



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

## HERRAMIENTA DE VISUALIZACIÓN DE RIESGO DE PACIENTES COVID-19 EN UNA ZONA URBANA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

JUAN PABLO ALVARADO GLENA

PROFESOR GUÍA:  
FERNANDO ORDÓÑEZ PIZARRO

MIEMBROS DE LA COMISIÓN:  
ÁNGEL JIMÉNEZ MOLINA  
ANDRÉS GARRIDO CASTRO

Este trabajo ha sido parcialmente financiado por la  
Agencia Nacional de Investigación y Desarrollo (ANID)

SANTIAGO DE CHILE  
2022

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL  
INDUSTRIAL  
POR: JUAN PABLO ALVARADO GLENA  
FECHA: 2022  
PROF. GUÍA: FERNANDO ORDÓÑEZ

## **HERRAMIENTA DE VISUALIZACIÓN DE RIESGO DE PACIENTES COVID-19 EN UNA ZONA URBANA**

El presente trabajo de título se enmarca en el proyecto ANID: COVID 0251 adjudicado por el Centro de Sistemas Públicos (CSP) de la Universidad de Chile. El objetivo principal de este trabajo consiste en el desarrollo de un prototipo de herramienta de visualización de casos confirmados y probables por COVID-19 para siete comunas de la Región Metropolitana de Santiago, las cuales componen el Servicio de Salud Metropolitano Sur Oriente (de ahora en adelante, SSMSO).

El trabajo de título busca crear una herramienta que ayude en la toma de decisiones del SSMSO respecto a las labores de trazabilidad y búsqueda activa de casos entre los habitantes que componen sus comunas de incidencia. Para esto, la herramienta facilitará el análisis de los casos registrados como ingresos por COVID-19 en cualquiera de sus establecimientos de salud entre Abril 2020 y Mayo 2021. A partir de lo anterior, la utilidad de la herramienta radica en asignar prioridad a ciertas zonas de la Región Metropolitana, de acuerdo a dos variables relevantes:

1. Riesgo: Representado como la probabilidad de que un paciente confirmado o probable contraiga un cuadro grave por COVID-19 de cada macrozona delimitada.
2. Cantidad de Adultos Mayores: Número de adultos mayores que habitan cada macrozona delimitada. Lo anterior cobra relevancia debido a la vulnerabilidad que tiene la tercera edad al estar expuesta al virus [1].

Para la realización de la herramienta, se integraron los resultados provenientes de dos desarrollos realizados como parte del proyecto ANID: un modelo de normalización de direcciones/geocodificación y también un modelo de riesgo clínico para caracterizar pacientes. El primero de ellos permite encontrar las latitudes y longitudes de los pacientes declarados como confirmados y probables por COVID-19, mientras que el segundo realiza una predicción del riesgo que cada persona en particular enfrenta con respecto a la adopción del virus y generar un cuadro de gravedad, en base a sus propios antecedentes clínicos declarados y características personales.

Con ambos resultados, se implementó el prototipo de herramienta de visualización del riesgo de desenlace por cuadro grave para casos COVID-19 confirmado y probable, donde destaca la elaboración de un mapa de calor que da cuenta de la probabilidad de cuadro grave de COVID-19 de las personas confirmadas y probables que habitan en las comunas del SSMSO.

*Para todas las personas que sufrieron producto de esta pandemia.*

# Agradecimientos

Mi familia ocupa un lugar esencial en mi vida y es por ello que les adjudico las primeras palabras de agradecimiento. Muchas gracias a mis padres y a mis hermanas, quienes apoyaron este largo y duro proceso desde mis raíces hasta este período más bien oscuro de la humanidad llamado pandemia. Gracias por cada palabra de aliento, por darme la educación necesaria que me permitiera realizar este trabajo y por el inmenso amor que demuestran en cada uno de sus actos. El amor que ustedes me han demostrado es lo más valioso que tengo y tendré.

Sin lugar a dudas me gustaría agradecer a mi profesor guía Fernando Ordóñez y al tesista Matías Martínez Pérez, quienes integraron el equipo que desarrolló el modelo de geocodificación y el prototipo de herramienta de visualización de casos. Fueron muchas reuniones por zoom o meets que significaron la realización del proyecto, donde siempre primó el respeto y la buena onda por sobre todas las cosas. Muchas gracias por la paciencia y por la orientación recibida. Agradecer también a Andrés Garrido por permitirme el uso de los resultados del modelo de riesgo que implementó para el proyecto, fue de mucha ayuda en la construcción de la plataforma.

A mis amigas y amigos de la universidad, las cuales son personas que me llevo con mucho cariño después de este período como estudiante de la FCFM ya que en ellas y ellos encontré una fuente de confianza que atesoro. Gracias por cada momento de risa, por los viajes y por la jarana. ¡Que viva la buena vida y la poca vergüenza!

Por último, agradecer el legado que realizó en mi la Universidad de Chile, como institución y como club de fútbol del cual soy hinchita. Académicamente, ya son más de 10 años ligado a esta institución, realizando cursos de escuela de invierno cuando era un estudiante de colegio hasta esta última instancia del pregrado. Futbolísticamente, ya es toda una vida siguiendo, alentando y vibrando con el león. No todo podía ser alegría, y espero que en un futuro pueda volver a leer este fragmento y ver al club de fútbol con la identidad y los valores que les fueron arrebatados por amiguismos empresariales. Y para ello, es necesario que la Universidad de Chile recupere lo que siempre fue suyo.

Desde que era chico, mi viejo me enseñó una frase que nos caracteriza a todas y todos quienes seguimos esta pasión:

“Si no se sufre, no es la U”

Y como el legado no es solo futbolístico, comprenderán que esta también aplica para mi estancia en la FCFM. Pero así como se sufre, también se disfruta (y mucho) el proceso.

# Tabla de Contenidos

<b>1. Introducción</b>	<b>1</b>
1.1. El COVID-19 . . . . .	1
1.1.1. Situación a nivel país . . . . .	3
1.2. Proyecto ANID: COVID 0251 . . . . .	5
1.3. El SSMSO . . . . .	6
1.4. Pregunta de Investigación, Objetivos Generales y Específicos . . . . .	8
1.5. Alcances y limitaciones . . . . .	9
1.6. Estructura . . . . .	11
<b>2. Marco Teórico</b>	<b>12</b>
2.1. Definiciones relacionadas al COVID-19 . . . . .	12
2.2. La geocodificación . . . . .	13
2.3. Asignación de riesgo . . . . .	15
2.4. Herramientas de visualización desarrolladas . . . . .	18
<b>3. La muestra</b>	<b>20</b>
3.1. Datos importantes . . . . .	20
3.2. EDA . . . . .	22
3.2.1. Personas . . . . .	22
3.2.2. Exámenes . . . . .	24
3.2.3. Direcciones . . . . .	26
<b>4. El modelo de geocodificación</b>	<b>29</b>
4.1. Proceso de normalización de direcciones . . . . .	30
4.2. Agrupación del territorio . . . . .	31
4.3. La asignación de personas a manzanas-entidades . . . . .	33
<b>5. El modelo de riesgo</b>	<b>36</b>
5.1. Riesgo clínico del COVID-19 . . . . .	36
5.2. El modelo de riesgo utilizado . . . . .	36
<b>6. La herramienta de visualización</b>	<b>40</b>
6.1. Nueva agregación de datos: las macrozonas . . . . .	40
6.1.1. Partición . . . . .	40
6.1.2. Datos incorporados . . . . .	42
6.2. La implementación de la herramienta de visualización . . . . .	43
6.2.1. El mapa . . . . .	43
6.2.2. La interfaz . . . . .	44

<b>7. Conclusiones</b>	<b>47</b>
7.1. Resultados obtenidos . . . . .	47
7.2. Discusión . . . . .	48
7.3. Trabajo futuro . . . . .	48
7.3.1. Robustecimiento del modelo de riesgo . . . . .	48
7.3.2. Mejoras en la aplicación . . . . .	49
<b>Bibliografía</b>	<b>57</b>

# Índice de Tablas

1.1.	Razón de contactos por caso confirmado en las comunas que componen el Servicio de Salud Metropolitano Sur Oriente según el 125° Informe Epidemiológico del MINSAL. . . . .	8
2.1.	Variables cuantitativas del estudio . . . . .	16
2.2.	Variables cualitativas del estudio . . . . .	16
3.1.	Variables en la tabla PERSONAS de la muestra de datos. . . . .	21
3.2.	Variables en la tabla EXÁMENES de la muestra de datos. . . . .	21
3.3.	Variables en la tabla DIRECCIONES de la muestra de datos. . . . .	21
3.4.	Resumen con la proporción de casos entregadas en el 125° Informe Epidemiológico versus lo contemplado en la muestra, para cada comuna de interés. . . . .	28
5.1.	Resumen de las variables consideradas en el modelo logístico de riesgo implementado por Andrés Garrido. . . . .	38
6.1.	Variables de interés para la conformación de la plataforma de visualización. . . . .	42
6.2.	Variables incorporadas a nivel manzana-entidad . . . . .	43
7.1.	Resultados de la regresión logística implementada por Andrés Garrido. . . . .	51
7.2.	Odd Ratios de las variables del modelo. . . . .	51

# Índice de Ilustraciones

1.1.	Medidas para prevenir el contagio del virus, elaborado por el Gobierno de Chile.	2
1.2.	Cantidad de tests realizados por caso confirmado en el mundo, publicado por Our World in Data. . . . .	4
1.3.	Defunciones mensuales por COVID-19 en Chile, obtenido de la página web del MINSAL en Marzo del 2022. . . . .	4
1.4.	Áreas funcionales del SSMSO. . . . .	7
1.5.	Porcentaje de la población con esquema de vacunación inicial completo (1 o 2 dosis según corresponda), publicado por Our World in Data. . . . .	10
2.1.	Ejemplo de geocodificación de dirección en Boston. Ejemplo tomado de PubNub.	14
2.2.	Odd Ratios con IC del 95 % de confianza del estudio. Obtenido del mismo. . .	17
3.1.	Distribución de la edad de las personas que componen la muestra. . . . .	23
3.2.	Distribución de género de las personas que componen la muestra. . . . .	23
3.3.	Distribución de ingresos registrados a lo largo del período. . . . .	24
3.4.	Número de exámenes PCR, desagregados en sus distintos resultados. . . . .	24
3.5.	Número de exámenes PCR tomados en cada mes del período de la muestra. . .	25
3.6.	Número de exámenes PCR con resultado positivo registrados cada mes del período de la muestra. . . . .	25
3.7.	Diagrama de embudo que explica el recorte del total de direcciones. . . . .	27
3.8.	Número de pacientes confirmados/probables que habitan cada comuna del SSMSO. . . . .	28
4.1.	Proporción de casos encontrados con cada una de las APIs. . . . .	30
4.2.	Ejemplo de dirección normalizada. . . . .	31
4.3.	Posibles divisiones del territorio. Extraída de la página web del INE. . . . .	32
4.4.	Ejemplo de caso desfavorable donde se asigna a una manzana-entidad equivocada. Extraído de la memoria de Matías Martínez P. . . . .	35
5.1.	Lógica del modelo de riesgo. . . . .	37
5.2.	Modelo de riesgo desarrollado por Andrés Garrido. . . . .	37
6.1.	Macrozonas que delimitan la geografía del SSMSO. Elaboración propia. . . . .	41
6.2.	Mapa desarrollado para el SSMSO. . . . .	44
6.3.	Prototipo de Herramienta de Visualización del riesgo de desenlace de cuadro grave por COVID-19. . . . .	45
7.1.	Divisiones comunales de Lo Espejo, La Cisterna, San Ramón y La Granja. . .	52
7.2.	Zonas Censales de Lo Espejo y La Cisterna. . . . .	53
7.3.	Manzana-Entidad de la comuna de Lo Espejo. . . . .	53
7.4.	Plataforma de visualización de casos de la Universidad de John Hopkins. . . .	54
7.5.	Dashboards realizados por la OMS. . . . .	54
7.6.	Dashboards realizados por Mayo Clinic. . . . .	55



7.7.	Plataforma de visualización de casos del Centro Nacional de Epidemiología (CNE) de España. . . . .	55
7.8.	Plataforma de visualización de casos desarrollada por DATA UC. . . . .	56
7.9.	Visor Territorial COVID-19 del Ministerio de Bienes Nacionales. . . . .	56

# Capítulo 1

## Introducción

### 1.1. El COVID-19

En Diciembre del 2019, se notificaron casos puntuales de pacientes que habían contraído una neumonía desconocida en la localidad de Wuhan, China. Posterior a investigaciones científicas en la zona, el día 20 de Enero del 2020 la Comisión Nacional de Salud de China confirmó que el virus que producía este tipo de neumonía, el SARS-CoV-2, podía transmitirse entre seres humanos que hubieran tenido algún tipo de contacto [2]. Ante la rápida propagación del virus en el territorio chino y posteriormente en el resto de países del mundo, la Organización Mundial de la Salud (OMS) declara una emergencia de salud pública de importancia internacional el día 30 de Enero, para posteriormente declararlo como pandemia el día 11 de Marzo de 2020, cuando se informó que el virus había causado 4.291 muertes, con 118.000 casos en 114 países distintos [3].

El virus cabe dentro de la categoría de coronavirus, los cuales son responsables de causar enfermedades en humanos y animales. A la fecha de hoy se han reconocido 7 tipos distintos de coronavirus, entre ellos los encargados de generar el síndrome respiratorio agudo grave (SARS-CoV) en China durante el 2002-2003 y el síndrome respiratorio de oriente medio (MERS-CoV) en la península Arábiga el 2012. A diferencia de los otros coronavirus el SARS-CoV-2 presenta diferencias en sus genomas, los cuales producen una facilidad de contagio y letalidad mayor, produciendo la emergencia sanitaria mundial actual.

Desde la irrupción del virus en nuestra sociedad, el ser humano ha realizado hallazgos relevantes para poder frenar la curva de contagios en momentos oportunos y también la tasa de mortalidad que estaba teniendo desde los inicios de su aparición. Principalmente, se ha descubierto que el virus se transmite entre personas a través de pequeñas gotas de saliva llamadas las gotas de Flügge, las cuales se emiten al hablar, estornudar, toser o respirar, no solamente al estar en contacto con otra persona que porte el virus, sino que también si es que una persona no contagiada entra en contacto con una superficie contaminada y lleva la parte de la piel que entró en contacto a la cara o a las mucosas. A la vez, métodos efectivos para prevenir la propagación del virus consisten en lavarse periódicamente las manos, cubrir la boca al toser, el distanciamiento físico, el uso de mascarillas y el aislamiento/seguimiento de personas que hayan contraído o sean sospechosos de portar el virus. La figura 1.1 resume las medidas para prevenir el contagio del coronavirus, afiche elaborado por el Gobierno de Chile.

# Prevención contagio Coronavirus

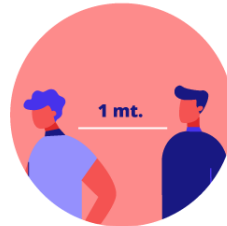
MINISTERIO DE SALUD  
PLAN DE ACCIÓN  
**CORONAVIRUS**  
COVID-19



**Lavado de manos**  
mínimo 20 segundos  
con agua y jabón.



**Estornudar o toser**  
con el antebrazo  
o en un pañuelo  
desechable.



**Mantener una distancia**  
mínima de **1 metro**  
entre usted y cualquier  
persona que tosa  
o estornude.



**Evitar tocarse**  
**los ojos,** la nariz  
y la boca.



**No compartir bombilla,**  
**vaso** o cubiertos  
con otras personas.



**Evitar saludar**  
**con la mano**  
**o dar besos.**

Figura 1.1: Medidas para prevenir el contagio del virus, elaborado por el Gobierno de Chile.

Entre los principales síntomas que produce la enfermedad, se encuentran:

- Cefálea o dolor de cabeza
- Tos seca
- Mialgias o malestar en los músculos
- Anosmia o pérdida del olfato
- Disgeusia/Ageusia o pérdida del gusto
- Congestión nasal

- Disnea o dificultad para respirar
- Fiebre
- Diarrea

Los síntomas anteriores cobran relevancia en el análisis desarrollado en este trabajo, dado a que el modelo de riesgo los considera como un *input* importante para predecir el egreso de un paciente contagiado por COVID-19. Lo anterior se aborda en mayor detalle en el capítulo 5: “El modelo de riesgo”.

### 1.1.1. Situación a nivel país

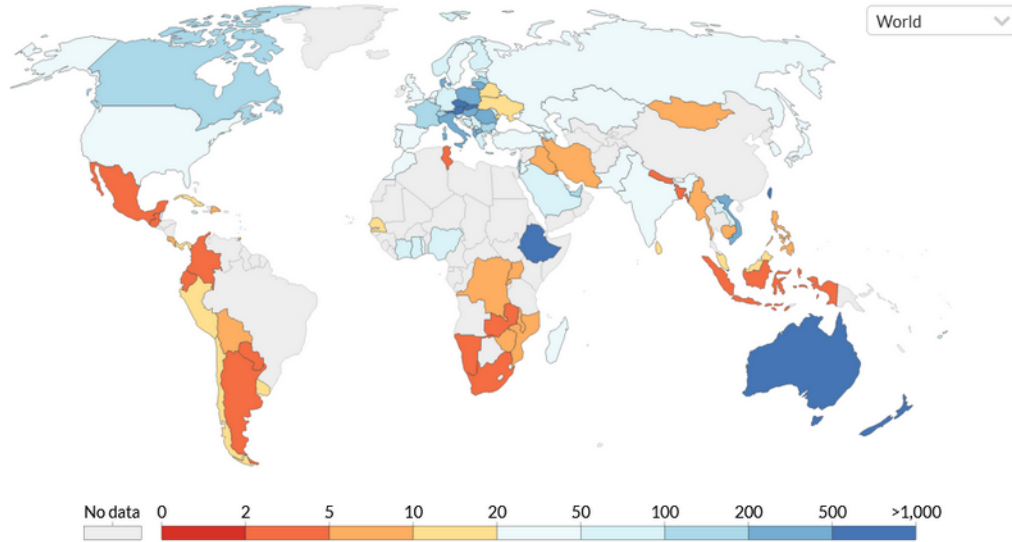
Antes de la pandemia, Chile gastaba aproximadamente un 8% de su PIB en salud [4], es decir aproximadamente unos \$23.219 millones de dólares. Sin embargo, en el marco de la pandemia del SARS-CoV-2 se estima que el gasto público en salud ha aumentado en un 11% respecto del año pasado a la misma fecha [5]. Lo anterior se complementa con lo difícil que resulta realizar tanto las labores de trazabilidad como las de seguimiento de pacientes, dado a que se cuenta con recursos humanos limitados, lo que ha hecho insostenible el seguimiento total de pacientes confirmados o probables de COVID-19. Adicionalmente, las distintas formas en que la enfermedad se manifiesta en las personas, a través de casos sintomáticos o asintomáticos, dificultan las labores previamente señaladas, ya que se alcanzan proporciones ligeramente distintas [6] que imposibilitan abarcar la totalidad de casos confirmados o probables.

En relación a la búsqueda activa de casos, a fecha 5 de Julio del 2021 (días posteriores al último caso de la muestra) Chile realizaba una cantidad media-baja de exámenes PCR antes de encontrar un caso confirmado en relación a los demás países del mundo. La figura 1.2 muestra el mapa de tests PCR tomados por cada caso confirmado de COVID-19 por país, donde a fecha 5 de Julio del 2021, se indica que Chile realizaba en promedio 18 tests de COVID-19 antes de encontrar un caso confirmado por COVID-19, muy por debajo de países como Austria (5.038), Australia (3.507) y Nueva Zelanda (2.684). La BBC destaca la correlación que existe entre los países que apuestan por el análisis masivo de su población y el llegar a un alto nivel de control de la pandemia [7].

## Tests conducted per new confirmed case of COVID-19, Jul 5, 2021

Shown is the daily number of tests for each new confirmed case. This is a rolling 7-day average.

Our World  
in Data



Source: Official data collated by Our World in Data - Last updated 6 July 2021, 12:20 (London time)

OurWorldInData.org/coronavirus • CC BY

Note: Comparisons of testing data across countries are affected by differences in the way the data are reported. Daily data is interpolated for countries not reporting testing data on a daily basis. Details can be found at our Testing Dataset page.

► Jan 6, 2020 ○ Jul 5, 2021

Figura 1.2: Cantidad de tests realizados por caso confirmado en el mundo, publicado por Our World in Data.

Solo en Chile, el virus ha cobrado la vida de 56.580 personas en total desde sus inicios hasta Marzo del 2022, alcanzando su máximo histórico de muertes mensuales el mes de Junio del 2020, con 6.498 muertes. Si se observa el período de muestra en el que se construyó la herramienta, este número asciende a las 38.081. En la figura 1.3 se puede observar los decesos producto del virus desde su irrupción en Chile, donde se observa que a pesar de ver una notable baja de muertes en la segunda mitad del año 2021 y Enero del 2022, este número vuelve a ir en aumento, situándose en valores por sobre las 2.000 personas para los meses de Febrero y Marzo, a pesar de la avanzada campaña de vacunación que ha ocurrido en el país.

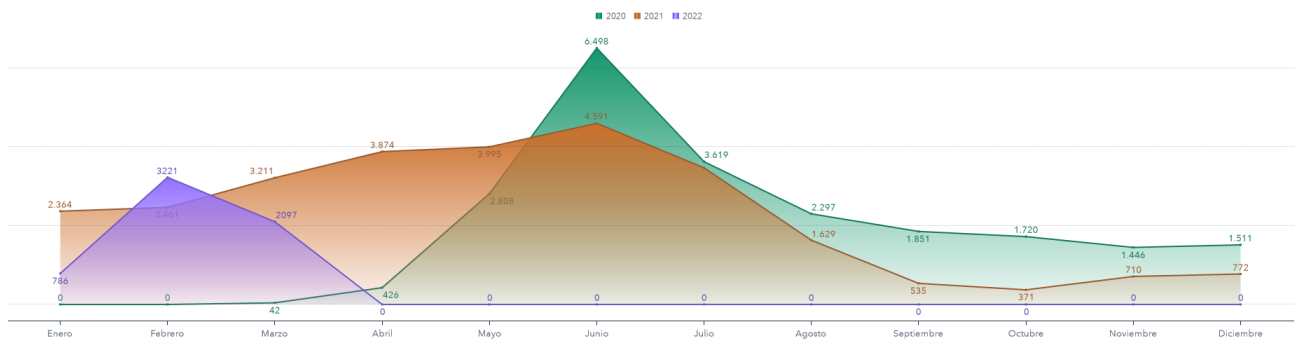


Figura 1.3: Defunciones mensuales por COVID-19 en Chile, obtenido de la página web del MINSAL en Marzo del 2022.

