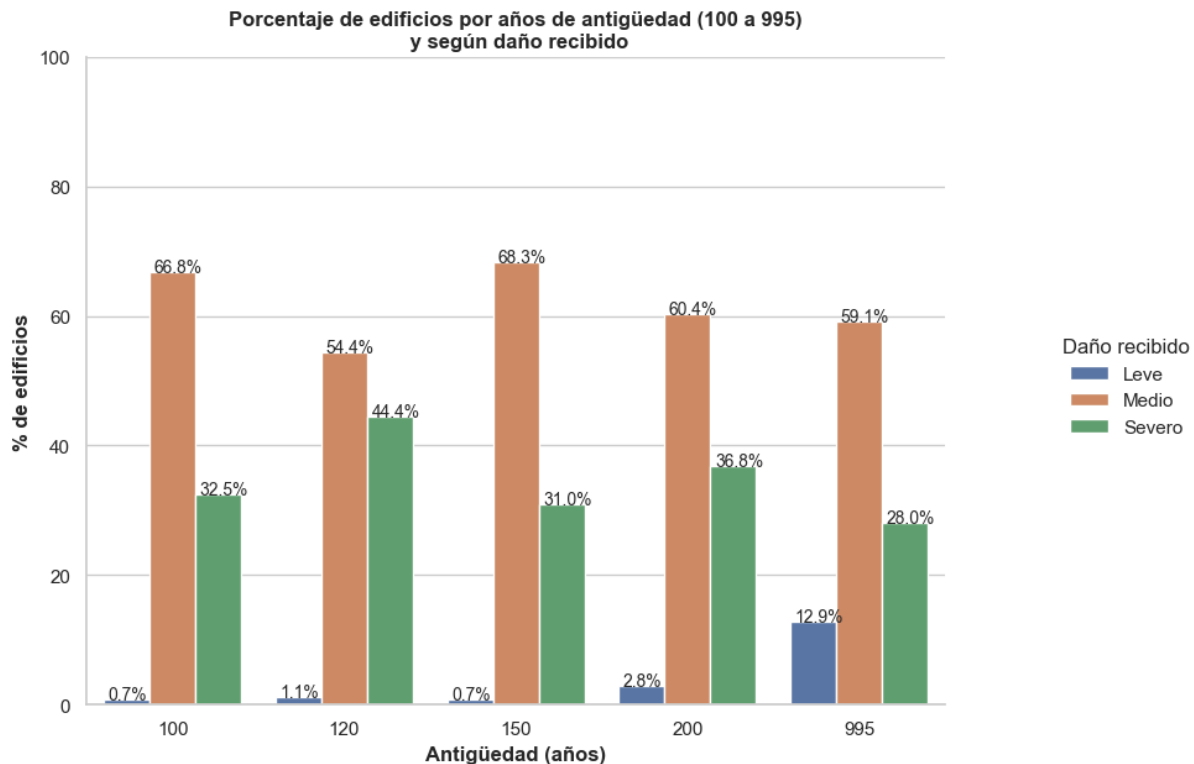


Figura 26: Porcentaje de edificios por años de antigüedad (100 a 995) y según daño recibido

Para el rango de 0-50 años de antigüedad, se puede ver que arriba del 50% de los edificios sufrieron un grado de daño de nivel medio, mientras que un porcentaje mayor al 30% sufrieron daño severo. Es interesante observar como a medida que la antigüedad va aumentando, la cantidad de edificios que sufrieron daños leves va disminuyendo al mismo tiempo que la cantidad de viviendas que sufrieron un daño grave va aumentando.

Si analizamos aquellos edificios con una antigüedad mayor a 100, podemos ver que esta tendencia se sigue manteniendo. Sin embargo, aquellos edificios con una antigüedad de 995 años tienen un comportamiento similar a los edificios con antigüedad entre 0 y 10 años. Esto puede atribuirse a que dichos edificios pueden ser históricos con una constante remodelación y/o reacondicionamiento con el fin de hacerlos más resistentes a daños sísmicos considerando que Nepal es una de las regiones sísmicas más activas del mundo⁴. Al mismo tiempo, creemos que este grupo de 995 años agrupa también edificios para los cuales se desconocen su antigüedad pero que se cree que son bastante añejos.

En general, podemos decir que los edificios con menor antigüedad sufrieron menos daño severo.

⁴ [Why Nepal is so vulnerable to quakes](#)

11.2. Analizando la altura y área para los edificios con 995 años

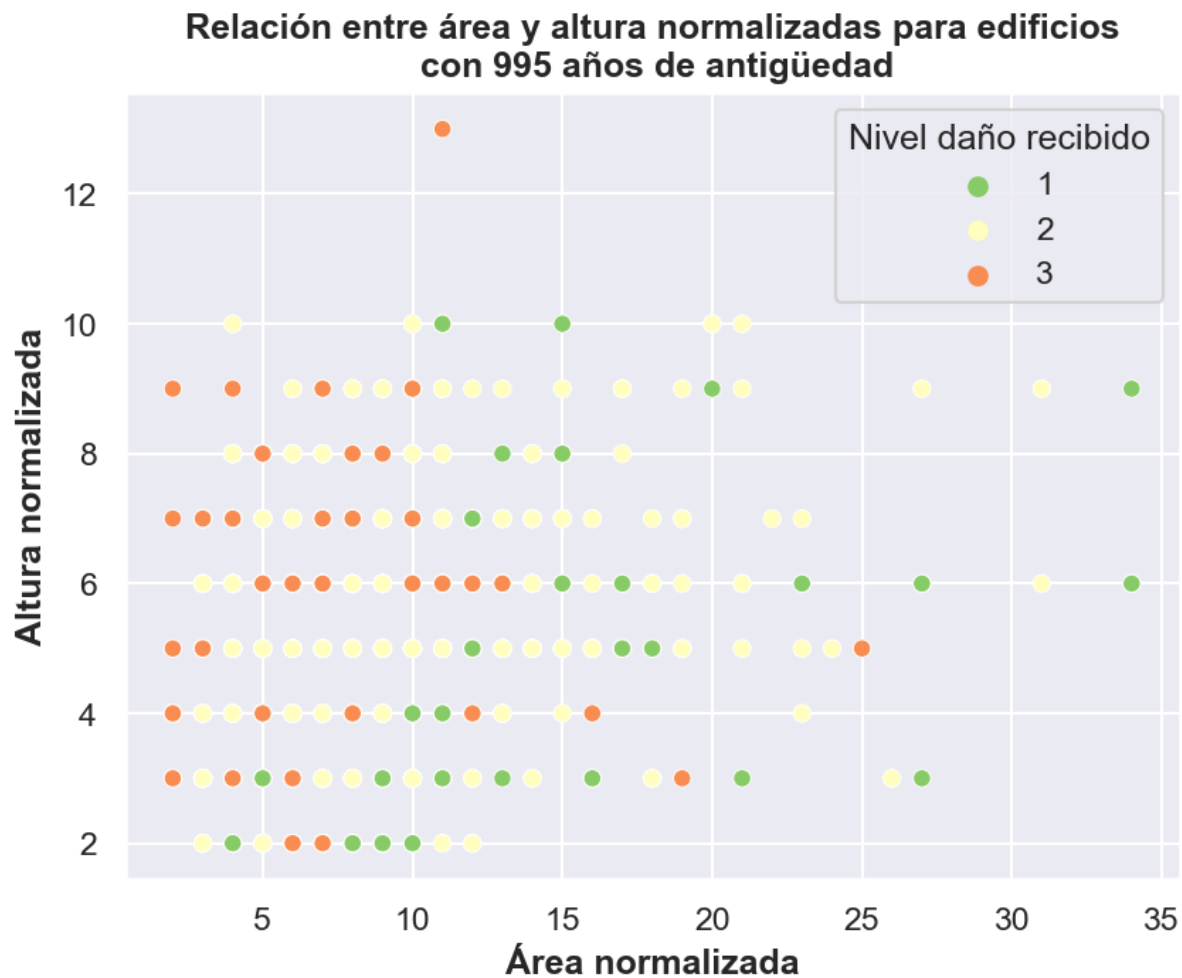


Figura 27: Relación entre área y altura normalizada para edificios con 995 años de antigüedad

Vemos que los edificios con antigüedad alrededor de 995 años que recibieron un daño severo (daño nivel 3) suelen tener un área normalizada menor a 15 y una altura normalizada menor a 10. A medida que la altura va aumentando, son escasos los edificios con un nivel de daño leve (daño nivel 1). Es decir, la altura tiene más nivel de influencia que el área para estos edificios a la hora de entender que nivel de daño recibieron.

