



Analítica de datos para la optimización del agendamiento de citas médicas

Andrés Alejandro Rodríguez Lozano

Universidad Jorge Tadeo Lozano
Facultad de Ciencias Naturales e Ingeniería,
Departamento de Ingeniería de Sistemas
Bogotá, Colombia
2020

Analítica de datos para la optimización del agendamiento de citas médicas

Andrés Alejandro Rodríguez Lozano

Tesis de grado presentada como requisito parcial para optar al título de:
Magister en ingeniería y analítica de datos

Director:
Ph.D., Sebastián Zapata Ramírez

Universidad Jorge Tadeo Lozano
Facultad de Ciencias Naturales e Ingeniería,
Departamento de Ingeniería de Sistemas
Bogotá, Colombia
2020

Resumen

E Call You S.A.S. es una empresa dedicada a dar soluciones como centro de contactos y BPO (*Bussiness process outsourcing*), que busca mejorar las capacidades que ofrece a sus principales usuarios en el sector salud, dentro de los que se encuentran: clínicas, IPS (Institución prestadora de servicios) y centros de salud en su modelo de atención a negocios, teniendo en cuenta al consumidor final, en este caso los pacientes de citas médicas, que requieren un trato humano, amable y ágil para solucionar sus necesidades relacionadas con la salud.

Con el fin de contribuir al desarrollo de labores en el modelo de negocio de E Call You, se pretende obtener un modelo de analítica óptimo y eficiente, para lograrlo se utilizan herramientas de la analítica de datos, que permiten realizar una segmentación adecuada de los grupos de pacientes recurrentes de los centros médicos y a su vez, desarrollar un modelo predictivo que refleje el comportamiento de asignación de citas médicas acorde con el perfil del paciente y las necesidades de este.

En el desarrollo del proyecto se tiene en cuenta la captura y gestión de información en un entorno de bodega de datos, junto a herramientas para el análisis de estos datos que le permitan a la compañía mostrarse como líder del agendamiento en el sector salud. Por lo anterior, se llevan a cabo modelos de Aprendizaje automático de agrupamiento de variables (*Clustering*) diseñados para la segmentación de los pacientes, como también, se pronostica los volúmenes de agendamiento de citas por especialidad.

Palabras clave: Paciente, Servicios médicos, Segmentación, análisis predictivo, optimización, visualización de datos.

Abstract

E Call You S.A.S. is a company who works as a contact center solutions and BPO (business process outsourcing) seeks to improve the capabilities it offers to its main clients in the health sector like health medical centers, in its business to business model, taking into account the final consumer, in this case medical appointment patients, who have a humane, friendly and agile treatment to solve their health-related needs.

In order to contribute to the E Call You business model, the aim is to obtain a more effective and efficient model, to ensure that data analysis tools are used, which allow adequate segmentation of groups of recurring patients from criticism and in turn, develop a predictive model that reflects the behavior of medical appointment allocation according to the patient's profile and patient needs.

To develop this project, includes the capture and management of information in a data warehouse environment that is taken into account, along with tools for the analysis of this data that will allow it to show the company as the program leader in the health sector. Therefore, machine learning models of clustering variables will be carried out for the segmentation of patients, as well, the scheduling volumes of appointments by specialty are forecast.

Finally, web views registered on the company page were designed, as an added value to the management of the contact center, it was selected to monitor and control the campaign by customers and leaders of the operation, delivering daily information updates.

Keywords: Patient, health care, consumer, clustering, predictive analysis, data optimization, data visualization.

Contenido

Resumen	IV
Lista de figuras	IX
Lista de tablas	IX
1 Introducción	2
2 Metodología	5
2.1 Estudio teórico del problema	5
2.2 Almacenamiento centralizado de la información: Implementación e integración de las fuentes de datos relacionales	5
2.3 Minería de datos	6
2.4 Estudio teórico del problema: Segmentación de los pacientes y optimización de procesos internos de contacto y comunicación, utilización de modelos de aprendizaje automático para el análisis de datos.	6
2.5 Aprendizaje automático para la predicción de necesidades médicas	7
2.6 Visualización de la información	7
3 Planteamiento del problema	8
3.1 Justificación	8
3.2 Objetivo general	9
3.3 Objetivos específicos	9
4 Marco teórico	10
4.1 Modelos analíticos	10
4.2 Aplicaciones en centros de contacto que prestan servicios de atención al usuario y agendamiento de citas médicas	16
5 Desarrollo del trabajo	20
5.1 Estructura de los datos	20
5.2 Inteligencia de negocios	26
5.3 Analítica predictiva y aprendizaje automático	30
5.4 Desarrollo de producto	41
5.5 Progreso	46