Respecto a la trazabilidad en el país, y de acuerdo al 125° Informe Epidemiológico [8] publicado el 2 de Junio de 2021 por el Ministerio de Salud respecto de la Estrategia Nacional

de Testeo, Trazabilidad y Aislamiento, Chile está trazando solamente a 2,5 contactos por cada confirmado, menos de la mitad de los contactos por caso confirmado que sugiere trazar la OMS con el fin de poder tener un mayor control de la pandemia [9]. A Enero de este año, este número descendió 1,1 según el 170° Informe Epidemiológico.

Es decir dado a que los montos gastados en salud han ido en aumento producto de la pandemia, sumado a las oportunidades de mejora que presentan tanto el programa de búsqueda activa como el de trazabilidad nacional, es que surge la concepción de un prototipo de herramienta de visualización de casos activos como una de las alternativas de solución para lograr una mejor trazabilidad a un menor costo, privilegiando los esfuerzos económicos en quienes sean más vulnerables al COVID-19 gracias a la noción de riesgo que esta otorga.

## 1.2. Proyecto ANID: COVID 0251

Dados los antecedentes recopilados y a la necesidad que tenía (y aún tiene) la sociedad por investigar y desarrollar soluciones a los problemas que el virus contrae, la Agencia Nacional de Investigación y Desarrollo (ANID) lanzó el día 29 de Abril del 2020 un concurso [10] para la asignación de recursos para proyectos de investigación ligados al SARS-CoV-2.

Así, la presente investigación se enmarca en el proyecto COVID 0251, financiado por la ANID. El proyecto se titula “Sistema integrado de información para el seguimiento domiciliario de pacientes COVID-19 en servicios de salud”, cuyo director es Richard Weber, Profesor Titular de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. En este proyecto participaron investigadores del Departamento de Ingeniería Industrial y profesionales del Centro de Sistemas Públicos de esta unidad.

El objetivo general de este proyecto es: “aumentar la efectividad del seguimiento a los pacientes COVID-19 confirmados, los casos sospechosos, los casos probables y sus contactos mediante una plataforma analítica escalable que integre diferentes fuentes de información, incluyendo la declarada por pacientes, y entregue analítica relevante para potenciar la toma de decisiones”. Es posible encontrar más información sobre el proyecto en el siguiente enlace: <https://www.sistemaspublicos.cl/gproyecto/covid0251/>

El proyecto, con una duración total de 10 meses, propone la elaboración de un sistema tecnológico que permita potenciar y escalar el seguimiento domiciliario, apoyando la toma de decisiones para la asignación de recursos en red, considerando tres elementos: una plataforma de seguimiento de pacientes mediante autoreporte, algoritmos de integración de los diferentes sistemas de información disponibles en los servicios de salud y un sistema de visualización y análisis prospectivo de la información mediante inteligencia artificial. En su totalidad, el proyecto contempló tres áreas de trabajo, cuyos objetivos se detallan a continuación:

1. Auto-Reporte: Se buscó generar un sistema de seguimiento multicanal de pacientes en cuarentena domiciliaria, poniendo el enfoque en que estos puedan reportar de manera autónoma sus síntomas diariamente, mediante canales digitales.
2. Integración: El trabajo consistió en automatizar la integración de las diferentes fuentes de información relevantes para el monitoreo de pacientes confirmados/probables y potenciales contactos.

3. Analítica: A través de esta unidad se creó un sistema de monitoreo prospectivo con la información integrada, lo cual facilitaría la toma de decisiones con respecto al seguimiento de pacientes COVID-19, aumentando su eficiencia y eficacia.

En concreto, el trabajo de título se enmarca en la unidad de analítica del proyecto. Dentro de la unidad de analítica, el equipo de trabajo conformado por el investigador Fernando Ordóñez y los tesisistas Matías Martínez y Juan Alvarado, tuvo la responsabilidad de realizar una herramienta de visualización geográfica de pacientes COVID-19 en conjunto con información demográfica que facilite la toma de decisiones para mitigar la pandemia, la cual se dividió en dos etapas: en primera instancia, se procedió a encontrar la latitud y longitud de pacientes mediante el procesamiento de las direcciones que estos declaraban al momento de ingresar a un establecimiento de salud y en segunda instancia se ubicaron a estos pacientes en un mapa para poder llevar a cabo un registro y seguimiento de estos. Ambas etapas del proyecto son complementarias y utilizan la analítica para desprender información que sea útil a la hora de tomar decisiones, siendo la primera de ambas la etapa encargada por Matías Martínez en su tema de memoria y la segunda etapa el tema de investigación de este trabajo.

### **1.3. El SSMSO**

La investigación se realiza en conjunto con el Servicio de Salud Metropolitano Sur Oriente, el cual compone a los establecimientos de salud pública de la zona sur-oriente de Santiago. La figura 1.4 resume las áreas funcionales del SSMSO en la Región Metropolitana de Chile. En particular, el SSMSO compone a los establecimientos de las comunas de La Florida, La Granja, La Pintana, Pirque, Puente Alto, San José de Maipo y San Ramón. Esta red se divide en tres áreas funcionales: Cordillera, Santa Rosa y La Florida.

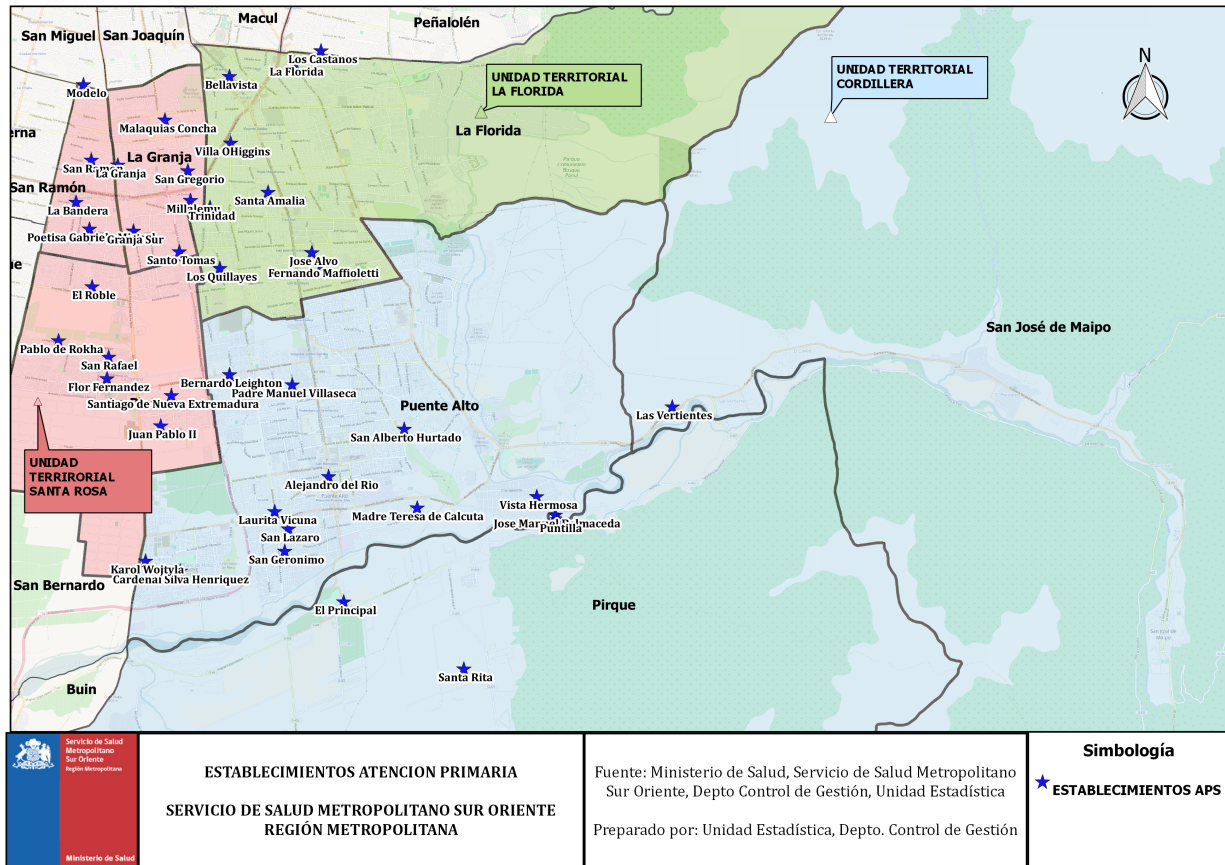


Figura 1.4: Áreas funcionales del SSMSO.

De acuerdo a la cuenta pública participativa del SSMSO [11] en el año 2019, la Red Pública de Salud Sur Oriente está compuesta por 4 hospitales: Complejo Asistencial Sótero del Río, Hospital Clínico Dra. Eloísa Díaz de La Florida, Complejo Asistencial San José de Maipo y Hospital Padre Hurtado. Además, cuentan con el Centro de Referencia de Salud del Hospital Provincia Cordillera, con el Centro Metropolitano de Sangre y Tejidos, con los establecimientos de Atención Primaria de Salud de las comunas de La Granja, San Ramón, La Pintana, La Florida, Puente Alto, Pirque y San José de Maipo y finalizando con establecimientos con convenios especiales como los tres CESFAM Ancora UC, el Hospital Josefina Martínez y la Clínica Familia. A la vez, en la cuenta pública se señala que aproximadamente 1.875.179 personas utilizaron la Red Sur Oriente durante el 2019, lo cual refleja un aumento del 23% con respecto a la cantidad de personas que utilizaron dependencias del servicio el año 2018.

Por otro lado, con respecto al rendimiento de la trazabilidad de las comunas del SSMSO en el marco de la pandemia, se desprende que el proyecto a realizar cobra especial sentido, considerando que en el 125° Informe de Estrategia Nacional Testeo, Trazabilidad y Aislamiento previamente señalado, se declara que la región Metropolitana tiene una razón promedio de trazabilidad de contactos por cada confirmado menor al promedio nacional, siendo esta de 2 frente a los 2,5 del ámbito nacional. Más aún, para las comunas del SSMSO se observa que solo las comunas de La Granja y La Florida se encuentran con un coeficiente igual o superior al promedio de la región Metropolitana, mientras que las comunas restantes tienen coeficientes menores, lo cual es un primer indicio de la dificultad que se ha tenido en esas



zonas para trazar a los contactos de las personas que se diagnostican como confirmados y/o probables. La tabla 1.1 resume la razón de contactos por cada caso confirmado en las comunas donde se implementará la herramienta de visualización, según lo establecido en el 125° Informe Epidemiológico elaborado por el MINSAL [8].

Comuna	Razón de contactos por caso
La Florida	2.2
La Granja	2.0
La Pintana	1.9
Pirque	1.1
Puente Alto	1.7
San José de Maipo	1.2
San Ramón	1.6

Tabla 1.1: Razón de contactos por caso confirmado en las comunas que componen el Servicio de Salud Metropolitano Sur Oriente según el 125° Informe Epidemiológico del MINSAL.

Por las características de la unidad de analítica del proyecto, la contraparte directa del proyecto es la Unidad de Salud Digital (USD) del SSMSO, la cual se encarga de liderar los procesos de digitalización de la información y la inclusión de nuevas tecnologías en los establecimientos que conforman la red asistencial de salud. Esta unidad se encuentra liderada por Juan Cristóbal Morales y René Lagos, y como misión declaran mejorar la calidad de la atención de la salud a través de la implementación e integración de herramientas tecnológicas, con el objetivo de ser referentes nacionales en el área informática clínica [12].

## 1.4. Pregunta de Investigación, Objetivos Generales y Específicos

A partir del contexto nacional e internacional previamente expuesto, la pregunta de rigor que motiva la realización de esta memoria consiste en si es posible generar una herramienta de apoyo a la toma de diversas decisiones importantes en el manejo de la pandemia, entre ellas la labor de trazabilidad y búsqueda activa, otorgando una noción de priorización. Esto quiere decir que en vez de tener una lógica de revisión caso por caso, se cuente con un sistema que asigne prioridades de seguimiento a aquellas personas que el virus les representa una amenaza más allá de los síntomas leves que este puede llegar a presentar, incluso pudiendo llegar a comprometer sus vidas. En línea con la anterior, otra pregunta que motiva la realización de esta investigación consiste en si es posible agregar la situación de cada una de las personas contagiadas o probables, con el fin de llevar un control más global de las zonas delimitadas por el SSMSO y facilitar la labor de las y los trazadores.

Dada la importancia documentada de la trazabilidad y la búsqueda activa en el control de la pandemia y al propósito del proyecto, el desarrollo de una herramienta de visualización de la situación sanitaria actual no solamente brindaría una ayuda a la hora de tomar decisiones de manera fundamentada a partir del análisis de datos, sino que también lo anterior podría

significar en una reducción de los gastos públicos asignados a la salud, entendiendo que desde la irrupción del virus estos han ido en aumento.

Así, el objetivo general del trabajo consiste en apoyar la toma de decisiones de trazabilidad y búsqueda activa, incorporando nociones de priorización de casos para asignar mejor los recursos públicos con los que se cuenta, para lo cual se desarrolla un prototipo de herramienta de visualización de casos confirmados y probables, con el fin de brindar una solución que permita hacer un seguimiento inteligente de los focos de contagio más riesgosos que pueda tener una comunidad de acuerdo a sus antecedentes personales y clínicos. Este objetivo busca hacerse cargo de los bajos índices de trazabilidad reportados a nivel país y a la vez busca proporcionar un soporte a la toma de decisiones a la hora de realizar el seguimiento de pacientes.

Para realizar la herramienta de visualización que buscará responder al objetivo general, primero se deberá ubicar a cada uno de los pacientes según la dirección que hayan declarado al momento de atenderse en cualquier establecimiento de salud del SSMSO, proceso que será detallado con mayor profundidad en el capítulo 4: El modelo de geocodificación. Posteriormente, se procederá a caracterizar el riesgo clínico de aquellas personas que en la base de datos figuran como casos confirmados y probables por COVID-19, con el fin de asignar la prioridad a aquellos casos en los que sea más probable que se genere un cuadro grave. Para ello, se tomarán en cuenta diversos antecedentes declarados al momento del ingreso en cualquiera de los establecimientos clínicos del servicio, tales como la edad, si el paciente declara tener asma o diabetes, entre otros. Lo anterior se profundizará en el capítulo 5: El modelo de riesgo, donde se explicarán todas las variables consideradas para finalmente otorgar una probabilidad a cada paciente de contraer un cuadro grave producto de un contagio con el virus.

Para poder llevar a cabo la aplicación, se definen los siguientes objetivos específicos que guiarán el desarrollo del trabajo a realizar:

1. Geocodificar a los casos confirmados y probables con fechas de diagnóstico con el fin de ubicarlos en un mapa. Este trabajo fue realizado en conjunto con Matías Martínez Pérez, tesista del proyecto COVID 0251 de la Agencia Nacional de Investigación y Desarrollo.
2. Estudiar el riesgo asociado a cada uno de estos casos, considerando antecedentes clínicos.
3. Asignar cada uno de los casos a unidades censales mayores para poder tener una noción más general del comportamiento de la pandemia en las comunas del SSMSO.
4. Programar y testear la herramienta de visualización de casos confirmados/probables.

## 1.5. Alcances y limitaciones

El resultado esperado de este trabajo de título es la implementación de un prototipo de herramienta de visualización, la cual a su vez es el producto final para la contraparte. Por lo tanto, el alcance de esta memoria es hasta la implementación de la plataforma y la posterior entrega del código. No se considera dentro del alcance de la memoria la integración de la aplicación en la plataforma de seguimiento ni en ningún otro sistema de información del servicio. Dado el tiempo limitado con el cual se cuenta para el desarrollo de la aplicación,

se dejarán de lado las actualizaciones de datos permanentes de la aplicación, por lo que se dará principal énfasis en entregar un producto automatizado, el cual mediante códigos de tratamiento de bases de datos pueda realizar el análisis prometido.

Otro alcance importante al trabajo de título consiste en la consideración del esquema de vacunación realizado en el país, debido a la diferencia sustancial de datos que hay entre la cantidad de personas que manifestaron un ingreso en el sistema de salud y los datos de personas vacunadas actualizadas a la fecha. En particular Chile inició el período de inoculación masiva a fines del 2020, comenzando por los adultos mayores y trabajadores de la salud. Ya para el mes de Mayo del año 2021, Chile contaba con aproximadamente el 42 % de su población total con esquema inicial de vacunación completo, como se refleja en la figura 1.5, siendo el segundo país más rápido en inocular a su población solo por detrás de Israel.

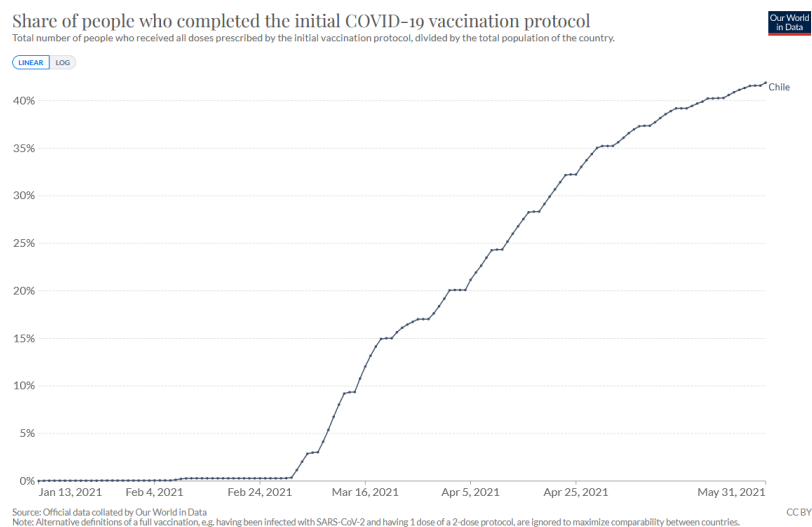


Figura 1.5: Porcentaje de la población con esquema de vacunación inicial completo (1 o 2 dosis según corresponda), publicado por Our World in Data.

Dado lo anterior, es necesario afirmar que para el desarrollo de la plataforma de visualización de casos riesgosos para el SSMSO no se consideró el esquema de vacunación para medir el posible riesgo que podría tener un paciente que haya registrado un ingreso en algún establecimiento de salud del servicio, es decir solo se consideraron antecedentes personales y síntomas declarados. Esto se debe a que se disponía de una muestra acumulada, por lo que los datos más lejanos en el tiempo, principalmente del año 2020, no iban a contar con información asociada a las vacunas debido a que en esas fechas aún se estaba investigando acerca de su eficacia.

Por otro lado, una limitación relevante consiste en que la herramienta permite su validación solo en las comunas de interés, ya que si bien el análisis se puede extrapolar a más comunas de Chile, la visualización contemplará solamente el tratamiento de las bases de datos suministradas por el servicio.

Por último, otra limitación que existe para la implementación de la plataforma consiste

en la procedencia de los datos. Existen comunas del SSMSO que no utilizaron los sistemas de seguimiento comunes desde su aparición, produciendo que la cantidad de casos confirmados o probables registrados en el período de la muestra se encuentren subestimados en relación a la incidencia total que hubo con el virus en esas comunas. Es decir, el análisis se efectuará netamente con la información que suministró el servicio, por lo que la información visualizada puede no coincidir con la situación real observada en los días correspondientes.

## 1.6. Estructura

Se expone a continuación un resumen de la estructura del presente trabajo, con el fin de facilitar la comprensión del lector:

**Capítulo 2: Marco teórico**, donde se expone el *background* de los modelos de geocodificación y de riesgo necesarios para realizar la herramienta, así como también se ejemplifican algunas aplicaciones de ambos modelos en problemas relacionados a la vida cotidiana o a la pandemia. También, se revisan algunos ejemplos de herramientas de visualización creadas en contextos de pandemia.

**Capítulo 3: La muestra**, donde se explican las variables incluidas en la base de datos suministrada por el SSMSO. En esta sección se caracterizará a las personas que registraron una visita a un establecimiento de salud mediante un análisis exploratorio de los datos, así como también se procederá a destacar las variables relevantes a considerar tanto en el modelo de geocodificación como en el modelo de riesgo.

**Capítulo 4: El modelo de geocodificación**, en el cual se explicará a grandes rasgos como se realizó la ubicación de pacientes dentro de un mapa generado en Python mediante OpenStreetMap y Google Maps. En esta sección se detallará el proceso de normalización de direcciones realizado en conjunto con Matías Martínez, junto con el posterior procesamiento de estas y la asignación de cada persona a unidades censales más grandes, con el objetivo de llevar un seguimiento más agregado de la evolución del virus en las comunas de interés.

**Capítulo 5: El modelo de riesgo**, sección donde se detalla el modelo de riesgo clínico utilizado a lo largo de la ejecución del proyecto. Se procederá a revisar y ejemplificar cuales son los inputs y outputs de este modelo y como este se relaciona con la herramienta de visualización de casos desarrollada.

**Capítulo 6: La herramienta de visualización**, donde se detallará la implementación de la herramienta diseñada para el SSMSO. Se comenzará revisando algunos ejemplos de herramientas similares utilizadas a lo largo del mundo en el marco de la pandemia, para luego profundizar en la herramienta de visualización de casos riesgosos realizada mediante la librería *streamlit* de Python.

**Capítulo 7: Conclusiones**, se exponen las principales conclusiones del trabajo, verificando el cumplimiento de los objetivos planteados y respondiendo las preguntas de investigación inicialmente declaradas.

# Capítulo 2

## Marco Teórico

Para realizar la herramienta de visualización de casos riesgosos se hace necesario distinguir entre algunos conceptos claves para entender el por qué se seleccionan solo a un subconjunto de personas, las calificadas como confirmadas y probables, y no así a todas las personas de la muestra, así como también comprender cual es el período de tiempo a considerar para una persona que fue diagnosticada como confirmada o probable, ya que se entiende que una persona es contagiante solo durante un período limitado de tiempo. Todo lo anterior se revisará a lo largo de este capítulo en la sección de definiciones relacionadas al COVID-19.

Además de los conceptos claves relacionados al COVID-19, durante la elaboración de la herramienta de visualización se utilizan dos técnicas ampliamente estudiadas en la literatura que son necesarias para lograr el objetivo general del proyecto: la geocodificación y la asignación de riesgo a cada persona. Ambas técnicas tienen múltiples aplicaciones a problemas de la vida cotidiana del ser humano, las cuales serán profundizadas en las secciones respectivas.

Al final de este capítulo se revisarán algunos ejemplos de herramientas de visualización desarrolladas tanto a nivel local como a nivel global, comparando sus niveles de agregación de datos y estadísticas visualizables, con el fin de obtener una noción de similitud entre las plataformas e identificar posibles espacios de diferenciación.

