

TP1 - Análisis Exploratorio

[7506] Organización de Datos

Curso 1

Primer cuatrimestre de 2021

Alumno	Padrón	Email
González Matías Ignacio	104913	maigonzalez@fi.uba.ar
Vázquez Lareu Román	100815	rlareu@fi.uba.ar
Figueroa Matias	92498	maafigueroa@fi.uba.ar

Índice

1. Introducción	2
2. Datos	2
2.1. Train Values	2
2.2. Train Labels	2
2.3. Estado de datos	2
3. Conteos	2
3.1. Cantidad de familias afectadas	2
3.2. Escuelas y centros de salud afectados	4
3.3. Edificios Históricos	5
3.4. Construcciones con usos secundarios	5
3.5. Grado de daño de acuerdo al tipo de techo	8
3.6. Grado de daño de acuerdo al tipo de construccion	9
3.7. Grado de daño de acuerdo a la cantidad de pisos	11
3.8. Grado de daño de acuerdo a los cimientos	12
3.9. Rc engineered y cimientos i	12
3.10. Grado de daño de acuerdo a la configuración del plano	13
4. Hipótesis	13
4.1. Los edificios más viejos y mas altos sufrieron más daños	13
4.2. Mientras mayor sea el cociente altura/área mayor será el daño	14
4.3. Las estructuras de concreto son menos propensas a daños de grado 3	15
4.4. Hay zonas especificas que reciben mas daño grave	17
4.5. Las estructuras de barro son mas propensas al daño	18
4.6. Las edificaciones de Bamboo de mas de 2 pisos son mas inestables y reciben mucho daño	20
4.7. El terremoto afecta mas a las edificaciones antiguas	21
5. Conclusiones	22
6. Links	23

1. Introducción

En el año 2015 Nepal fue afectado por el terremoto Gorkha, un sismo que registró una magnitud de 7.8 en la escala Richter y tuvo su epicentro en la ciudad de Kathmandu. Aproximadamente 600,000 estructuras en el centro y pueblos aledaños fueron dañadas o destruidas. Un análisis posterior al sismo llevado por la Comisión Nacional de Planeamiento de Nepal comunicó que la pérdida total económica ocasionada por el terremoto fue de aproximadamente \$7 mil millones (USD; NPC, 2015).

El presente informe reúne la documentación del análisis exploratorio sobre los datos sobre el terremoto.

2. Datos

Se cuenta con dos archivos: **Train values** y **Train labels**. En el primero se encuentran datos sobre la estructura identificada por su id. En el segundo encontramos datos sobre el grado de daño para cada estructura.

2.1. Train Values

Hay 38 columnas de distintas features de cada edificación. Entre ellas encontramos columnas de tipo entero: Geo ID, pisos, antigüedad, area normalizada, altura normalizada y cantidad de familias. También columnas de tipo categorico: condicion del terreno, tipo de cimientos, tipo de techo, tipo de piso, orientacion, tipo de diseño sismico, estado legal del terreno. Por ultimo hay columnas booleanas que indican materiales y usos del edificio.

2.2. Train Labels

En este archivo se encuentra el grado de daño para cada edificio.

2.3. Estado de datos

Se realizo un analisis sobre los datos en donde se observo que estan completos y no hay datos faltantes.

Se observó también una anomalía con respecto a la antigüedad de algunas edificaciones: existe una brecha entre 200 y 995 años. Se asumió para el desarrollo del trabajo práctico que 995 años representa edificios muy antiguos que no se sabe con certeza su antigüedad.

Los datos entonces se separaron en edificios con antigüedad de hasta 200 años y edificios con antigüedad de 995 años.

3. Conteos

A continuación se detallan diferentes recuentos o totales sobre diferentes parámetros. A partir de estos, luego se generaron hipótesis para la siguiente sección.

3.1. Cantidad de familias afectadas

Una porcion de edificios era hogar de un número de familias. Se muestra a continuacion la distribucion de familias afectadas.

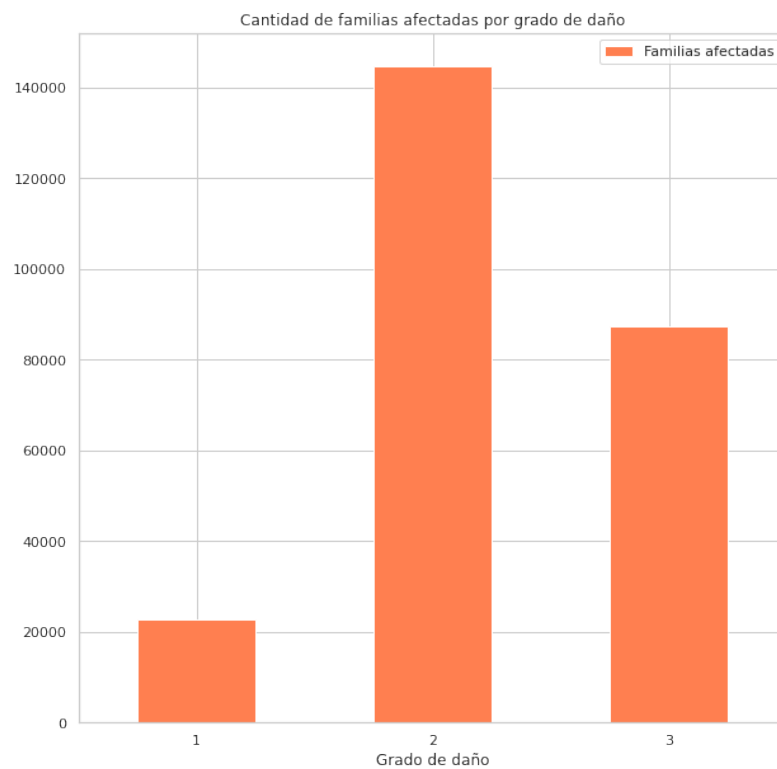


Figura 1: Familias Afectadas

El total de familias afectadas por el terremoto según los datos fue de 256.418, de las cuales mas de 140.000 sufrieron un daño de grado 2 y mas de 80.000 sufrieron un daño de grado 3. Solo una minoría de un poco mas de un 10 % sufrió un daño de grado 1.

