

PANEL DE VISUALIZACIÓN WEB AUTOMATIZADO PARA LA TOMA DE DECISIONES DE NEGOCIO EN UNA INSTITUCIÓN DE SALUD

POR: PHIERA CONSTANZA VALENZUELA ESCOBAR

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESORES GUÍA:

Dra. Loreto Bravo Celedón y Dr. Jorge Castillo Sepúlveda

Diciembre 2021
SANTIAGO, CHILE.

RESUMEN

El análisis de datos como metodología de trabajo proporciona una vista integral sobre la situación de una organización, ya sea respecto a la actualidad como respecto al pasado histórico de la misma. El uso de herramientas de gestión y análisis de datos, en conjunto con herramientas de data science, como la automatización y el machine learning, proporcionan una ventaja competitiva al encaminar a todas las áreas de la institución en la misma dirección para cumplir con las metas de la empresa.

El objetivo principal de este trabajo es implementar un sistema de visualización de indicadores para una organización de salud, cuya actualización se realice de manera diaria, automatizada y que pueda ser utilizada para obtener información relevante en la toma de decisiones, el planteamiento de nuevos proyectos y la mejora de la rentabilidad.

El proyecto se desenvuelve dentro del sistema de negocios de una institución clínica, donde se abordan temas como los tipos de servicios de la clínica en cuestión, los indicadores que son utilizados actualmente por sus áreas de gestión, la implementación de un servicio web y la metodología aplicada para desarrollar un indicador basado en machine learning.

Dentro de los resultados se muestran las visualizaciones utilizadas para representar cada uno de los indicadores descritos, se presenta el servicio web puesto en marcha, y se entregan los resultados del modelo de machine learning para segmentación de pacientes. Como resultado de este indicador se logra identificar que el grupo de pacientes que genera mayores ingresos está compuesto principalmente por pacientes de edad avanzada, que suelen asistir a la clínica para atención de consultas médicas y derivan en alguna prestación quirúrgica de alta complejidad.

INDICE DE CONTENIDOS

1.	INTRODUCCIÓN	1
2.	HIPÓTESIS Y OBJETIVOS	3
2.1.	HIPÓTESIS	3
2.2.	OBJETIVO GENERAL.....	3
2.3.	OBJETIVOS ESPECÍFICOS	3
3.	MARCO TEÓRICO:.....	4
3.1.	SOFTWARE EMPRESARIAL O DE NEGOCIOS.....	4
3.1.1.	<i>Origen de los Dashboards:</i>	<i>5</i>
3.1.2.	<i>Tipos de dashboards y Métricas de desempeño.....</i>	<i>6</i>
3.2.	UTILIZACIÓN DE DASHBOARDS EN UNA INSTITUCIÓN DE SALUD	8
3.2.1.	<i>Conceptos del negocio:.....</i>	<i>8</i>
3.2.2.	<i>Descripción de la Base de Datos disponible</i>	<i>12</i>
4.	METODOLOGÍA:	14
4.1.	DESCRIPCIÓN DE INDICADORES:	14
4.1.1.	<i>Reporte de Ventas.....</i>	<i>14</i>
4.1.2.	<i>Cantidad de Pacientes Atendidos.....</i>	<i>17</i>
4.1.3.	<i>Distribución de Venta por Aseguradora</i>	<i>18</i>
4.1.4.	<i>Demografía de Pacientes</i>	<i>19</i>
4.2.	IMPLEMENTACIÓN DE LA PLATAFORMA WEB:.....	20
4.2.1.	<i>Adquisición de la base y preparación de los datos:.....</i>	<i>20</i>
4.2.2.	<i>Levantamiento del Servicio Web:</i>	<i>21</i>
4.2.3.	<i>Diseño del Front End de la Página Web:</i>	<i>23</i>
4.3.	MODELO DE SEGMENTACIÓN DE PACIENTES:	25
4.3.1.	<i>Descripción del modelo:</i>	<i>25</i>

4.3.2.	<i>Datos de entrada al modelo:</i>	26
4.3.3.	<i>Reducción de Dimensionalidad:</i>	27
4.3.4.	<i>Segmentación de pacientes:</i>	29
5.	PRESENTACION Y ANÁLISIS DE RESULTADOS	31
5.1.	VISUALIZACIÓN Y ANÁLISIS DE INDICADORES	31
5.1.1.	<i>Reporte de Ventas Semanal</i>	31
5.1.2.	<i>Cantidad de Pacientes Atendidos</i>	35
5.1.3.	<i>Distribución de Venta por Asegurador</i>	37
5.1.4.	<i>Demografía de Pacientes</i>	40
5.2.	PLATAFORMA WEB IMPLEMENTADA:	44
5.2.1.	<i>Reporte de Ventas Semanal:</i>	44
5.2.2.	<i>Reporte de Ventas por Isapre:</i>	45
5.2.3.	<i>Reporte de Cantidad de Pacientes:</i>	46
5.2.4.	<i>Reporte Demográfico:</i>	46
5.2.5.	<i>Reporte de Segmentación de Pacientes:</i>	47
5.3.	MODELO DE SEGMENTACIÓN DE PACIENTES:	47
5.3.1.	<i>Reducción de dimensionalidad por FAMD:</i>	48
5.3.2.	<i>Segmentación de pacientes por Kmeans:</i>	50
5.3.3.	<i>Descripción de clusters de pacientes:</i>	52
6.	CONCLUSIONES	56
6.1.	CONCLUSIONES	56
6.2.	LIMITACIONES:	58
6.3.	TRABAJO A FUTURO:	59
7.	BIBLIOGRAFÍA	60

INDICE DE FIGURAS

FIGURA 1: CONEXIÓN A BASE DE DATOS ORACLE MEDIANTE EL IDE JUPYTER NOTEBOOK.....	20
FIGURA 2: DESCARGA DE TABLAS MEDIANTE CONSULTAS SQL EN JUPYTER NOTEBOOK.	21
FIGURA 3: CREACIÓN DEL ENDPOINT “HOME” MEDIANTE FASTAPI EN JUPYTER NOTEBOOK.....	22
FIGURA 4: LEVANTAMIENTO DEL SERVICIO WEB A UN SERVIDOR LOCAL MEDIANTE UVICORN	22
FIGURA 5: EJEMPLO DE PÁGINA WEB SENCILLA IMPLEMENTADA EN SERVIDOR LOCAL.	22
FIGURA 6: CARGA DEL TEMPLATE “PAGE.HTML” AL ENDPOINT “HOME” MEDIANTE JINJA2TEMPLATE.....	23
FIGURA 7: TEMPLATE BASE CARGADO EN LA PÁGINA WEB, QUE CUENTA CON BARRA DE NAVEGACIÓN INTERACTIVA.	24
FIGURA 8: TABLA DE DATOS DE ENTRADA PARA EL MODELO DE MACHINE LEARNING NO-SUPERVISADO.	26
FIGURA 9: CÓDIGO PARA OBTENER DOS COMPONENTES PRINCIPALES MEDIANTE LA FUNCIÓN PRINCE.FAMD.....	28
FIGURA 10: CÓDIGO PARA OBTENER LA VARIANZA EXPLICADA DE UN SISTEMA DE 7 VARIABLES DE ENTRADA.....	28
FIGURA 11: CÓDIGO PARA REALIZAR CLUSTERIZACIÓN MEDIANTE KMEANS CON K = 5.....	30
FIGURA 12: EJEMPLO DE MÉTODO DEL CODO CON RESULTADOS PARA K = 3. [24]	30
FIGURA 13: REPORTE DE VENTAS SEMANAL AÑO 2021.	31
FIGURA 14: VENTA SEMANAL NORMALIZADA ENTRE 0 Y 1 PARA CADA LÍNEA DE NEGOCIO.	33
FIGURA 15: REPORTE DE VENTAS SEMANAL AÑOS 2019, 2020 Y 2021.	34
FIGURA 16: A) CANTIDAD DE PACIENTES ATENDIDOS EN EL AÑO 2021 DE MANERA SEMANAL. B) VENTA PROMEDIO DE PACIENTES ATENDIDOS EN EL AÑO 2021 DE MANERA SEMANAL	36
FIGURA 17: DISTRIBUCIÓN DE VENTA POR ASEGURADOR EN EL AÑO 2021.	37
FIGURA 18: DISTRIBUCIÓN DE VENTA POR ISAPRES EN EL AÑO.....	38
FIGURA 19: EVOLUCIÓN MENSUAL DE DISTRIBUCIÓN PORCENTUAL DE VENTA POR ASEGURADORA.....	39
FIGURA 20: DISTRIBUCIÓN DE VENTA POR EDAD DE PACIENTES AÑO 2021.	41
FIGURA 21: DISTRIBUCIÓN GEOGRÁFICA DE PACIENTES ATENDIDOS EN EL AÑO 2021.	43
FIGURA 22: DASHBOARD DEL REPORTE DE VENTAS IMPLEMENTADO EN EL SERVICIO WEB.	44
FIGURA 23: DASHBOARD DEL REPORTE DE VENTAS POR ASEGURADORA IMPLEMENTADO EN EL SERVICIO WEB.	45

FIGURA 24: DASHBOARD DEL REPORTE DE CANTIDAD DE PACIENTES IMPLEMENTADO EN EL SERVICIO WEB.....	46
FIGURA 25: DASHBOARD DEL REPORTE DEMOGRÁFICO IMPLEMENTADO EN EL SERVICIO WEB.....	47
FIGURA 26: A) VARIANZA EXPLICADA PARA CADA COMPONENTE PRINCIPAL. B) VARIANZA TOTAL ACUMULADA.	48
FIGURA 27: REPRESENTACIÓN 2D DE LOS DOS COMPONENTES PRINCIPALES SELECCIONADOS.	49
FIGURA 28: TÉCNICA DEL MÉTODO DEL CODO APLICADO AL PROYECTO.	50
FIGURA 29: CLUSTERIZACIÓN DE DATOS POR KMEANS CON CANTIDAD DE CLUSTERS IGUAL A 5.	51

INDICE DE TABLAS

TABLA 1: AGRUPACIÓN DE PACIENTES POR RANGO ETARIO.	19
TABLA 2: PORCENTAJE DE ATENCIONES, DE VENTA Y FACTOR DE RENTABILIDAD POR RANGO ETARIO.	42
TABLA 3: VALORES PROMEDIOS PARA EL GRUPO 0.	52
TABLA 4: DISTRIBUCIÓN DE ASEGURADORAS DE SALUD EN EL GRUPO 0.	52
TABLA 5: DISTRIBUCIÓN DE SEXO EN EL GRUPO 0.....	53
TABLA 6: VALORES PROMEDIOS PARA EL GRUPO 1.	53
TABLA 7: DISTRIBUCIÓN DE ASEGURADORAS DE SALUD EN EL GRUPO 1.	53
TABLA 8: DISTRIBUCIÓN DE SEXO EN EL GRUPO 1.....	53
TABLA 9: VALORES PROMEDIOS PARA EL GRUPO 2.	54
TABLA 10: DISTRIBUCIÓN DE ASEGURADORAS DE SALUD EN EL GRUPO 2.	54
TABLA 11: DISTRIBUCIÓN DE SEXO EN EL GRUPO 2.....	54
TABLA 12: VALORES PROMEDIOS PARA EL GRUPO 3.	54
TABLA 13: DISTRIBUCIÓN DE ASEGURADORAS DE SALUD EN EL GRUPO 3.	54
TABLA 14: DISTRIBUCIÓN DE SEXO EN EL GRUPO 3.....	54
TABLA 15: VALORES PROMEDIOS PARA EL GRUPO 4.	55
TABLA 16: DISTRIBUCIÓN DE ASEGURADORAS DE SALUD EN EL GRUPO 4.	55
TABLA 17: DISTRIBUCIÓN DE SEXO EN EL GRUPO 4.....	55

1. INTRODUCCIÓN

El análisis de datos y el big data se han convertido en una necesidad competitiva a la hora de tomar decisiones en industrias y negocios de toda índole. Su objetivo es dotar de conocimiento a las organizaciones y transformar la información en oportunidades para el negocio, ya sea para nuevos desarrollos o para mejorar el rendimiento actual de la institución [1].

Una organización que pertenece a la industria de la salud suele contar con una gran cantidad de datos, puesto que contiene información clínica, comercial, operacional, farmacéutica y transaccional de los pacientes [2]. Manejar esta cantidad de información podría significar grandes avances a nivel clínico, por ejemplo, el poder determinar de manera más acertada el diagnóstico y el tratamiento de cada paciente mediante el uso de herramientas basadas en machine learning como apoyo a los profesionales clínicos [2]. Por otro lado, el análisis de datos en conjunto con otras tecnologías de la información también puede ser empleado para apoyar a áreas de gestión y negocios de una clínica. Algunos de estos sistemas podrían ser, por ejemplo, sistemas que segmentan pacientes, sistemas que entregan información sobre rentabilidad, sistemas para el control de abastecimiento y sistemas de control sobre los profesionales clínicos[2].

Las herramientas de gestión y análisis de datos pueden utilizarse en distintos rubros de la industria, más aún en empresas del ámbito privado. En el rubro de la salud en particular, suele presentarse mayoritariamente en clínicas y aseguradoras privadas, lo cual se debe principalmente al interés de estas instituciones en mejorar la rentabilidad, además de que poseen mayores recursos como para invertir en nuevos sistemas de análisis y gestión.

A partir de esta necesidad se comienzan a implementar paneles de control y visualización de indicadores, los cuales aportan información a distintas áreas de una organización y ayudan a mejorar el control y eficiencia de cualquier actividad. Este tipo de herramientas en conjunto con un sistema de automatización podría significar una gran ventaja, logrando encaminar a todas las áreas de la empresa a una misma dirección para cumplir con los objetivos de la empresa [3].

Por esta razón es que se propone un sistema de visualización de indicadores para una organización de salud, que facilite la obtención de información en la gestión y toma de decisiones, y que además se caracterice por ser un sistema automatizado.

Por ello, en el presente proyecto, se describirá paso a paso como llevar a cabo la implementación del sistema de visualización. En primera instancia, se darán a conocer los indicadores utilizados por el área comercial y de control de gestión de la institución, los cuales se analizarán y representarán mediante visualizaciones que sean fácil entendimiento para los usuarios. Los datos a utilizar en estos dashboards se adquirirán de la base de datos Oracle de la institución clínica, la cual se tratará mediante lenguaje de programación en Python para generar los gráficos de indicadores. A continuación, se levantará un servicio web mediante el framework FastAPI, con el que se crearán las distintas páginas de la web que contendrá cada visualización previamente creada. Y finalmente, se implementará un modelo de machine learning no-supervisado para agrupar los distintos tipos de paciente, y poder descifrar cuál de los grupos generados es el que genera mayores ingresos. Este será el último de los indicadores que será añadido al servicio web ya implementado.

2. HIPÓTESIS Y OBJETIVOS

2.1. Hipótesis

El uso de cuadros de mando integral en conjunto con la ciencia de datos, son una potente herramienta que puede facilitar la obtención de información, mejorar el análisis de indicadores y apoyar a la toma de decisiones.

2.2. Objetivo General

Implementar un panel de visualización web que, mediante el análisis de datos, la automatización y el machine learning, facilite la obtención de información para medir, rectificar o potenciar a las distintas áreas de una institución de salud.

2.3. Objetivos Específicos

- Describir los indicadores predefinidos por las áreas comercial y control de gestión que se implementarán en el panel de visualización web.
- Diseñar una plataforma web a nivel técnico (lenguajes, tecnologías) y de usuario (visualización, funcionalidades) que permita presentar las visualizaciones obtenidas.
- Implementar un modelo de clusterización basado en machine learning, para agrupar e identificar a los pacientes que generan mayor rentabilidad, y que pueda incorporarse al panel web como un nuevo indicador.

3. MARCO TEÓRICO:

3.1. Software empresarial o de negocios

En la mayoría de las organizaciones suele contarse con múltiples unidades de negocio, divisiones o departamentos, donde cada uno de ellos se caracteriza por sus propios procesos, estrategias y sistemas. En consecuencia, suelen producir redundancia e incluso conflictos en la información que ha sido generada, dando como resultado, una organización que “gira” en múltiples direcciones sin una estrategia clara. [3]

Con el fin de lograr el enfoque de las distintas áreas de una empresa, se incentiva la necesidad de un sistema de información que mejore la comunicación entre ellas, informando de forma concisa las estrategias de la empresa y entregando objetivos clave de manera individualizada. Las medidas de desempeño en conjunto con retribuciones por el buen comportamiento y el reforzamiento necesario encaminarán a los empleados a lograr las metas y estrategias de la organización. [3]

Con este objetivo en mente, se crea la idea del software empresarial o de negocios, que corresponde a cualquier tipo de software diseñado especialmente para ofrecer apoyo a una organización, midiendo y mejorando su productividad. Este término abarca una extensa variedad de aplicaciones informáticas como por ejemplo programas de contabilidad y ofimática, sistemas de planificación de recursos empresariales (*ERP: Enterprise Resource Planning*), programas de gestión de clientes (*CRM: Customer Relationship Management*), programas de administración de cadena de suministros (*SCM: Supply Chain Management*), y muchos otros que brindan soporte a las empresas. [4]

Lo que propone [3] es que un software empresarial, como los paneles de desempeño, deben contar con las siguientes características:

- **Monitorear** los procesos y actividades críticas utilizando métricas de rendimiento del negocio que active alertas cuando surjan problemas.
- **Analizar** la causa raíz de los problemas mediante la exploración de información relevante y oportuna desde las múltiples perspectivas y en varios niveles de detalle.
- **Gestionar** personas y procesos para mejorar la toma de decisiones, optimizar el rendimiento y orientar a la organización a la dirección correcta.