6 Conclusiones	47
7 Referencias bibliográficas	49

Lista de Figuras

4-1	Técnicas de aprendizaje automático, elaboración propia con datos de [14]. . .	13
4-2	CRISP-DM para minería de datos, elaboración propia con información [22]. .	16
5-1	Pipeline de datos, Elaboración propia con información de [22].	20
5-2	Estructura de flujo de datos en aplicativos para E Call You S.A.S., elaboración propia basado en [22].	22
5-3	Datos relevantes gestión Clínica de la Colina, archivo personal elaborado para E Call You S.A.S.	24
5-4	Gestión histórica de interacciones por intervalos de 15 minutos, archivo personal elaborado para E Call You S.A.S.	27
5-5	Oportunidad en la asignación de citas médicas, archivo personal elaborado para E Call You S.A.S.	28
5-6	Volumen de agendamiento de citas médicas, archivo personal elaborado para E Call You S.A.S.	29
5-7	Correlación de variables con mapa de calor, archivo personal elaborado para E Call You S.A.S.	30
5-8	Clúster de variables usando K-medios con gráfica de R en Power BI, archivo personal elaborado para E Call You S.A.S.	33
5-9	Ficha de segmentación de usuarios de acuerdo con agrupamiento por clústeres, archivo personal elaborado para E Call You S.A.S.	36
5-10	Pronóstico de llamadas por fecha, modelo Auto Regresivo de orden 10, archivo personal elaborado para E Call You S.A.S.	37
5-11	Pronóstico de llamadas por fecha, modelo Auto Regresivo con promedio móvil de orden 10, archivo personal elaborado para E Call You S.A.S.	38
5-12	Pronóstico de llamadas por fecha, modelo TBATS de orden 10, archivo personal elaborado para E Call You S.A.S.	39
5-13	Formulario de registro usuario y contraseña para cliente interno y externo, archivo personal elaborado para E Call You S.A.S.	42
5-14	Gestión transaccional, indicadores de atención y abandono, atención de llamadas por intervalos, archivo personal elaborado para E Call You S.A.S. . .	43
5-15	Inteligencia de negocios e indicadores de efectividad Clínica de la Colina, archivo personal elaborado para E Call You S.A.S.	44

Lista de Tablas

4-1	Algoritmos más usados para el modelamiento con Aprendizaje automático, elaboración propia basado en [8, 9 y 10].	12
5-1	Agendamientos E Call You S. A. S. Diccionario de datos, Archivo personal elaborado para E Call You S.A.S.	25
5-2	Transacciones E Call You S. A. S. Diccionario de datos, Archivo personal elaborado para E Call You S.A.S.	26
5-3	Variables seleccionadas para el agrupamiento de segmentos, Archivo personal elaborado para E Call You S.A.S.	31
5-4	Segmentos de pacientes generados, Archivo personal elaborado para E Call You S.A.S.	35
5-5	Resumen de documentos generados en el desarrollo del producto, Archivo personal elaborado para E Call You S.A.S.	45
5-6	Resumen avance del proyecto, Archivo personal elaborado para E Call You S.A.S.	46

1 Introducción

Durante los más de 10 años de operación de E Call You S.A.S. como compañía prestadora de servicios de gestión de contacto, servicio y agendamiento de necesidades médicas para las Clínicas Country y la Colina, ha realizado una diligente ejecución de actividades con buenos resultados; pero se evidenció recientemente la necesidad de almacenar la información de una manera eficiente, entenderla y procesarla garantizando la manipulación correcta de datos personales y generando información clave para la toma de decisiones de la compañía a través de la implementación de tecnologías de información [1].

En el desenvolvimiento de la gestión de la compañía han aparecido ineficiencias en la forma en que se captura la información proveniente de la operación, los resultados procesados y entregados por el equipo de datos a los clientes internos y externos y debido a esto, es necesario continuar avanzando en el análisis y uso de herramientas integrales (diferente al uso de archivos de Excel con ligeros análisis descriptivos y estáticos) que permitan consolidar la información y a la que se pueda acceder en un almacenamiento omnipresencial en la nube o los servidores de la compañía, usando motores de bases de datos SQL donde se pueda extraer datos con consultas rápidas, sin remitirse a archivos que se encuentran en un almacenamiento descentralizado ubicado en los equipos de cómputo o discos duros externos del departamento de datos.

Este departamento se compone tradicionalmente por personal con el cargo de “*Datamarshall*”, que se limitaba a gestionar día a día las necesidades operativas del centro de contacto sin tener las capacidades y el tiempo suficiente para desarrollar proyectos de mayor envergadura que generen valor a la compañía y sus clientes.

El propósito de este trabajo es optimizar la productividad de la compañía E Call You S.A.S. en el agendamiento de citas médicas mediante la implementación de un modelo de analítica de datos que logre segmentar y entender el comportamiento de los pacientes, así como pronosticar el volumen de agendamiento.

Con el fin de perfeccionar el servicio y optimizar el agendamiento de citas médicas se necesita tomar un rol activo en la gestión de información, mejorando su almacenamiento y generando una segmentación que permita entender las características de los usuarios, necesidades médicas del grupo objetivo, tendencias en la asignación de citas y preferencias en el contacto

con las Clínicas.

