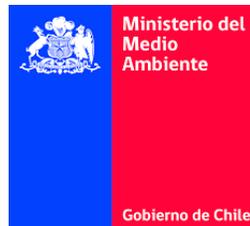


LICITACIÓN N° 608897-83-LE20

ANTECEDENTES GENERADOS EN PROCESOS DE FISCALIZACIÓN Y SANCIÓN REALIZADOS POR LA SUPERINTENDENCIA DEL MEDIO AMBIENTE EN EL MARCO DEL D.S. N°38/11 DEL MMA - NORMA DE EMISIÓN DE RUIDOS GENERADO POR FUENTES QUE INDICA INFORME FINAL

PREPARADO PARA:



PROYECTO N°5074					
VERSIÓN	FECHA	DESCRIPCIÓN	ELABORACIÓN	REVISIÓN	APROBACIÓN
A	19.11.2020	Elaboración Inicial	IGS	MGD	

SANTIAGO, NOVIEMBRE DE 2020

ÍNDICE DE CONTENIDOS

1	Introducción	3
2	Objetivos	3
2.1	Objetivo general	3
2.2	Objetivos específicos	3
3	Metodología	3
3.1	Revisión de la información recibida	3
3.2	Recopilación de la información faltante	5
3.3	Procesamiento de datos	8
3.3.1	Limpieza y codificación de textos y fechas	9
3.3.2	Extracción automática de información relevante	9
3.3.3	Creación de categorías de Actividad Económica y asignación a cada denuncia	11
3.3.4	Creación de variables adicionales necesarias	16
3.3.5	Optimización de la estructura de datos	17
3.3.6	Revisión final y validación de bases de datos	17
3.4	Conteos y cálculos requeridos	19
3.5	Creación de las visualizaciones de datos	20
3.6	Plataforma interactiva y códigos computacionales	22
4	Propuesta de Categorías de Actividad Económica	23
4.1	Actividades de esparcimiento	24
4.2	Actividades productivas	29
4.3	Actividades comerciales	33
4.4	Faenas Constructivas	36
4.5	Actividades de servicio	39
4.6	Elementos de infraestructura	42
4.7	Dispositivos (maquinarias y equipos)	44
5	Estadística Descriptiva de la Información	45
5.1	Denuncias	49
5.2	Fiscalizaciones	56
5.3	Proceso sancionatorio	60

5.4	Actividad económica	66
5.5	Mediciones	69
5.6	PDC	77
6	Discusión y Conclusiones	84
7	Bibliografía	90
8	Profesionales participantes	91

1 INTRODUCCIÓN

En el marco del proyecto **Antecedentes generados en procesos de fiscalización y sanción realizados por la Superintendencia del Medio Ambiente en el marco del D.S. N°38/11 del MMA - norma de emisión de ruidos generado por fuentes que indica** para el Ministerio del Medio Ambiente, este documento corresponde al Informe Final del servicio adjudicado por Gerard Ingeniería Acústica SpA (Control Acústico) en licitación pública N° 608897-83-LE20.

2 OBJETIVOS

2.1 Objetivo general

- Generar antecedentes técnicos y administrativos que permitan sustentar las modificaciones que resulten del proceso de revisión del D.S. 38/11 del MMA para la elaboración del Anteproyecto.

2.2 Objetivos específicos

- Conocer los procesos de fiscalización y sanción generados por la Superintendencia del Medio Ambiente en el marco del D.S.38/11 del MMA.
- Cuantificar los grados de incumplimiento por tipo de actividad económica.
- Conocer los costos asociados a partir de las multas establecidas por la Superintendencia del Medioambiente.

3 METODOLOGÍA

3.1 Revisión de la información recibida

Se revisó el documento con formato Excel y nombre “Planilla_fiscalizaciones_sanciones_ruidos_TODO.xlsx” (en adelante “Planilla de referencia”) que entregó el mandante para la realización del estudio con el objetivo de identificar en qué lugar específico y en qué formato (numérico tabulado, texto tabulado, texto libre, etc.) se encontraba la información requerida: Región, Comuna, Actividad Económica (AE), estado del Proceso Sancionatorio (PS), Niveles de ruido, Periodo, Zona,

Máximos permisibles, Exceso, Multa, existencia de Programa De Cumplimiento (PDC), Medidas de control de ruido comprometidas y su Valor monetario total involucrado.

Para ello se hizo una inspección visual de las tablas en conjunto con un análisis exploratorio del contenido de sus distintas columnas mediante el software RStudio, permitiendo de este modo identificar tres distintos estados de la información:

- Aquella que puede ser extraída y utilizada directamente: Unidad Fiscalizable, Región, Comuna, estado del PS, presencia de PDC, Multa, Fechas de ingreso, fiscalización y formulación de cargos.
- La que está presente, pero que requiere un procesamiento previo a su utilización: categorías para los niveles AE (desde columnas Hechos, Denuncias, UF_Nombre, etc.) para todas las denuncias.
- La información que no está presente y debe ser recopilada por otros medios: Medidas de control de ruido comprometidas y Valor monetario total involucrado en los PDC (documentos alojados en el sitio web). Clasificación LOSMA (Ley Orgánica de la Superintendencia de Medioambiente), Niveles de ruido, Periodo, Zona, Máximos permisibles, Exceso y Fecha de medición (columna Hecho en tablas contenidas en las fichas de sanción en el sitio web de cada PS).

Como resultado de la revisión de la Planilla de referencia se detectaron los siguientes puntos:

- La hoja de denuncias tiene 3 filas erróneas (Id denuncia 13133, 13476, 13735), puesto que dentro de su columna Hechos está alojada la información de otras denuncias que debieran tener cada una su propia fila. Esto se corrigió y de un total de 5708 filas de denuncias se pasó a 5759.
- La hoja de fiscalizaciones tiene 3 filas duplicadas. Luego de corregir se pasó de 700 a 697 fiscalizaciones.
- Si bien la hoja de PS cuenta con 624 filas, varias de ellas están duplicadas. Luego de remover las duplicadas se pasó a un total de 457 casos.
- De las 457 filas del punto anterior, se detectó que 432 estaban efectivamente asociados al D.S. N°38/2011 del MMA, por lo que se descartó el resto. Posteriormente se detectó que 8 casos de la hoja de fiscalizaciones tenían hipervínculos a PS que no están contenidos en la hoja de original de PS en la Planilla de referencia, los que de

todas formas fueron agregados al listado consolidado. De este modo, se identificaron 440 PS a partir de la información recibida.

3.2 Recopilación de la información faltante

Como se dijo en el apartado anterior, fue necesario recopilar información adicional acerca de los PS desde el sitio web del Sistema Nacional de Fiscalización Ambiental¹ (SNIFA), especialmente a lo referido con los Programas de Cumplimiento (PDC) y la información de mediciones asociadas a la fiscalización. Para ello, se realizaron las siguientes actividades:

- A. Se desarrollaron y ejecutaron códigos computacionales para la descarga automática y selectiva de la información contenida en el sitio web, siguiendo los hipervínculos incluidos en la Planilla de referencia asociados a los PS y aplicando técnicas de *web scraping* (paquete *rvest*, software RStudio). Esto permitió la descarga de todos los documentos asociados al PS (Formulación de cargos, PDC, Resolución Sancionatoria, entre otros, en formatos PDF o Word, RAR, ZIP, etc.), completando cerca de 33 GB de peso en disco duro, así como también la descarga de la información contenida en las tablas y textos del sitio mismo: Fechas, Instrumento infringido, Hecho, Estado, Sanción, Multa, Clasificación LOSMA, Expediente Fiscalización, etc. Esto último se llevó a cabo con el fin de actualizar y contrastar la información al compararla con el contenido de la Planilla de referencia. La última actualización se realizó el 18 de octubre de 2020.

¹ <https://snifa.sma.gob.cl>

Ilustración 1: Captura de pantalla de SNIFA (PS 2285)

#	Nombre Documento	Tipo Documento	Fecha	Link
1	Formulación de Cargos	Formulación de Cargos	13-08-2020	Descargar
2	Designa Fiscal Instructor	Otros	13-08-2020	Descargar
3	Notificación FdC titular	Notificación Formulación de Cargos	19-08-2020	Descargar
4	Notificación FdC denunciante	Otros	19-08-2020	Descargar
5	Res. Ex. N°1	Resuelve Solicitud Ampliación de Plazos (Descargos y PDC Presentación)	13-08-2020	Descargar
6	Programa de Cumplimiento	PDC Presentación	11-09-2020	Descargar
7	Escrito titular	Otros	11-09-2020	Descargar
8	Escrito titular	Otros	14-10-2020	Descargar

#	Hecho	Instrumento Infringido	Infracción (Art.35 LOSMA)	Clasificación (Art. 36 LOSMA)
1	La obtención, con fechas 14 de mayo de 2019 y 01 de julio de 2019, de Niveles de Presión Sonora Corregidos (NPC) de 56 dB(A) y 51 dB(A), respectivamente, ambas mediciones efectuadas en horario nocturno, en condición interna, con ventana abierta, en un receptor sensible ubicado en Zona III.	NE:38/2011 Link Detalle	h) El incumplimiento de las Normas de Emisión, cuando corresponda	Leves <i>Hecho, actos u omisiones que contravengan cualquier precepto o medida obligatorios y que no constituyan infracción gravísima o grave, de acuerdo con lo previsto en los números anteriores.</i>

B. Puesto que algunos documentos asociados a los PDC corresponden archivos PDF que contienen imágenes escaneadas y sin un formato uniformado, y por lo tanto la tarea de extraer la información de manera automática es de baja precisión aun cuando se utilicen técnicas de OCR, se optó por recopilar la información de los PDC de manera manual. A la fecha de la última descarga de documentos, se identificaron 225 PS que presentaron PDC, para los cuales se recopiló la siguiente información: Estado PDC, Medidas de control de ruido propuestas, Valor económico total declarado, Fechas asociadas, etc.. Para la categorización de las medidas de control de ruido se utilizó lo dispuesto en el formulario oficial de presentación de PDC, en su sección Acciones (en la Ilustración 2 se aprecia un ejemplo con las 12 categorías existentes). Dado que los valores económicos en algunos casos se

declaran en Unidades de Fomento (UF), se utilizó el valor de conversión indicado en el sitio web del [Banco Central de Chile](#), según fecha de presentación del PDC. Para unificar la información, todos los valores CLP fueron convertidos al valor UF correspondiente.

Finalmente, durante esta recopilación manual también se registró la información de las mediciones asociadas a estos PS que presentaron PDC, con tal de utilizar posteriormente esta información para evaluar la precisión los resultados de la extracción automática de la información y su validación (ver capítulo 3.3.6). De este modo, para la medición que representó el mayor exceso a los máximos permisibles se recopiló Fecha, Periodo, Zona, Nivel de Presión sonora Corregido (NPC) y Exceso.

Ilustración 2: Captura de pantalla de identificador de acciones comprometidas de los PDC.

4. ACCIONES COMPROMETIDAS:	
N° Identificador	1
<p>Acciones <i>Marque una de las siguientes medida(s) a implementar para reducir el ruido. Si desea marcar más de una, realizar en tabla siguiente.</i></p>	<p><input type="checkbox"/> Barrera acústica: Consiste en una barrera con un material cuya densidad debe ser superior a los 10 Kg/m², la cual se debe instalar lo más cerca posible de la fuente para ser efectiva.</p> <p><input type="checkbox"/> Encierros acústicos: Considera la elaboración de una construcción que encierre la fuente, con murallas tipo sándwich con acero de 2 mm en ambas caras, material anticorrosivo alquídico, y núcleo de lana de vidrio de 50 mm de espesor y 32 Kg/m³ de densidad superficial. El panel de acero interior debe ser perforado en un 60%.</p> <p><input type="checkbox"/> Puerta acústica: Se basa en la construcción de una puerta acústica tipo sándwich, de características similares al encierro acústico. Esto es, ambas caras de acero de 2 mm, con núcleo de 50 mm de espesor y densidad superficial de 32 Kg/m³. Esta debe tener un marco perimetral estructural y pomeles que soporten el peso de esta.</p> <p><input type="checkbox"/> Celosía acústica: Corresponden a un conjunto de celosías acústicas para la parte inferior de la puerta, construida con acero galvanizado.</p> <p><input type="checkbox"/> Silenciador tipo Splitter: Los silenciadores tipo Splitter se utilizan a la salida de ductos de aire, y similares, para evitar la propagación del ruido emitidos por esos.</p> <p><input type="checkbox"/> Termopanel: Corresponden, en la generalidad, a vidrios dobles que proveen una reducción sonora de $R_w = 26$ dB. Se destaca el hecho que estos deben contar con un montaje que permita un cierre hermético de la habitación.</p> <p><input type="checkbox"/> Limitador acústico: Son equipos electrónicos que se incluyen dentro de la cadena electroacústica, que permiten limitar el nivel de potencia acústica que genera el sistema en su totalidad.</p> <p><input type="checkbox"/> Recubrimiento con material de absorción de paredes, piso o techumbre: El recubrimiento con material aislante de ruido es una medida que está orientada en evitar que existan reflexiones de las ondas de sonido. Esta medida debe ser instalada en sectores donde no exista riesgo de deterioro y debe pasar por un tratamiento contra incendios. La atenuación máxima que se espera por medio de esta medida es de 2 dBA. Los materiales más utilizados son las espumas acústicas de poliestireno y la lana mineral.</p> <p><input type="checkbox"/> Reubicación de equipos o maquinaria generadora de ruido: Realizar la reubicación de los equipos o maquinaria, desplazando el instrumento emisor de ruido a un sector donde no genere superaciones al D.S. N°38/2011 en receptores cercanos.</p> <p><input type="checkbox"/> Cambio en la actividad: Realizar el cambio de la actividad productiva, por otra que no genere emisión de ruidos molestos.</p> <p><input type="checkbox"/> Traslado o cierre de la unidad fiscalizable: Realizar el cambio de ubicación de la actividad o el cierre definitivo del establecimiento actividades en el sector.</p> <p><input checked="" type="checkbox"/> Otras medidas (indicar todas las otras medidas que usted considere necesarias y que se implementarán antes de la medición final de presión sonora): Se procederá a</p>

3.3 Procesamiento de datos

El procesamiento de datos es una de las etapas más demandantes e importantes a lo largo de un proyecto de ciencia de datos y este proyecto no fue la excepción. A continuación, se detallan todos los pasos realizados que fueron necesarios

para que los datos estén limpios, ordenados y con la estructura óptima antes de realizar los cálculos y visualizaciones. En todos los casos se utilizó el software RStudio.

3.3.1 Limpieza y codificación de textos y fechas

Se uniformaron y agruparon los textos asociados a variables cualitativas mediante la solución de distintos problemas comunes como lo es la presencia de: puntos, comas u otros símbolos, espacios en blanco extras, el uso arbitrario de mayúsculas y minúsculas, más de una manera de escribir la misma palabra (ejemplo: “Stgo.”, “Santiago”, “santiago”), entre otros. Lo anterior fue especialmente relevante para los atributos categóricos tales como Región, Comuna, estado del PS, estado del PDC, clasificación LOSMA, etc. así como también para las columnas con textos largos (Hechos).

Para el uso de caracteres se consideró el protocolo de codificación UTF-8, permitiendo una correcta manipulación de caracteres especiales del idioma español como tildes, cremillas y la letra “ñ”. La clara definición de este parámetro es fundamental para el uso posterior de la base de datos en herramientas computacionales.

En el caso de las fechas se utilizó el formato ISO 8601, el cual establece la siguiente nomenclatura: “YYYY-MM-DD”, es decir, 4 caracteres numéricos para el año (Y: Year), guión, 2 caracteres numéricos para el mes (M: Month), guión y 2 caracteres numéricos para el día (D: Day).

Para estas tareas se utilizaron principalmente los paquetes *dplyr*, *stringr* y *lubridate* del lenguaje R.

3.3.2 Extracción automática de información relevante

Se desarrollaron y ejecutaron códigos computacionales para extraer automáticamente la información relevante para los casos en los que esto fuera posible. Para el caso de la información que ya estaba presente en la Planilla de referencia se aplicaron filtros condicionales de manera directa a los datos tabulados, mientras que para aquellos casos donde era necesario un procesamiento previo se aplicaron técnicas de procesamiento de lenguaje natural mediante [Expresiones Regulares](#) (ER) y/o Minería de Texto.

El principal caso de aplicación de ER correspondió a la extracción automática de la información de mediciones, la que para la mayoría de los casos está contenida en el texto de la columna Hecho, recopilada desde el sitio web de la SNIFA. Allí se tienen textos como el del ejemplo de la Ilustración 1, en el cual es posible identificar fechas de medición, periodo, zona del D.S. N°38/2011 del MMA y valores NPC. Aplicando las ER es posible extraer de manera aislada y correcta esta información para luego calcular el nivel máximo permisible y el exceso en dB asociado con tal de contar con toda la información de interés con un alto nivel de precisión (ver capítulo 3.3.6). Para estas tareas se utilizaron principalmente los paquetes *dplyr*, *stringr*, *stringi* y *rebus* del lenguaje R.

Si bien este tipo de caso corresponde a la mayoría, hay otros en los cuales no hay una situación ideal y que por lo tanto no permiten una extracción automática precisa. Para estos casos, se aplicó una recopilación manual, los cuales son:

- Solo existe información del exceso o superación al máximo permisible y nada más.
- Zona rural sin información de ruido de fondo ni máximo permisible.
- Redacción compleja o incompleta, no permitiendo encontrar patrones o reglas que permitan aplicar las ER de forma correcta. Para estos casos se priorizó no equivocarse en lugar de recoger toda la información.

En los casos de PS con más de una medición asociada, se consideró toda la información de aquella medición correspondiente al mayor exceso [dB] encontrado, a excepción de la fecha ya que para efectos de análisis del tiempo transcurridos entre hitos del PS se consideró relevante la fecha de la primera medición, aun cuando ésta no represente el mayor exceso.

Por otro lado, con el objetivo de categorizar los niveles de AE se realizaron diferentes tareas para la extracción de información desde la columna de texto Hecho de la hoja denuncias de la Planilla de referencia, de las cuales algunas fueron automatizadas mediante códigos y reglas lógicas mientras que otras fueron realizadas de forma manual. Dentro de las tareas por código se tiene a las que tienen relación con Minería de Texto: *tokenización* de texto para conteo y ranking de palabras individuales y reglas de asociación para encontrar patrones de aparición simultánea y jerarquías entre las palabras. En el siguiente capítulo se presentan mayores detalles de estos procesos.

3.3.3 Creación de categorías de Actividad Económica y asignación a cada denuncia

Para categorizar las fuentes de ruido se tomó como base las actividades económicas para las cuales aplica el Decreto Supremo N° 38/2011 del MMA según se indica en la Resolución Exenta N° 867 de la Superintendencia de Medio Ambiente, "Protocolo Técnico para la Fiscalización del D.S MMA 38/2011 y Exigencias Asociadas al Control del Ruido en Instrumentos de Competencia de la SMA" que se muestra en la siguiente tabla:

Ilustración 3: Imagen de Resolución Exenta N° 867 de la Superintendencia de Medio Ambiente con indicación de actividades económicas.

Tabla 2 – Fuentes Emisoras de ruido afectas a la Norma de Emisión:

Fuente	Descripción	Artículo
Actividades Productivas	Industrias, depósitos, talleres, bodegas y similares, así como actividades de extracción u obtención de productos, como las actividades agrícolas, ganaderas, forestales, extractivas, mineras y similares	Art. 6°, punto 1.
Actividades Comerciales	Instalaciones destinadas principalmente a la compraventa de mercaderías, productos y/o servicios diversos.	Art. 6°, punto 2.
Actividades de Esparcimiento	Instalaciones destinadas principalmente a la recreación, el deporte, el ocio, la cultura y similares. Ejemplo de esto pueden ser recintos deportivos, discotecas, centros culturales, entre otros.	Art. 6°, punto 3.
Actividades de Servicio	Instalaciones destinadas al servicio, público o privado, de salud, de educación, de seguridad, social, comunitario, religioso, servicios profesionales, y similares.	Art. 6°, punto 4.
Infraestructura de Transporte	Estaciones ferroviarias, terminales de transporte terrestre, recintos marítimos, portuarios y aeroportuarios, y similares. Se incluyen dentro de estas fuentes los dispositivos asociados a las redes de infraestructura de transporte.	Art. 6°, punto 10 a).
Infraestructura Sanitaria	Plantas de captación, tratamiento de agua potable o de aguas servidas, de aguas lluvia, rellenos sanitarios, estaciones exclusivas de transferencia de residuos, y similares; y redes tales como distribución de agua potable o de aguas servidas, evacuación de aguas lluvia, y similares.	Art. 6°, punto 10, b).
Infraestructura Energética	Instalaciones de generación, distribución o almacenamiento de energía, combustibles o telecomunicaciones; y redes de distribución o conducción de energía, combustibles o telecomunicaciones.	Art. 6°, punto 10, c).
Faenas Constructivas	Actividades de construcción, reparación, modificación, alteración, reconstrucción o demolición, entre otros.	Art. 6°, punto 12.
Dispositivos	Maquinarias, equipos o aparatos, tales como generadores eléctricos, calderas, compresores, equipos de climatización, de ventilación, de extracción, y similares, o compuesto por una combinación de ellos.	Art. 6°, punto 8.

Se consideraron 3 niveles, el primero, nivel 1 coincide con la primera columna de la tabla antes mostrada, "Fuente". Luego como guía para el nivel 2 se utilizó la columna "Descripción" y posteriormente se agregó donde fuese posible una tercera categoría, denominada "Nivel 3" la cual se obtuvo a partir de herramientas de análisis de datos y de criterio experto.

Dichos niveles de categorías fueron asignados de manera individual a cada denuncia. El listado de categorías fue validado por la contraparte técnica y puede ser apreciado en detalle en el capítulo 4. Para el caso de las faenas de construcción se desarrolló un cuarto nivel, asociado a la maquinaria que efectuaba la actividad denunciada, esto solo cuando en la columna Hechos se mencionó explícitamente alguna maquinaria.

En primer lugar, para las denuncias que tienen asociada una fiscalización en la Planilla de referencia, se asignó su Nivel 2 correspondiente según la columna "UF_CategoríaEconómica", permitiendo asignar de manera directa también su Nivel 1 asociado.