Un panel de desempeño implementado de buena manera será una poderosa herramienta que puede lograr transformar una organización de bajo rendimiento en una de alto rendimiento.

3.1.1. Origen de los Dashboards:

Los paneles o tableros de desempeño se han utilizado desde hace mucho tiempo en algunos tipos de industria como por ejemplo en automóviles y otros vehículos, sin embargo, en la mayoría de las industrias la tendencia comenzó alrededor de la década de 1980, con la idea de tener un tablero ejecutivo con el que conducir las empresas desde una sola sala de reuniones [3].

Durante la década de 1990, las organizaciones comenzaron a evaluar diferentes formas para entregar información crítica de manera directa y oportuna a los usuarios del área de negocios, lo que fue llamado como Inteligencia de Negocios [3].

En el año 1992, David Norton y Robert Kaplan expusieron la metodología de Cuadro de Mando Integral en un artículo para el Harvard Business Review, donde se estudiaba la medición del desempeño en organizaciones cuyos activos intangibles eran un factor esencial en la ecuación

de la creación de valor. En esta nueva metodología se propuso analizar al menos cuatro perspectivas para abarcar todo el espectro de indicadores de la empresa: perspectiva financiera, perspectiva cliente, perspectiva interna y perspectiva de aprendizaje [5].

Ya en el 2004, en un estudio realizado por The Data Warehousing Institute (*TDWI*), se mostró que el 51% de las organizaciones utilizaban un cuadro de mando integral y que otro 17% se encontraba en procesos de diseño. Este estudio indicó también, que casi un tercio de las organizaciones que ya utilizan un cuadro de mando, lo utiliza como su principal aplicación a la hora de realizar informes y análisis de datos [3].

En la actualidad, son aún más las empresas que utilizan cuadros de mando integral (*CMI*) basados en la propuesta de Kaplan y Norton, los cuales se identifican mayoritariamente con el nombre de *Dashboards* [5].

3.1.2. Tipos de dashboards y Métricas de desempeño

Como se menciona anteriormente, la mayoría de las organizaciones ya han implementado un sistema de gestión de algún tipo, y un tercio de ellos lo utiliza como su principal método analítico. Esto abarca una gran gama de rubros y posibilidades de aplicación, por lo que [3] propone una clasificación según el rol que desempeñan en una empresa:

- **Dashboards Operativos:** Permiten a los trabajadores realizar un seguimiento de los procesos operativos básicos y enfatizan las funciones de monitoreo más que el análisis o la gestión. Suele utilizarse para tomar decisiones a corto plazo, a menudo reactivas, y que pueden tener un impacto sobre las operaciones diarias.

- **Dashboards Tácticos:** Ayudan a los gerentes y analistas a rastrear y estudiar la información, ya sea de actividades, procesos o proyectos departamentales. Se enfatizan las funciones de analítica, investigación y estudio de tendencias. Suele utilizarse con un enfoque proactivo y pueden tener un impacto en decisiones a mediano plazo.
- **Dashboards Estratégicos:** Apoyan a la dirección ejecutiva de una empresa a medir su progreso en comparación al presupuesto y objetivos establecidos. Se enfatizan las funciones de gestión, y se incentiva la colaboración entre gerentes y personal de distintas unidades y departamentos. Su objetivo es servir como apoyo para la toma de decisiones a largo plazo, respecto a las metas y objetivos de la empresa.

El tipo de dashboard a utilizar dependerá exclusivamente del rol al que se quiera enfocar y de quienes serán los usuarios finales. Por otro lado, para implementar un cuadro de mando también es necesario determinar las métricas o indicadores a analizar (*KPI: Key Performance Indicator*). Sin embargo, ¿Cómo se definen las métricas más efectivas? Esta respuesta es importante, dado que serán los indicadores quienes enfocarán de mejor manera a los trabajadores a realizar las tareas que los ejecutivos consideren más importantes para el éxito del negocio. [3]

A pesar de que la selección de indicadores es de prioridad durante la creación de un dashboard, suelen ser definidos mayoritariamente por el directorio ejecutivo de una empresa, los cuales a la par, se encargan de definir metas y objetivos.

3.2. Utilización de Dashboards en una Institución de Salud

El objetivo de este proyecto es implementar un cuadro de mando integral en un servicio web, que tenga la capacidad de actualizarse de manera diaria y que sirva como apoyo para la toma de decisiones de una institución de salud. El rol de este cuadro de mando integral será del tipo táctico, el cual estará compuesto por distintos indicadores que ya han sido definidos por las áreas comercial y de control de gestión de la empresa. Para comenzar con la implementación del dashboard, en primera instancia, se deberá entender cómo funciona el sistema de negocios e ingresos de esta institución clínica de alta complejidad.

3.2.1. Conceptos del negocio:

A grandes rasgos, este sistema está compuesto por tres líneas de negocio: servicio ambulatorio, servicio hospitalario y servicio de urgencias, donde cada una de ellas, genera ingresos, costos y tiene sus propios criterios para ser analizados:

a) Línea de negocio ambulatorio:

Se considera ambulatorio a todas las atenciones que no requieren de internación clínica o cuya atención es diurna. Un paciente ambulatorio puede acudir a un centro de salud, llevar a cabo las sesiones correspondientes al tratamiento y luego regresar a su hogar. Dentro de estas atenciones se consideran las consultas médicas de los distintos especialistas, exámenes de laboratorio, exámenes de imágenes, procedimientos de baja complejidad como biopsias y medicina transfusional, e incluso, atenciones domiciliarias [6].

b) Línea de negocio hospitalario:

Se considera hospitalario a la atención de un paciente que requiere de internación clínica o pernocte en la institución. Las causas de internación clínica pueden ser variadas, como por ejemplo, una derivación luego de realizada una intervención quirúrgica que no es ambulatoria y que requiere reposo, una consulta de urgencia que derivó en hospitalización debido a su alta complejidad, y algunas hospitalizaciones que se requieren para casos específicos como oncología y cuidados paliativos. Además, se deben considerar hospitalarias, todas las prestaciones que se hayan realizado dentro de la estadía del paciente en hospitalización, como por ejemplo, exámenes de imagen y laboratorio, consultas de kinesiólogía y psicología, etc.

c) Línea de negocio urgencias:

Se considera urgencias a la atención de un paciente que se presenta a la institución con síntomas de riesgo vital o secuelas funcionales graves, las cuales deben ser atendidas de manera inmediata e impostergable [7]. Las atenciones por urgencias suelen tener un carácter inicial como ambulatorio, ya que se suelen realizar prestaciones tales como consultas médicas, imágenes y exámenes de laboratorio. En el caso de que la complejidad de la urgencia requiera alguna intervención quirúrgica u hospitalización, las prestaciones realizadas pierden el carácter de ambulatorio y se ingresan como prestaciones dentro de un ingreso hospitalario.

Algunos de los indicadores que se utilizarán para analizar el negocio, pueden ser aplicados de manera transversal a los tres servicios, incluso considerando que necesitan diferentes criterios para su conteo y análisis. Por otro lado, habrá algunos casos donde los indicadores serán aplicados de manera específica a una de las tres áreas, como por ejemplo sería el conteo de cirugías realizadas en hospitalización y derivaciones de urgencias, criterio que no puede ser evaluado en ambulatorio.

Otro indicador que será evaluado debido a su impacto comercial corresponde a la influencia de las aseguradoras en la institución clínica. Las aseguradoras de salud son entidades encargadas de financiar las atenciones y beneficios en salud a sus afiliados, mediante un sistema de seguros que se paga con una prima mensual.

Para el caso de FONASA (*Fondo Nacional de Salud*), que corresponde a la aseguradora de salud estatal, el valor de la prima corresponderá al monto de la cotización obligatoria, es decir, el 7% del sueldo imponible. Por otro lado, la bonificación de las prestaciones dependerá del tramo de riesgo social en el que se encuentra la persona afiliada, variando entre el 80% y el 100% de bonificado [8]. En el caso de las ISAPRES (*Instituciones de Salud Previsional*), que corresponden a entidades privadas, la prima mensual corresponderá como mínimo al monto de la cotización obligatoria, es decir, un 7% hacia arriba. Esto se debe a que los planes de salud privados son variados, con valores elevados y con diferentes porcentajes de bonificación que dependen del gusto de cada consumidor [9].

A pesar de que los planes de seguro de salud son contratados por cada paciente de manera individual, son las aseguradoras las que realizan convenios con las clínicas, generando instancias para que sus afiliados sean derivados hacia estas mismas. Algunas de estas instancias

pueden ser descuentos especiales para ciertas atenciones, o incluso, derivaciones de manera directa como ocurre en el caso de las GES (*Garantías Explícitas en Salud*) [10], donde las aseguradoras le indican a cada paciente donde deben atender su patología de manera obligatoria. Debido a lo anterior, es de gran importancia para una institución clínica mantener buenos convenios con las aseguradoras, e idealmente convertirse en prestadores preferentes para obtener la mayor cantidad de derivaciones de pacientes.

Finalmente, y de manera similar a como se analiza la rentabilidad por cada aseguradora, es importante evaluar la rentabilidad por tipo de paciente, para lo cual suelen influir variables del tipo demográficas. La demografía es una ciencia que estudia estadísticamente las características de una población para clasificarlas de distintas maneras, como puede ser la edad, el sexo, el lugar donde residen, condiciones laborales, ingreso per cápita y otros [11].

Esta información puede ser de gran utilidad al hacer un estudio de mercado sobre cuáles son los tipos de pacientes que generan mayores ingresos. Al definir un mercado objetivo, se podrán implementar proyectos y campañas dirigidas a estos grupos que generan mayor rentabilidad, y como consecuencia, aumentar los ingresos. Esta información también servirá para identificar cuáles son las áreas clínicas de una institución que reciben menor cantidad de pacientes o que generan menores ingresos. Luego, se podrá evaluar cada caso para determinar si lo más conveniente sea dejar de invertir recursos en estas áreas para obtener los mejores resultados, por ejemplo.

Son variadas las áreas de estudio en una institución clínica, sin embargo, para efectos de este trabajo solo se abordará el análisis de ventas a nivel de las tres líneas de negocio: ambulatorio, hospitalario y urgencias, además del análisis de pacientes y de aseguradoras.

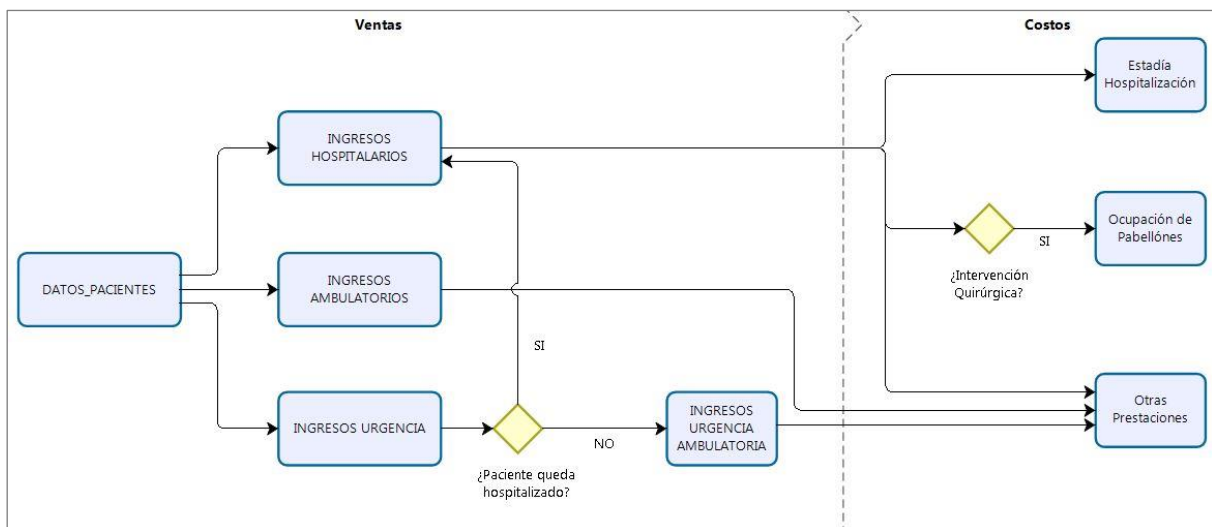
3.2.2. Descripción de la Base de Datos disponible

Para alimentar los dashboards de indicadores que se quieren implementar, se utilizará la base de datos de la institución de salud. Esta está compuesta por múltiples tablas de datos que funcionan en base a un modelo relacional:

- **Base de Datos de Pacientes:** Base de datos con las características demográficas de cada paciente como edad, sexo y lugar de residencia. Cuenta con información del asegurador y planes de salud, de gran importancia si se quiere realizar un análisis comercial.
- **Devengado Ingresos Hospitalarias:** Base de datos transaccional donde aparecen todos los consumos asociados a un ingreso de paciente hospitalario. Estos consumos pueden ser prestaciones quirúrgicas, prestaciones ambulatorias, utilización de camas y consumo de medicamentos e insumos, cada uno de ellos con su precio de venta y fecha de consumo.
- **Base Prestaciones Ambulatorias:** Base de datos transaccional que cuenta con todas las prestaciones realizadas por un paciente que asiste al servicio ambulatorio de la clínica como por ejemplo consultas médicas, exámenes de imagen, laboratorio, kinesiología, psicología y algunos procedimientos de baja complejidad que no requieren hospitalización.
- **Base Quirúrgica:** Base de datos de cada prestación quirúrgica que se realiza en un pabellón. Contiene información de la hora de ingreso y egreso de pabellón, la prestación realizada y los clínicos participantes: cirujanos, arsenaleros y anestesiistas.

- **Base de Estadía:** Base de datos detallada de la estadía del paciente. Cuenta con información sobre todas las camas en las que estuvo un paciente, los traslados realizados, fecha de inicio y término de estadía en una cama, y servicio/unidad clínica en donde se ubicaba.

Cabe destacar que, las bases de datos mencionadas son una forma simplificada de describir los datos existentes, como se muestra en el Esquema 1. En la realidad, el sistema cuenta con más de 100 tablas relacionadas entre sí, que comunican las áreas de operaciones, clínica y financiera, para entregar resultados aterrizados y entendibles. Esto se realiza de manera interna mediante el HIS (*Sistema de Información Hospitalario*) de la institución, sistema utilizado por los funcionarios para ingresar toda la información referente a las atenciones clínicas, y que luego será descargada por los analistas e informáticos para transformarla en un modelo de datos de fácil entendimiento.



Esquema 1: Modelo simplificado del sistema de datos de la institución clínica.

4. METODOLOGÍA:

La metodología de este proyecto se desarrollará acorde a los objetivos específicos definidos. El primero de ellos corresponde a la descripción de indicadores utilizados por la institución clínica. Estos se describirán y representarán mediante distintas visualizaciones que se mostrarán en el servicio web.

A partir de la descripción de indicadores, se presentará la forma de implementación del servicio web. En este capítulo, se describe el modo de adquisición de las bases y el lenguaje de programación utilizado para preparar los datos y visualizaciones. Dentro de esta sección, además, se presentará la metodología asociada a la creación de página web y al levantamiento del servicio local, que implica conocimientos de lenguaje de programación en HTML.

Finalmente, y de acuerdo con el tercer objetivo específico definido, se presentará el último indicador del servicio web. Este corresponde a un clusterizador de pacientes basado en un modelo de machine learning no supervisado, con el que se espera identificar cuáles son los grupos de pacientes que generan mayores ingresos.

4.1. Descripción de indicadores:

4.1.1. Reporte de Ventas

Como primera instancia y debido a su importancia a nivel gerencial, se replicará el Reporte de Ventas de la institución. El reporte de ventas es un informe que le permite a las empresas tener una visión más completa de sus actividades comerciales. Su objetivo es evaluar periódicamente

la evolución del estado de la organización para encaminar a la empresa a tomar mejores decisiones, e incluso, prever situaciones mediante el análisis de reportes históricos [12].

En la organización en estudio, el reporte de ventas suele emitirse de manera semanal, donde se entrega información sobre la evolución de la venta semanal y mensual. Para el caso del mes en curso, se entrega además, un estimativo del cierre de mes que se calcula principalmente en base a fórmulas lineales. Finalmente, éste se compara con la venta presupuestada y se sacan conclusiones dependiendo si se alcanzó el presupuesto o no.

Este reporte suele estar afecto a una gran cantidad de errores debido a que depende netamente de la persona a cargo de realizar los pasos del proceso, que suelen realizarse mediante Excel y un motor de bases de datos Oracle. Por esta razón, el objetivo principal de este proyecto es disponer de un reporte en la web, cuya actualización se realice de manera diaria y automatizada, con el fin de optimizar los resultados y reutilizar las horas hombre en nuevos proyectos.