### 2.1. Definiciones relacionadas al COVID-19

Para entender la selección de las personas a considerar en la plataforma, se debe distinguir a las personas que realmente signifiquen un potencial problema para otros seres humanos. Basado en un documento [13] elaborado por la Organización Mundial de la Salud, se definen todos los casos posibles por los cuales puede ser catalogada una persona con respecto a la patología:

1. Para ser catalogado como **caso sospechoso**, debe cumplir al menos uno de los siguientes escenarios:
  - La persona cumple con el criterio clínico y epidemiológico de la enfermedad (revisar en Anexos la sección “Criterio Clínico y Epidemiológico del COVID-19”).
  - La persona presenta síntomas de infección respiratoria aguda grave.
  - La persona es asintomática y no cumple con el criterio epidemiológico, pero cuenta con un resultado positivo del test rápido de antígenos.

2. Para ser catalogado como **caso probable**, debe cumplir al menos uno de los siguientes escenarios:
  - La persona cumple con el criterio clínico y es contacto estrecho de un caso confirmado o probable.
  - La persona es un caso sospechoso con imágenes de tórax que sugieren patologías por COVID-19.
  - La persona presenta síntomas de anosmia (pérdida del olfato) y ageusia (pérdida del gusto)
  - La persona fallece producto de una dificultad respiratoria anterior a la muerte y que además fue un contacto estrecho de un caso probable o confirmado
3. Para ser catalogado como **caso confirmado**, debe cumplir al menos uno de los siguientes escenarios:
  - La persona presenta un resultado positivo en un test de amplificación de ácidos nucleicos (NAAT por su sigla en inglés)
  - La persona presenta un resultado positivo del test rápido de antígenos y a la vez cumple con la definición de caso probable, o con alguno de los dos primeros escenarios de un caso sospechoso.
  - La persona presenta un resultado positivo del test rápido de antígenos y a la vez es un contacto estrecho de un caso probable o confirmado.

Dado lo anterior, un caso sospechoso de COVID-19 no cuenta con evidencia suficiente como para poder significar un real peligro, no así los casos probables y confirmados, los cuales si cumplen con evidencia que permite inferir que se porta el virus. Con respecto al tipo de ingresos, se hace una distinción entre los casos índices y secundarios: en el “Protocolo de coordinación para acciones de vigilancia epidemiológica durante la pandemia COVID-19 en Chile: Estrategia nacional de Testeo, Trazabilidad y Aislamiento” [14] desarrollado por el Ministerio de Salud, se indica que los casos índice corresponden a personas con cuadro clínico que cumple con las características del caso sospechoso, probable o confirmado de COVID-19 y que su detección da inicio a la investigación epidemiológica e identificación de contactos estrechos, es decir son casos a los cuales no se le puede detectar una fuente previa de contagio. Por otro lado, los casos secundarios son personas a las cuales se les puede identificar su fuente de contagio. Es por ello que ambos tipos de pacientes son candidatos a localizar si es que se trata de casos confirmados o probables.

Por último, según las definiciones del mismo protocolo anterior, se considera el período en que un paciente es activo por COVID-19 como 14 días posteriores al inicio de los síntomas, período en el cual se hará necesario localizar al paciente para frenar su movimiento.

## 2.2. La geocodificación

Como primer paso para lograr un seguimiento de los pacientes confirmados/probables registrados, se hace necesario localizarlos según el domicilio que estos hayan registrado al momento de hacer su ingreso en el establecimiento de salud. Para ello se utiliza el proceso de geocodificación o *geocoding*, el cual consiste en convertir un texto de dirección en un par

ordenado de coordenadas de latitud y longitud, las cuales pueden ser representadas en un mapa. Actualmente existen distintas herramientas o programas que realizan el proceso, como lo son Google Maps, OpenStreetMap, ArcGIS, QGIS, entre otros. La figura 2.1 ilustra un ejemplo del proceso de geocodificación:

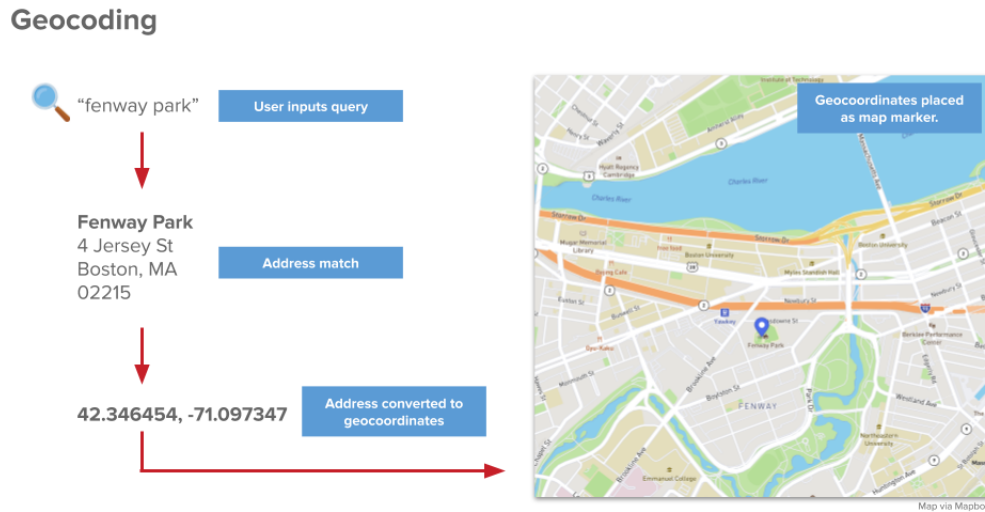


Figura 2.1: Ejemplo de geocodificación de dirección en Boston. Ejemplo tomado de PubNub.

Las aplicaciones en problemas de la vida cotidiana de la geocodificación son variadas e incluyen casos como lo son la localización de objetos u artículos mediante la logística, el ruteo de vehículos (comúnmente conocido como VRP o *Vehicle Routing Problem* por sus siglas en inglés), *geomarketing* para entender el comportamiento de clientes de determinadas zonas del mundo y orientar productos hacia clientes dentro de datos demográficos específicos, entre otras. En particular, para el área de la salud en la cual está inserta el propósito de esta memoria, mediante la geocodificación es posible evaluar el acceso de los pacientes a los centros de salud y a la vez estudiar patrones epidemiológicos de una enfermedad. Por ejemplo, se ha estudiado la aplicación de la geocodificación en pacientes con cáncer [15] donde los autores revisan tres métodos específicos de geocodificación:

1. Asignación de una observación (paciente con cáncer) a una unidad geográfica sin especificar la ubicación específica dentro de la unidad.
2. Interpolación de coordenadas basada en una distancia proporcional entre la dirección en un registro y el rango de direcciones existentes en una calle específica.
3. *Parcel Matching Method* donde cada "paquete" o registro cuenta con una dirección y se le asigna una coordenada basada en el centroide de la ubicación

Donde se señala que el último método es el más certero y más utilizado en ciudades y condados de Estados Unidos. Más aún, para el contexto de pandemia actual ya se ha estudiado la utilidad de la geocodificación para el control de la pandemia. Dos investigadores de la Universidad de Batna, Algeria, destacan en su estudio [16] la creación y utilización de

una herramienta de geocodificación utilizada en la ciudad de Ain Touta, donde no se cuenta con un sistema efectivo de direccionamiento geoespacial y que catalogan como un problema de gestión urbana común en Algeria. El objetivo de esta herramienta consistía en reducir el estrés de los sistemas de salud de la ciudad, debido a que gracias a esta se podría llevar a cabo un seguimiento particular a los pacientes en sus domicilios, evitando el colapso de los sistemas de salud de la ciudad. Como resultados, observaron una contribución significativa al éxito de las cuarentenas y del aislamiento social, así como también una mejora sustantiva en la flexibilidad y la capacidad del sistema de salud, pasando de 40 pacientes atendidos en hospital a más de 200 pacientes atendidos mediante salud a domicilio.

### 2.3. Asignación de riesgo

Por otro lado, para poder comprender quienes de las personas localizables en la herramienta de visualización necesitan especial foco de atención de acuerdo a sus características personales y clínicas, se hace necesario otorgar una noción de riesgo a cada paciente con el fin de poder diferenciarlos de acuerdo a su potencial gravedad al adquirir el virus. En estadística se utilizan comúnmente los modelos de regresión, los cuales permiten estimar las relaciones de dependencia entre una variable dependiente  $Y$  y un set de  $m$  variables independientes  $X_i$ . Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijos.

Nuevamente, el hecho de asociar un riesgo predicho a alguna persona o cliente de acuerdo a un set de características que presenta es un problema cotidiano en múltiples industrias, en las que destacan principalmente las industrias financieras y de seguros, donde a través de distintos parámetros se puede predecir, por ejemplo, el riesgo de que un cliente no pueda seguir pagando un préstamo bancario (conocido como el riesgo de *default*) o, por otro lado, se puede calcular el precio adecuado que debe pagar un cliente por la póliza de seguros que contrata de acuerdo al riesgo que este presente de tener que cobrar dicha póliza.

Por ejemplo, el 5 de Mayo del 2020 se publicó un estudio [17] donde se realizó un modelo de predicción de riesgo de crédito de una muestra de clientes de una institución bancaria en Portugal. El objetivo primordial era predecir el comportamiento de pago de estos clientes, para así tener más certeza acerca de si terminarían cayendo en *default* financiero (clientes riesgosos) o no. Para este ejemplo en particular, como la variable a explicar consiste en si el cliente entre en *default* o no (variable binaria), la variable dependiente  $Y$  sigue una distribución Bernoulli:  $Y_i \sim Ber(\mu_i)$ . Esta variable intenta ser predecida a partir de un set de variables cuantitativas y cualitativas:



Variable	Descripción
<i>Contracted Capital</i>	Capital negociado en el acuerdo (Euros)
<i>Capital Outstanding</i>	Capital que se debe al banco (Euros)
<i>Spread</i>	Diferencia entre tasa de interés cobrada y pagada por el banco (%)
<i>Term</i>	Período de tiempo del préstamo (años)
<i>Monthly Installment</i>	Cuota mensual a pagar al banco (Euros)
<i>Age</i>	Edad del prestatario (años)
<i>Seniority</i>	Años en que el prestatario ha sido cliente del banco (años)
<i>Credit Cards</i>	Número de tarjetas de crédito que posee el prestatario

Tabla 2.1: Variables cuantitativas del estudio

Variable	Descripción
<i>Sex</i>	Género del prestatario
<i>Marital Status</i>	Estado civil del prestatario
<i>Salary</i>	Indica si el salario se recibe en una cuenta del banco o no
<i>Other Credit</i>	Indica si el cliente tiene otros créditos
<i>Tax Echelon</i>	Indica el escalón de ingreso del prestatario
<i>Default</i>	Indica si el cliente se encuentra actualmente en <i>default</i>

Tabla 2.2: Variables cualitativas del estudio

A partir de las variables anteriores, las autoras proponen un modelo de regresión logística para explicar el riesgo de *default*:

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9$$

Donde  $\mu$  es la media de la variable *default* (y representa la probabilidad de *default*),  $X_1$  es la variable *Spread*,  $X_2$  es la variable *Term*,  $X_3$  es la variable *Age*,  $X_4$  es la variable *Credit Cards*,  $X_5$  es una variable *dummy* para cuando *Salary* = 1,  $X_6$  es una variable *dummy* para cuando *Tax Echelon* = 2,  $X_7$  es una variable *dummy* para cuando *Tax Echelon* = 3,  $X_8$  es una variable *dummy* para cuando *Tax Echelon* = 4 y  $X_9$  es una variable *dummy* para cuando *Tax Echelon* = 5. Como conclusiones relevantes del estudio, se observa que las variables relevantes para explicar el riesgo de *default* fueron *Spread*, *Term*, *Age*, *Credit Cards*, *Salary* y *Tax Echelon* y que a la vez el riesgo de *default* aumentaba si lo hacían las variables *Spread*, *Term* y *Age*, pero disminuía si aumentaba la variable *Credit Cards*. Con respecto a las variables *Salary* y *Tax Echelon*, se observó que los clientes que recibían su salario en la misma institución financiera tenían menor riesgo de *default* que aquellos que lo recibían en otra institución, mientras que los clientes de escalones salariales más bajos son más propensos a entrar en riesgo de *default* que los de escalones más altos. Lo anterior ilustra como el riesgo puede ser explicado a partir de distintas variables tanto intrínsecas de la persona como externas.

Más aún, para el contexto de pandemia actual se ha revisado el riesgo de pacientes de acuerdo a las posibles comorbilidades que estos presentan. Un estudio [18] realizado en México

con datos de aproximadamente medio millón de personas concluye que el riesgo de mortalidad aumenta aditivamente de acuerdo a la presencia de más de una comorbilidad asociada a cada paciente, entre las que destacan la presencia de obesidad, diabetes, hipertensión, enfermedades renales crónicas y enfermedades cardiovasculares. Los autores utilizan un modelo bayesiano jerárquico donde la variable dependiente, al igual que en el ejemplo anterior, fue un indicador binario que indicaba si la persona contagiada por COVID-19 moriría o no, la cual distribuía Bernoulli. Las variables utilizadas para la predicción radica en información socio-demográfica (edad, sexo, origen étnico), datos geográficos (estado y municipio de residencia), variables auxiliares del paciente (asma, tabaquismo, inmunosupresores y otras enfermedades), e información sobre el momento del diagnóstico e intervención sanitaria para controlar la dependencia del tiempo. En la figura 2.2 se presentan los *odds ratio* producto del estudio, los cuales indican la posibilidad que una condición de salud o enfermedad se presente en un grupo de población frente al riesgo que ocurra en otro. Para este estudio, refleja la relación de cada variable característica de los grupos de interés (diabéticos, hipertensos, fumadores, entre otros) con el riesgo de mortalidad.

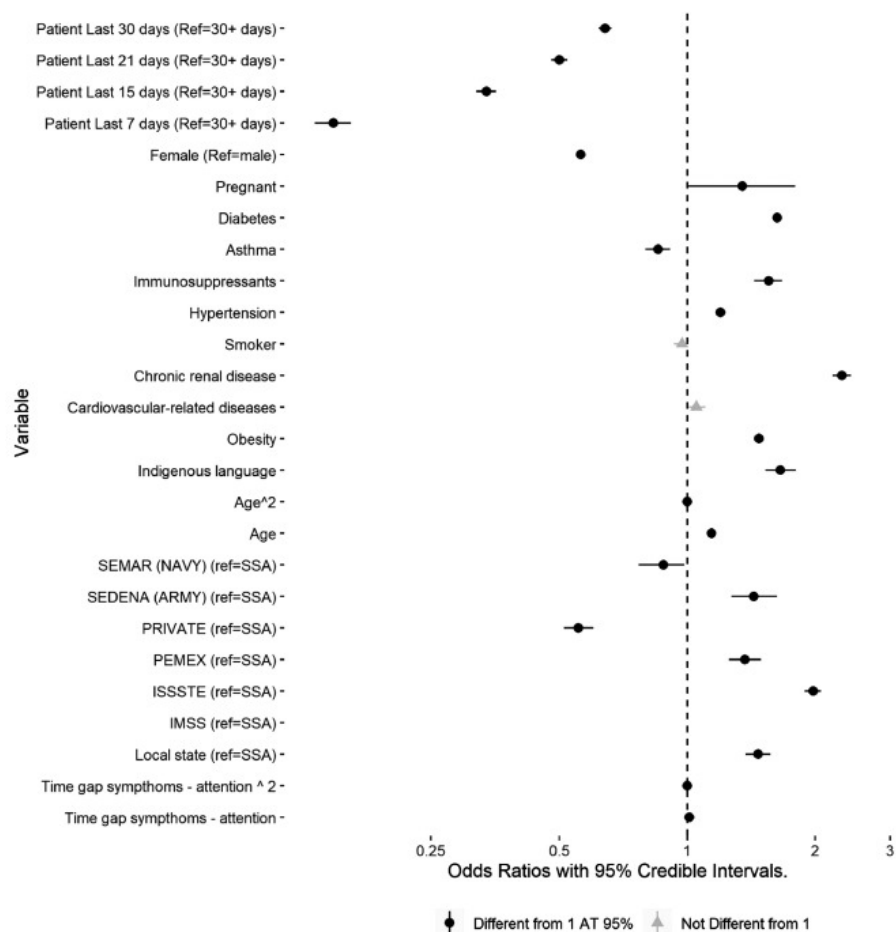


Figura 2.2: Odd Ratios con IC del 95 % de confianza del estudio. Obtenido del mismo.

Como conclusiones, obtienen que la diabetes, la obesidad, la hipertensión y la enfermedad renal crónica aumentaron las probabilidades de muerte relacionada con la COVID-19, donde la enfermedad renal crónica tuvo el mayor efecto entre las comorbilidades. El estudio también señala la diferencia entre la atención en distintos establecimientos de salud, donde pacientes

atendidos en clínicas privadas y dentro del sistema de salud de la marina tenían menores chances de morir que aquellos que se atendían en establecimientos públicos.

## 2.4. Herramientas de visualización desarrolladas

Con respecto a las plataformas de visualización de casos desarrolladas en el marco de la pandemia, existen múltiples organizaciones y universidades que han realizado toda clase de mapas interactivos para conocer estadísticas referentes al COVID-19. A la vez, existen mapas desagregados a nivel de países, a nivel de estados, comunas, llegando hasta cuadrantes urbanos.

Un ejemplo de plataforma de COVID-19 mundial es la desarrollada por la Universidad de Johns Hopkins [19], la cual cuenta con datos relacionados a la totalidad de casos, totalidad de muertes y totalidad de vacunas suministradas históricamente y en los últimos 28 días para cada uno de los países, llegando a un nivel comunal y en el caso de Estados Unidos, a un nivel urbano. A la vez, la OMS [20] creó un dashboard virtual donde se pueden visualizar la cantidad de casos, muertes y estado de vacunación histórico, en los últimos 7 días, entre otros.

Por otro lado, Mayo Clinic [21] realizó un mapa para Estados Unidos, con datos respecto a la cantidad de casos diarios por 100.000 personas, totalidad de casos, tasa de positividad y tasa de fatalidad para cada uno de los estados que componen el país. Otro ejemplo de plataforma de casos a nivel más regional es la desarrollada por el Centro Nacional de Epidemiología (CNE) de España [22], el cual desagrega información relacionada a la incidencia de casos por 100.000 habitantes para todas las regiones de España.

Finalmente, a nivel comunal y cuadrante urbano existen algunas plataformas desarrolladas en Chile como lo es la plataforma desarrollada por el equipo DATA UC de la Pontificia Universidad Católica de Chile y el Visor Territorial COVID-19 del Ministerio de Bienes Nacionales [23]. La primera permite ver la cantidad de casos activos y la tasa de activos por cada 100 mil habitantes tanto a nivel comunal como en cada región de Chile y la segunda de ellas permite ver los casos activos, confirmados, recuperados, índices de positividad y de hospitalizados UCI, a nivel de cuadrante urbano, lo cual quiere decir que se generan divisiones cuadrangulares a nivel de ciudad o pueblo. Lo anterior cobra relevancia para la posterior segmentación que procederá a realizarse en la plataforma implementada, dado a que se cuenta con información a nivel de manzana-entidad por lo que el siguiente paso corresponderá agruparlas para simplificar el número de focos de contagio a visualizar.

Al revisar todas las plataformas anteriores, un atributo común que comparten radica en visualizar la información a través de mapas de calor con respecto a la densidad de casos o nivel de vacunación en algún país, estado, comuna y ciudad. Esto cobra relevancia ya que la plataforma incluirá un mapa de calor que represente el riesgo de desenlace por cuadro grave de COVID-19 en las distintas agrupaciones zonales que resulten de la segmentación de las comunas del SSMSO. Por otro lado, se identifican oportunidades de diferenciación respecto al nivel de agregación de la información, llegando a ser de carácter individual y a la vez de carácter más grupal, mediante la información suministrada por el Censo. De este modo, ambos datos se complementan, otorgando la noción de riesgo a partir de los datos individuales y la cantidad de adultos mayores a través de datos del censo, variables que permitirán realizar

una priorización en la atención y en los sitios en los cuales realizar búsqueda activa, según la situación epidemiológica en determinadas zonas de la capital. En la sección de Anexos “Herramientas de visualización desarrolladas” se encuentran imágenes de cada una de las plataformas revisadas anteriormente.

# Capítulo 3

## La muestra

### 3.1. Datos importantes

Para la realización de esta memoria se cuenta con una base de datos suministrada por el SSMSO, la cual registra información de los pacientes desde Abril del 2020 hasta Mayo del 2021. Esta base de datos se compone por tres tablas de interés:

1. **PERSONAS:** Detalla a las personas que han hecho ingreso a algún establecimiento de salud del servicio por motivos relacionados al COVID-19. En ella se encuentran las variables:
  - *IDPER* como identificador único de cada paciente.
  - *IDINGRESO* como identificador único del ingreso al establecimiento.
  - *TIPOINGRESO* la cual señala el motivo del ingreso del paciente. Esta variable puede ser Confirmado Índice, Confirmado Secundario, Probable Índice, Probable Secundario, Contacto, Sospechoso o Descartado.
  - *EDAD* y *GÉNERO*, variables descriptivas de cada paciente particular.
  - *FECHAINICIOCUARENTENA* y *FECHAINICIOSINTOMAS*, variables que señalan el inicio de cuarentena y de síntomas del paciente registrado respectivamente.
  - *TRABAJADORDESALUD* y *CONTACTODIRECTO*, variables que toman el valor 1 si la persona es trabajador de la salud o fue contacto directo y 0 en caso contrario.
2. **EXÁMENES:** Detalla los exámenes PCR que han sido tomados a las distintas personas de las cuales se tiene registro. En ella se encuentran las variables:
  - *IDPER* como identificador único de cada paciente.
  - *IDINGRESO* como identificador único del ingreso al establecimiento.
  - *FECHARESULTADOEXAMEN* la cual señala la fecha en que se obtuvo el resultado del PCR tomado al paciente.
  - *RESULTADOEXAMEN* la cual toma los valores 1, 2, 3, 4 o 5 dependiendo del resultado obtenido: Positivo, Negativo, Pendiente, No Concluyente o Problemas con la muestra respectivamente.
3. **DIRECCIONES:** Detalla la dirección y comuna que declara el paciente al momento de ser atendido en un centro de salud. En ella se encuentran las variables:

- *IDPER* como identificador único de cada paciente.
- *DIRECCION* la cual detalla la dirección que declara el paciente.
- *COMUNA* la cual detalla la comuna que declara el paciente.

Las siguientes tablas resumen la información previamente explicada:

Tabla PERSONAS	
Variable	Tipo de variable
<i>IDPER</i>	Identificador Único
<i>IDINGRESO</i>	Identificador Único
<i>TIPOINGRESO</i>	Texto o String
<i>EDAD</i>	Numérica
<i>GÉNERO</i>	Texto o String
<i>FECHAINICIOCUARENTENA</i>	Fecha
<i>FECHAINICIOSINTOMAS</i>	Fecha
<i>TRABAJADORDESALUD</i>	Binaria
<i>CONTACTODIRECTO</i>	Binaria

Tabla 3.1: Variables en la tabla PERSONAS de la muestra de datos.