Es importante comprender los factores externos e internos que influyen sobre la operación de la compañía, esta información contribuye a una mejora en los análisis que se llevan a cabo actualmente. Se debe determinar cuáles son las características de los usuarios que tienen relación con la clínica, cuánto tiempo están dispuestos a esperar para recibir atención médica, cuánto dinero pueden pagar, la influencia que puede tener la zona desde la cual se desplazan para asistir a una cita médica; el impacto que puede tener la coyuntura actual en la decisión para acudir a una cita médica de manera presencial; ¿éstos pacientes desean tener más puntos de contacto o continúan prefiriendo el agendamiento mediante una línea telefónica?; ¿el grupo etario de estos consumidores los hace afines a otro tipo de contacto más tecnológico?; con lo anterior, es posible segmentar a los consumidores de servicios de salud y ofrecerles un servicio óptimo y de calidad que se traduzca en mejores niveles de satisfacción asignados por el paciente y un mayor número de agendamientos, mejores gestiones para las clínicas y empresas prestadoras de servicios de salud.

En cuanto a factores internos, se puede determinar cuáles son los asesores con mejores resultados para abordar a los distintos tipos de pacientes; del mismo modo, observar en qué franjas horarias hay una mayor recepción de llamadas y cómo influye en la operación y con ello identificar qué asesores tienen un mejor desempeño; si dadas las características de los usuarios, es una buena estrategia contactar a los pacientes por mensajes de texto, mensajería instantánea, llamada telefónica o correo; si se debe tener en cuenta la ubicación de los pacientes y su relación geográfica con las sedes de las clínicas.

El uso de tecnologías de ciencia de los datos permite dar valor a la información más allá de almacenarla y presentar una fotografía de lo acontecido; usando técnicas de estadística descriptiva frente a lo que ya ocurrió en la operación. Con el aprendizaje automático se busca usar la capacidad computacional y herramientas de visualización con las que cuenta la compañía para evidenciar tendencias en los grupos de datos, que permitan entender al paciente y encontrar predicciones frente al resultado del comportamiento de la categoría.

La solución a las problemáticas planteadas anteriormente consiste en la utilización del siguiente modelo de analítica para E Call you S.A.S:

1. Estudio teórico: realizar un estudio de las necesidades, casos de uso y soluciones teóricas del problema.
2. Almacenamiento centralizado de la información: Organizar la información de forma confiable en un entorno de Big Data mediante una bodega de datos para tener acceso inmediato a la información histórica de la campaña.

3. Minería de datos: Usar la información histórica para monitorear cómo se desarrolla la gestión de agendamiento de citas médicas y con ello, tomar decisiones de forma oportuna e identificar posibles alertas frente a la gestión de las necesidades de la campaña manteniendo la excelencia en los KPI (*key performance indicator*) claves de la operación.

4. Segmentación: Generar analítica de datos con modelos que examinen la mejor forma de agrupar los pacientes de acuerdo con sus características sociodemográficas y comportamentales frente al interés por los servicios médicos.

5. Aprendizaje automático: Pronosticar los volúmenes de agendamiento de citas médicas por especialidad para dimensionar efectivamente la cantidad de asesores para la atención a los pacientes.

6. Visualización de la información: Exponer de forma clara los resultados de los procesos de analítica de datos desarrollados en etapas anteriores del proyecto con el fin de soportar la toma de decisiones dentro de la empresa y ver dinámicamente los resultados de la gestión a los interesados.

2 Metodología

El proyecto se soluciona en 6 etapas mencionadas anteriormente, se realizan de forma lineal, usando durante su implementación metodologías ágiles [2] para lograr resultados con mejoras iterativas en cortos periodos de tiempo.

2.1. Estudio teórico del problema

1. Estudio de los conceptos básicos relacionados con Estadística, Big Data, inteligencia artificial, aprendizaje automático y herramientas útiles para la implementación de proyectos de analítica de datos.
2. Estudio de los conceptos y aplicaciones en la gestión de servicio al paciente, agendamiento de salud, optimización en el servicio y comportamiento de los usuarios de servicios médicos.

2.2. Almacenamiento centralizado de la información: Implementación e integración de las fuentes de datos relacionales

En primer lugar, se organizan los documentos relacionados con la gestión y agendamiento de citas médicas, información sociodemográfica de los pacientes del año 2019 y lo que va corrido de 2020 en una base de datos materializada que permita la consulta rápida y extracción de información en los modelos de analítica.

Con el fin de alcanzar el objetivo principal, se crean bases en SQL, alimentadas directamente por el proveedor de aplicativos para desempeñar labores en el centro de contacto mediante disparadores automatizados para enviar información en horas de la madrugada con datos del día anterior, permitiendo el manejo de grandes volúmenes de datos en bases transaccionales que ofrecen métodos sencillos de actualización y centralización de datos.

De ahí, que esta etapa del proyecto es clave para generar consultas de información de forma ágil, automatizable y diferencial en la forma de extraer la información, comparado con la

forma en que tradicionalmente se realiza y comparte información de gestión de datos.

En consecuencia, al implementar un entorno de información en la nube para la recolección, almacenamiento y análisis de datos es necesario realizar un correcto diseño de la base de datos, de acuerdo con las variables claves para entender el comportamiento de la gestión de la campaña.

Por lo tanto, se debe materializar esos datos que se disponían en tablas de Excel migrándola a información gestionada en bases SQL, obtener la jerarquía de información y organización de datos en los servidores de la compañía y gestionar un correcto manejo en calidad de la información contenida usando diccionario de datos y manuales de trabajo referentes a las bases de datos, tablas, vistas y reportes.