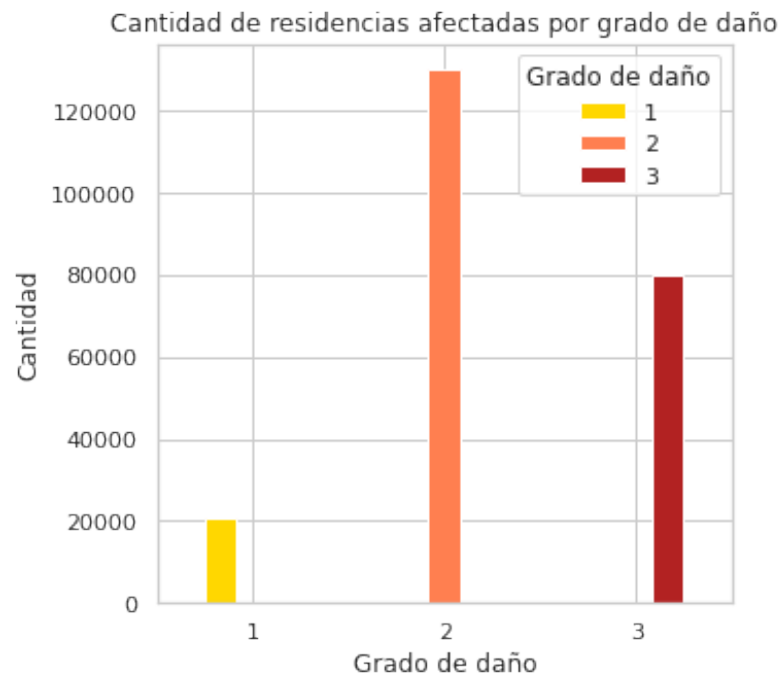


Figura 2: Residencias Afectadas

3.2. Escuelas y centros de salud afectados

Se muestra a continuación la distribución de escuelas y centros de salud afectados.

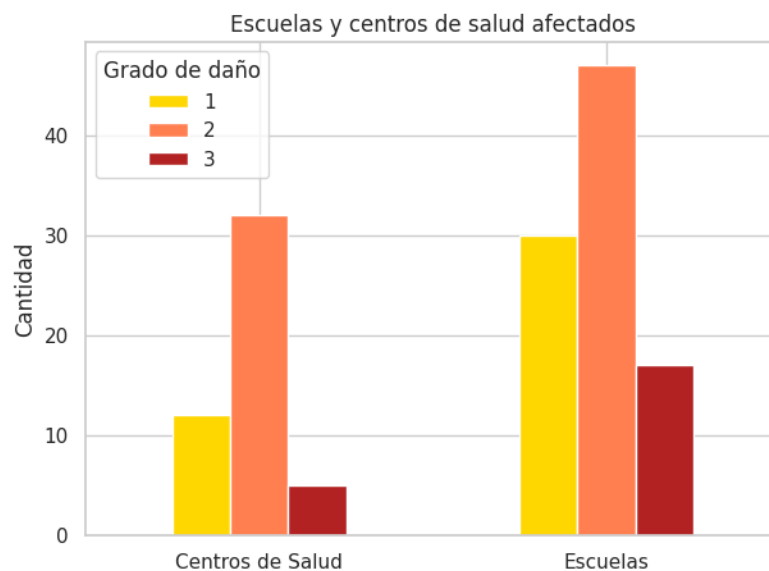


Figura 3: Escuelas y centros de salud

Se puede observar que la distribución del grado de daño en los edificios usados para educación y salud es diferente a la de edificios ocupados principalmente para vivienda. En los primeros, la mayoría de los grados de daños está en 1 o 2 mientras que para los de vivienda está en 2 y 3. Esto puede ser por los materiales, diseño o dimensiones de la edificación.

3.3. Edificios Históricos

Consideramos edificios históricos a los cuales cuya antigüedad en los datos es de 995 años.

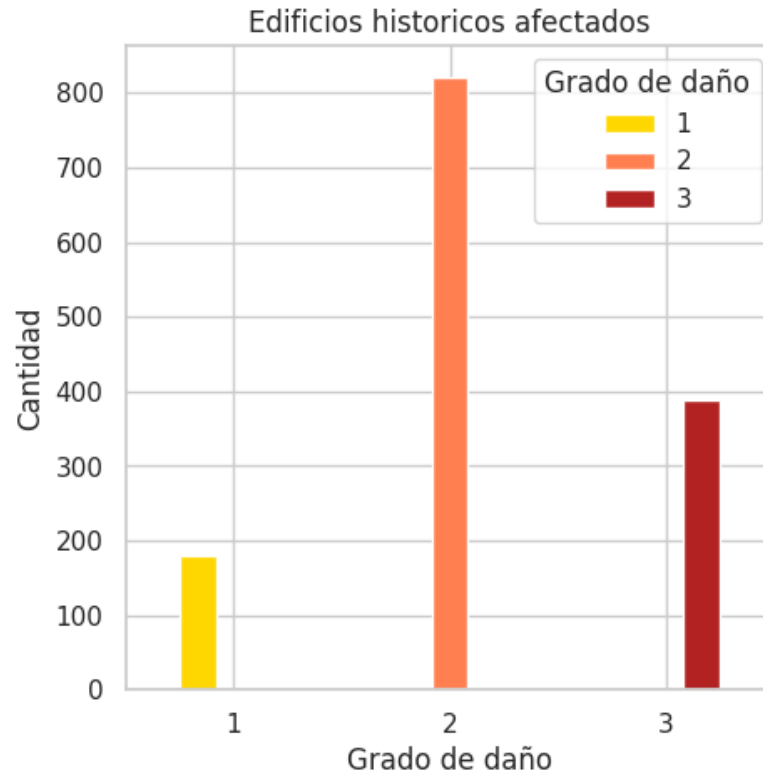


Figura 4: Grado de daño de edificios históricos

Estos edificios presentan una distribución de daño similar a la de las residencias.

3.4. Construcciones con usos secundarios

Se observa que en su mayoría, las edificaciones afectadas no poseían un uso secundario

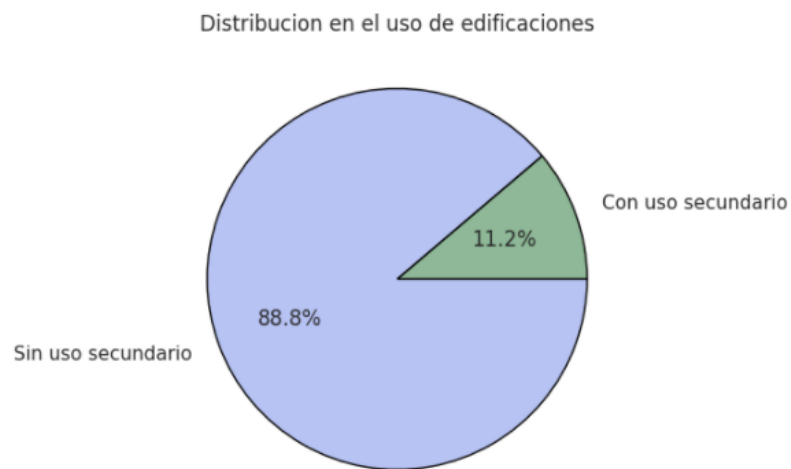


Figura 5: Construcciones con usos secundarios

De las construcciones con uso secundario, la distribucion es la siguiente:

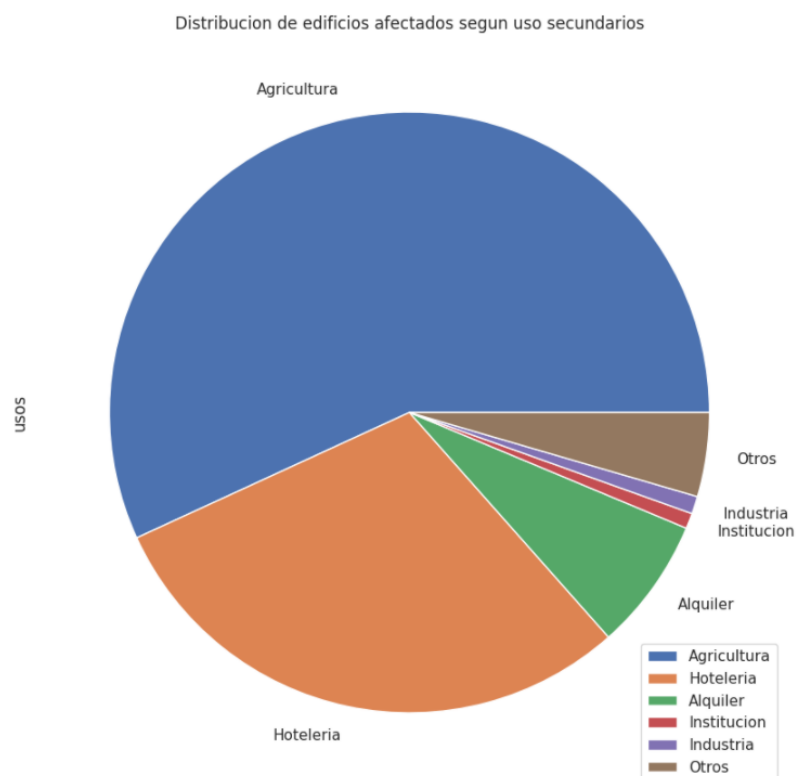


Figura 6: Distribucion de construcciones con uso secundario

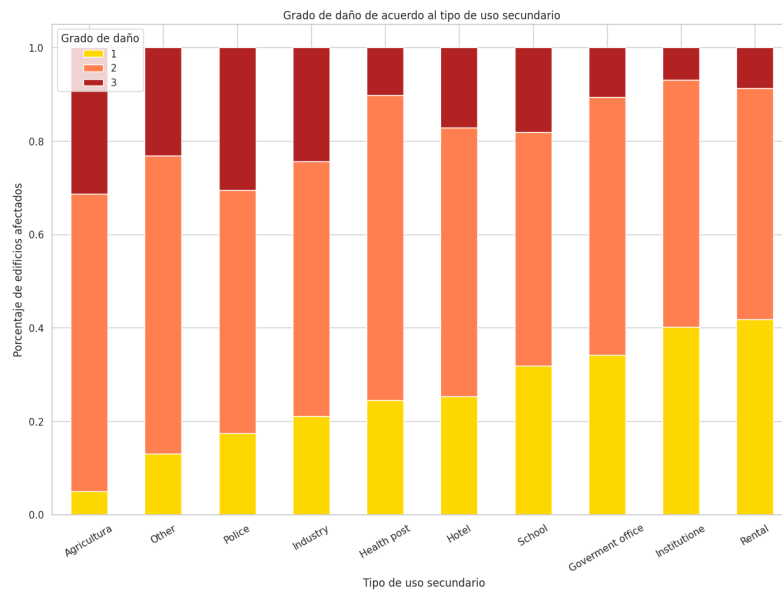


Figura 7: Grado de daño por tipo de uso.

Se observa que en cuanto a daños de **grado 1**, las edificaciones dedicadas a Agricultura fueron las menos afectadas, mientras que con casi un 0.4, las de renta e instituciones fueron las mas golpeadas por este grado de daño. Si bien varia levemente, los daños de **grado 2** representaron un porcentaje similar en las distintas categorias. Se ve a Policia y Colegios como los menos afectados por este tipo de daño. Finalmente, ya si mas marcado, las edificaciones dedicadas a agricultura y Policia son las que presentan un mayor porcentaje de daño de **grado 3**. En conclusion, las de renta e instituciones fueron las que mejor soportaron el terremoto.

3.5. Grado de daño de acuerdo al tipo de techo

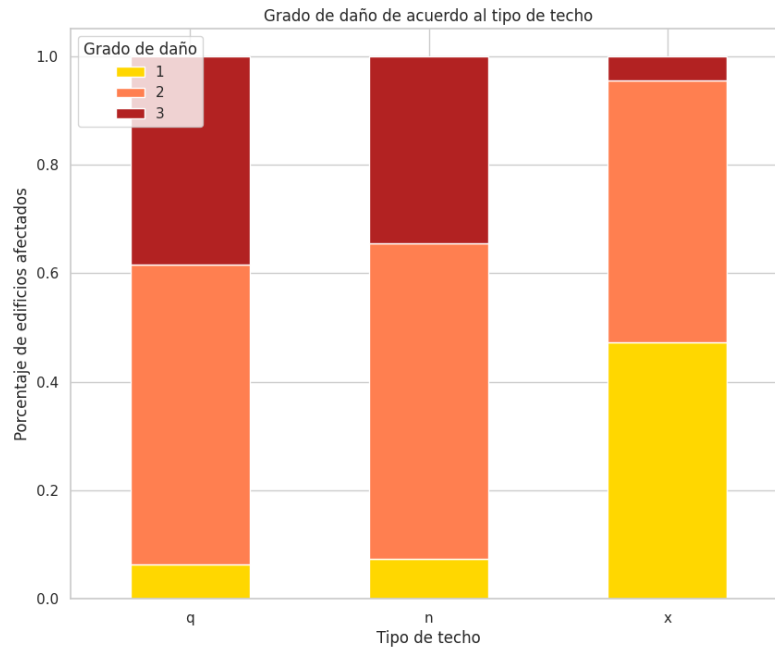


Figura 8: Grado de daño de acuerdo al tipo de techo.