Tabla EXÁMENES	
Variable	Tipo de variable
<i>IDPER</i>	Identificador Único
<i>IDINGRESO</i>	Identificador Único
<i>FECHARESULTADOEXAMEN</i>	Fecha
<i>RESULTADOEXAMEN</i>	Numérica

Tabla 3.2: Variables en la tabla EXÁMENES de la muestra de datos.

Tabla DIRECCIONES	
Variable	Tipo de variable
<i>IDPER</i>	Identificador Único
<i>DIRECCION</i>	Texto o String
<i>COMUNA</i>	Texto o String

Tabla 3.3: Variables en la tabla DIRECCIONES de la muestra de datos.

La información que condensan estas tres tablas sirven como *input* para desarrollar tanto los modelos de geocodificación como de riesgo de pacientes, lo cual posteriormente se condensa en la herramienta de visualización y seguimiento de casos riesgosos desarrollada para el servicio.

Es relevante destacar que solo se utilizaron registros confirmados o probables para el desarrollo de la herramienta de visualización. Además, de las personas confirmadas y probables

también se utilizó un subconjunto de personas, correspondientes a las personas que fue posible localizar con el algoritmo desarrollado por Matías Martínez. Lo anterior se debe a la existencia de *missing values* y direcciones incompletas o mal escritas que el algoritmo no fue capaz de ubicar dentro del mapa. Lo anterior se profundizará en el capítulo 4: el modelo de geocodificación, donde se detallarán los principales hallazgos del modelo desarrollado en colaboración con Matías Martínez en el contexto del proyecto ANID.

## 3.2. EDA

En primera instancia se realizó un análisis exploratorio de los datos, con el objetivo de comprender y caracterizar a las personas que han ingresado a los establecimientos de salud del SSMSO. El período de muestras de exámenes PCR registrados comprende las fechas entre Abril del 2020 hasta Mayo del 2021, considerando que una misma persona (*IDPER*) puede haberse tomado más de un examen en el rango de fechas.

Para facilitar el entendimiento, a continuación se muestran los resultados del análisis a partir de cada una de las tablas señaladas al inicio del capítulo:

### 3.2.1. Personas

A partir de los datos otorgados, se observa que la muestra contiene la información de los resultados de exámenes PCR, antecedentes clínicos y personales de 145.454 personas distintas, los cuales representan un 9,72% de la población total de las comunas de interés. Según el 125° Informe Epidemiológico, a fecha del 31 de Mayo del 2021 la población contagiada por COVID-19 en las comunas de interés asciende a las 147.803 personas, número que dista ligeramente de lo entregado en la base de datos. Por otro lado, estas personas realizaron 172.887 ingresos a establecimientos de salud que componen el SSMSO, cuyos resultados del examen pudo haber sido positivo, negativo, pendiente, no concluyente o problemas con la muestra. Es decir, existen personas que ingresaron más de una vez al sistema debido a que pudieron haberse contagiado y/o sido un caso probable más de una vez.

La figura 3.1 muestra que la muestra se compone en su mayoría por jóvenes entre 15 y 29 años. los cuales representan un cuarto de la muestra, seguidos de las personas entre 30 y 44 años, las cuales representan un 22.4% de la muestra. El resto de grupos etarios con una proporción menor al 20% la componen las personas entre 45 y 59 años, las personas con 60 años o más y las personas entre 0 a 14 años, con proporciones de la muestra del 19.8%, 18% y 14.9% respectivamente. A la vez, la figura 3.2 muestra la distribución por género donde, de la totalidad de pacientes registrados, el 53% corresponde a personas del género femenino, 45.7% corresponde a personas del género masculino y 1.2% corresponde a personas que al momento de ingresar no declararon su género.

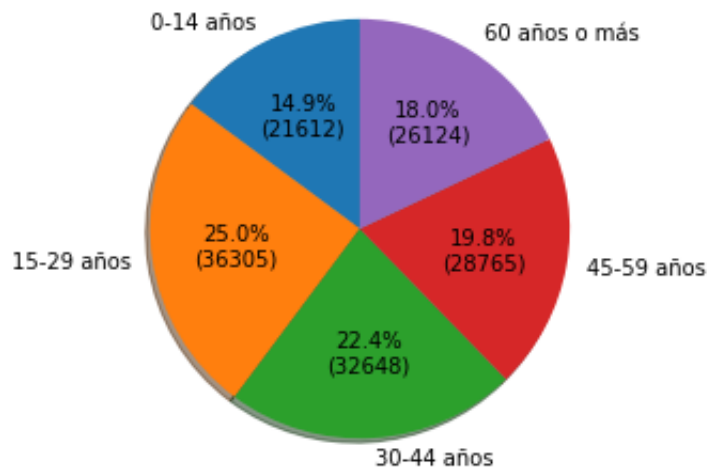


Figura 3.1: Distribución de la edad de las personas que componen la muestra.

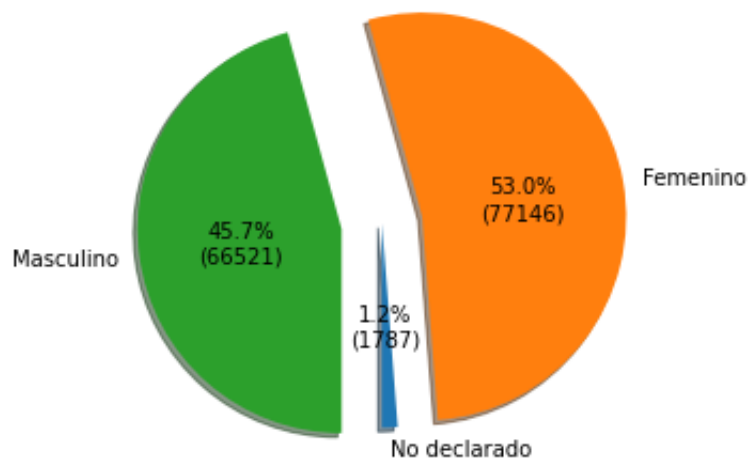


Figura 3.2: Distribución de género de las personas que componen la muestra.

A continuación se muestra la distribución de los ingresos realizados por las personas a lo largo del período de la muestra. De este gráfico, se observa que la mayoría de ingresos registrados corresponden a casos descartados índice, es decir personas cuyo ingreso a un establecimiento del SSMSO se debe a que cumplen con características de caso sospechoso, probable o confirmado y cuyo origen de estas características no pueda ser determinado previamente y que a la vez su resultado PCR no registra tener la enfermedad de acuerdo a los criterios mencionados en el marco teórico. Le siguen ingresos por contacto, confirmados índice, sospechosos, descartados, probables secundario, confirmados secundario, probables índice y descartados secundario.



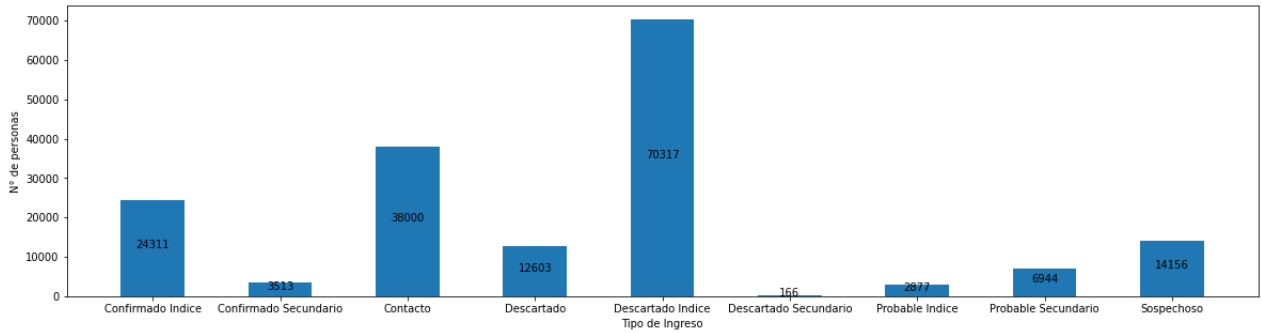


Figura 3.3: Distribución de ingresos registrados a lo largo del período.

### 3.2.2. Exámenes

Con respecto a los exámenes registrados en el período de muestra, la base de datos incluye la información de 278.384 resultados de exámenes PCR realizados en distintos establecimientos de salud del SSMSO. El siguiente gráfico muestra la distribución de todos los resultados obtenidos durante el período de análisis, donde en su mayoría se presentan exámenes con resultado pendiente a la hora del ingreso, seguido de exámenes negativos y positivos. Solo una poca cantidad de exámenes reciben resultados no concluyentes o presentan problemas con los resultados.

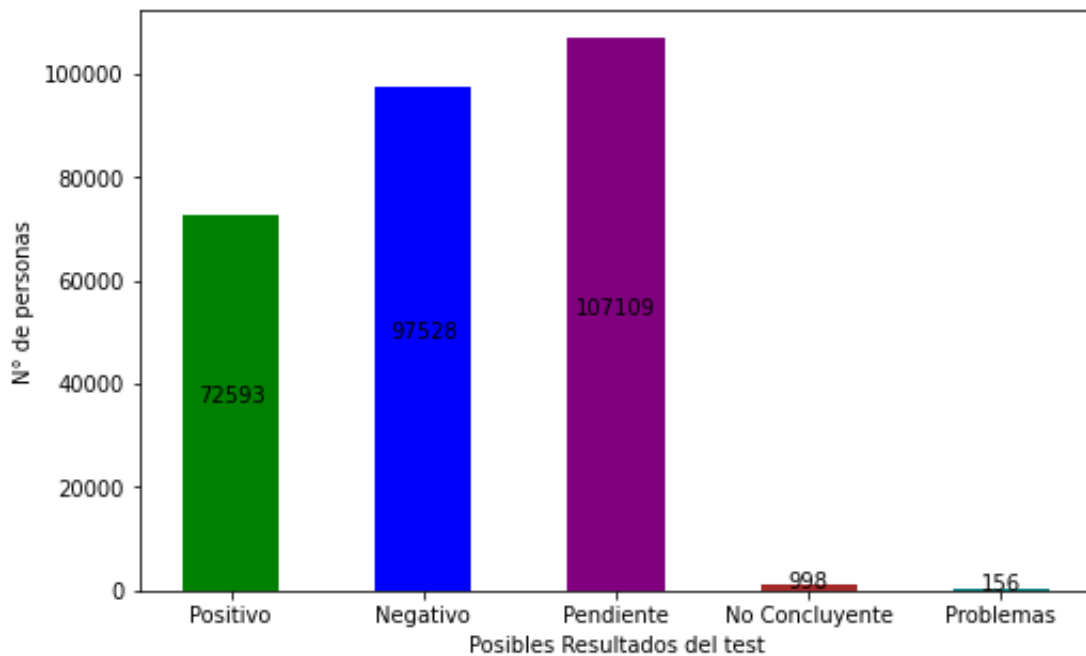


Figura 3.4: Número de exámenes PCR, desagregados en sus distintos resultados.

Por otro lado, se observa la distribución de exámenes tomados en cada mes desde Abril 2020 hasta Mayo 2021. De la figura 3.6 se observa que el mes con más exámenes PCR tomados

durante el año 2020 fue en Agosto con 36.443 resultados, mes en que coincide el término de la primera ola de contagios en Chile producto de la estabilidad de muertes diarias por el virus, al situarse a fecha 8 de Mayo por primera vez bajo las 50 muertes diarias según las cifras oficiales del gobierno [24]. Por otro lado, en 2021 fue el mes de Abril, con 47.631 exámenes, seguido de cerca por el mes de Marzo con 42.209, coincidentemente durante estos meses se vivía la segunda parte de la segunda ola de contagios en el país, la cual se manifestó producto de las celebraciones de fin de año y el período de vacaciones en Chile, junto con la apertura parcial de focos turísticos del país. Si bien en Marzo del 2021 ya se detectaba un aumento considerable de los casos y muertes asociadas al virus, el 9 de abril del 2021 se alcanzaron los 9.000 casos nuevos diarios, mientras que el día 13 abril se llegó a un máximo histórico de 141 defunciones diarias para ese momento.

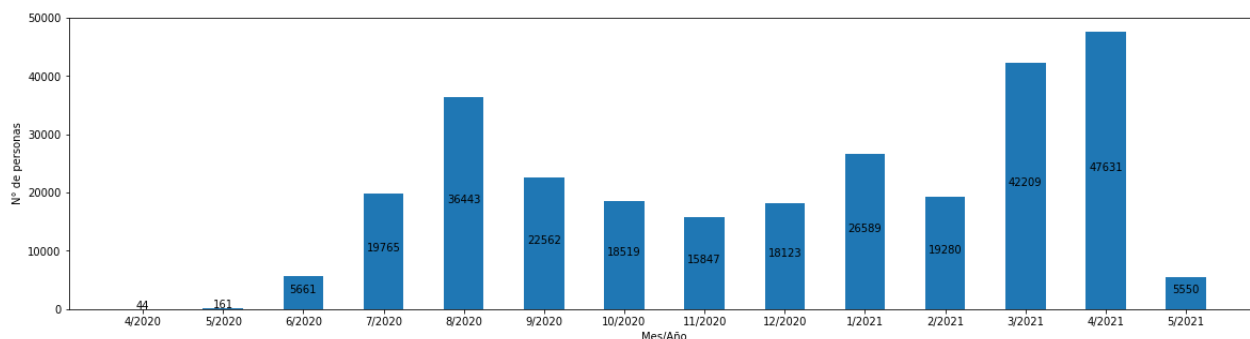


Figura 3.5: Número de exámenes PCR tomados en cada mes del período de la muestra.

Por otro lado, se observa que los casos positivos alcanzaron su máximo el mes de Julio para el año 2020 y el mes de Abril para el año 2021 en los establecimientos de salud del SSMSO, lo cual coincide con la realidad nacional explicada anteriormente.

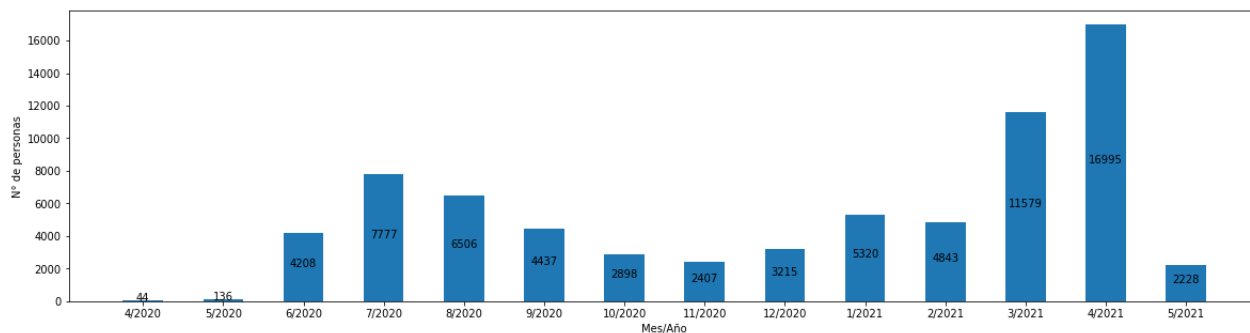


Figura 3.6: Número de exámenes PCR con resultado positivo registrados cada mes del período de la muestra.

### 3.2.3. Direcciones

Con respecto a las direcciones de los pacientes, se cuentan con un subconjunto de 30.250 direcciones de los pacientes únicos catalogados como Confirmados y Probables (sean índice o secundario). Si bien la distribución de ingresos muestra que la suma de estas personas equivale a 37.645 personas, para efectos del análisis exploratorio de datos el primer número es el relevante ya que a este se le eliminaron los duplicados de personas y/o las direcciones vacías, es decir el segundo número cuenta con personas que hayan ingresado como caso confirmado o sospechoso más de una vez y personas sin una dirección declarada.

Por otro lado, de los 30.250 casos confirmados y probables, se tiene que solo 25.667 personas corresponden a habitantes de comunas que componen el SSMSO y que a la vez fue posible otorgarles un riesgo con el modelo, debido a que parte de los 30.250 pacientes fueron asignados a localidades ajenas a las comunas del SSMSO y/o no contaban con registros claros acerca de sus síntomas declarados, por lo que las personas restantes se excluyeron del análisis debido al alcance de la memoria y el trabajo conjunto con el servicio de salud. La figura 3.8 resume el recorte de direcciones para la realización del análisis exploratorio:

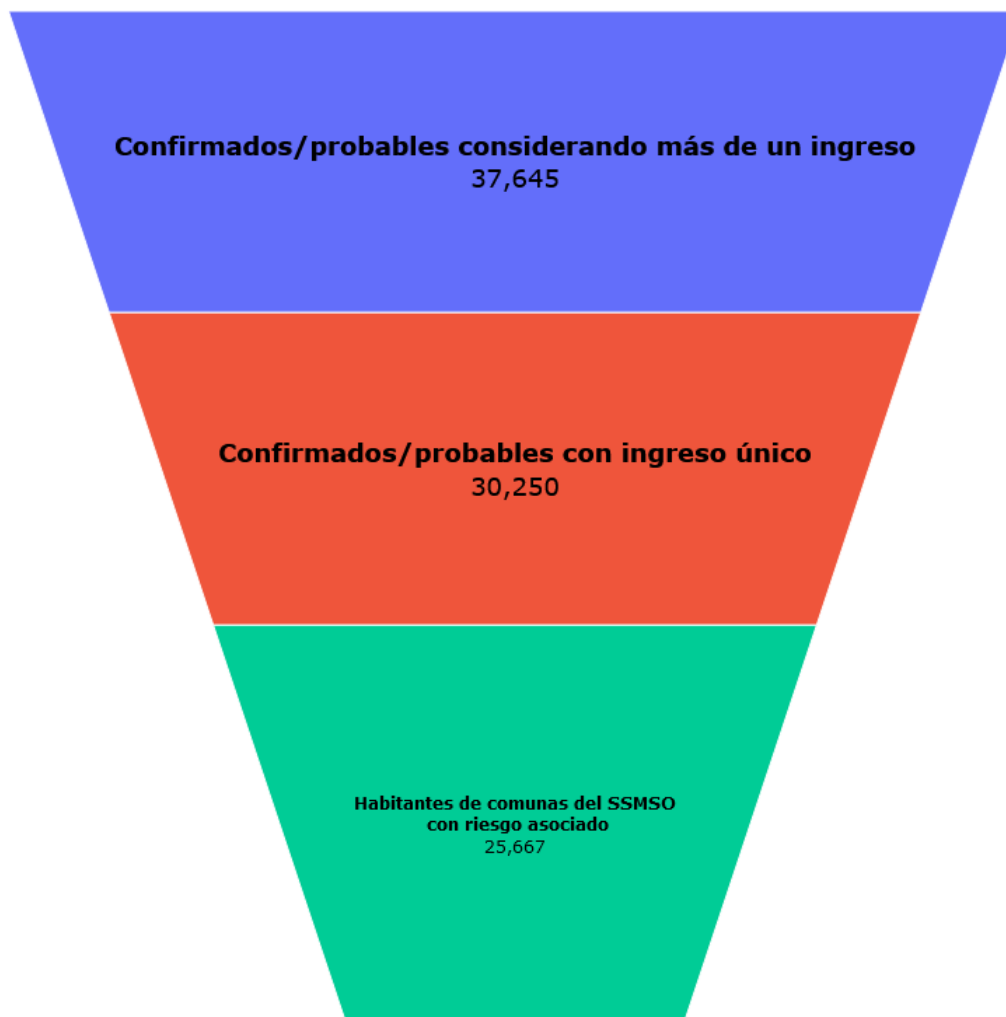


Figura 3.7: Diagrama de embudo que explica el recorte del total de direcciones.

A continuación se muestra la distribución de las comunas de residencia de las 25.667 personas únicas confirmadas/probables de la muestra en la figura 3.9. Se desprende que la comuna con mayor cantidad de casos a lo largo del período de muestra es La Florida, seguida de La Pintana y en menor medida la comuna de La Granja. Es relevante notar que la comuna con más habitantes de Chile, Puente Alto, presenta un número bajo de pacientes confirmados o probables en la muestra. Según el encargado de la unidad de salud del SSMSO, René Lagos, esto se debe a que los habitantes de la comuna no utilizaban la plataforma para registrar los casos, por lo cual es probable que este número, para efectos de la realidad epidemiológica de la comuna, esté subestimado. Por otro lado, se observa que las comunas de San José de Maipo y Pirque cuentan con una cantidad pequeña de personas cuyo resultado del test PCR fue positivo o fueron catalogadas como casos probables, a lo largo del período de la muestra. Al igual que en el caso de Puente Alto, en estas comunas la plataforma no fue utilizada continuamente, sumado a que estas comunas cuentan con una menor cantidad de habitantes que explican estos números tan bajos.

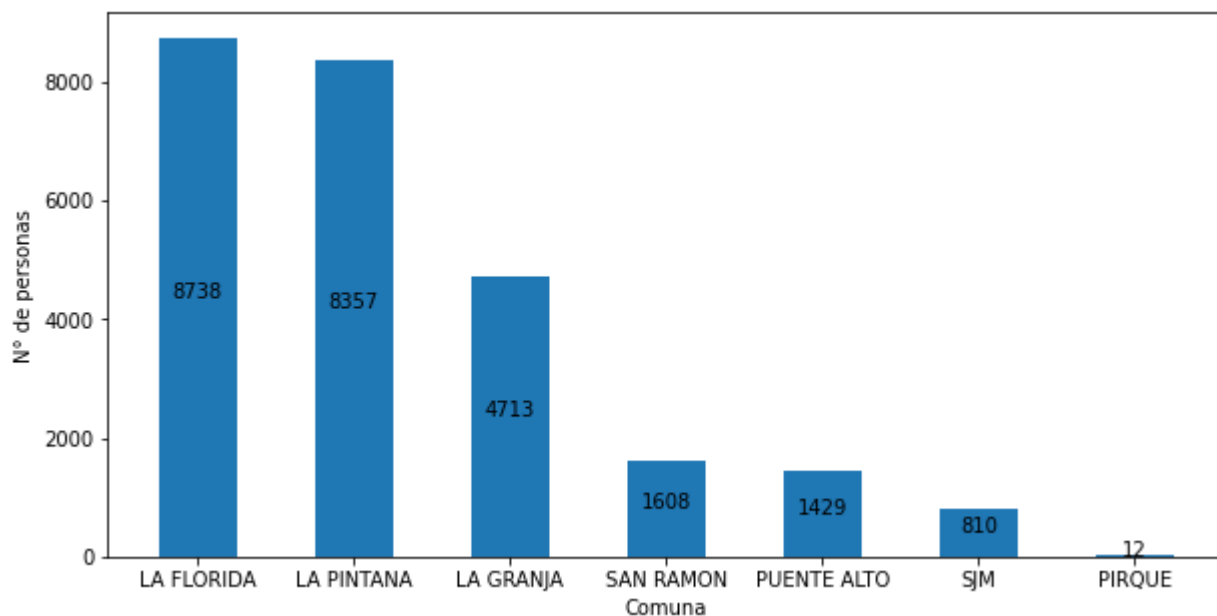


Figura 3.8: Número de pacientes confirmados/probables que habitan cada comuna del SSMSO.