12. Análisis por cantidad de pisos de la vivienda

12.1. ¿Los edificios con mayor cantidad de pisos sufrieron un mayor daño?

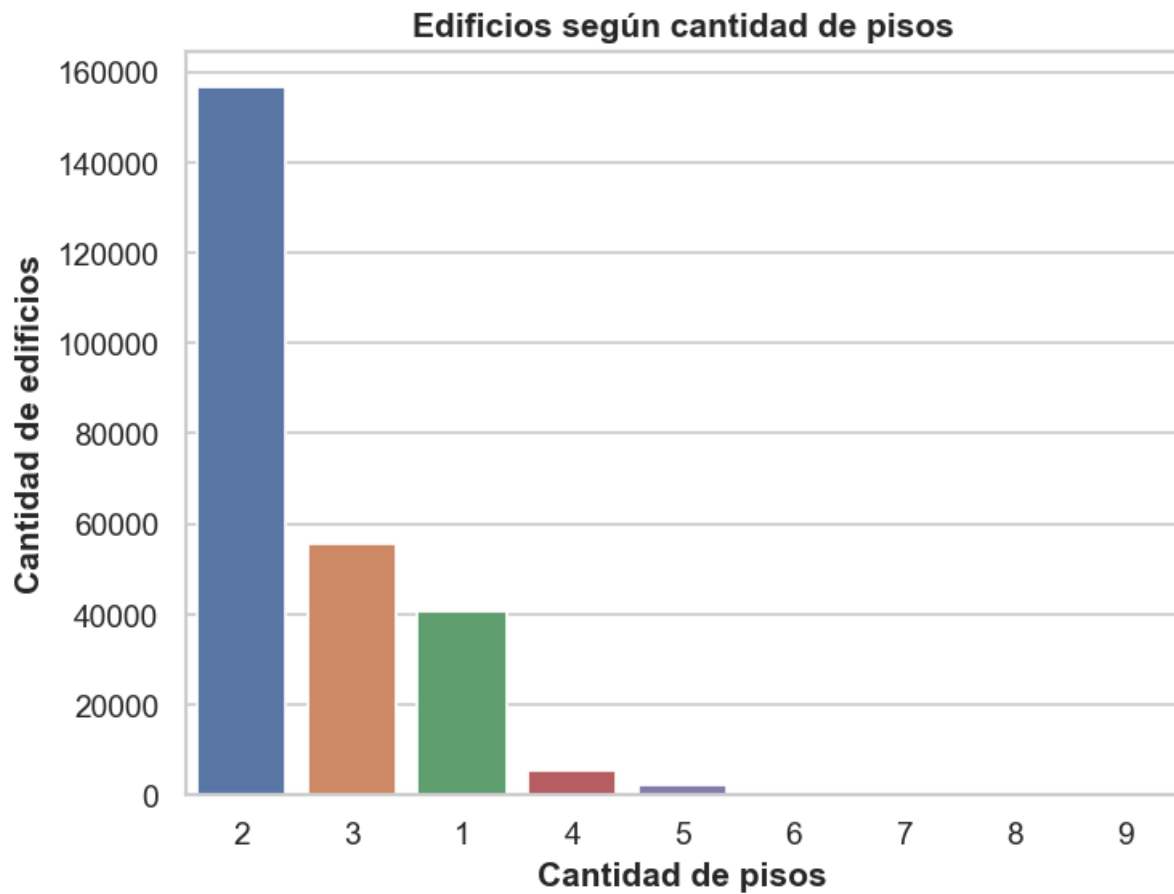


Figura 28: Edificios según cantidad de pisos

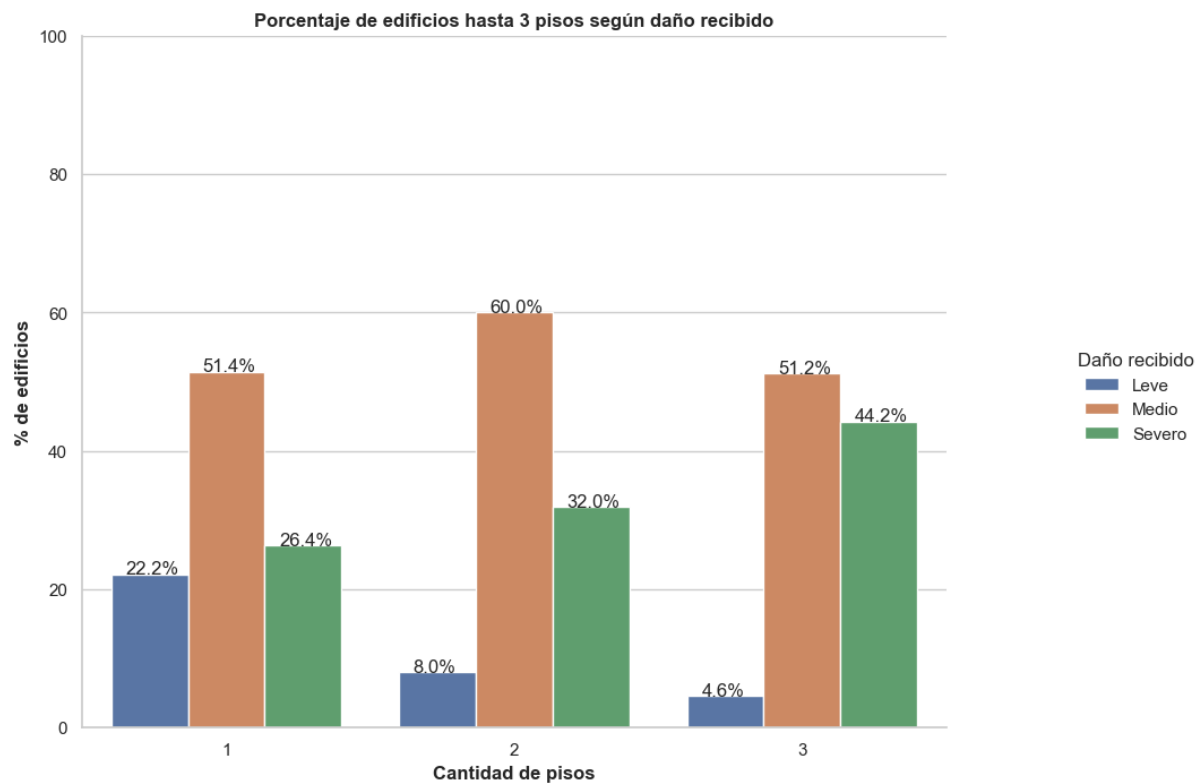


Figura 29: Porcentaje de edificios de hasta 3 pisos según daño recibido

Si solo nos quedamos con las viviendas que tienen hasta 3 pisos, se observa claramente que a mayor cantidad de pisos, mayor fue el daño recibido. Es notable ver que la diferencia del 14% entre el porcentaje de edificios de 1 y 2 pisos proviene principalmente en la disminución del porcentaje de viviendas que sufrieron un daño leve, aumentando más el porcentaje de la cantidad de edificios que sufrieron un daño medio que los que sufrieron un daño severo. Mientras que si comparamos las viviendas con 2 y 3 pisos, vemos que se disminuye tanto las viviendas con daño leve y medio para así aumentar en 12% aproximadamente las viviendas con daño severo. Todo esto puede deberse a que dada las condiciones en que se construyen las viviendas en Nepal, y viendo que predominan las viviendas fabricadas, parcial o totalmente, con barro-piedra y fabricadas con timber; sumado a que probablemente no se cumplan las condiciones mínimas necesarias para que dichas viviendas sean resistentes a un daño sísmico, el aumento de pisos en una vivienda genera condiciones para que dicha vivienda sufra daños mayores.

12.2. ¿El plan de configuración para diseño sísmico ayudó a los edificios a recibir menos daño?

