



UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA Y
CIENCIAS APLICADAS

CARACTERIZACIÓN Y ANALISIS DEL
MERCADO DE BODEGAS EN RM
CHILE, DESDE LA PERSPECTIVA DE
LOS USUARIOS

JULIO SEBASTIÁN HENRÍQUEZ GIMÉNEZ

MEMORIA PARA OPTAR AL TITULO DE
INGENIERO CIVIL INDUSTRIAL

PROFESOR GUÍA: SANTIAGO RAFAEL TRUFFA
SOTOMAYOR

SANTIAGO, FEBRERO DE 2023

UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA Y
CIENCIAS APLICADAS

AGRADECIMIENTOS

Quiero partir agradeciendo a mis profesores guías Santiago Truffa y Christian Wegmann, por haber confiado en mi trabajo y haberme apoyado con su conocimiento, sin el cual esta memoria no se podría haber llevado a cabo de esta forma. En segundo lugar, quisiera agradecer al equipo docente de la Universidad de los Andes, que es gracias a ellos que hoy cuento con las herramientas necesarias para poder llevar a cabo tareas de esta envergadura y gracias a ellos poder desarrollarme de manera integral.

En tercer lugar, quisiera agradecer a mi familia y amigos, los cuales siempre estuvieron allí como un pilar fundamental durante transcurso de mi carrera y siempre fueron el empuje para poder completar todos los desafíos propuestos a lo largo de esta. Sé que esto se mantendrá a lo largo de los futuros desafíos a los que me vea enfrentado.

Es por lo anteriormente expuesto que no puedo expresar nada más que gratitud a todos estos actores visibles en mi desarrollo como profesional, pero tampoco quisiera dejar de lado a todos los actores que ya sea con un saludo, una indicación, o un buen gesto me hicieron mirar con humildad el camino emprendido.

ÍNDICE GENERAL

RESUMEN	VI
1. INTRODUCCIÓN	1
1.1 MOTIVACIÓN.....	1
1.2 OBJETIVOS DE LA MEMORIA	2
1.2.1 Objetivo general	2
1.3 ALCANCES.....	3
1.4 METODOLOGÍA	3
2. MARCO TEÓRICO Y REVISIÓN BLIBLIOGRÁFICA.....	1
2.1 MARCO TEÓRICO	1
2.1.1 QGIS software	4
2.1.2 Stata software	5
2.1.3 RStudio Software.....	5
2.1.4 <i>Python</i>	5
2.1.5 Clusterización de datos.....	5
2.1.6 Métodos para elegir número de <i>clusters</i>	6
2.1.7 Distancia Euclidiana	7
3. METODOLOGÍA.....	8
3.1 DEFINICIÓN DE VARIABLES RELEVANTES A LEVANTAR	8
3.2 TRATAMEINTO BASES DE DATOS	9
3.2.1 Asociación de normativas, distritos y zonas de Santiago con direcciones de centros de bodegaje	9
3.2.2 Cruce de bases de datos SII	10
3.2.3 Geocodificación de direcciones contribuyentes personas jurídicas	12
3.2.4 Cruce de datos para armar base final.....	16
4. BASE DE DATOS	1
4.1 CAMPOS CONTENIDOS EN LA BASE DE DATOS.....	1
5. ANÁLISIS DE CLUSTER Y RESULTADOS.....	4
5.1 PREPARACIÓN DE LOS DATOS PARA ANÁLISIS	4
5.1.1 transformación de variables categóricas a <i>dummies</i>	4
5.1.2 Normalización de los datos y elección del número de <i>clusters</i>	5
5.1.3 Aplicación de kmeans para la generación de los clusters	7
5.2 RESULTADOS	7
5.2.1 Promdios de las distancias de infraestructura critica de los centros logísticos por <i>cluster</i>	7
5.2.2 Variables categoricas presentes por <i>cluster</i>	9
6. CONCLUSIONES Y RECOMENDACIONES	13
7. BIBLIOGRAFÍA	14
8. ANEXOS.....	17
ANEXO 1. TABLA DE VARIABLES RELEVANTES A LEVANTAR.....	18

ANEXO 2. DO-FILE UTILIZADO EN EL CRUCE DE BASES DE DATOS DEL SII.....	22
ANEXO 3: CANTIDAD DE REGISTROS POR COMUNA.....	25
ANEXO 4: MACRO UTILIZADA PARA EL TRASPASO DE COORDENADAS DE MULTIPLES VARIABLES	27
ANEXO 5: PORCENTAJE DE PRIMERAS DOS CORRIEDAS DE GEOCODING	28

ÍNDICE DE TABLAS

TABLA 1: <i>CRITERIA FOR DESCRIBING ULS</i>	3
TABLA 2: CANTIDAD DE TRABAJADORES INFORMADOS Y DISTANCIA PROMEDIO A CENTRO DE BODEGAJE MAS CERCANO POR <i>CLUSTER</i>	7
TABLA 3: DISTANCIA PROMEDIO DE CENTRO DE BODEGAJE A SUPERMERCADO Y METRO MAS CERCANOS POR <i>CLUSTER</i>	8
TABLA 4: DISTANCIA PROMEDIO DE MULTITIENDA Y CENTRO COMERCIAL A CENTRO DE BODEGAJE POR <i>CLUSTER</i>	8
TABLA 5: COMUNA Y DISTRITO CENSAL POR <i>CLUSTER</i>	9
TABLA 6: TIPO DE CONTRIBUYENTE PRESENTE POR <i>CLUSTER</i>	9
TABLA 7: RUBRO ECONOMICO PRESENTE POR <i>CLUSTER</i>	10
TABLA 8: CORREDOR DE CENTRO LOGISTICO PRESENTE POR <i>CLUSTER</i>	10
TABLA 9: TIPO DE CAPITAL PROPIO PRESENTE POR <i>CLUSTER</i>	11
TABLA 10: TRAMO DE FACTURACIÓN PRESENTE POR <i>CLUSTER</i>	11
TABLA 11: CLASE DE CENTRO LOGÍSTICO PRESENTE POR <i>CLUSTER</i>	12

ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1: METODOLOGÍA DE TRABAJO.....	4
ILUSTRACIÓN 2: DISTANCIA EUCLIDIANA.	7
ILUSTRACIÓN 3: <i>SPATIALJOIN</i> ZONAS RM Y CENTROS DE BODEGAJE.	10
ILUSTRACIÓN 4: CANTIDAD DE REGISTROS POR COMUNA.....	11
ILUSTRACIÓN 5: MUESTRA DE DIRECCIÓN CON ERROR DE TIPEO.	12
ILUSTRACIÓN 6: MUESTRA DE ".CSV" A GEOCODIFICAR.	13
ILUSTRACIÓN 7: MUESTRA DE REPORTE DE GEOCODIFICACIÓN.	14
ILUSTRACIÓN 8: EJEMPLO DE PRIMERA CORRIDA DE GEOCODING COMUNA DE SANTIAGO.	15
ILUSTRACIÓN 9: PLANO CON COORDENADAS DE COMUNAS ASIGNADAS A JULIO HENRÍQUEZ.....	16
ILUSTRACIÓN 10: TIPO DE VARIABLE.	4
ILUSTRACIÓN 11: INFORMACIÓN DATA SET <i>DUMMIES</i>	5
ILUSTRACIÓN 12: REGLA DEL CODO PARA DETERMINAR CANTIDAD DE <i>CLUSTERS</i>	6
ILUSTRACIÓN 13: ANÁLISIS DE SILUETA PARA LA DETERMINACIÓN DEL NÚMERO DE <i>CLUSTERS</i> . .	6
ILUSTRACIÓN 14: ETIQUETAS DE LOS <i>CLUSTERS</i> Y TAMAÑO DE CADA UNO.	7

RESUMEN

La temática que se trata en esta memoria es la caracterización y análisis del mercado de bodegas en la región metropolitana de Chile, esto aterrizado desde la perspectiva de sus usuarios, en este caso los contribuyentes personas jurídicas los cuales se relacionan de manera directa con los centros logísticos. Ambas partes tienen características propias las cuales no se han visibilizado como se relacionan y es por eso que surge una necesidad de buscar las conexiones entre ambos actores.

Para la realización de esto se utilizó tanto información proporcionada por la consultora *Global Properties Solutions* como información pública del Servicio de Impuestos Internos, fue en base a esta más la información disponible en la reportería de la industria la nos encamino hacia los principales *drivers* presentes en este mercado y poner el foco sobre lo que se estudia a lo largo de este escrito.

Una vez tratada la información y teniendo una base de datos robusta con sustento es que se realizaron análisis pertinentes de acuerdo a los alcances de esta memoria y mediante el análisis de *clusters* utilizando el algoritmo *kmeans* se logra agrupar a los contribuyentes personas jurídicas con los principales centros logísticos de la región metropolitana, en base a características en común

1. INTRODUCCIÓN

El propósito de este capítulo es presentar el tema que se aborda en este trabajo y destacar su relevancia. Con este fin, se dan a conocer los objetivos, los alcances y la metodología seleccionada.

1.1 Motivación

El mercado de bodegas ha experimentado un crecimiento considerable en los últimos años, debido al comercio electrónico que se potencio desde en octubre del 2019 y, posteriormente, por el confinamiento causado por el surgimiento del virus COVID-19. Ambos fenómenos generaron una considerable disminución en la operación de las tiendas físicas, lo que obligó a las distintas marcas a buscar alternativas para seguir vendiendo (Santiago, 2020).

En relación con anterior cada vez es más importante tratar la actividad logística, debido a los constantes cambios y adaptaciones que ha tenido en los últimos años, principalmente en la Región Metropolitana, donde cada vez aumenta más la densidad poblacional por m². Según Gajardo Polanco (2019) “La población total de la RMS debiera aumentar de 7,29 millones de personas en 2015 hasta 8,84 millones en 2035” lo que se traduce en un crecimiento global de 21,5%. Este hecho, impacta directamente en la necesidad de las empresas por contar con centros logísticos. Sin embargo cada vez hay mayor escaseces de suelos para proyectos, además de una disminución de la vacancia de los centros de bodegaje ya consolidados, entre el primer y segundo semestre de 2021 (Global Property Solutions, 2021) , con un incipiente aumento en los precios de los suelos.

Ante el escenario expuesto en los puntos anteriores es qué se vuelve cada vez más auspicioso generar un análisis acabado de todas las aristas relacionadas a los fenómenos logísticos presentes en la Región Metropolitana, puesto que este es un mercado que se está consolidando, además de presentar una oferta diversa de centros logísticos, en características, servicios y dimensiones, además de una gran variedad de clientes con múltiples necesidades para diversas industrias. Condiciones ideales a la hora de recopilar datos y buscar patrones, para que posterior su estudio se puedan identificar patrones de conducta y necesidades tipo

asociadas a las distintas industrias, cuestión de gran utilidad a la hora de querer replicar los modelos a otras regiones del país.