En el reporte de ventas de la institución se analizan las tres áreas de negocio: ambulatorio, hospitalario y urgencias, donde, para cada una de ellas, se estima la venta semanal, mensual y se compara con el presupuesto. A pesar de esto, en el presente proyecto, solo se mostrarán los ingresos de manera semanal, sin realizar una comparación con el presupuesto.

a) Línea de negocio ambulatorio:

Para esta línea de negocios, la transacción o venta asociada a una atención, se contabiliza durante el mismo día de realizada la prestación, y se puede ver reflejada en la base de datos de prestaciones ambulatorias. Esta base cuenta con cada código de prestación realizada, la fecha de atención, la cantidad, el monto y el identificador del paciente.

b) Línea de negocio hospitalario:

Para este caso, las transacciones asociadas a un ingreso o internación hospitalaria se ven reflejadas en el devengado productivo de la institución, puesto que la fecha de pago no corresponde a la fecha de consumo o de realización de las prestaciones. Esto se debe a que ciertos factores se deben cumplir antes de liquidar una cuenta hospitalaria: primero, el paciente debe estar dado de alta para cerrar una cuenta con todas las atenciones que este haya consumido, y segundo, luego de confeccionada la cuenta por parte de la institución de salud, ésta debe ser enviada al asegurador del paciente para su evaluación y aceptación conforme de lo cobrado.

Dado lo anterior, la base de datos utilizada para contabilidad la venta de atenciones hospitalarias corresponde al devengo productivo de la institución, la cual cuenta con los días camas que estuvo el paciente internado, las cirugías realizadas, los medicamentos e insumos consumidos y las prestaciones realizadas, cada una con la fecha en que se realizó, la cantidad, el valor de la venta y el identificador del paciente que se atendió.

c) Línea de negocio urgencias:

Las atenciones por urgencias suelen tener un carácter inicial como ambulatorio, por lo que las atenciones suelen verse reflejadas dentro de la base de datos de prestaciones ambulatorias, con un respectivo identificador de urgencias que las diferencien. Cuando el paciente de urgencias es derivado a hospitalización, las prestaciones realizadas pierden el carácter de ambulatorio y se ingresan como prestaciones dentro de un ingreso hospitalario, es decir, que se verán reflejadas en el devengado.

Lograr definir los costos de las prestaciones tiene una mayor complejidad, debido a que depende del tipo de prestación y no del área de negocio involucrada. Algunas de ellas, se costean de forma directa, como por ejemplo los exámenes de imagen y laboratorio, cuyo costo está previamente definido. Otras prestaciones dependerán de la estadía del paciente, como por ejemplo, el día cama de hospitalización y el uso de pabellones en cirugía, cuya contabilización es a partir de las horas de uso. Y así sucesivamente, existen diferentes criterios de definición de costos para las múltiples prestaciones que realiza la clínica.

Con esta información, junto a los precios de venta que contienen las demás bases de datos, es posible establecer los márgenes por cada una de las áreas clínicas y prestaciones. Sin embargo, debido a la complejidad de definición de costos de las prestaciones, solo se analizará la venta para cada línea de negocios.

4.1.2. Cantidad de Pacientes Atendidos

A partir de las bases de datos con las que se obtiene el reporte de ventas, ya sea ambulatorio, hospitalario o de urgencias, es posible obtener la cantidad de pacientes que se atendieron.

Dado que un paciente puede realizarse más de una prestación ambulatoria durante el mismo día, como suele suceder el caso de una consulta médica en conjunto con exámenes de imágenes, se contará como un único paciente atendido en el día. Luego de identificados los Rut de pacientes atendidos en cada día, se obtendrá la suma, ya sea semanal o mensual, para emitir los reportes que se quieran mostrar.

En el caso de las atenciones hospitalarias, el conteo de pacientes se realizará considerando la cantidad de ingresos hospitalarios que se realizaron en cada día. Cada uno de estos ingresos

corresponde a un paciente, los cuales se sumarán para obtener la cantidad de pacientes por semana y por mes respectivamente.

Para el caso de las atenciones de urgencia, se aplicará el criterio de conteo de Rut de igual manera como se evalúa la cantidad de pacientes en el área ambulatoria. En el caso de que un paciente se derive a hospitalización, el paciente no se contará como ingreso de urgencias si no que como hospitalario.

4.1.3. Distribución de Venta por Aseguradora

Como se mencionó anteriormente, es de gran importancia para una institución clínica conocer cuál de las aseguradoras es la que genera mayores ingresos, para optimizar los convenios y relaciones.

Para cubrir este requerimiento, se presenta el reporte de distribución de venta por aseguradores, que entrega en valores porcentuales la proporción de ingresos generados por cada aseguradora.

Para emitir este reporte se utiliza la base de datos de ventas y transacciones, información que viene asociada a los pacientes y que permite obtener la información ligada a las aseguradoras.

La distribución de venta se calculará en base al total de la venta y se representará de manera mensual a desde el año 2019 hasta julio 2021. De esta forma se podrá evaluar la evolución mensual a través de los últimos años y los distintos escenarios que se han hecho presente, como por ejemplo, el estallido social de octubre 2019 y la pandemia por Covid-19 que inició en marzo de 2020.

4.1.4. Demografía de Pacientes

De igual manera en que se obtiene la distribución de venta por aseguradoras, es posible obtener distintos indicadores asociados a las características demográficas de los pacientes, como, por ejemplo:

- Indicador de distribución geográfica, para identificar donde reside la mayoría de los pacientes actuales.
- Indicador de distribución de venta por rango etario y sexo, para identificar cuál de los grupos de pacientes genera la mayor cantidad de ingresos.

Este último indicador no solo servirá para identificar que grupo de pacientes genera mayores ingresos, sino que también, para identificar cuanto frecuentan los pacientes la clínica, y cuantos ingresos generan en promedio por visita.

Para determinar este indicador, lo primero será crear agrupaciones por rango etario, para facilitar la representación de los datos en los gráficos. Los grupos generados son los siguientes:

Grupo	Edad
1	[0-15]
2	[16-30]
3	[31-45]
4	[45-60]
5	[61-75]
6	[76-]

Tabla 1: Agrupación de pacientes por rango etario.

4.2. Implementación de la Plataforma Web:

4.2.1. Adquisición de la base y preparación de los datos:

El almacenamiento de la base de datos de la institución clínica se encuentra externalizado a una empresa que no pertenece al corporativo, la cual mantiene respaldos de la información tanto en el DataCenter como en la nube. Junto con el almacenamiento, esta empresa cuenta también con los computadores para la carga y la gestión de la base de datos, a los cuales se accede mediante una máquina virtual.

La base de datos corresponde a un modelo relacional, que fue pensada para ser aplicada de manera transversal a las múltiples clínicas integradas al holding de empresas al que pertenece la institución, donde cada una de ellas, tiene sus propios parámetros de acceso a la información. Para acceder a la base de datos, se utiliza un motor Oracle, específicamente mediante el software de gestión de bases de datos SQL Developer de Oracle. Sin embargo, para efectos del proyecto, se realizará la adquisición de datos con el IDE (*Integrated Drive Electronics*) Jupyter Notebook, mediante lenguaje de programación en Python. En este, se accede a las bases de Oracle mediante la librería **cx_Oracle**, creando una conexión de la siguiente manera:

```
import cx_Oracle
conexion=cx_Oracle.connect("usuario/password@host:puerto/servicio")
```

Figura 1: Conexión a base de datos Oracle mediante el IDE Jupyter Notebook

Luego de definida la conexión a la base de datos, es posible cargar consultas SQL y obtener resultados en formato DataFrame, lo cual facilita la lectura, análisis y modificación de las tablas.

```
import pandas as pd
dataframe=pd.read_sql("SELECT * FROM TABLA",con= conexion)
```

Figura 2: Descarga de tablas mediante consultas SQL en Jupyter Notebook.

A partir de esta herramienta de trabajo, se realiza la lectura de las tablas necesarias para emitir los indicadores. A continuación, mediante funciones tradicionales de Python como **Numpy**, **Pandas** y **Datetime**, se realiza la limpieza de datos y preparación de los indicadores a presentar. Finalmente, se utiliza la librería **Matplotlib** para generar las imágenes que serán levantadas en el servicio web. Estas imágenes se guardan en el ordenador local, y se renuevan cada vez que se ejecute el código de actualización de la base de datos.

4.2.2. Levantamiento del Servicio Web:

Las imágenes generadas mediante el programa anterior son las que darán forma al dashboard de indicadores que se presentará mediante una plataforma web.

Para levantar el servicio web, se seguirá utilizando el IDE Jupyter Notebook en conjunto con **FastAPI**, un framework caracterizado por ser de alto rendimiento, robusto, intuitivo y fácil a la hora de programar [13].

Mediante FastApi es posible crear distintos EndPoint en la web, que corresponden a las ubicaciones de cada servicio, cuya dirección será utilizada para dirigir los mensajes emitidos desde el cliente [14]. En cada EndPoint habilitado se presentarán indicadores relacionados formando dashboards. Como ejemplo, se muestra la creación del EndPoint “**home**” que retorna el string “Hola Mundo”:

```
from fastapi import FastAPI|  
  
app = FastAPI(title='Implementando un servicio web en FastAPI')  
  
@app.get("/")  
async def home():  
    return "Hola Mundo"
```

Figura 3: Creación del EndPoint “home” mediante FastApi en Jupyter Notebook.

Luego, mediante **Uvicorn**, un servidor ASGI (*Asynchronous Server Gateway Interface*) que proporciona la conexión a la interfaz web, se levanta el servidor local con los EndPoint que fueron creados mediante FastApi, de la siguiente manera:

```
import uvicorn  
  
host = "127.0.0.1"  
uvicorn.run(app, host=host, port=8080)
```

Figura 4: Levantamiento del servicio web a un servidor local mediante Uvicorn

De esta forma el servidor queda funcionando con los EndPoint que fueron definidos, pero que hasta el momento se encuentran vacíos. A continuación, se muestra como se ve el EndPoint “home” creado como ejemplo en el servidor local:

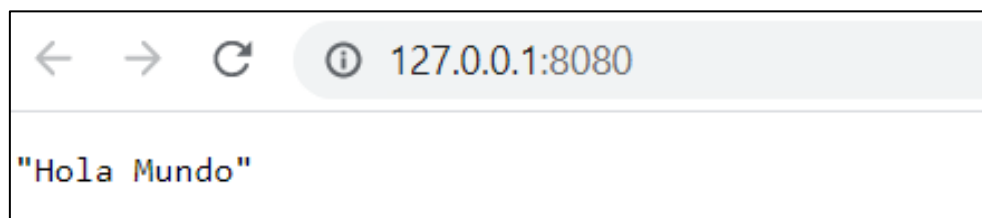


Figura 5: Ejemplo de página web sencilla implementada en servidor local.

4.2.3. Diseño del Front End de la Página Web:

Hasta el momento, los EndPoint definidos se encuentran vacíos, por lo que es necesario cargar archivos en formato HTML que contengan la información y las imágenes que se desean mostrar. Para simplificar esta etapa se utilizarán templates prediseñados de HTML. Estos corresponden a modelos de páginas web ya desarrolladas y requieren solo algunas modificaciones para obtener el formato de la página deseada.

Para ello se utilizarán nuevas librerías y funciones, como **Jinja2Templates** [15], que permite cargar los template de HTML directamente en el EndPoint creado por FastAPI. Además de lo anterior, se utilizarán funciones como **Request** de **starlette.request** que obtiene información sobre la dirección IP del cliente y crea correctamente la conexión al servidor [16].

A continuación, se muestra cómo se carga un template al EndPoint “home” creado anteriormente:

```
from fastapi import FastAPI
from fastapi.responses import HTMLResponse
from starlette.templating import Jinja2Templates
from starlette.staticfiles import StaticFiles
from starlette.requests import Request

app = FastAPI(title='Implementando un servicio web en FastAPI')
app.mount("/static", StaticFiles(directory="static"), name="static")
templates = Jinja2Templates(directory="templates")

@app.get("/", response_class=HTMLResponse)
async def home(request: Request):
    data = {'text': '<div id="top">\n<h2>Hola Mundo: </h2>\n'}
    return templates.TemplateResponse("page.html", {"request": request, "data": data})
```

Figura 6: Carga del template “page.html” al EndPoint “home” mediante Jinja2Template.

El template utilizado fue descargado de Bootstrap, un kit de herramientas de código abierto para desarrollar Front-End de páginas web [17], donde es posible encontrar una gran variedad de modelos de distintos colores, diseños y funcionalidades. El template seleccionado se compone principalmente por un fondo blanco con la capacidad de desplazarse, junto con una barra de navegación interactiva con la que se puede acceder a los distintos EndPoint que fueron definidos. A continuación, se muestra el template seleccionado, sin ninguna modificación adicional:



Figura 7: Template base cargado en la página web, que cuenta con barra de navegación interactiva.

Para poder cargar los dashboard en el servicio web, si será necesario realizar modificaciones al template de la Figura 7. Para ello, se sobrescribirán los archivos mediante el IDE Spyder o incluso con algún editor de texto, con el cual se añadirán las instrucciones de cargar los gráficos de indicadores y la información que deseamos visualizar en el servicio web.

4.3. Modelo de Segmentación de Pacientes:

4.3.1. Descripción del modelo:

Acorde al tercero de los objetivos específicos, se presentará el indicador de segmentación de pacientes, cuya función será agrupar a los pacientes y poder identificar cuál de estos grupos es el que genera mayores ventas. A diferencia de los indicadores anteriores, este KPI no será de características descriptivas, más bien, será un clusterizador automático obtenido a través de un modelo de machine learning.

El machine learning o aprendizaje automático, por su traducción al español, consiste en entrenar un sistema computacional para extraer información de los datos de manera automatizada. Existen dos grandes tipos de algoritmos de aprendizaje automático: los de aprendizaje supervisado y los de aprendizaje no-supervisado [18].

Para entrenar un modelo de aprendizaje supervisado, el usuario deberá entregar información que tenga los datos de entrada y de salida deseada, como consecuencia, el algoritmo se entrenará hasta encontrar la manera de producir la salida deseada dada una entrada específica. En otras palabras, el algoritmo será capaz de crear una salida para una entrada que nunca había visto sin necesidad del recurso humano [18].

Por otro lado, para entrenar un modelo de aprendizaje no supervisado, el usuario entregará información de los datos de entrada sin necesidad de proporcionar los datos de salida al algoritmo. Como resultado, estos algoritmos crean una nueva representación de los datos que facilitan el análisis en comparación con los datos originales, sin embargo, como estos datos no contienen información de salida, es difícil estimar cual debería ser el resultado correcto [18].

El modelo que se utilizará para definir el indicador de segmentación de pacientes se encuentra compuesto por sistemas de aprendizaje no-supervisado. Esto se debe a que la información actual no contiene etiquetas de salida, y en efecto, su objetivo será agrupar a los pacientes y poder identificar cuál de los grupos es el que genera mayores ventas. Los sistemas que se utilizarán corresponden a la Reducción de Dimensiones por FAMD (*Análisis Factorial de Datos Mixtos*) y la clusterización por KMeans.

4.3.2. Datos de entrada al modelo:

Los datos de entrada que se utilizarán para el entrenamiento del modelo de machine learning corresponden a la totalidad de los pacientes atendidos durante el año 2021. Cada uno de estos pacientes tiene un identificador único, con el cual es posible realizar el seguimiento de la cantidad de atenciones realizadas en los tres servicios de la clínica: ambulatorio, hospitalario y de urgencias. Con esta información se crea la tabla de datos de entrada del modelo que contiene las atenciones solicitadas por el paciente y algunos de sus datos personales, como se muestra en la Figura 8:

	ATENCIONES_AMB	ATENCIONES_HOSP	ATENCIONES_URG	EDAD	SEXO_PACIENTE	NOMBRE_ISAPRE	NOMBRE_COMUNA
0	1	0	0	65.0	M	FONASA	VITACURA
1	3	0	0	82.1	F	ISAPRE CRUZ BLANCA	LAS CONDES
2	4	0	1	65.0	F	ISAPRE BANMEDICA S.A.	LAS CONDES
3	2	0	0	122.0	F	ISAPRE COLMENA	VITACURA
4	1	0	0	34.9	M	PARTICULAR	VITACURA
...
202686	1	0	0	33.4	M	ISAPRE BANMEDICA S.A.	CON CON
202687	1	0	0	2.0	M	ISAPRE CONSALUD	RENCA
202688	1	0	0	0.1	M	ISAPRE CRUZ BLANCA	LO BARNECHEA
202689	0	0	1	3.5	F	ISAPRE CONSALUD	RECOLETA
202690	0	0	1	1.2	M	ISAPRE CONSALUD	MAIPU

Figura 8: Tabla de datos de entrada para el modelo de machine learning no-supervisado.

4.3.3. Reducción de Dimensionalidad:

La tabla de datos de entrada de la Figura 8 está compuesta por 202.690 filas, correspondientes a cada uno de los pacientes atendidos, y por 7 columnas de características tanto del tipo categórico como numérico.

Para representar este tipo de indicador mediante gráficos será necesario reducir sus siete variables de entrada a un sistema de dos dimensiones, ya que actualmente no es factible la visualización de estos datos. Para ello, se implementará la primera de las transformaciones de aprendizaje no-supervisado del modelo: la reducción de dimensiones. Esta transformación tomará la representación de los datos originales, que consta de muchas características, y encontrará una nueva forma de representarlos con menos características, manteniendo solo las de mayor importancia o esencialidad [18].