Se desprende del gráfico que las edificaciones con un tipo de uso "**n**" o "**q**" sufrieron de manera similar tanto daños de **grado 1** como de **grado 3**, siendo en su mayoría severamente afectadas por el terremoto. Por el contrario, las edificaciones con tipo de techo "**x**" sufrieron de igual manera daños de **grado 1 y 2**; únicamente un 0.05 de ellas sufrieron daños de **grado 3**. De esta manera se concluye que soportaron de mejor manera el terremoto las edificaciones de tipo de techo "**x**"

3.6. Grado de daño de acuerdo al tipo de construcción

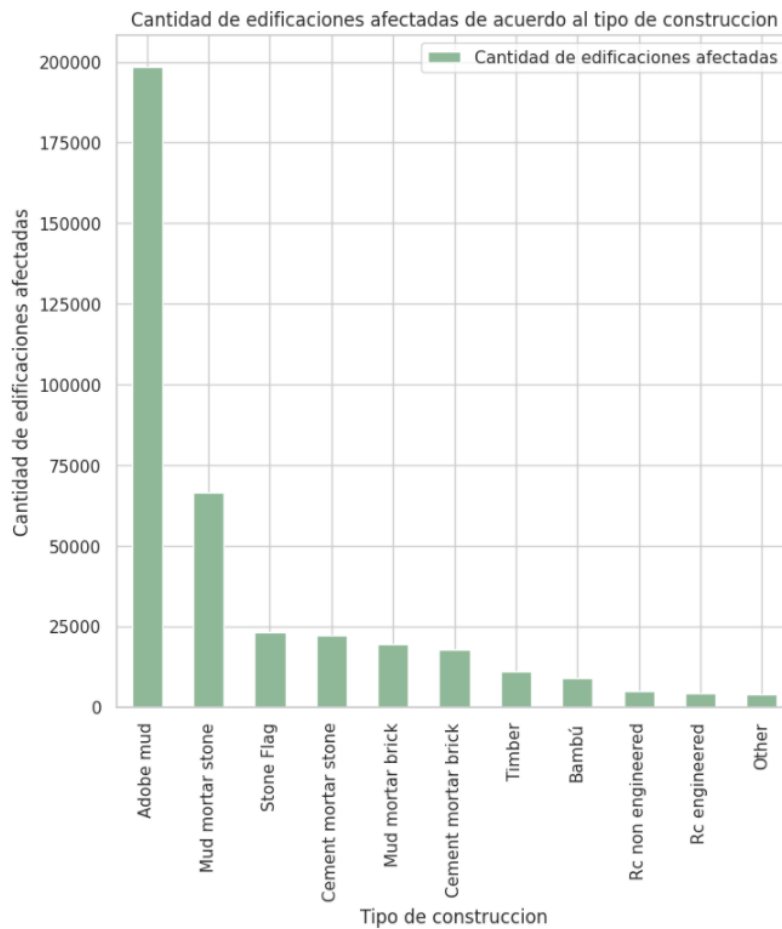


Figura 9: Cantidad de edificios dañados de cada tipo de construcción.

En su mayoría las construcciones son de adobe o presentan al mismo como material primario de la edificación. Se cree que las construcciones de adobe sufriran en mayor medida las consecuencias del terremoto que, por ejemplo, las RC engineered".

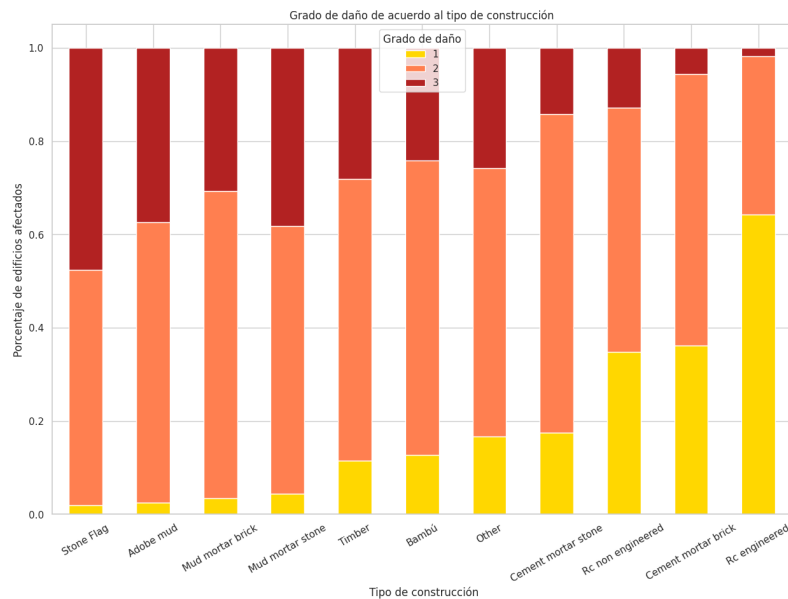


Figura 10: Grado de daño de acuerdo al tipo de construcción

Del gráfico se interpreta que las edificaciones cuyo porcentaje de daño de **grado 3** fue mayor fueron las que poseían una superestructura de "Stone Flag", junto con Adobe Mudz "Mud mortar stone", seguidas por las de "Mud mortar Brick". Se podría decir que las superestructuras de tipo adobe no resistieron de buena manera a los daños de **grado 3**. Muy claramente se ve que las edificaciones que mejor resistieron fueron las de estructura Rc engineered": prácticamente sin daños de **grado 3** y con mas del 0.6 con daños de **grado 1**

3.7. Grado de daño de acuerdo a la cantidad de pisos

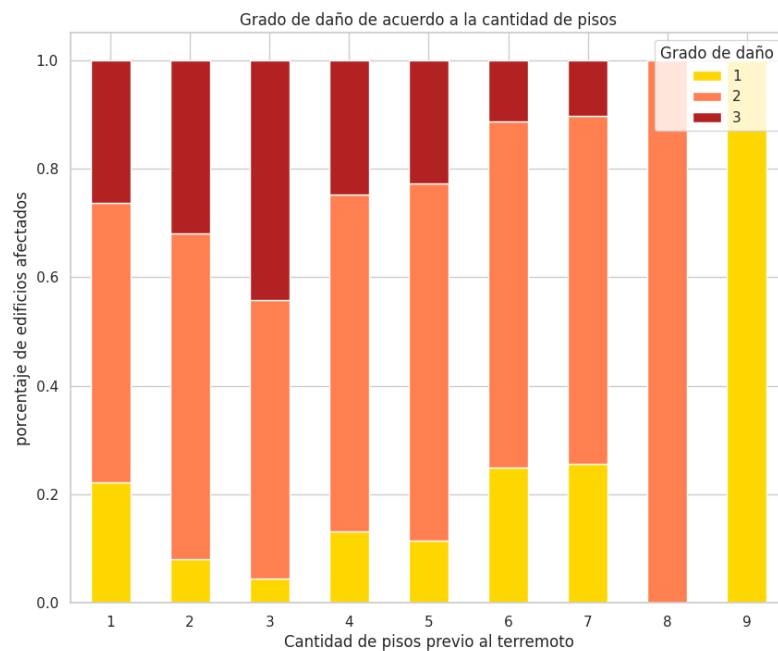


Figura 11: Grado de daño de acuerdo a la cantidad de pisos

Podría interpretarse que los edificios de pocos pisos, debido a su baja altura, sufriran menos el terremoto; mientras que se supone que edificios mas altos ya presentan una estructura capaz de resistir en mejor medida los terremotos. Claramente las edificaciones que mas sufrieron el terremoto fueron las de 3 pisos