La tabla 3.4 resume la población total de las comunas de interés y a la vez contrasta la información obtenida en el 125° Informe Epidemiológico con lo obtenido en la muestra de datos suministrada por el SSMSO. En esta destacan la poca representatividad obtenida para las comunas de Puente Alto y Pirque, lo cual refuerza lo señalado en el párrafo anterior con respecto a la utilización de la plataforma por parte de aquellas comunas.

Comuna	Población Total	N° de casos acumulados según 125° I.E	N° de casos confirmados/probables según la muestra	Porcentaje de representatividad
La Florida	402.433	34.984	8.738	24,9 %
La Pintana	189.335	21.972	8.357	38,0 %
La Granja	122.557	14.197	4.713	33,1 %
San Ramón	86.510	9.329	1.608	17,2 %
Puente Alto	645.909	60.516	1.429	2,3 %
San José de Maipo	18.644	1.495	810	54,1 %
Pirque	30.433	2.706	12	0,4 %

Tabla 3.4: Resumen con la proporción de casos entregadas en el 125° Informe Epidemiológico versus lo contemplado en la muestra, para cada comuna de interés.

# Capítulo 4

## El modelo de geocodificación

Para poder realizar la labor de geocodificar a los pacientes confirmados y probables, se decidió utilizar el lenguaje de programación *Python* debido a su polifuncionalidad a la hora de poder trabajar con grandes cantidades de datos, ya que al ser un lenguaje de código abierto se pueden descargar y utilizar distintos *packages* realizados por externos que permiten desarrollar las aplicaciones que el usuario requiera, en este caso el *geocoding*. Para realizar ello, se debe vincular Python con las distintas herramientas que realicen la geocodificación a través de APIs (*Application Programming Interface*) [25], lo cual permite que los servicios y productos se comuniquen entre sí y aprovechen los datos y la funcionalidad de los demás a través de una interfaz documentada.

En el marco de esta memoria, y en conjunto con Matías Martínez Pérez, se procedió a utilizar las APIs Open Street Map y Google, debido a la facilidad con las que se pueden integrar con Python, con el objetivo de poder procesar un conjunto de más de 25.000 direcciones y poder obtener cada una de las latitudes y longitudes de las personas confirmadas o probables, lo cual es esencial a la hora de poder llevar un seguimiento de este tipo de pacientes. Cabe destacar que el rendimiento de ambas APIs difiere en precisión y también en costo, dado a que la API de Google Maps tiene un período de prueba gratuito para posteriormente ser de carácter pago, mientras que la API de Open Street Map es completamente gratuita, sin embargo no es tan certera y veloz como la primera a la hora de procesar y encontrar las direcciones como pares ordenados de latitudes y longitudes. Para visualizar lo anterior, se realizó un experimento aleatorio donde se compara el rendimiento de ambas APIs a la hora de geocodificar 100 direcciones de la muestra de datos, llegando a la conclusión de que la API de Google no solo encuentra un mayor porcentaje de estas a través de todos los métodos de procesamiento de direcciones comparados, sino que también es más certera con la ubicación y a la vez se demora menos en procesar los strings ingresados. Los resultados de este trabajo pueden ser revisados en la memoria pertinente [26], dado a que lo relevante en esta memoria es el siguiente paso, el cual consiste en utilizar las direcciones para situarlas en un mapa.

En resumen, a pesar de que la API de Google Maps sea superior a la de Open Street Maps en todos los ámbitos posibles, el hecho de que esta tenga un límite de búsquedas en su versión gratuita obliga a tener que utilizar la otra API para así llegar a entregar una menor cantidad de direcciones a la API más certera, en concreto las direcciones que sean más difíciles de encontrar. La política de cobro de Google para la API de Geocoding es proporcional al número de solicitudes que esta deberá procesar, y considerando que en este trabajo de título

se busca ubicar el equivalente a 25.667 direcciones correspondientes a pacientes confirmados/probables con residencia en alguna comuna del SSMSO, esto equivaldría a \$130 USD por mes[27], lo cual no convendría pagar considerando que el algoritmo con ambas APIs logra geocodificar todas las direcciones de manera gratuita, así como también considerando que este número en el tiempo iría en aumento, por lo que el precio también se encarecería.

Así, dadas las restricciones de búsqueda por parte de la API más eficiente (Google Maps), para poder encontrar una latitud y una longitud a cada una de las direcciones, se realizó el siguiente procedimiento:

- En primera instancia, se procesa cada dirección a través de la API de OSM, donde el programa espera a que la API encuentre la dirección entregada para poder otorgar la latitud y la longitud correspondiente y así pasar a la siguiente dirección.
- Si es que la API de OSM no fue capaz de encontrar la dirección entregada, se realiza un segundo procesamiento pero esta vez a través de la API de Google Maps. El objetivo de esta dinámica es entregar la menor cantidad posibles de direcciones a esta API, con el fin de que se pueda realizar el procesamiento de manera gratuita.

Al geocodificar estas direcciones, y a pesar de intentar encontrar primero cada una de las direcciones con la API de Open Street Map, se lograron encontrar 9.287 direcciones a través de esta y 16.380 direcciones a través de la API de Google Maps. Por otro lado, es posible concluir que con el algoritmo desarrollado fue posible ubicar a cada una de las personas confirmado/probable de la muestra a lo largo del tiempo.

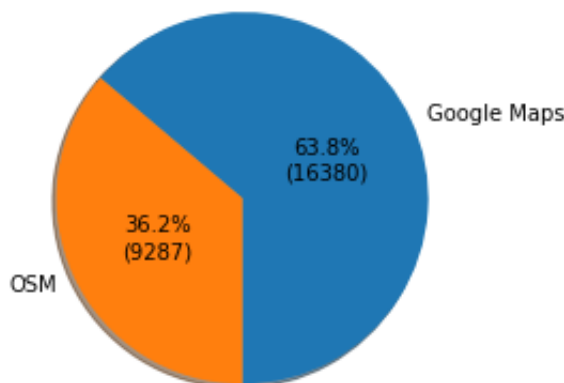


Figura 4.1: Proporción de casos encontrados con cada una de las APIs.

## 4.1. Proceso de normalización de direcciones

A pesar de la diferencia que existe entre APIs, este proceso requiere de un procesamiento previo de las direcciones en la base de datos, debido a que los *strings* de las direcciones declarados por los pacientes en la base de datos cuentan con una serie de caracteres que dificultan el procesamiento de estas en las APIs, lo que a su vez genera que una cantidad

importante de direcciones no puedan ser encontradas, ni por Open Street Map ni por Google. A continuación se detalla un ejemplo de la dirección declarada por uno de los pacientes y como el tratamiento del string ayuda a normalizar esta dirección, con el objetivo de dejar una dirección más simple para poder procesar por las distintas APIs.



Figura 4.2: Ejemplo de dirección normalizada.

De lo anterior, se observa que al *string* inicial se le quitan caracteres y datos innecesarios, tales como el código postal, la ciudad y el país, junto con el caracter # para señalar el número de la vivienda. De esta forma, y posterior al procesamiento previo, todas las direcciones declaradas en la base de datos siguen la estructura “Calle - Número - Comuna”, lo cual facilita su búsqueda. Para mayor profundización en el método de normalización de direcciones, se sugiere revisar la memoria de Matías Martínez.

Lo anterior permite encontrar a la totalidad de las personas confirmadas/probables de la muestra. Sin embargo, se debe tener en consideración que la cantidad de personas encontradas sobrepasan las 25.000, lo cual puede dificultar las labores del personal de trazabilidad y seguimiento del SSMSO si es que se quiere realizar un seguimiento diario y de manera individual.

Dado lo anterior y con el objetivo de facilitar las labores del personal a cargo, se hace necesario realizar una agrupación de la cantidad de personas geocodificadas. Por lo anterior, se utilizaron datos cartográficos del INE para asignar a cada una de las personas geocodificadas a unidades censales más grandes.

## 4.2. Agrupación del territorio

Luego de obtener la latitud y longitud de cada uno de los pacientes geocodificados, se procedió a utilizar datos provenientes del INE para poder agrupar a los pacientes en unidades mayores, lo cual permitiese a su vez poder llevar a cabo un seguimiento más global de la situación del SSMSO.

Los últimos datos registrados a nivel domiciliario datan del año 2017, correspondientes al último Censo realizado en el país, el cual condensa información a distintos niveles, siendo los más relevantes:



- División Política Administrativa (DPA): División del país normada mediante la Constitución Política, la cual resume la información de Chile en 16 regiones, 56 provincias y 346 comunas, siendo las comunas la unidad básica de administración del territorio.
- División Geográfica Censal (DGC): División del territorio comunal en distritos, los cuales pueden ser de carácter urbano, rural o mixto. A su vez en el área urbana se reconocen “zonas censales” compuestas por manzanas, mientras que en el área rural se reconocen “localidades”, las cuales se componen por entidades.

En la siguiente ilustración se puede visualizar las divisiones existentes:

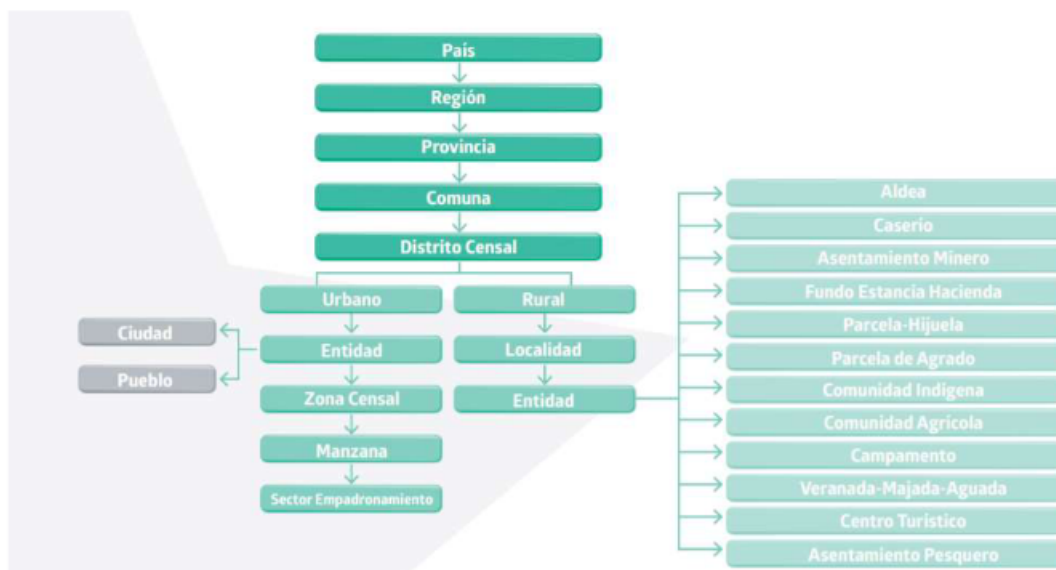


Figura 4.3: Posibles divisiones del territorio. Extraída de la página web del INE.

Debido a las características geográficas de las comunas que componen el SSMSO, se realizó el análisis dentro de las zonas urbanas y rurales, lo que significa que se agrupó a las personas de estas localidades en la zona censal más pequeña dentro de las agrupaciones de territorio, la cual corresponde a la manzana censal en el radio urbano y a la entidad en el radio rural. Una manzana censal es la unidad geográfica básica con fines estadísticos que conforman zonas censales en áreas urbanas, estas contienen grupos de viviendas contiguas o separadas, edificios, establecimientos y/o predios, delimitados por rasgos geográficos, culturales y naturales [28]. Por otro lado, la entidad rural consiste en un asentamiento humano concentrado o disperso que posee 1.000 o menos habitantes, o entre 1.001 o 2.000 habitantes, donde más del 50% de la población que declara haber trabajado se dedica a actividades primarias [29].

A pesar de que estos datos fueron recabados hace aproximadamente 5 años, el INE cuenta con una plataforma de datos geoestadísticos [30], la cual actualiza periódicamente los datos correspondientes a manzanas-entidades del territorio chileno, tanto los datos cartográficos (es decir, que zonas delimitan cierta manzana u entidad) y también datos cuantitativos de la población residente en esas unidades censales. Algunos ejemplos de estos últimos datos son la cantidad de personas, total de hombres y mujeres, total de personas mayores a 65 años, cantidad de hogares, entre otros. La Región Metropolitana cuenta con 51.070 manzanas-entidades

(48.860 manzanas y 2.210 entidades), de las cuales 51.012 coincidieron con respecto a los datos iniciales del Censo 2017, donde solo coincidía 47.752 manzanas-entidades. En la sección de Anexos, se encuentran las figuras que representan los distintos grados de divisiones visualizables en la plataforma de datos geoestadísticos del INE, desde la comuna hasta la manzana censal.

Ya contando con las zonas censales y manzanas-entidades de toda la Región Metropolitana, el problema posterior radica en agrupar a las personas confirmadas/probables geocodificadas a la manzana-entidad correspondiente, de acuerdo a la latitud y longitud encontrada en el proceso.

### 4.3. La asignación de personas a manzanas-entidades

Para llevar a cabo la asignación de todos los pacientes confirmados/probables del SSMSO a su manzana-entidad correspondiente, se pensaron en dos métodos:

1. Asignación por distancia del paciente a la manzana más cercana: Para realizar este método, se debe definir primero el punto de la manzana-entidad desde donde se haría el cálculo de cercanía con respecto al paciente, ya que en ocasiones estos polígonos comprenden varios metros de área. Se decidió utilizar el centroide del polígono como punto de referencia de cada uno de ellos, debido a la facilidad para realizar aquella labor dado a que Python cuenta con una función para ello. Luego de lo anterior, se hace necesario calcular la distancia de cada paciente a cada uno de los centroides de las distintas manzanas-entidades consideradas, las cuales corresponden a las de las comunas de interés del SSMSO, para así poder asignar aquella distancia más pequeña, la cual sería su centroide más cercano. Para ello, se utilizó la fórmula del semiverseno, la cual es útil para calcular la distancia entre 2 puntos definidos por su latitud y longitud sobre una superficie esférica, lo cual es de utilidad considerando una aproximación esférica a la forma de la Tierra. La fórmula del semiverseno se define como:

$$\text{semiversin}\left(\frac{d}{R}\right) = \text{semiversin}(\varphi_1 - \varphi_2) + \cos(\varphi_2) \cos(\varphi_1) \text{semiversin}(\lambda_2 - \lambda_1) \quad (4.1)$$

Con la función semiversin definida por la expresión:

$$\text{semiversin}(\theta) = \text{sen}^2\left(\frac{\theta}{2}\right) \quad (4.2)$$

Donde:

- $d$  es la distancia entre dos puntos (sobre un círculo máximo de la esfera)
- $R$  es el radio de la esfera, que en este caso sería el radio medio de la Tierra
- $\varphi_1$  es la latitud del punto 1 en radianes

- $\varphi_2$  es la latitud del punto 2 en radianes
- $\lambda_1$  es la longitud del punto 1 en radianes
- $\lambda_2$  es la longitud del punto 2 en radianes

Si llamamos “h” al lado derecho de la ecuación 4.1, se puede despejar el valor de la distancia d:

$$d = R \operatorname{semiversin}^{-1}(h) = 2R \operatorname{arcsin}(\sqrt{h}) \quad (4.3)$$

Lo cual es calculable considerando que se cuenta con todos los datos restantes. Así, con la ecuación 4.3 se pueden calcular todas las distancias que se requieren para así poder elegir la distancia más pequeña a cada una de las manzanas-entidades, para cada uno de los pacientes.

2. Asignación por encontrarse al interior del polígono formado por la manzana: En este método se verifica si es que las coordenadas del domicilio del paciente se encuentran dentro o no de alguno de los polígonos de las manzanas existentes. Esto también puede ser logrado como una función de Python, la cual entrega un valor booleano si es que el paciente se encuentra dentro o no de cada uno de los polígonos.

Ambos métodos permiten una asignación del paciente a alguna manzana-entidad y ambos tienen sus ventajas y desventajas. El primero tiene la gran ventaja de asignar a cada uno de los pacientes a una manzana-entidad, aunque esto puede tener consecuencias negativas ya que la manzana-entidad podría quedar en una posición muy lejana con respecto a la posición real del paciente, lo cual se produciría en el caso de que la geocodificación haya fallado y haya sido localizado en una posición errónea, donde ninguna manzana-entidad estuviera realmente cerca. Lo anterior puede solucionarse situando un límite de distancia entre la posición identificada del paciente y su manzana-entidad más cercana, con el fin de que pacientes queden asignados a manzanas-entidades no correspondidas. A la vez, puede ocurrir que el paciente se encuentre en una posición cercana a, por ejemplo, una manzana-entidad muy grande y otra muy pequeña, por lo cual el método seguramente le asignará la manzana-entidad más pequeña debido a que su centroide estará indudablemente más cercano que el de la manzana-entidad grande, a pesar de que el paciente pueda estar inserto en la manzana-entidad más grande. A continuación se puede visualizar este caso particular.

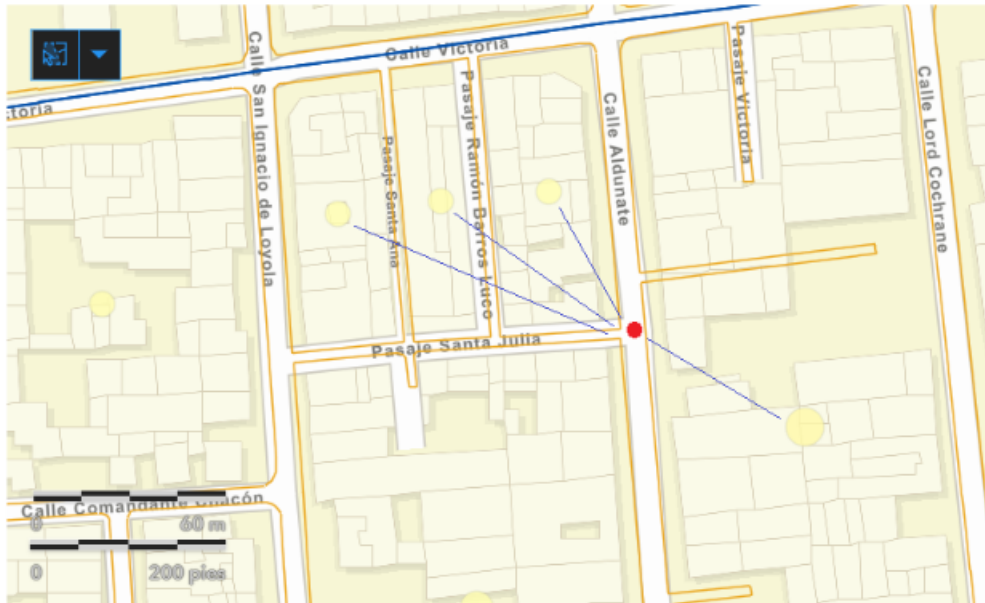


Figura 4.4: Ejemplo de caso desfavorable donde se asigna a una manzana-entidad equivocada. Extraído de la memoria de Matías Martínez P.

El otro método tiene la desventaja de que por construcción de datos las manzanas no incluyen las calles, es decir existe un espacio vacío entre cada manzana, lo cual puede implicar que si un paciente fue localizado muy cercano a la calle o en la calle misma, este quedará fuera de la manzana ya que no estará dentro de sus límites, pese a que en realidad si pertenezca a ella. Esto es un resultado frecuente del proceso de geocodificación, dado a que existen ocasiones en que la aplicación encuentra la calle pero no el número del domicilio particular, por lo que el programa entregará coordenadas relativas a la calle misma. Por lo tanto, se utilizaron ambos métodos aprovechando sus ventajas y desventajas: se partió asignando a los pacientes mediante el segundo método, donde los pacientes que pertenecían a la manzana-entidad correspondiente si fueron correctamente asignados, mientras que aquellos que no haya sido posible asignarlos mediante este método se procesaron con el primero, el cual los asigna a la que tengan más cerca, siempre y cuando esta distancia más cercana no exceda los 300 metros de distancia con el paciente que se busca asignar.

Posterior a la asignación de pacientes confirmados y probables a su manzana-entidad más cercana con los métodos anteriormente señalados, de los 25.667 pacientes que habitan comunas del SSMSO y con riesgo asociado, el proceso de geocodificación señalado logró asignar a 25.373 pacientes (98.85 % del total) a manzanas-entidades cercanas a su latitud y longitud encontrada. De los 294 pacientes no asignados, 276 pacientes no fueron asignados producto de que la distancia más cercana al centroide de una manzana-entidad del SSMSO superaba los 300 metros y 18 pacientes no fue posible geocodificarlos correctamente debido a que no fue posible encontrar o su latitud o su longitud con la dirección que declararon. Es decir, en ambos casos se produjo un error en el proceso producto de que la dirección declarada no fue posible procesarla correctamente, ya sea por errores en esta, nombres o números incompletos o direcciones declaradas que si bien en la base de datos salían registradas como parte de alguna de las 7 comunas del SSMSO, su resultado obtuvo una latitud y longitud de otra comuna ajena.

# Capítulo 5

## El modelo de riesgo

### 5.1. Riesgo clínico del COVID-19

A pesar de que se pueden llegar a presentar síntomas similares a un resfriado común y corriente, en ocasiones el portar el virus puede desencadenar en un cuadro de gravedad que puede incluso causar la muerte. Nadie tiene la suerte escrita con respecto a la enfermedad, pero si existen personas que son mucho más vulnerables que otras. Según la OMS [31], el riesgo clínico de padecer COVID-19 aumenta significativamente si es que la persona tiene más de 60 años o si es que la persona padece afecciones médicas subyacentes, tales como hipertensión arterial, problemas cardíacos o pulmonares, diabetes, obesidad o cáncer. Cabe destacar que en este tipo de personas aumenta la probabilidad de tener un cuadro grave de la enfermedad, sin embargo cualquier persona de cualquier edad puede enfermar de COVID-19 y presentar un cuadro grave o morir.