A nivel de programación, como indica [19], actualmente existen variadas técnicas para reducción de dimensionalidad, como por ejemplo:

- **Análisis de componentes principales (PCA):** Técnica de reducción de dimensionalidad más utilizada debido a su simplicidad para extraer la información relevante de un conjunto de datos. Suele utilizarse cuando todas las variables de entrada corresponden a datos numéricos.
- **Análisis de correspondencia múltiple (MCA):** Técnica de reducción de dimensionalidad que se utiliza específicamente para características categóricas.

- **Análisis factorial de datos mixtos (FAMD):** Técnica de reducción de dimensionalidad utilizada para conjuntos de datos que contenta características tanto cuantitativas como cualitativas, es decir, un conjunto de datos con características numéricas y categóricas.

Debido a que el conjunto de datos contiene características cuantitativas y cualitativas, se realizará la reducción de dimensiones mediante la técnica de Análisis Factorial de Datos Mixtos. La reducción se realizará en el IDE Jupyter Notebook a través de la función **prince.FAMD** de Python [20], que recibe como parámetros la cantidad de componentes principales deseados y la tabla de datos de entrada original. El código utilizado se muestra a continuación:

```
import prince
ca = prince.FAMD(n_components=2)
componentes_principales = ca.fit_transform(df_inicial)
```

Figura 9: Código para obtener dos componentes principales mediante la función prince.FAMD.

Para efectuar la reducción de dimensiones es importante conocer cuál es el porcentaje de información que se está manteniendo en el sistema y cuanto se está descartando. Para conocer este valor se recurre a la Proporción de Varianza Explicada de cada componente principal [21], la cual entrega el porcentaje de información que contiene cada uno de ellos. A continuación, se muestra el método **explained_inertia_** de **FAMD** que se utiliza para ello:

```
import prince
ca = prince.FAMD(n_components=7)
variance = ca.fit(df_inicial).explained_inertia_
```

Figura 10: Código para obtener la varianza explicada de un sistema de 7 variables de entrada.

El resultado de este código será una lista con siete valores porcentuales que representan la cantidad de información que contiene cada uno de los siete componentes principales de la tabla de origen.

De esta manera será posible determinar cuanta información se está rescatando al haber seleccionado solo dos de los componentes principales de los datos, los que formarán la nueva tabla de datos de entrada para el modelo de segmentación de pacientes.

4.3.4. Segmentación de pacientes:

Con la nueva tabla de datos de origen obtenida se procede a la segunda transformación de aprendizaje no-supervisado del modelo: la segmentación o clusterización de pacientes. Los algoritmos de agrupación tienen la funcionalidad de segmentar el conjunto de datos en grupos llamados clústeres. El objetivo es dividir los datos de tal manera que los puntos dentro de un mismo grupo sean muy similares y los puntos en diferentes grupos sean diferentes. De manera similar a los algoritmos de clasificación, los algoritmos de agrupamiento predicen a qué grupo pertenece un punto en particular. [20]

La técnica de clusterización de datos utilizada en el proyecto corresponde a **Kmeans** [22] de la librería **Scikitlearn**. Esta recibe como parámetros a la base de datos obtenida por la reducción de dimensionalidad y a la cantidad de grupos en que se desean separar los datos, cuyo resultado será la asignación de un “grupo” para cada fila de la tabla de datos, es decir, para cada paciente. A continuación, se observa un ejemplo sobre como ejecutar Kmeans en el IDE Jupyter Notebook:

```

from sklearn.cluster import KMeans
n_clusters=5
kmeans=KMeans(n_clusters=n_clusters)
kmeans.fit(df)
result_clusterizacion=kmeans.predict(df)

```

Figura 11: Código para realizar clusterización mediante Kmeans con $k = 5$.

En el ejemplo anterior se observa que la cantidad de clústers es igual a cinco, valor que ha sido seleccionado arbitrariamente para el ejemplo. Sin embargo, para un sistema específico, será necesario definir el número óptimo de clústers.

Para escoger el número de clústers a utilizar se empleará el **Método del Codo**, técnica basada en medir la distancia intra-clúster promedio del sistema. En Kmeans, cuanto más grande es el número de clústers (k), la distancia intra-clúster tiende a disminuir y los clústers son más compactos. Sin embargo, no es factible utilizar un número de clústers infinitos, por lo que se deberá encontrar un punto medio, donde se utilice el menor valor de k que satisfaga que un incremento de k no siga mejorando la distancia media intra-clúster [23].

En el ejemplo de la Figura 12 se aprecia como la evolución de la distancia intra-clúster promedio genera un cambio brusco con un valor de $k = 3$, el cual correspondería al valor de k óptimo [24].

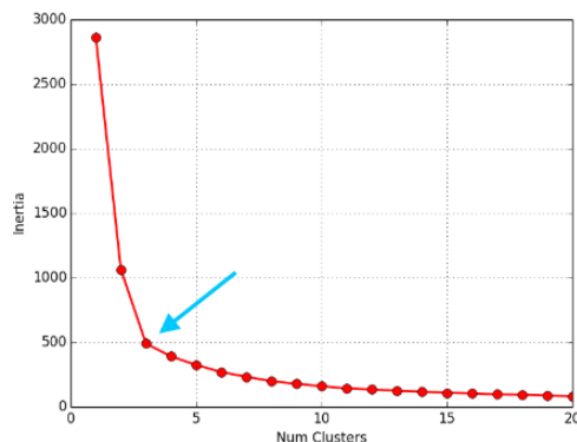


Figura 12: Ejemplo de método del codo con resultados para $k = 3$. [24]

5. PRESENTACION Y ANÁLISIS DE RESULTADOS

En el presente capítulo se presentarán todos los resultados asociados a la metodología descrita previamente. Se abordará la visualización de cada indicador descrito, se presentarán los resultados del servicio web implementado y para finalizar, se mostrará el resultado del modelo de machine learning para Segmentación de Pacientes.

Los datos para analizar corresponden a toda la información transaccional de pacientes generados a partir del 01 de enero del 2021 hasta el 31 de noviembre del 2021.

5.1. Visualización y análisis de indicadores

5.1.1. Reporte de Ventas Semanal

Mediante el reporte de ventas de la institución, es posible evaluar la evolución de las actividades comerciales semana a semana, tanto para la línea de negocio hospitalaria, ambulatoria y de urgencias. Dada su función de visualización y análisis, se utiliza para investigar las bajas o alzas imprevistas, y así mismo intentar anteponerse a una situación similar en el futuro.

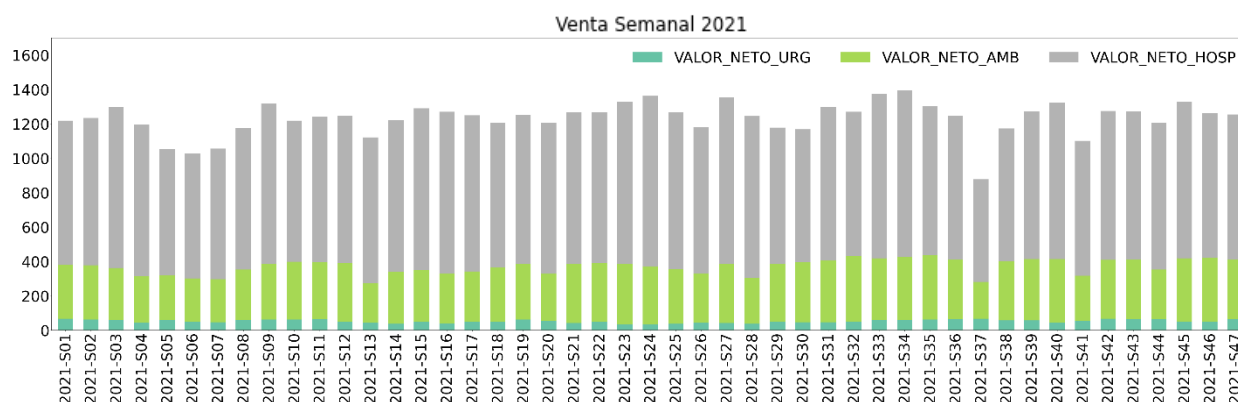


Figura 13: Reporte de ventas semanal año 2021.

Al analizar el gráfico de la Figura 13, es posible observar que durante las semanas s05, s06 y s07 se presenta una disminución en relación con las demás semanas. Al verificar mediante el calendario, se puede ver que estas semanas corresponden al mes de febrero donde, a nivel general, suele haber una menor cantidad de personas en la región, dejando como consecuencia, una disminución en la venta de febrero. Por el contrario, en las semanas s08 y s09 de marzo se observa un alza en la venta, que corresponden a las semanas de inicios de clase, donde la mayoría de las personas vuelven de sus vacaciones.

Con esta información se trata de reflejar como el calendario laboral y académico de las personas puede afectar directamente a la venta de la institución. Otro ejemplo de aquello se presenta con los días feriados, como sucede por ejemplo, en la semana s13 del gráfico que corresponde al festivo religioso *Pascua de Resurrección* y en la s37 que corresponde a *Fiestas Patrias*.

Al observar la Figura 13 se ve que la venta Ambulatoria es la que principalmente se ve afectada. Esto se debe a que el negocio ambulatorio depende directamente de los días hábiles que se tenga para atender, por otro lado, la venta hospitalaria y de urgencias, dependen en muy poca medida de los días hábiles. Las cirugías y hospitalizaciones suelen estar calendarizadas de antemano, es decir, que los días festivos si afectarán a la venta, pero no en la misma medida que en la venta ambulatoria. Mientras que, en el lado de las urgencias, que corresponden a sucesos independientes e imprevistos, no se verá la afectada la venta.

Para visualizar de mejor manera la influencia de los días festivos, se presenta la Figura 14 que cuenta con tres gráficos normalizados entre 0 y 1. Cada uno de ellos representa una de las líneas de negocio, donde se puede verificar la tendencia de la venta respecto a las semanas con días festivos del año.

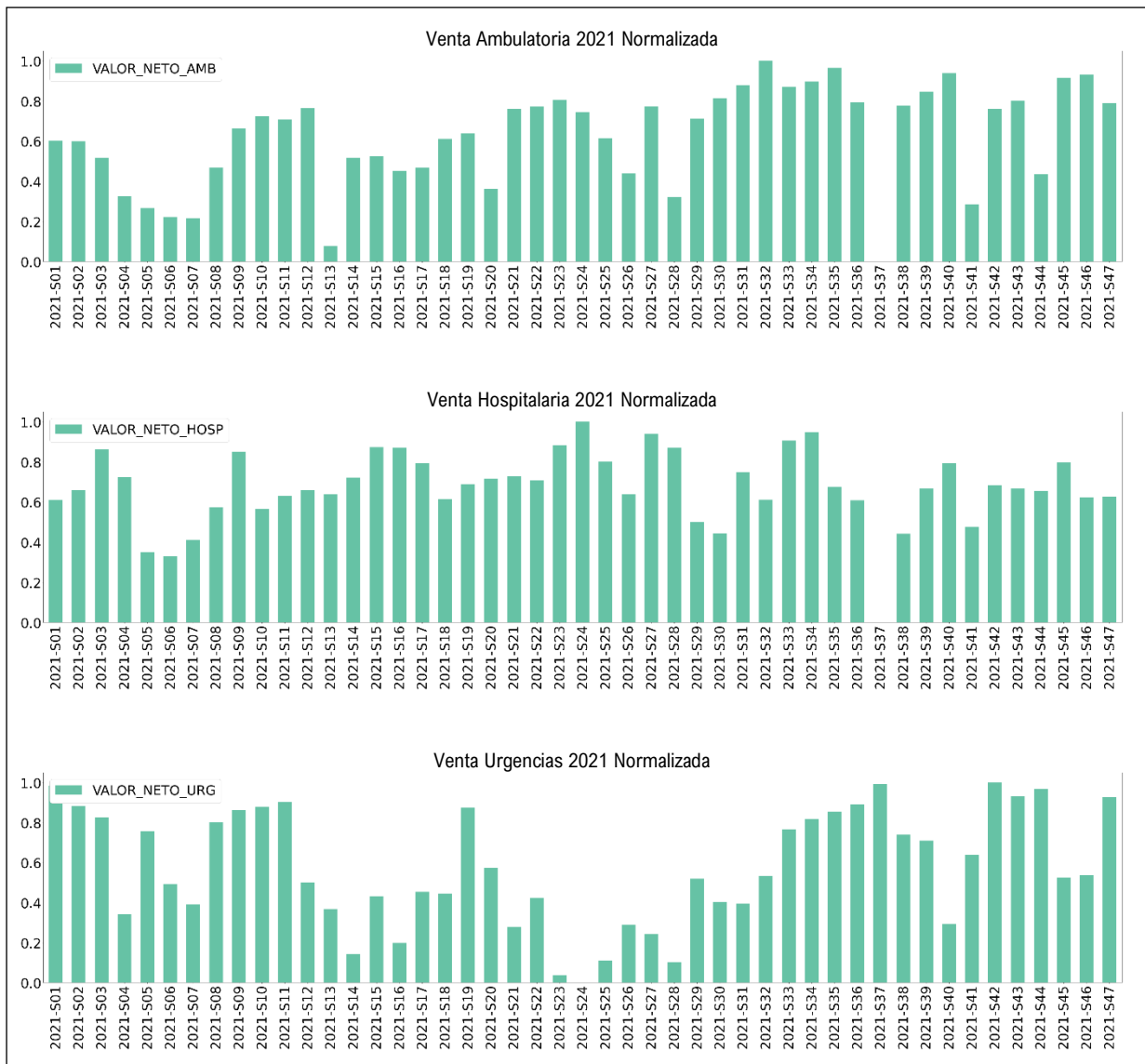


Figura 14: Venta semanal normalizada entre 0 y 1 para cada línea de negocio.

Al normalizar los valores de venta entre 0 y 1, se pueden identificar los valores máximo y mínimo de la venta semanal del año. Como resultado, se observa que el menor registro de ventas durante el año, para las líneas ambulatoria y hospitalaria, corresponde a las S37, que corresponde al festivo de *Fiestas Patrias*. Por el contrario, la venta de urgencias no se ve afectada por la S37, y que en efecto, es difícil identificar un patrón de comportamiento en el año.

Al conocer cómo afectan estas variables en la venta de una organización, es posible predecir como serán los resultados para los próximos años. A pesar de esto, existen escenarios imprevistos que pueden desviar los resultados, como por ejemplo, el escenario de una pandemia. Mediante la Figura 15 se puede ver la comparación entre los resultados obtenidos en el año 2019 (escenario normal), el año 2020 (escenario pandemia) y el año 2021 (escenario actual).

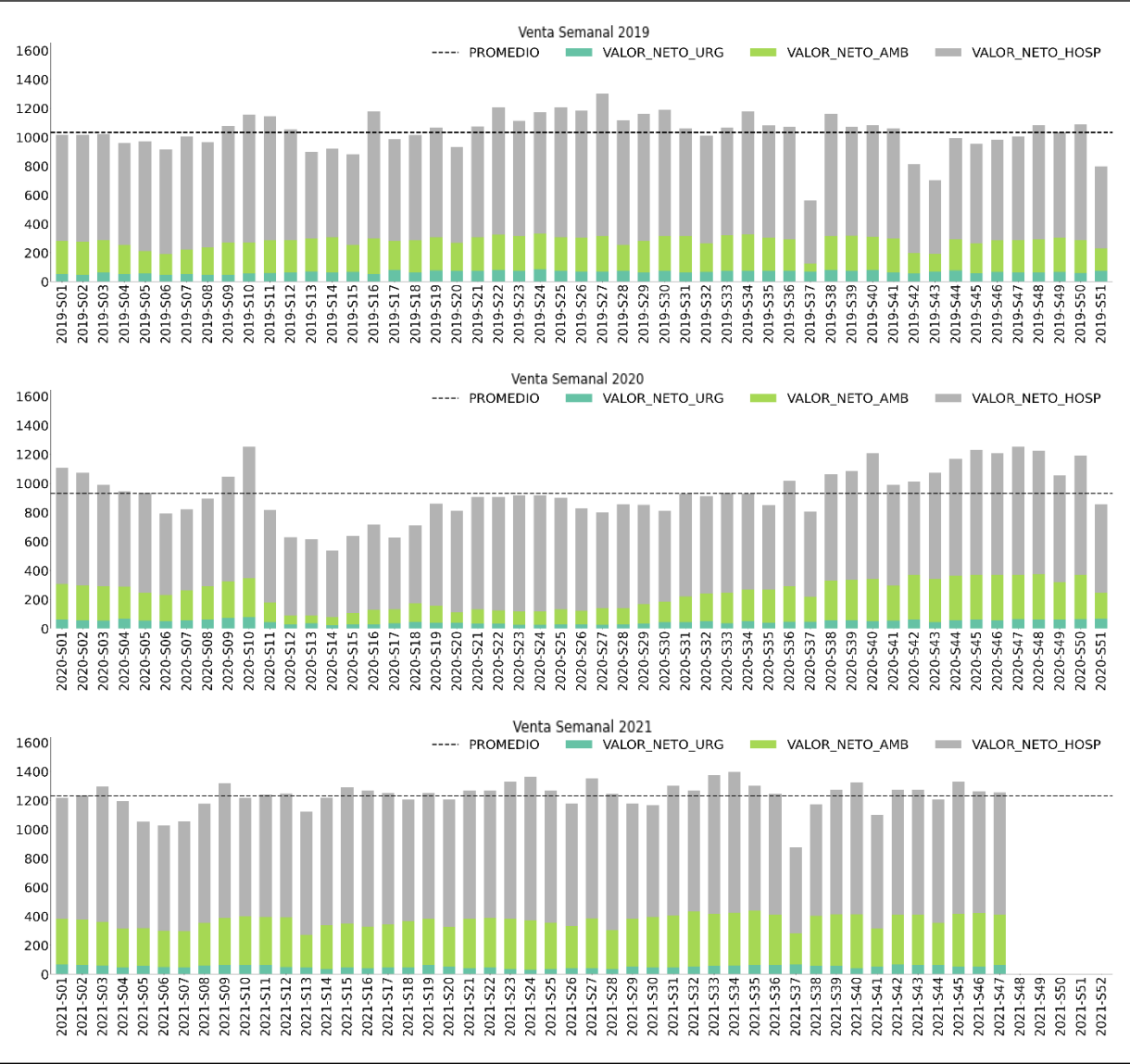


Figura 15: Reporte de ventas semanal años 2019, 2020 y 2021.