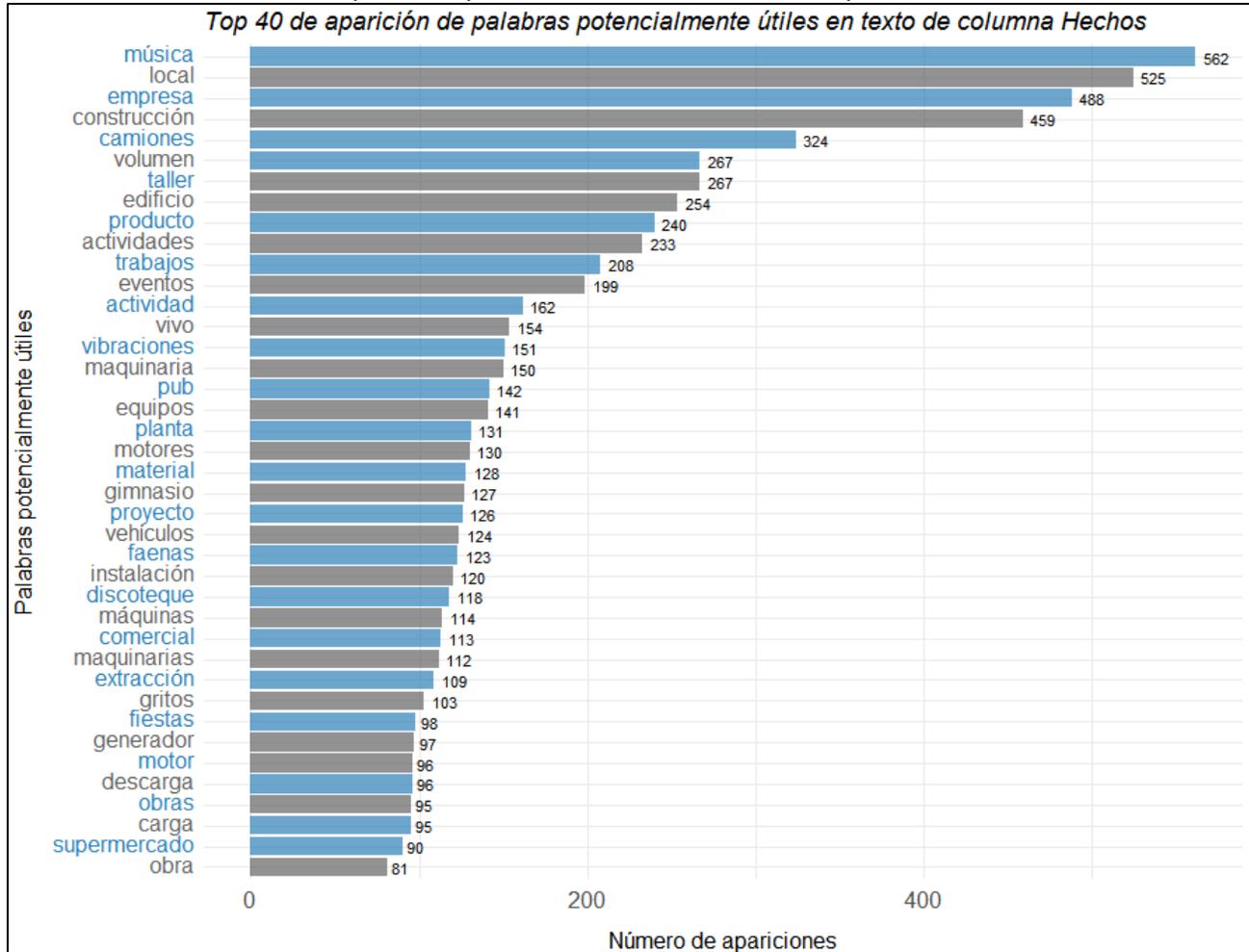
Posteriormente se buscó la manera de definir las categorías de Nivel 3 y su respectiva asignación a cada una de las denuncias. A continuación, se detalla cronológicamente cada uno de los pasos realizados para ello:

1. Se realizó una *tokenización* del texto contenido en la columna Hechos, permitiendo el conteo y ranking de palabras individuales. De este modo fue posible identificar qué palabras son las más frecuentes y cuáles de ellas son potencialmente útiles para su posterior proceso, análisis y asignación de categorías. Para esto se utilizó el paquete *tidytext* del lenguaje R y el software RStudio. Se consideraron todos los aspectos de limpieza de texto mencionados en el capítulo 3.3.1 además de remover las *stopwords*² del idioma español enfatizando la búsqueda de sustantivos.

A continuación se presentan dos visualizaciones: En primer lugar una nube con las palabras más frecuentes del total de 9593 palabras obtenidas en el conteo preliminar, y en segundo lugar el ranking top 40 de las 186 palabras identificadas como candidatas o potencialmente útiles:

² Palabras generalmente cortas que no aportan al análisis. Entre ellas están los artículos, pronombres, preposiciones, conjunciones, etc.

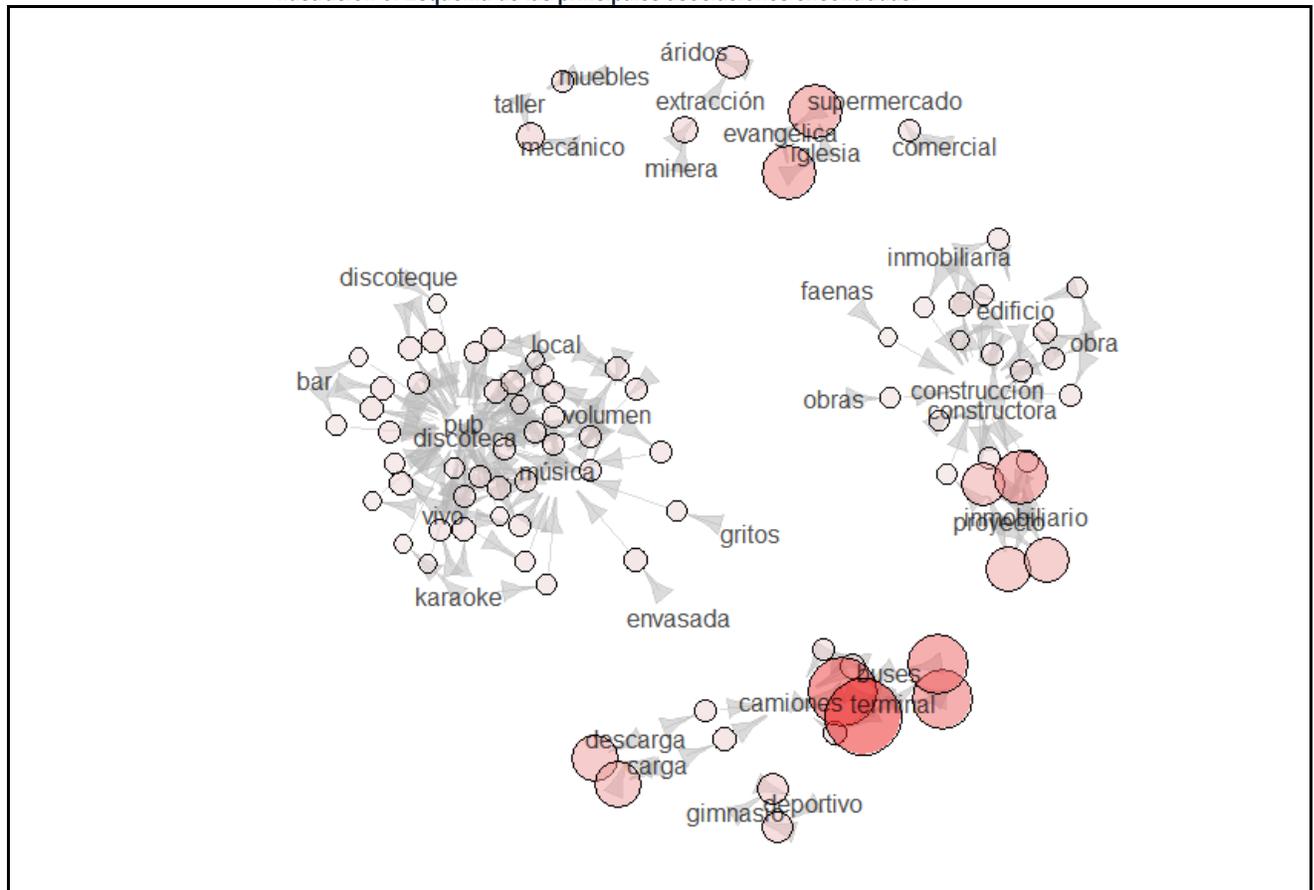
Ilustración 5: Jerarquización de palabras identificadas como candidatas o potencialmente útiles.



- Se calcularon reglas de asociación para las palabras candidatas, en conjunto con algunos términos frecuentes en las columnas "Denuncias", "UnidadFiscalizable", "CategoriaEconomica" y "SubcategoriaEconomica". De este modo fue posible identificar qué conjuntos de palabras cuya aparición simultánea pueden resultar en asignaciones directas y definitivas y cuáles grupos permiten acotar la búsqueda al realizar una segunda lectura ya agrupada. Para esto se utilizó el algoritmo *apriori* con los paquetes *arules* y *avizrules* del lenguaje R.

A continuación un esquema que muestra las principales asociaciones encontradas. El tamaño de los círculos indica la frecuencia relativa de la relación y la nitidez del color indica la exclusividad de la misma.

Ilustración 6: Esquema de las principales asociaciones encontradas.



Se pueden apreciar las siguientes agrupaciones:

- “Taller” con “muebles” y “mecánico”
- “Terminal” con “buses”
- “Iglesia” con “evangélica”
- “Camiones” con “carga” y “descarga”
- “Construcción” o “constructora” por un lado con “proyecto” e “inmobiliario” y por otro lado con “obras”, “edificio” y “faenas”.

- “Gimnasio” con “deportivo”
 - “Extracción” con “áridos” y “minera”
 - “Pub” y “discoteca” con “música”, “volumen”, “vivo”, “karaoke”, “envasada”, “gritos”, “local”, “bar”, etc.
3. Se consideró un proceso de *stemming* para considerar la raíz de las palabras en lugar de varias palabras similares. Un ejemplo de esto es usar la raíz “constru” en vez de “construcción”, “constructora”, “construyendo” o bien “inmobiliar” en vez de “inmobiliaria”, “inmobiliario”, “inmobiliarias” e “inmobiliarios”.
 4. Se crearon reglas lógicas de asignación directa para los casos más sencillos de categorías Nivel 3 dada su alta granularidad. Algunos de ellos son pub, restaurant, gimnasio, discoteca, supermercado, cafetería, terminal de buses, iglesia, entre otros. Así se realizó una asignación automática vía código de todos los niveles.
 5. Finalmente, se realizó una asignación manual de todas las denuncias restantes basándose en los resultados de los pasos previos (conteos y agrupaciones), el ojo experto y el detalle encontrado en las distintas columnas con texto de la Planilla de referencia. Además, se revisaron las asignaciones efectuadas de manera automática corrigiendo en caso de ser necesario. Las denuncias para las que no fue posible la asignación de alguna categoría de alguno o todos los niveles, fueron asignadas a las respectivas categorías “Otros” correspondientes a cada nivel según cada caso, siempre categorizando una determinada denuncia la mayor cantidad de niveles posible dejando el “otros” para cuando ya no es posible una mayor precisión.

3.3.4 Creación de variables adicionales necesarias

Se crearon variables adicionales para el uso óptimo de los datos posteriormente en la etapa de cálculos y visualizaciones. Entre ellas están las variables booleanas con salidas SI/NO asociadas a las respuestas de las preguntas: “¿Tiene información de fiscalización asociada?”, “¿Hay información de PS asociado?”, “¿Presentó PDC?”, etc.

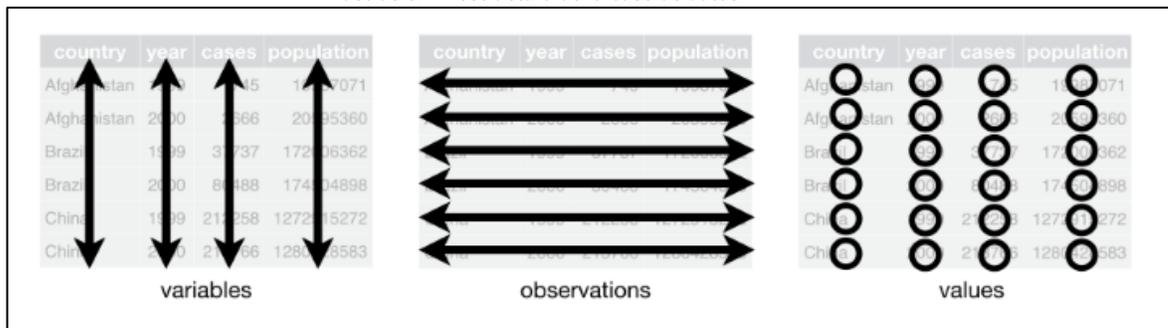
Adicionalmente se crearon las variables de diferencia de tiempo en meses entre algunas fechas de distintos hitos de interés. A saber: fecha de ingreso de denuncia, fecha de primera fiscalización, fecha de formulación de cargos (inicio del PS) y fecha de término de PS, solo en el caso de que este efectivamente haya terminado.

3.3.5 Optimización de la estructura de datos

Una vez extraída y definida toda la información requerida se optimizó la estructura de la base de datos, teniendo tablas ordenadas bajo el estándar Tidy Data³, cuyos preceptos principales son:

- Cada celda tiene un valor asociado a una sola información.
- Cada muestra u observación (en este caso cada denuncia) está en una sola fila.
- Cada variable está en una sola columna.

Ilustración 7: estructura de la base de datos.



Para estas tareas se utilizaron las librerías *dplyr* y *tidyr* del lenguaje R.

De este modo, para los cálculos, las visualizaciones y el desarrollo de la aplicación interactiva se trabajó con distintas tablas según el tipo de datos que contienen. Estas son: Denuncias, Fiscalizaciones, Procesos Sancionatorios, Mediciones, Actividades Económicas.

3.3.6 Revisión final y validación de bases de datos

Mediante el uso de software especializado en análisis de datos (RStudio) se revisó la base de datos con el objetivo de validar el proceso de extracción automática de la información, la limpieza de texto y el contenido completo de las distintas

³ [Tidy Data, Hadley Wickham](#)

tablas. Para ello se realizó la búsqueda de distintos errores como, por ejemplo: valores NA o celdas vacías, anomalías numéricas, conteos generales para exploración y obtención de valores únicos para cada variable cualitativa.

Los principales hallazgos corresponden a:

- Valores NA en algunas columnas debido principalmente a la ausencia de la información tanto en la Planilla de referencia como en el sitio de la SNIFA y sus documentos. En estos casos se optó por mantener de todas formas las filas asociadas, teniendo en cuenta las posibles diferencias que pueden ocurrir al realizar conteos o relacionar las tablas entre sí utilizando esas columnas específicas. La plataforma interactiva desarrollada simplemente no considera en los conteos y visualizaciones los casos cuando algún atributo no tiene valor declarado.
- Un caso con valor numérico anómalo asociado al valor total económico de PDC y 2 valores anómalos en caso de Multas. En ambos casos valores muy altos en comparación al resto. De todas formas, se optó por mantener la información tal y como fue declarada.

Un aspecto importante de la metodología de extracción automática de la información de mediciones detallada en el capítulo 3.3.2., es la capacidad de extracción en cuanto a volumen y precisión que tiene el algoritmo utilizado. Para evaluar dichos conceptos se consideró como referencia la muestra de 225 PS cuya información de mediciones fue recopilada manualmente (para precisión) y los 432 PS con los que se contó inicialmente desde la Planilla de referencia (para volumen). Es así como se obtienen los siguientes resultados de % de extracción y % de precisión para cada parámetro relevante:

Tabla 1: Resumen de resultados de extracción.

Parámetro	Volumen: % de salidas que arroja el algoritmo (total considerado 432 PS)	Precisión: % de info. extraída correctamente en comparación a la referencia (total considerado 225 PS)
Fecha 1era MED	89%	95%
Horario	98%	95%
Zona DS38	92%	99%

Límite	85%	95%
NPC	84%	95%
Exceso	79%	96%

Para todos los casos se obtuvo una precisión no menor al 95%. De este modo se validó la precisión del algoritmo, aceptando un margen de error máximo del 5% dependiendo de cada parámetro. Esto se debe tener en cuenta al observar e interpretar los resultados asociados a mediciones.

En relación con el volumen de extracción, el resultado se consideró aceptable y, tal y como se dijo previamente, la información faltante se recopiló de forma manual. Una de las ventajas de la aplicación de este tipo de herramientas es que justamente permite la extracción automática de alto % de la información de manera eficiente y rápida, ahorrando tiempo y recursos ante cada actualización. Si se quisieran mejorar los porcentajes de volumen y precisión de la extracción, se deben aplicar mejoras en relación con la manera en que se registra y dispone la información (ver capítulo 6).

3.4 Conteos y cálculos requeridos

Se realizaron todos los cálculos de estadísticas para cada denuncia, fiscalización y procesos sancionatorios, agrupadas según las siguientes variables cualitativas:

- Categorías de Actividad Económica de la Unidad Fiscalizable.
- Zonas homologadas al DS38/11 MMA.
- Periodo de medición
- Año.
- Región.
- Comuna.
- Estado de fiscalización.
- Estado del proceso sancionatorio.

- Clasificación de gravedad según LOSMA
- Presencia de PDC y su estado correspondiente en caso de existir
- Medidas de control declaradas en PDC.

También se indican los resultados para las distintas variables cuantitativas como NPC, excesos sobre el máximo permisible, multas, valor total declarado en PDC, tiempo transcurrido entre distintos hitos, etc. Esto en relación con las distintas variables cualitativas antes mencionadas.

Se presentan tablas con valores absolutos (sumas y cantidades), porcentajes y medianas, siendo coherente con las cifras significativas, ordenando de manera descendente u alfabética dependiendo de cada caso.

Para estas tareas se utilizó principalmente el paquete *dplyr* del lenguaje R y el software RStudio.

3.5 Creación de las visualizaciones de datos

Para crear las visualizaciones se consideró el objetivo que se busca con ellas y/o el tipo de dato. A continuación, una tabla resumen con ejemplos de los principales casos utilizados:

Tabla 2: Tipos de visualización según tipos de datos u objetivo

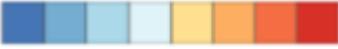
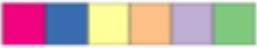
Objetivo	Tipo de dato	Tipo de Visualización	Ejemplo
Comparar, o rankear u observar evolución	Magnitudes agregadas (sumas, promedios, etc.)	Barras	
Observar proporciones respecto a un total	Muestras individuales o agregaciones (sumas, promedios, etc.)	Donas	

		<i>Treemap</i>	
Observar correlación	<i>Muestras individuales</i>	Dispersión	
		Burbujas	
		<i>Heatmap</i>	
Observar distribución	<i>Muestras individuales</i>	Caja y bigotes (<i>boxplot</i>)	
		Histograma	

		Densidad	
--	--	----------	---

Se utilizaron escalas o paletas de colores que tengan relación con la variable que se está mapeando, para lo cual en general se tienen 3 casos:

Tabla 3: Tipos de escalas de colores y ejemplos

Tipo de variable y escala de colores	Ejemplo
Divergente (de un polo a otro polo)	
Secuencial (de un mínimo a un máximo)	
Catagóricas (cualitativas)	

Adicionalmente, se privilegió el uso de paletas de colores para las variables categóricas con no más de 9 colores diferentes dado que a partir de ese valor aparecen dificultades de percepción visual del ser humano para distinguir entre distintos colores.

Para estas tareas se utilizaron principalmente los paquetes *ggplot2*, *ggiraph* y *highcharter* del lenguaje R y el software RStudio.

3.6 Plataforma interactiva y códigos computacionales

Se creó una plataforma para la exploración y visualización interactiva de los datos y sus resultados, utilizando el paquete *shiny* del lenguaje R y el software RStudio. De este modo fue posible definir con flexibilidad la interfaz gráfica y el flujo de

datos que permitan al usuario encontrar intuitivamente la información que busca. Se consideraron 6 pestañas principales, incluyendo: denuncias, fiscalizaciones, procesos sancionatorios, actividades económicas, mediciones y PDC.

Además de pestañas específicas ya mencionadas, se incluyeron filtros para Fechas, Región, presencia o ausencia de unidad fiscalizable, fiscalización y sanción. Así, el usuario puede elegir qué tipo de información y variables quiere ver, además de tener la posibilidad de navegar dentro de la misma herramienta por las tablas que contienen los datos con los que se construyen las visualizaciones, pudiendo cambiar el orden de las filas según los valores de cada columna, hacer búsquedas específicas de texto o números e incluso acceder directamente al sitio web de la SNIFA al clicar en los hipervínculos presentes en las tablas.

Las visualizaciones incluidas son del tipo indicado en el capítulo anterior, dependiendo de la variable mapeada y el objetivo que se busca. Se sugiere explorar las visualizaciones situando el cursor sobre los distintos elementos de cada una, puesto que así se despliega información adicional de manera interactiva.

El acceso a la plataforma es posible mediante la siguiente ruta:

<https://ismaelgomez.shinyapps.io/DS38-snifa-mma-2020/>

La información de usuario y clave será enviada vía correo electrónico de forma privada directamente al mandante, así como el acceso a las bases de datos y todos los códigos computacionales desarrollados para las tareas de web scraping, procesamiento de datos y plataforma interactiva dispuestos en un repositorio también privado.

4 PROPUESTA DE CATEGORÍAS DE ACTIVIDAD ECONÓMICA

A continuación se muestran las categorías de Nivel 1 en el cual se dividen las 5759 denuncias consideradas en el análisis acumuladas entre septiembre de 2014 y septiembre de 2020.

Tabla 4: Categorías de nivel 1, cantidad y porcentaje del total.

Nivel 1	Cantidad	Porcentaje
---------	----------	------------

Actividades de esparcimiento	1595	28%
Actividades productivas	1482	26%
Actividades comerciales	1004	17%
Faenas constructivas	693	12%
Actividades de servicio	369	6%
Elementos de infraestructura	321	6%
Otros	154	3%
Dispositivos (maquinarias y equipos)	120	2%
No aplica	21	< del 1%
Total	5759	100%

En los apartados siguientes (de 4.1 a 4.7) se muestran las categorías nivel 3 agrupadas según su categoría nivel 2. Todas se indican con el número de apariciones en las denuncias. Cada apartado muestra un determinado nivel 1.

4.1 Actividades de esparcimiento

Tabla 5: Cantidad de denuncias por categoría para cada nivel.

Cultura	17
anfiteatro	1
centro cultural	9

cine	1
estudio de grabación	3
teatro	3

Tabla 6: Cantidad de denuncias por categoría para cada nivel.

Deporte	285
academia de artes marciales	3
autódromo	1
cancha	24
cancha de tenis	1
club de planeadores	3
club de rayuela	1
club deportivo	14
complejo deportivo	8
estadio	6
estadio atlético	1

gimnasio	198
hipódromo	3
otros	9
polideportivo	1
recinto deportivo	12

Tabla 7: Cantidad de denuncias por categoría para cada nivel.

Ocio	1234
bar / restobar	83
cabaret	12
carro simulador	2
casino	16
centro de eventos	84
centro gastronómico	1
club	2
club de yates	1
discoteque	485

entretenimiento	24
festival	2
fiesta	5
fiesta / carnaval / festival	1
fuelle de soda	11
karaoke	4
local nocturno	71
otros	15
paseo	1
peña folclórica	1
pub	275
salón de eventos	5
salón de pool	3
vivienda	130

Tabla 8: Cantidad de denuncias por categoría para cada nivel.