2.3. Minería de datos

En esta etapa se logrará conformar un panorama completo de la estructura que contienen los datos frente a su tipo, dimensiones, forma, características, como también el posible uso que se le puede dar mediante la limpieza y exploración de datos en Python.

1. Exploración completa de datos con visualizaciones descriptivas.
2. Limpieza de datos.
3. Entender las tendencias de los datos.

2.4. Estudio teórico del problema: Segmentación de los pacientes y optimización de procesos internos de contacto y comunicación, utilización de modelos de aprendizaje automático para el análisis de datos.

1. Adelantar un análisis de los datos priorizando las características sociodemográficas de los pacientes, los servicios médicos que solicitan, el desembolso de dinero para su pago y los métodos de contacto que prefieren.

1. Crear reportes interactivos que permitan a los líderes de la compañía conocer de forma recurrente el comportamiento de la campaña y percibir alertas en la gestión cuando las cosas no estén operando bien.

3. Diseñar la planta óptima necesaria para atender a los pacientes que se comunican y sus

necesidades sin disminuir los porcentajes de operación pactados como metas de los KPI's de la campaña.

2.5. Aprendizaje automático para la predicción de necesidades médicas

A fin de diseñar un análisis prescriptivo se usan métodos predictivos de Aprendizaje automático que deduce el comportamiento histórico de la campaña; se puede predecir en este caso qué especialidades serán más demandas para agendamiento de citas en las siguientes semanas.

Así mismo, entender cómo se generarán los beneficios de la compañía frente a los acuerdos de operación y niveles de satisfacción pactados, logrando encontrar las mejoras que se deben considerar en la operación; para efectuar compromisos y resultados inmediatos.

El proceso de analítica se realizará de la siguiente manera:

1. Entender la estructura interna de los datos.
2. Conectar las bases de datos que pueden generar información adicional útil de los pacientes recurrentes de la clínica.
3. Implementar nuevas variables que permitan definir modelos de clasificación y construir pronósticos frente al volumen de llamadas y agendamiento de citas que se realizan diariamente.
4. Ejecutar la validación de los modelos para garantizar el ajuste de las predicciones.

2.6. Visualización de la información

Se crea un acceso a la información en la página web de la compañía para los líderes administrativos por parte del cliente externo y los gerentes, directores y jefes de operación, permitiéndoles visualizar los resultados del proceso de analítica de datos.

3 Planteamiento del problema

E Call You S.A.S. ofrece soluciones como centro de contactos y servicios de BPO (*Business process outsourcing*) especializada en el servicio al usuario y agendamiento en salud, cuenta con una estrecha relación con clínicas y compañías que desarrollan actividades en este sector, las cuales en la operatividad gestionan datos de contacto, información sociodemográfica y necesidades médicas de sus pacientes.

Tradicionalmente, la empresa se ha limitado a la generación de reportes descriptivos y utilización limitada de los datos que recibe, realizando pequeños análisis y guardando la información en documentos independientes, sin tener la capacidad de explotar los hallazgos de esta información para convertirse en un actor relevante, un aliado estratégico y apoyar con información que aporte en la toma de decisiones del negocio.

Hoy cada vez más empresas que laboran en el mundo del servicio al usuario y el agendamiento de citas de salud quieren usar modelos analíticos para optimizar su desempeño, pero muchas de las compañías que intentan hacerlo se limitan al uso de Bots conversacionales, reduciendo el uso de asesores de servicio en muchas ocasiones.

Este tipo de soluciones de inteligencia artificial y apoyo a la gestión telefónica también se implementan en la compañía, pero en los meses recientes se ha visto la necesidad de usar aprendizaje automático, con modelos de analítica que apoyen el desarrollo del negocio con una gestión optimizada frente a los objetivos de negocio de la compañía posicionando así a la organización con un perfil innovador, que le permita diferenciarse de otras empresas que operan en el sector.

3.1. Justificación

La compañía E Call You S.A.S. en el último año ha gozado un proceso de innovación, reinvención de las técnicas, métodos y mejora en la forma en la que se solucionan las necesidades de los clientes. Para lo anterior, se plantea adquirir un conocimiento profundo de las personas que están detrás de los datos (pacientes) y quienes tramitan la atención y servicio al usuario (asesores), en el agendamiento de citas médicas, con el fin de mejorar la eficiencia, productividad de la compañía, el entendimiento del paciente de esta forma generando valor

agregado al cliente de la compañía.

En consecuencia, se desarrolla un modelo pronóstico y segmentación usando analítica de datos para mejorar la gestión de campaña, que promueve generar un entendimiento inteligente y eficiente de la gestión.

3.2. Objetivo general

Desarrollar herramientas de inteligencia de negocios que mediante la implementación de modelos de analítica de datos para segmentar, entender el comportamiento de los pacientes y pronosticar el volumen de contactos para agendamiento de citas médicas, permitan optimizar la productividad de la compañía E Call You S.A.S.