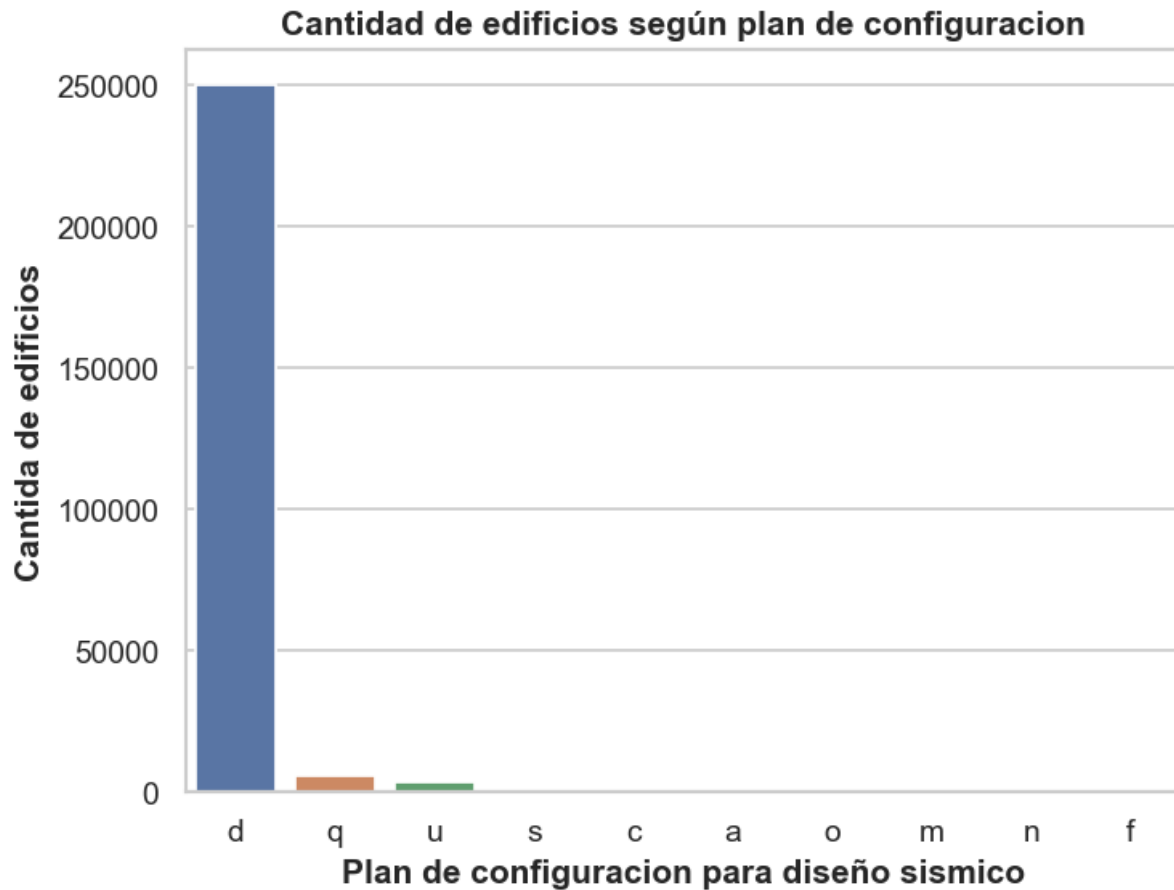


Figura 30: Cantidad de edificios según plan de configuración para diseño sísmico

Para la categoría de plan de configuración para diseño sísmico, el valor “d” es la que gobierna el dataset abarcando casi la totalidad del mismo. Al mismo tiempo, los valores “q” y “u” tienen un poco menos de 5000 registros. Para el análisis que nos compete, nos quedamos con estos tres valores para el conjunto de viviendas con 1-3 pisos.

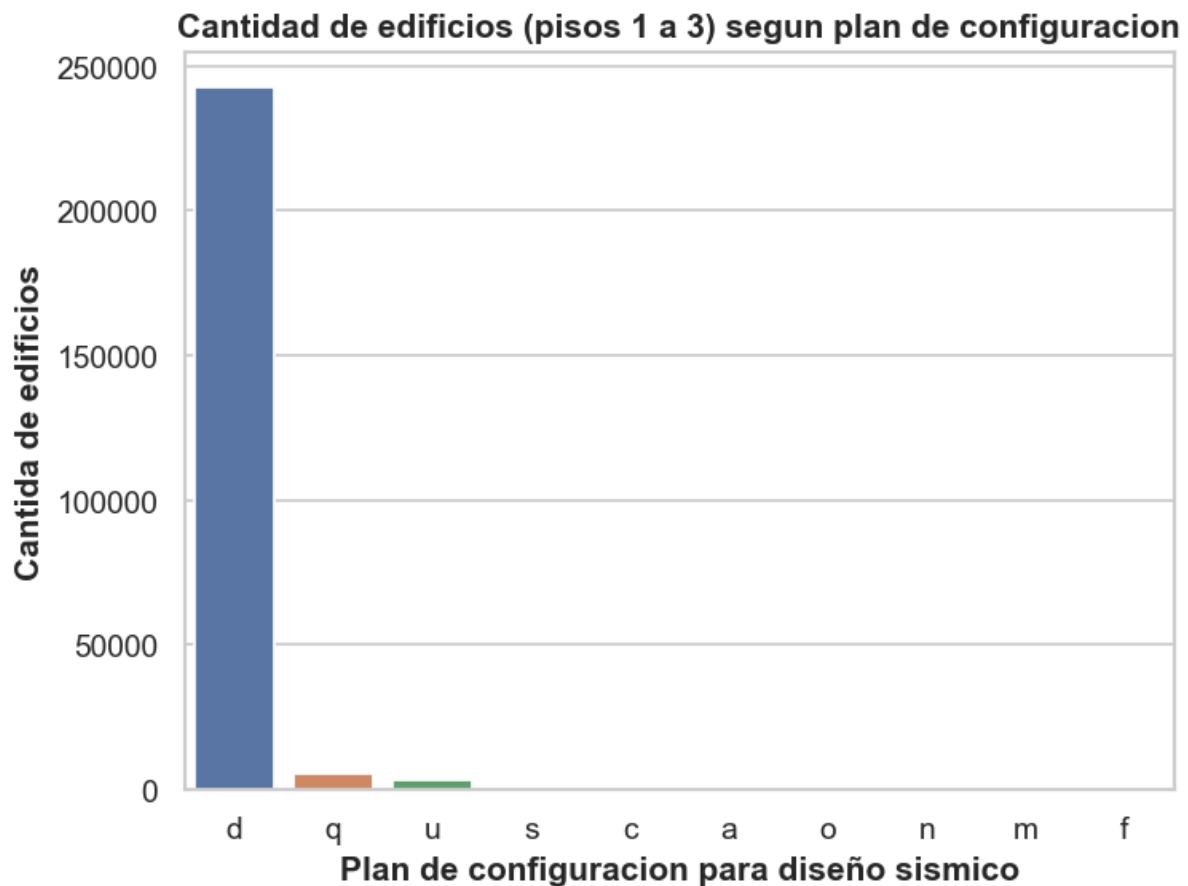


Figura 31: Cantidad de edificios con pisos entre 1 y 3 según plan de configuración para diseño sísmico

Como presuponíamos, para las viviendas con una cantidad de pisos entre 1 y 3, vemos una distribución a la figura 31 dándonos como resultado que el plan de configuración para diseño sísmico que más predomina es del tipo “d”. Entendemos que esta categoría es aquella en donde las viviendas tienen nulo o muy poco diseño para resistir daños sísmicos.

A continuación, nos quedamos con los valores “d-q-u” dentro del conjunto de viviendas que tienen entre 1-3 pisos.

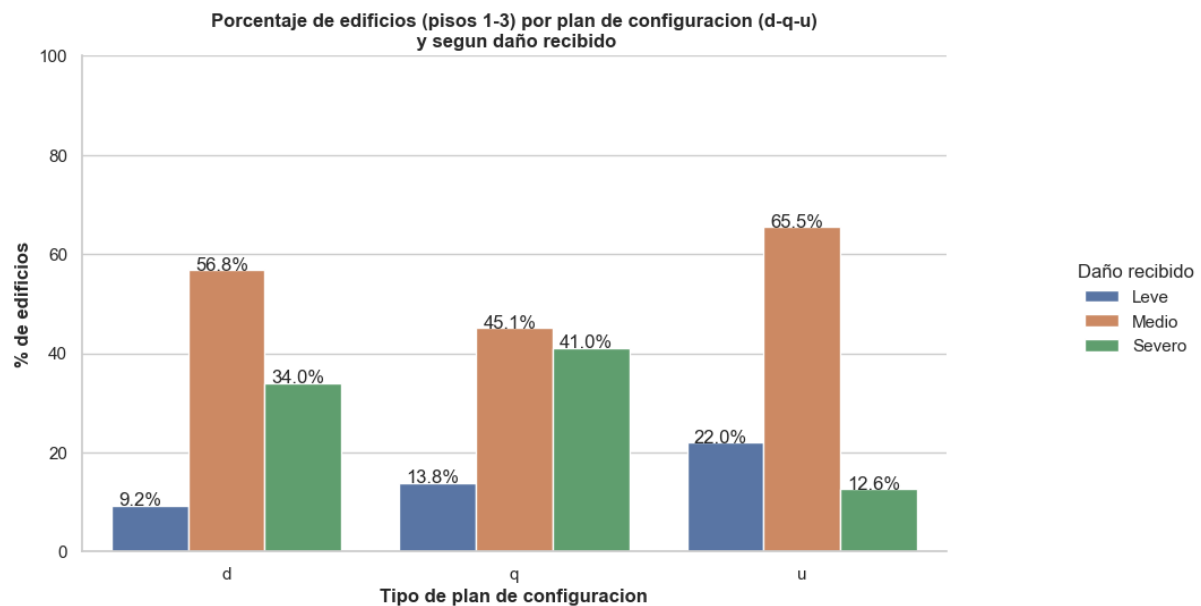


Figura 32: Porcentaje de edificios con pisos entre 1 y 3 para tipo de planes de configuración para diseño sísmico d, q y u según daño recibido

Por cuestiones ya mencionadas, predomina el daño medio para estos planes de configuraciones. No obstante, si nos centramos únicamente en los daños leve y severo nos encontramos con la información que la categoría “u” es aquella que tiene el mayor porcentaje de viviendas con daño leve y el menor porcentaje de viviendas con daño severo. Es probable que esta categoría sea aquella que nos indica que las viviendas tienen algún tipo de diseño para daños sísmicos (mayor nivel que la categoría “d”).