Otros	1
--------------	----------

Tabla 9: Cantidad de denuncias por categoría para cada nivel.

Recreación	58
camping	5
centro recreativo	2
circo	14
club de campo	1
complejo turístico	1
lago	2
otros	3
parque	11
parque de diversiones	2
piscina	7
pistas de karting	4
playa	1
plaza cívica	4
quinta de recreo	1

4.2 Actividades productivas

Tabla 10: Cantidad de denuncias por categoría para cada nivel.

Bodega	23
empaque	1
frigorífico	22

Tabla 11: Cantidad de denuncias por categoría para cada nivel.

Depósito	35
acopio	31
depósito de materiales	4

Tabla 12: Cantidad de denuncias por categoría para cada nivel.

Extracción	17
pesca	17
Extracción agrícola	107
agroindustrias agrícola	107
Extracción forestal	19
forestal	19
Extracción ganadera	27

agroindustrias ganadera	24
ganadería	3
Extracción minera	246
áridos	39
minera	206
otros	1

Industria	586
agroindustrias	6
aserradero	36
aserradero / procesadora de madera	6
cristalería	1
embalaje industriales	1
envasadora	2
fundición	1

laboratorio	6
metalúrgica	13
molino	10
otros	10
papelera	2
planta / fábrica	481
planta de cal	1
procesadora	1
procesadora de madera	2
procesadora y envasado de agua	2
talabartería	5

Tabla 13: Cantidad de denuncias por categoría para cada nivel.

otros	6
--------------	----------

Tabla 14: Cantidad de denuncias por categoría para cada nivel.

Taller	416
asfaltos	4

astillador	1
barraca	14
barraca y aserradero	1
carpintería	5
chatarrería	16
estructuras metálicas	45
galpón	9
hojalatería	8
imprensa	10
maestranza	66
metalmecánica	7
otros	53
reparación de herramientas	7
taller aluminio y cristales	5
taller artesanías	5

taller de costura	3
taller de muebles	6
taller de pintura	4
taller de soldadura	10
taller mecánico	117
tornerí-a	11
vulcanización	9

4.3 Actividades comerciales

Tabla 15: Cantidad de denuncias por categoría para cada nivel.

Mercadería	382
alimentos	7
automotora	8
centro comercial	19
comercializadora y distribuidora	3
distribuidora	27

feria	2
importadora	4
importadora y exportadora	2
leñería	8
local comercial	119
mall	18
mercado	1
mini market	9
mueblería	18
otros	11
supermercado	126

Tabla 16: Cantidad de denuncias por categoría para cada nivel.

Productos	165
carnicería	16
heladería	3
panadería	144

pastelería	2
------------	---

Tabla 17: Cantidad de denuncias por categoría para cada nivel.

Servicios	457
alimentos	1
banco	8
cafetería	10
call center	1
carro de comida	2
centro de belleza	1
contratista	1
desabolladuría	18
encomiendas	8
estación de servicios	15
farmacia	2
ferretería	12
hospedaje	2

hotel	39
hotel canino	2
lavado de autos	20
lavandería	12
otros	10
peluquería	1
productora	1
productora musical	2
restaurant	269
restaurant y hotel	1
sala de ensayo	4
transporte de carga	12
veterinaria	3

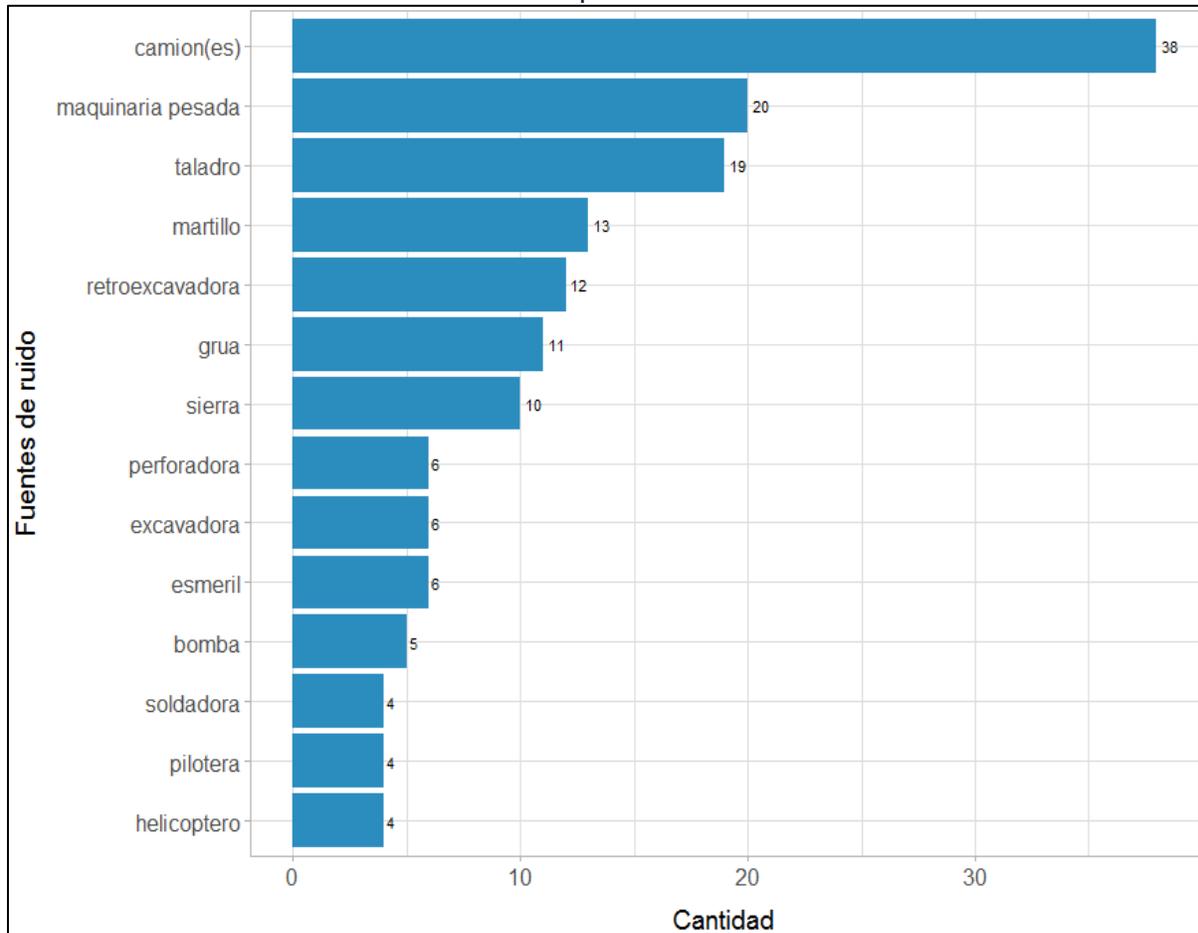
4.4 Faenas Constructivas

Tabla 18: Cantidad de denuncias por categoría para cada nivel.

Construcción	654
industria	2
infraestructura	105
inmobiliario	288
otros	259
Demolición	37
industria	1
infraestructura	3
inmobiliario	17
otros	16
Reparación	2
infraestructura	2

Se consideró el levantamiento de información específico para las Faenas Constructivas con el objetivo de generar un cuarto nivel de categorías asociado a fuentes de ruido cuando fuera posible. Del total de 693 denuncias categorizadas como Faenas Constructivas se encontraron 121 casos (17.5%) con al menos una fuente de ruido declarada. A continuación, una ilustración con el ranking de las fuentes más recurrentes (sólo aquellas con más de 3 apariciones):

Ilustración 8: Nivel 4 para faenas constructivas.



4.5 Actividades de servicio

Tabla 19: Cantidad de denuncias por categoría para cada nivel.

Comunitario	27
comunitario	24
juzgado	1
municipalidad	2
Educación	122
academia de baile	8
academia de música	1
colegio	51
escuela	14
establecimiento educacional	6
guardería infantil	2
instituto	11
jardín infantil	10
liceo	7
universidad	12

Profesionales	10
asesorías	4
oficina comercial	6
Religioso	129
Salud	68
central de abastecimiento del sistema nacional de servicios de salud	1
centro de cáncer	1
centro de diagnóstico	1
centro de diálisis	2
centro de rehabilitación	3
centro kinésico	1
centro radiológico	2
clínica	39
clínica dental	2
comunidad terapéutica	3
consulta dental	1

hospital	9
instituto de salud pública	1
óptica	1
otros	1
Seguridad	7
bomberos	3
comisaria	1
ejército	1
otros	1
seguridad social	1
Social	6
diario	1
dirección de aseo y ornato de la municipalidad	1
radio	3
tesorería general	1

4.6 Elementos de infraestructura

Tabla 20: Cantidad de denuncias por categoría para cada nivel.

Energéticos	110
instalación de generación	47
instalación de transmisión	57
otros	6
Sanitaria	32
centro de reciclaje	9
estación de transferencia	1
manejo de residuos	2
sanitaria	17
vertedero	3
Telecomunicaciones	31
antena	20
otros	11

Transporte	148
aeródromo	3
aeropuerto	3
buses	35
camión	15
estacionamiento	14
ferrocarril	6
metro	12
otros	9
puerto	29
terminal de buses	21
terminal de taxis	1

4.7 Dispositivos (maquinarias y equipos)

Tabla 21: Cantidad de denuncias por categoría para cada nivel.

Bombas	22
Calderas	7
Climatización	25
Compresores	2
Extracción	15
Generadores eléctricos	26
Alarma	2
Otros	21

Cabe señalar que no fue posible categorizar 154 denuncias por información poco concluyente, las cuales se dejaron en "Otros" de nivel 1 y otras 21 donde se determinó claramente que no aplicaba el DS38 las cuales se dejaron en nivel 1 "no aplica".

5 ESTADÍSTICA DESCRIPTIVA DE LA INFORMACIÓN

En este capítulo se muestran los resultados principales obtenidos a partir de la sistematización de los datos de denuncias, fiscalizaciones y procesos sancionatorios. En la siguiente tabla se muestran las fechas de inicio y fin de la información que fue entregada para realizar esta consultoría además de información adicional provista por el consultor.

Tabla 22: fechas de inicio y fin de la información disponible.

Ítem	Fecha de inicio	Fecha final	Información adicional
Denuncias	Enero de 2013	Julio de 2020	-
Fiscalizaciones	Enero de 2014	Septiembre de 2020	Octubre de 2020*
Procesos sancionatorios	Septiembre de 2014	Septiembre de 2020	Octubre de 2020*

*Se actualizó el estado de los procesos sancionatorios

En las siguientes tablas se muestra el resumen de los datos obtenidos en la presente consultoría desde el punto de vista de las denuncias, considerando primero una agrupación por Actividad Económica y luego considerando las regiones del país.

Tabla 23: Resumen general de los datos obtenidos, ordenados por Actividades económicas.

Actividad Económica	Cantidad	Con UF identificable	Fiscalizadas	Sancionadas	NPC Diurno [dB(A)]	NPC Nocturno [dB(A)]	Exceso Diurno [dB]	Exceso Nocturno [dB]	Mediana PDC UF	Mediana multa UTA
actividades de esparcimiento	1595	730	232	92	69	61	5.5	13	174.4	8.0
actividades productivas	1482	696	168	27	67	54	9	6	747.7	5.9
actividades comerciales	1004	413	137	34	68.5	59	6	10	158.9	2.7
faenas constructivas	693	324	86	15	74	68	12	23	399.1	19.0
actividades de servicio	369	150	44	12	68	61.5	7	13	253.9	1.0
elementos de infraestructura	321	183	65	3	68	45	4	7	883.4	5.0
otros	154	4	2	1	67	58	15	13	549.8	1.0
dispositivos	120	35	10	5	67	53	7	4	2107.5	4.0
no aplica	21	3	0	0	0	0	0	0	0.0	0.0

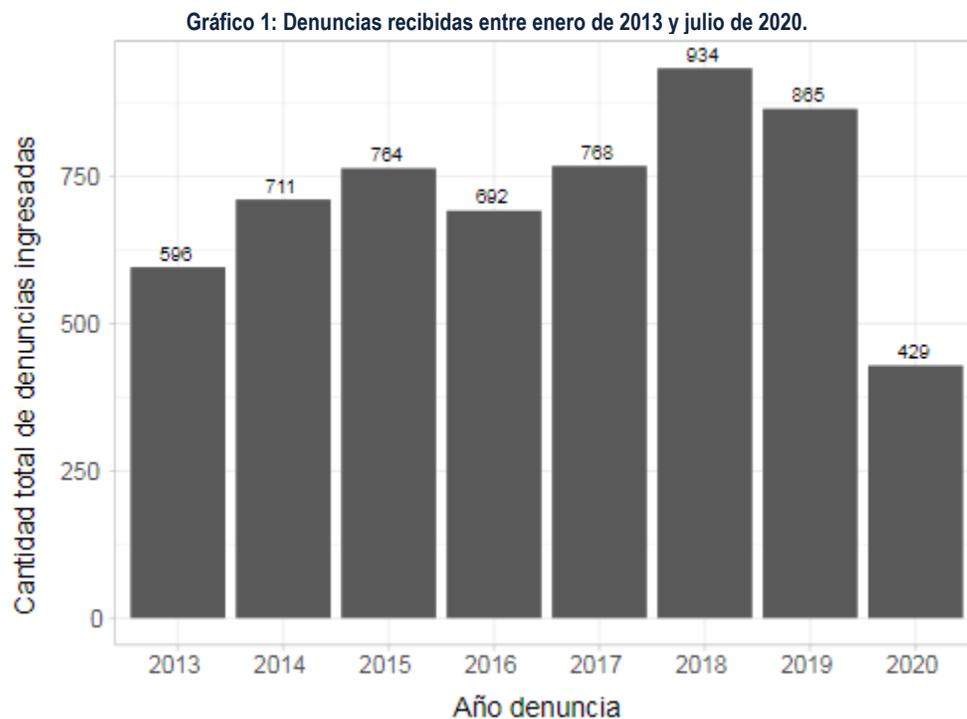
Tabla 24: Resumen general de los datos obtenidos, ordenados por Región.

División administrativa	Cantidad	Con UF identificable	Fiscalizadas	Sancionadas	NPC Diurno [dB(A)]	NPC Nocturno [dB(A)]	Exceso Diurno [dB]	Exceso Nocturno [dB]	Mediana PDC UF	Mediana multa UTA
Región Metropolitana	2268	1086	319	75	72	57	10	9	366.2	12
Región del Biobío	515	260	79	13	71	62	14	12	737.2	3.8
Región de Valparaíso	457	90	29	1	67	58	7	12.5	352.3	9
Región de la Araucanía	324	97	26	12	70	58	5	7	206	2.5
Región del Maule	322	115	36	12	67	60	2	11	129.7	1
Región de Coquimbo	311	187	66	5	69	45	10	7	104.4	7
Región de Antofagasta	296	110	28	4	78.5	63	18.5	18	232.7	13
Región del Libertador General Bernardo O'Higgins	216	69	1	2	0	59	0	14	0	33.2
Región de los Lagos	199	62	7	3	71	52.5	15	7.5	3046.2	1
Región de Tarapacá	168	83	35	14	69	62	5	16	49.2	6.75
Región de Arica y Parinacota	161	64	22	13	68.5	61	4.5	12	77.9	1.4
Región de Los Ríos	153	79	38	6	68	61.5	16.8	13	697.95	2.5
Región de Atacama	130	85	21	17	95	59	35	12	61.6	6
Región de Ñuble	120	72	19	7	61	57	9	10.6	106.3	5

Región de Magallanes y la Antártica Chilena	63	38	11	4	0	57	0	7.5	63.7	4.45
Región de Aysén del General Carlos Ibáñez del Campo	52	41	7	1	0	63	0	18	98.8	8

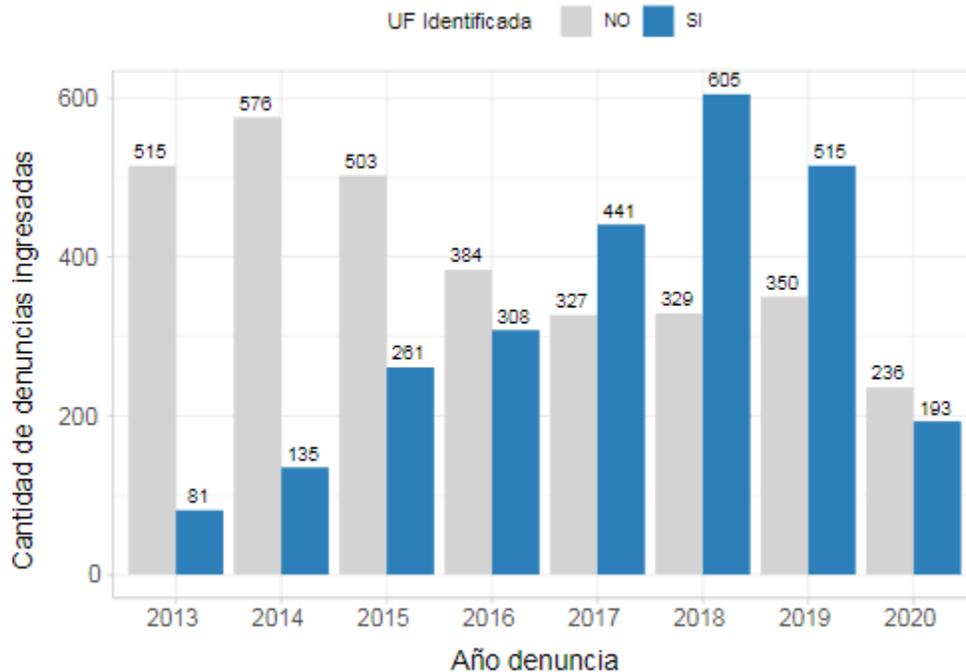
5.1 Denuncias

Las denuncias recibidas entre enero de 2013 y julio de 2020 fueron 5759 y se muestran en el siguiente gráfico agrupadas por el año de su ingreso.



De las denuncias procesadas, 2539 cuentan con Unidades de Fiscalización (UF) identificadas. En el siguiente gráfico se muestra la evolución desde el 2014 hasta 2020 de las denuncias separadas en las que cuentan UF y las que no.

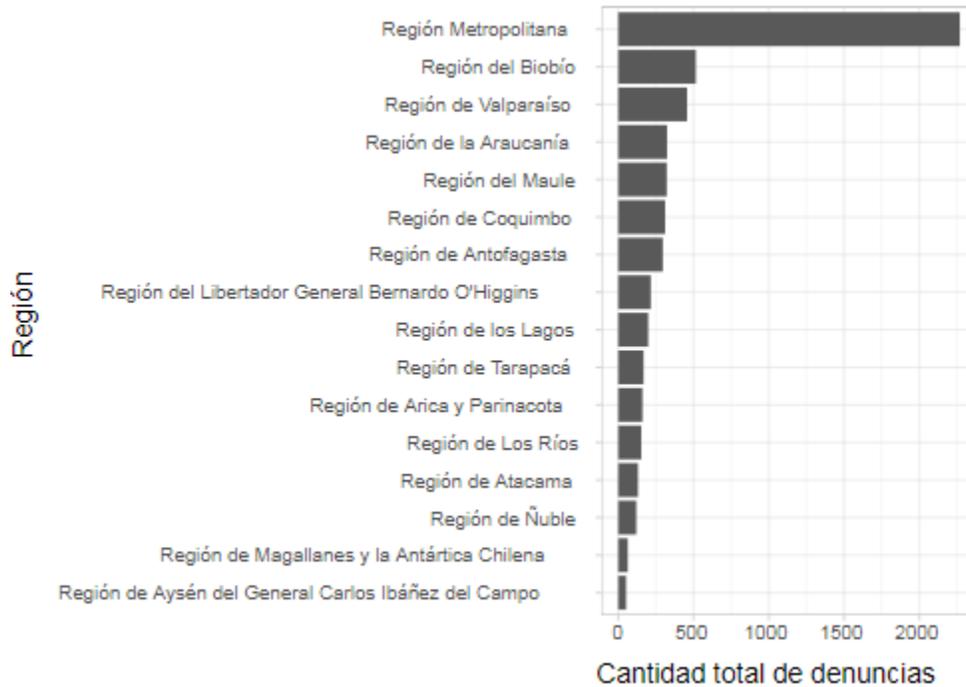
Gráfico 2: Denuncias con y sin Unidades de Fiscalización (UF) identificadas.



Se observa que el valor más alto de denuncias con UF se da el 2018 con 605 denuncias, mientras que sin UF el año 2014, si consideramos el total de denuncias con y sin UF, el valor más alto se obtiene el 2018 con 934 denuncias, considerando en el promedio diario se obtiene un valor de 2.6 denuncias diarias, es decir, aproximadamente una denuncia cada 9 horas.

En el siguiente gráfico se muestra la distribución de las denuncias en la división administrativa principal del país, las regiones.

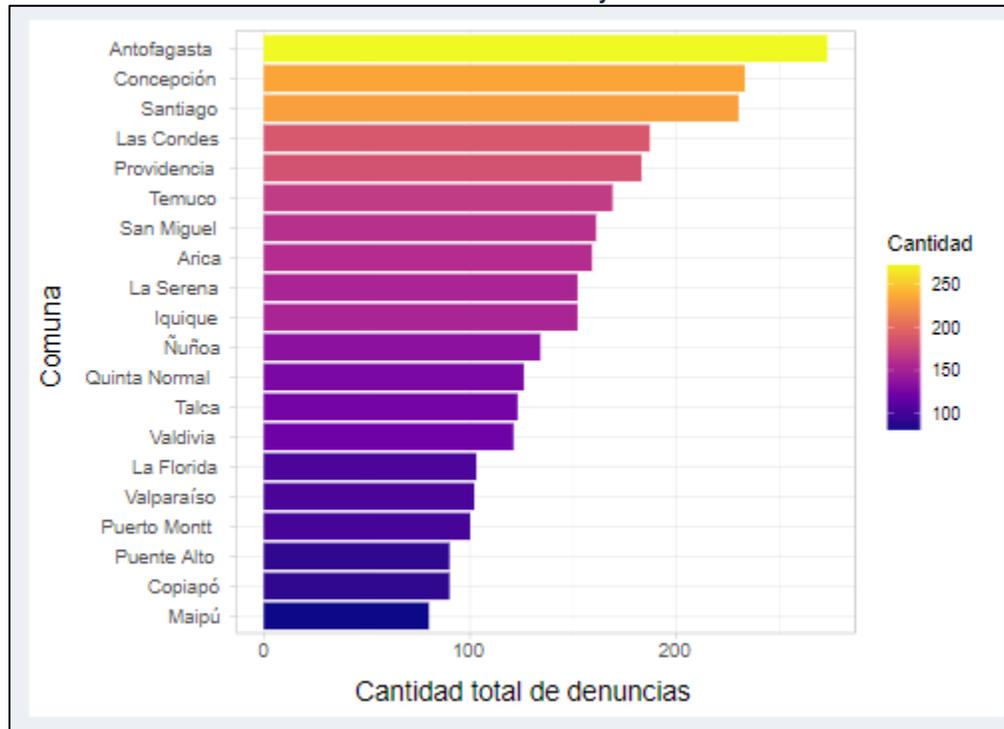
Gráfico 3: Denuncias por Región.