1.2 OBJETIVOS DE LA MEMORIA

1.2.1 Objetivo general

El objetivo específico de este trabajo es lograr caracterizar el mercado de bodegas de la Región Metropolitana de Santiago de Chile desde características y perspectivas relevantes para los usuarios, es decir, identificar de manera clara distintos perfiles de clientes y sus preferencias y patrones de aglomeración, con esto se espera llenar un vacío que existe respecto de esta información.

Objetivos específicos

- a) Generar un contexto especializado que presente que son los centros logísticos, definir finalidades y características generales.
- b) Identificar los principales *drivers* del mercado de bodegas.
- c) Seleccionar las variables de interés, tanto externas cómo internas a los centros logísticos, definir su naturaleza y métrica correspondiente.
- d) Realizar levantamiento de datos con el propósito de analizar el mercado de bodegas de la Región Metropolitana de Chile.
- e) Ejecutar una clusterización de los tipos de clientes de bodegas de la Región Metropolitana de Chile, de acuerdo con la afinidad de sus características.
- f) Entregar conclusiones y comentarios respecto de los hallazgos de esta investigación.

1.3 ALCANCES

En esta memoria, se pretende realizar un estudio relevante y actualizado del mercado de las bodegas de la Región Metropolitana de Chile, desde una mirada de los usuarios y las características que estos esperan de los centros logísticos. Todo esto con el fin de poder segmentar la oferta y la demanda de acuerdo con necesidades específicas.

Los datos recopilados fueron proporcionados por la consultora Global Property Solutions (en adelante GPS) y obtenidos a través del Servicio de Impuestos Internos (en adelante SII). En este último caso se utilizó información pertinente para el año 2020, con el fin de que esta fuera actualizada y completa.

Los *softwares*, tanto comerciales como de código abierto, se seleccionaron debido al alto nivel de detalle; lo que está acorde con lo que se espera conseguir en esta memoria. Además, permiten hacer análisis acabados de grandes cantidades de información y están consolidados en la industria.

1.4 METODOLOGÍA

Con el fin de conseguir el objetivo general y los objetivos específicos, se utilizó siguiente metodología:



Ilustración 1: Metodología de Trabajo.

Fuente: Elaboración Propia.

2. MARCO TEÓRICO Y REVISIÓN BIBLIOGRÁFICA

En este capítulo se presentan los fundamentos que apoyan la relevancia de la perspectiva de los usuarios en un mercado logístico que ha estado en auge durante los últimos años, además se da a conocer cómo otros países han resuelto y evolucionado en torno a la expansión logística y, a su vez, de que forma el mercado chileno ha clasificado y presentado patrones similares de descentralización del fenómeno. En el marco teórico se explican conceptos y modelos relevantes para la lectura de este escrito, junto con descripciones de programas usados para el análisis correspondiente.

2.1 MARCO TEÓRICO

Es clave diferenciar las características de una ubicación que atrae actividades de almacenamiento, principalmente bodegaje, de una instalación de tipo tránsito directo, o centro de distribución. El avance en la tecnología de los últimos años ha impulsado la constante mejora de la cadena de suministros. Esto ha obligado a la industria a migrar desde los grandes almacenes tradicionales a los nuevos centros de distribución que cuentan con grandes superficies para así facilitar el flujo de las distintas mercancías (Hesse & Rodrigue, 2004).

Entre los usos que se les da a los almacenes, está el manejo de carga que incluye: clasificación, consolidación y fraccionamiento, procesamiento de pedidos, etiquetado y/o empaquetado (Sakai, et al., 2020).

El propósito es minimizar el tiempo y el costo del transporte. La logística que subyace al proceso genera ventajas competitivas, por lo tanto, la elección del lugar de almacenamiento es de tipo estratégica, ya que esta requiere de grandes inversiones de capital y costos de mano de obra.

La descentralización de los centros de bodegaje se puede explicar, en parte, por la ganancia de los usuarios con la reducción de los costos de los suelos en relación con los de transporte, además, actualmente se manejan volúmenes mucho mayores. Hay una tendencia de que los nuevos proyectos de centros de distribución y bodegas se acercan al aeropuerto en Los Ángeles, CA.

(Sanggyun, 2020) en la actualidad hay que transportar grandes volúmenes de bienes a los costos más bajos, lo que se explica por los grandes avances de la tecnología de los últimos años, junto con la reducción de los costos de transporte. Debido a la consolidación, enrutamiento y gestión de la carga, se mejoró la flexibilidad de la ubicación para el transporte de carga y las instalaciones logísticas.

Los principales factores por considerar en el estudio de la ubicación de las bodegas son los siguientes: espacio adecuado, servicio al cliente, conexiones de tráfico favorables con proveedores y mercados clave, fácil acceso a autopistas, cercanía a los distintos puertos y mano de obra calificada. (Rajesh Kr, et al., 2018). La decisión sobre dónde colocar la bodega es considerada de carácter irreversible por los grandes costos que esto implica. En una encuesta sobre costos logísticos en Europa se observó que el costo de almacenamiento es el doble que el de inventario, y en Estados Unidos es, prácticamente, el mismo. Factores de infraestructura, tales como electricidad y el suministro de agua y la configuración de las telecomunicaciones, desempeñan un papel importante en el funcionamiento eficiente de un almacén (MacCarthy & Atthriawong, 2003).

La constante mejora en las autopistas es clave para entender por qué las bodegas migran a la periferia en busca de reducir costos. En este sentido, la gran demanda en la zona norte de Santiago. Se explica posiblemente por sus accesos a las autopistas, cercanía al aeropuerto y menores costos respecto de zonas más centrales. Los centros de distribución ofrecen servicios, además del simple arriendo del lugar, acceso controlado a los recintos, servicios de seguridad como cámaras guardias, ajustabilidad, es decir posibilidad de crecer dentro del mismo lugar, estacionamientos y zona de carga y descarga, posibilidad de implementación de oficinas de tipo modular, baños privados, etc.

Clasificación de bodegas GPS

A continuación, se presentan los principales antecedentes que diferencian un centro de distribución según Global Properties Solutions (2021).

Superficie mínima de un centro de bodegaje: 6.500 m².

Bodegas clase 1: superficie máxima ilimitada, superficie mínima de 800 m², altura mínima de 8 m.

Bodegas clase 2: superficie máxima de 1.800 m², superficie mínima de 200 m².

Los espacios logísticos urbanos se consideran como una interfaz entre las distintas cadenas de suministros y el entorno urbano. Estos articulan a los productores, distribuidores y consumidores. (Boudoin, et al., 2013). Es fundamental estudiar el diseño de estos espacios, luego sus características, principalmente su tamaño, funcionalidad y la cobertura con la que cuentan para facilitar la coordinación de los flujos del sistema logístico y por lo tanto mejorar la economía en los distintos negocios y ciudades. Con el fin de caracterizar los centros logísticos, se proponen 5 condiciones comunes para clasificarlos, *Functionalities, Stakeholders, Loading Volume, Type of merchandise, Type of company (related to type of service)*. (Meza-Peralta, et al., 2020).

Tabla 1: *Criteria for describing ULS.*

Table 3. Criteria for describing ULS.

#	Classification criteria	Description
1	Functionality	The specific features of the ULS that make it possible to achieve logistics management objectives. Types of functionalities: Materials management, (all stages of the chain), physical distribution, Cross Docking, storage and cross docking with storage.
2	Dimension	Length, extension or size of a ULS. Types of Dimension: 5164,5 m ² , 5610,6 m ² , 9599,1 m ² , 12,986,7 m ² , 450,000 m ² .
3	Loading volume	Load volume capacity of a ULS. Types of volume: (Depending on the size of the ULS).
4	Stakeholders	Group or organisation that is part or is affected by the design or location of a ULS, obtaining some benefit. Types of Stakeholders: Logistic service providers (Logistics operators), Wholesalers, retailers, private and public authorities, population and foreign investors.
5	Management of goods' flows.	Type of product that moves at a certain point. It applies to entries and exits from a warehouse, through a network or a logistics system. Types of flows: As for the type of cargo.

Fuente: (Meza-Peralta, et al., 2020)

2.1.1 QGIS software

Es una herramienta que provee información geográfica de forma libre y es de código abierto, es decir, es un Sistema de Información Geográfica (GIS) que rige bajo la licencia de “*General Public License*” (GNU), que soporte múltiples formatos. Por lo anterior se decidió implementar esta solución para trabajar con los datos del SII. De esta manera, se realizó el *geocoding* de las direcciones de las personas jurídicas como, a su vez, el cálculo de distancias a infraestructura crítica de los centros de bodegaje (QGIS.org, 2023).

Este *software*, además, permite la instalación de múltiples paquetes de complementos para ejecutar realizar actividades específicas, que se detalla a continuación:

- a) *Plugin NNJoin: Nearest neighbour join* se utiliza para unir pares de vectores, uno dado como *input* y la capa¹ con la cual se desea unir este. Deben tener características en común, con la finalidad de unirlos en la menor distancia. Todas las formas geométricas posibles se encuentran soportadas por este complemento y esta entrega como resultado un nuevo vector con la misma geometría y sistema de coordenadas inscritos por el vector de *input*.
- b) *Plugin MMQGIS*: Es una colección de operadores disponibles para ser aplicados a las distintas capas. Este complemento, basado en el lenguaje de programación *Python*, se utiliza para la manipulación de capas vectoriales. En este se debe entregar como *input* un archivo de formato .csv que se podrá unir con la información de la capa. Finalmente, retorna un nuevo archivo, también en formato .csv. Entre las principales aplicaciones, están las de unir capas, geocodificación², conversiones geométricas, *buffering*, entre otras.

¹ Capa: se refiere a un formato de datos vectorial. El utilizado por QGIS es el de tipo *shape* “.shp”, este contiene la localización de elementos de tipo geográfico y atributos que estén asociados a ellos.

² Geocodificación: es el proceso mediante el cual se busca transformar una descripción de una ubicación (dirección, nombre) en coordenadas específicas asociadas dentro de un sistema previamente definido.

2.1.2 Stata software

Es un programa estadístico utilizado para la ciencia de datos. Permite la manipulación de grandes cantidades de datos, debido a la visualización de estos, sus principales estadísticos y una generación de reportes de forma automática. A su vez *Stata* es una plataforma multiformato, por lo que soporta la entrada de múltiples archivos para su posterior manipulación y acepta un posterior almacenado, en una amplia variedad de formatos compatibles (StataCorp, 2023).

a) *Do-file*: es un editor de texto integrado en *Stata*. Como su nombre lo indica, se utiliza para generar una lista de comandos por ejecutar de manera secuencial por *Stata*. Generalmente, se utiliza, cuando se desean aplicar una serie de comandos a múltiples variables.

2.1.3 RStudio Software

Es un ambiente de desarrollo integrado (*IDE*, por sus siglas en inglés) gratuito que utiliza el lenguaje de programación R, es un programa de código abierto para la ciencia de datos, investigación científica y la adopción de un lenguaje técnico. Además incluye un destacador de sintaxis con editor integrado, el cual admite la ejecución directa de código, herramienta de gráficos, historial de comandos, control sobre los elementos presentes en la interfaz, y la capacidad de movilizarlos de acuerdo con los requerimientos (R Foundation, 2023).