En los gráficos de la Figura 15 se comparten los mismos valores del eje Y para los tres años. De esta forma, se puede apreciar como la pandemia afecta en los resultados del año 2020.

Por otro lado, si comparamos la venta de los años 2019 y 2021, se puede ver que existe un comportamiento similar, y que incluso, aun con la pandemia presente en el año 2021, se observan mejores ingresos que en el 2019. Esto podría el ser resultado de la experiencia y de algún nuevo proyecto que se haya implementado en el año 2021.

5.1.2. Cantidad de Pacientes Atendidos

El reporte sobre la cantidad de pacientes atendidos que se observa en la Figura 16.a, se entrega de manera semanal y segmentado en las tres líneas de negocio, de la misma manera en que se muestra el reporte de ventas. Sin embargo, para contabilizar la cantidad de pacientes de las distintas líneas de negocio, se utilizaron criterios diferentes:

- **Ambulatorio:** Se cuenta la cantidad de pacientes atendidos en el día, independiente de la cantidad de consultas o exámenes que se cada uno de ellos se haya realizado.
- **Hospitalario:** Se cuenta la cantidad de ingresos hospitalarios que haya habido en el día, independientes de la cantidad de prestaciones que pueda realizarse durante toda su estadía.
- **Urgencias:** Se cuenta la cantidad de pacientes que ingresan por el área de urgencias, sin embargo, si el paciente es derivado a hospitalización, será contabilizado en el área hospitalaria y no en urgencias.

Como se puede ver, la mayoría de los pacientes que asisten a la clínica corresponden a pacientes que se realizan prestaciones ambulatorias, ya sean consultas médicas, exámenes, imágenes, etc. Sin embargo, esto no se relaciona con los porcentajes de distribución vistos en la Figura 15.

Para analizar esta información, se utiliza la Figura 16.b donde se entrega el valor de venta promedio por paciente, cuyos resultados, corresponden a la división entre la venta total de cada segmento y la cantidad de pacientes de cada segmento.

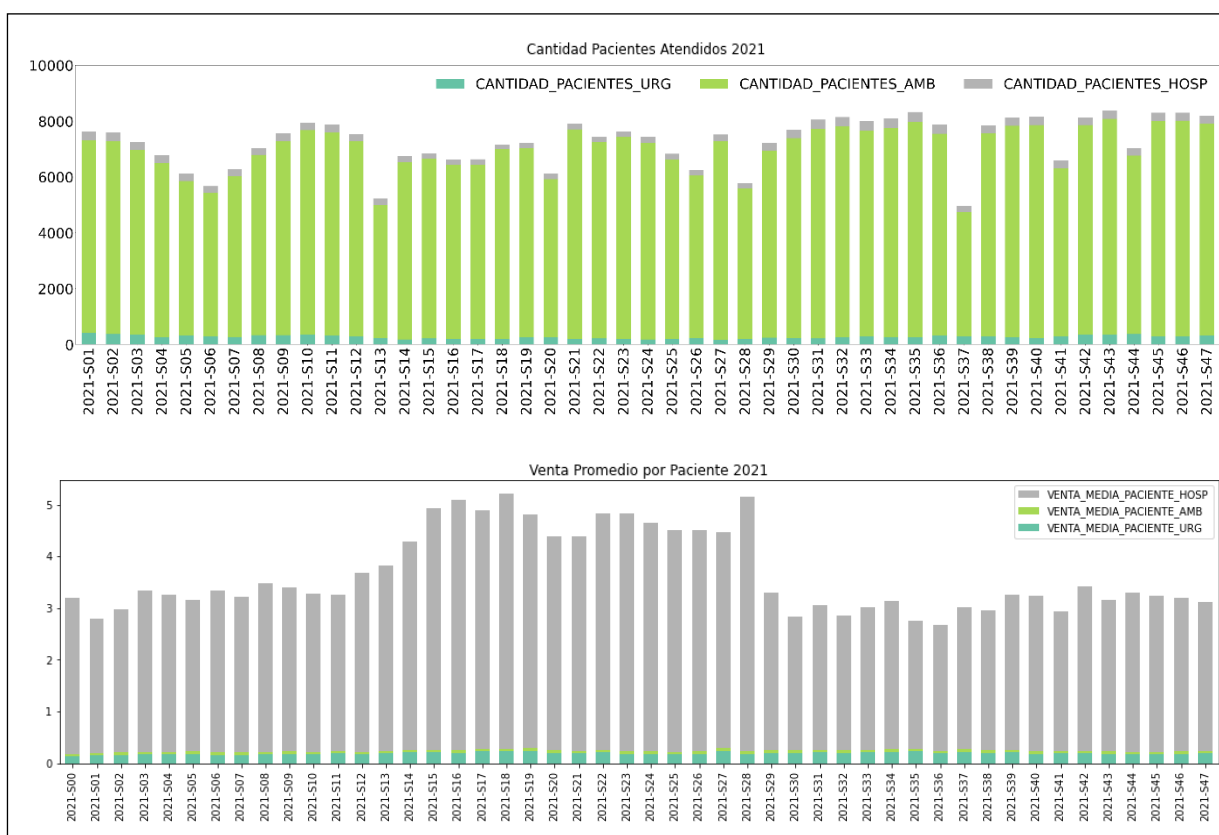


Figura 16: a) Cantidad de pacientes atendidos en el año 2021 de manera semanal.
b) Venta promedio de pacientes atendidos en el año 2021 de manera semanal.

Pese a que la mayoría de los pacientes atendidos proviene del área ambulatoria, se puede ver con claridad que no es el área que genera mayores ingresos a la clínica, como se muestra en la figura 16.b. A pesar de esto, es importante destacar que el área de prestaciones ambulatorias es un pilar fundamental a la otra de atraer nuevos pacientes a la clínica, los que en primera instancia llegan por una consulta médica y luego se derivan a un procedimiento hospitalario. Sin

prestaciones ambulatorias, la cantidad de cirugías y hospitalizaciones, que son el área del negocio que genera mayores ingresos, disminuiría en gran medida.

5.1.3. Distribución de Venta por Asegurador

El reporte sobre distribución de venta por asegurador pretende entregar información sobre cuál de las empresas aseguradoras que trabajan con la institución es el mayor cliente. Mediante el gráfico de la Figura 17, se entrega información sobre el ingreso generado por cada aseguradora respecto del total, como un valor porcentual en cada línea de negocios.

Como resultado, se obtuvo que las aseguradoras privadas generan un 73% de las ventas totales, en comparación con un 26% para Fonasa, lo cual es coherente con los análisis estadísticos sobre la distribución de aseguradoras por comunas realizado por la Superintendencia de Salud [25][26].

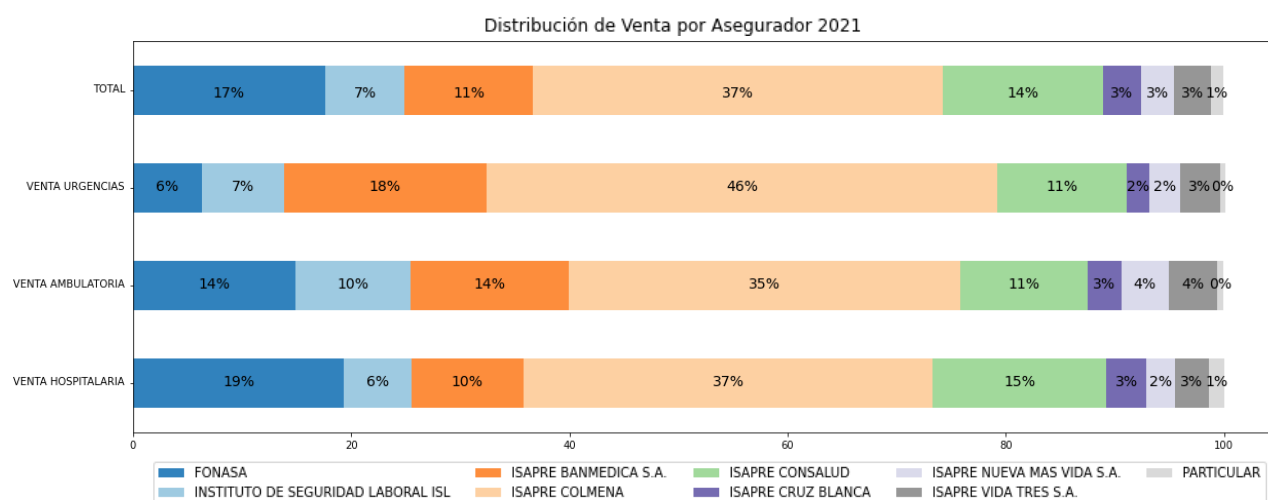


Figura 17: Distribución de venta por asegurador en el año 2021.

De este análisis, se considerarán los afiliados a Isapres y Fonasa de las comunas del sector SurOriente de la Región Metropolitana de Santiago: Las Condes, La Reina, Lo Barnechea, Ñuñoa, Providencia y Vitacura, donde se obtuvo lo siguiente:

- 631.748 beneficiarios de Isapres, de acuerdo con el Boletín Estadístico de Isapres de diciembre 2015 generado por la Superintendencia de Salud [25].
- 300.449 beneficiarios de Fonasa, de acuerdo con el Boletín Estadístico de Fonasa de diciembre 2015 generado por la Superintendencia de Salud [26].

Es decir que, de la población ubicada en el sector Sur Oriente de Santiago, el 32% se encuentra afiliada a Fonasa y el 68% se encuentra afiliada a Isapres, resultados similares a lo obtenido en el gráfico de distribución de venta por asegurador.

En la Figura 18, se presenta un gráfico de distribución de venta por asegurador que no considera los ingresos generados por Fonasa, con el fin de definir la influencia de las aseguradoras privadas. Como resultado, se obtiene que la Isapre de mayor influencia corresponde a Consalud con el 48% del mercado, seguido de Colmena y Cruz Blanca con un 14% cada uno.

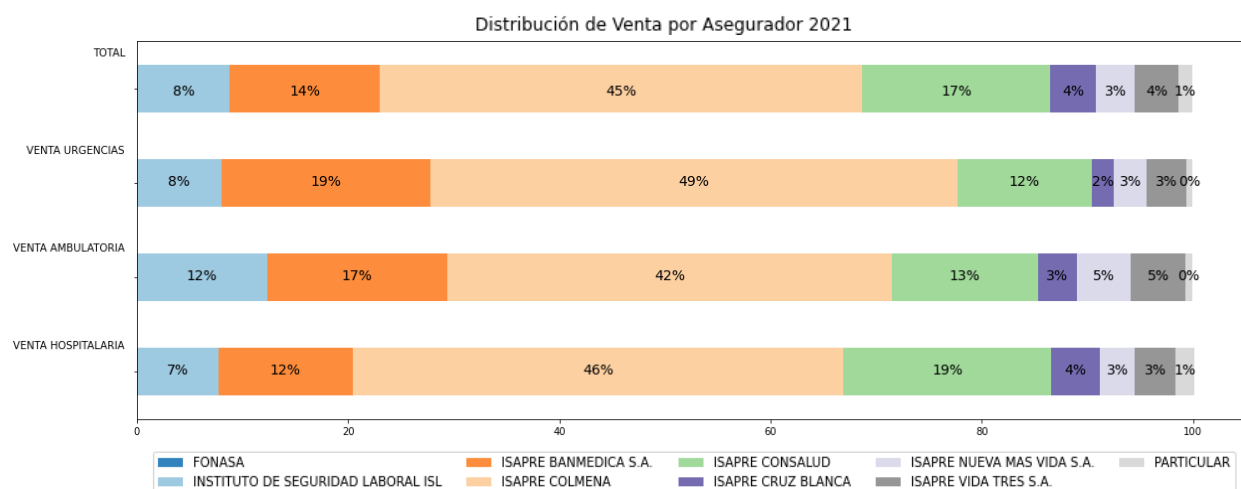


Figura 18: Distribución de venta por Isapres en el año.

Para validar esta información, se realizó una comparación con el estudio de mercado realizado por la Superintendencia de Salud “Prestadores de Salud, Isapres y Holdings” [27], donde se analiza la participación de mercado de las Isapres. En este estudio se informa que el 22% de los cotizantes de Isapres a nivel país pertenece a Consalud, seguido de un 20% para Cruz Blanca, 20% para Banmédica y 16% para Colmena.

Se puede observar que la mayoría de los pacientes atendidos en la institución provienen de Consalud, lo cual se debe a que ambos participantes, tanto la Isapre Consalud como la clínica en análisis, pertenecen al mismo holding de empresas, lo cual genera convenios internos y aumenta la derivación de pacientes por parte de la Isapre. Respecto al resto de Isapres, se puede ver que existe relación con los valores del estudio realizado por la superintendencia de salud.

Para finalizar el análisis de distribución de ventas por Isapre, se decide realizar un gráfico evolutivo mensual a partir del año 2019 y de esta forma, poder visualizar los cambios que ha sufrido la distribución de venta y si tiene relación con los escenarios que han surgido durante este periodo.

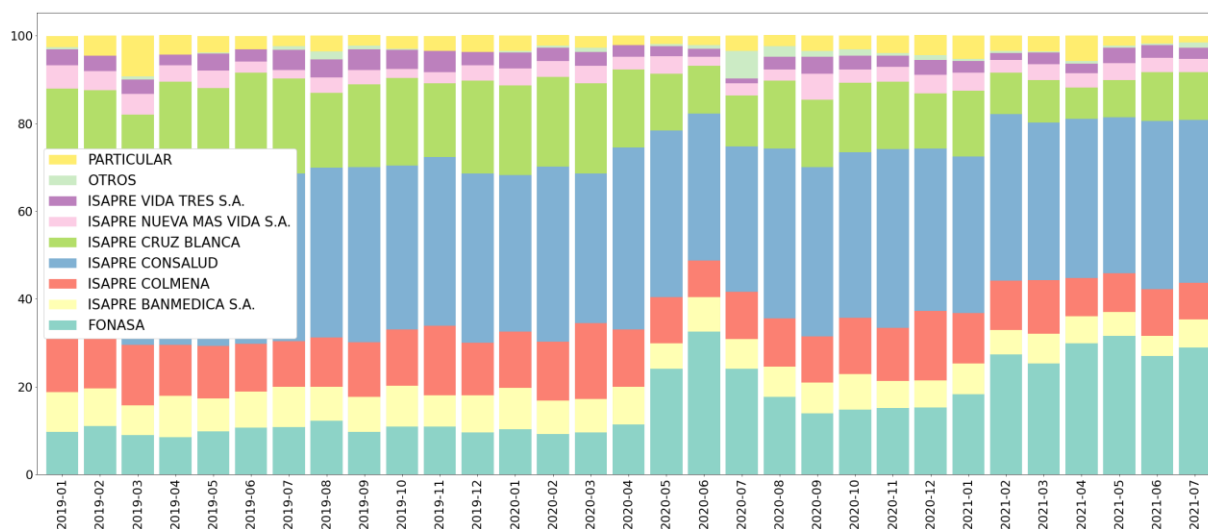


Figura 19: Evolución mensual de distribución porcentual de Venta por Aseguradora.

En la Figura 19 se puede ver que efectivamente hay variaciones durante los últimos años. Lo más destacable corresponde al alza de venta de Fonasa en el segundo trimestre del año 2020 y el primer semestre del año 2021, lo cual se relaciona directamente a las olas de contagio de covid-19 en el país. En estos periodos, se declaró el colapso total de los hospitales públicos del país, de acuerdo con lo informado por distintos noticieros [28], [29] y [30], donde se indica además que el Ministerio de Salud firmó acuerdos con clínicas privadas para que recibieran pacientes con covid-19. Dado lo anterior, se puede concluir que el alza de ventas generada por Fonasa efectivamente corresponde a derivaciones de pacientes Covid-19 que no pudieron ser atendidos en hospitales.

5.1.4. Demografía de Pacientes

El reporte sobre demografía de pacientes contiene información que puede ayudar a identificar que pacientes son los que se están atendiendo en la institución, y cuál de estos grupos es el que genera mayor rentabilidad.