3.8. Grado de daño de acuerdo a los cimientos

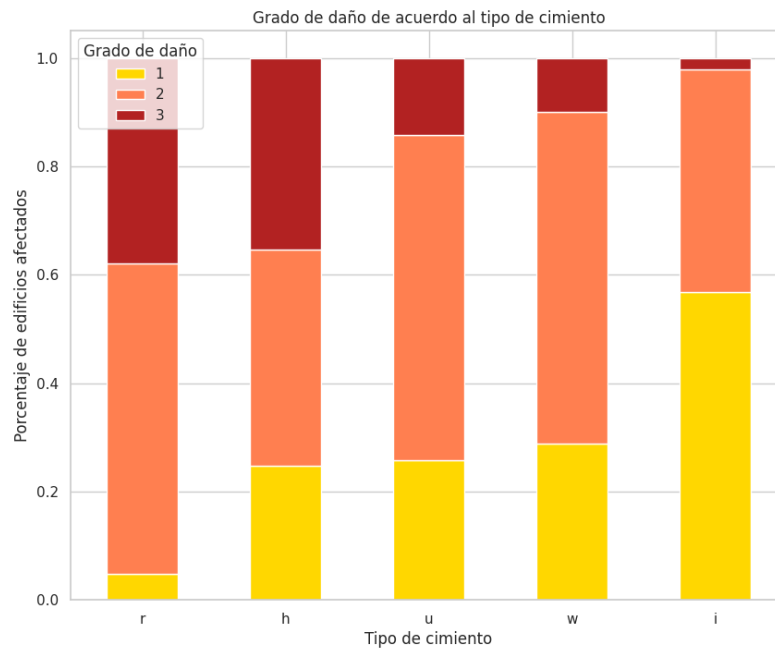


Figura 12: Grado de daño de acuerdo a los cimientos

Si bien no se tiene conocimiento de que representa cada tipo de cimiento, es claro que el que mejor resistió el impacto del terremoto fue el de tipo **i**. De la misma manera, el que peor desempeño tuvo fue la de tipo **r**

3.9. Rc engineered y cimientos i

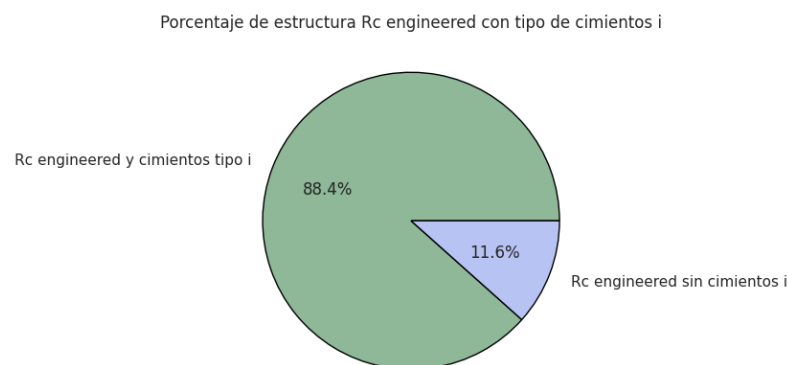


Figura 13: Porcentaje de Rc engineered con cimientos i

A partir de lo anteriormente visto, puede resultar interesante sumar al análisis que, en general, el tipo de estructura que mejor resistió que fue la Rc engineered posee en su mayoría un tipo de cimientos **i**, que también resulta ser la configuración de piso que mejor desempeño tuvo en el terremoto.

3.10. Grado de daño de acuerdo a la configuración del plano

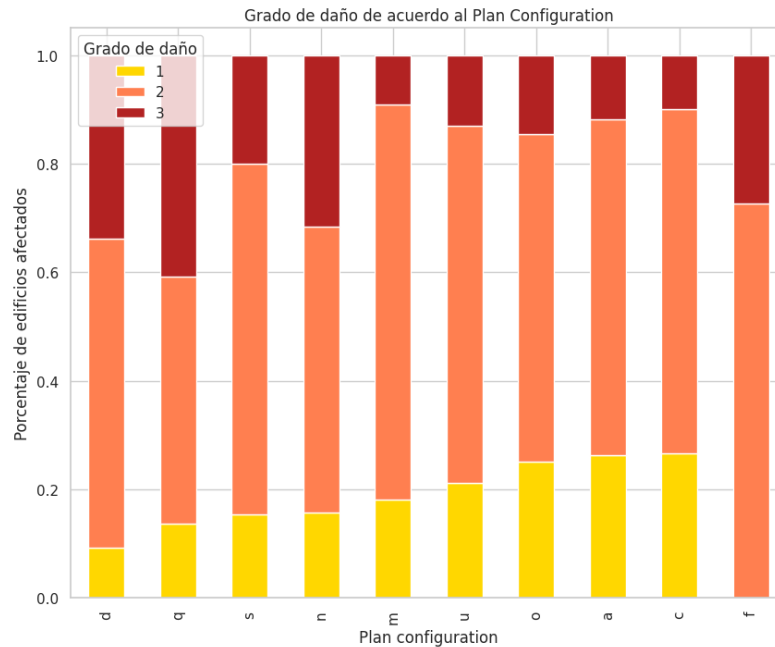


Figura 14: Grado de daño de acuerdo a la configuración del plano

Este gráfico puede resultar interesante ya que el 'Plan configuration' apunta al diseño sísmico de la estructura. No se puede concluir que se hayan obtenido resultados claramente satisfactorios con algún tipo de configuración, pero sí que los tipos 'o', 'a' y 'c' tuvieron resultados levemente mejores al resto. Por el contrario, los de tipo 'q' tuvieron claramente un daño mayor al resto de las edificaciones.

4. Hipótesis

A continuación se presentan las diferentes hipótesis planteadas en la exploración de los datos.

4.1. Los edificios más viejos y más altos sufrieron más daños

Para este análisis se filtraron los datos: usando edificios de hasta 200 años de antigüedad y de hasta 8 pisos. Para edificios que no cumplen estas condiciones no habían los suficientes datos.

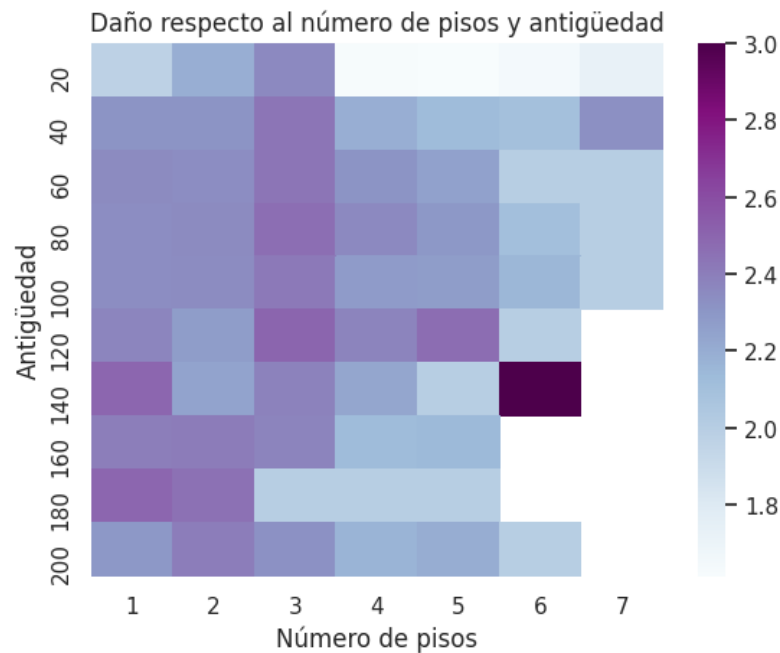


Figura 15: Los edificios más viejos y mas altos sufrieron más daños.