2.1.4 Python

Es un lenguaje de programación de código abierto con múltiples propósitos, este se eligió porque que admite datos multiformato y permite ejecutar cambios de manera versátil y ágil sobre estos en grandes volúmenes (Python Software Foundation, 2023).

2.1.5 Clusterización de datos

Es un proceso que permite la asociación de grupos de entidades similares, es decir un método de agrupamiento. Este se ejecuta por medio de un procedimiento estadístico multivariante. Se inicia

con un conjunto de datos que contiene información sobre una muestra de entidades e intenta reorganizarlas en grupos relativamente homogéneos. En virtud de lo anterior es que el uso de *clusters* está relacionado principalmente con las siguientes 4 actividades (Aldenderfer & Blashfield, 1984):

- a) Desarrollo de tipologías o clasificaciones.
- b) Investigación de mapas conceptuales que agrupan entidades de manera útil.
- c) Generación de hipótesis por exploración de datos.
- d) Aplicación test de hipótesis para corroborar si los patrones definidos a través de procedimientos alternativos se encuentran en realidad presentes en la muestra elegida.

2.1.6 Métodos para elegir número de *clusters*

- a) El método del codo: este busca la minimización de la distancia *intra-cluster* y maximizar las distancias *inter-cluster*, es decir busca las distancias medidas de las observaciones con respecto de su respectivo centroide. Es una técnica visual en la que se busca identificar el punto tal que ante un incremento del número de *clusters* la variación en la distancia *inter-cluster* no mejore sustancialmente.
- b) Análisis de silueta: esta mide la calidad del agrupamiento de acuerdo a la distancia de separación entre los *clusters*, es decir que tan cerca están los puntos de un *cluster* con respecto de los puntos del siguiente en una escala que va [-1, 1] siendo -1 muestras asignadas a un *cluster* erróneo y cercano a 1 que se encuentran lejanos a los *clusters* vecinos siendo la fórmula del coeficiente la siguiente:

$$S = \frac{b - a}{\max(a, b)}$$

a = distancia media *intra-cluster*.

b = distancia media a las observaciones al *cluster* más cercano.

(Gonzalo, 2019)

2.1.7 Distancia Euclidiana

Es relevante hacer la distinción entre lo que es el camino recorrido y la distancia entre dos puntos, ya que estos en múltiples oportunidades pueden variar de acuerdo con las necesidades del instante. La distancia Euclidiana es la distancia física en línea recta, en otras palabras, es la longitud del camino más corto posible que se pueda trazar entre dos puntos. La siguiente figura ilustra un ejemplo de un posible caso de distancia recorrida vs la distancia Euclidiana.

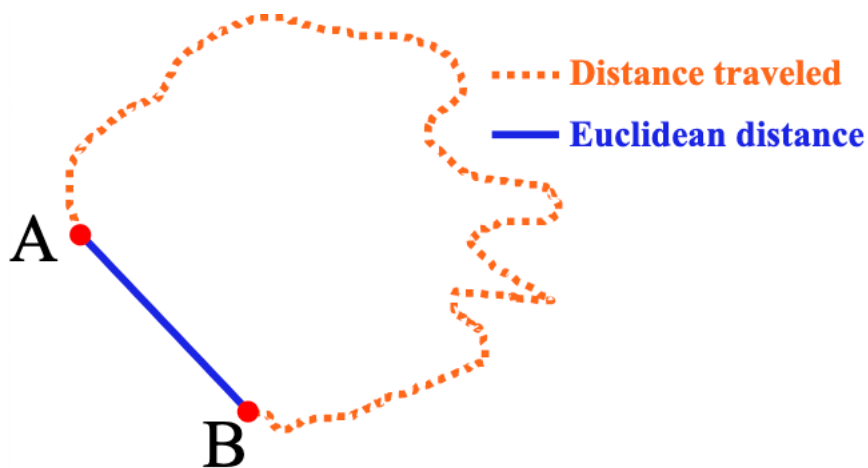


Ilustración 2: Distancia Euclidiana.

Fuente: (Euclides, 2022)

3. METODOLOGÍA

En este capítulo, se presenta la metodología realizada para la obtención de datos y su depuramiento de acuerdo con las necesidades requeridas. El propósito es realizar un análisis posterior y, de esta forma, conseguir información útil para la caracterización del mercado de bodegas de la Región Metropolitana de Santiago de Chile.

En relación con el trabajo de tratamiento, cruce y depuración de bases de datos, es importante señalar que este proceso se realizó junto con Tomás Apraiz y Gastón Larrea. El grupo abordó las temáticas, con los cuales se compartió problemas hermanos, por lo que debimos recopilar información de los mismos lugares y realizar los mismos tratamientos, por lo que se decidió unir fuerzas para lograr abarcar la mayor cantidad de información posible y de manera expedita. El grupo usó las mismas bases de datos, pero cada uno de los integrantes abordó distintas problemáticas.

3.1 DEFINICIÓN DE VARIABLES RELEVANTES A LEVANTAR

En primer lugar, a partir de la información de los *drivers* disponibles en el mercado de bodegas de Chile se definieron las variables de interés por levantar y su efecto sobre este mercado. En segundo lugar, se elaboró una tabla (ver anexo 1) en la que se muestra un consolidado de todas las variables consideradas como importantes. Estas se clasificaron en internas (con características intrínsecas de una bodega o centro de bodegaje) o externas (características relevantes de zona, principalmente geográficas).

La tabla cuenta con las siguientes columnas:

- a) Variables: se indica el nombre de la variable por investigar.
- b) Métricas: En esta columna, se encuentra la unidad de medición o la naturaleza de la variable.

- c) La tenemos: esta columna muestra si la variable contiene la información requerida, que, en caso contrario, habrá que requerirla con la ayuda de G.P.S.
- d) Fuente: en esta columna se indican las fuentes de las cuales se levantarán los datos, es decir, si la van a conseguir los investigadores o si se solicitará a otra empresa.
- e) Descripción: en esta columna, se muestra una descripción de la variable que se desea analizar.

3.2 TRATAMIENTO BASES DE DATOS

3.2.1 Asociación de normativas, distritos y zonas de Santiago con direcciones de centros de bodegaje

Para comenzar con este proceso se utilizó el programa de Google Earth Pro para transformar la base de datos entregada por GPS, de formato ESRI³ a formato *Shapefile* para lograr hacerlo compatible con *QGIS*. Una vez realizado el cambio de formato se logró importar la base de datos, la cual contenía direcciones de los principales centros de bodegaje de la Región Metropolitana y luego unirla mediante la utilización del complemento *NNJoin* en su opción de “*SpatialJoin*” con el fin de lograr incorporar las normativas, distritos y zonas de Santiago vigentes a cada centro logístico.

³ ESRI: es un formato de archivo creado por la empresa Environmental Research Institute Inc. El cual consiste en un archivo de tipo representación cartográfica.



Ilustración 3: *SpatialJoin* zonas RM y centros de bodegaje.

Fuente: Elaboración Propia.

En la Ilustración 3 se puede apreciar en color verde las distintas zonas de Santiago y en color amarillo, los principales centros de bodegaje industrial de la Región. Una vez habiendo realizado este proceso para todas las características mencionadas anteriormente se prosiguió con la obtención y depuración de bases de datos del SII.

3.2.2 Cruce de bases de datos SII

En primer término, se tuvo que descargar de la página del SII las bases de datos de “Nomina de direcciones históricas de contribuyentes personas jurídicas” y la de “Nomina de empresas personas jurídicas año comercial 2020” y comenzar a limpiarlas a modo de obtener los datos relevantes para la investigación, para esto se utilizó el *software* de *Stata* debido a la gran cantidad de datos con los que se estaba trabajando.

Estas bases contenían un universo en principio de 686.030 datos correspondientes a la base de “Nomina de empresas personas jurídicas año comercial 2020” y de 1.048.575 para la base de “Nomina de direcciones históricas de contribuyentes personas jurídicas” en su archivo correspondiente a los domicilios.

Primero se les dio formato a los títulos de los campos relevantes, ya que estos no eran reconocibles por temas de tildes y ñ, una vez realizado esto se prosiguió con la eliminación de los datos pertenecientes a regiones distintas de la Metropolitana, además de eliminar las direcciones que no se encontraran vigentes. Finalmente, luego de depuradas las bases se realizó la unión de estas utilizando la clave “Rut” como vinculo consiguiéndose un total de 360.994 direcciones vigentes para contribuyentes personas jurídicas dentro de la región metropolitana. En el anexo 2 se presenta el código utilizado en el do file de Stata para lograr el cruce.

De un total de 360.994 direcciones con *match*, estas se distribuyeron dentro de las 52 comunas de la región Metropolitana de Santiago de Chile de la siguiente forma presentada en la Ilustración 4 (el detalle se encuentra en el anexo 3):

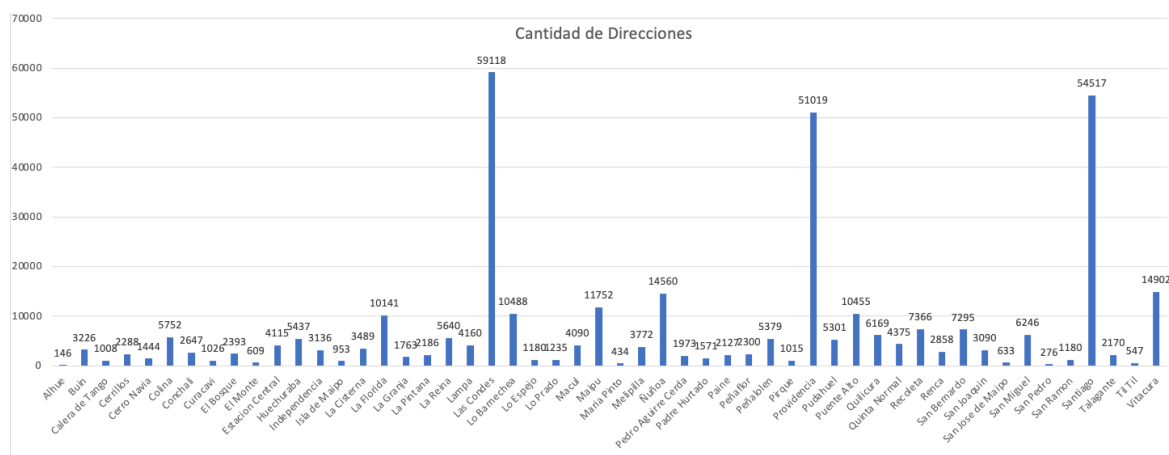


Ilustración 4: Cantidad de registros por comuna

Fuente: Elaboración propia

3.2.3 Geo codificación de direcciones contribuyentes personas jurídicas

Una vez conseguidas y filtradas las direcciones correspondientes a los contribuyentes personas jurídicas de la region metropolitana vigentes al año 2020, se continuo con el proceso de geo codificación, y el tratamiento de la base de datos.

a) Normalización de base de direcciones: acá se tuvo que partir un proceso el cual consistió en dejar las direcciones rescatadas del SII de una forma tal que se pudieran reconocer, y posteriormente ser encontradas por un complemento de cartografía. En primer lugar se comenzó por identificar todas las direcciones las cuales su numeración y nombre de calle se encontrarán en la misma casilla de Excel. En la siguiente ilustración se muestra un ejemplo de este problema.