Mediante el gráfico de distribución de venta por edad que se observa en la Figura 20, es posible determinar que la mayor cantidad de ingresos proviene del rango etario de 31 a 45 años con un 26% de las ventas, seguido del rango etario de 61 a 75 años con un 23% de las ventas. Por otro lado, la agrupación de pacientes que menos ingreso genera corresponde a los pacientes más jóvenes del rango 0 a 15 años con un 6% de las ventas.

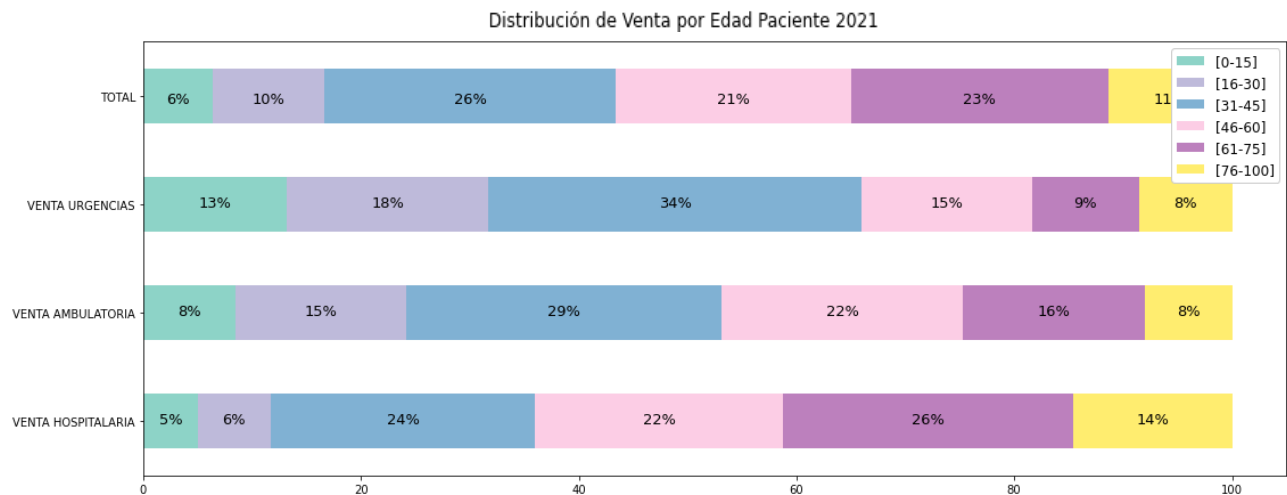


Figura 20: Distribución de venta por edad de pacientes año 2021.

Para complementar esta información, en la Tabla 2 se puede observar que, además de contener el porcentaje de venta de cada grupo de pacientes, también tiene el porcentaje atenciones para cada grupo. Esta información puede ser utilizada para definir quienes son los pacientes que más frecuentan la clínica, y además, para determinar cuál de estos grupos es el que genera mayores ingresos en una única visita, lo que corresponde al *Factor de rentabilidad*.

Por ejemplo, si se tiene que un grupo etario asiste a la clínica en un 1% de las atenciones totales y genera un 1% del total de la venta, significa que su factor de rentabilidad es igual a 1. Por otro lado, si otro grupo etario asiste a un 1% del total de las atenciones, pero genera un 2% del total de la venta, significa que es más rentable y que su factor de rentabilidad sería igual a 2.

En la tabla 2 se puede ver a modo de resumen, que los pacientes que más asisten a la clínica y que más ventas generan, corresponde al rango de entre 31 y 45 años, sin embargo, el grupo de pacientes que tiene un mayor factor de rentabilidad corresponde al grupo de entre 61 y 75 años.

Edad	% Atenciones	% Venta	Factor de Rentabilidad
[0-15]	10.9 %	6.4 %	0.6
[16-30]	16.6 %	10.2 %	0.6
[31-45]	29.1 %	26.8 %	0.9
[45-60]	20.6 %	21.6 %	1.0
[61-75]	15.1 %	23.6 %	1.6
[76-]	7.7 %	11.4 %	1.5

Tabla 2: Porcentaje de atenciones, de venta y factor de rentabilidad por rango etario.

Por otro lado, en la Figura 21 se presenta otra característica demográfica, que hace referencia a la residencia de los pacientes que se han atendido en la clínica durante el año 2021. En esta imagen se puede ver que la mayoría de los pacientes atendidos se concentra en la zona Oriente de la Región Metropolitana, no obstante, es posible apreciar una gran cantidad de pacientes proveniente de otras comunas.

El análisis de la distribución geográfica de los pacientes puede ser utilizado para realizar convenios con distintas municipalidades, mejorando los precios para los pacientes afiliados de cada comuna y aumentar la cantidad de atenciones realizadas. El modelo de negocio sería enfocado en aumentar el volumen de atenciones a cambio de precios más bajos, por ejemplo. Mediante la información del tipo demográfico, es posible definir nuevas estrategias de marketing en una organización, estableciendo metas enfocadas en el segmento de clientes que genera mayores ingresos, e incluso, tratar de disminuir los recursos destinados a las áreas que generan menor rentabilidad.

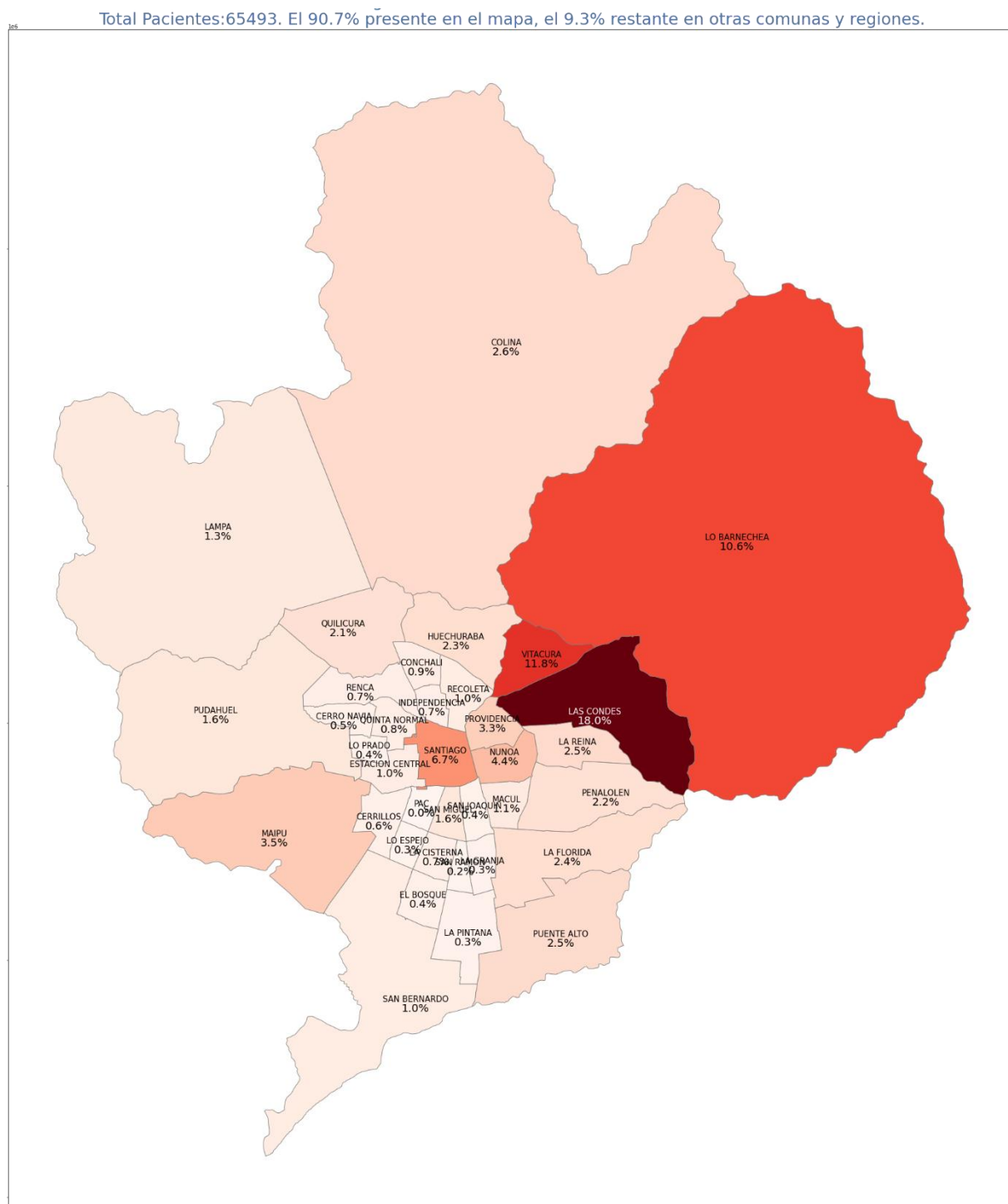


Figura 21: Distribución geográfica de pacientes atendidos en el año 2021.

5.2. Plataforma Web Implementada:

El servicio web implementado, cuenta con cuatro EndPoint activos donde se mostrarán los indicadores que fueron definidos. Mediante una barra de selección interactiva desarrollada mediante Bootstrap, es posible acceder a cada uno de los dashboard que se muestran a continuación.

5.2.1. Reporte de Ventas Semanal:

En esta sección se puede visualizar el EndPoint referente al Resumen de ventas. En este se muestran unos recuadros que entregan información simplificada sobre la venta acumulada semanal, mensual y anual en cada uno de los segmentos de negocio. Además, mediante los gráficos de barras apiladas que fueron definidos anteriormente, se muestra la evolución de venta semanal del año en curso, como se muestra en la Figura 22.

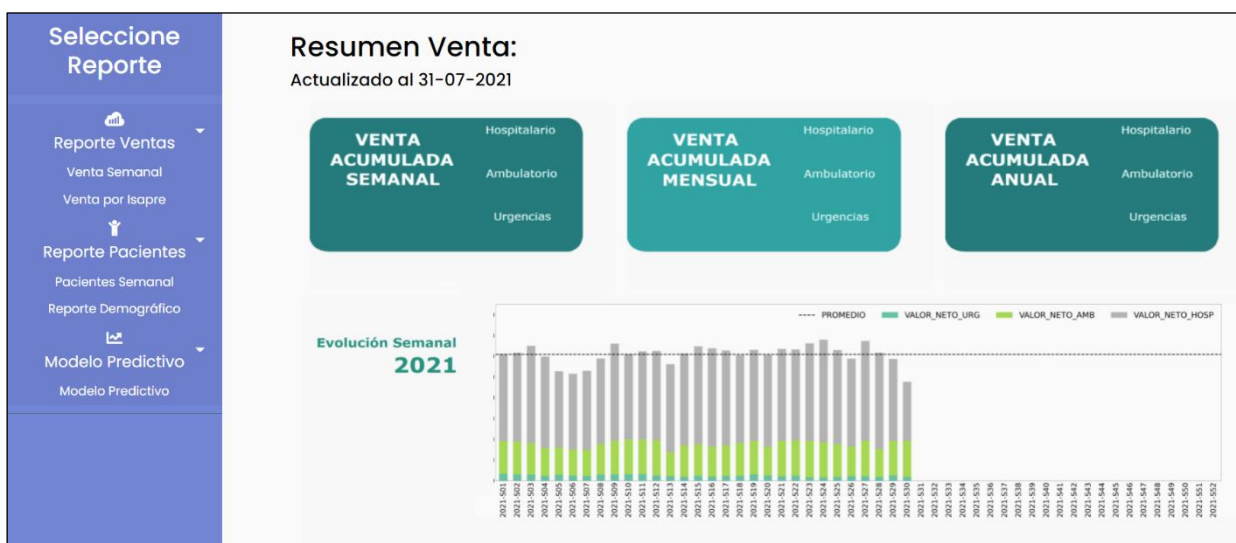


Figura 22: Dashboard del Reporte de ventas implementado en el servicio web.

5.2.2. Reporte de Ventas por Isapre:

La página que se muestra en la Figura 23 entrega información sobre la venta generada por cada una de las aseguradoras. En primera instancia y como imagen principal, se ve la distribución de venta por cada aseguradora a nivel porcentual de la venta total, sin embargo, bajo esta imagen se encuentra un sistema de acordeón o pestañas desplegables, que muestran de manera monetaria la venta específica de cada una de las aseguradoras como evolutivo en el tiempo.

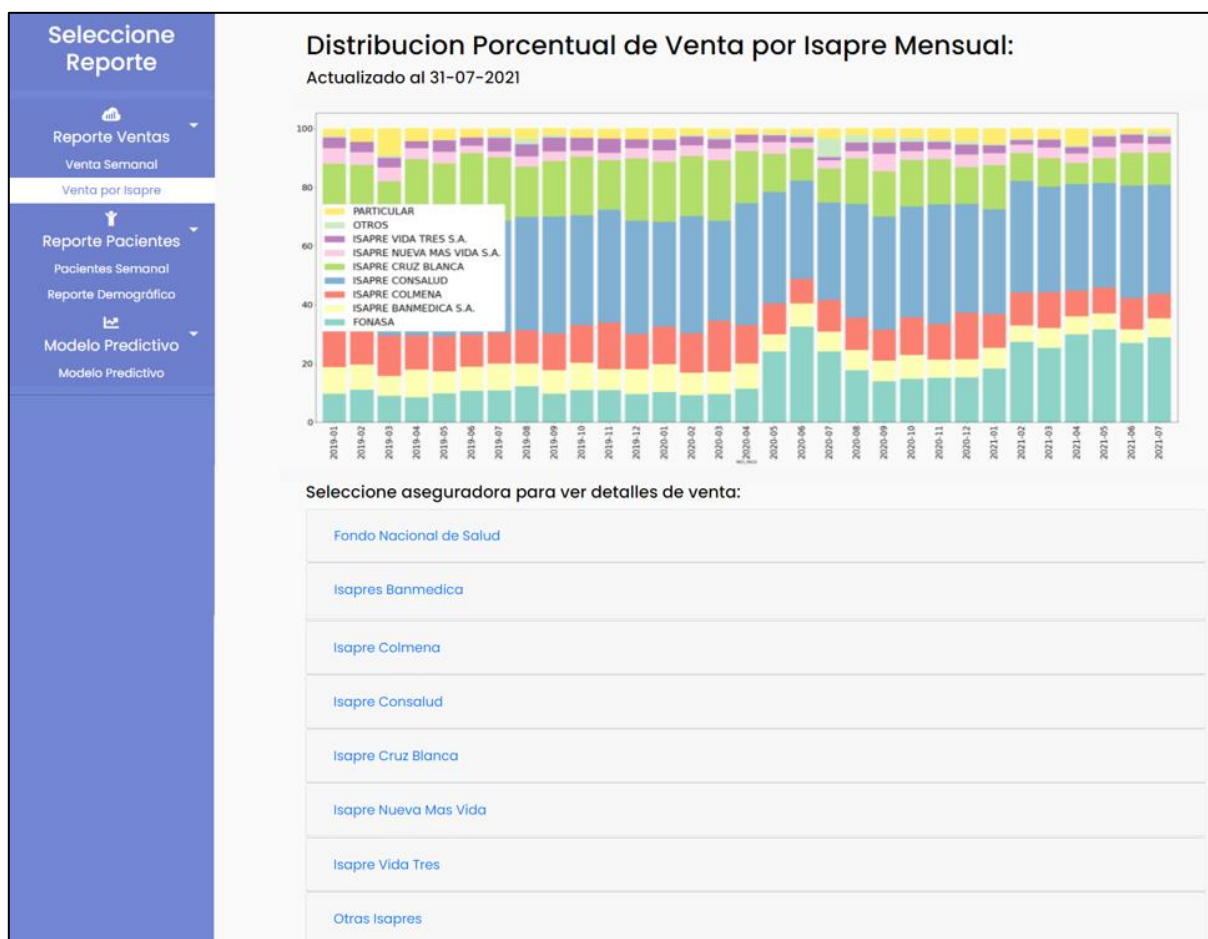


Figura 23: Dashboard del Reporte de ventas por aseguradora implementado en el servicio web.