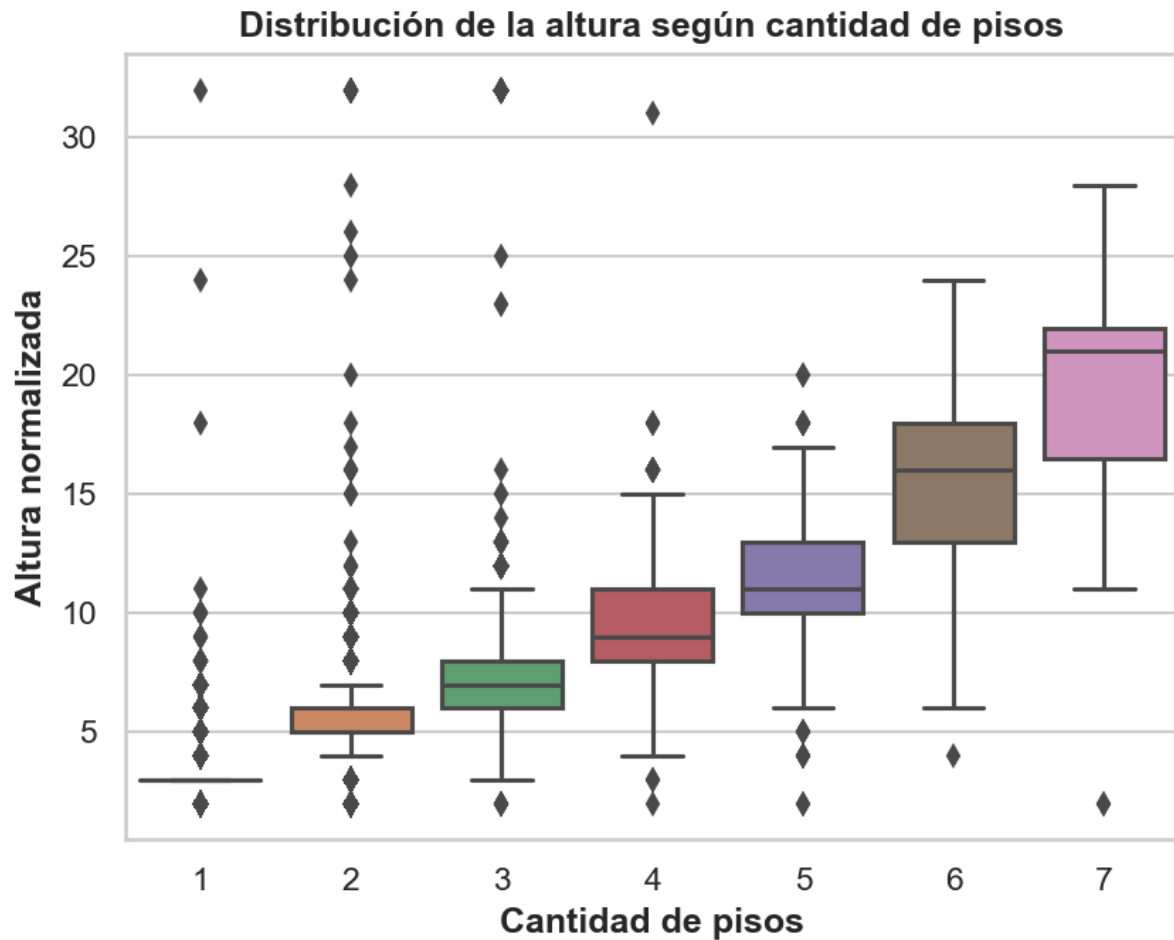


Figura 33: Distribución de la altura normalizada según la cantidad de pisos

Como pensábamos, a mayor cantidad de pisos mayor cantidad de altura normalizada. Una observación es que no se superponen los cuantiles dadas las distintas cantidad de pisos salvo para los casos de 4-5 y 6-7, haciéndonos notar que para edificios con una cantidad de pisos elevada, el cambio de altura no es tan brusco. Sin embargo, desconocemos cómo fue el cálculo de la altura pero entendemos que se normalizó para no tener una cantidad tan dispar de valores posibles para la columna de `height_percentage`.

13. Análisis por tipo de construcción usado en sus otros pisos

En esta sección analizaremos los distintos tipos de construcciones usados para los pisos que no fueran planta baja.

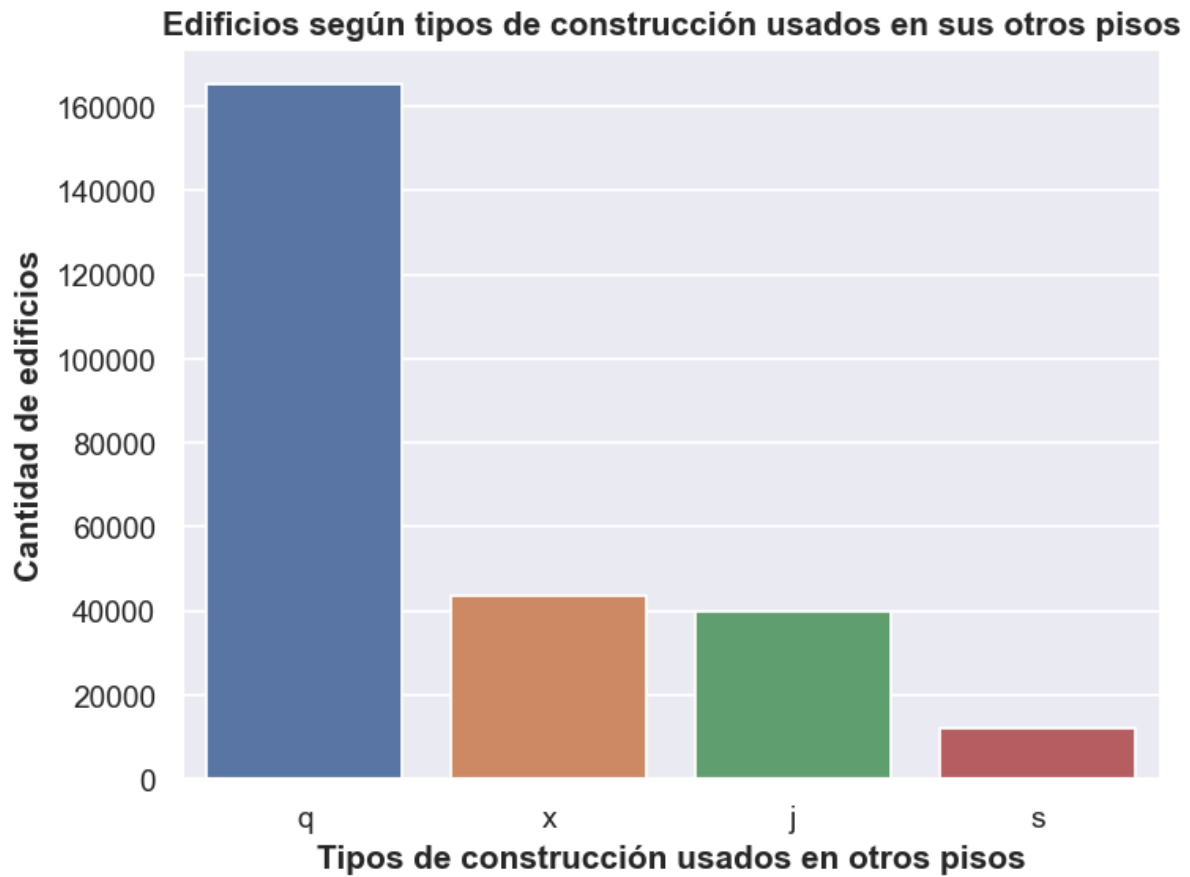


Figura 34: Cantidad de edificios según tipos de construcción usados en sus otros pisos

Se puede observar, que la gran mayoría de los edificios fueron construidos con el tipo “q”.