3.3. Objetivos específicos

- Implementar y dar seguimiento a un entorno de “*Big Data*” que permita el almacenamiento de la información en Bodega de datos en el lenguaje SQL, accediendo de forma eficiente a los datos.
- Desarrollar dos modelos de Aprendizaje automático para incrementar y mejorar el entendimiento y producción del agendamiento de citas en salud.
- Integrar en una aplicación web estos desarrollos para dar valor agregado a la compañía permitiendo a los clientes monitorear la gestión en agendamiento de citas en el sector salud.

4 Marco teórico

4.1. Modelos analíticos

En la búsqueda de la comprensión y contextualización de la información relevante referente al campo de estudio de este trabajo, se tienen en cuenta conocimientos de áreas de estudio como economía, estadística y analítica de datos, que en conjunto contribuyen al cumplimiento de los objetivos planteados en el documento.

Este estudio parte de entender la Microeconomía como eje fundamental del comportamiento de los agentes que intervienen en el mercado; centrándose específicamente en el consumidor (consumidor de servicios de salud o que se halla bajo atención o tratamiento médico, definido paciente [3]) que experimenta necesidades a solucionar, en un entorno en el que intervienen otros agentes, como: El productor, que ofrece bienes para solventar la demanda en un mercado imperfecto, debido a múltiples asimetrías que aparecen en el mercado, tales como: problemas en la comunicación o información incompleta, desplazamientos largos y altos precios de servicios que afectan a los individuos del mercado que buscan generar utilidad (productores u oferentes) para su propio ser y ofrecer beneficios para el consumidor [4].

A causa de lo anterior, el estudio de la Economía del Comportamiento es clave en el desarrollo de este documento, puesto que busca identificar las diferentes variables que afectan la toma de decisiones de los seres humanos, teorías como lo es la aversión al riesgo, que percibe más atractivas las ganancias que pueden obtener en las decisiones que realiza frente a las pérdidas y castigos que recibiría, a pesar de tener el mismo valor y que caracterizan su comportamiento [5].

Para lograr agrupar a los consumidores de acuerdo con su comportamiento, características sociodemográficas compartidas, necesidades a atender y respuesta frente a los incentivos propuestos, es necesario recurrir a métodos de estudio que puedan encontrar diferencias en los grupos de pacientes que se deseen analizar y comparar. Es indispensable utilizar técnicas para el análisis de datos como la regresión cuantílica, que se encarga de realizar modelos econométricos de estudio, que conserva las diferencias entre los grupos, clúster o audiencias de pacientes, y permite generar hallazgos extrapolables dentro de individuos que comparten ciertas características específicas referentes a la zona de la ciudad donde viven, al nivel de ingresos, las patologías preexistentes en sus historias clínicas, por ejemplo [6].

En efecto, la elección de los algoritmos para crear, entrenar el modelo de aprendizaje y generar información relevante se realiza de acuerdo con clasificaciones de alto nivel que están relacionados con la cantidad de datos históricos conocidos previamente, esto permitiría la generación de modelos Supervisados, es decir, se conoce el comportamiento histórico de los datos [7]. Por el contrario, los modelos no supervisados se realizan cuando contamos con menos información sobre las variables.

Si se desea predecir una variable continua, como un precio, temperatura, altura, peso, se podrían utilizar modelos supervisados como regresiones lineales o múltiples que permiten conocer los valores futuros de una variable dependiente, basado en el comportamiento previo de otras variables independientes y que afectan el desempeño de la variable dependiente. Otras técnicas supervisadas son los árboles de decisión, que utilizan condiciones lógicas en la búsqueda de categorizar y representar una serie de procedimiento en la resolución de problemas y bosques aleatorios que son una combinación de árboles de decisión [7].

De tal manera que, cuando se dispone de menos información sobre los datos se realizan modelos no supervisados con técnicas de agrupamiento o clustering, estos modelos permiten generar inferencias de acuerdo con la cercanía de los vectores; las técnicas más relevantes son las k-medios las cuales corresponden a una agrupación de datos en conjuntos, donde las observaciones pertenecen a grupos que tengan un valor medio cercano a la observación específica [8].

En cambio, con variables de tipo categóricas utilizadas para la clasificación de información, como la decisión de encasillar un correo como *spam* o no *spam*, se podría utilizar modelos supervisados de clasificación de información con técnicas como *K-nearest neighbors (KNN)* que estiman la probabilidad de que un elemento pertenezca a una clase o casilla que permita reconocimiento de patrones. La técnica de análisis no supervisado son las Regresiones Logísticas, que mediante transformaciones matemáticas permiten modelar la probabilidad de un evento como función de variables explicativas [9]. Las máquinas de vectores de soporte o SVM por sus siglas en inglés, son un conjunto de algoritmos útiles en la resolución de problemas de clasificación y regresión. Otra técnica de modelación supervisada es el *Naive Bayes* que es un clasificador probabilístico que usa el teorema de Bayes, para asumir la presencia o ausencia de una característica frente a otras, de acuerdo con la clase de variable usada [9].

De forma antagónica, los modelos no supervisados de variables categóricas, que disponen una menor cantidad de información histórica utilizan técnicas como el análisis de asociación que utilizan tablas de contingencia y matrices de correlación para definir las relaciones que existen entre variables categóricas estimando magnitudes en las que estas se corresponden.