5.2.3. Reporte de Cantidad de Pacientes:

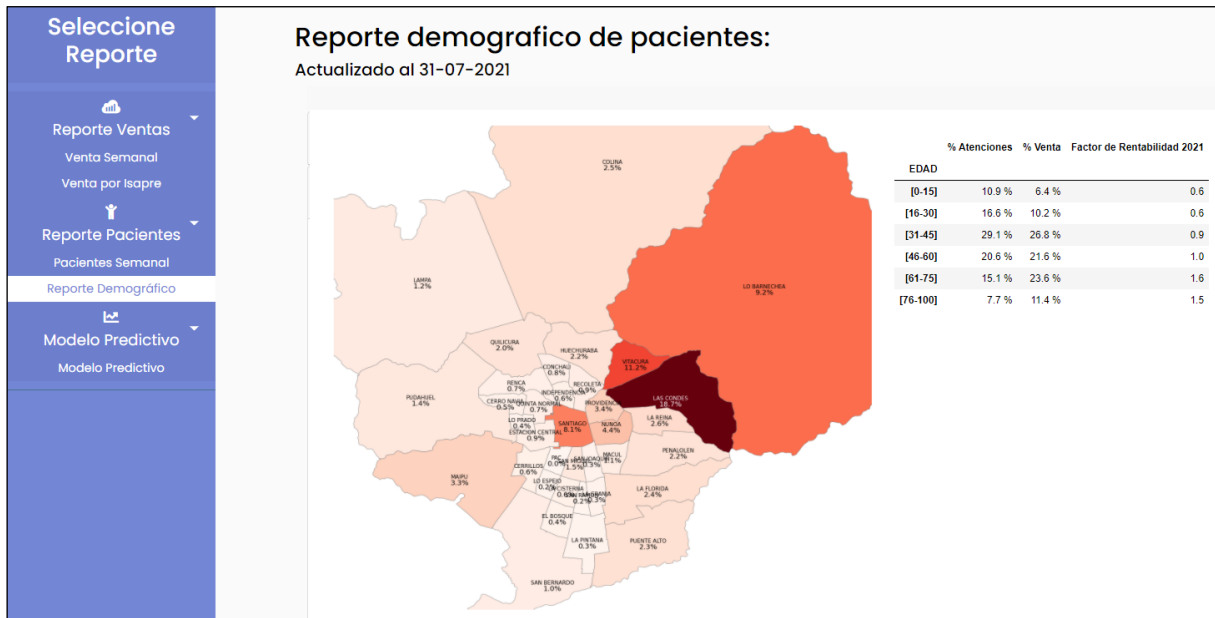
En la siguiente página, y de manera similar a como se representa el reporte de ventas, es posible visualizar los recuadros que entregan información sobre la venta acumulada semanal, mensual y anual, además del gráfico de barras apiladas que muestra el evolutivo de cantidad de pacientes atendidos por segmento de negocio, como se puede ver en la Figura 24.



Figura 24: Dashboard del Reporte de cantidad de pacientes implementado en el servicio web.

5.2.4. Reporte Demográfico:

En la Figura 25, que corresponde al endpoint del Reporte Demográfico de Pacientes, se pueden ver los dos indicadores del tipo demográfico que fueron definidos. El primero de ellos correspondiente a la distribución geográfica de las residencias de los pacientes, y el segundo, una tabla de rangos etarios que muestra información respecto al porcentaje de venta por grupo, porcentaje de atenciones por grupo y el factor de rentabilidad.



5.2.5. Reporte de Segmentación de Pacientes:

El último endpoint que se puede ver dentro de la barra de selección interactiva en la Figura 24, corresponde al indicador de Segmentación de Pacientes, el cual se desarrollará a continuación.

5.3. Modelo de Segmentación de Pacientes:

El modelo de segmentación de pacientes tiene la función de crear un indicador para identificar cuál de los grupos de pacientes es el que genera mayores ingresos.

El modelo de machine learning que se utilizó para crear este indicador se encuentra compuesto por dos técnicas de aprendizaje no-supervisado: la Reducción de Dimensiones por FAMD (Análisis Factorial de Datos Mixtos) y la clusterización por KMeans.

5.3.1. Reducción de dimensionalidad por FAMD:

La reducción de dimensiones se utiliza para transformar un sistema de múltiples variables a uno que contenga menor cantidad de dimensiones.

Para llevar a cabo la reducción es necesario conocer cuál es el porcentaje de información que contiene cada uno de los componentes del sistema. Para ello se recurre a la Proporción de Varianza Explicada que, a través del método `explained_inertia_` de FAMD, entrega el porcentaje de información de cada componente, como se muestra a continuación:

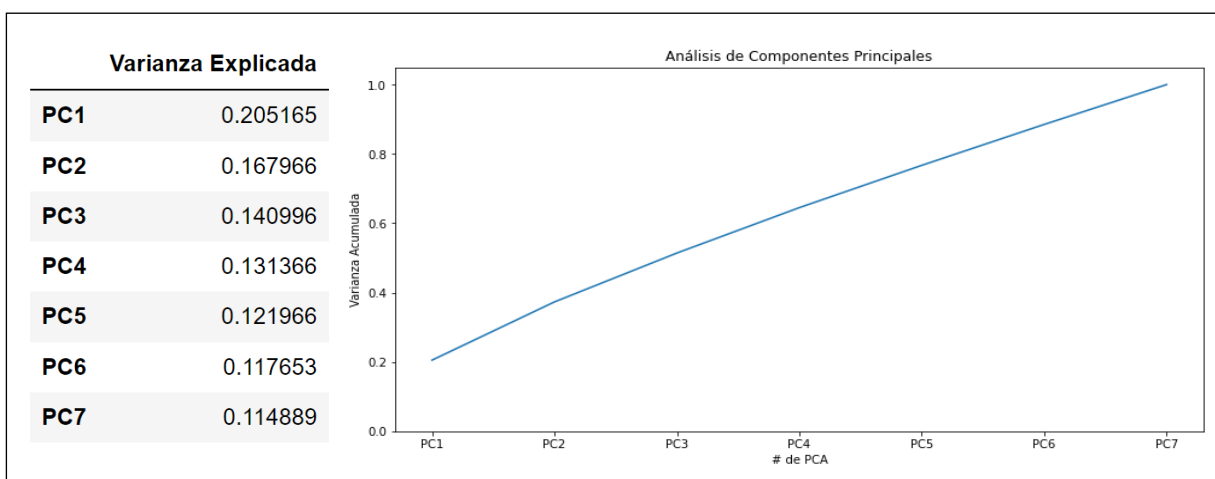


Figura 26: a) Varianza explicada para cada componente principal. b) Varianza total acumulada.

En la Figura 26 se puede ver la varianza explicada para cada uno de los componentes principales, los cuales, al sumarlos, dan como resultado el 100% de la información. El objetivo de la reducción de dimensionalidad es poder representar los datos mediante una visualización, por lo tanto, se seleccionarán los dos primeros componentes principales del sistema, que corresponden al 37% de los datos. A pesar de que esta información no es suficiente, se decide continuar con la implementación del modelo hasta completar el análisis de resultados.

En la Figura 27 se puede ver la representación en 2D de los componentes principales seleccionados, donde cada uno de los puntos corresponde a cada paciente del data set original representado por sus dos componentes principales, es decir, el 37% de la información original.

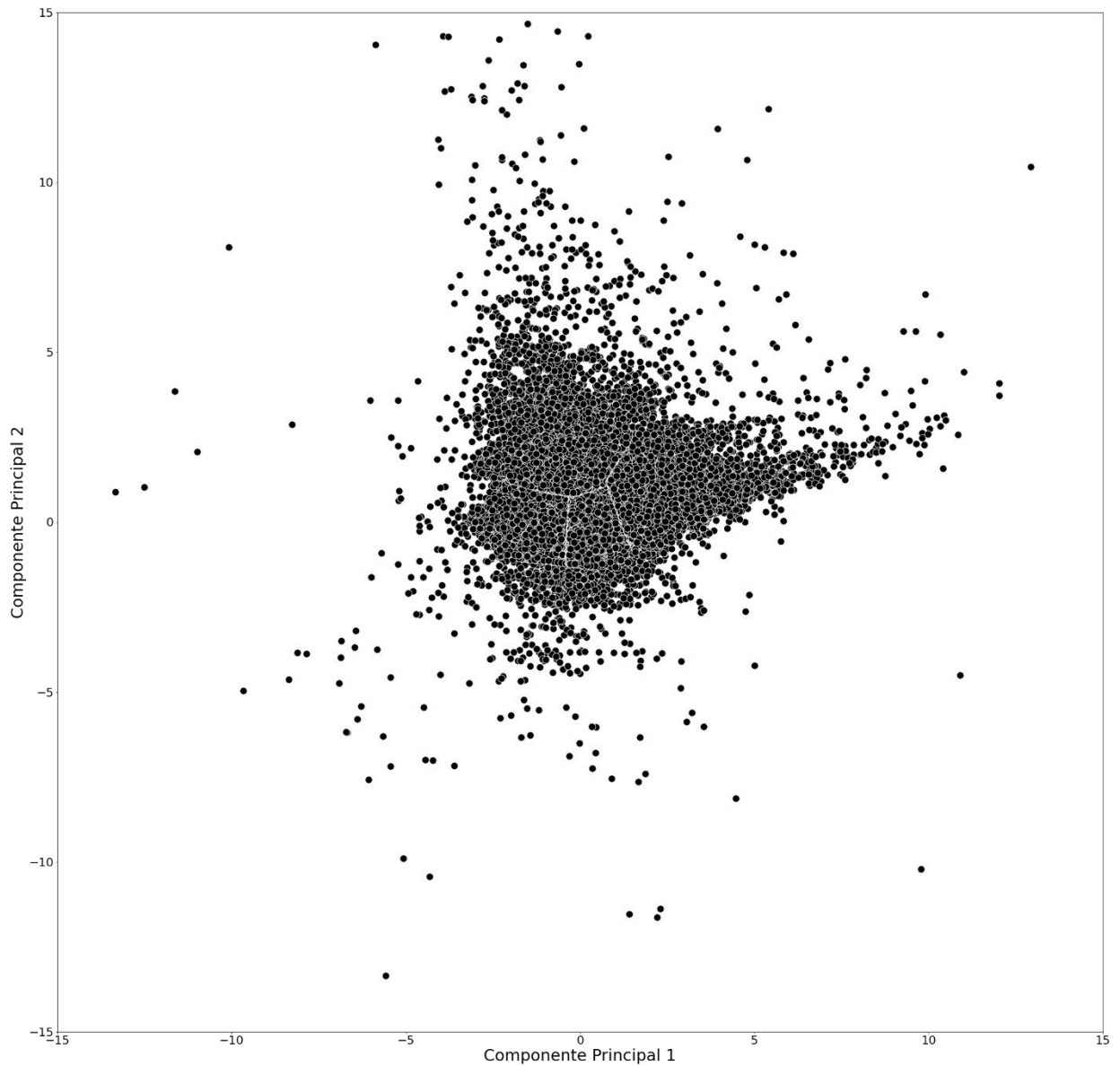


Figura 27: Representación 2D de los dos componentes principales seleccionados.

5.3.2. Segmentación de pacientes por Kmeans:

La información utilizada como entrada para el modelo de segmentación, corresponde a la tabla compuesta por los dos componentes principales que fueron seleccionados a partir de la reducción de dimensionalidad.

La técnica de segmentación utilizada en el proyecto corresponde a Kmeans, cuyo objetivo es segmentar el conjunto de datos en grupos más pequeños llamados clúster. Mediante esta técnica se espera obtener una asignación de clúster para cada uno de los pacientes del dataset.

Para aplicar la técnica de Kmeans, primero será necesario definir la cantidad de clústers óptima para el sistema. Para ello se emplea el Método del Codo, técnica que simula la distancia intra-clúster promedio del sistema con diferentes valores de k y los representa mediante una visualización 2D con la que se pretende identificar en qué valor de k se genera un quiebre o “codo” en la curva. Este quiebre representa el punto óptimo de clústers del sistema, el cual corresponde al menor valor de k donde ya no se perciba una mejora sustancial al aumentar el valor de k .

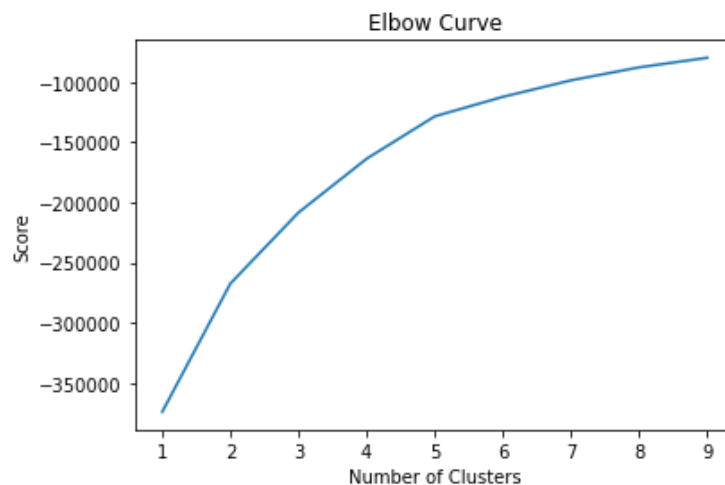


Figura 28: Técnica del método del codo aplicado al proyecto.

En la Figura 28 se observa el método del codo aplicado al proyecto, donde se ve que no existe un quiebre evidente en la curva, siendo las mejores opciones de selección los valores 4 o 5. Con el valor de $k = 5$ seleccionado para el proyecto, se procede con la visualización de los datos. Como el objetivo de este indicador es evaluar los grupos de pacientes que generan mayores ingresos, se decide agregar el Valor de Venta como tercer componente del gráfico, convirtiendo el indicador en una visualización 3D, como se ve en la Figura 29.

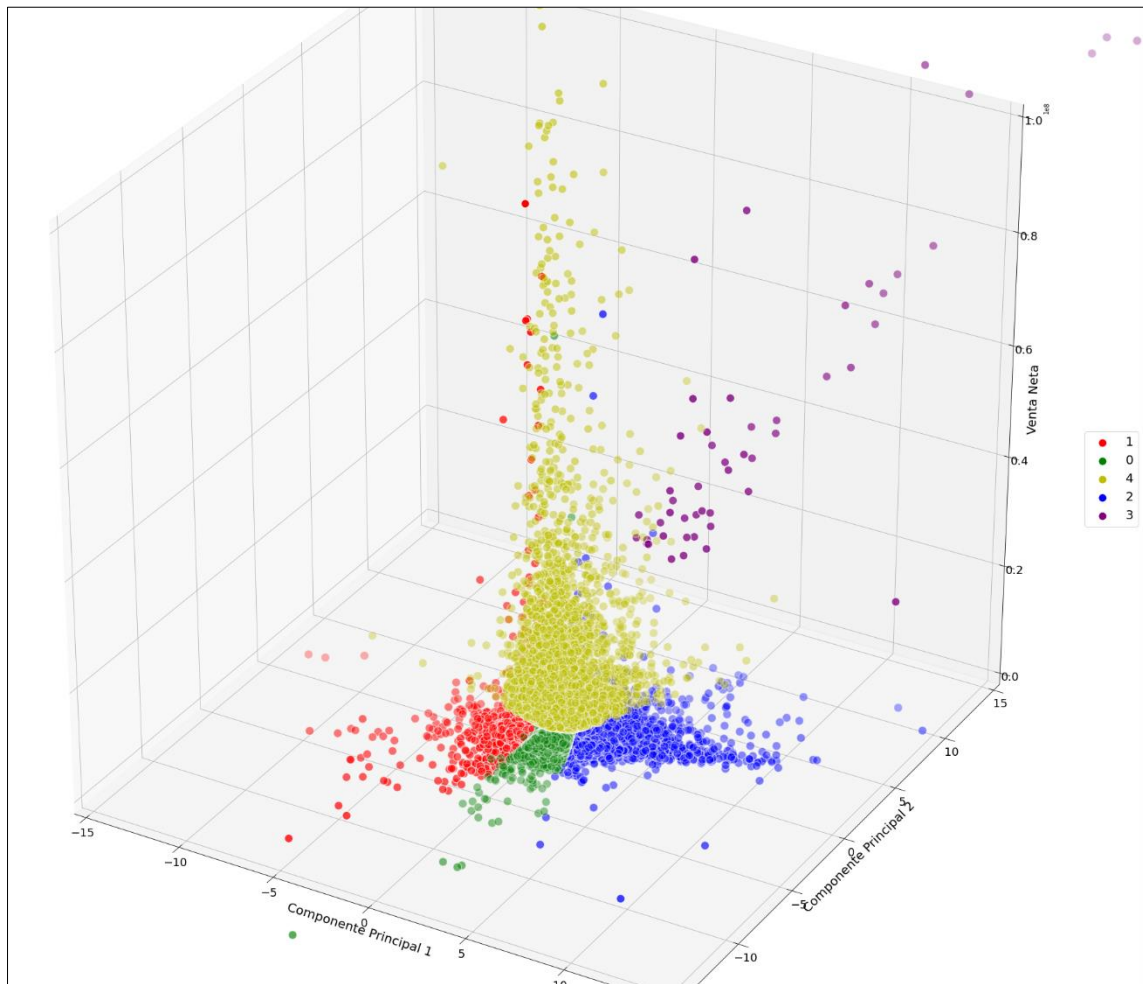


Figura 29: Clusterización de datos por Kmeans con cantidad de clústers igual a 5.

La visualización de la Figura 29, servirá para identificar cuáles son los distintos grupos de pacientes y cuáles de ellos son los que generan mayores ingresos, que, para el caso de estudio, corresponderían al grupo 4 (amarillo) y al grupo 3 (morado). Por el contrario, los grupos de pacientes que generarían menos ingresos corresponderían a los grupos 0 (rojo), 1 (azul) y 2 (verde).

5.3.3. Descripción de clústers de pacientes:

Luego de identificar los grupos de pacientes que generan mayor rentabilidad, se realiza un análisis para determinar cuáles son las características de estos grupos, y que hacen que se generen mayores o menores ingresos.

A continuación, se presenta una descripción resumen que entrega información sobre cada uno de los clústers de la Figura 29:

- Grupo 0: Es el clúster que genera menores ingresos en promedio. Se caracteriza por que sus pacientes suelen realizar 1 sola atención ambulatoria en promedio, lo cual le otorga la menor venta entre los grupos.

Grupo	Edad	Valor Neto	Atenciones Ambulatorias	Atenciones Hospitalarias	Atenciones Urgencias
0	29	\$81.582	1	0	0

Tabla 3: Valores promedios para el grupo 0.