Se observa que las tres regiones con más denuncias corresponden a RM, Biobío y Valparaíso con el 39.4%, 8.9% y 7.9% del total respectivamente.

A continuación se muestran las comunas con mayor concentración de denuncias

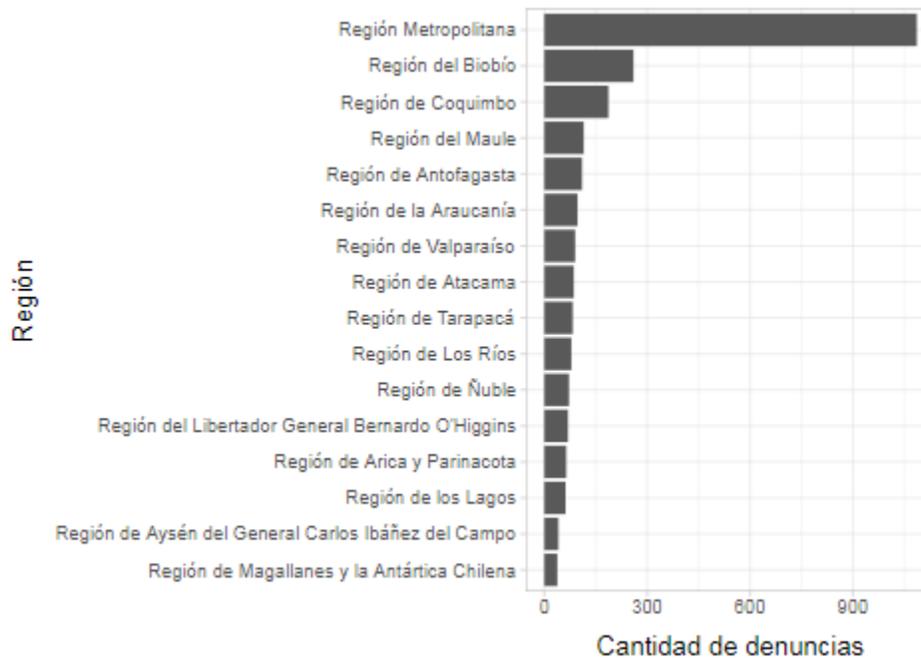
Gráfico 4: Denuncias en las comunas con mayor cantidad de denuncias.



Se observa que las comunas con mayor cantidad de denuncias no se ubican en la Región Metropolitana y son Antofagasta (4.7%) y Concepción (4.0%). Siendo las comunas de Santiago, Las Condes y Providencia la tercera, cuarta y quinta respectivamente.

A continuación se efectúa el mismo análisis pero considerando exclusivamente las denuncias a las que se les relaciona una Unidad Fiscalizable.

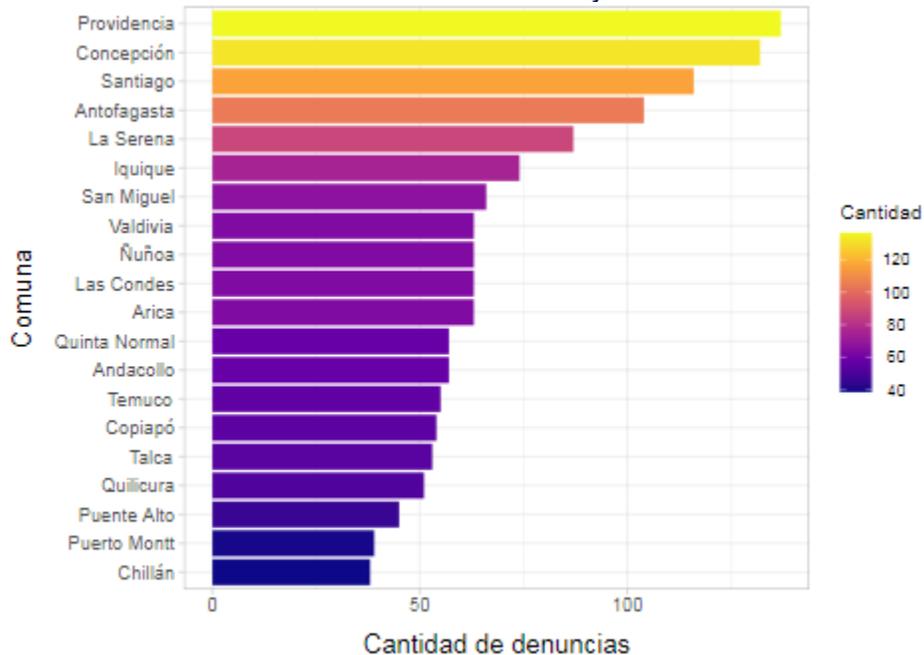
Gráfico 5: Denuncias con UF por Región.



El anterior gráfico muestra solamente las denuncias donde la UF está identificada. Se observa que la Región Metropolitana concentra una proporción muy significativa de las denuncias con el 42.8% de estas. La segunda y tercera corresponden a Biobío (10.2%) y Coquimbo (7.4%) siendo la Región de Magallanes y la Antártica Chilena la última con un 1.5% de las denuncias con UF identificadas.

En el siguiente gráfico se muestra en orden descendente las denuncias con UF identificada por comuna considerando las comunas con una mayor cantidad de denuncias.

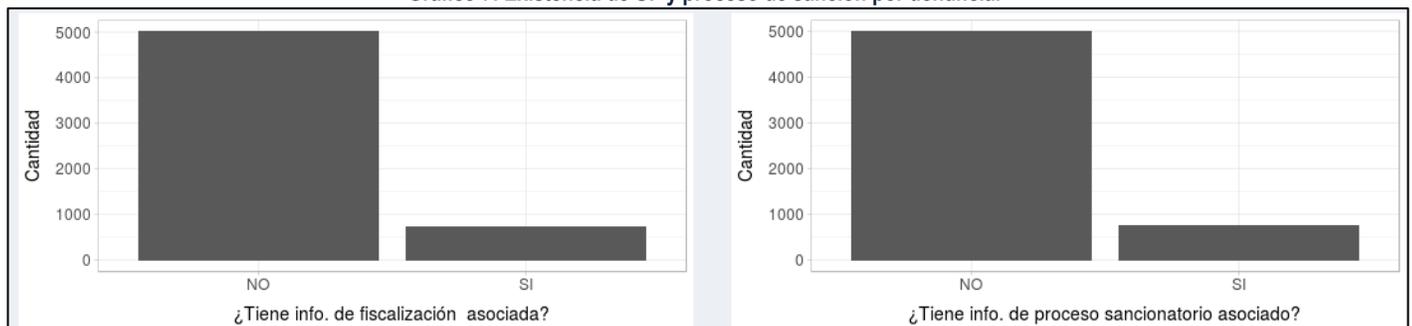
Gráfico 6: Denuncias con UF en las comunas con mayor cantidad de denuncias.



Se observa que las tres comunas con mayor cantidad de denuncias corresponden a Providencia (5.4%), Concepción (5.2%) y Santiago (4.6%). De las 10 primeras, se observa que 6 corresponden a la Región Metropolitana, mientras que las comunas de Antofagasta, La Serena, Iquique y Valdivia cuentan con números de denuncias significativas.

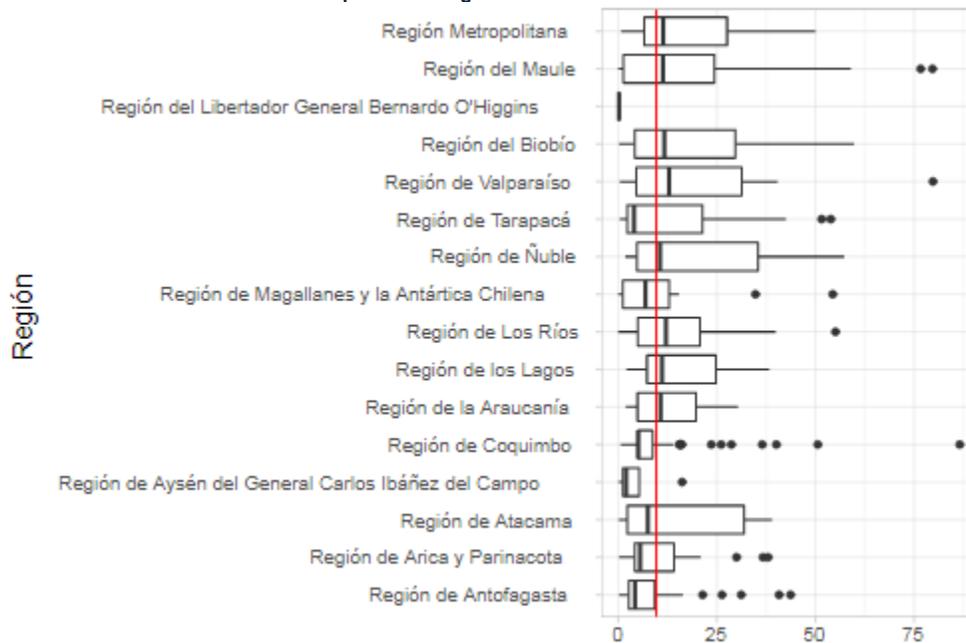
Por otro lado, tal y como se observa en los siguientes gráficos, del total de denuncias menos de un 15% de las denuncias cuentan con una fiscalización o proceso sancionatorio asociado.

Gráfico 7: Existencia de UF y proceso de sanción por denuncia.



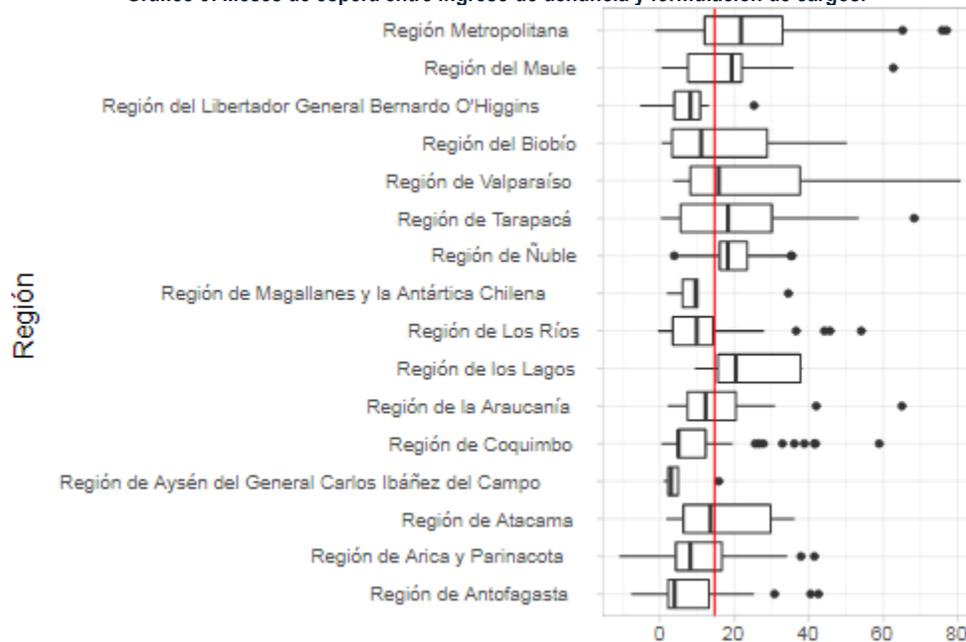
En los siguientes gráficos se muestran los tiempos de demora entre los hitos más importantes del proceso.

Gráfico 8: Meses de espera entre ingreso de denuncia e inicio de fiscalización.



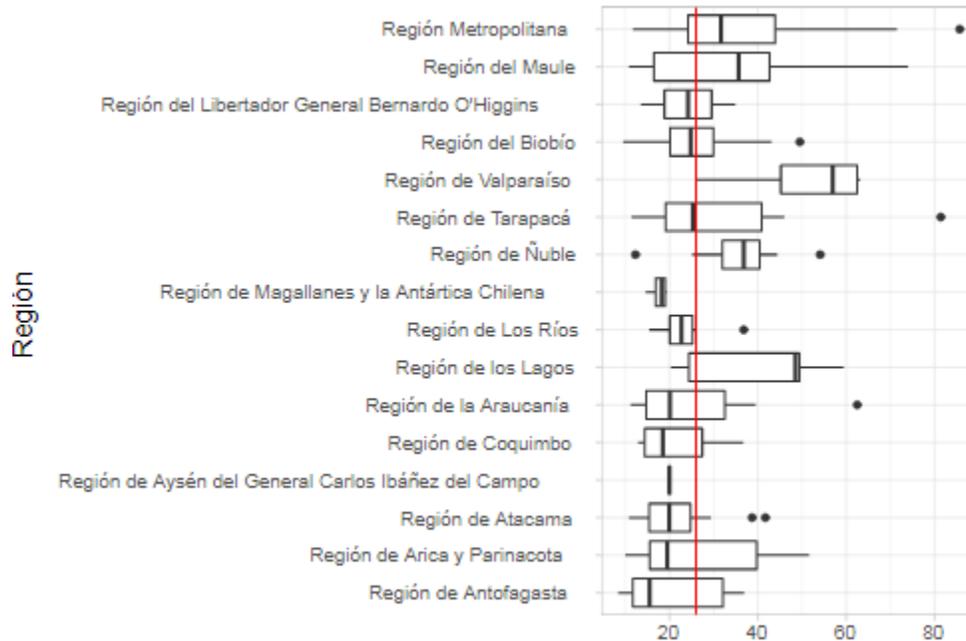
En el gráfico anterior se observa que la mediana del tiempo de espera entre la denuncia y el inicio de la fiscalización considerando todas las regiones es de 9.9 meses.

Gráfico 9: Meses de espera entre ingreso de denuncia y formulación de cargos.



La mediana del tiempo de espera entre denuncia y formulación de cargos es de 14.7 meses, siendo la Región Metropolitana la que tiene la mediana del tiempo de espera más extenso.

Gráfico 10: Meses de espera entre ingreso de denuncia y finalización del proceso de sanción.

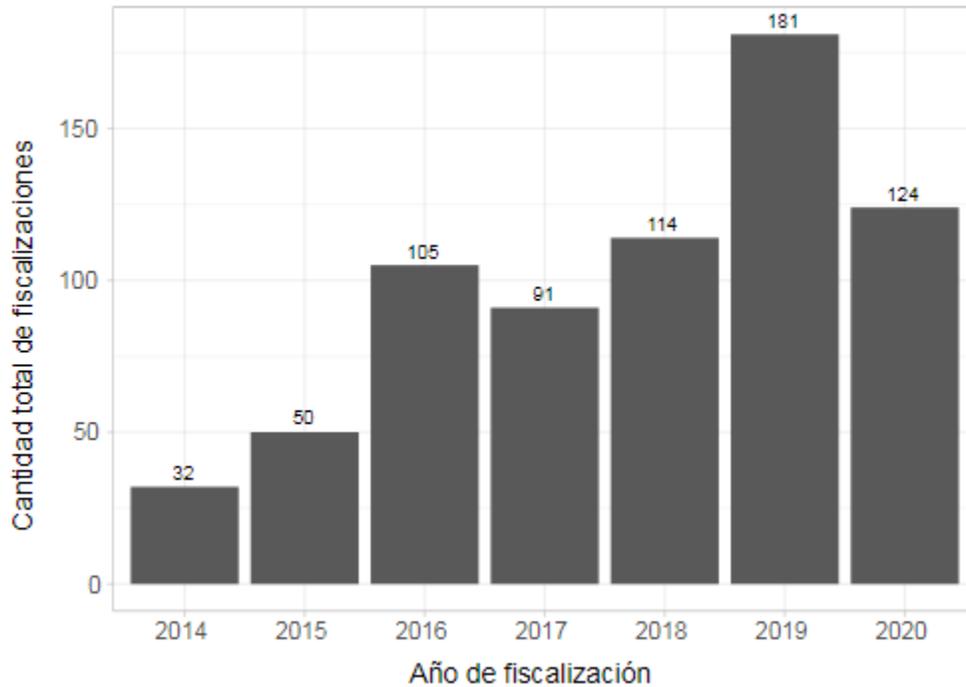


En el gráfico anterior se observa el tiempo transcurrido completo entre la denuncia y la finalización del proceso sancionatorio, la mediana es de 25.9 meses.

5.2 Fiscalizaciones

El total de fiscalizaciones efectuadas en el periodo de estudio es de 697. En el siguiente gráfico se muestra la evolución de las fiscalizaciones anuales realizadas en el territorio nacional desde el 2014 hasta 2020.

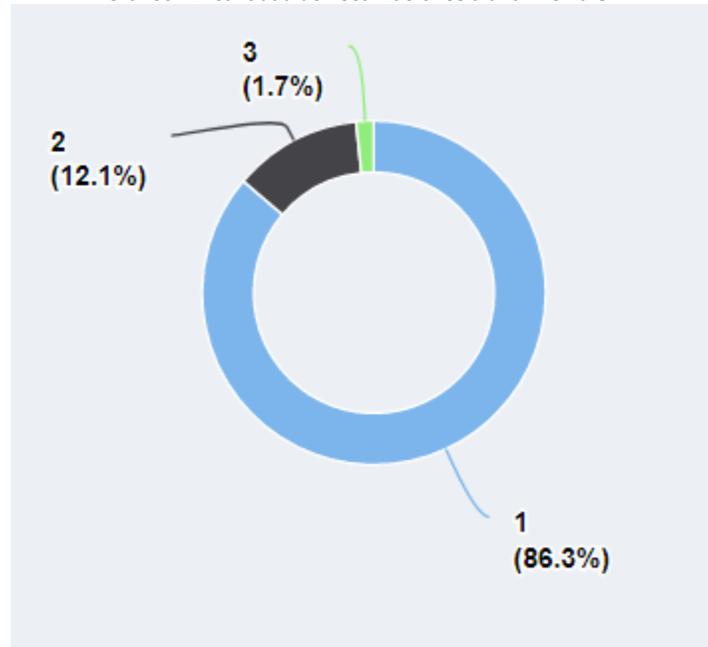
Gráfico 11: Evolución de fiscalizaciones entre 2014 y 2020.



Se observa que el valor más alto se da en el 2019 con 181 fiscalizaciones y el valor más bajo se da en el 2014 con 32 fiscalizaciones.

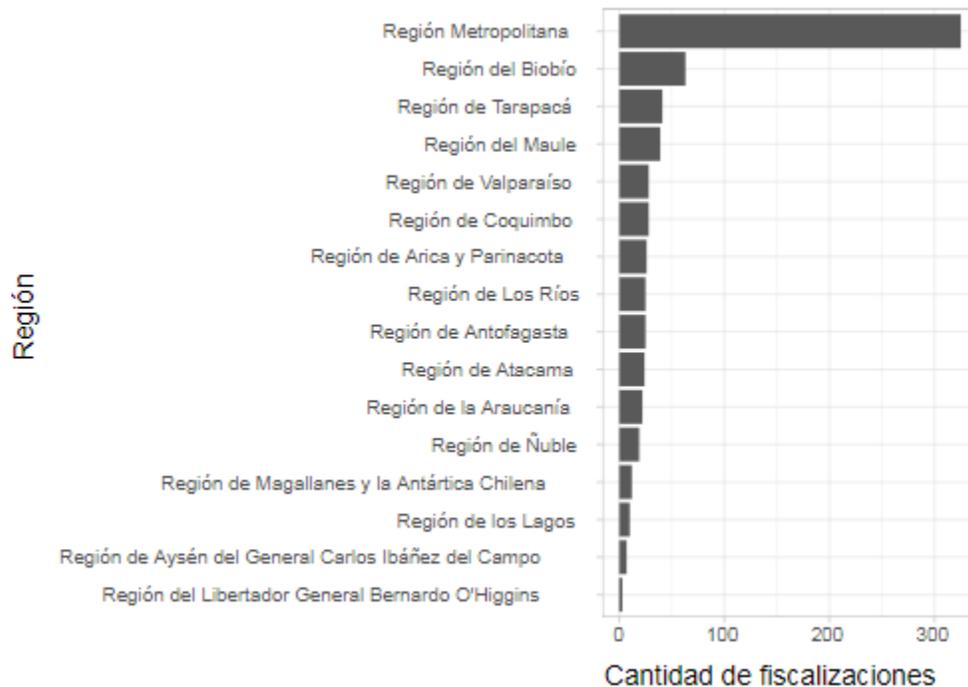
En el siguiente gráfico se muestra la cantidad de fiscalizaciones a una misma UF, en donde se puede observar que 521 UF presentan sólo una fiscalización mientras que menos de 83 UF presentan 2 o más fiscalizaciones.

Gráfico 12: cantidad de fiscalizaciones a una misma UF.



En el siguiente gráfico se muestra la distribución de las fiscalizaciones en las distintas regiones de Chile.

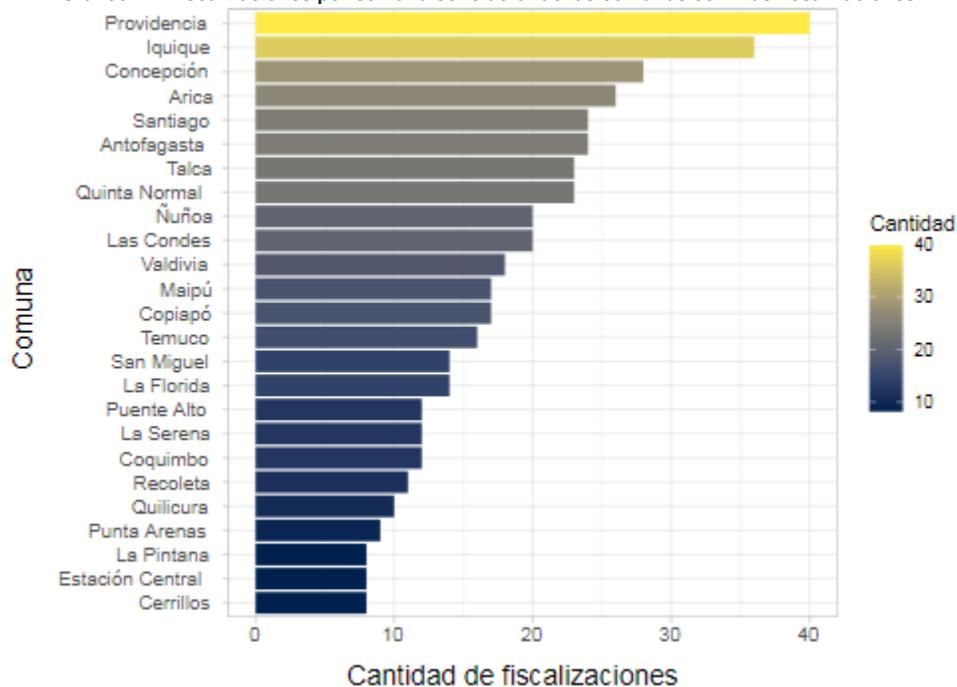
Gráfico 13: Fiscalizaciones por Región.