En línea con lo revisado en los capítulos anteriores, teniendo a cada persona localizada y asignada a su manzana-entidad más cercana se procederá a caracterizar su riesgo clínico, donde será relevante considerar los antecedentes que presente cada paciente como por ejemplo obesidad, disnea, entre otros. Como se explicó anteriormente, a pesar de que existan grupos más vulnerables que otros, cualquier persona puede presentar un cuadro grave de la enfermedad, por lo que las nociones de riesgo que se presentarán en el capítulo posterior se realizarán a través de probabilidades, las cuales representarán la posibilidad de que aquella persona caracterizada presente un cuadro grave de la enfermedad.

### 5.2. El modelo de riesgo utilizado

Luego de encontrar las latitudes y longitudes de cada una de las personas que hicieron su registro en algún establecimiento de salud del SSMSO mediante la geocodificación y posteriormente asignarlas a una unidad censal más grande, el segundo punto necesario para lograr la realización de la herramienta de visualización consiste en caracterizar a cada una de estas personas según sus antecedentes personales y clínicos, para así poder otorgar una noción de riesgo de cuadro de gravedad adecuada a cada uno de los pacientes.

Para la realización de lo anterior, se utilizó el modelo de riesgo desarrollado por uno de los integrantes de la unidad de analítica del Proyecto ANID: COVID 0251, Andrés Garrido,

el cual otorga una probabilidad de desenlace de cuadro grave tomando en cuenta factores relevantes de cada paciente: algunos declarados por cada paciente como lo son sus síntomas, algunos inherentes de cada paciente como lo es su edad y otros relativos al estado del seguimiento que se ha efectuado a partir de los días que han transcurrido desde el primer síntoma. Considerando esas variables como *input*, el modelo busca predecir el estado de egreso del paciente según los egresos declarados en la base de datos suministrada por el SSMSO, los cuales pueden ser “Alta”, “Hospitalizado” o “Fallecido”.

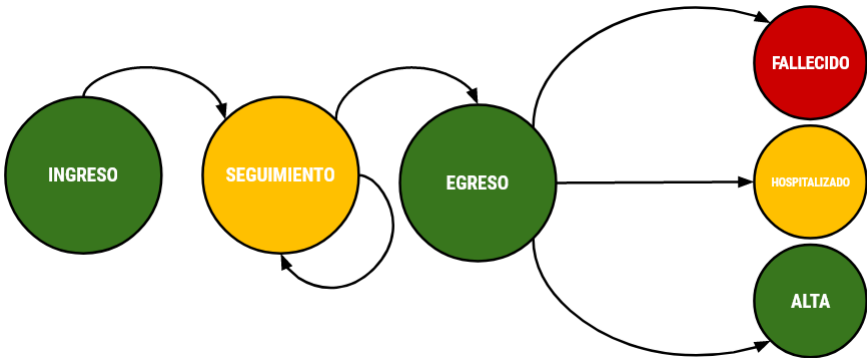


Figura 5.1: Lógica del modelo de riesgo.

La siguiente figura resume lo considerado por el modelo, donde se analizan los antecedentes clínicos declarados junto con la edad y el seguimiento de cada paciente desde el primer síntoma para determinar en que estado de peligro se encuentra cada paciente. Lo anterior se representa mediante un coeficiente que toma valores desde el 0 al 1, siendo el valor más cercano al 1 un paciente con una alta probabilidad de padecer un cuadro grave de COVID-19. La lógica del semáforo ilustra que los pacientes con alto coeficiente deben ser priorizados en la atención, mientras que pacientes con coeficientes más bajos pueden tener un riesgo de desencadenar en un cuadro clínico desfavorable de igual manera, pero en base al aprendizaje del modelo y a pacientes anteriores es más probable que eso no ocurra.



Figura 5.2: Modelo de riesgo desarrollado por Andrés Garrido.

Dado lo anterior, el modelo de riesgo representa el siguiente modelo de regresión logística:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15}$$

En la tabla 5.1 se describen todas las variables utilizadas en el modelo logit.

Variable	Tipo de Variable	Descripción
$p_i$	Numérica entre 0 y 1	Probabilidad de cuadro grave por COVID-19
$X_1$	Binaria	Paciente declara Hipertensión Arterial como antecedente o no
$X_2$	Binaria	Paciente declara Asma como antecedente o no
$X_3$	Binaria	Paciente declara Diabetes como antecedente o no
$X_4$	Binaria	Paciente declara Disnea como antecedente o no
$X_5$	Binaria	Paciente declara Disgeusia/Ageusia como antecedente o no
$X_6$	Binaria	Paciente declara Congestión Nasal como antecedente o no
$X_7$	Binaria	Paciente declara Compromiso de Estado General (CEG) como antecedente o no
$X_8$	Binaria	Paciente declara Decaimiento como antecedente o no
$X_9$	Binaria	Paciente declara Diarrea como antecedente o no
$X_{10}$	Binaria	Paciente declara Otros Antecedentes como antecedente o no
$X_{11}$	Binaria	Paciente tiene menos de 15 años o no
$X_{12}$	Binaria	Paciente tiene entre 45 y 65 años o no
$X_{13}$	Binaria	Paciente tiene más de 65 años o no
$X_{14}$	Binaria	Han transcurrido menos de 5 días desde el primer síntoma o no
$X_{15}$	Binaria	Han transcurrido más de 15 días desde el primer síntoma o no

Tabla 5.1: Resumen de las variables consideradas en el modelo logístico de riesgo implementado por Andrés Garrido.

Donde cabe destacar que se dejaron afuera del modelo las variables *dummies* correspondientes a “Paciente tiene entre 15 y 45 años” y “Han transcurrido entre 5 y 15 días desde el primer síntoma” para darle consistencia al modelo. En la sección de Anexos, tablas 7.1 y 7.2, se encuentra un resumen de los resultados de la regresión logística realizada por Andrés, así como también los Odd Ratios con intervalo de confianza del 95% para cada una de las variables. Como conclusiones más relevantes del modelo, se observa que las variables que más aumentan la probabilidad de un posible cuadro de gravedad son  $X_{13}$  (“Paciente tiene más de 65 años”),  $X_7$  (“Paciente declara Compromiso de Estado General (CEG) como antecedente”),  $X_4$  (“Paciente declara Disnea como antecedente”) y  $X_2$  (“Paciente declara Asma como antecedente”). Lo anterior es consistente con lo señalado por la OMS, y cobra sentido que pacientes con dificultades respiratorias sean más propensos a tener más dificultades para sortear de buena manera la enfermedad.

Por otro lado, las variables que más disminuyen la probabilidad de cuadro grave son  $X_5$  (“Paciente declara Disgeusia/Ageusia como antecedente”) y  $X_6$  (“Paciente declara Congestión Nasal como antecedente”), probablemente porque se trate de síntomas comunes para cualquier caso COVID-19, tanto cuadro leve como grave y por consiguiente, tanto para casos dados de alta como casos hospitalizados y fallecidos.

Con respecto a los Odd Ratios, se observa una alta asociación entre la probabilidad de cuadro grave y las variables *ASMA*, *DISNEA*, *COMP\_ESTADO\_GRAL* y *EDAD\_100*, donde

personas con estos antecedentes son 3,17; 3,26; 5,097 y 6,693 veces más propensas a contraer un cuadro grave de COVID-19 en relación a quienes no presenten estos antecedentes, respectivamente.

Con lo anterior concluye el proceso de caracterización del paciente, es decir se tiene la latitud, longitud y probabilidad de riesgo para cada uno de los casos confirmados y probables que habitan en el SSMSO, por lo que se cuenta con el *input* necesario para poder realizar la herramienta de visualización.



# Capítulo 6

## La herramienta de visualización

Con la información anteriormente recopilada, se tienen los antecedentes necesarios para comenzar a trabajar en el cumplimiento del objetivo general de esta memoria. Para ello se comenzará revisando algunas aplicaciones similares desarrolladas en otros lugares del mundo, con el fin de entender su usabilidad y a la vez su utilidad en el marco de la pandemia.

Posterior a ello, y en línea con la segunda pregunta de investigación declarada, se procederá a explicar el motivo de una nueva agregación de los datos, pasando de pacientes agregados en manzanas-entidades a pacientes agregados en “macrozonas”, las cuales agrupan a más de una manzana-entidad. A la vez, se detallará la información que contiene cada una de estas zonas geográficas delimitadas.

Luego, se explicará el cómo fue implementada la herramienta de visualización. Para ello, se comenzará por revisar los paquetes de Python que permitieron la realización de esta como lo son *folium*, *geopandas* y *streamlit*. A la vez, se revisarán los datos visualizables por la persona encargada de realizar la trazabilidad a este análisis por macrozonas, como lo es el índice de riesgo que presentan los pacientes que habitan esa macrozona, la cantidad de adultos mayores que habitan en esta macrozona, entre otros.

Finalmente, se comentará acerca de los resultados obtenidos al entregar la herramienta a la Unidad de Salud Digital (USD) del SSMSO, así como también se comentará acerca de los alcances y limitaciones que esta herramienta presenta, concluyendo con una propuesta de trabajo futuro que llegue a robustecer la herramienta entregada.

### 6.1. Nueva agregación de datos: las macrozonas

#### 6.1.1. Partición

Según los datos obtenidos del INE, las comunas que componen el SSMSO cuenta con aproximadamente 10.903 manzanas-entidades, las cuales se procedieron a agregar en macrozonas que delimitan el espacio geográfico del SSMSO, las cuales consisten en cuadrantes urbanos equivalentes de aproximadamente 2x2 kilómetros y que alcanzan a cubrir todas las comunas salvo San José de Maipo, que por su lejanía y distinta distribución de viviendas con respecto a las otras comunas se decidió tratar como una zona aparte, delimitada por un triángulo

que alberga todas las distintas zonas desde donde se tiene registro de un posible contagio. Lo anterior tiene como objetivo facilitar la gestión y usabilidad de la plataforma, para así ayudar a llevar un registro más general.

Para realizar lo anterior, se calcularon las distancias (ver ecuación 4.3) entre los puntos de mayor y menor latitud y longitud resultantes del proceso de geocodificación de pacientes, es decir los puntos más al norte, sur, este y oeste desde donde algún paciente declaró su domicilio para las comunas que no fueran San José de Maipo, lo cual terminó generando un rectángulo de aproximadamente 14.13 x 10.61 km de largo y ancho. Por lo mismo, se procedió a dividir la diferencia de latitud y longitud en 7 y 5 sectores respectivamente, generando un total de 35 cuadrantes urbanos que delimitan todas las latitudes y longitudes resultantes del proceso de geocodificación de cada paciente confirmado y probable. Del mismo modo, para el caso de San José de Maipo se calculó la distancia haversiana entre tres de los cuatro puntos resultantes de combinar los pares de latitud y longitud máxima y mínima, dejando de lado el punto de máxima latitud y longitud, ya que según la geografía del lugar, corresponde a un sector montañoso donde no se encuentra población que habite el sector. Lo anterior genera que, a diferencia de los 35 cuadrantes idénticos, el polígono que delimite la comuna sea un triángulo. Se utilizó el paquete *mplleaflet*, el cual permite graficar y dibujar polígonos en un mapa realizado por Leaflet, una biblioteca de JavaScript de código abierto que se utiliza para crear aplicaciones de mapas web.

Similar a la lógica del Visor Territorial COVID-19 revisado en la sección anterior, ahora las manzanas-entidades de las comunas se encuentran agrupadas en estas “macrozonas”, las cuales condensarán la analítica de los pacientes que habiten en cada una de las manzanas-entidades contenidas en cada una de estas macrozonas. La figura 6.1 ilustra las delimitaciones de las 36 macrozonas resultantes (35 cuadrantes urbanos y el triángulo que delimita la comuna de San José de Maipo).

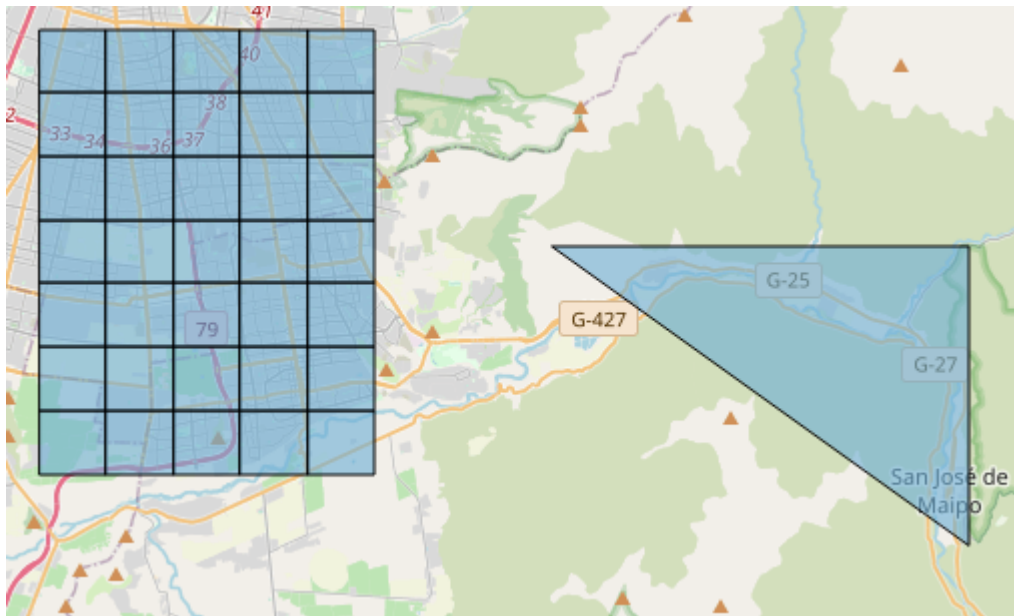


Figura 6.1: Macrozonas que delimitan la geografía del SSMSO. Elaboración propia.

## 6.1.2. Datos incorporados

Combinando los datos relacionados de cada paciente con los datos relacionados a cada manzana-entidad extraídos del INE, se obtiene una base de datos con la información de 25.373 pacientes geocodificados a manzanas-entidades, con las siguientes variables de interés representadas en la tabla 6.1:

Variable	Descripción	Fuente de la información
<i>IDPER</i>	Identificador único de la persona	Data de Paciente
<i>Edad</i>	Edad del paciente	Data de Paciente
<i>FechaInicioCuarentena</i>	Fecha en la que el paciente inició su cuarentena	Data de Paciente
<i>Dirección</i>	Dirección registrada por el paciente	Data de Paciente
<i>LatitudPaciente</i>	Latitud de la dirección del paciente	Modelo de geocodificación
<i>LongitudPaciente</i>	Longitud de la dirección del paciente	Modelo de geocodificación
<i>Pred_Proba_eg</i>	Probabilidad de riesgo predicha de desenlace de cuadro grave por COVID-19 del paciente	Modelo de riesgo
<i>MANZENT</i>	Identificador único de la manzana-entidad a la cual pertenece el paciente	Datos geoestadísticos INE
<i>Comuna</i>	Comuna de la manzana-entidad	Datos geoestadísticos INE
<i>total_viviendas</i>	Total de viviendas registradas en la manzana-entidad	Datos geoestadísticos INE
<i>total_personas</i>	Total de personas que habitan la manzana-entidad	Datos geoestadísticos INE
<i>edad_65ym</i>	Cantidad de adultos mayores que habitan en la manzana-entidad	Datos geoestadísticos INE
<i>LatitudCentroide</i>	Latitud del centroide de la manzana-entidad	Datos geoestadísticos INE
<i>LongitudCentroide</i>	Longitud del centroide de la manzana-entidad	Datos geoestadísticos INE

Tabla 6.1: Variables de interés para la conformación de la plataforma de visualización.

Posteriormente, se agrupó la información de estas 25.373 personas a nivel de *MANZENT*, es decir se sintetizó toda la información de pacientes a nivel de manzanas-entidades. Como resultado se obtiene que la totalidad de pacientes se agruparon en 4.719 de las 10.903 manzanas-entidades extraídas de los Datos geoestadísticos del INE. La tabla 6.2 sintetiza las variables que contienen las 4.719 manzanas-entidades representadas.

Variable	Descripción
<i>MANZENT</i>	Identificador único de la manzana-entidad
<i>prom_edad</i>	Promedio de edad de los pacientes que habitan esa manzana-entidad
<i>riesgo_manzana</i>	Promedio del riesgo predicho de los pacientes que habitan esa manzana-entidad
<i>num_personas_manzanas</i>	Número de personas que habitan esa manzana-entidad, según el INE
<i>num_casos</i>	Número de casos confirmados/probables que habitan esa manzana-entidad
<i>Comuna</i>	Comuna de la manzana-entidad, según el INE
<i>total_viviendas</i>	Total de viviendas registradas en la manzana-entidad, según el INE
<i>edad_65ym</i>	Cantidad de adultos mayores que habitan en la manzana-entidad, según el INE
<i>LatitudCentroide</i>	Latitud del centroide de la manzana-entidad
<i>LongitudCentroide</i>	Longitud del centroide de la manzana-entidad

Tabla 6.2: Variables incorporadas a nivel manzana-entidad

Finalmente, agrupando a todas las manzanas-entidades en las macrozonas previamente explicadas, resultan un total de 36 tablas (1 por macrozona), las cuales se componen a partir de toda la información anterior de cada manzana-entidad que se encuentre inserta en el cuadrante urbano (o triángulo) que la delimita. Con la información anterior ya sintetizada, se procederá a implementar la herramienta de visualización, incorporando un mapa de calor con el riesgo de las macrozonas delimitadas y realizando la interfaz de visualización de información importante para cada macrozona particular a través de Python.

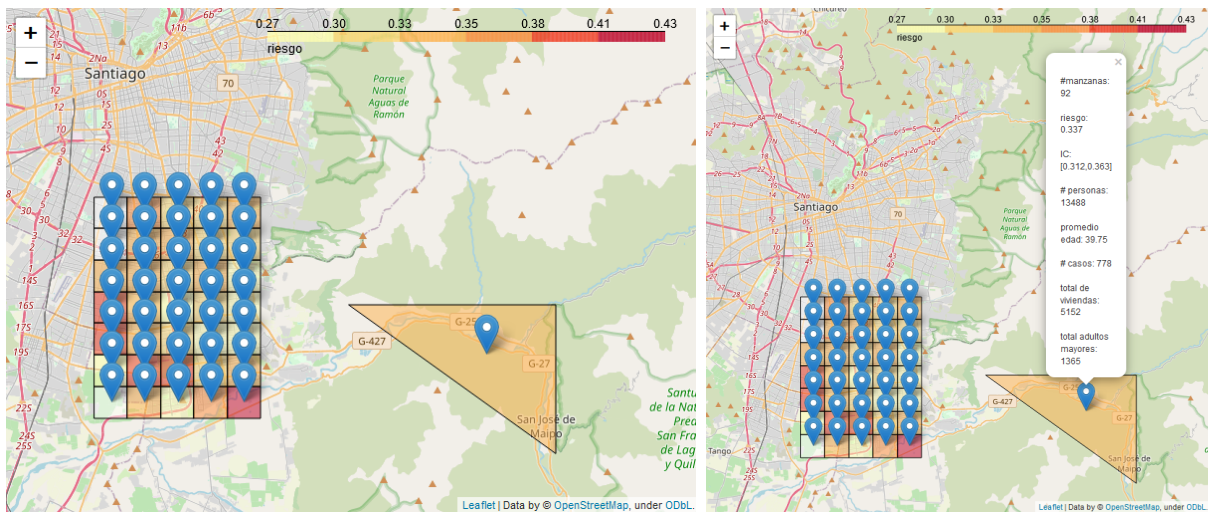
## 6.2. La implementación de la herramienta de visualización

### 6.2.1. El mapa

Ya con el nivel de agregación deseado para cada uno de las macrozonas, se procedió a utilizar dos librerías necesarias para tratar datos geográficos y poder representarlos en mapas: *geopandas* (versión 0.9.0) y *folium* (versión 0.11.0). Por un lado, el propósito del primero es simplificar el trabajo con datos geoespaciales, permitiendo realizar operaciones geoespaciales en el paquete *pandas* a través de *Geodataframes* y *Geoseries*, permitiendo almacenar información por medio de polígonos definidos por pares ordenados de latitud y longitud. Por otro lado, el segundo de ellos permite interactuar con los datos geoespaciales y así plasmarlos en mapas interactivos de la biblioteca Leaflet, los cuales son integrables a sitios web.

De esta manera, mediante el paquete *geopandas* se almacenó toda la información relevante

a los polígonos de las macrozonas urbanas delimitadas para así poderlos dibujar en un mapa realizado mediante Leaflet, incorporando marcadores que permitan desplegar texto que resume el número de manzanas-entidades con pacientes confirmados/probables, el riesgo promedio de los pacientes que habitan en esa macrozona, el número de casos confirmados/probables que habitan la macrozona, y datos censales correspondientes al número de personas, total de viviendas y total de adultos mayores que habitan en toda esa macrozona, es decir la suma de todas las personas, viviendas y adultos mayores de cada una de las manzanas-entidades que componen una macrozona. La lógica de incorporar el número de adultos mayores consiste en otorgar una posible opción de priorización a las labores de trazabilidad, dada su vulnerabilidad con respecto al contagio con el virus, basado en el modelo desarrollado y en la literatura previamente revisada.



(a) Mapa de calor del riesgo de cada macrozona. Elaboración propia.

(b) Datos desplegados por cada una de las macrozonas.

Figura 6.2: Mapa desarrollado para el SSMSO.

Se puede observar que mientras más roja se encuentre la macrozona, mayor es el riesgo que se presenta y a medida que se va esclareciendo la tonalidad a colores más amarillos, el riesgo de esa macrozona disminuye. Lo anterior resume la probabilidad de cuadro grave de los pacientes geocodificados en manzanas-entidades que se encuentren contenidas en la macrozona correspondiente.

### 6.2.2. La interfaz

Luego de contar con los mapas ya implementados, el paso final consiste en realizar una interfaz donde se puedan ejecutar los algoritmos previamente programados, para que el usuario objetivo que utilice la plataforma pueda visualizar los datos. Para ello, se utilizó el paquete *streamlit*, el cual permite realizar aplicaciones web compartibles a partir de scripts programados en Python. De esta forma, es posible realizar aplicaciones de diversa índole, entre ellas mapas geográficos, dashboards, entre otras.

Para poder desplegar el mapa previamente elaborado, primero se procede a transformar la base de datos con la información de todas las manzanas-entidades en formato *pickle*, los cuales facilitan y hacen más rápida la lectura de archivos, con el fin de optimizar la velocidad del algoritmo. Ya con el archivo *pickle*, se procede a realizar la agrupación en macrozonas y posteriormente sintetizar la información en cuadros de textos que se despliegan al presionar los marcadores localizados en el centro de cada una de las macrozonas. La figura 6.3 permite visualizar el prototipo de herramienta de visualización elaborada, junto con la información que esta despliega.