Grupo	% Banmédica	% Consalud	% Colmena	% Cruz Blanca	% Fundación	% Nueva Mas Vida	% Vida Tres	% Fonasa	% Particular
0	11.8	23.5	15.1	11.0	0.2	3.2	3.1	19.8	12.1

Tabla 4: Distribución de aseguradoras de salud en el grupo 0.

Grupo	% Mujeres	% Hombres
0	44.6	55.4

Tabla 5: Distribución de sexo en el grupo 0.

- **Grupo 1:** Es el segundo grupo con menor venta promedio del sistema. Se caracteriza por que sus pacientes suelen realizar en promedio 2 atenciones ambulatorias y la mayoría de sus pacientes pertenecen a Fonasa. Además, debido al bajo valor neto promedio, se puede concluir que este tipo de pacientes tiende a realizarse atenciones ambulatorias simples como consultas, exámenes de laboratorio, imágenes u otros.

Grupo	Edad	Valor Neto	Atenciones Ambulatorias	Atenciones Hospitalarias	Atenciones Urgencias
1	56	\$147.225	2	0	0

Tabla 6: Valores promedios para el grupo 1.

Grupo	% Banmédica	% Consalud	% Colmena	% Cruz Blanca	% Fundación	% Nueva Mas Vida	% Vida Tres	% Fonasa	% Particular
1	8.6	18.3	11.1	9.2	0.2	3.4	5.0	44.4	4.3

Tabla 7: Distribución de aseguradoras de salud en el grupo 1.

Grupo	% Mujeres	% Hombres
1	46.1	53.9

Tabla 8: Distribución de sexo en el grupo 1.

- **Grupo 2:** Es el tercer grupo con menor venta promedio del sistema. Se caracteriza por que sus pacientes suelen realizar en promedio 1 atención en urgencia y la mayoría de sus pacientes pertenecen a Consalud.

Grupo	Edad	Valor Neto	Atenciones Ambulatorias	Atenciones Hospitalarias	Atenciones Urgencias
2	24	\$420.803	0	0	1

Tabla 9: Valores promedios para el grupo 2.

Grupo	% Banmédica	% Consalud	% Colmena	% Cruz Blanca	% Fundación	% Nueva Mas Vida	% Vida Tres	% Fonasa	% Particular
2	9.0	44.5	18.1	13.2	0.0	1.8	2.6	3.8	5.7

Tabla 10: Distribución de aseguradoras de salud en el grupo 2.

Grupo	% Mujeres	% Hombres
2	49.5	50.5

Tabla 11: Distribución de sexo en el grupo 2.

- **Grupo 3:** Es el grupo con mayor venta promedio del sistema. Se caracteriza por su alta tasa de atenciones hospitalarias, junto con 1 atención ambulatoria promedio. Se podría concluir que este tipo de paciente de edad avanzada asiste a la clínica para atención de consultas médicas que derivan en algún tipo de cirugía compleja con hospitalización.

Grupo	Edad	Valor Neto	Atenciones Ambulatorias	Atenciones Hospitalarias	Atenciones Urgencias
3	61	\$13.180.412	1	16	0

Tabla 12: Valores promedios para el grupo 3.

Grupo	% Banmédica	% Consalud	% Colmena	% Cruz Blanca	% Fundación	% Nueva Mas Vida	% Vida Tres	% Fonasa	% Particular
3	0	7.9	2.0	72.5	2.0	7.8	0	5.9	2.0

Tabla 13: Distribución de aseguradoras de salud en el grupo 3.

Grupo	% Mujeres	% Hombres
3	29.4	70.6

Tabla 14: Distribución de sexo en el grupo 3.

- **Grupo 4:** Es el segundo grupo con mayor venta promedio del sistema. Se caracteriza por que sus pacientes suelen realizar en promedio 2 atenciones ambulatorias, pero a diferencia del Grupo 1, la mayoría de sus pacientes pertenecen a Consalud. Debido al alto valor neto promedio, se podría concluir que este tipo de pacientes tiene a realizarse procedimientos o cirugías de baja complejidad que se consideren ambulatorias.

Grupo	Edad	Valor Neto	Atenciones Ambulatorias	Atenciones Hospitalarias	Atenciones Urgencias
4	46	\$3.954.008	2	0	0

Tabla 15: Valores promedios para el grupo 4.

Grupo	% Banmédica	% Consalud	% Colmena	% Cruz Blanca	% Fundación	% Nueva Mas Vida	% Vida Tres	% Fonasa	% Particular
4	6.8	31.0	13.0	17.6	0.3	3.7	2.9	17.8	5.1

Tabla 16: Distribución de aseguradoras de salud en el grupo 4.

Grupo	% Mujeres	% Hombres
4	46.1	53.9

Tabla 17: Distribución de sexo en el grupo 4.

Mediante estas tablas es posible caracterizar cada uno de los grupos resultantes de la clusterización, donde se pueden identificar valores promedio como la edad, la cantidad de atenciones y las aseguradoras con mayor participación en cada segmento.

6. CONCLUSIONES

6.1. Conclusiones

El presente proyecto comienza con el estudio de softwares empresariales, presentando su función, origen y como derivan en los dashboard que actualmente se utilizan para medición de desempeño. A partir de la premisa sobre la ventaja competitiva que genera el uso de dashboard para la gestión de una institución, es que se propuso la implementación de un panel de visualización web para una institución de salud.

Dentro de la metodología para llevar a cabo el proyecto, se abordó la descripción y preparación de los indicadores de desempeño utilizados actualmente por la clínica. Esta etapa del proyecto consumió una gran parte del tiempo destinado a programación, debido a la limpieza, preparación y procesamiento de los datos para obtener los indicadores. Y no solo por el tiempo dedicado a la implementación, sino también por el tiempo que requiere en ejecutarse el código cada vez que se desean actualizar los datos, el cual abarca alrededor de 2 a 3 horas.

Luego, se presentó la metodología asociada a la implementación del servicio web, donde para realizar la conexión entre el IDE Jupyter Notebook y la base de datos Oracle, solo se requirieron los datos de acceso a la base de datos, otorgados por el área de TI de la empresa. El mayor impedimento para avanzar en esta etapa se debió a la falta de experiencia en el área y de cómo realizar la continuidad entre cada sistema de programación.

Durante de la presentación de resultados, se explicó cómo funciona el sistema de negocio de la clínica y como se ve afectado mediante distintos factores externos, como por ejemplo, el escenario generado a partir de la pandemia de covid-19.

Además de lo anterior, mediante el análisis de resultados se pudo visualizar varios de los indicadores propuestos, donde se determinaron las áreas, pacientes y aseguradores de salud que generan mayores ingresos a la clínica. A nivel de las aseguradoras, se tuvo que Isapre Consalud es quien genera mayores ingresos debido a que pertenece al mismo holding de la clínica. A nivel de grupo de pacientes se tuvo que en general los adultos mayores son los que generan mayores ingresos, debido a que suelen presentar patologías graves y que requieren mayores cuidados.

Para finalizar el análisis de resultados, se presentaron los indicadores obtenidos a partir del modelo de Segmentación de Pacientes. En primera instancia se analizó la varianza explicada de cada uno de los siete componentes principales que componen el sistema donde se identificó que, los dos primeros componentes requeridos para implementar el sistema de visualización 2D, solo contenían el 37% de la información total de los datos. A pesar de esto, se decidió continuar con la implementación de Kmeans.

Para representar el indicador de segmentación de pacientes se utilizaron los dos componentes previamente seleccionados como ejes X y Z, y la venta neta como eje Y. El resultado de esto fue un gráfico 3D clusterizado por Kmeans, donde se pudo identificar cuáles fueron los grupos que generan mayores ingresos.

Como resultado, se obtuvo que los grupos 3 y 4 son los tipos de pacientes que generan una mayor cantidad de ingresos en promedio, el primero de ellos caracterizado por ser pacientes de edad avanzada que suelen sufrir un alto número de cirugías y hospitalizaciones, mientras el segundo por ser un grupo compuesto mayoritariamente de pacientes Consalud, que asisten constantemente al servicio ambulatorio.

A grandes rasgos, el desarrollo del proyecto tiene la finalidad de facilitar y automatizar la extracción de información mediante indicadores, resultados y conclusiones, que se utilicen para la toma de decisiones, la medición de desempeño, la mejora de procesos y para encaminar a todas las áreas de la empresa en una misma dirección.

De entre todas las decisiones que se van tomando en la dirección de una empresa, la mayoría depende principalmente de la experiencia y la intuición de los dirigentes, sin embargo, mediante un sistema de indicadores como el que se presenta, será de mayor facilidad y confiabilidad el tomar decisiones para encaminar a una empresa.

6.2. Limitaciones:

Una de las limitaciones presentadas en el proyecto tiene que ver con el nivel de procesamiento y descarga de los datos. En primera instancia se estimó que el panel de visualización web fuera en tiempo real, sin embargo, la descarga de bases de datos considera un gran consumo de recursos de red para la institución, a nivel de tiempo y de ancho de banda.

Como consecuencia se toma la determinación de que la actualización se realice como un bash de manera diaria, en algún horario donde no exista mayor flujo de información ni se utilicen en gran medida los sistemas HIS.

La segunda de las limitaciones corresponde a la imposibilidad de acceder a las bases de datos de diagnóstico del paciente. Esto se debe a la Ley de deberes y derechos del paciente, donde se estipula que su información médica no sea entregada a personas no relacionadas a su atención

médica [31]. Esto quiere decir que no se puede acceder a estados de diagnóstico del paciente, como por ejemplo su peso corporal, estatura, resultados de laboratorio, enfermedades, etc.

Debido a esta limitación no es posible realizar análisis predictivos de alta precisión, como por ejemplo sería el estimar cuales son las atenciones que se realizará un paciente dado un diagnóstico o cuales son las posibilidades de que un paciente padezca una enfermedad o no.

La tercera y última de las limitaciones se relaciona con el modelo de segmentación de pacientes implementado. Dado que el modelo utilizado corresponde a un método de aprendizaje no-supervisado, es difícil poder verificar si los resultados del modelo son correctos o no. Esto se debe a que no existen etiquetas previas de los datos con que comparar los resultados para poder obtener una métrica de desempeño.

6.3. Trabajo a Futuro:

Con el servicio web ya diseñado y testeado por los usuarios clave de la institución, se pretende elevar el proyecto a nivel productivo. Esto tendrá varias implicancias a nivel jerárquico, ya que requerirá de gestiones y permisos entre las áreas involucradas de la institución, como por ejemplo, el área TI de la clínica, el área gerencial y algunas áreas contraparte del corporativo al que pertenece la empresa.

Luego de que el servicio se encuentre levantado, se procedería a una etapa de enriquecimiento del panel, donde se pudieran abarcar nuevas áreas a las que proveerles indicadores para gestionar sus procesos, como por ejemplo, áreas de gestión operacional, áreas de abastecimiento y farmacia, etc.

7. BIBLIOGRAFÍA

- [1]. Tabares, L. F., & Hernández, J. F. (2014). “Big Data Analytics: Oportunidades, Retos y Tendencias”.
- [2]. MA Mach y MS Abdel-Badeeh, (2010) "Técnicas inteligentes para la inteligencia empresarial en el cuidado de la salud", Décima Conferencia Internacional sobre Diseño y Aplicaciones de Sistemas Inteligentes, El Cairo, 2010, págs. 545-550, doi: 10.1109 / ISDA.2010.5687209. <https://ieeexplore.ieee.org/document/5687209>
- [3]. W. Eckerson, W. (2005). Performance Dashboards: Measuring, Monitoring, and Managing Your Business. <http://library.lol/main/39D0B84F1D6E7391E30DBE19B72A0F28>
- [4]. Colaboradores de Wikipedia. (2021, 25 octubre). Software empresarial. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Software_empresarial
- [5]. Andrade Garda, J., Riveira Otero, M., & Gómez Caderno, I. (2019). Pasado, presente y futuro del Cuadro de Mando Integral.
- [6]. Pérez Porto, J., & Gardey, A. (2010). “Definición de ambulatorio”. <https://definicion.de/ambulatorio/>
- [7]. “¿Qué es una atención de urgencia o emergencia vital?”. Superintendencia de Salud. <https://www.supersalud.gob.cl/consultas/667/w3-article-2981.html>
- [8]. Fonasa. “Tramos Fonasa 2020”. <https://www.fonasa.cl/sites/fonasa/tramos>
- [9]. Granel, M. (2020, 20 enero). ¿Qué es ISAPRE? ¿Cómo funcionan las Isapres? Rankia. <https://www.rankia.cl/blog/mejores-opiniones-chile/3890646-que-isapre-como-funcionan-isapres>
- [10]. Garantías Explícitas en Salud (GES). (2021). Orientación en Salud. Superintendencia de Salud, Gobierno de Chile. <https://www.supersalud.gob.cl/difusion/665/w3-propertyvalue-1962.html>
- [11]. INE. (2021). Características de la población. <https://www.ine.cl/ine-ciudadano/definiciones-estadisticas/poblacion/caracteristicas-de-la-poblacion>
- [12]. Palacios, D. (2021). “Cómo crear un reporte de ventas (incluye ejemplos)”. <https://blog.hubspot.es/sales/reporte-de-ventas>

- [13]. Tiangolo. (2018). FastApi Framework. <https://fastapi.tiangolo.com/>
- [14]. Sun Microsystems. (2007, 18 septiembre). What is a Server-Side Endpoint? The WSIT Tutorial. https://docs.oracle.com/cd/E17802_01/webservices/webservices/reference/tutorials/wsit/doc/Initialization2.html
- [15]. Template Designer Documentation — Jinja Documentation (3.0.x). (2021). Template Designer Documentation. <https://jinja.palletsprojects.com/en/2.11.x/templates/>
- [16]. 4 funciones avanzadas útiles en FastAPI. (2020, 26 noviembre). ICHI.PRO. <https://ichi.pro/es/4-funciones-avanzadas-utiles-en-fastapi-264493979768375>
- [17]. Otto, M. J. T. (2021). Build fast, responsive sites with Bootstrap. Bootstrap. <https://getbootstrap.com/>
- [18]. C. Müller, A., & Guido, S. (2016). Introduction to Machine Learning with Python. O'Reilly Media Inc.
- [19]. Wijaya, C. Y. (2021, 12 octubre). 5 Must-Know Dimensionality Reduction Techniques via Prince. Towards Data Science. <https://towardsdatascience.com/5-must-know-dimensionality-reduction-techniques-via-prince-e6ffb27e55d1>
- [20]. Prince. (2020, 6 octubre). PyPI. <https://pypi.org/project/prince/#factor-analysis-of-mixed-data-famd>
- [21]. Rodrigo, J. A. (2017, junio). Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE. Ciencia de Datos. https://www.cienciadedatos.net/documentos/35_principal_component_analysis
- [22]. Pedregosa, F. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [23]. Gonzales, A. (2019, 8 marzo). Segmentación utilizando K-means en Python. MachineLearningParaTodos.com. <https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/>
- [24]. Ricardo Moya García. (2016, septiembre). Machine Learning(en Python), con ejemplos. <https://docplayer.es/150852042-Machine-learning-en-python-con-ejemplos-ricardo-moya-garcia-phd-version-septiembre-2016.html>

- [25]. Biblioteca Digital. Superintendencia de Salud. Gobierno de Chile. (diciembre 2015). “Boletín Estadístico de Fonasa” [Conjunto de datos]. https://www.supersalud.gob.cl/documentacion/666/articles-13905_recurso_1.xlsx
- [26]. Biblioteca Digital. Superintendencia de Salud. Gobierno de Chile. (diciembre 2015). “Boletín Estadístico de Isapres” [Conjunto de datos]. <https://www.ine.cl/ine-ciudadano/definiciones-estadisticas/poblacion/caracteristicas-de-la-poblacion>
- [27]. Superintendencia de Salud. (2013, diciembre). Prestadores de Salud, Isapres y Holdings: ¿Relación Estrecha? https://www.supersalud.gob.cl/documentacion/666/articles-8826_recurso_1.pdf
- [28]. Carvajal G., C. (2021, marzo 31). Funcionarios de Salud denuncian total colapso en hospitales de la Región Metropolitana. Diario UChile. <https://radio.uchile.cl/2021/03/31/funcionarios-de-salud-denuncian-total-colapso-en-hospitales-de-la-region-metropolitana/>
- [29]. Editorial La República S.A.S. (2021, 10 junio). Santiago, la capital de Chile, se mueve hacia el cierre total con hospitales colapsados. Diario La República. <https://www.larepublica.co/globoeconomia/santiago-la-capital-de-chile-se-mueve-hacia-el-cierre-total-con-hospitales-colapsados-3183566>
- [30]. Mostrador, E. (2021, 4 febrero). Clínica Las Condes insiste al Minsal que le pague pacientes derivados por covid-19 y corre riesgo de ser intervenida. El Mostrador. <https://www.elmostrador.cl/dia/2021/02/04/clinica-las-condes-insiste-al-minsal-que-le-pague-pacientes-derivados-por-covid-19-y-corre-riesgo-de-ser-intervenida/>
- [31]. Ley de derechos y deberes. (2012, octubre). Superintendencia de Salud, Gobierno de Chile. <https://www.supersalud.gob.cl/consultas/667/w3-propertyvalue-4185.html>