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
										numero	bloque	departamen	villapoblacio	ciudad	vigencia2	_merge	comuna2
1	como sub, rubro, e actividad, pc region	provincia	comuna	vigencia	fecha	tipodireccion	calle										
2	RCIO A VENTA AL PC VENTA AL PC XIII REGION	Santiago	LAS CONDES	S	24/08/21	DOMICILIO	LA CAPITANIA 80 OF 108 PS 1										
3	RCIO A VENTA AL PC VENTA AL PC XIII REGION	Chacabuco	LAMPA	S	16/09/20	DOMICILIO	LT 153 MIRADOR DEL VALLE BATUCO										
4	RCIO A VENTA AL PC OTRAS ACTIV XIII REGION	Chacabuco	COLINA	S	13/10/20	DOMICILIO	EL CASTILLO LT 2 C 3										
5	SPORT OTRAS ACTIV TRANSPORT XIII REGION	Melipilla	MARIA PINTI	S	22/01/20	DOMICILIO	LT 2 ST C PI SAN ANDRES										
6	DADES FONDOS Y S FONDOS Y S XIII REGION	Santiago	LAS CONDES	S	01/09/20	DOMICILIO	LA CAPITANIA 80 OF 108 PS 1										
7	TRIA FABRICADOI FABRICADOI XIII REGION	Melipilla	MELIPILLA	S	09/04/20	DOMICILIO	SECT B 1 H 2 - 3 STA ELISA										
8	DADES ACTIVIDADES OTRAS ACTIV XIII REGION	Santiago	MACUL	S	12/07/21	DOMICILIO	DR M BARROS BORGONO 71 OF 1105 11P										
9	RCIO A VENTA AL PC OTRAS ACTIV XIII REGION	Santiago	PROVIDENCI	S	15/10/20	DOMICILIO	DR M BARROS BORGONO 71 OF 1105 11P										
10	CULTIVO CULTIVO DE CULTIVO DE XIII REGION	Melipilla	SAN PEDRO	S	11/02/20	DOMICILIO	LT 5 FDO CEMENTERIO										
11	DADES ACTIVIDADES OTRAS ACTIV XIII REGION	Melipilla	MELIPILLA	S	26/10/20	DOMICILIO	LT 15 - A - 6 PC 15 BOLLENAR										
12	TRIA FABRICADOI FABRICADOI XIII REGION	Santiago	EL BOSQUE	S	22/04/21	DOMICILIO	CAN LO HERRERA										
13	DADES ACTIVIDADES OTRAS ACTIV XIII REGION	Melipilla	SAN PEDRO	S	23/09/20	DOMICILIO	LT 5 A LT 5 LOS AROMOS INT										
14	IMACK ACTIVIDADES ACTIVIDADES XIII REGION	Santiago	LAS CONDES	S	16/09/20	DOMICILIO	LA CAPITANIA 80 OF 108 PS 1										
15	RCIO A VENTA AL PC VENTA AL PC XIII REGION	Santiago	LAS CONDES	S	28/09/20	DOMICILIO	LA CAPITANIA 80 OF 108 PS 1										
16	DADES OTRAS ACTIV OTROS SERV XIII REGION	Santiago	LAS CONDES	S	24/01/20	DOMICILIO	LA CAPITANIA 80 OF 108 PS										

Ilustración 5: Muestra de Dirección con error de tipeo.

Fuente: Elaboración Propia.

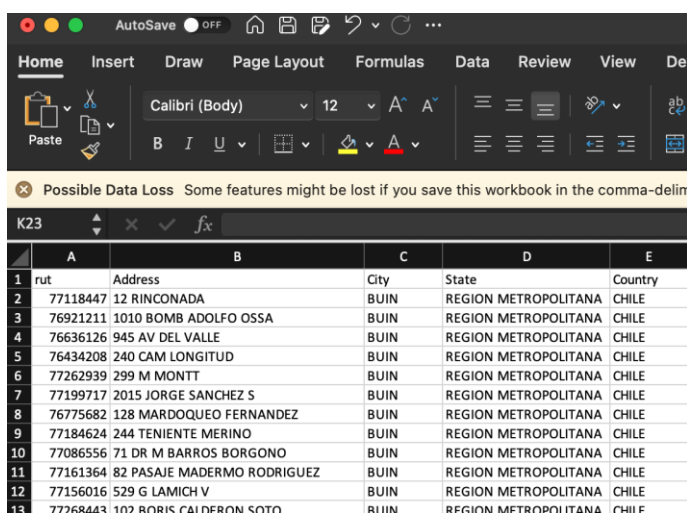
En esta parte se dividió la totalidad de direcciones encontradas entre mis tres compañeros a modo de poder realizar la normalización de manera más ágil. Las direcciones que tuvieran el problema de estar contenidas tanto el nombre de la calle junto con el resto de los campos fue trasladados a un archivo Excel el cual fue provisto por nuestro profesor guía, Christian Wegmann, llamado “SEPARA LETRA DE NUMERO.xls” el cual como su nombre lo indica es una macro la cual tiene como función separar el texto de los números.

Gracias a este Excel se simplifico en gran parte lo que pudo haber sido una revisión de uno a uno. Luego de correr esta macro fue necesario volver a revisar los datos para asegurarse que no tuvieran nuevos errores y una vez teniendo separado y solucionado el tema de los nombres de las calles y

los números se prosiguió con la eliminación de las direcciones que simplemente ya sea debido a que estaban mal ingresadas o algún otro problema, no iban a poder ser encontradas por un complemento de cartografía.

b) Geo codificación e inspección visual: una vez conseguida la normalización de las direcciones anteriormente mencionadas se prosiguió con la aplicación de un filtro a nivel comunal el cual sirvió para facilitar una inspección visual una vez realizados los *geocoding* de estas. Posteriormente con mis compañeros nos dividimos las comunas a geo codificar de manera equitativa.

- i. Dar formato a las direcciones: para poder utilizar el complemento *MMQGIS*, era necesario tener ciertos campos para poder buscar las coordenadas de cada dirección, estos eran “*Address*”, que correspondía a la concatenación del número de la dirección seguido del nombre de la calle, “*City*”, el cual correspondía a la comuna que se deseaba buscar, “*State*”, que era la región de interés y finalmente “*Country*”, refiriéndose al país a buscar.



	A	B	C	D	E
1	rut	Address	City	State	Country
2	77118447	12 RINCONADA	BUIN	REGION METROPOLITANA	CHILE
3	76921211	1010 BOMB ADOLFO OSSA	BUIN	REGION METROPOLITANA	CHILE
4	76636126	945 AV DEL VALLE	BUIN	REGION METROPOLITANA	CHILE
5	76434208	240 CAM LONGITUD	BUIN	REGION METROPOLITANA	CHILE
6	77262939	299 M MONTT	BUIN	REGION METROPOLITANA	CHILE
7	77199717	2015 JORGE SANCHEZ S	BUIN	REGION METROPOLITANA	CHILE
8	76775682	128 MARDOQUEO FERNANDEZ	BUIN	REGION METROPOLITANA	CHILE
9	77184624	244 TENIENTE MERINO	BUIN	REGION METROPOLITANA	CHILE
10	77086556	71 DR M BARROS BORGONO	BUIN	REGION METROPOLITANA	CHILE
11	77161364	82 PASAJE MADERMO RODRIGUEZ	BUIN	REGION METROPOLITANA	CHILE
12	77156016	529 G LAMICH V	BUIN	REGION METROPOLITANA	CHILE
13	77768443	102 BORIS CALDERON SOTO	BUIN	REGION METROPOLITANA	CHILE

Ilustración 6: Muestra de ".csv" a geo codificar.

Fuente: Elaboración Propia.

Una vez corrida la geo codificación, proceso que demoraba entre 2 a 8 horas por corrida dependiendo de la cantidad de direcciones ingresadas, se entregaba un reporte con la cantidad de direcciones que fueron efectivamente encontradas. Se muestra un ejemplo de esto en la siguiente ilustración.

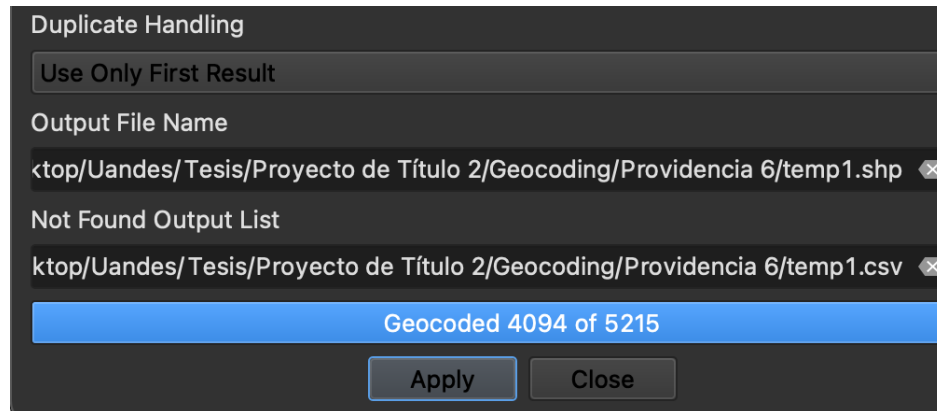


Ilustración 7: Muestra de reporte de geocodificación.

Fuente: Elaboración Propia.

- ii. Inspección visual de resultados: una vez realizada la primera corrida por comuna de geo codificación se realizaba una inspección visual de si los puntos encontrados estaban efectivamente sobre la comuna objetivo o si estos se habían ubicado sobre otras, en el caso de que se encontraran fuera de los límites de la comuna se debió eliminar esas observaciones para luego analizar cómo proceder respecto de cuál era la falla de dicha dirección, en la siguiente ilustración se observa un ejemplo donde gran parte de los puntos caen fuera de la comuna.

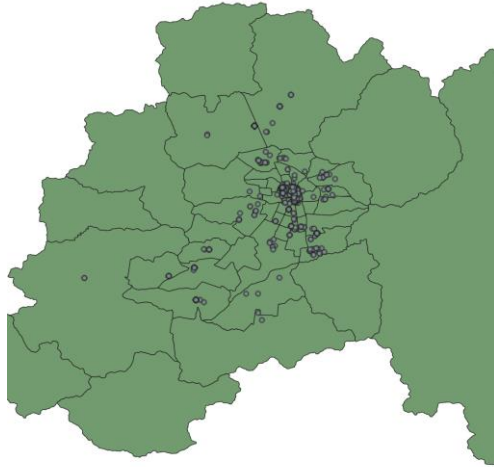


Ilustración 8: Ejemplo de Primera corrida de geocoding comuna de Santiago.