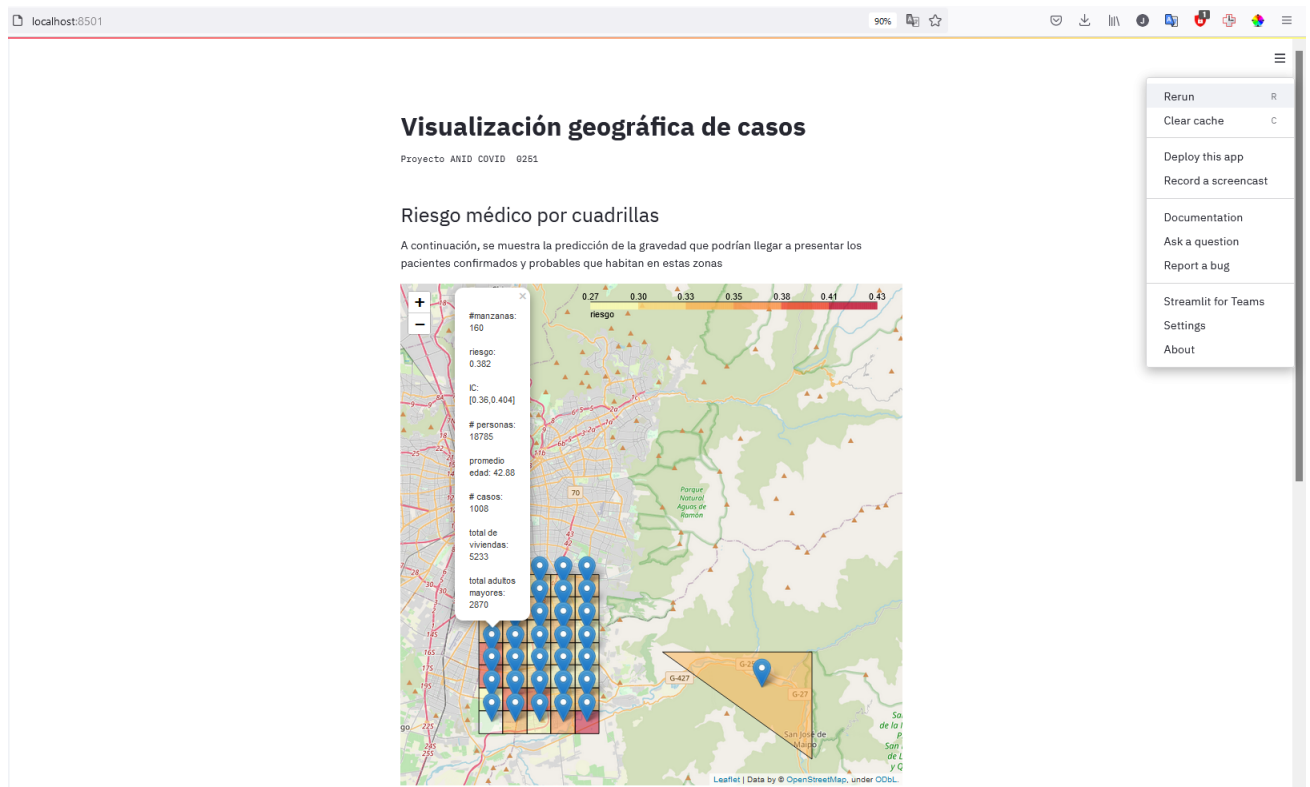


Figura 6.3: Prototipo de Herramienta de Visualización del riesgo de desenlace de cuadro grave por COVID-19.

Finalmente, y tal como se mencionó en los alcances del trabajo de título, no se considera contemplar una conexión directa a sistemas con información actualizada, por lo que solo es posible visualizar la información de casos entregada en el período en el que se considera el análisis. Con esto en consideración, se programó el despliegue automático del mapa a través de la aplicación, la cual requiere de los siguientes pasos para poder ser desplegada:

1. **Generar el input del modelo de riesgo:** Para ello, se debe filtrar la base de datos única que condensa la información del modelo de geocodificación y el modelo de riesgo (ver Tabla 6.1) por *FechaInicioCuarentena*, es decir, se debe seleccionar el rango de fechas del período en el cual se quiera desplegar la información. Lo anterior puede ser realizado a través de Microsoft Excel. Posterior a ello se debe agrupar la información en manzanas-entidades, lo cual se realiza automáticamente con un script entregado, el

cual trata la base de datos filtrada previamente y guarda una planilla Excel con la información a nivel de manzana-entidad.

2. **Convertir la base de datos de manzana-entidad en archivo *pickle*:** Lo anterior se realiza con otro script automático que toma la base de datos filtrada y la convierte en un archivo .pkl, el cual se guarda en la misma ubicación que el script que contiene la aplicación web.
3. **Ejecutar la herramienta por *streamlit*:** Con la base de datos anterior ya transformada, se debe ejecutar la aplicación web mediante un terminal que reconozca los paquetes *geopandas*, *folium* y *streamlit*, lo cual generará el despliegue del mapa y la información recabada en ese período de tiempo por el SSMSO.

# Capítulo 7

## Conclusiones

### 7.1. Resultados obtenidos

El resultado más relevante es el cumplimiento del objetivo general: el desarrollo del prototipo de herramienta de visualización de casos confirmados y/o probables por COVID-19 con los datos suministrados por el SSMSO. A diferencia de las herramientas previamente revisadas en la primera sección de este capítulo, este mapa cuenta con la gran diferencia de otorgar no solo información a un nivel de cuadrantes urbanos para las comunas del SSMSO, sino que también información censal relevante como la cantidad total de personas que habitan esa macrozona, incluyendo a los adultos mayores, el promedio de edad y el número de viviendas que existen en las macrozonas. Esto permite complementar el seguimiento por riesgo, añadiéndole una dimensión social que pesa a la hora de la toma de decisiones: en base a la evidencia revisada, mientras mayor se hace la persona, más probable es que padezca un cuadro grave de COVID-19. Lo anterior responde a las dos preguntas de investigación manifestadas en la sección 1.4:

1. ¿Es posible generar una herramienta de apoyo a la toma de diversas decisiones importantes en el manejo de la pandemia?
2. ¿Es posible agregar la información de cada una de las personas contagiadas o probables por COVID-19?

Con respecto a la plataforma en sí, se pueden llegar a visualizar zonas donde no existe color, debido a que en las fechas buscadas ningún paciente identificado como confirmado o probable por COVID-19 habitaba en alguna de las manzanas-entidades de esa macrozona particular. Por otro lado, evidentemente existen zonas mucho más concentradas, donde la población es incluso mayor debido al radio urbano delimitado por manzanas-entidades más pobladas. Esto se debe a que solamente una porción de las manzanas-entidades totales de las comunas del SSMSO fueron utilizadas, dejando a muchas otras de lado producto de que ningún paciente habitaba en una de ellas según los registros.

Finalmente, la herramienta fue puesta a disposición a la contraparte, sin embargo como se trataba de un prototipo de plataforma de seguimiento no se tiene registro acerca de su utilización por el mismo SSMSO, lo cual fue señalado como alcance del trabajo de título.



## 7.2. Discusión

Con respecto a la elaboración del mapa, se concluye que Python ofrece buenas alternativas para tratar datos geoestadísticos, facilitando la implementación de herramientas de visualización tales como mapas de calor. Es necesario destacar que el rendimiento y precisión del mapa generado depende mucho de las personas naturales, ya que es probable que se cometan errores involuntarios que pueden entorpecer el proceso de geocodificación y caracterización de riesgo. Ejemplo de lo anterior podrían ser pacientes que den direcciones equivocadas o que la escriban de forma defectuosa, así como también de pacientes que no manifiesten todos los antecedentes o síntomas que realmente tienen.

El código implementado logra realizar el mapa y mostrarlo automáticamente en la aplicación web desarrollada en *streamlit*, siempre y cuando se le suministren los datos correspondientes a las manzanas-entidades (revisados en la sección 6.2.2). El problema radica en que, así como existen algunas variables que no cambiarán conforme pase el mediano plazo, como lo son todas las variables extraídas de las bases geoestadísticas del INE, también existen datos móviles que se van actualizando conforme va avanzando el tiempo como lo son el número de casos, riesgo de manzana y promedio de edad de los pacientes. Como el SSMSO otorgó bases de datos con la lista de ingresos acumulada en distintas épocas de tiempo, constantemente se hacía necesario correr el algoritmo de geocodificación y de riesgo para nuevos pacientes encontrados, razón por la cual en la sección de alcances se había mencionado previamente la no consideración de actualizaciones de la plataforma, llegando a trabajar con una de las últimas versiones entregadas. En síntesis, es posible correr los mapas y la aplicación web implementada, pero previamente se hace necesario correr los algoritmos para generar la información que el prototipo utilizará para visualizar, lo cual requiere de intervención humana.

Es precisamente por el hecho de que la plataforma depende mucho de las personas para su correcta ejecución, que esta no es completamente automatizada en esta etapa de prototipaje. Sería relevante para la implementación del modelo que el proceso de ingreso de pacientes a establecimientos de salud fuera mucho más autónomo y centralizado, ya que el mapa refleja paulatinamente de mejor manera la situación actual conforme se cuente con más información de casos activos y/o probables, y del mismo modo, los modelos de geocodificación y de riesgo entregarían resultados más robustos si se minimizan los potenciales errores señalados al inicio de esta sección.

## 7.3. Trabajo futuro

### 7.3.1. Robustecimiento del modelo de riesgo

Hoy en día, la realidad que se vive con respecto al virus dista de lo ocurrido en el período de tiempo del cual se tienen datos para el análisis de esta investigación. Si bien al inicio de la pandemia preocupaba la letalidad del virus, con el paso del tiempo y la avanzada campaña de inoculación realizada en Chile y algunas partes del mundo se ha logrado volver progresivamente a una normalidad moderada, producto de que con las vacunas ha sido más fácil poder controlar la disponibilidad de camas críticas, evitando el colapso de los sistemas de salud.

Sería relevante considerar este proceso en una próxima estimación del riesgo de desenlace de cuadro grave, debido a la cantidad de gente vacunada a fecha de hoy con alguna dosis del COVID-19 y a la evidencia existente [32] respecto a la efectividad de las nuevas dosis para mantener un mejor cuidado de la población.

A la vez, otra manera para complementar aún más el pronóstico generado por el modelo sería utilizar otro tipo de variables. Ejemplo de lo anterior puede ser utilizar variables demográficas, socioeconómicas, educacionales, entre otras.

### **7.3.2. Mejoras en la aplicación**

Existen oportunidades de mejora para una segunda versión del prototipo desarrollado, entre las que se encuentran una mejor agregación poligonal para la comuna de San José de Maipo y el desarrollo de nuevas métricas analíticas como las implementadas en las herramientas de visualización desarrolladas estudiadas en la sección 6.1. A la vez, sería relevante que la aplicación visualizara en segunda instancia no solo la macrozona delimitada, sino también a las manzanas-entidades que conforman esa macrozona en forma de polígonos y desagregar la información a ese nivel, lo cual ayudaría a llevar tanto un registro agregado, de fácil interpretación, y un registro específico con más detalles individuales.

A la vez, la plataforma otorgaría un mayor valor si permitiera integrarse con otros sistemas digitales de salud públicos o privados. De esta forma se podrían obtener no solo información relacionada a los pacientes por COVID-19, sino que podría implementarse para visualizar el estado de incluso otras enfermedades o propagaciones. Idealmente, se podría seguir comparando las distintas APIs de geocodificación, ya que en la realización de esta memoria solo se estudiaron dos alternativas.

Finalmente, el automatizar el proceso de geocodificación y de asignación de riesgo mediante código e interacción de bases de datos otorgaría mayor simplicidad y beneficiaría la usabilidad de la herramienta de visualización.

# Anexos

## Criterio Clínico y Epidemiológico del COVID-19

Según la definición de casos del COVID-19 realizada por la OMS [13] revisada en el marco teórico, se especifica que:

1. Para que una persona cumpla con el criterio clínico del COVID-19 debe:
  - Tener síntomas de fiebre y tos; o
  - Tener indicios de, por lo menos, tres o más de los siguientes síntomas: Fiebre, tos, debilidad/fatiga general, dolor de cabeza, mialgia, dolor de garganta, coriza, disnea, anorexia/náuseas/vómitos, diarrea o estado mental alterado.
2. Para que una persona cumpla con el criterio epidemiológico del COVID-19 debe:
  - Residir o trabajar en una zona con alto riesgo de transmisión de virus, como lo son espacios cerrados en entornos residenciales, entornos humanitarios como campamentos y entornos similares a campamentos para personas desplazadas; en cualquier momento dentro de los 14 días anteriores al inicio de los síntomas; o
  - Residir o viajar a un área con transmisión comunitaria en cualquier momento dentro de los 14 días antes del inicio de los síntomas; o
  - Trabajar en cualquier entorno de atención de la salud, incluso dentro de los establecimientos de salud o dentro de la comunidad en cualquier momento dentro de los 14 días anteriores al inicio de los síntomas.

## Resultados y Odd Ratios de la regresión logística

Variable	Coefficiente	Std. Error	P> z
<i>Constante</i>	-1.4971	0.017	0.000
<i>HTA</i>	0.031	0.022	0.162
<i>ASMA</i>	1.1557	0.036	0.000
<i>DIABETES</i>	0.5576	0.024	0.000
<i>DISNEA</i>	1.1818	0.031	0.000
<i>DISGEUSIA_AGEUSIA</i>	-1.4033	0.052	0.000
<i>CONGESTION_NASAL</i>	-1.1842	0.057	0.000
<i>COMP_ESTADO_GRAL</i>	1.6286	0.069	0.000
<i>DECAIMIENTO</i>	0.4365	0.038	0.000
<i>DIARREA</i>	-0.3516	0.058	0.000
<i>OTROS_ANT</i>	0.5681	0.018	0.000
<i>EDAD_15</i>	-0.8735	0.043	0.000
<i>EDAD_65</i>	0.8544	0.021	0.000
<i>EDAD_100</i>	1.9010	0.024	0.000
<i>DS_5</i>	0.3532	0.02	0.000
<i>DS_15</i>	0.5226	0.044	0.000

Tabla 7.1: Resultados de la regresión logística implementada por Andrés Garrido.

Variable	5 %	95 %	Odd Ratio
<i>Constante</i>	0.216	0.232	0.224
<i>HTA</i>	0.988	1.077	1.032
<i>ASMA</i>	2.961	3.407	3.176
<i>DIABETES</i>	1.667	1.830	1.747
<i>DISNEA</i>	3.069	3.463	3.260
<i>DISGEUSIA_AGEUSIA</i>	0.222	0.272	0.246
<i>CONGESTION_NASAL</i>	0.274	0.342	0.306
<i>COMP_ESTADO_GRAL</i>	4.448	5.839	5.097
<i>DECAIMIENTO</i>	1.435	1.668	1.547
<i>DIARREA</i>	0.628	0.788	0.704
<i>OTROS_ANT</i>	1.705	1.827	1.765
<i>EDAD_15</i>	0.384	0.454	0.417
<i>EDAD_65</i>	2.255	2.449	2.350
<i>EDAD_100</i>	6.381	7.020	6.693
<i>DS_5</i>	1.368	1.481	1.424
<i>DS_15</i>	1.547	1.838	1.686

Tabla 7.2: Odd Ratios de las variables del modelo.

## Zonas Censales de la plataforma de Datos Geoestadísticos del INE

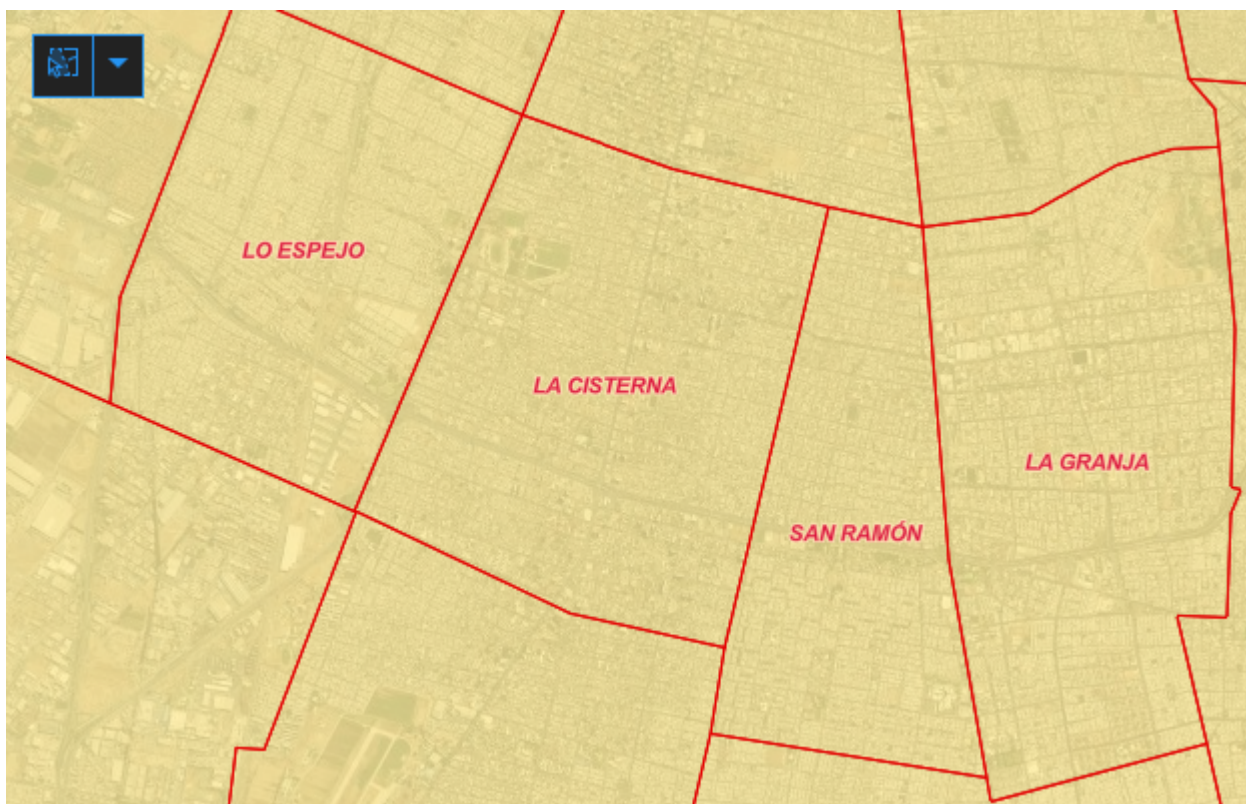


Figura 7.1: Divisiones comunales de Lo Espejo, La Cisterna, San Ramón y La Granja.



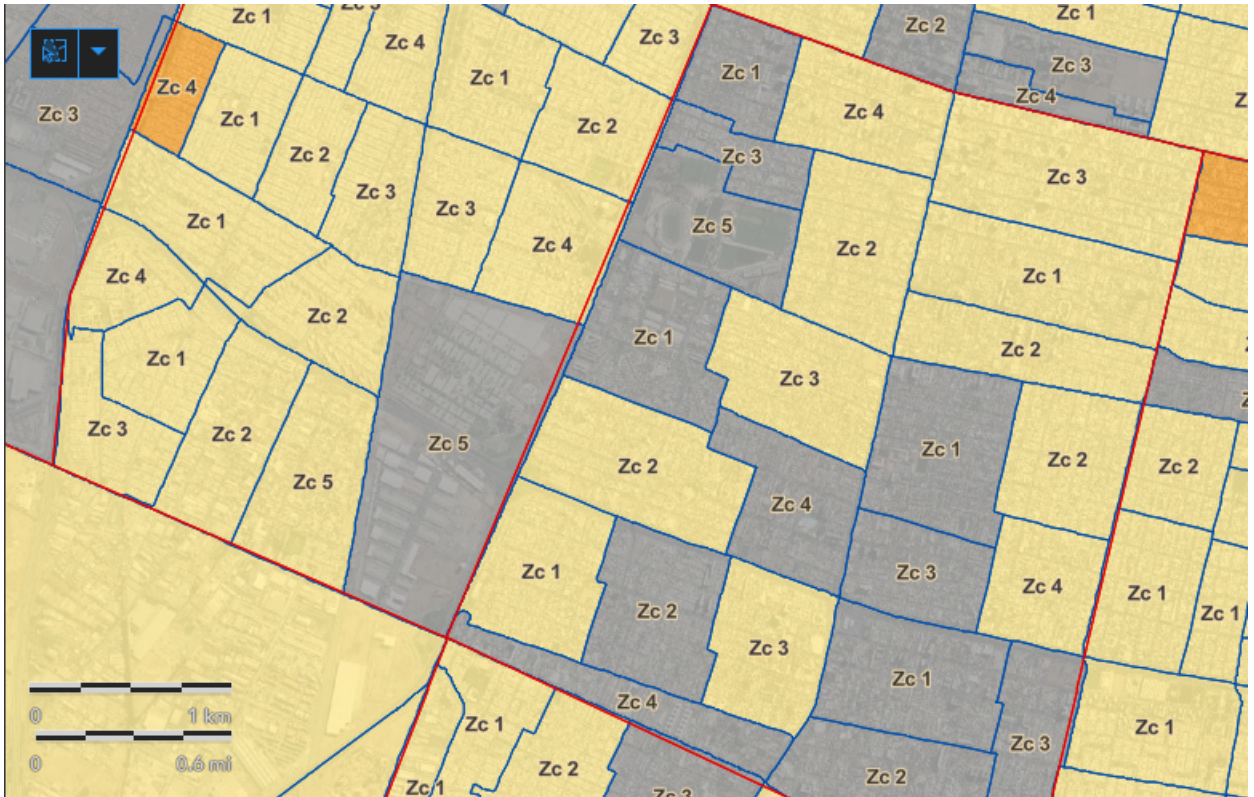


Figura 7.2: Zonas Censales de Lo Espejo y La Cisterna.