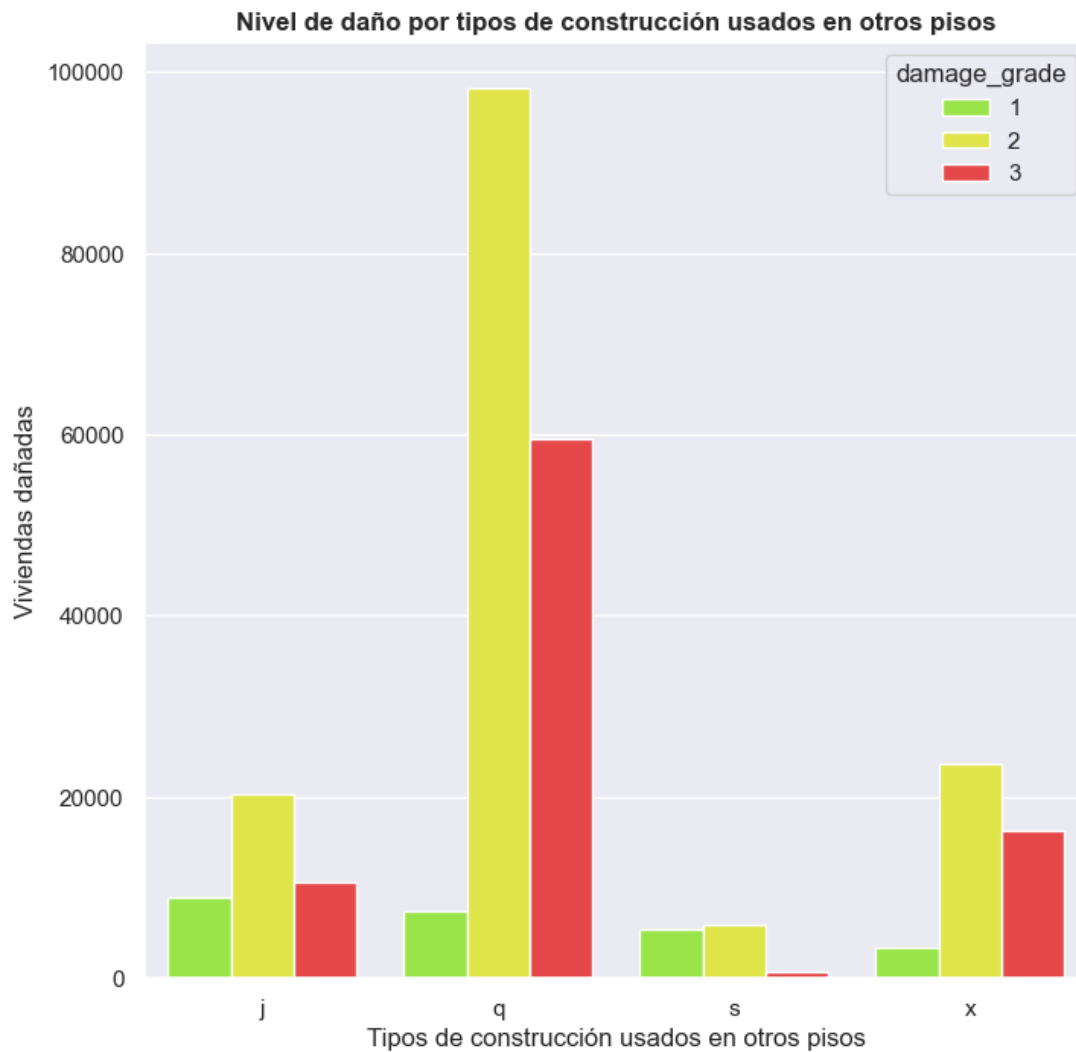


Figura 35: Cantidad de viviendas dañadas según el tipo de construcción usado en sus otros pisos

Dado que la mayoría de los edificios fueron contruidos de tipo “t” para los pisos superiores, no sorprende que este sea el más afectado por el terremoto.

14. Análisis por tipo de techo

En esta sección analizaremos los distintos tipos de techo que se encontraron en la zona afectada y buscaremos si hubo algún tipo más propenso a recibir mayores daños.

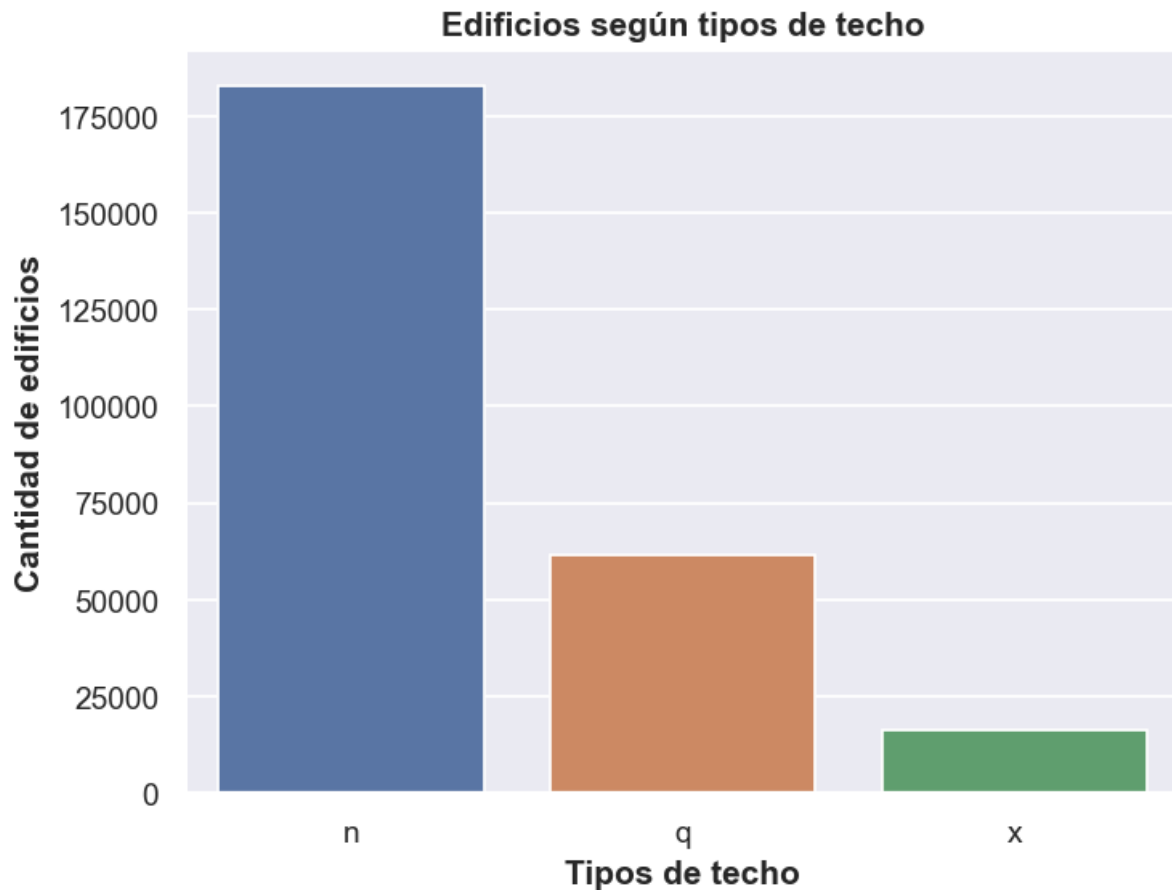


Figura 36: Cantidad de edificios según los tipos de techos

A partir de este gráfico podemos observar que la mayoría de los techos encuestados eran de tipo “n”, seguidos de los de tipo “q” y por último de tipo “x”. Debido a que no contamos con referencias acerca de estas letras, no podemos hacer un análisis más exhaustivo acerca del por qué de este fenómeno.