Fuente: Elaboración propia.

- iii. Con el resultado del *geocoding* ya eliminados los registros mal puestos se procedió a realizar la descarga con la tabla de atributos en formato .csv para migrar las coordenadas encontradas a la base de datos original, debido a la gran cantidad de atributos que se debían traspasar se realizó una macro en Excel a modo de poder agilizar el proceso y no consumir tantos recursos computacionales (anexo 4).
- iv. Repetición de normalización: una vez habiendo realizado los geocoding a todas las comunas y habiendo migrado las coordenadas bien ubicadas se debió repetir el proceso de normalización de las direcciones para todas aquellas que no tuvieran y así poder encontrar nuevos errores de modo de poder obtener la mayor cantidad de direcciones con coordenadas de manera correcta, por esto se debió volver a aplicar el proceso anterior a los contribuyentes personas jurídicas que no contaran con coordenadas asociadas a su Rut. En el anexo 4 se presenta una tabla resumen de las dos primeras corridas de geocoding.

En la siguiente ilustración se puede apreciar un plano de las comunas de la región Metropolitana de Chile una vez terminado de asignar coordenadas al tercio de la base de datos del SII que tuvo asignada en la separación que realizamos con Tomás Apraiz y Gastón Larrea.

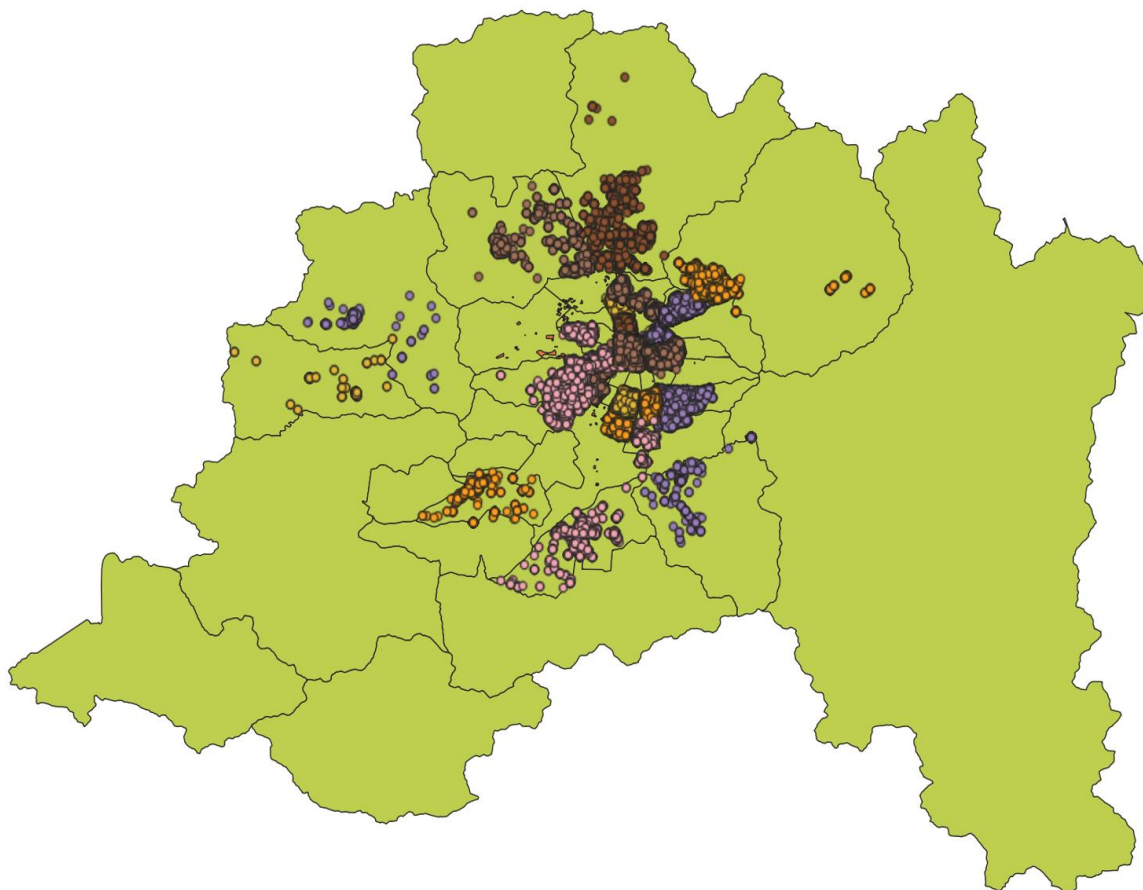


Ilustración 9: Plano con coordenadas de comunas asignadas a Julio Henríquez.

Fuente: Elaboración propia.

3.2.4 Cruce de datos para armar base final

Una vez conseguido la mayor cantidad de contribuyentes personas jurídicas asociadas con una dirección y coordenadas se procedió a realizar el cruce de estas con la capa de bodegas proporcionada por GPS y se notó que pocos registros caían coincidentes con los polígonos de los centros logísticos, por lo que se decidió asociar a los contribuyentes personas jurídicas de acuerdo a su comuna y distrito censal y en base a eso realizar el cruce con los polígonos de los centros

logísticos, y tomando esos filtros en consideración se realizó un *SpatialJoin* para calcular las distancias y encontrar cual era el centro logístico más cercano a cada contribuyente persona jurídica

Finalmente, y teniendo presente el análisis que se desea aplicar sobre los datos se procedió a armar la base de datos final con la mayor cantidad de información que se pudo recopilar para caracterizar las principales variables relevantes para los contribuyentes personas jurídicas en relación a los centros logísticos.

4. BASE DE DATOS

En este capítulo se detalla la base de datos que se utilizó para armar los clusters de contribuyentes personas jurídicas con las principales características de relevancia levantadas sobre los centros logísticos y su entorno.

4.1 CAMPOS CONTENIDOS EN LA BASE DE DATOS

- a) Rubro Económico: este dato proveniente del SII se genera a través de un código declarado por el contribuyente durante la operación renta de cada año, este puede ser común entre varios contribuyentes y no representa necesariamente la actividad económica principal de este.
- b) Distancia a Infraestructura de interés: estas son distancias que fueron calculadas con el fin de poder agrupar de acuerdo a necesidades e intereses en común. Estas son:
 - i. Distancia del contribuyente al centro de bodegaje más cercano.
 - ii. Distancia de los centros de bodegaje a centros comerciales, Homocentros, multitiendas, supermercados más cercanos.
 - iii. Distancia de los centros de bodegaje a la estación de metro más cercana.
- c) Cantidad de trabajadores dependientes informados: como lo dice su nombre es la cantidad de trabajadores a los que el contribuyente persona jurídica tiene declarados recibiendo renta, esta variable es interesante de revisar en conjunto a las demás para encontrar patrones.
- d) Tipo de contribuyente: definición del SII en la cual se clasifica a los profesionales ya las personas que ejercen actividades u operaciones lucrativas de forma independiente, además de las sociedades de profesionales, anónimas, etc. De esta definición se ve la tributación y ésta puede ser común a múltiples contribuyentes.
- e) Nombre del centro de bodegaje más cercano: en este campo se encuentra el código y nombre del centro de bodegaje que se encuentre más cercano al contribuyente persona jurídica, este se hizo según compartieran comuna y distrito censal.

- f) Comuna-Distrito: campo que mezcla las comunas y los distritos de los contribuyentes personas jurídicas, este campo fue clave a la hora de relacionar la información de los contribuyentes a la de los centros logísticos.
- g) Corredor centro logístico: este campo nos indica en que sector dentro de la región metropolitana se encuentra ubicado el centro de bodegaje: Norponiente, Norte, Oriente, Poniente o Sur. Este campo también puede ser común a varios centros logísticos.
- h) Tramo capital propio: este campo proviene del SII, este puede ser tanto positivo como negativo, va en una escala del 1 al 10 en ambos casos y por temas de dimensionalidad se combinaron, el negativo yendo del 1 al 10 y el positivo del 11 al 20. Esta definición también puede ser común en varios contribuyentes.
- i. $0 < \text{Capital Propio} \leq 10 \text{ UF}$
 - ii. $10 < \text{Capital Propio} \leq 25 \text{ UF}$
 - iii. $25 < \text{Capital Propio} \leq 50 \text{ UF}$
 - iv. $50 < \text{Capital Propio} \leq 100 \text{ UF}$
 - v. $100 < \text{Capital Propio} \leq 250 \text{ UF}$
 - vi. $250 < \text{Capital Propio} \leq 500 \text{ UF}$
 - vii. $500 < \text{Capital Propio} \leq 1.000 \text{ UF}$
 - viii. $1.000 < \text{Capital Propio} \leq 2.000 \text{ UF}$
 - ix. $2.000 < \text{Capital Propio} \leq 10.000 \text{ UF}$
 - x. $10.000 < \text{Capital Propio}$
- (Servicio de Impuestos Internos, 2022)
- i) Tramo de Facturación: este dato proviene del SII este proviene de las ventas anuales y no representa necesariamente su valor económico real. Esta definición puede ser común a varios contribuyentes.
- i. Sin Información: corresponde a contribuyentes cuya información tributaria declarada, no permite determinar un monto estimado de ventas.
 - ii. 1er Rango Micro Empresa: 0,01 a 200,00 UF Anuales
 - iii. 2do Rango Micro Empresa: 200,01 a 600,00 UF Anuales

- iv. 3ro Rango Micro Empresa: 600,01 a 2.400,00 UF Anuales
- v. 1er Rango Pequeña Empresa: 2.400,01 a 5.000,00 UF Anuales
- vi. 2do Rango Pequeña Empresa: 5.000,01 a 10.000,00 UF Anuales
- vii. 3er Rango Pequeña Empresa: 10.000,01 a 25.000,00 UF Anuales
- viii. 1er Rango Mediana Empresa: 25.000,01 a 50.000,00 UF Anuales
- ix. 2do Rango Mediana Empresa: 50.000,01 a 100.000,00 UF Anuales
- x. 1er Rango Gran Empresa: 100.000,01 a 200.000,00 UF Anuales
- xi. 2do Rango Gran Empresa: 200.000,01 a 600.000,00 UF Anuales
- xii. 3er Rango Gran Empresa: 600.000,01 a 1.000.000,00 UF Anuales
- xiii. 4to Rango Gran Empresa: más de 1.000.000,01 UF Anuales

(Servicio de Impuestos Internos, 2022)

- j) Clase: definición de GPS en la que los centros logísticos se dividen en clase 1 y clase 2 según sus dimensiones y altura de viga. Esta definición puede ser común a más de un centro logístico.

5. ANÁLISIS DE CLUSTER Y RESULTADOS

En este capítulo se dará a conocer la mitología empleada para realizar el análisis de cluster, además de presentar los resultados de este y sus principales características distintivas consolidadas.