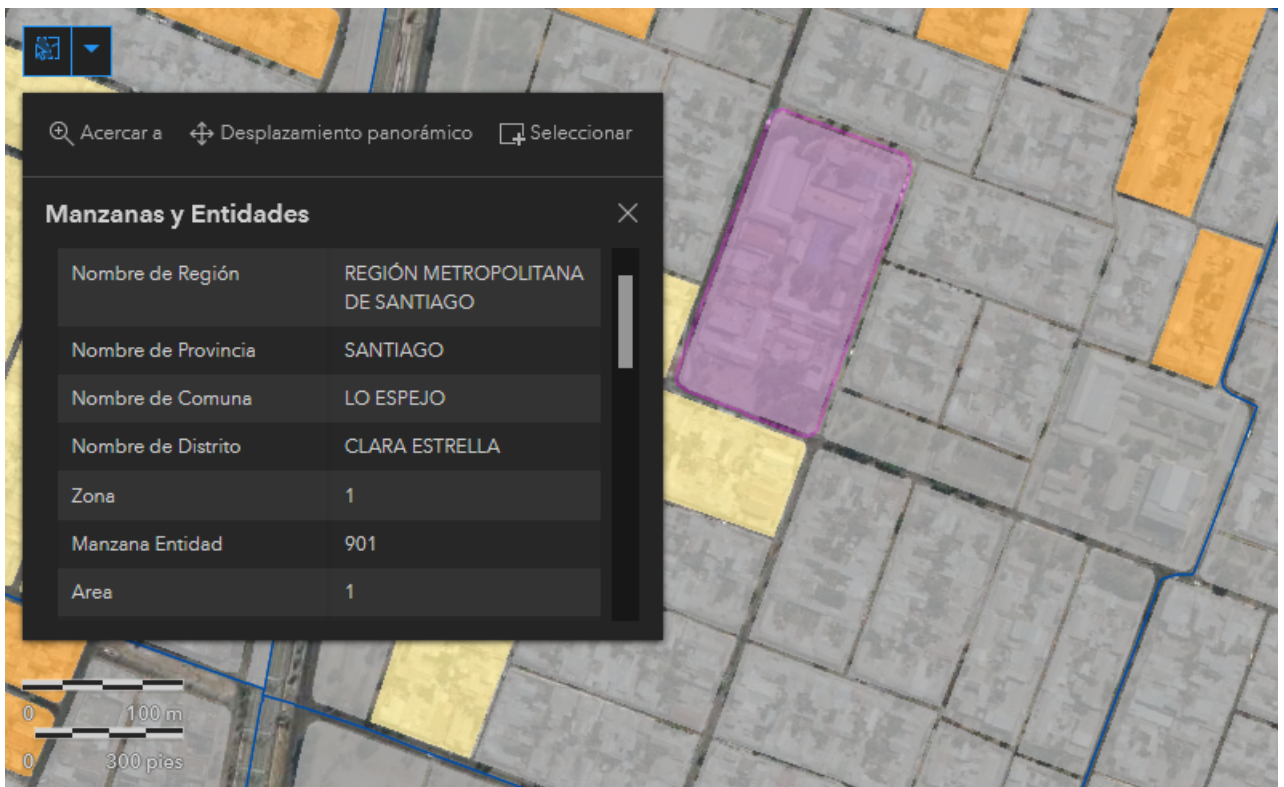


Figura 7.3: Manzana-Entidad de la comuna de Lo Espejo.

# Herramientas de visualización desarrolladas

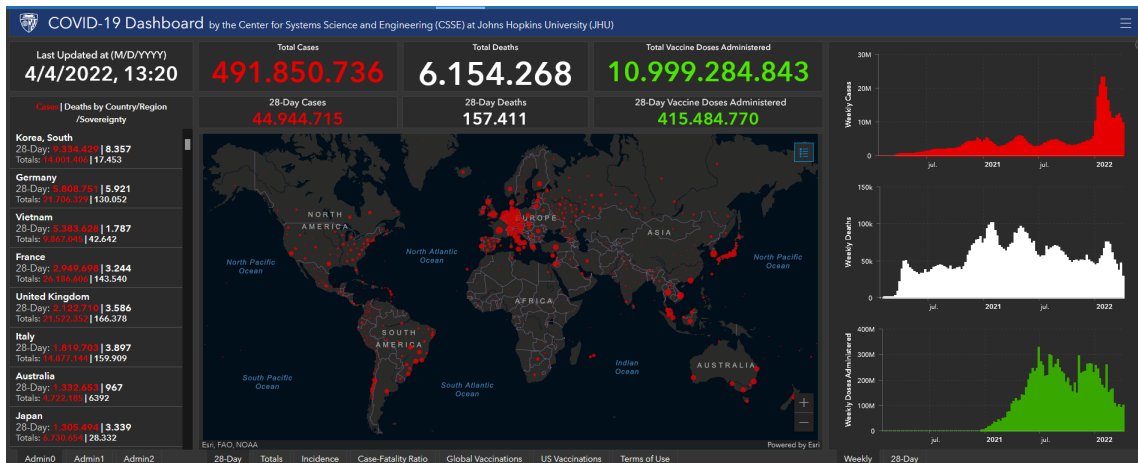


Figura 7.4: Plataforma de visualización de casos de la Universidad de John Hopkins.

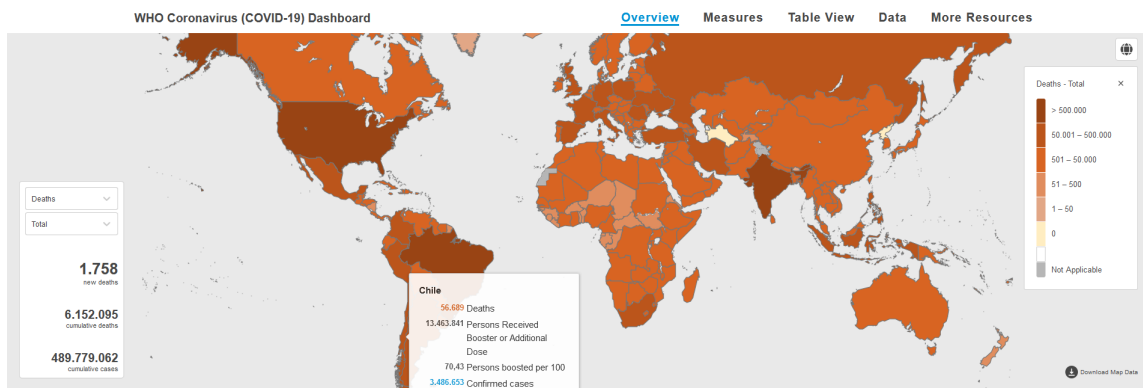


Figura 7.5: Dashboards realizados por la OMS.

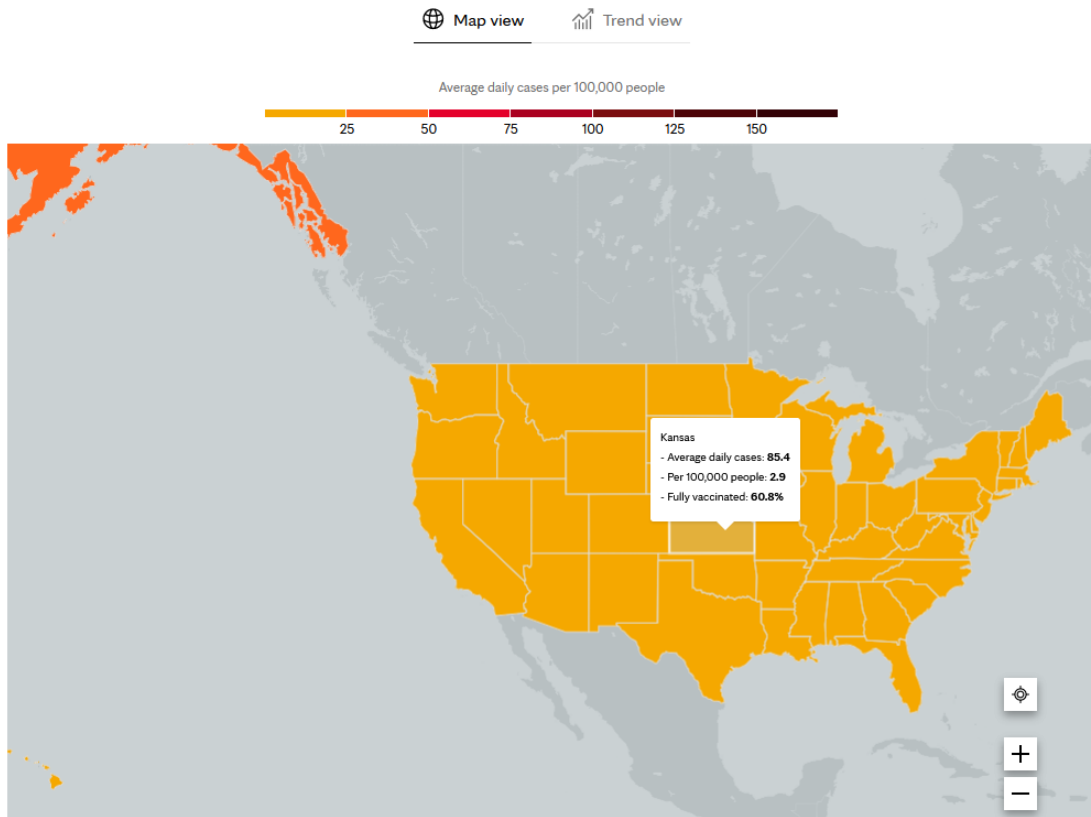


Figura 7.6: Dashboards realizados por Mayo Clinic.

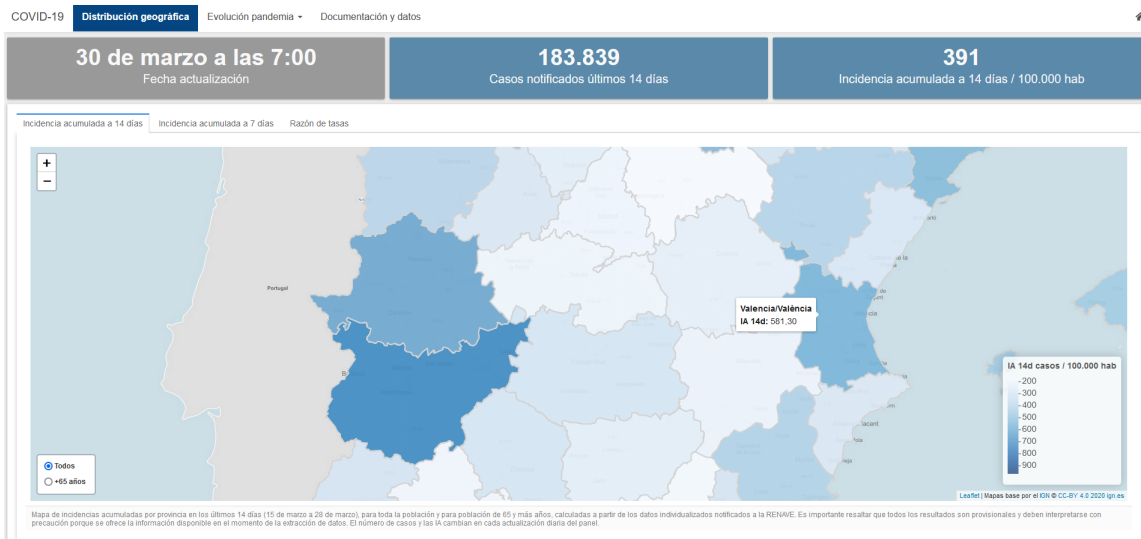


Figura 7.7: Plataforma de visualización de casos del Centro Nacional de Epidemiología (CNE) de España.



## Mapa de casos activos por comunas

Seleccione una región para obtener un mapa de calor de los casos activos de Covid-19 en cada una de sus comunas, donde la intensidad del color indica la cantidad o tasa de casos activos.

Toque o pase el cursor sobre las comunas para ver sus nombres y cantidad de casos activos

Elija una región para graficar el mapa:

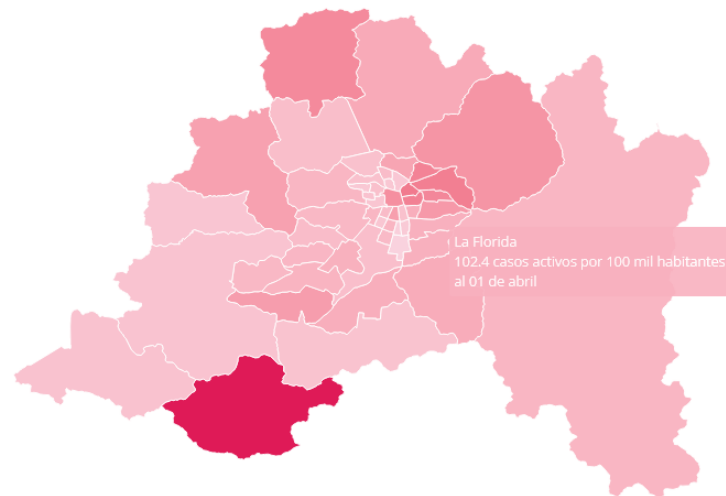
Metropolitana

Utilice los botones para elegir si graficar los casos activos o la tasa de casos por cada 100 mil habitantes (que permite una mejor comparación entre regiones ya que se ajusta a las diferencias de población).

Cifra para graficar:

- Casos activos por comuna
- Tasa de activos por cada 100 mil habitantes

Región de  
01 de abril



Mesa de datos COVID-19, casos activos por fecha de inicio de síntomas y comuna  
Ministerio de Ciencia, Tecnología, Conocimiento e Innovación

Figura 7.8: Plataforma de visualización de casos desarrollada por DATA UC.

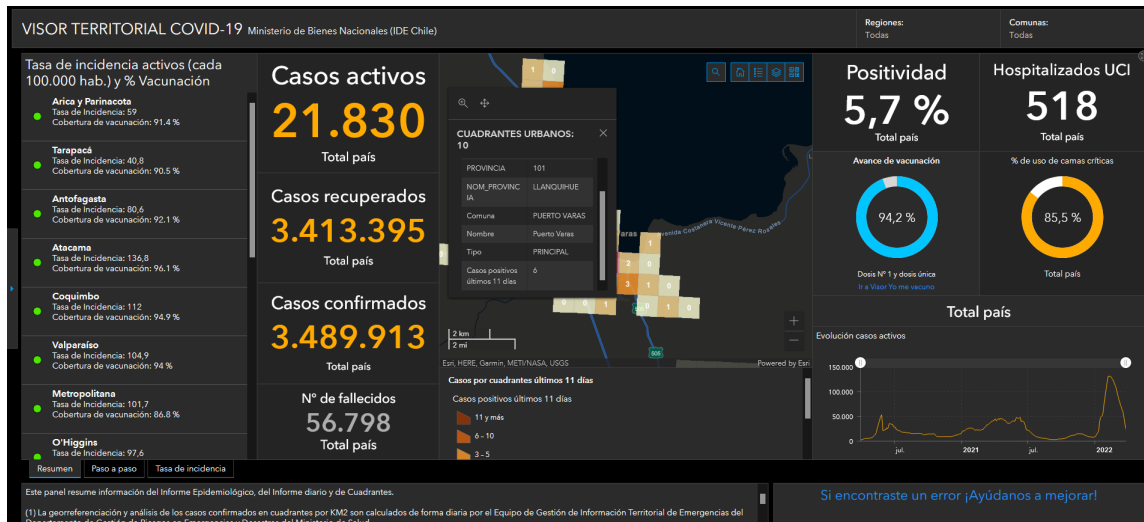


Figura 7.9: Visor Territorial COVID-19 del Ministerio de Bienes Nacionales.

# Bibliografía

- [1] Sasson, I. (2021). Age and COVID-19 mortality. *Demographic Research*, 44, 379-396.
- [2] El Comercio. China confirma que la neumonía de Wuhan puede transmitirse entre humanos, 2020. [en línea] <<https://www.elcomercio.com/tendencias/salud/china-neumonia-wuhan-contagio-personas.html>>. [consulta : 8 de Junio 2021].
- [3] OMS. Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020, 2020. [en línea] <<https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>>. [consulta : 8 de Junio 2021].
- [4] María Luisa Cisternas. Gasto público en salud: la falencia que pone a Chile en riesgo frente a la pandemia, 2020. [en línea] <<https://radio.uchile.cl/2020/04/19/gasto-publico-en-salud-la-falencia-que-pone-a-chile-en-riesgo-frente-a-la-pandemia/>>. [consulta : 8 de Junio 2021].
- [5] Manuel Henríquez. “No podemos saber cuánto ha gastado el gobierno en esta emergencia”, 2020. [en línea] <<https://www.ciperchile.cl/2020/08/19/no-podemos-saber-cuanto-ha-gastado-el-gobierno-en-esta-emergencia/>>. [consulta : 8 de Junio 2021].
- [6] Clínicas de Chile. Hasta el 45 % de los casos de COVID-19 pueden ser asintomáticos, 2020. [en línea] <<https://www.clinicasdechile.cl/noticias/hasta-el-45-de-los-casos-de-covid-19-pueden-ser-asintomaticos/>>. [consulta : 8 de Junio 2021].
- [7] BBC News Mundo. Coronavirus: 5 lugares que han aplicado estrategias exitosas contra la pandemia del Covid-19, 2020. [en línea] <<https://www.bbc.com/mundo/noticias-internacional-52067658>>. [consulta : 10 de Julio 2021].
- [8] Ministerio de Salud de Chile. 125º Informe Epidemiológico de COVID-19 e Informe de Estrategia Nacional Testeo, Trazabilidad y Aislamiento, 2021. [en línea] <<https://www.minsal.cl/125o-informe-epidemiologico-de-covid-19-e-informe-de-estrategia-nacional-testeo-trazabilidad-y-aislamiento/>>. [consulta : 8 de Junio 2021].
- [9] Felipe Elorrieta y Camilo Becerra Rodríguez. Trazabilidad: El factor clave que fue olvidado, 2021. [en línea] <<https://www.ciperchile.cl/2021/04/06/trazabilidad-el-factor-clave-que-fue-olvidado/>>. [consulta : 8 de Junio 2021].
- [10] Agencia Nacional de Investigación y Desarrollo. Concurso para la Asignación Rápida de Recursos para Proyectos de Investigación sobre el Coronavirus (covid-19) año 2020, 2020. [en línea] <<https://www.anid.cl/concursos/concurso/?id=379>>. [consulta : 9 de Septiembre 2021].
- [11] SSMSO. Cuenta Pública Participativa Gestión, 2019. [en línea] <<https://www.ssmso.cl/>>

- FilesComunicaciones/ResumenEjecutivoCuentaPúblicaParticipativaSSMSO2019.pdf* >. [consulta : 8 de Junio 2021].
- [12] Unidad de Salud Digital del SSMSO. Misión y Visión. [en línea] <<https://saluddigital.ssmso.cl/quienes-somos/mision-y-vision-3/>>. [consulta : 8 de Junio 2021].
- [13] World Health Organization. WHO COVID-19 Case definition, 2020. [en línea] <[https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance\\_Case\\_Definition-2022.1](https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2022.1)>. [consulta : 10 de Noviembre 2021].
- [14] Ministerio de Salud de Chile. PROTOCOLO DE COORDINACIÓN PARA ACCIONES DE VIGILANCIA EPIDEMIOLÓGICA DURANTE LA PANDEMIA COVID-19 EN CHILE: ESTRATEGIA NACIONAL DE TESTEO, TRAZABILIDAD Y AISLAMIENTO, 2020. [en línea] <<https://www.minsal.cl/wp-content/uploads/2020/07/Estrategia-Testeo-Trazabilidad-y-Aislamiento.pdf>>. [consulta : 06 de Marzo 2022].
- [15] Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M., Zimmerman, D. L. (2006). Geocoding in cancer research: a review. *American journal of preventive medicine*, 30(2), S16-S24.
- [16] Akakba, A., Lahmar, B. (2020). The use of Geocoding for Home Healthcare Application and Management an Epidemic Situation. Case of COVID-19 Virus Outbreak. *Geographica Pannonica*, 24(4).
- [17] Costa e Silva, E., Lopes, I. C., Correia, A., Faria, S. (2020). A logistic regression model for consumer default risk. *Journal of Applied Statistics*, 47(13-15), 2879-2894.
- [18] Najera, H., Ortega-Avila, A. G. (2021). Health and institutional risk factors of COVID-19 mortality in Mexico, 2020. *American journal of preventive medicine*, 60(4), 471-477.
- [19] Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE), 2022. [en línea] <<https://coronavirus.jhu.edu/map.html>>. [consulta : 04 de Abril 2022].
- [20] World Health Organization. WHO Coronavirus (COVID-19) Dashboard, 2022. [en línea] <<https://covid19.who.int/>>. [consulta : 04 de Abril 2022].
- [21] Mayo Clinic. U.S. COVID-19 map: What do the trends mean for you?, 2022. [en línea] <<https://www.mayoclinic.org/coronavirus-covid-19/map>>. [consulta : 04 de Abril 2022].
- [22] Centro Nacional de Epidemiología de España. Panel COVID-19, 2022. [en línea] <<https://cneccovid.isciii.es/covid19/>>. [consulta : 04 de Abril 2022].
- [23] Gobierno de Chile. Visor territorial COVID-19, Ministerio de Bienes Nacionales (IDE Chile), 2022. [en línea] <<https://covid.visorterritorial.cl/>>. [consulta : 04 de Abril 2022].
- [24] Gobierno de Chile. Cifras Oficiales COVID-19, 2022. [en línea] <<https://www.gob.cl/coronavirus/cifrasoficiales/>>. [consulta : 07 de Marzo 2022].
- [25] IBM. What is an Application Programming Interface (API), 2020. [en línea] <<https://www.ibm.com/cloud/learn/api>>. [consulta : 09 de Marzo 2022].
- [26] Matías Martínez. Memoria para optar al título de Ingeniero Civil Industrial: Métodos eficaces de búsqueda y corrección de direcciones de pacientes Covid-19, 2022.
- [27] Google. Planes de precios y coste de las APIs - Google Maps Platform, 2022. [en línea]

- <<https://mapsplatform.google.com/intl/es/pricing/>>. [consulta : 15 de Agosto 2022].
- [28] INE. Definición de Manzana Censal, 2015. [en línea] <<https://geoarchivos.ine.cl/File/boletines/DefiniciónManzanaCensal.pdf>>. [consulta : 09 de Marzo 2022].
- [29] INE. Ciudades, Pueblos, Aldeas y Caceríos, 2019. [en línea] <[https://geoarchivos.ine.cl/File/pub/Cd\\_Pb\\_Al\\_Cs\\_2019.pdf](https://geoarchivos.ine.cl/File/pub/Cd_Pb_Al_Cs_2019.pdf)>. [consulta : 09 de Marzo 2022].
- [30] INE. Plataforma de Datos Geoestadísticos, 2020. [en línea] <<https://ine-chile.maps.arcgis.com/apps/dashboards/e8292e6a13814b6b8bcfd3415ef4eb02>>. [consulta : 09 de Marzo 2022].
- [31] OMS. Información básica sobre la COVID-19, 2021. [en línea] <<https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>>. [consulta : 09 de Marzo 2022].
- [32] Ministerio de Salud de Chile. Chile realiza primer estudio a nivel mundial sobre uso de dosis de refuerzo en vacunas inactivadas: Aumenta efectividad en prevención y hospitalización por COVID-19, 2021. [en línea] <<https://www.minsal.cl/chile-realiza-primer-estudio-a-nivel-mundial-sobre-uso-de-dosis-de-refuerzo-en-vacunas-inactivadas-aumenta-efectividad-en-prevencion-y-hospitalizacion-por-covid-19/>>. [consulta : 04 de Abril 2022].