La técnica *Hidden Markov Model* es un método estadístico que asume un modelo que busca determinar los parámetros ocultos de una cadena de parámetros observables. Estos valores encontrados se pueden emplear para llevar a cabo análisis sucesivos permitiendo el reconocimiento de patrones [10].

Relacionado con los modelos de aprendizaje automático, se debe tener en cuenta para este proyecto el entendimiento de series temporales para pronosticar el comportamiento futuro de variables de acuerdo a los acontecimientos vividos anteriormente. Estos modelos explican el hecho de que los puntos de datos tomados a lo largo del tiempo pueden tener una estructura interna (como la correlación automática, la tendencia o la variación estacional) que debe tenerse en cuenta [11]. Un modelo auto-regresivo integrado de promedio móvil o ARIMA, Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes [12].

Adicionalmente, frente a series de tiempo se estudia la Predicción de series temporales con múltiples estacionalidades utilizando TBATS (acrónimo de Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components), la cual permite entender que cada estacionalidad está modelada por una representación trigonométrica basada en la serie de Fourier. Una de las principales ventajas de este enfoque es que solo requiere 2 estados semilla, independientemente de la duración del período [13]. La tabla [4-1] resume los algoritmos y herramientas estadísticas más usados para el aprendizaje automático.

Tabla 4-1: Algoritmos más usados para el modelamiento con Aprendizaje automático, elaboración propia basado en [8, 9 y 10].

Tipo de variable	Modelos supervisados	Modelos no supervisados
Continua	Regresión lineal, Regresión múltiple, Series de tiempo, Árboles de decisión, Bosques aleatorios.	Clustering o agrupamiento, K-medios
Categorica	Clasificación de información, KNN, Regresión logística o de probabilidad, Naive Bayes	Análisis de asociación, Hidden Markov Model

Los anteriores métodos de análisis permitirán generar hallazgos extraídos de los datos después de crear, probar, entrenar y seleccionar los mejores modelos para las necesidades y la información disponible.

Para elegir cuáles técnicas de Aprendizaje automático son las más adecuadas para usar, se debe pensar en las características de los datos, así como el tipo de algoritmos que pueden solucionar el problema. Para esto se parte del aprendizaje supervisado y no supervisado, luego el interés del estudio, enfocado a predecir un comportamiento, encontrar relaciones entre variables o clasificar las tendencias de los datos [14], como se observa su clasificación en la figura [4-1].

4-1.