A continuación veremos el nivel de daño de cada tipo:

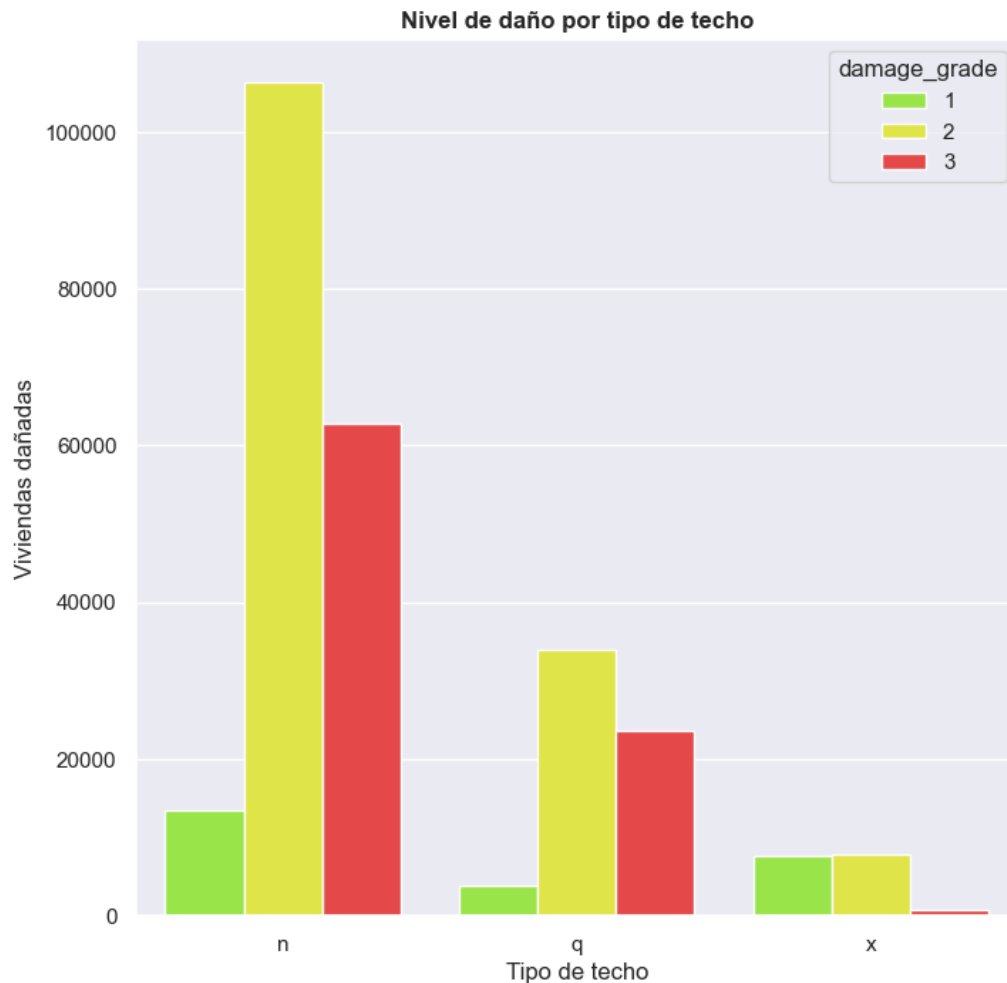


Figura 37: Cantidad y nivel de viviendas dañadas según tipo de techo

Podemos observar que las viviendas con tipo de techo “n” fueron las más afectadas, seguidas de las de tipo “q” y luego las de tipo “x”. Teniendo en cuenta el gráfico anterior esto no nos sorprende, puesto que es concordante que el tipo de techo más habitual encontrado en la encuesta, sea el que más se repita en las edificaciones más afectadas, así como sucedía en el análisis de los tipos de construcciones de los pisos superiores (distintos de los de planta baja).

15. Análisis de cantidad de familias por vivienda

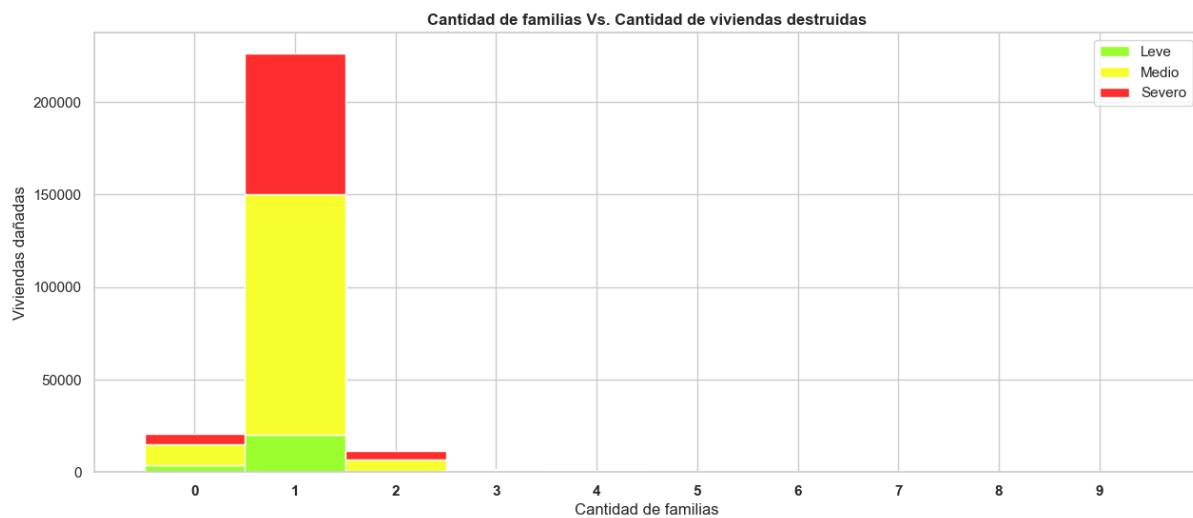


Figura 38: Cantidad de viviendas dañadas según la cantidad de familias habitando

Las viviendas donde se registran un grupo familiar fueron las que sufrieron más daño por amplia diferencia, tanto de nivel medio como de nivel severo. Le siguen aquellas donde se registran dos familias o ninguna, pero con más de un 90% de diferencia con el grupo mencionado anteriormente. Esto debe ser porque dichas propiedades (las de 0 grupo familiar) fueron usadas para otro propósito.

En cuanto a nuestra hipótesis inicial, es decir, si existe alguna relación entre la cantidad de viviendas dañadas y la cantidad de familias viviendo en ellas, podemos concluir que no hay tal relación. Los datos de las encuestas no aportan al análisis de la problemática sino que solo nos da información sobre la distribución poblacional la cual consideramos no tiene relación con los daños causados por el sismo.

16. Análisis del estado legal de la propiedad

En el siguiente análisis, veremos si el estado legal de la propiedad tuvo alguna incidencia en el nivel de daño sufrido por la propiedad debido al sismo.