5.1 PREPARACIÓN DE LOS DATOS PARA ANÁLISIS

5.1.1 transformación de variables categóricas a *dummies*

Para esto se comenzó separando las variables numéricas de las categóricas, ya que estas últimas se debieron transformar en *dummies*, ya que para agrupar los datos en *clusters* se utilizó el método de *kmeans*, esta es una técnica estadística multivariante de tipo algoritmo no supervisado, ya que los grupos de forman por características en común, y este recibe como parámetros de entrada datos numéricos.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23701 entries, 0 to 23700
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   RubroEconomico                           23701 non-null  object
1   # DeTrabajadoresDependientesInformados  23701 non-null  float64
2   TipoDeContribuyente                     23701 non-null  object
3   NombreCentroBodegajeMasCercano          23701 non-null  object
4   DistCentroDeBodMasCercano (m)            23701 non-null  float64
5   ComunaDistrito                          23701 non-null  object
6   DistCentroDeBodSupermercados (m)        23701 non-null  float64
7   CorredorBodega                          23701 non-null  object
8   DistCentroDeBodMetro (m)                23701 non-null  float64
9   DistCentroDeBodMultitienda (m)          23701 non-null  float64
10  DistCentroDeBodCentroComercial (m)       23701 non-null  float64
11  TramoCapitalPropio                      23701 non-null  object
12  TramoFacturacion                        23701 non-null  object
13  Clase                                   23701 non-null  object
dtypes: float64(6), object(8)
memory usage: 2.5+ MB
```

Ilustración 10: Tipo de variable.

Fuente: Elaboración propia.

Cómo se observa en la Ilustración 10 hay 8 variables categóricas a transformar en *dummies*, las que una vez transformadas pasan a ser 235 variables tipo *dummy*.