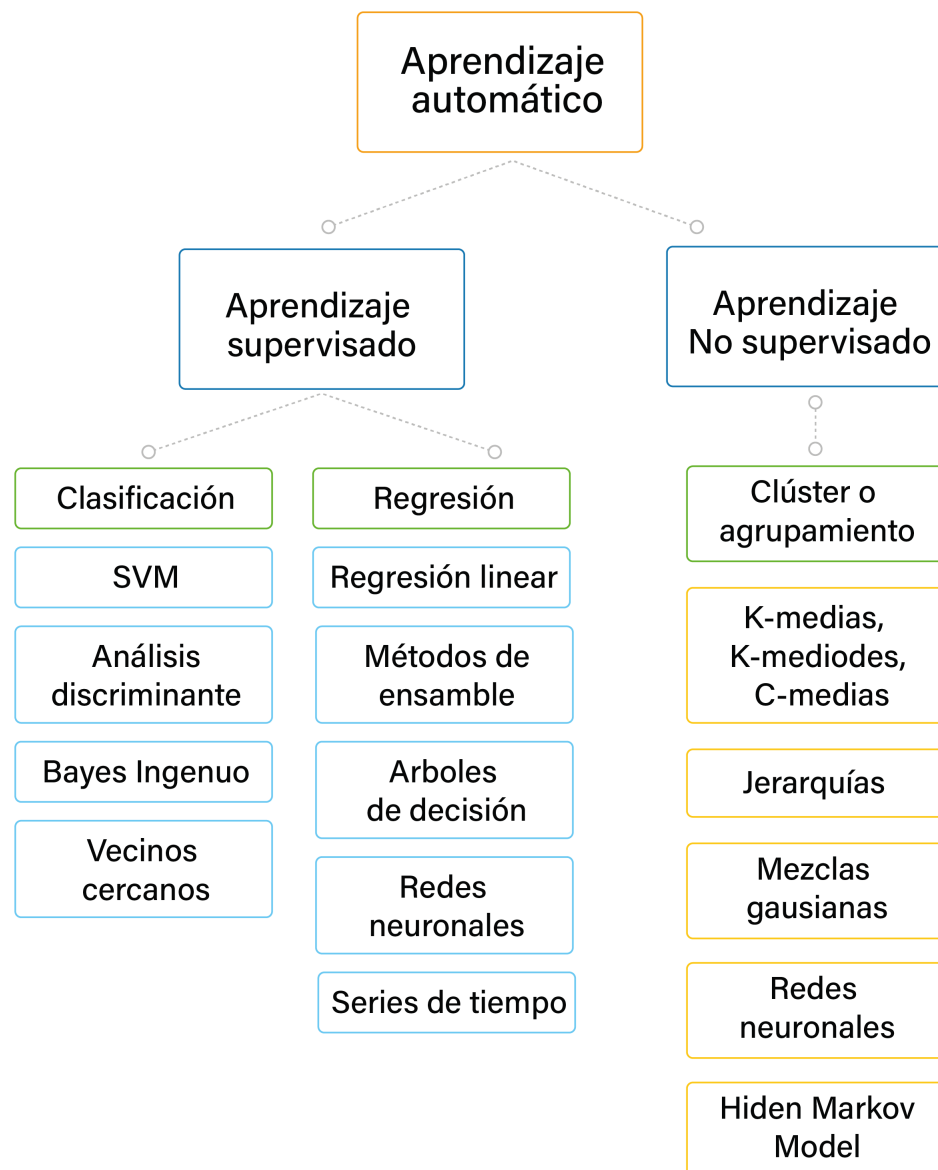


Figura 4-1: Técnicas de aprendizaje automático, elaboración propia con datos de [14].

Python y R son los lenguajes de programación e IDEs (Integrated Development Environment) más ampliamente usados para la realización de limpieza de datos y generación de modelos de Aprendizaje automático en el mundo [15].

R es un lenguaje de programación y ambiente de desarrollo para el análisis estadístico y generación de gráficos. R provee una amplia variedad de métodos estadísticos en diferentes librerías, incluyendo regresiones lineales y no lineales, pruebas estadísticas, análisis de series de tiempo, clasificación, clustering, entre otros. Gracias a diferentes paquetes y librerías creados para una fácil limpieza y minería de datos hace que sea uno de los ambientes más utilizados. Con el aumento de la utilización del Aprendizaje automático y la relativa facilidad para aprender el lenguaje la hacen la indicada para analizar volúmenes manejables de información [16].

No obstante, y debido al rápido crecimiento de las bases de datos se pretende realizar Aprendizaje automático en el corto y mediano plazo a través de una herramienta llamada Python, el rival directo de R, que desde su creación como lenguaje de programación ha sido un ambiente de desarrollo colaborativo que se viene perfilando como una gran herramienta para el manejo de grandes volúmenes de datos y la generación de Aprendizaje automático sobre el conjunto de datos que no serían fácil de manipular en otras herramientas de análisis [17]. Actualmente, ambas herramientas, R y Python se integran fácilmente con lenguajes como SQL e incluso con visualizadores de información como Power BI, gracias a los paquetes y librerías desarrollados en estos entornos colaborativos [17].

De forma complementaria en el proceso de análisis y visualización de información, existen herramientas gratuitas y de pago desarrolladas por grandes compañías para generar reportes con los datos almacenados, herramientas como Shiny, que es un paquete de R diseñado para desarrollar fácilmente y de forma interactiva aplicaciones web conectadas directamente con datos y documentos de R alojados en páginas web. Shiny permite mejorar sus aplicaciones con temas CSS, widgets programados en html y gestos de Javascript [18].

Con relación al caso práctico de este proyecto y para el rápido desarrollo e implementación en producción de los reportes se hace uso de la herramienta de Microsoft para el análisis empresarial y visualización de información Power BI lanzada en 2014 creada para mejorar las capacidades de inteligencia empresarial gracias a la fácil integración con motores de bases de datos, fuentes de información y lenguajes de programación para el análisis de datos y Aprendizaje automático [19].

Es por ello, que es importante revisar la gestión manipulada de meses atrás para evaluar aquellas dificultades con las que contó la compañía frente al almacenamiento. Se llegó la conclusión que era necesario la implementación de lenguajes de programación y herramientas

que simplifiquen la forma en la que se guarda la información. Generalmente son datos planos, fácilmente transformables en tablas relacionales que pueden contener un gran volumen de información y ocupando el menor espacio posible en los servidores de la compañía.

En lo que respecta a la actividad de realizar una consulta en la estructura que se manejaba en la compañía implicaba obtener reportes en aplicativos web de las herramientas usadas en la operación, elegir periodos de tiempo, guardar esta información en documentos de Excel y adicional a lo anterior, crear reportes que implicaban mucho tiempo en la gestión de un analista.

Sin embargo, se identificó que el verdadero problema residía al tratar de buscar un agendamiento de un paciente en diferentes meses de gestión, esa tarea podría implicar la apertura de múltiples documentos y el uso poco eficiente del procesamiento de la máquina para saltar entre documentos independientes y lograr encontrar a un usuario. Para resolver estos inconvenientes, existen motores de bases datos que permiten el almacenamiento de grandes volúmenes de datos, junto a la fácil y rápida consulta de datos específicos en cortas sentencias de código en lenguaje de programación SQL (Structured query language) que tarda segundos en generar resultados que de otra manera es poco eficiente, este lenguaje fue diseñado para administrar y recuperar información de sistemas de bases de datos relacionales [20].