Se observa que la Región Metropolitana concentra la mayor cantidad de fiscalizaciones realizadas en el país con un 46.7% del total, seguido por la Región del Biobío (9.1%) y la Región de Tarapacá (5.9%). La Región del Libertador General Bernardo O'Higgins presenta el número más bajo de fiscalizaciones con solamente 3, un 0.4% del total.

En el siguiente gráfico se muestra en orden descendente las fiscalizaciones por comuna considerando las comunas con una mayor cantidad de fiscalizaciones.

Gráfico 14: Fiscalizaciones por comuna considerando las comunas con más fiscalizaciones.

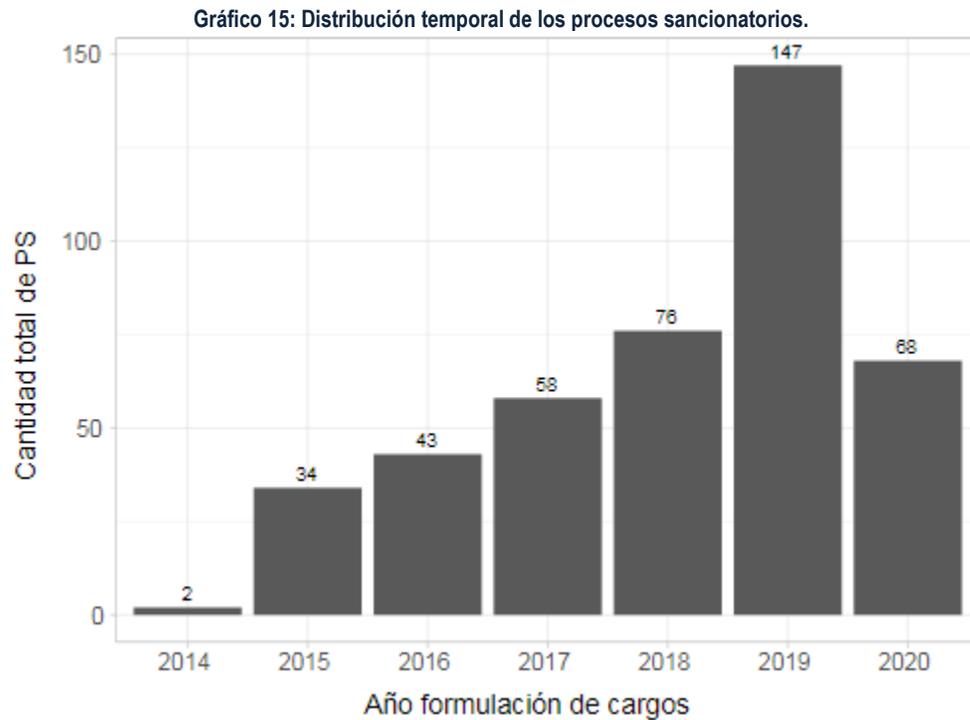


Se observa que las tres comunas con mayor cantidad de fiscalizaciones corresponden a Providencia (5.7%), Iquique (5.2%) y Concepción (4%). De las 10 primeras, se observa que 5 corresponden a la Región Metropolitana.

Por último, se observa que de las 697 fiscalizaciones, el 71.4% tiene asociado un proceso sancionatorio.

5.3 Proceso sancionatorio

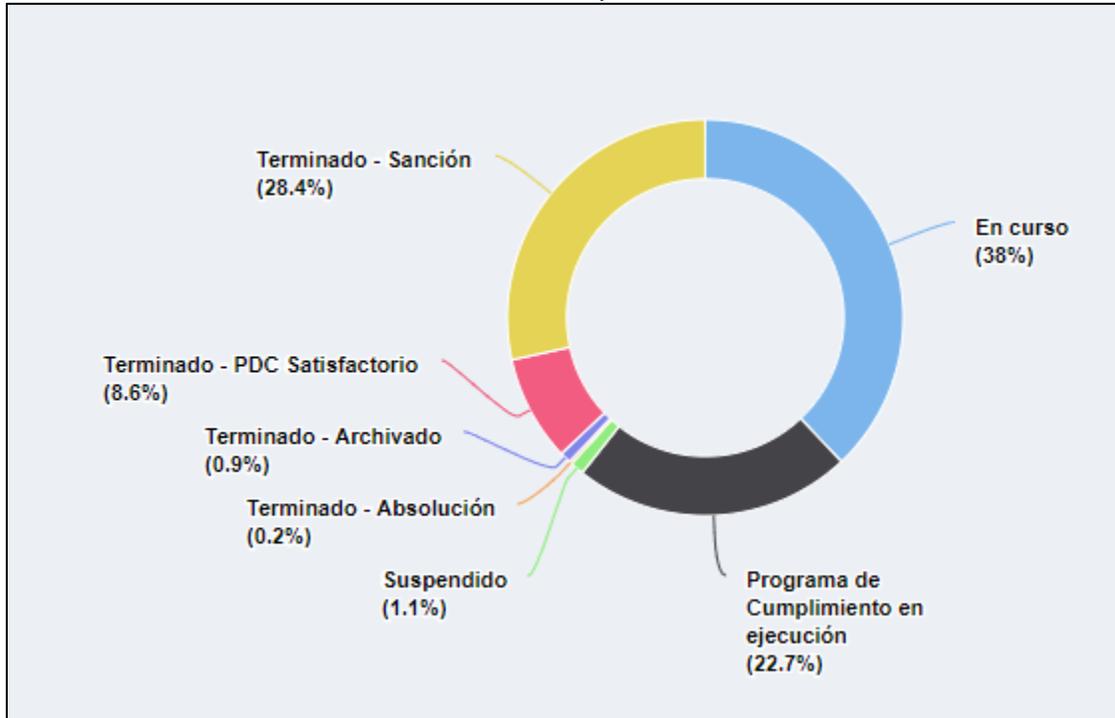
Son 440 los procesos sancionatorios considerados en el presente estudio. En el siguiente gráfico se muestra como se han distribuido desde el 2014 al 2020



Se observa un incremento continuo hasta el 2019, año en el cual se da el valor más alto llegando casi a las 150 formulaciones de cargos.

En el siguiente gráfico se muestra el estado actual (octubre de 2020) de los procesos sancionatorios.

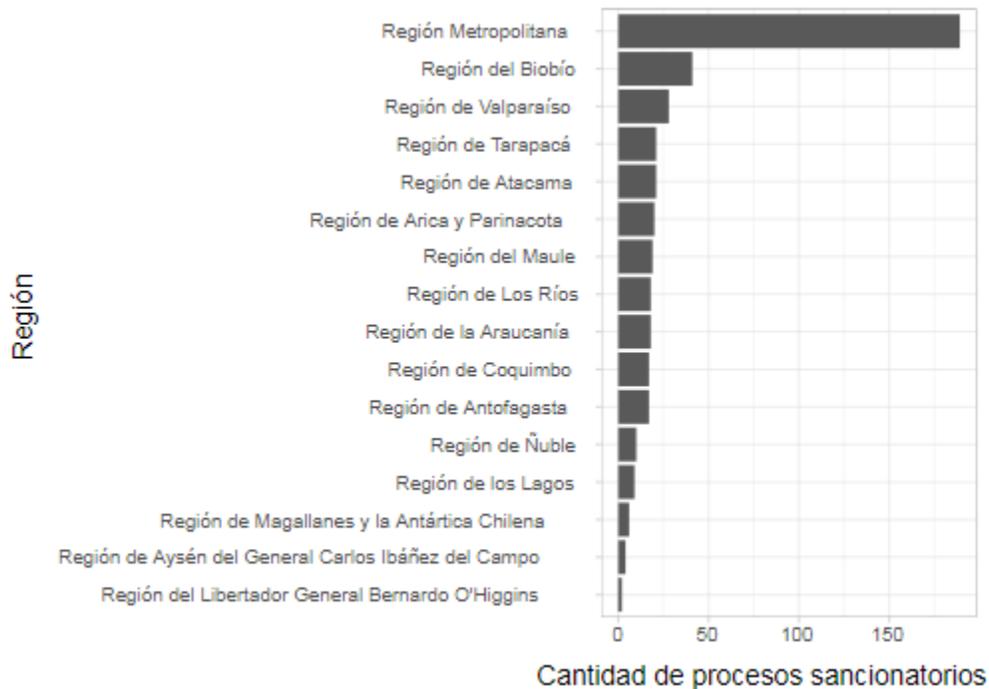
Gráfico 16: Estado actual de los procesos sancionatorios.



Se observa que la mayoría de los procesos sancionatorios se encuentran en curso (38%), seguido por aquellos que han terminado en sanción (28%), mientras que el 8.6% de los procesos han terminado con un PDC.

En el siguiente gráfico se muestra la distribución de procesos sancionatorios en las distintas regiones de Chile.

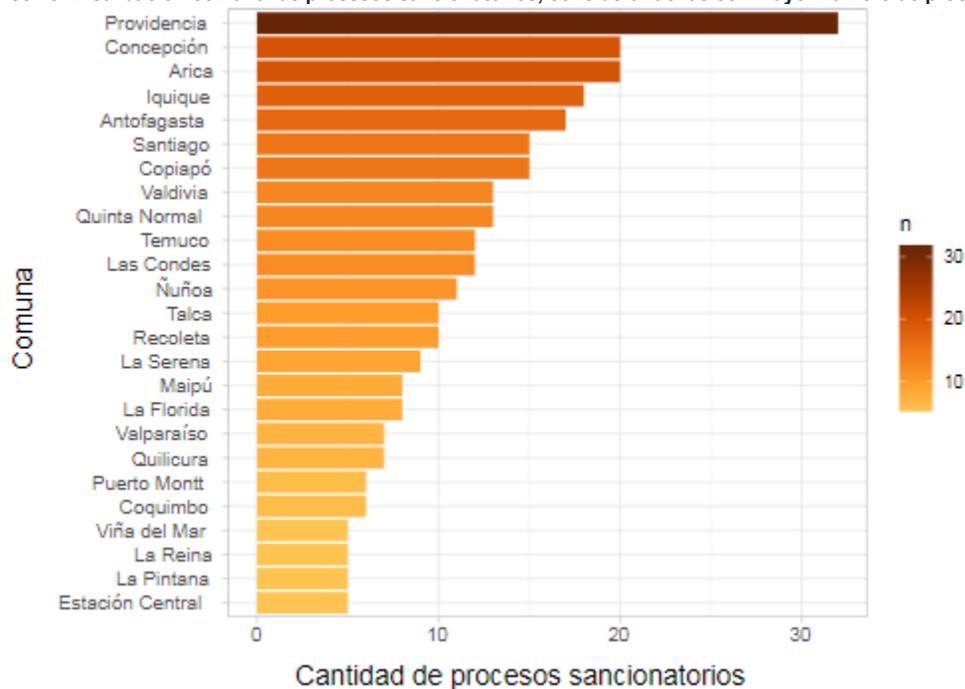
Gráfico 17: Distribución regional de los procesos sancionatorios.



Se observa que la Región Metropolitana concentra la mayor cantidad de procesos sancionatorios realizados en el país (46%), seguido por la Región del Biobío (9.3%) y la Región de Valparaíso (6.4%).

En el siguiente gráfico se muestran los procesos sancionatorios por comuna considerando las comunas con una mayor cantidad de procesos sancionatorios.

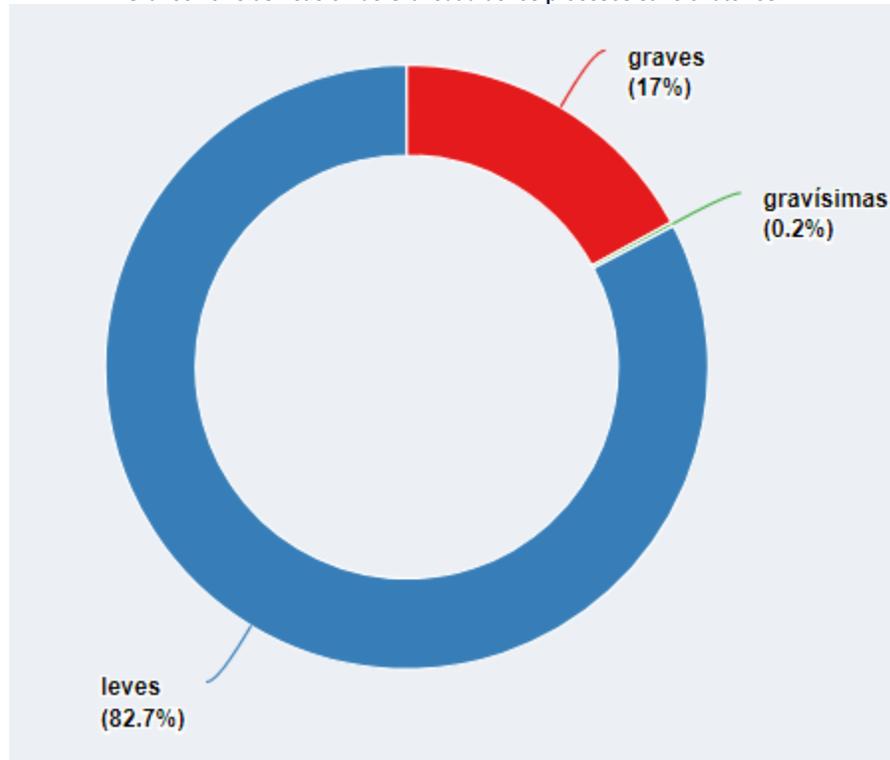
Gráfico 18: Distribución comunal de procesos sancionatorios, considerando las con mayor número de procesos.



Se observa que las tres comunas con mayor cantidad de denuncias corresponden a Providencia (7.3%), Concepción (4.5%) y Arica (4.5%). De las diez primeras, se observa que tres corresponden a la Región Metropolitana.

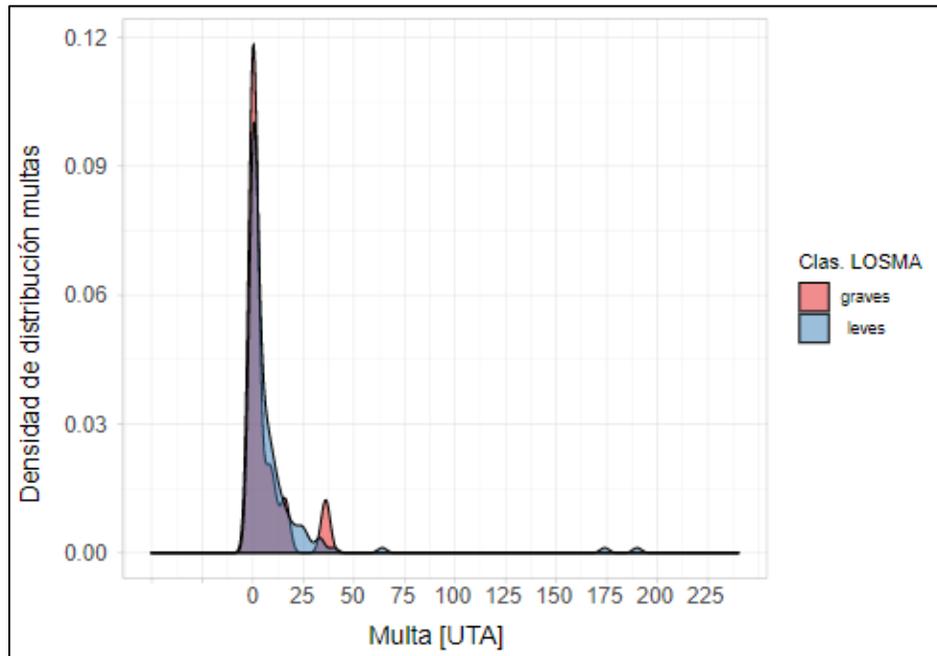
En el siguiente gráfico se observa que la mayoría de los procesos sancionatorios han sido categorizados como leves (82.7%), mientras que los que han sido categorizados como graves corresponden al 17%. Las gravísimas, muy escasas, componen el 0.2% de los casos.

Gráfico 19: Clasificación de Gravedad de los procesos sancionatorios.



En el siguiente gráfico se muestra el rango de valores asociados a los procesos sancionatorios graves y leves.

Gráfico 20: Densidad de distribución de valores de multa en UTA.

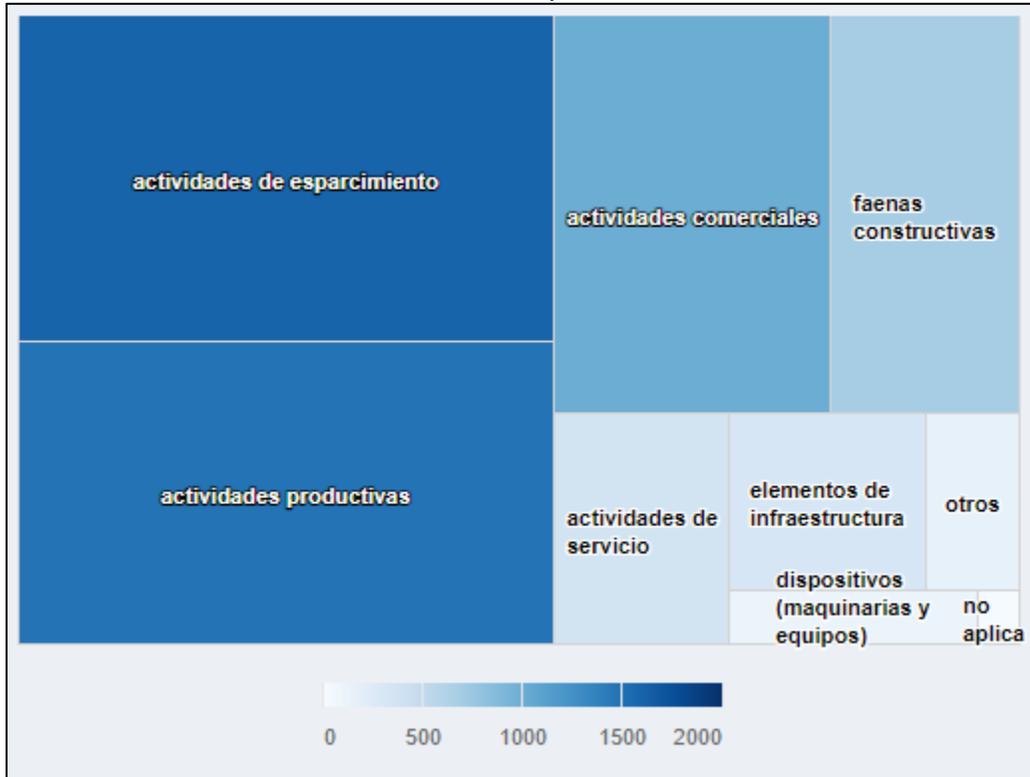


Se observa que la mayoría de los procesos sancionatorios tanto los clasificados leves como graves, el valor asociado a su multa es cercano a 0 [UTA], también se aprecia que las multas de mayor valor están asignadas a procesos sancionatorios leves. Por lo tanto, la clasificación de la gravedad de la infracción no indica necesariamente el valor de la multa. Por último, se menciona que la suma total de las multas da por resultado 1357 UTA, que convertido a peso chileno utilizando el valor de noviembre de 2020 correspondiente a 608.088 pesos cada una nos da una suma de 825.175.416 pesos.

5.4 Actividad económica

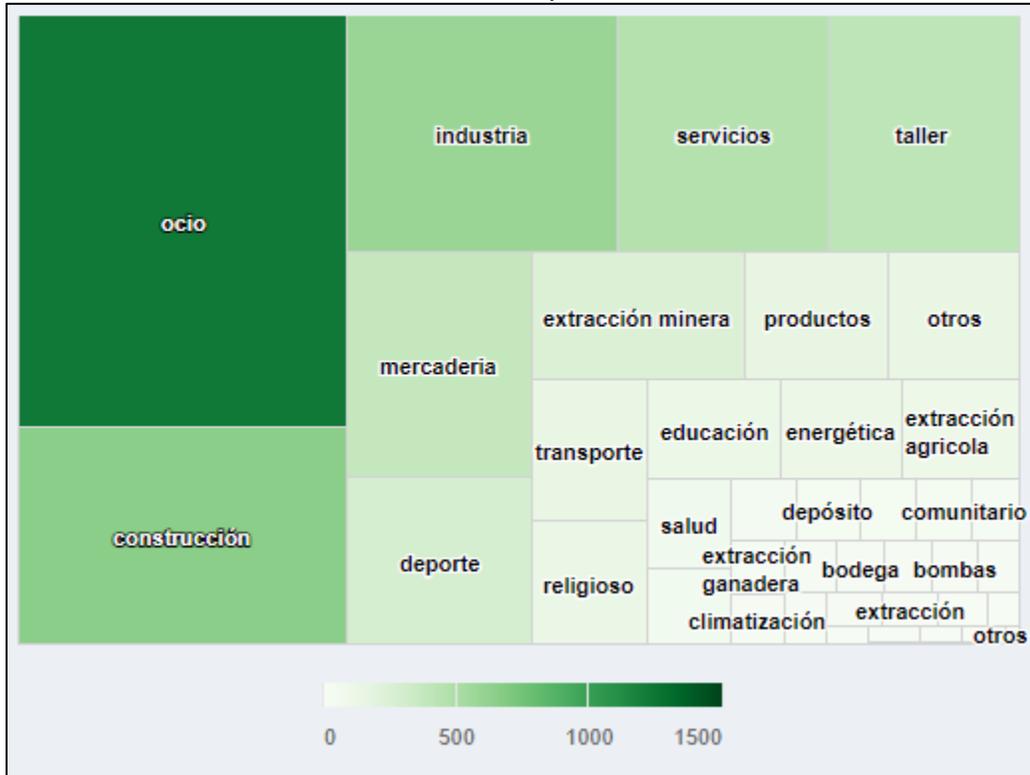
Las actividades económicas están divididas en tres niveles tal como se indica en el capítulo 4 Propuesta de Categoría de Actividad Económica. En la siguiente ilustración se muestran las divisiones del primer nivel. El área de los rectángulos indican su jerarquización de cantidad de denuncias.

Gráfico 21: Distribución de denuncias por actividad económica. Nivel 1.