```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 23701 entries, 0 to 23700  
Columns: 241 entries, # DeTrabajadoresDependientesInformados to Clase_UNO  
dtypes: float64(6), uint8(235)  
memory usage: 6.4 MB
```

Ilustración 11: Información Data set *dummies*.

Fuente: Elaboración propia.

5.1.2 Normalización de los datos y elección del número de *clusters*

Cómo el método *kmeans* busca minimizar las distancias euclidianas de las observaciones respecto de los centros definidos, es necesario normalizar los datos, ya que no tendría ningún sentido comparar la distancia de los centros logísticos con el tramo de facturación de los contribuyentes personas jurídicas, al tratarse de variables distintas tanto en su naturaleza, como en su finalidad.

Para la normalización de los datos se utilizó el método *MinMaxScaler ()* del paquete de *sklearn* de python, el cual busca traducir cada variable a una magnitud equivalente, se pondera la desviación estándar de la variable por la diferencia de su valor máximo con el mínimo y a este resultado se le suma el mínimo.

Una vez normalizados los valores de la base de datos de *dummies* se procedió a aplicar el método del codo y el análisis de silueta para determinar el número óptimo de *clusters* para las observaciones.

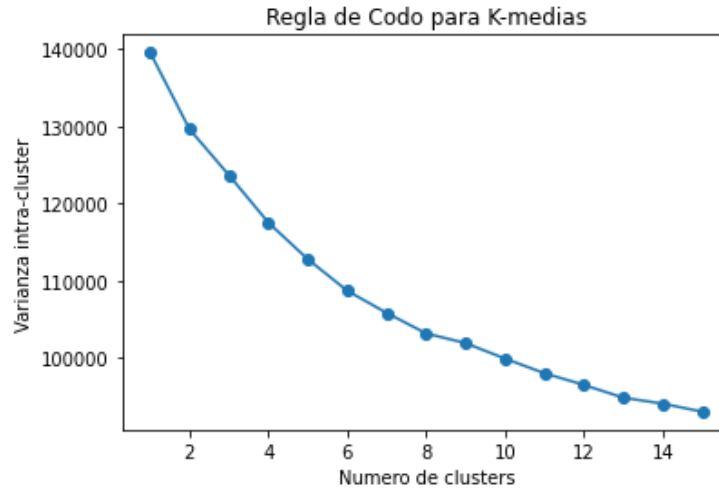


Ilustración 12: Regla del codo para determinar cantidad de *clusters*.

Fuente: Elaboración propia.

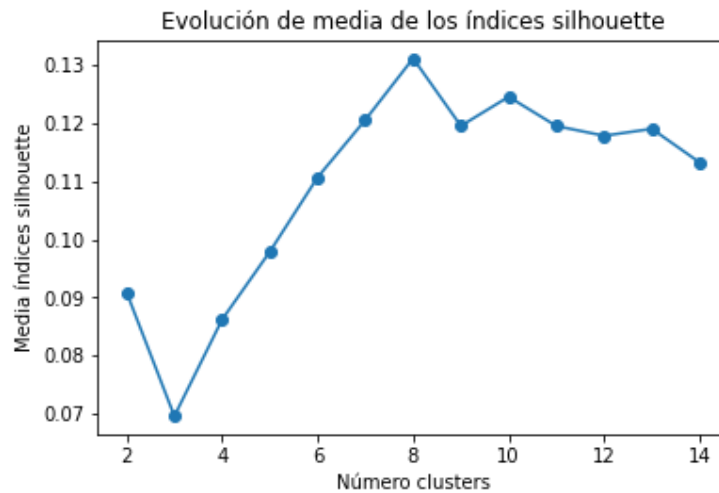


Ilustración 13: Análisis de silueta para la determinación del número de *clusters*.

Fuente: Elaboración propia.

Como se puede apreciar en las ilustraciones 12 y 13 el número óptimo de *clusters* para la muestra es de 8, ya que para esta cantidad se maximiza el indicador de silueta y se minimiza la varianza *intra-cluster* para el aumento en la cantidad, es decir es el codo.

5.1.3 Aplicación de kmeans para la generación de los clusters

Una vez determinado como número óptimo de clusters 8 se aplicó el algoritmo kmeans formándose los grupos y etiquetándose de la siguiente forma:

	size
0	1686
1	4216
2	2498
3	4577
4	3604
5	2172
6	3021
7	1927

Ilustración 14: Etiquetas de los *clusters* y tamaño de cada uno.

Fuente: Elaboración propia.

5.2 RESULTADOS

5.2.1 Promedios de las distancias de infraestructura critica de los centros logísticos por *cluster*

A continuación, se presentan tablas con los resultados promedio de las distancias

Tabla 2: Cantidad de trabajadores informados y distancia promedio a centro de bodegaje más cercano por *cluster*.

Cluster Labels	# De Trabajadores Dependientes Informados	Promedio de DistCentroDeBodMasCercano (m)
0	12	1708,52
1	41	876,95
2	23	602,76
3	19	1439,09
4	13	1977,35
5	33	824,62
6	26	671,06
7	9	960,06
Total	23	1158,81

Fuente: Elaboración propia.

Tabla 3: Distancia promedio de centro de bodegaje a supermercado y metro más cercanos por *cluster*.

Cluster Labels	Promedio de DistCentroDeBodSupermercados (m)	Promedio de DistCentroDeBodMetro (m)
0	1385,95	2415,97
1	2174,08	4530,81
2	961,70	924,61
3	1457,53	6745,33
4	994,90	3996,01
5	3135,66	5059,10
6	1431,14	4622,67
7	1047,60	7638,64
Total	1574,39	4659,43

Fuente: Elaboración propia.

Tabla 4: Distancia promedio de multitienda y centro comercial a centro de bodegaje por *cluster*.

Cluster Labels	Promedio de DistCentroDeBodMultitienda (m)	Promedio de DistCentroDeBodCentroComercial (m)
0	2474,43	2898,21
1	3996,75	2468,99
2	2258,73	2315,79
3	1415,00	2321,04
4	3699,31	3795,45
5	4474,65	2984,61
6	3340,32	3745,04
7	5066,23	5246,36
Total	3208,55	3092,22

Fuente: Elaboración propia.

5.2.2 Variables categóricas presentes por *cluster*

A continuación, se presenta el detalle de las variables categóricas que se encuentran consideradas en casa *cluster*.

Tabla 5: Comuna y distrito censal por *cluster*.

Cluster	0	1	2	3	4	5	6	7
ComunaDistrito			CERRILLOS_1			CERRO NAVIA_1		
	COLINA_3				COLINA_3			
	CONCHALI_8				CONCHALI_8			
	LAMPA_4		EST CENTRAL_2		LAMPA_4			
	LAMPA_5			LAMPA_5		LAMPA_5		
			LO ESPEJO_7					
		MACUL_4						
		MACUL_5						
				MAIPU_14				
	MAIPU_5		MAIPU_15					
			MAIPU_5					
	PENALOEN_2		NUNOA_9					
				PUDAHUEL_6				
				PUDAHUEL_7				PUDAHUEL_7
				PUDAHUEL_8	PUDAHUEL_8	PUDAHUEL_8		PUDAHUEL_8
	QUILICURA_1				QUILICURA_1	QUILICURA_1		QUILICURA_1
	QUILICURA_2			QUILICURA_2		QUILICURA_2		QUILICURA_2
	QUILICURA_3				QUILICURA_3	QUILICURA_3		
	QUINTA NORMAL_3							
	RENCA_2							
	RENCA_4							
	RENCA_7					RENCA_7		RENCA_7
	SAN BERNARDO_11		SAN BERNARDO_11					
			SAN BERNARDO_12					
			SAN BERNARDO_6					
			SAN BERNARDO_8					
	SAN JOAQUIN_2		SAN JOAQUIN_2					
	SAN JOAQUIN_3		SAN JOAQUIN_3					
						SANTIAGO_18		

Fuente: Elaboración propia.

Tabla 6: Tipo de Contribuyente presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
TipoContribuyente		INSTITUCIONES FISCALES	INSTITUCIONES FISCALES	INSTITUCIONES FISCALES	INSTITUCIONES FISCALES		INSTITUCIONES FISCALES	
			MUNICIPALIDADES	MUNICIPALIDADES				MUNICIPALIDADES
	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO	ORG. SIN FINES DE LUCRO
	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL	PERSONA JURIDICA COMERCIAL
	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA	SIN PER. JURIDICA
		SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS	SOCIEDADES EXTRANJERAS

Fuente: Elaboración propia.

Tabla 7: Rubro económico presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
RubroEconomico	A1	A1	A1	A1	A1	A1	A1	A1
	A2	A2	A2	A2	A2	A2	A2	A2
	A3	A3	A3	A3	A3	A3	A3	A3
	A4	A4	A4	A4	A4	A4	A4	A4
	A5	A5	A5	A5	A5	A5	A5	A5
	A6	A6	A6	A6	A6	A6	A6	A6
	A7	A7	A7	A7	A7	A7	A7	A7
			A8	A8	A8		A8	A8
	A9	A9	A9	A9	A9	A9	A9	A9
	A10	A10	A10	A10	A10	A10	A10	A10
	A11	A11	A11	A11	A11	A11	A11	A11
	A12	A12	A12	A12	A12	A12	A12	A12
	A13	A13	A13	A13	A13	A13	A13	A13
	A14	A14	A14	A14	A14	A14	A14	A14
	A15	A15	A15	A15	A15	A15	A15	A15
	A16	A16	A16	A16	A16	A16	A16	A16
	A17	A17	A17	A17	A17	A17	A17	A17
	A18	A18	A18	A18	A18	A18	A18	A18
	A19	A19	A19	A19	A19	A19	A19	A19
	A20	A20	A20	A20	A20	A20	A20	A20

Fuente: Elaboración propia.

Tabla 8: Corredor de centro logístico presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
CorredorBodega							Nor Poniente	Nor Poniente
		Norte				Norte		
	Oriente		Oriente					
					Poniente			Poniente
	Sur			Sur				

Fuente: Elaboración propia.

Tabla 9: Tipo de capital propio presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
TramoCapitalPropio	CATORCE	CATORCE	CATORCE	CATORCE	CATORCE	CATORCE	CATORCE	CATORCE
	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO
	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO
	DIECINUEVE	DIECINUEVE	DIECINUEVE	DIECINUEVE	DIECINUEVE	DIECINUEVE	DIECINUEVE	DIECINUEVE
	DIECIOCHO	DIECIOCHO	DIECIOCHO	DIECIOCHO	DIECIOCHO	DIECIOCHO	DIECIOCHO	DIECIOCHO
	DIECISEIS	DIECISEIS	DIECISEIS	DIECISEIS	DIECISEIS	DIECISEIS	DIECISEIS	DIECISEIS
	DIECISIETE	DIECISIETE	DIECISIETE	DIECISIETE	DIECISIETE	DIECISIETE	DIECISIETE	DIECISIETE
	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ
	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE
	DOS	DOS	DOS	DOS	DOS	DOS	DOS	DOS
	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE
	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO
	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE
	QUINCE	QUINCE	QUINCE	QUINCE	QUINCE	QUINCE	QUINCE	QUINCE
	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS
	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE
	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE
	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES
	UNO	UNO	UNO	UNO	UNO	UNO	UNO	UNO
	VEINTE	VEINTE	VEINTE	VEINTE	VEINTE	VEINTE	VEINTE	VEINTE

Fuente: Elaboración propia.

Tabla 10: tramo de facturación presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
TramoFacturacion	UNO	UNO	UNO	UNO	UNO	UNO	UNO	UNO
	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES
	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE	TRECE
	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE	SIETE
	SIES	SIES	SIES	SIES	SIES	SIES	SIES	SIES
	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE	ONCE
	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO
	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE	NUEVE
	DOS	DOS	DOS	DOS	DOS	DOS	DOS	DOS
	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE	DOCE
	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ	DIEZ
	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO
	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO	CINCO

Fuente: Elaboración propia.

Tabla 11: Clase de centro logístico presente por *cluster*.

Cluster	0	1	2	3	4	5	6	7
Clase	DOS					DOS		DOS
		UNO	UNO	UNO	UNO		UNO	

Fuente: Elaboración propia.

6. CONCLUSIONES Y RECOMENDACIONES

Son las conclusiones de la memoria, y las recomendaciones a futuro, ya sea a la empresa donde se ha resuelto el problema, o a personas que quieran continuar la investigación.

7. BIBLIOGRAFÍA

Aldenderfer, M. S. & Blashfield, R. K., 1984. Cluster Analysis. En: M. S. Lewis-Beck, ed. California: Sage Publications, Inc, pp. 07-14.

Boudoin, D., Morel, C. & Gardat, M., 2013. Supply Chains and Urban Logistics Platforms. En: J. Gonzalez-Feliu, F. Semet & J. Routhier, edits. *Sustainable Urban Logistics: Concepts, Methods and Information Systems*. s.l.:Springer, Berlin, Heidelberg, pp. 1-20.

Euclides, E., 2022. Euclides. [En línea]
Available at: <https://euclides.org/distancia/>
[Último acceso: 11 Octubre 2022].

Gajardo Polanco, S., 2019. Región Metropolitana de Santiago análisis de proyecciones de población INE período 2015-2035, Santiago de Chile: Seremi de Desarrollo Social y Familia Metropolitana.
Global Property Solutions, 2021. Centros de Bodegaje, Reporte Global de Mercado, Santiago de Chile: Global Property Solutions.

Gonzalo, Á., 2019. machinelearningparatodos. [En línea]
Available at: <https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/#:~:text=1%2D%20El%20m%C3%A9todo%20del%20codo%3A&text=El%20m%C3%A9todo%20del%20codo%20utiliza,intra%2Dcluster%20tiende%20a%20disminuir.>
[Último acceso: 21 Enero 2023].

Hesse, M. & Rodrigue, J.-P., 2004. The transport geography of logistics and freight distribution. *Journal of Transport Geography*, 12(3), pp. 171-184.

MacCarthy, B. & Atthriawong, W., 2003. Factors effecting location decisions in international operations-A Delphi study. *International Journal of Operations and Production Management*, 23(7).

Meza-Peralta, K., Gonzalez-Feliu, J., Montoya-Torres, J. R. & Khodadat-Saryzady, A., 2020. A unified typology of urban logistics spaces as interfaces for freight transport. *Supply Chain Forum: An International Journal*, 21(4), pp. 274-289.

Python Software Foundation, 2023. *Python*. [En línea]
Available at: <https://www.python.org/>
[Último acceso: 21 Enero 2023].

QGIS.org, 2023. *QGIS*. [En línea]
Available at: <https://www.qgis.org/en/site/>
[Último acceso: 21 Enero 2023].

R Foundation, 2023. *The R Project for Statistical Computing*. [En línea]
Available at: <https://www.r-project.org/>
[Último acceso: 21 Enero 2023].

Rajesh Kr, S., Nikhil, C. & Nikhil, S., 2018. Selection of warehouse location for a global supply chain: A case study. *IIMB Management Review*, 30(4), pp. 343-356.

Sakai, T., Bezait, A. & Heitz, A., 2020. Location factors for logistics facilities: Location choice modeling considering activity categories. *Journal of Transport Geography*, Volumen 85.

Sanggyun, K., 2020. Warehouse location choice: A case study in Los Angeles, CA. *Journal of Transport Geography*, Volumen 88.

Santiago, C. d. C. d., 2020. *Panel semanal comportamiento de consumo por efectos COVID19*, Santiago: Camara de Comercio de Santiago.

Servicio de Impuestos Internos, 2022. *sii.cl*. [En línea]

Available at:

https://www.sii.cl/sobre_el_sii/nominapersonasjuridicas.html#:~:text=1er%20Rango%20Gran%20Empresa%3A%20100.000,1.000.000%2C01%20UF%20Anuales

[Último acceso: 21 Enero 2023].

StataCorp, 2023. Stata. [En línea]

Available at: <https://www.stata.com/>

[Último acceso: 21 Enero 2023].

8. ANEXOS

1. Tabla de variables relevantes a levantar
2. Do-file utilizado en cruce de bases de datos del SII
3. Macro utilizada para el traspaso de coordenadas de múltiples variables
4. Porcentaje de primeras dos corridas de geocoding

Anexo 1. Tabla de variables relevantes a levantar

	Variables	Métricas	La tenem os	Fuente	Descripción
Externas	Tipo de rubro	Categórica (primaria, secundaria, terciaria)	NO	SII/GPS	Es una variable categórica: 1 (Actividades relacionadas con materias primas), 2 (Actividades relacionadas a la producción), 3 (Actividades relacionadas a distribución).
	Conexión a autopistas	Distancia en metros	SI	Elaboración Propia	Es la distancia desde la entrada del centro logístico a la autopista más cercana.
	Cercanía a puerto marítimo	Distancia en metros	SI	Elaboración Propia	Es la distancia desde la entrada del centro logístico al puertos marítimo más cercano.
	Cercanía a aeropuerto	Distancia en metros	SI	Elaboración Propia	Es la distancia desde la entrada del centro logístico al aeropuerto más cercano.
	Vacancia del sector	Tasa de ocupación (m2 desocupados/ m2 del total del proyecto)	NO	GPS	Es la tasa de ocupación en m2 que tienen los distintos proyectos logísticos.
	Conexión para los trabajadores	Distancia en metros al paradero o estación de	SI	Elaboración Propia	Es la distancia desde la entrada del centro logístico al paradero más cercano.

		metro más cercana			
	Costo del suelo	Costo por m2 del proyecto	SI	SII	Es el costo que tiene el suelo total del proyecto.
	Composición física del distrito	m2	NO		m2 por cada distrito.
	Composición sociodemográ fica del distrito		NO		
	Composición de contribuyente s tributarios por distrito		NO		
	Normativa Urbana del distrito	Categórica (tipos de suelos)	SI	Plano regulador de cada comuna	Acá se muestra bajo que plano regulador se rige cada proyecto.
	Tamaño patio del maniobras del proyecto	m2	NO	GPS	Es el tamaño en m2 del patio de maniobras del proyecto
Internas	Altura hombro	metros	NO	GPS	Es la altura de la viga maestra por las bodegas.
	Cannon de arriendo	UF por m2	NO	GPS	Es el precio de arriendo del m2 de bodega en el centro logístico
	Acceso controlado	Binaria	SI	Elaboraci ón Propia	Ver si el centro logístico tiene acceso controlado o no.

Cámaras de seguridad y guardias	Binaria	SI	Elaboración Propia	Ver si el centro logístico tiene cámaras y guardias o no.
Oficinas	Binaria	SI	Elaboración Propia	Ver si el centro logístico tiene oficinas o no.
Casino de alimentación	Binaria	SI	Elaboración Propia	Ver si el centro logístico tiene casinos para alimentación o no.
Arriendo de maquinaria	Binaria	SI	Elaboración Propia	Ver si el centro logístico arrienda maquinaria como brazos articulados y alzadores eléctricos o no.
Arriendo de grúas horquilla	Binaria	SI	Elaboración Propia	Ver si el centro logístico arrienda grúas horquillas o no.
Arriendo de racks	Binaria	SI	Elaboración Propia	Ver si el centro logístico arrienda racks o no.
Obras para clientes	Binaria	SI	Elaboración Propia	Ver si el centro logístico construye oficinas, baños y realiza trabajos en las bodegas o no.
Baños privados	Binaria	SI	Elaboración Propia	Ver si la bodega del centro logístico tiene baño privado o no.
Estacionamiento de camiones	Binaria	SI	Elaboración Propia	Ver si la bodega del centro logístico tiene estacionamiento de camiones o no.
Muelle de carga	Binaria	NO	GPS	Ver si la bodega del centro logístico tiene muelle de carga o no.
Clase	Categorica (1,2)	SI	GPS	Ver de acuerdo al tamaño de las bodega a que clase pertenece Bodegas clase 1 : superficie máxima ilimitada, superficie mínima de 800 m2, altura

					mínima de 8 m. Bodegas clase 2 : superficie máxima de 1.800 m2, superficie mínima de 200 m2. (Global Property Solutions, 2021).
--	--	--	--	--	---

Anexo 2. Do-file utilizado en el cruce de bases de datos del SII