Del heatmap no se puede concluir en que la hipótesis es verdadera. Si se puede observar que la mayoría de edificios con una antigüedad de menos de 20 años sufrieron en promedio menos daño que edificios más viejos.

Se observa un pico en edificios de entre 120 y 140 años con una altura de 6 pisos. Esto es probablemente un resultado de haber caído en la trampa de *la ecuación más peligrosa del mundo*.

4.2. Mientras mayor sea el cociente altura/área mayor será el daño

Se calcula el cociente altura/área de cada edificio y se realiza un scatter plot para obtener la línea de regresión con respecto al grado de daño.

Se excluyen los edificios cuyo cociente altura/área sea mayor a 4 ya que no hay una cantidad importante de ellos.

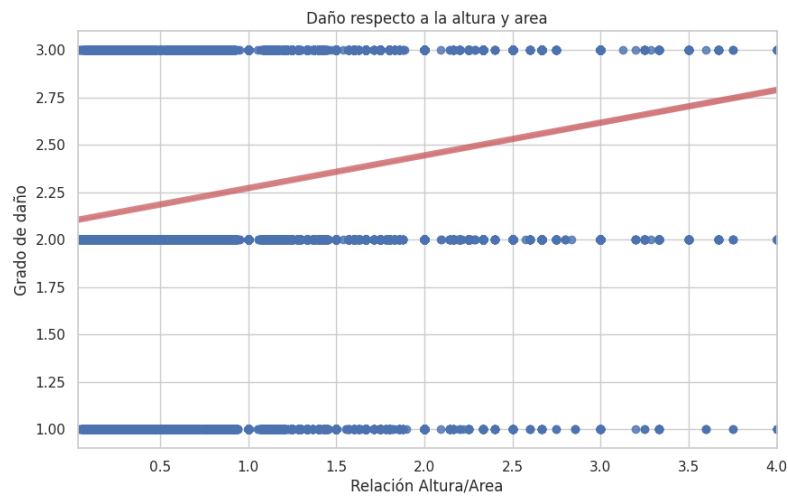


Figura 16: Línea de regresión daño por altura/area.

El gráfico obtenido concuerda con la hipótesis inicial. Tiene sentido físico ya que el centro de masa aumenta su altura y al disminuir el área ocupada hace que el edificio oscile en mayor medida. Concluimos que mientras mayor sea el cociente altura/area, mayor será el daño.

4.3. Las estructuras de concreto son menos propensas a daños de grado 3

Se filtro las edificaciones que estan hechas con concreto, y se agrupo por grado de daño. Obtuvimos que solo un 9,8 porciento de las edificaciones sufrieron daños de grado 3 (fig 17). Cuando vemos que el resto de las estructuras tienen un promedio de 34,9 porciento de daño grado 3. Podemos concluir que la hipotesis es cierta.

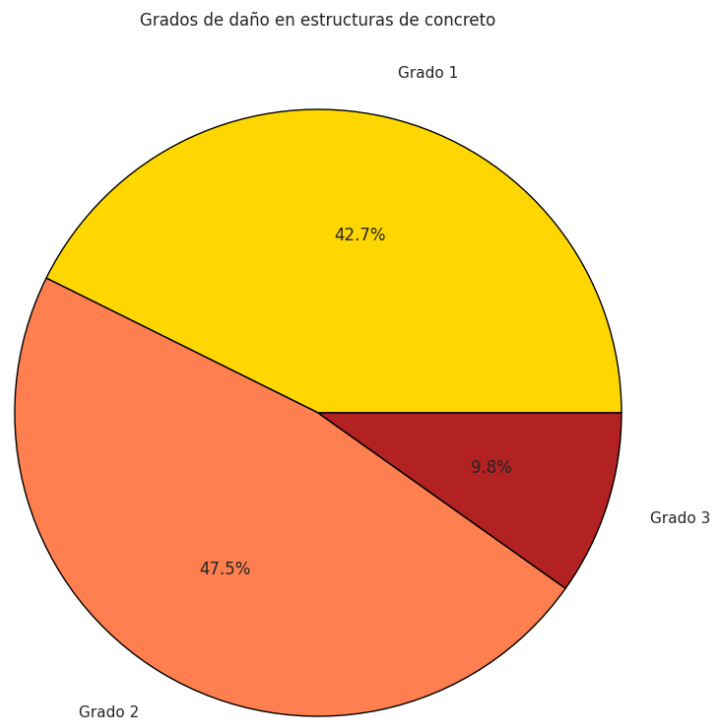


Figura 17: Grado de daño de estructuras con concreto.

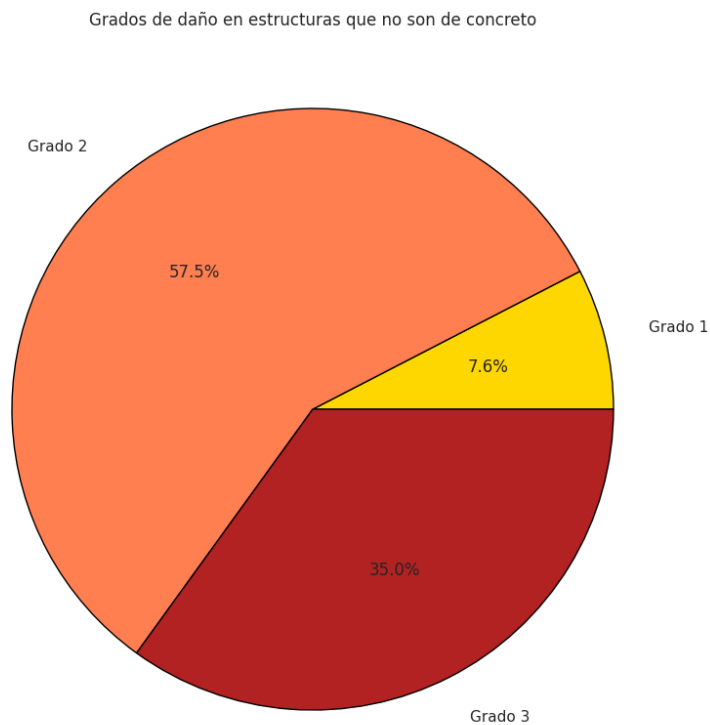


Figura 18: Grado de daño de estructuras sin concreto.