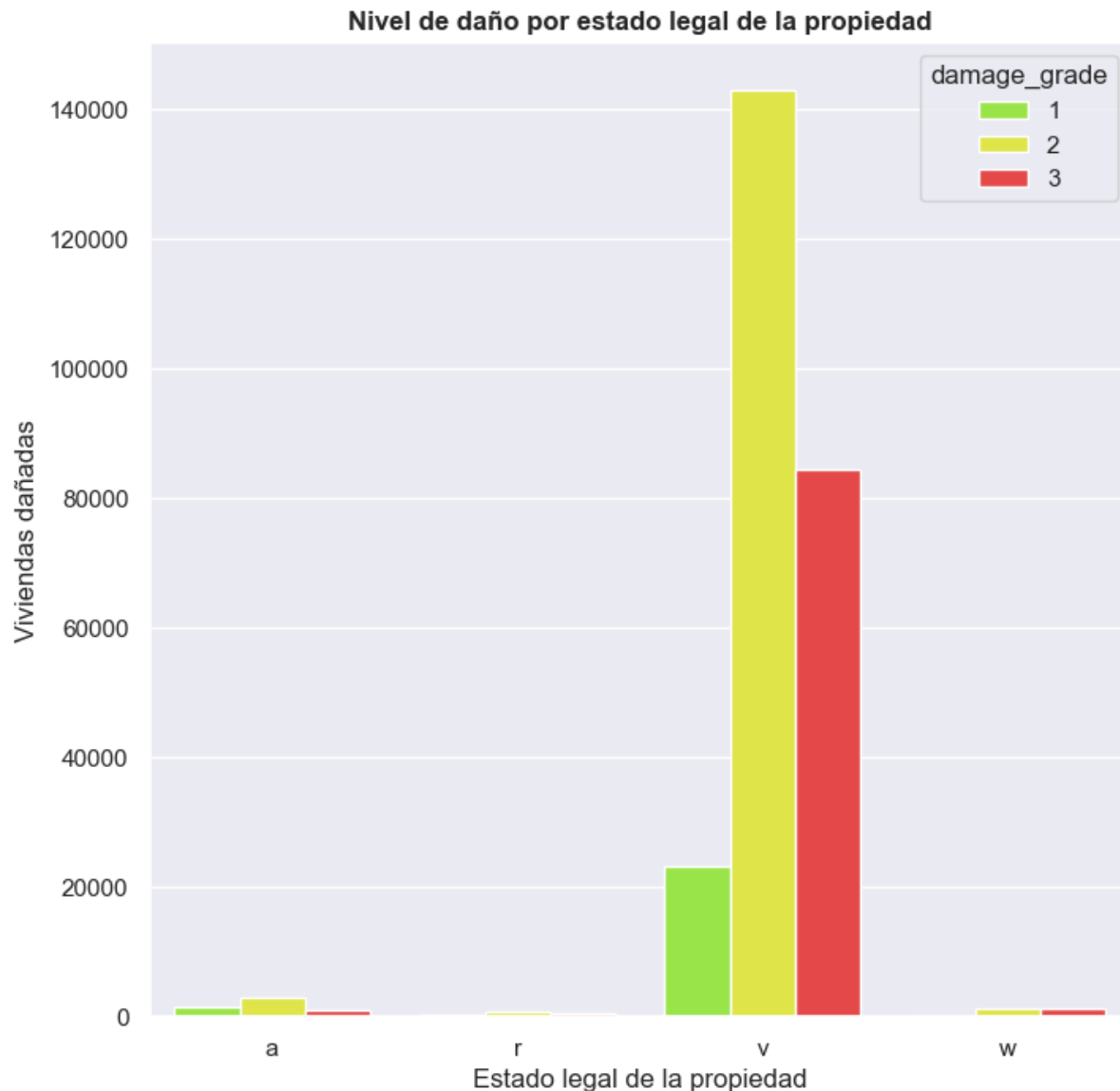


Figura 39: Cantidad y nivel de viviendas dañadas según su estado legal

Aparentemente, se concentran la mayor cantidad de viviendas afectadas (en los tres niveles) en las propiedades cuyo estado legal se informa como “v”. Debido a que no contamos con más información acerca del significado de los tipos de estado legal de la propiedad, no podemos conjeturar acerca del por qué de este fenómeno.

17. Análisis por usos secundarios de los edificios

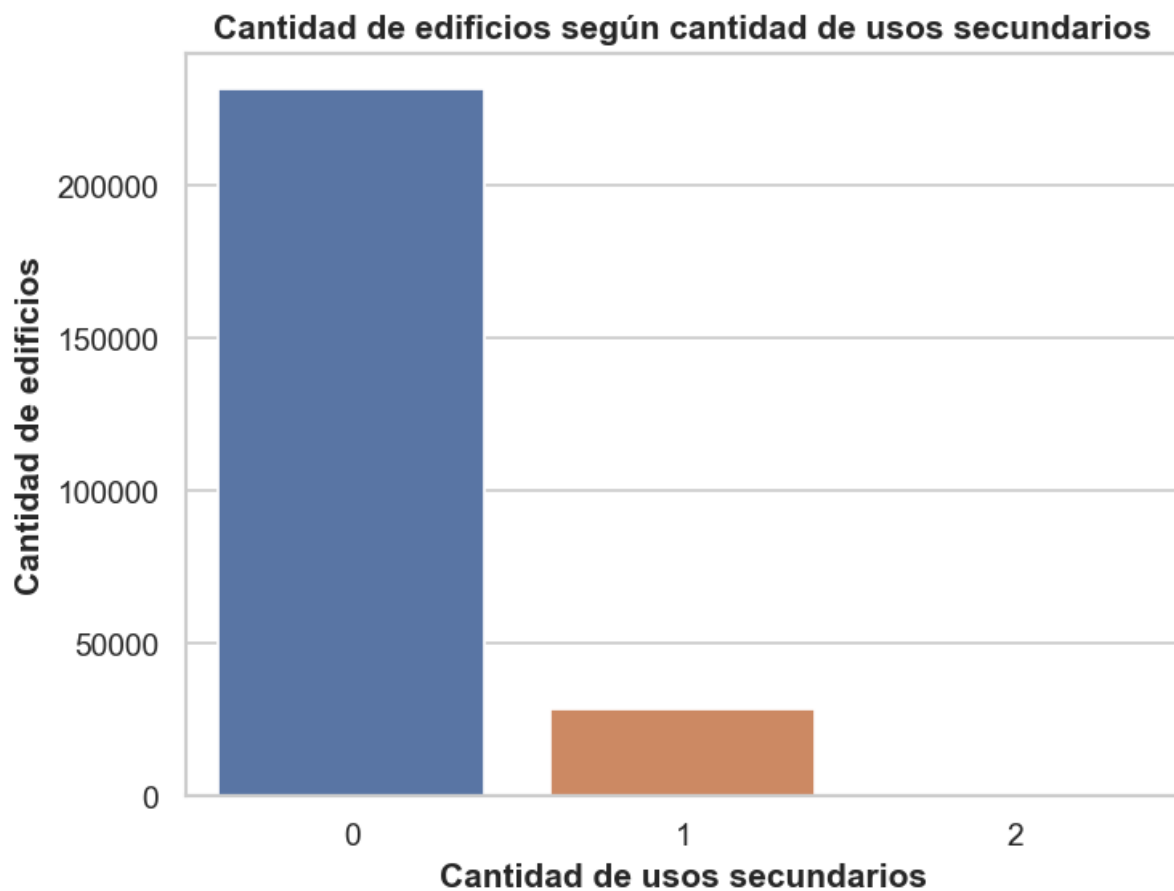


Figura 40: Cantidad de edificios según cantidad de usos secundarios

La gran mayoría de los edificios de este dataset aparentemente son viviendas, mientras que alrededor de unas 25.000 edificaciones tienen algún uso secundario en particular. Vamos a ver cuáles son esos usos secundarios y ver si teniendo alguna relación con el gobierno recibieron menos daños implicando que poseen un mejor tipo de estructura al tener mayor presupuesto disponible.