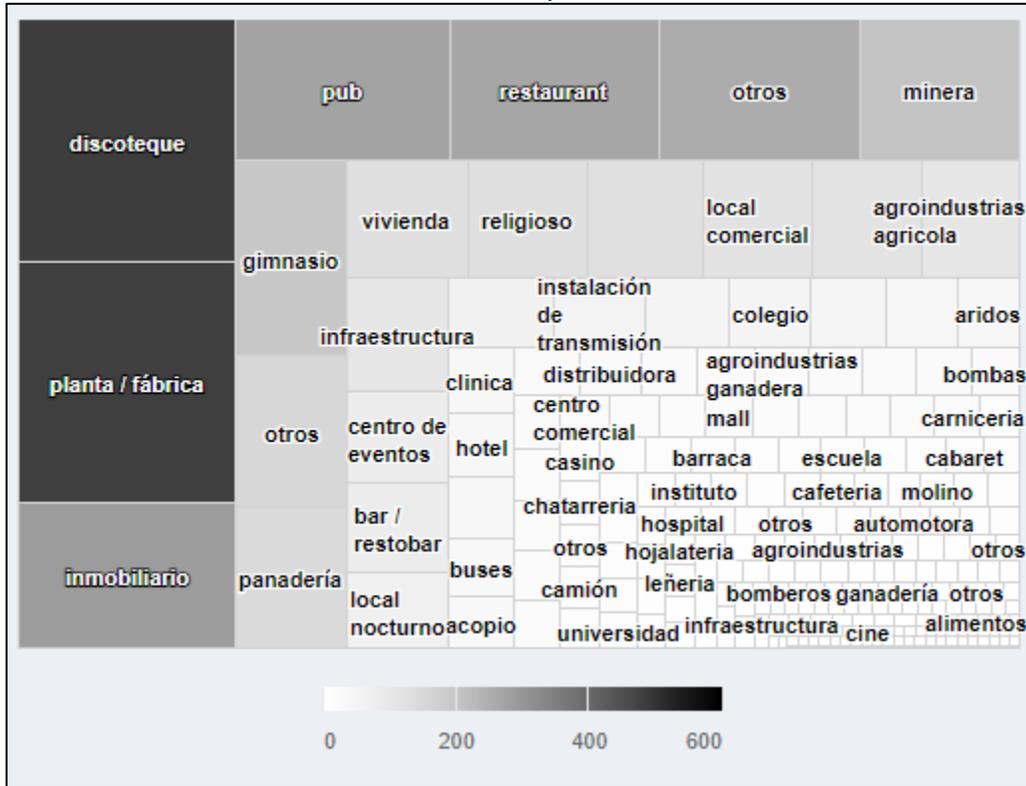
En el gráfico anterior se observa que actividades de esparcimiento es la actividad nivel 1 que presenta la mayor cantidad de denuncias, siendo un total de 1595 representando el 27.7% del total de este nivel, seguido por actividades productivas con 1482 denuncias representando un 25.7% y actividades comerciales con 1004 denuncias representando un 17.4%.

Gráfico 22: Distribución de denuncias por actividad económica. Nivel 2.



En el gráfico anterior se observa que ocio es la actividad nivel 2 que presenta la mayor cantidad de denuncias, con un total de 1234 representando el 77.4% del total del nivel 1 (actividades de esparcimiento), seguido por construcción con 654 denuncias representando un 94.4% del nivel 1 (faenas constructivas) e industria con 586 denuncias representando un 39.5% de su nivel 1 (actividades productivas).

Gráfico 23: Distribución de denuncias por actividad económica. Nivel 3.

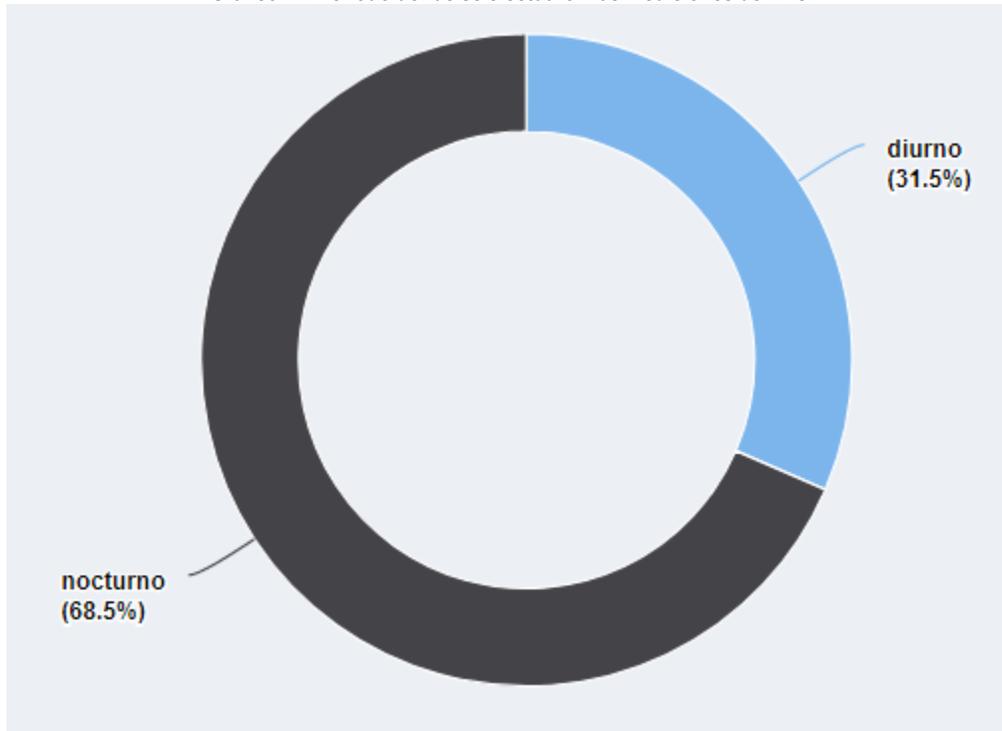


En el gráfico anterior se observa que discoteque es la actividad nivel 3 que presenta la mayor cantidad de denuncias, con 485 representando el 39.3% del total del nivel 2 respectivo (ocio), seguido por planta / fábrica con 481 denuncias representando un 82.1% de su nivel 2 (industria) e inmobiliario con 288 denuncias representando un 44% de su nivel 2 (construcción).

5.5 Mediciones

Las mediciones analizadas fueron 440, de las cuales a 438 se le encontró información con respecto a en qué periodo fueron realizadas. 300 de estas fueron evaluaciones nocturnas y las restantes 138 evaluaciones diurnas.

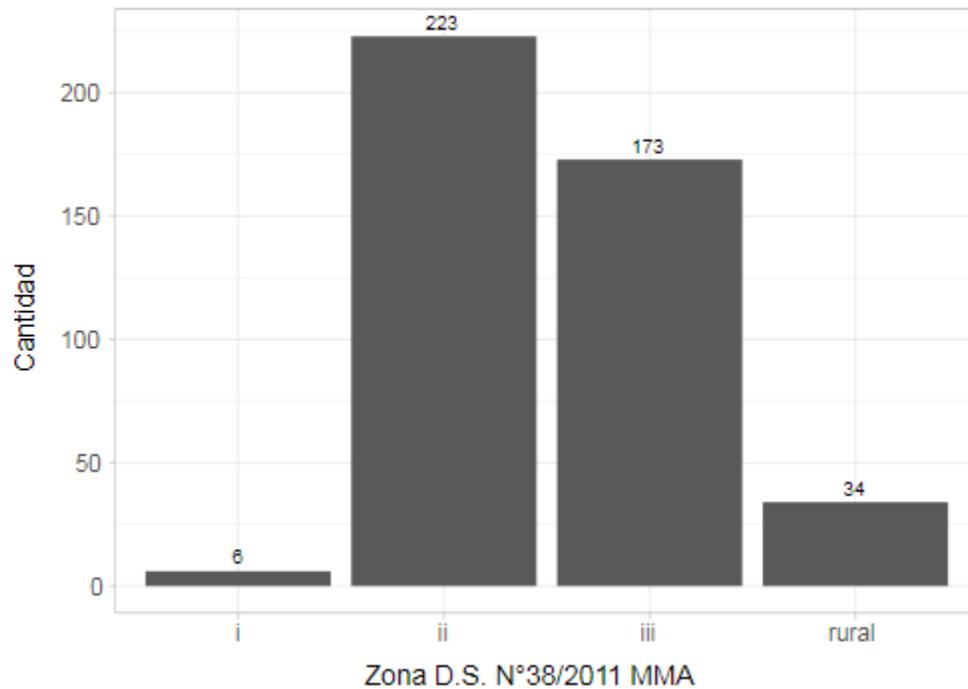
Gráfico 24: Periodo donde se efectuaron las mediciones de NPC.



En el gráfico anterior se observa que la cantidad de mediciones efectuadas en horario nocturno con un 68.5% del total sobrepasan en cantidad a las realizadas en horario diurno (31.5%).

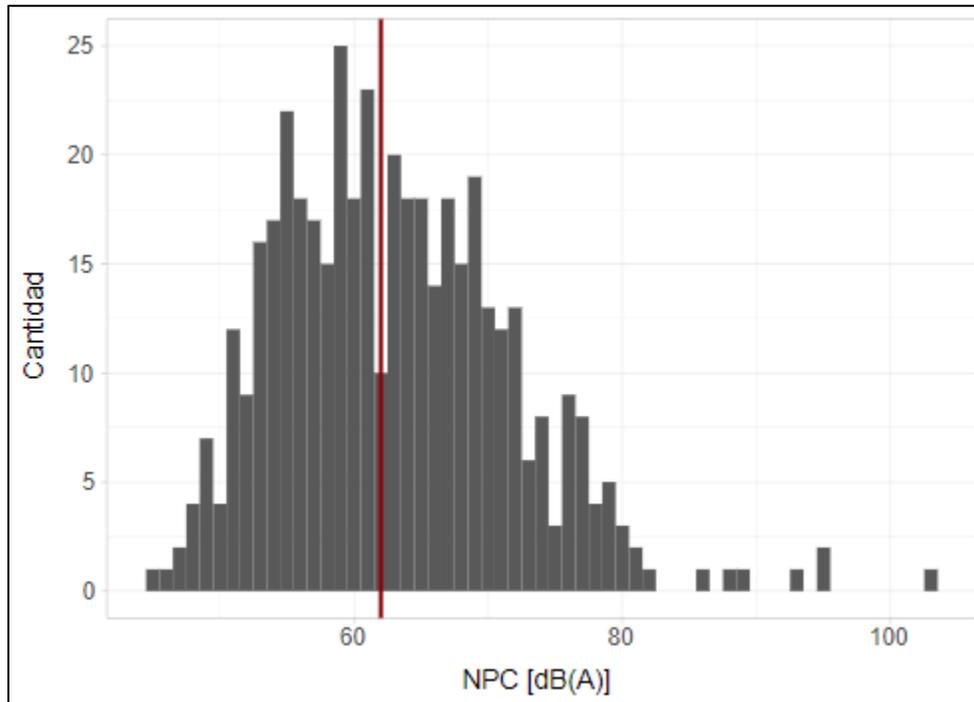
En el siguiente gráfico se muestra la cantidad de mediciones que se han realizado por zona. Se puede observar que la mayoría de las mediciones se han llevado a cabo en zona II con un 51.1% del total y por el contrario zona I (1.4%) es la que presenta menor cantidad de mediciones.

Gráfico 25: Zona según el DS38 en el que se efectuaron las mediciones de NPC.



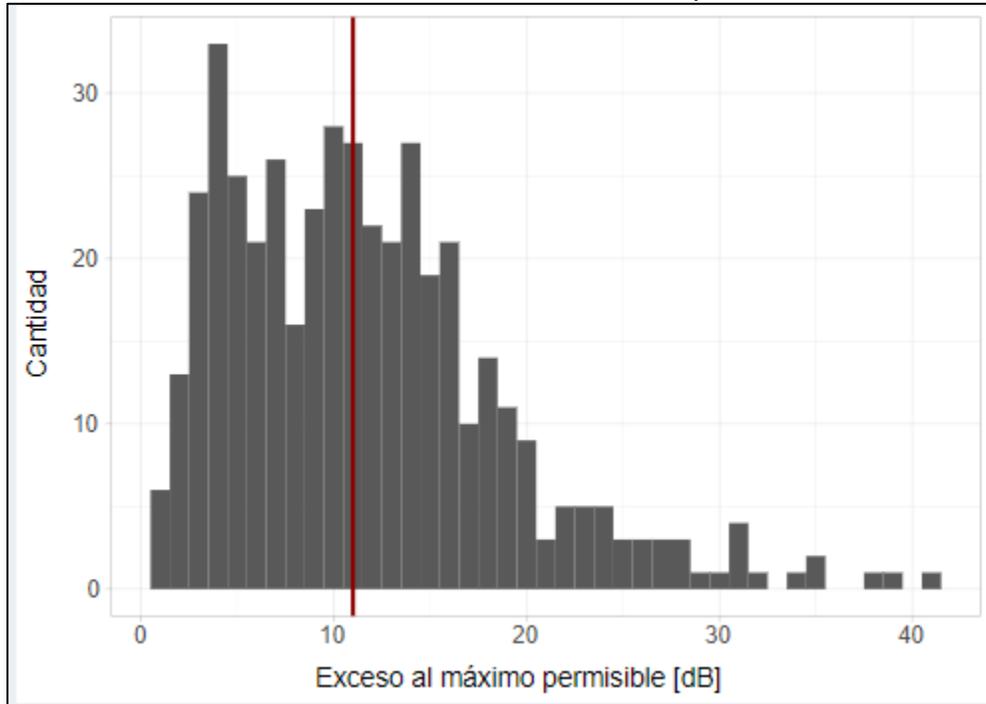
Tal como se observa en el siguiente gráfico la mayoría de los resultados de las mediciones de NPC se concentran entre los 55 y 70 [dB(A)], se observa también que las mediciones con mayor nivel están por sobre los 100 [dB(A)]. Además, se aprecia que la mediana de las mediciones se encuentra en los 62 [dB(A)].

Gráfico 26: Distribución de los NPC medidos.



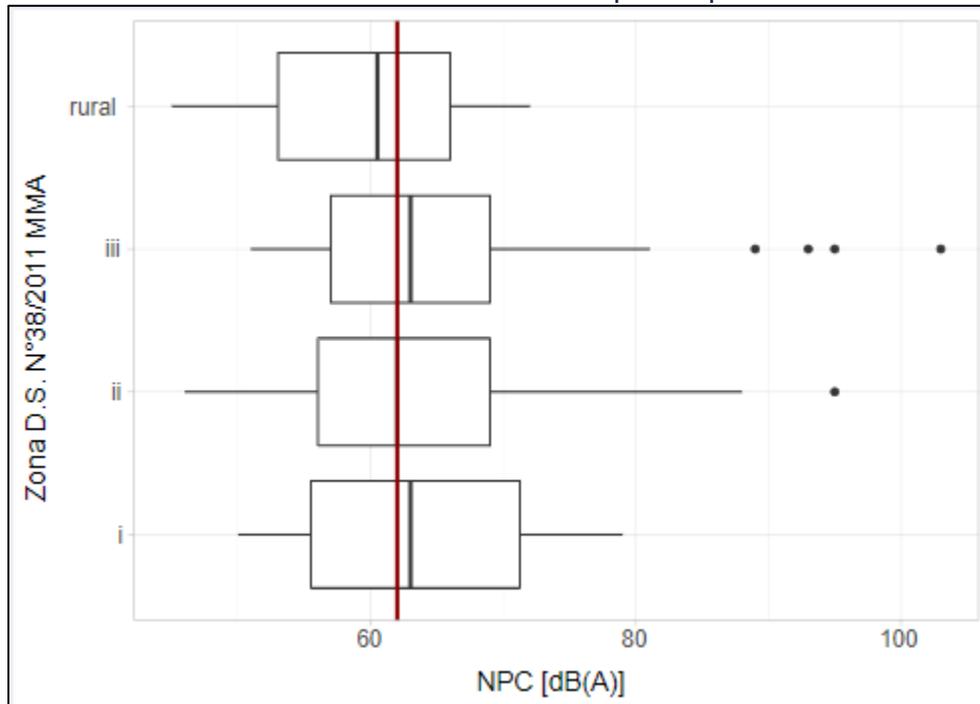
En los siguientes gráficos se muestran los excesos al máximo permisible.

Gráfico 27: Distribución de los excesos al máximo permitido.



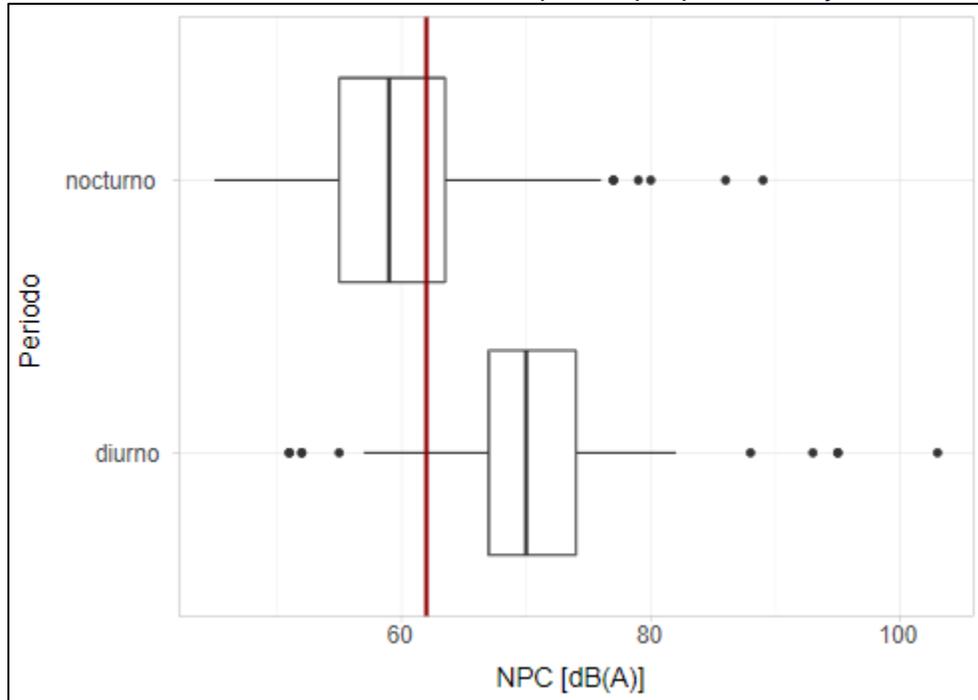
Se observa que el valor que más se repite en cuanto a exceso es 4 [dB], siendo más de 30 los casos en que la superación al máximo permisible corresponde a ese valor, además se aprecia que la mediana del exceso al máximo permisible es 11 [dB]

Gráfico 28: Distribución de los excesos al máximo permitido por zonas DS38.



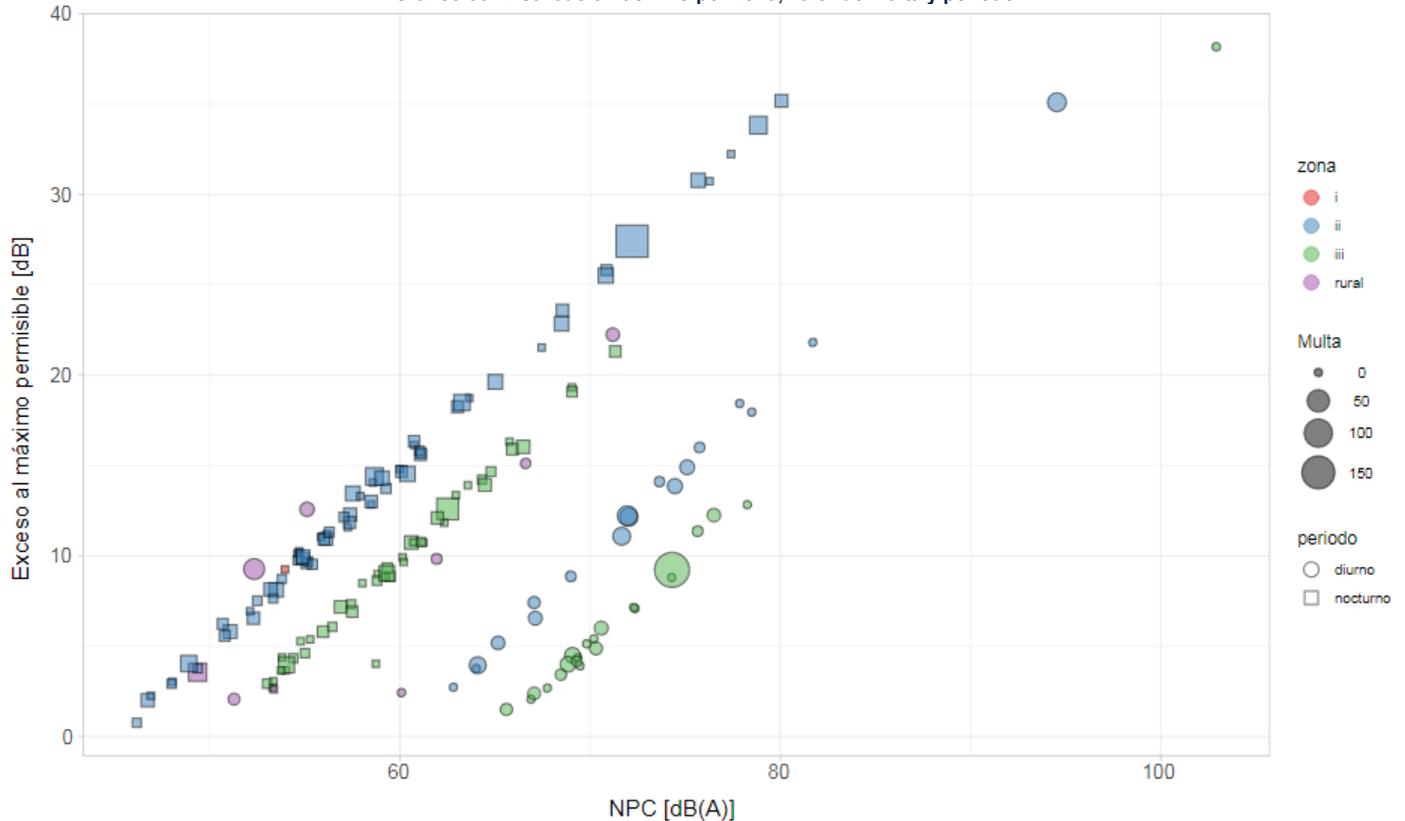
En el gráfico anterior se puede observar que la mediana del NPC para todas las zonas está por sobre los 60 [dB(A)], mientras que la mediana general para todas las mediciones se encuentra en los 62 [dB(A)].

Gráfico 29: Distribución de los excesos al máximo permitido para periodo diurno y nocturno.



En el anterior gráfico se puede apreciar que la mediana para los NPC en horario nocturno es de 59 [dB(A)], mientras que para el horario diurno es 70 [dB(A)], además se aprecia que la mediana general para todas las mediciones se encuentra en los 62 [dB(A)].

Gráfico 30: Distribución de NPC por zona, valor de multa y periodo.