4.4. Hay zonas específicas que reciben mas daño grave

Verificamos si hay zonas que recibieron mas daño grave a nivel porcentual. Para ello solo tomamos en cuenta las zonas que tienen mas de 100 edificaciones. Podemos ver en el grafico que las zonas 16 y 18 se destacan del resto. La 16 se puede ver que casi la totalidad de las edificaciones sufrieron daños medios y altos. Por otra parte la zona 18 tiene el mayor porcentaje de edificios con daños graves. Se podría pensar que estas zonas están mas cerca del epicentro del terremoto.

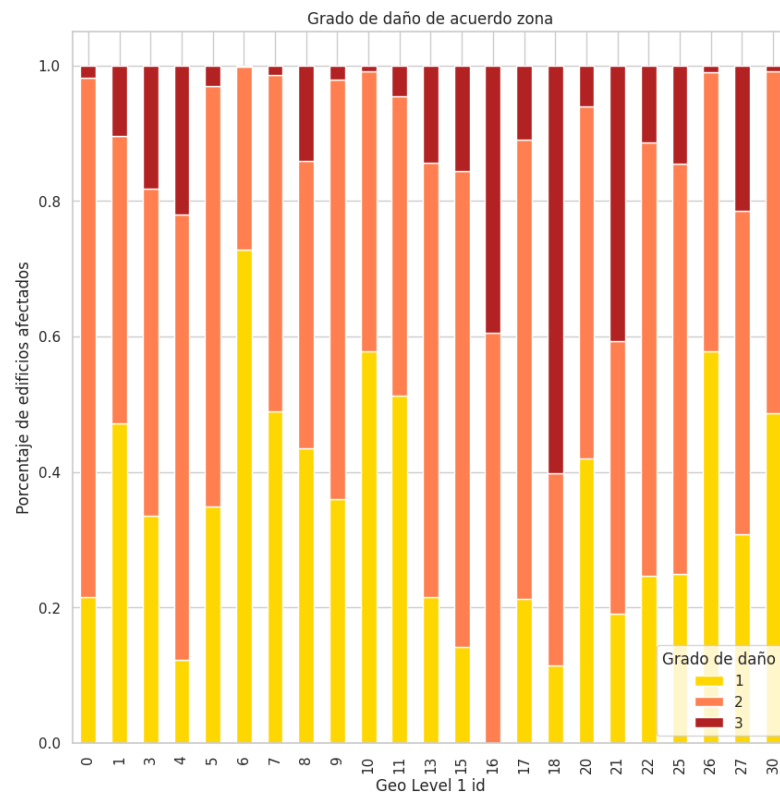


Figura 19: Daño por zona.

4.5. Las estructuras de barro son mas propensas al daño

Se filtro las edificaciones que estan hechas con barro, y se agrupo por grado de daño. Se realizo tambien lo opuesto, como se ve, las estructuras hechas de barro, son mucho mas propensas al daño que las que estan hechas de otro tipo de material.

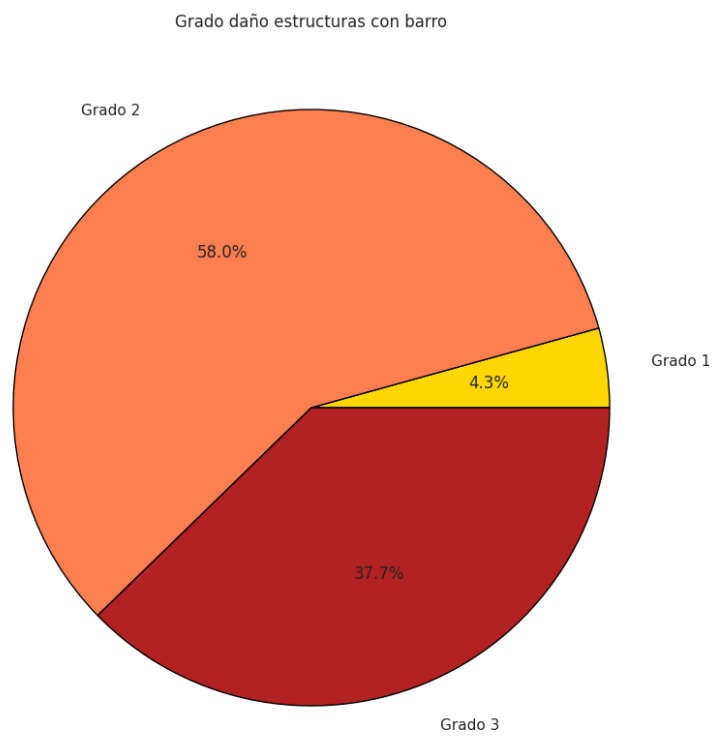


Figura 20: Grado de daño de estructuras de barro.

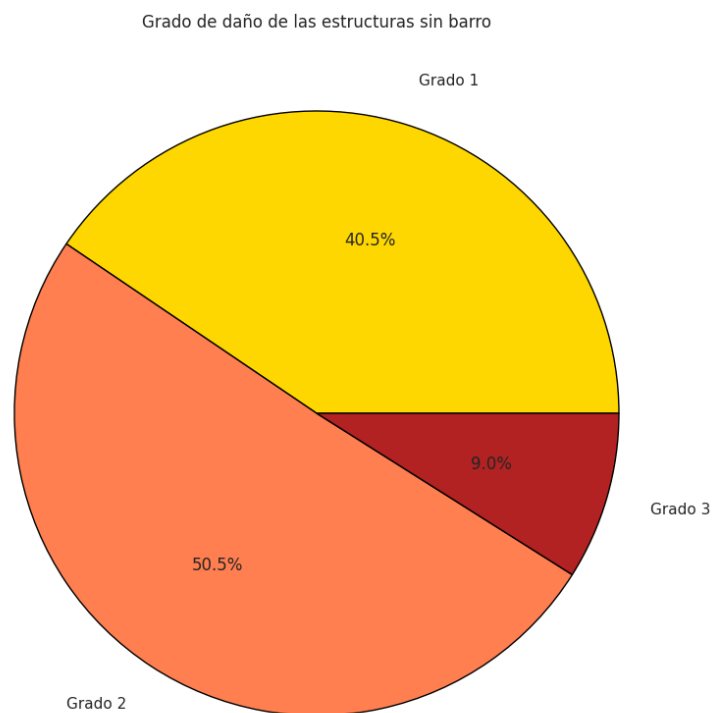


Figura 21: Grado de daño de estructuras sin barro.

4.6. Las edificaciones de Bamboo de mas de 2 pisos son mas inestables y reciben mucho daño

Se filtro las edificaciones de bamboo que tengan mas de 2 pisos, es notable que el daño recibido es importante para este tipo de edificaciones. Casi en su totalidad recibieron daños altos o medios.