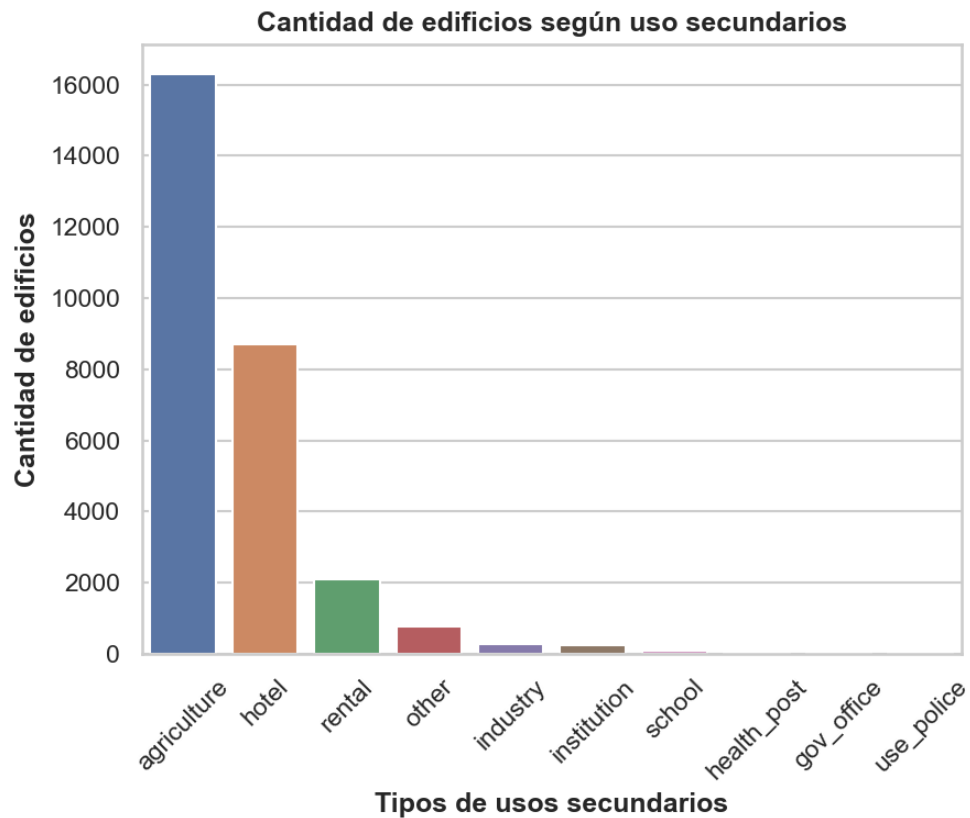


Figura 41: Cantidad de edificios según tipos de uso secundario

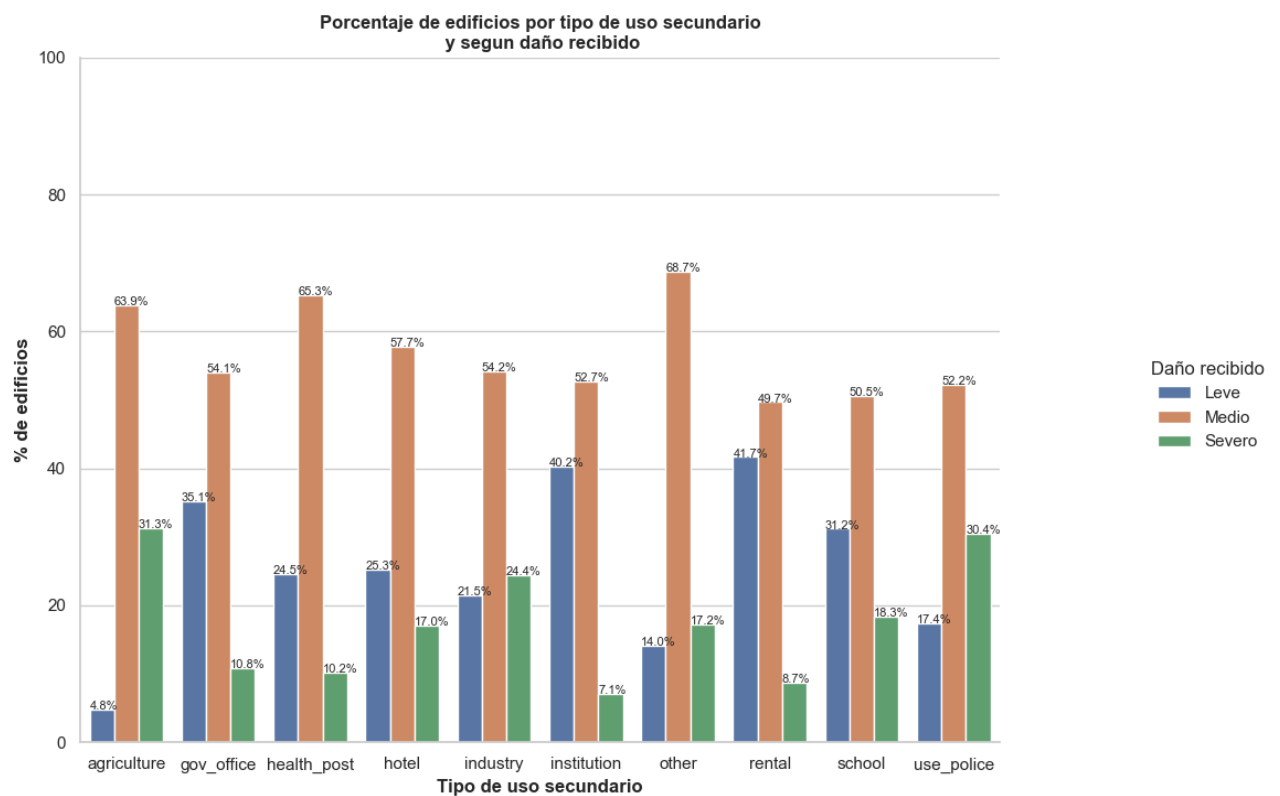


Figura 42: Porcentaje de edificios por tipo de uso secundario y según daño recibido

Ninguno de los tipos lograron tener un porcentaje de edificios con daños leves mayor a los otros dos tipos de daños, siendo que predomina el daño medio en todas las categorías. Es interesante ver como los tipos de oficina de gobierno, edificios relacionados con la salud, instituciones, escuelas y edificios alquilados tienen un alto porcentaje de edificios que recibieron un daño leve; esto puede estar relacionado con que de alguna manera (ya sea de parte del Estado o de forma privada) tienen un mayor presupuesto que pudo ser destinado a tener una mejor y más resistente estructura base.

18. Conclusiones

Antes de presentar las conclusiones obtenidas a partir del análisis del dataset presentado es necesario remarcar que los datos representan solo el **34% de la totalidad de los datos**⁵ de la encuesta realizada por Kathmandu Living Labs y Central Bureau of Statistics.

Se descubrió que la gran mayoría de las viviendas recibieron un **daño medio (nivel 2)**, y teniendo en cuenta que las viviendas de este dataset en su mayoría fueron fabricadas con barro-piedra, y sumado a que el epicentro fue en Katmandú (capital de Nepal) esperaríamos que la cantidad de viviendas con daño severo (nivel 3) fuera aún mayor a la presentada.

Debido a que las zonas correspondientes a los id's 6, 10, 17 y 26 del `geo_level_1` fueron las que acumularon la mayor cantidad de viviendas afectadas suponemos que estos abarcan a la capital de Nepal y sus alrededores.

Tener la posibilidad de tener más datos (ya sea de la totalidad de una región o regiones y sin truncamientos) podría hacer un análisis más rico, y si le agregamos la chance de tener más atributos no categóricos sería mucho mejor. Una de las dificultades que se nos presentó fue la enorme cantidad de features categóricos (ya sea de tipo booleano como si fuera One-Hot encodeado; o categórico ofuscado) con lo que en muchas ocasiones solamente agrupamos y contamos para observar un panorama general de los datos. De todas formas, realizamos el análisis de estos tipos categóricos ofuscados obteniendo cuáles fueron los de mayor o menor ocurrencia y a partir de ello concluir cuáles son los que inciden más o menos en el daño de las viviendas. Por otro lado, pudimos ahondar un poco más a la hora de encontrar cuáles tipos de estructuras son más o menos resistentes al momento de episodios sísmicos.

Al realizar un análisis exploratorio, es usual que las empresas brinden los datos de manera ofuscada para proteger la información por lo que el análisis se realiza de manera abstracta, como en este caso. Luego, dichas empresas se encargarán de interpretar el significado de los resultados obtenidos.

18.1. Insights

En primer lugar, en la figura 40 notamos que el dataset está compuesto básicamente por viviendas, con muy pocas edificaciones con alguna función en particular aparte. Para el grupo de viviendas encontramos varias causas para la insuficiente resistencia ante terremotos entre los que se incluyen:

- Elección inadecuada de materiales de construcción.
- Elección inadecuada de mortero (barro o cemento).
- Tipo de cimientos no adecuados.
- Falta de mantenimiento de estructuras viejas.

⁵ [2015 Nepal Earthquake: Open Data Portal](#)

- Mala utilización de los materiales elegidos en la unión de techos, plantas intermedias y planta baja.
- Poco conocimiento del comportamiento estructural de las viviendas.

La mayoría de las viviendas que fueron afectadas por el terremoto fueron aquellas en las que en sus materiales se encuentra alguna combinación de barro-piedra. Por otro lado, las estructuras que presentaron cemento-ladrillo o algún tipo de concreto reforzado en su composición, fueron las que mejor actuaron frente al terremoto. Entendemos que la gran cantidad de viviendas con barro-piedra, bambú o timber provienen de la población que se va mudando de zonas rurales a zonas urbanas, permitiéndoles construir un hogar en base a sus necesidades y al presupuesto disponible en el momento.

En segundo lugar, en la figura 3 analizamos las condiciones de las superficie terrestres en donde se edificaron las viviendas. Si sumamos lo obtenido en la figura 12, podemos decir que las condiciones de terreno en Nepal son muy diversas, y que esto puede generar que el diseño y los materiales puedan variar bastante entre viviendas, razón por la cual no se puede mantener un estándar a la hora de diseñar un plan resistente a terremotos que se adecue a todas las viviendas. De hecho, existe un diseño para terremotos desde 1994⁶, pero sólo se hace cumplir en ciertos municipios. Si tenemos en cuenta las figuras 25 y 26 vemos que los daños severos se van acrecentando a medida que la antigüedad de la vivienda es mayor; y esto está sumamente relacionado con que son viviendas construidas antes de 1994.

Por último, para reducir los daños en posibles terremotos las viviendas deben estar diseñadas por personal capacitado y especializado en este tipo de acontecimientos, los cuales deben cumplir con el código de diseño para terremotos. El control de calidad tanto al momento de construcción, como de revisión de viviendas ya fabricadas debe tomarse en serio para reducir paulatinamente el volumen de viviendas que no están adecuadas.

⁶ [NEPAL NATIONAL BUILDING CODE GUIDELINES FOR EARTHQUAKE RESISTANT BUILDING CONSTRUCTION: EARTHEN BUILD](#)