```
*excel pub_empresas_pj
rename aòocomercial ano_comercial
rename razûnsocial razon_social
rename rubroeconûmico rubro_economico
rename subrubroeconûmico sub_rubro_economico
rename actividadeconûmica actividad_economica
rename regiûn region
drop tramosegnventas nmerodetrabajadoresdependientesi fechainiciodeactividadesvigente
fechatèrminodegiro fechaprimerainscripciûndeactivid tipotèrminodegiro tipodecontribuyente
subtipodecontribuyente tramocapitalpropiopositivo tramocapitalpropionegativo

sort region
destring rut, replace
*eliminar primeros 32
drop if rut==0

encode region, gen(region2)
tab region2
tab region2, nolabel
drop if region2==1
drop if region2==2
drop if region2==3
drop if region2==4
drop if region2==5
*drop if region2==6
drop if region2==7
drop if region2==8
drop if region2==9
drop if region2==10
```

```
drop if region2==11
drop if region2==12
drop if region2==13
drop if region2==15
drop if region2==16
drop if region2==17
```

```
destring vigencia, replace
encode vigencia, gen(vigencia2)
label define eti 1 "N" 2 "S"
label values vigencia2 eti
drop if vigencia2==1
```

*unir bases de direcciones

```
append using 22
```

*unir base de contribuyentes con direcciones:

```
destring ano_comercial, replace
```

```
duplicates report rut ano_comercial
duplicates list rut ano_comercial
duplicates tag rut ano_comercial, gen(isdup)
edit if isdup //si queremos editar los datos
drop if isdup==1
drop if isdup==2
drop isdup
```

```
drop region2
```

```
merge 1:m rut ano_comercial using 2122
```

```
destring tipo_direccion, replace  
encode tipo_direccion, gen(tipo_direccion2)  
label define eti2 1 "Domicilio" 2 "Sucursal"  
label values tipo_direccion eti2
```

```
destring ano_comercial, replace
```

```
duplicates report rut  
duplicates list rut  
duplicates tag rut, gen(isdup)  
edit if isdup //si queremos editar los datos  
drop if isdup==1  
drop if isdup==2  
drop isdup
```

```
drop region2
```

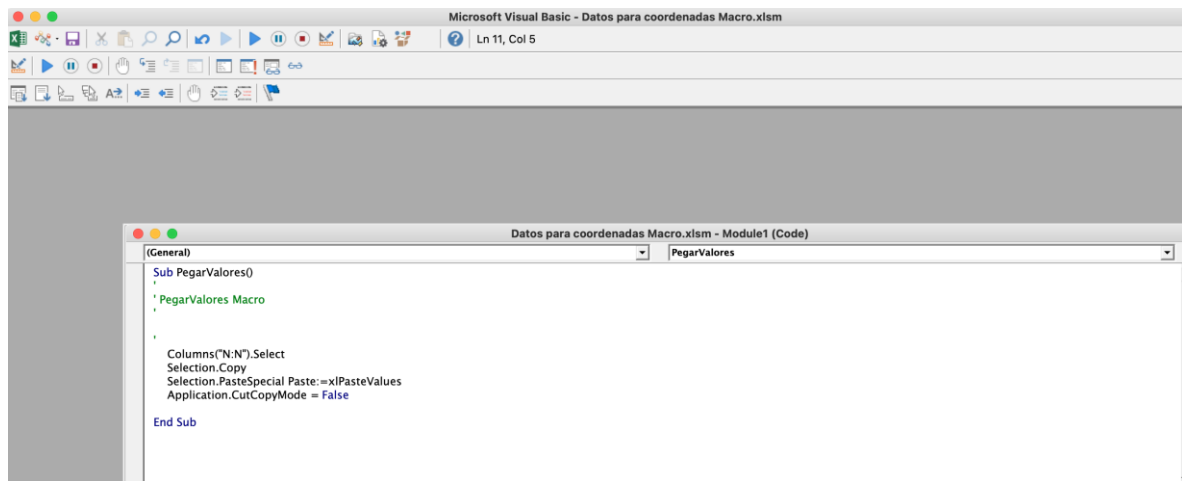
```
merge 1:m rut ano_comercial using 2122
```

Anexo 3: Cantidad de registros por comuna

Comuna	Cantidad de Direcciones
Alhue	146
Buin	3226
Calera de Tango	1008
Cerrillos	2288
Cerro Navia	1444
Colina	5752
Conchalí	2647
Curaca vi	1026
El Bosque	2393
El Monte	609
Estación Central	4115
Huechuraba	5437
Independencia	3136
Isla de Maipo	953
La Cisterna	3489
La Florida	10141
La Granja	1763
La Pantana	2186
La Reina	5640
Lampa	4160
Las Condes	59118
Lo Barnechea	10488
Lo Espejo	1180
Lo Prado	1235
Macul	4090
Maipú	11752
Maria Pinto	434
Melipilla	3772
Ñuñoa	14560
Pedro Aguirre Cerda	1973
Padre Hurtado	1571
Paina	2127
Peña flor	2300
Peñalolén	5379
Pirqué	1015

Providencia	51019
Pudahuel	5301
Puente Alto	10455
Quilicura	6169
Quinta Normal	4375
Recoleta	7366
Renca	2858
San Bernardo	7295
San Joaquin	3090
San Jose de Maipo	633
San Miguel	6246
San Pedro	276
San Ramón	1180
Santiago	54517
Talagante	2170
Til	547
Vitacura	14902
TOTAL	360994

Anexo 4: Macro utilizada para el traspaso de coordenadas de múltiples variables



Anexo 5: Porcentaje de primeras dos corridas de geocoding

1 geocode	# Direcciones encontradas	# Direcciones no encontradas	# Total	% Encontradas
Alhue	25	121	146	17,12328767
Buin	1135	2091	3226	35,18288903
Calera de Tango	152	856	1008	15,07936508
Cerrillos	1154	1134	2288	50,43706294
Cerro Navia	877	1444	1444	60,73407202
Colina	0	5752	5752	0
Conchalí	848	1799	2647	32,03626747
Curaca vi	168	858	1026	16,37426901
El Bosque	1375	1018	2393	57,45925616
El Monte	279	330	609	45,81280788
Estación Central	1299	2816	4115	31,56743621
Huechuraba	3468	1969	5437	63,78517565
Independencia	1536	1600	3136	48,97959184
Isla de Maipo	383	570	953	40,18887723
La Cisterna	2258	1231	3489	64,71768415
La Florida	6060	4081	10141	59,75742037
La Granja	1100	663	1763	62,39364719
La Pantana	511	1675	2186	23,37602928
La Reina	3490	2150	5640	61,87943262
Lampa	1238	2922	4160	29,75961538
Las Condes	34021	25097	59118	57,54761663
Lo Barnechea	5391	5097	10488	51,40160183
Lo Espejo	718	462	1180	60,84745763
Lo Prado	837	398	1235	67,77327935
Macul	749	3341	4090	18,31295844
Maipú	4761	6991	11752	40,51225323
Maria Pinto	83	351	434	19,12442396
Melipilla	1410	2362	3772	37,38069989
Ñuñoa	6312	8248	14560	43,35164835
P Aguirre Cerda	0	1973	1973	0
Padre Hurtado	510	1061	1571	32,46339911
Paina	577	1550	2127	27,1274095
Peña flor	710	1590	2300	30,86956522

Peñalolén	2761	2618	5379	51,32924335
Pirqué	275	740	1015	27,09359606
Providencia	0	51019	51019	0
Pudahuel	0	5301	5301	0
Puente Alto	4181	6274	10455	39,9904352
Quilicura	0	6169	6169	0
Quinta Normal	2031	2344	4375	46,42285714
Recoleta	0	7366	7366	0
Renca	1110	1748	2858	38,8383485
San Bernardo	3583	3712	7295	49,11583276
San Joaquín	1129	1961	3090	36,53721683
San José de Maipo	204	429	633	32,22748815
San Miguel	3795	2451	6246	60,75888569
San Pedro	5	271	276	1,811594203
San Ramón	698	482	1180	59,15254237
Santiago	11124	43393	54517	20,40464442
Talagante	933	1237	2170	42,99539171
Til Til	118	429	547	21,57221207
Vitacura		14902	14902	0
Sin Comuna	0	6	6	0
TOTAL	115382	246453	360994	34,55861866

2 geocode	# Direcciones encontradas	# Direcciones no encontradas	# Total	% Encontradas
Alhue	107	39	146	73,28767123
Buín	2870	356	3226	88,96466212
Calera de Tango	302	706	1008	29,96031746
Cerrillos	2153	135	2288	94,09965035
Cerro Navia	1401	43	1444	97,02216066
Colina	4352	1400	5752	75,66063978
Conchalí	2500	147	2647	94,44654326
Curacaví	712	314	1026	69,3957115
El Bosque	2334	59	2393	97,53447555
El Monte	277	332	609	45,48440066
Estación Central	2996	1119	4115	72,80680437
Huechuraba	5113	324	5437	94,04083134
Independencia	3007	129	3136	95,88647959

Isla de Maipo	888	65	953	93,17943337
La Cisterna	3379	110	3489	96,84723416
La Florida	9931	210	10141	97,9291983
La Granja	1713	50	1763	97,16392513
La Pantana	2126	60	2186	97,25526075
La Reina	4157	1483	5640	73,70567376
Lampa	1221	2939	4160	29,35096154
Las Condes	33961	25157	59118	57,4461247
Lo Barnechea	7783	2705	10488	74,20861937
Lo Espejo	841	339	1180	71,27118644
Lo Prado	957	278	1235	77,48987854
Macul	3424	666	4090	83,71638142
Maipú	8539	3213	11752	72,65997277
Maria Pinto	221	213	434	50,92165899
Melipilla	2297	1475	3772	60,89607635
Ñuñoa	12220	2340	14560	83,92857143
P Aguirre Cerda	1673	300	1973	84,79472884
Padre Hurtado	1436	135	1571	91,40674729
Paina	1072	1055	2127	50,39962388
Peña flor	1616	684	2300	70,26086957
Peñalolén	3961	1418	5379	73,63822272
Pirqué	654	361	1015	64,43349754
Providencia	48806	2213	51019	95,66240028
Pudahuel	4979	322	5301	93,9256744
Puente Alto	10254	201	10455	98,07747489
Quilicura	5891	278	6169	95,49359702
Quinta Normal	3675	700	4375	84
Recoleta	7112	254	7366	96,55172414
Renca	2715	143	2858	94,99650105
San Bernardo	6903	392	7295	94,62645648
San Joaquin	2939	151	3090	95,11326861
San Jose de Maipo	547	86	633	86,41390205
San Miguel	5985	261	6246	95,82132565
San Pedro	172	104	276	62,31884058
San Ramón	1171	9	1180	99,23728814
Santiago	48799	5718	54517	89,51152851
Talagante	1687	483	2170	77,74193548
Til Til	228	319	547	41,68190128

Vitacura	9419	5483	14902	63,20628104
Sin Comuna	4	2	6	66,66666667
TOTAL	293480	67478	360994	79,48190492