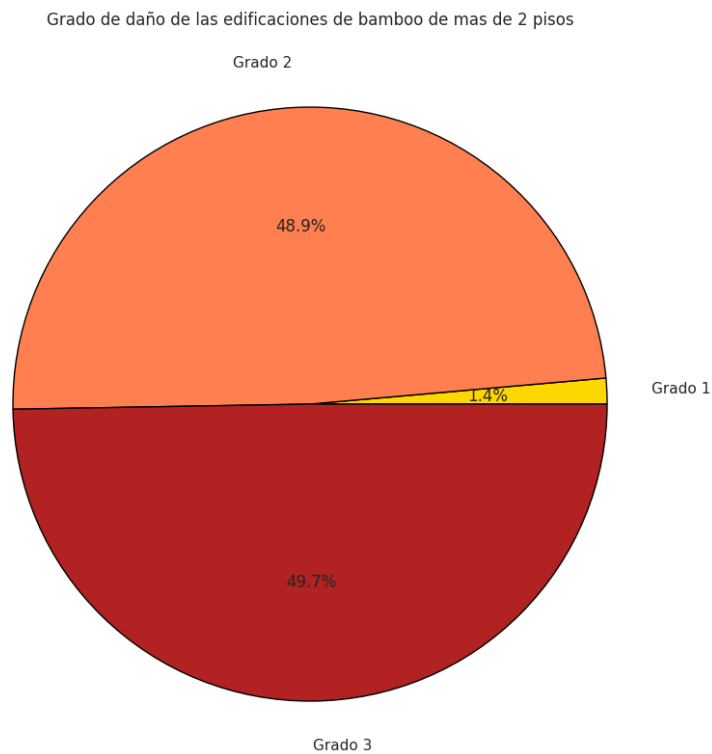


Figura 22: Grado de daño de estructuras de bamboo de más de 2 pisos.

4.7. El terremoto afecta mas a las edificaciones antiguas

Se recategorizo separando las edificaciones en grupos de 25 años, se puede ver que no hay un crecimiento notorio del grado de daño a medida que incrementa la edad, por lo tanto no podemos concluir que la hipotesis sea verdadera.

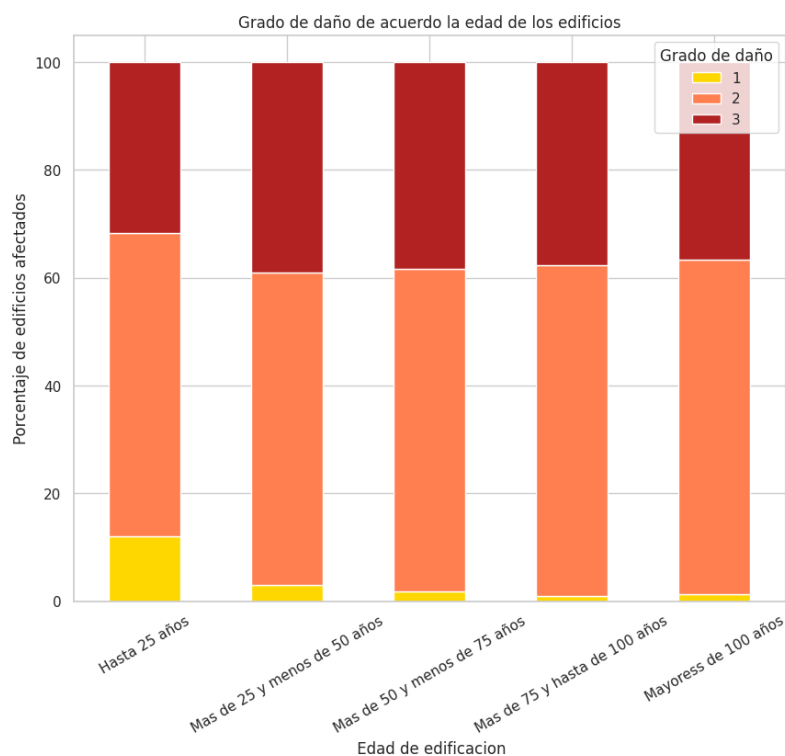


Figura 23: Grado de daño de acuerdo a la edad.

5. Conclusiones

Para cerrar el informe puede hacerse un breve punteo de la información obtenida a partir del análisis exploratorio:

- **Tipo de uso:** oportaron mejor el terremoto las edificaciones de renta e intituciones, peor fue agricultura, policia y 'otras'.
- **Tipo de piso:** aguantó claramente mejor el tipo x mientras que q y n presentaron los peores registros en cuando a grado de daño
- **Tipo de construcción:** aquellas de concreto aguantaron mejor que ninguna, casi sin grado de daño 3. Las peores por lejos fueron stone flag, adobe mud, mud mortar brick y mud mortar stone. Cabe destacar la alta correlacion entre R_c engineered y tipo de piso x
- **Cantidad de pisos:** las edificaciones de 9 pisos solo sufrieron daños de grado 1, las construcciones que peor la pasaron fueron los de 3 pisos. Entre 1-2 y 3-5 fue parejo.
- **Tipo de Fundación:** muy por encima del resto en cuanto al rendimiento las de tipo i, mientras que muy por debajo del resto las de tipo r.
- **Configuracion del plano:** no es tan claro pero le fue mejor a o,a y c y muy mal a d, q y n
- **Zonas geográficas:** de las tipo 1, podemos decir que hay dos zonas que se destacan del resto. La 16 que casi la totalidad de sus edificaciones sufrieron daños graves o medios, y la 18 con mayor porcentaje de edificios con daños altos.
- **Antigüedad de edificios:** No se verifica un agravamiento en los daños en relación con la antigüedad del edificio. Si se puede destacar que en general, los edificios de menos de 20 años de antigüedad sufrieron menos en promedio que el resto

- **Bamboo:** las edificaciones de este material, especialmente de 2 pisos en adelante, resistieron de muy mala manera el terremoto.
- **Altura y superficie:** se concluye, como se podía prever, que el crecimiento en altura sumado a un decrecimiento en superficie tiene por resultado un mayor grado de daño.
- **Estructuras de concreto:** sufrieron significativamente menos que el resto de las edificaciones
- **Epicentro o zonas más afectadas:** no fue pareja la distribución de los daños, concentrándose en áreas puntuales. Esto puede haber sido debido al epicentro del terremoto como a la densidad de edificaciones
- **Estructuras de barro:** contrario a las de concreto, aquellas como barro como componente principal sufrieron los peores daños

6. Links

- Repositorio en Github:
<https://github.com/matias-gonz/GorkhaExploracionDatos>
- Google Colab:
<https://colab.research.google.com/drive/1nmgSn-xTnRPWm18-wFIb6QorrHGAdKj0?usp=sharing>