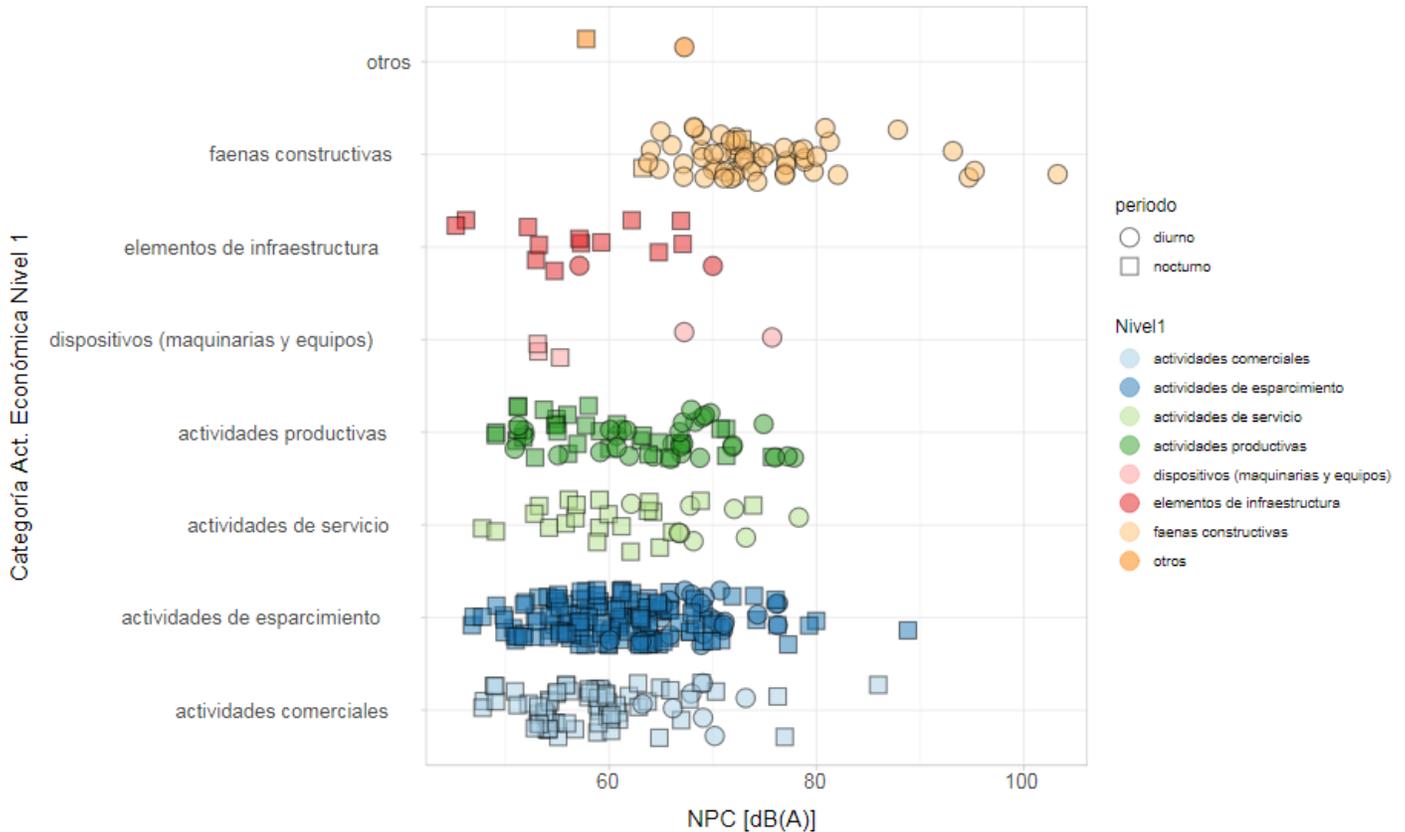


En el gráfico anterior se puede observar que las zonas que presentan mayor cantidad de excesos al máximo permisible son las zonas II y III, donde se aprecian sus respectivas diagonales separadas según el periodo. Los mayores excesos corresponden principalmente a Zona II, especialmente en periodo nocturno.

La multa de mayor valor (190 [UTA]) está dada para un exceso menor a 10 [dB] registrado en zona II. No se aprecia una correlación directa entre el NPC o el exceso registrado con el valor de la multa.

En el siguiente gráfico se muestran los distintos niveles en [dB(A)] registrados para cada actividad económica Nivel 1, tanto en periodo diurno como período nocturno.

Gráfico 31: : Distribución de NPC por actividad económica y periodo.



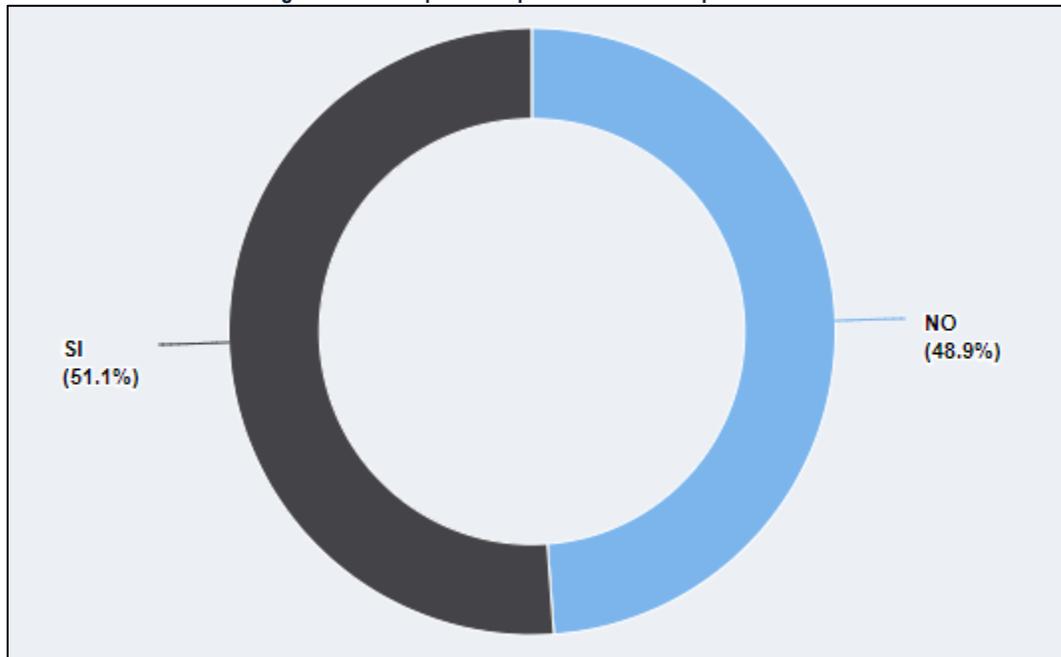
Se puede observar que para actividades productivas, actividades de servicio, actividades de esparcimiento y actividades comerciales los niveles se concentran entre 50 y 70 [dB(A)].

Las faenas constructivas se concentran entre 60 y 80 [dB(A)] obteniéndose los niveles más altos superando el NPC de 90 [dB(A)]. Para elementos de infraestructura se concentran entre 40 y 70 [dB(A)], mientras que para dispositivos (maquinarias y equipos) y Otros presentan mayor dispersión en los datos. También se observa un predominio de evaluaciones en periodo diurno para las faenas constructivas y nocturno para las actividades de esparcimiento y elementos de infraestructura.

5.6 PDC

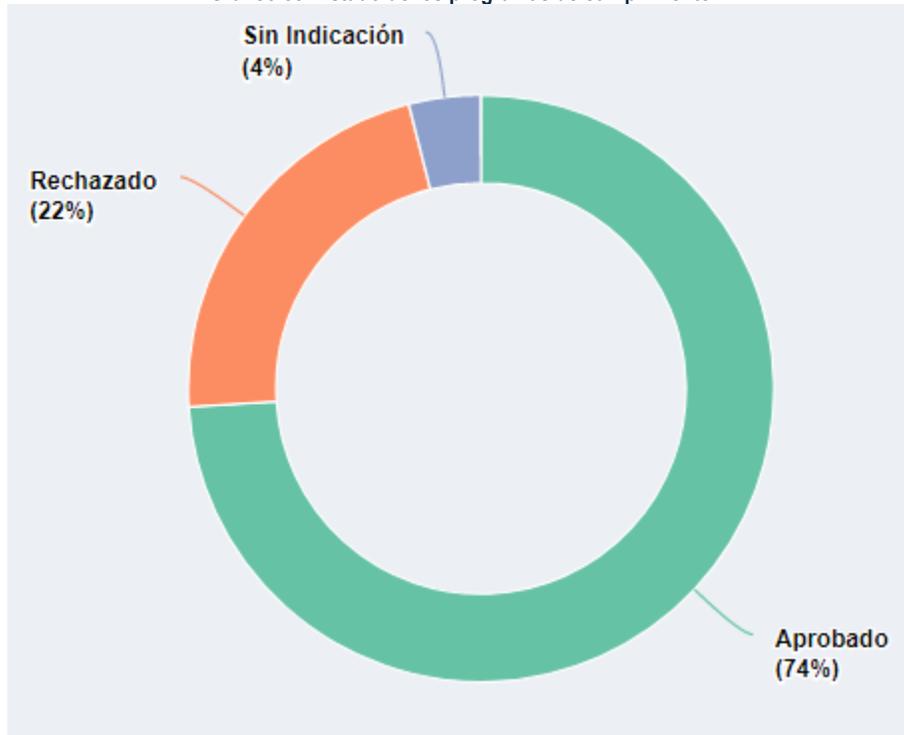
La cantidad de Programas de cumplimiento presentados en los procesos sancionatorios en el transcurso del 2014 al 2020 fueron 225 tal como se muestra en el siguiente gráfico, en donde se observa que representa al 51.1% del total, sin embargo, el porcentaje de aquellos que no presentan (PDC) es similar al de los que sí lo hacen siendo este un 48.9%.

Gráfico 32: Programas de cumplimiento presentados en los procesos sancionatorios.



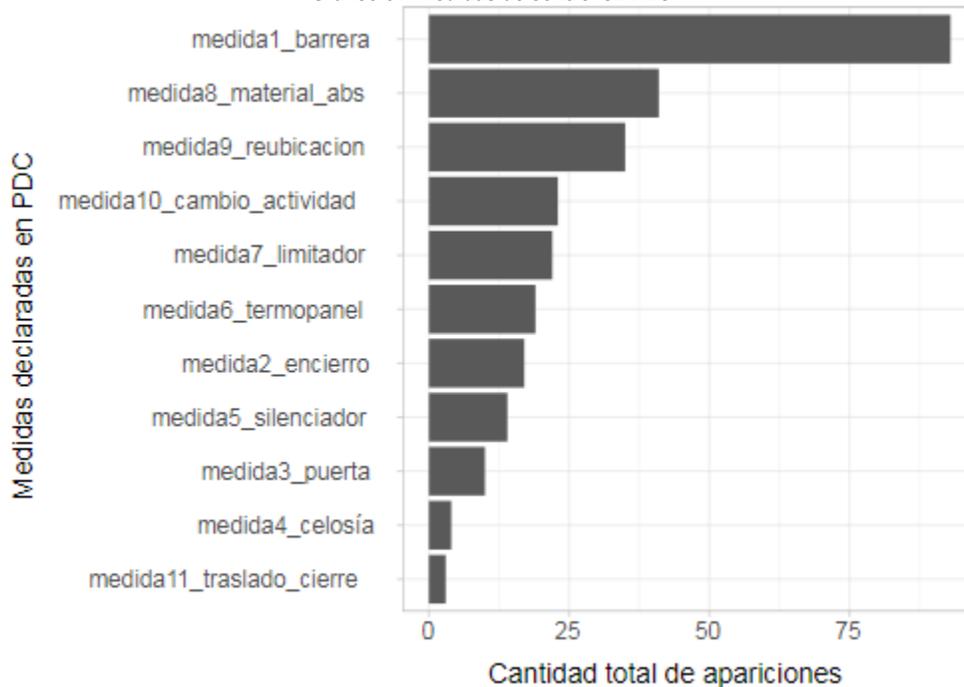
Como se observa en el siguiente gráfico, la mayoría, siendo el 74% del total de los PDC presentados, se encuentran actualmente aprobados, mientras que la minoría se encuentran sin indicación (4%).

Gráfico 33: Estado de los programas de cumplimiento.



El siguiente gráfico muestra las distintas medidas con las cuales se intentará controlar el ruido generado declaradas en los PDC junto a su cantidad total de apariciones.

Gráfico 34: Medidas de control en PDC.

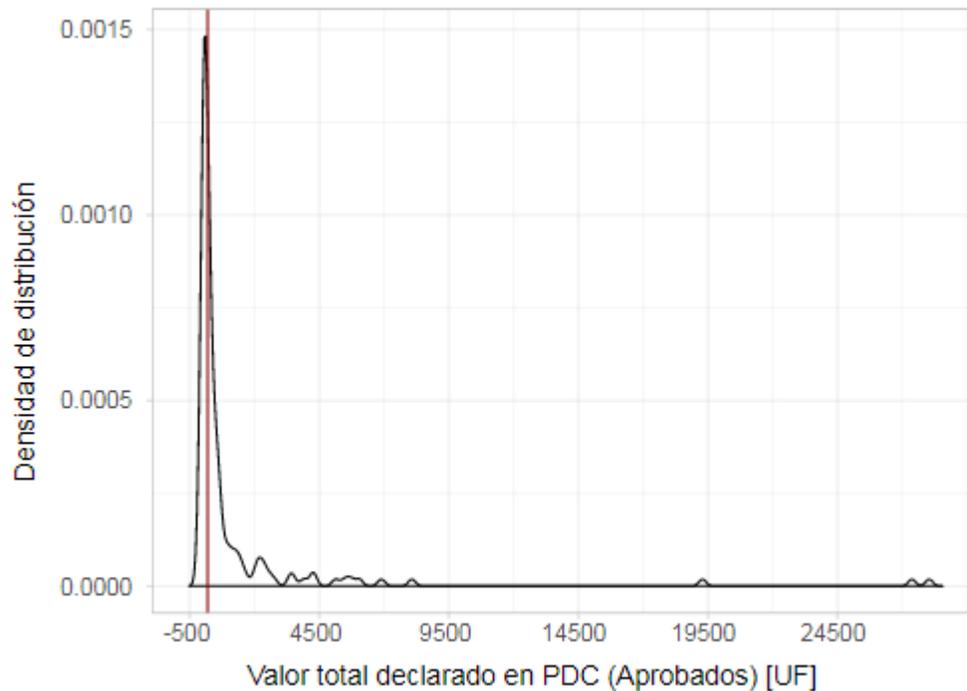


Se observa que la medida más declarada es la “barrera acústica” con un 33.1% de apariciones, seguida de “recubrimiento con material de absorción paredes, piso o techumbre” (14.6%) y la “reubicación de equipos o maquinaria generadora de ruido” (12.5%), por otro lado, la medida menos declarada es el “traslado o cierre de la unidad fiscalizable” con un 1.1% de apariciones.

También existe una medida número 12 correspondiente a “otros”, en donde se explicitan las medidas a realizar en caso de no corresponder a ninguna de las 11 anteriores. 127 casos (57%) indicaron al menos una medida en esta categoría, sin embargo, al ser declarada de forma escrita la variedad de casos es bastante amplia. Lo más recurrente encontrado fue la capacitación de personal.

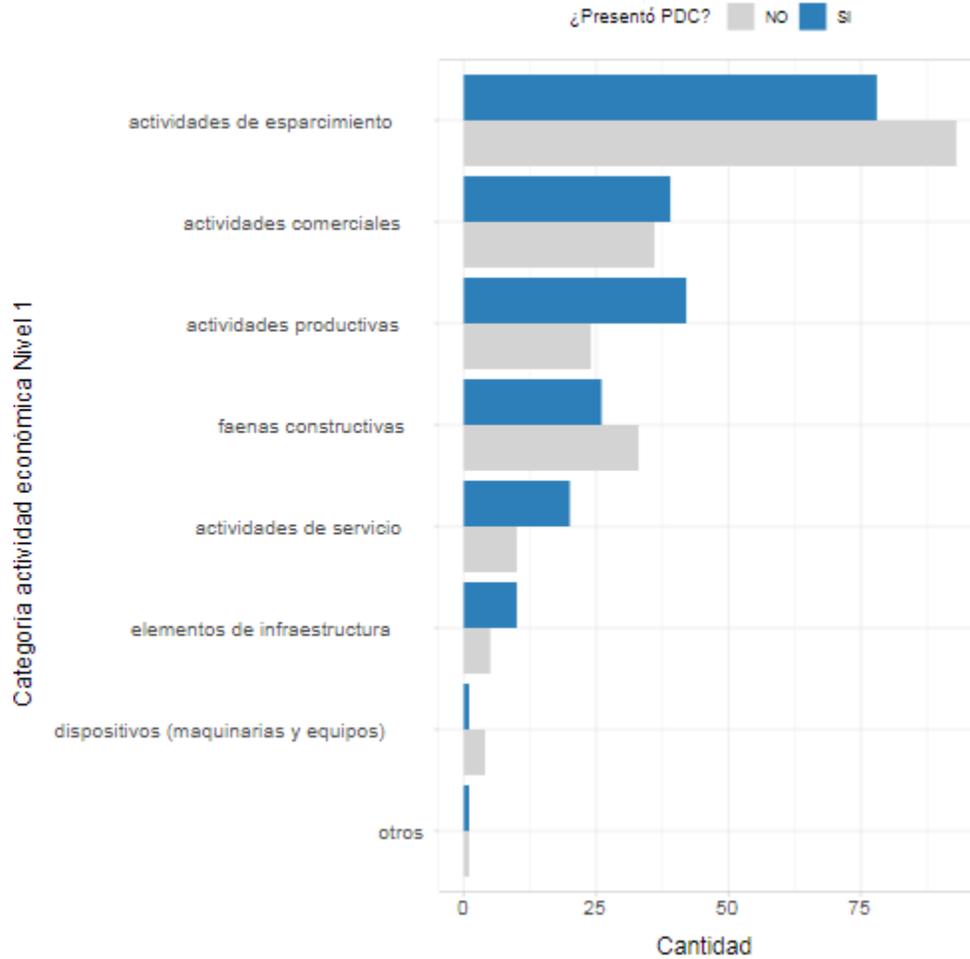
El siguiente gráfico muestra los distintos valores declarados en los PDC aprobados, la mediana de los valores es 192.5 UF y que los valores más elevados superan las 27000 UF.

Gráfico 35: Densidad de distribución de Valores de PDC.



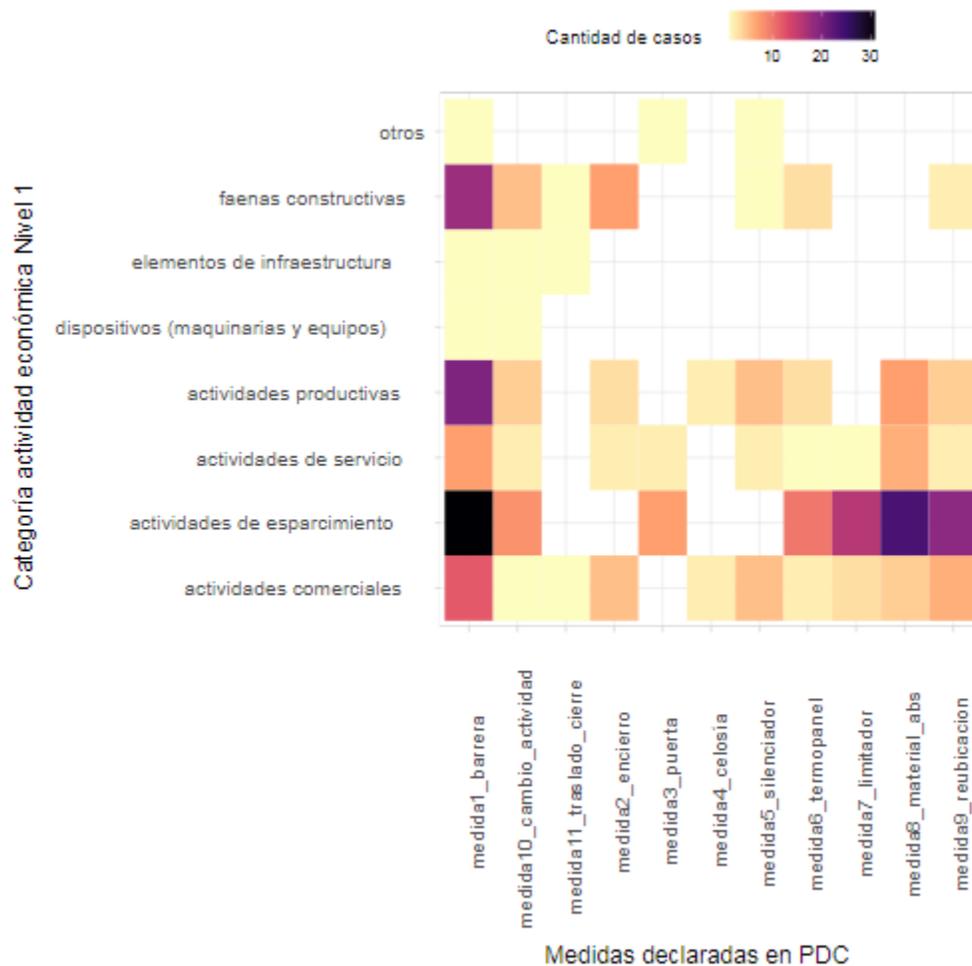
Tal como muestra el siguiente gráfico, las UF relacionadas a las actividades económicas de nivel 1 actividades de servicio y elementos de infraestructura son las que muestran mayor porcentaje de PDC presentados del total asignado a su nivel, siendo estos porcentajes 66.7% cada uno, sin embargo, la actividad que cuenta mayor cantidad de PDC presentados es actividades de esparcimiento, con 171 PDC correspondiente al 45.6% del total asignado a su nivel.

Gráfico 36: Existencia y número de PDC por Actividad Económica.



El siguiente gráfico muestra la cantidad de medidas declaradas en PDC según actividad económica de nivel 1.

Gráfico 37: Concentración de medidas de control por Actividad Económica.



Se observa en el gráfico anterior que para (1) faenas constructivas la medida más utilizada es la barrera acústica, 48.6% de apariciones en PDC de UF relacionadas a este nivel, para (2) elementos de infraestructura las medidas más utilizadas son barrera acústica, cambio en la actividad y traslado o cierre de la unidad fiscalizable, para cada una de ellas un 33.3% de apariciones en PDC de UF relacionadas a este nivel, para (3) dispositivos (maquinaria y equipos) las medidas más utilizadas son barrera acústica y cambio en la actividad, para cada una de ellas un 50% de apariciones en PDC de UF relacionadas a este nivel, para (4) actividades productivas la medida más utilizada es la barrera acústica mostrando un 41.7% de apariciones en PDC de UF relacionadas a este nivel, para (5) actividades de servicio la medida más utilizada es la barrera acústica mostrando un 28% de apariciones en PDC de UF relacionadas a este nivel, para (6) actividades de esparcimiento la medida más utilizada es la barrera acústica mostrando un 27% de apariciones en PDC de UF relacionadas a este nivel, pero además es la actividad que concentra el mayor % de medidas material absorbente,

reubicación y limitador, para (7) actividades comerciales la medida más utilizada es la barrera acústica mostrando un 29.3% de apariciones en PDC de UF relacionadas a este nivel. Además, la barrera acústica se encuentra presente como medida en todas las categorías de actividad de nivel 1, siendo las actividades de esparcimiento las cuales hacen mayor uso de ellas.

6 DISCUSIÓN Y CONCLUSIONES

Análisis y utilidad de la información

Con el análisis de los datos de las denuncias, fiscalizaciones y procesos sancionatorios se pudo obtener información interesante y útil para el estudio de la aplicación del Decreto Supremo N°38/2011 del MMA a lo largo de estos últimos 6 años, desde enero de 2014 hasta octubre de 2020.

Las denuncias fueron incrementando con el paso de los años, con excepción del 2016, hasta llegar a su nivel más alto el año 2018 con más de 900 denuncias, para luego bajar el 2019 y 2020 (para el caso del año 2020 los datos son incompletos), lo cual se podría explicar por ser años inusuales, ya sea por el “estallido social” como por la pandemia de Covid-19 que trajeron, toques de queda, restricciones a la libertad de circulación y trabajo lo cual podría haber influido en el curso normal de las denuncias por ruido.

Con respecto a los tipos de categorías económicas más denunciadas, la más importante corresponde a “actividades de esparcimiento” con un 27.7%, luego con un 25.7% se encuentran las “actividades productivas”. En el análisis del segundo nivel de categorización destacan el “ocio” para las actividades de esparcimiento e “industria” para las actividades productivas. Los pubs y discotecas corresponden a más de la mitad de las denuncias en el ítem “ocio” mientras que las plantas o fábricas corresponden a un porcentaje importante del nivel 2, “Industria”.

Por otro lado, las “actividades de esparcimiento” se mantienen como la actividad más fiscalizada con un 31.2% y las actividades productivas en segundo lugar con un 22.6% del total.

Para el caso de las AE de denuncias que terminaron sancionadas se obtuvo que un 48.7% corresponden a “actividades de esparcimiento”, es decir, cerca de la mitad de los casos sancionados corresponde a esta actividad. Gran parte de estas AE corresponden a la subcategoría Ocio y dentro de esta a Pubs y discotecas. Cabe señalar que las “actividades de esparcimiento” equivalen a menos de un tercio del total de las denuncias y son más de la mitad de los procesos sancionatorios.

Con respecto a las fiscalizaciones se observó que la Región Metropolitana concentra el 46.6% de 697 fiscalizaciones que se revisaron en el presente estudio. Sin embargo, en el análisis por comuna, se tiene que comunas como Iquique, Concepción y Arica cuentan con un alto número de fiscalizaciones en comparación con el resto de las comunas del país. Cabe señalar que en futuros estudios se debería cruzar la densidad poblacional por regiones y comunas con la información levantada en esta consultoría para una mejor comprensión de la distribución poblacional y geográfica de todos los descriptores considerados.

Con la herramienta interactiva creada es posible comparar las cantidades de denuncias y relacionarlas con fiscalizaciones y sanciones para una determinada actividad, por ejemplo en la siguiente tabla y gráfico se efectúa la comparación de los datos obtenidos para los casos específicos de Pub y Fábricas.

Es posible observar que las denuncias a los pubs son un poco más de la mitad que las de las fábricas, sin embargo, tienen casi el doble de fiscalizaciones y poco más que cuatro veces los procesos sancionatorios.

Tabla 25: Comparación de datos obtenidos para los casos específicos de Pub y Fábricas.

Actividad	N° Denuncias	N° Fiscalizaciones	N° Procesos terminados en sanción
Pub (N1: Esparcimiento, N2: Ocio)	275	106	34
Fábrica (N1: Productiva, N2: Industria)	481	62	8

Del análisis de los NPC se pudo obtener que la mediana para el periodo diurno es de 70 [dB(A)], mientras que para el periodo nocturno es de 59 [dB(A)]. Por su parte, para los excesos a los máximos permitidos, la mediana para el periodo diurno es de 9.0 [dB], mientras que para el periodo nocturno es de 11.4 [dB].

Con respecto a los tiempos de espera entre los distintos hitos del proceso de la aplicación del DS N°38/2011 del MMA tenemos la siguiente tabla.

Tabla 26: Resumen de tiempos de espera entre hitos del proceso.

Hitos principales de la evolución de la denuncia	Tiempos de espera en meses (mediana)	Tiempo de espera acumulado en meses (mediana)
Denuncia - Fiscalización	9.9	9.9
Fiscalización - Inicio proceso	4.8	14.7
Inicio de proceso - Fin de proceso	11.2	25.9

Se observa a partir de la tabla anterior que el tiempo de espera entre la fiscalización y el inicio del proceso sancionatorio es el menor, mientras que entre el inicio del proceso sancionatorio y el fin del mismo es el más extenso.

Con respecto a las multas, tenemos 168 casos que corresponden a procesos sancionatorios que tienen asociados un valor de multa, de los cuales 125 corresponden a aquellos que terminaron en sanción, mientras que los otros 43 terminaron con PDC satisfactorio o proceso archivado o en calidad de absolución, para los cuales la multa es cero. De los terminados

en sanción hay 19 con multa cero. De los 125 se tiene una suma de 1357 UTA, donde la mediana es 5 UTA y el promedio 11 UTA.

La mediana del valor total de las medidas incorporadas a los PDC aprobados es de 192.5 UF. El valor total de la suma de los costos de todos los PDC es de 189012 UF para 156 PDC. Cabe señalar que 12 PDC aprobados quedaron excluidos del análisis por no contar con la fecha de sus valores, lo que generó la imposibilidad de convertir estos a UF.

Para comparar los valores de las multas de los procesos sancionatorios con los valores de PDC se considera apropiado utilizar las medianas y el valor UF. Se utilizaron los valores de UTA de noviembre de 2020 (608088 pesos) y el valor UF del 18 de noviembre de 2020 (28949.28 pesos). En la siguiente tabla se muestra la comparación de ambos valores.

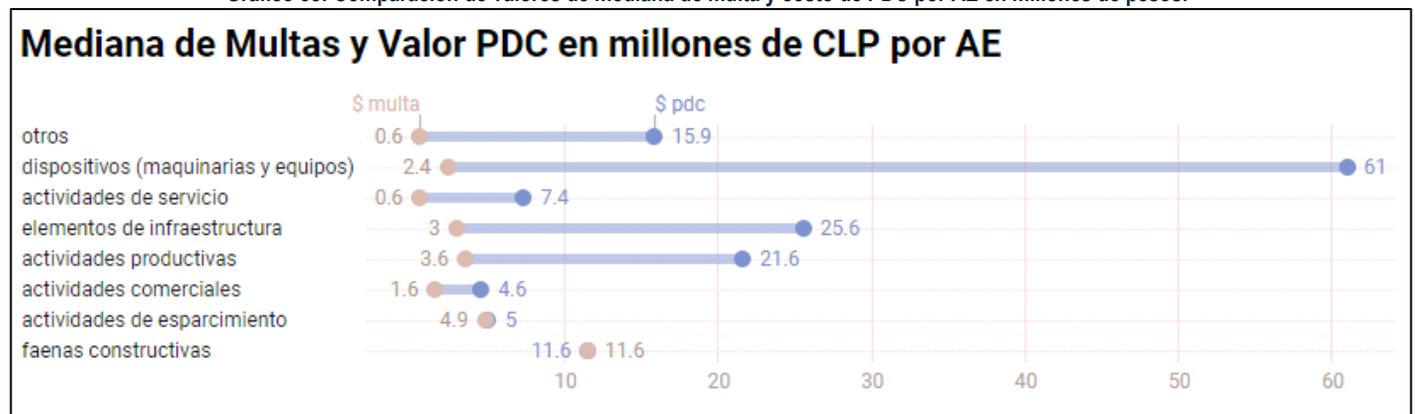
Tabla 27: Comparación de valores de multa y costo de PDC.

Ítem	Cantidad de casos considerados	Mediana valor	Indicador	Valor en pesos chilenos (nov.2020)	Valor UF
Valores Multa	125	5	UTA	3040440	105.0
Valores PDC	144	192.5	UF	5572736	192.5

Se observa que la mediana del valor de la multa corresponde a un 55% del valor de la mediana del PDC. Por otro lado, se evidenció en el presente trabajo que el valor de la multa no tiene correlación ni con el exceso al máximo permitido, ni con la clasificación LOSMA de gravedad, como se observa en los gráficos 20 y 30.

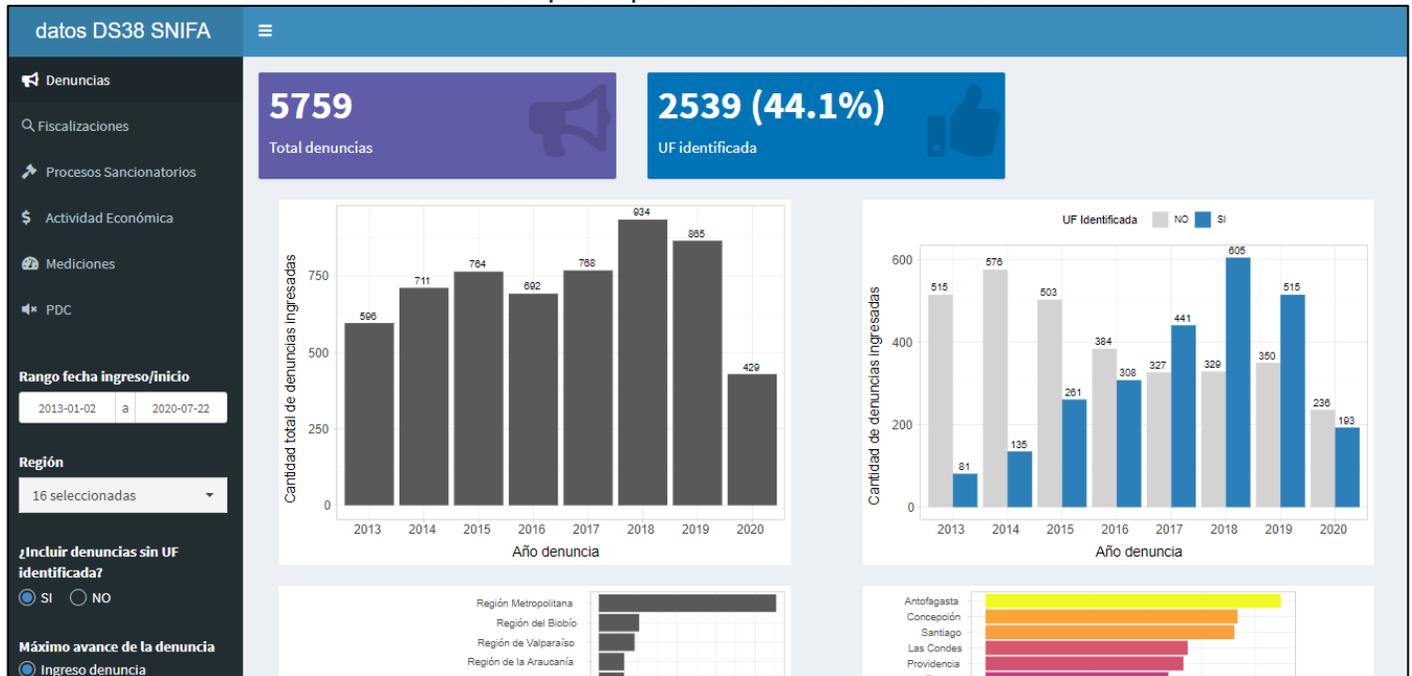
En el siguiente gráfico se muestra la comparación de los valores de multa y costo del PDC por actividad económica. Se observa que para el caso de actividades de esparcimiento y faenas constructivas los valores son muy similares, mientras que para los dispositivos el valor de los PDC es mucho mayor que el de las multas.

Gráfico 38: Comparación de valores de mediana de multa y costo de PDC por AE en millones de pesos.



Como parte de esta consultoría se creó una plataforma para la exploración y visualización interactiva de los datos y sus resultados, lo cual permite al usuario encontrar rápidamente la información que busca. Se incluyeron filtros para Fechas, Región, presencia o ausencia de unidad fiscalizable, fiscalización y sanción, además de pestañas específicas para la información de mediciones, PDC y actividades económicas. Se muestra a continuación parte de la pestaña denuncias de la plataforma mencionada.

Ilustración 9: Captura de pantalla de herramienta de visualización.



Los cálculos efectuados en el presente informe para relacionar variables se hicieron utilizando esta aplicación. Para explorar otras relaciones no contempladas en el presente informe se puede utilizar de forma eficiente e intuitiva.

Problemáticas en relación a la disposición y precisión de la información y propuestas de mejora

1.- La redacción de las denuncias, en algunas ocasiones, es efectuada de tal forma que las incertezas al categorizar son muy altas, ya sea porque se juntan muchas fuentes de distinto tipo en una sola denuncia, por ejemplo, un local que funciona como taller de día y en el que se efectúan eventos los fines de semana; o bien porque la información no es concluyente, como por ejemplo "Ruidos molestos calle", o también por faltas ortográficas y otros motivos.

Es por lo anterior que una propuesta de mejora sería utilizar algún método donde las denuncias sean formuladas de forma que sea necesario que elegir una palabra ya prevista y categorizada. Para esta labor se pueden utilizar como base las distintas categorías propuestas en el presente trabajo. De este modo se evita la posibilidad que el usuario elija "texto libre", lo que induce a errores y dificultades posteriores en el procesamiento. Esto se puede implementar como parte de

un formulario que considere dichas categorías mediante un menú que se despliegue y auto complete a medida que el usuario ingrese caracteres. Si se quiere, se puede considerar la opción “Otros” como único caso de texto libre, el cual debería ser revisado periódicamente con el objetivo de detectar potenciales nuevas categorías. Lo óptimo es que esto se aplique para las categorías de nivel 3, dado que así la asignación de niveles 2 y 1 se realiza de manera automática y directa.

Otra línea de desarrollo corresponde a la creación de categorías paralelas que considere las fuentes de ruido propiamente tal, ya que las categorías actuales están direccionadas en el sentido de identificar la Actividad Económica donde está inserta la fuente de ruido, de este modo un “ventilador”, podría pertenecer a cualquiera de las AE definidas en la Resolución Exenta N° 867, incluso “Dispositivos” la cual corresponde a una clasificación que tiene otro tipo de argumento más relacionada con la fuente de ruido. Esta definición híbrida puede ser un problema en el análisis y en la categorización efectiva. Una categorización debería considerar tanto la actividad económica como la fuente de ruido de manera paralela.

Una propuesta alternativa está enfocada en el desarrollo de un modelo de *machine learning* que utilice texto como entrada y estime las probabilidades de similitud para asignar categorías de AE como salida. El funcionamiento de este tipo de algoritmos es similar a los algoritmos de búsqueda de información entre documentos, como el que utiliza Google en su buscador. Para ello es necesario disponer de ejemplos etiquetados de denuncias para entrenar, testear y validar el modelo. Las más de 5 mil denuncias etiquetadas en este estudio pueden servir para dicha tarea, además de las eventuales futuras denuncias si se implementa el formulario con categorías predefinidas propuesto anteriormente. La aplicación de un modelo de este tipo podría ser probada también posteriormente en otros tipos de denuncias de ruido, tales como las realizadas en las municipalidades o carabineros.

2.- La información de mediciones está contenida principalmente en dos partes. Por un lado está en uno o varios de los documentos que forman parte de los expedientes de fiscalización y proceso sancionatorio (formulación de cargos, resolución sancionatoria, fichas de medición, etc.) y por otro lado en la columna Hecho del sitio web de la SNFA ligado a un PS en particular. En ambos casos hay dificultades para la extracción automática de la información. En el próximo punto se discute lo referido al contenido de los documentos. En este punto se enfatiza la redacción del texto de la columna Hecho, la cual presenta los siguientes obstáculos:

- Similar al caso de las denuncias, se tiene distintos estilos de redacción, ya que no existe un formato estandarizado en cuanto al contenido y orden de los elementos, al menos para los casos más antiguos.
- Los casos que tienen más de una medición, ya sea por fecha o receptor, integran todas las fechas y valores NPC asociados, dificultando la identificación y asociación de cada grupo por separado mediante algoritmos.
- Ausencia de información relevante en el texto redactado. El ruido de fondo casi nunca está presente, aun cuando se identifica zona rural, impidiendo el cálculo del límite permitido. En algunos casos no hay algún otro parámetro, como zona, horario, NPC, fecha o exceso

3.- Los documentos subidos al sistema presentan distintos tipos de archivo (PDF, texto en Word, archivos comprimidos en RAR o ZIP, etc.) y distintos formatos internos, ya sean estos textos, tablas, imágenes escaneadas, etc. Esto dificulta la extracción automática de datos con alta precisión, por lo que para cierto tipo de información no queda otra opción que

hacer recopilación manual, como por ejemplo medidas y valores de PDC, lo que implica un gasto de recursos considerable. Si bien se identificaron casos de archivos PDF con formato estandarizado para el ingreso de datos de mediciones y de medidas de PDC, para un uso óptimo de la automatización se requiere que la gran mayoría cumpla con esto. Adicionalmente, es importante tener claro que este enfoque también aumenta los requerimientos computacionales de espacio en disco y RAM dada la para el almacenamiento y proceso la cantidad de documentos involucrados. Si de todas formas se sigue con este tipo de registro, se sugiere considerar un cambio en la manera de ingresar la información de medidas de control dado que marcar cruces dentro de cuadrados no facilita su reconocimiento automático mediante técnicas OCR.

Todo indica que estas dificultades son naturales dado que la disposición de la información no se ideó inicialmente para cumplir el objetivo de acceder automáticamente a toda la información particular de ruido para su posterior análisis de datos, sino más bien para tener un registro histórico de consulta general y/o quizás específica, pero a nivel de fiscalizaciones y PS individuales. Es justamente ese punto el que se sugiere abordar primero para decidir cómo aplicar mejoras: ¿Para qué se quiere registrar la información?.

Se identifican dos principales caminos en relación a las mejoras: a) modificar la manera actual de registro para que esta se adapte de mejor manera a las necesidades de la extracción automática de la información, y/o b) agregar nuevos registros adicionales y específicos en los formularios y en el sitio web de modo que se pueda acceder directamente a la información, aparte de los registros ya existentes. En la línea del primer caso está la estandarización de la redacción y el formato de documentos subidos al sistema para mejorar la precisión de la extracción automática con expresiones regulares o la aplicación de modelos de *machine learning*, mientras que en la línea del segundo caso está la implementación de formularios predefinidos y la inclusión directa y aislada desde un inicio de la información que se quiere analizar después. Esto último por ejemplo en relación a la información de mediciones y de los PDC. Se pueden aplicar criterios de priorización, como por ejemplo el que se aplicó en este estudio para las mediciones: considerar solo el NPC asociado a la medición que representó el mayor exceso y la fecha de la primera medición (relevante como hito para cálculo de delay).

Finalmente, se identificó la falta de información de un número importante de fiscalizaciones dado que algunos PS no tenían su fiscalización ingresada. Se desconoce el motivo.

Por otro lado, sería interesante disponer de herramientas computacionales que permitan la actualización y validación periódica de la base de datos y la plataforma interactiva en función del contenido del sitio web de la SNIFA.

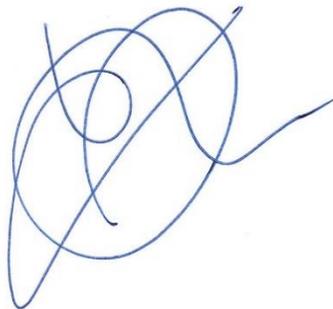
7 BIBLIOGRAFÍA

- Información facilitada por el mandante para realizar el estudio
- D.S. N°38/11 del MMA “Establece norma de emisión de ruidos generados por fuentes que indica”
- Tidy data, Hadley Wickham - Journal of Statistical Software
- Interactive Data Visualization, Foundations, Techniques, and Applications –Ward, Grinstein, Keim.
- Gómez I., Glisser M. y Padilla C., “Data analysis of noise complaints in Región Metropolitana, Chile“, 22nd International Congress on Acoustics, Buenos Aires, Argentina, September 5 – 9, 2016.
- Cuenta pública 2017, Superintendencia de Medio Ambiente, Gobierno de Chile.
- Ministerio del Medio Ambiente, Gobierno de Chile, “Gestión del Control del Ruido Ambiental en Chile“, Seminario Ruido Ambiental MMA, 2016.
- Ministerio del Medio Ambiente, Gobierno de Chile, “Informe Primera Encuesta Nacional de Medio Ambiente”, 2015.
- Ministerio del Medio Ambiente, Gobierno de Chile, "Elaboración de mapa de ruido del Gran Santiago mediante software de modelación", 2011.
- Ministerio de Medio Ambiente, Gobierno de Chile, “Estrategia para la Gestión del Control de Ruido Ambiental 2010-2014”.
- N. D. S. Academy, NYC Data Science Academy Blog, "NYC noise complaint visualization", Noviembre 2015. Available here.
- Decreto Supremo N°38/2011 del Ministerio del Medio Ambiente “Norma de Emisión de Ruidos Generados por Fuentes que Indica”.
- Ley N° 20.285 Sobre Acceso a la Información Pública, Ministerio Secretaría General de la Presidencia, Gobierno de Chile, Agosto 2008.
- Rainer Guski. “An analysis of spontaneous noise complaints”. Environmental Research 13(2):229-36. May, 1977.
- The Noise App. 2018 RH Environmental Limited. <https://www.thenoiseapp.com/#/>
- R Foundation for Statistical Computing, R Development Core Team, R: A language and environment for statistical computing. Vienna, Austria, 2008.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL: <http://www.rstudio.com/>. Version 1.1.447. RStudio: Integrated development environment for R. Boston, MA.

- Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59 (10), 1 - 23. doi: <http://dx.doi.org/10.18637/jss.v059.i10>
- Interactive Data Visualization: Foundations, Techniques, and Applications. A. K. Peters, Ltd. Natick, MA, USA 2010.
- A web tool for selecting colors for maps. Brewer, Cynthia A., 2002-2013. <http://www.ColorBrewer.org>
- 2017 UK Nuisance Noise Study. <https://www.lovemymvouchers.co.uk/2017-noise-study/>
- Diagnosing New York City's Noises with Ubiquitous Data. Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu , Yanchi Liu , Eric Chang. Microsoft Research.
- Instituto Nacional de Estadísticas. <http://nuevoportal.ine.cl>

8 PROFESIONALES PARTICIPANTES

Cargo	Nombre	Profesión
Jefe de Proyecto	Max Glisser Donoso	Ing. Civil en S. y Acústica Dipl. en Gestión Ambiental
Asesor Especialista en Análisis de Datos	Ismael Gómez Schmidt	Ing. Civil en S. y Acústica Dipl. en Ciencia e Ingeniería de Datos
Asesor	Benjamín Madueño Maringuer	Ingeniero en Sonido



MAX GLISSER DONOSO
INGENIERO CIVIL EN SONIDO Y ACÚSTICA
JEFE DE PROYECTO



ISMAEL GÓMEZ SCHMIDT
INGENIERO CIVIL EN SONIDO Y ACÚSTICA
ASESOR ESPECIALISTA EN ANÁLISIS DE DATOS