Por otro lado, la producción y el desarrollo de entregables en el área de sistemas y analítica de información implementan metodologías de trabajo ágiles para lograr soluciones tempranas a las necesidades de información que solicitan las compañías [16]. Esta forma de trabajo con entregas semanales o quincenales de desarrollos ha cambiado la forma de ver los resultados, generando victorias tempranas, que permiten satisfacer las necesidades del cliente e ir iterando mejoras en las solicitudes que se realizan en cada etapa buscando tener mejoras incrementales de producto [21].

La estructuración del proceso analítico del proyecto se sustenta sobre el modelo CRISP-DM para la minería de datos, siguiendo un esquema de análisis que parte de la definición de casos de uso frente a las necesidades específicas del negocio, continúa con una exploración de los datos disponibles para una correcta selección de herramientas y generación de algoritmos, con esta información se construye modelos de aprendizaje automático, se visualizan y despliegan resultados a la operación. Posteriormente se harán procesos de evaluación, seguimiento y mejora continua del proyecto [22]. Se presenta un entendimiento visual en la figura [4-2].

4-2.

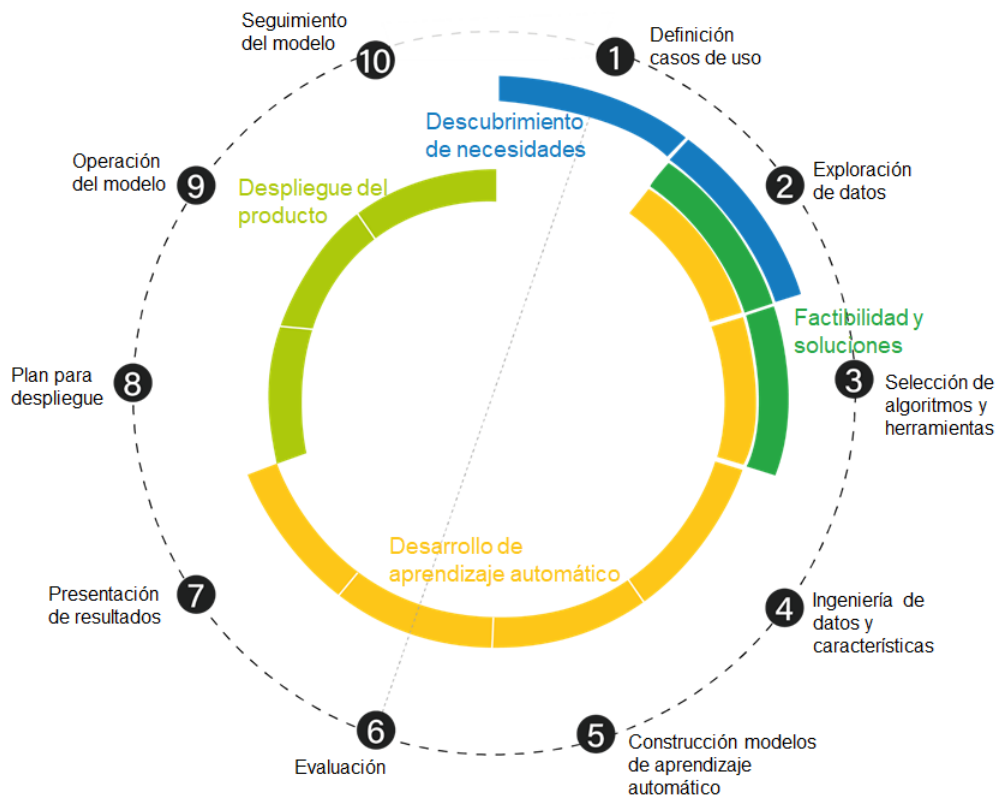


Figura 4-2: CRISP-DM para minería de datos, elaboración propia con información [22].

4.2. Aplicaciones en centros de contacto que prestan servicios de atención al usuario y agendamiento de citas médicas

Analizando los desarrollos en el mundo de los centros de contacto para la atención al usuario y agendamiento de citas en salud, la principal estrategia que se ha tomado es la implementación de bots conversacionales basados en tecnologías de la información que utilizan la inteligencia artificial para generar conversaciones naturales con los pacientes y con ello tener contacto con los usuarios en diferentes medios digitales como redes sociales y mensajería instantánea, con mensajes personalizados para la confirmación y reconfirmación de citas y procedimientos médicos que incrementan la contactabilidad [23].

El uso de modelos de Aprendizaje automático e inteligencia artificial es básico en algunos casos pero ha tenido una expansión exponencial en el mundo del centros de contacto, apoyando su contacto con los consumidores en el uso de canales digitales como chats embebidos en páginas web, chat, email y utilizando bots básicos para la toma de decisiones que apoyan

la gestión humana y la recopilación de información en bases de datos [24].

Los centros de contacto y BPO que operan en el país centran sus avances en analítica de datos basada en Aprendizaje automático en los siguientes pilares para la inteligencia de negocios:

1. Personalización mediante el análisis de las bases de datos: Los bots tienen acceso en tiempo real a las grandes bases de datos sobre clientes y usuarios, permitiéndoles ofrecer un servicio dedicado para cada consumidor.
2. Big Data: Usado para almacenar todas las actividades del usuario en internet y detectar patrones de comportamiento, consumo y compra en la red [25].
3. Modelos predictivos: La búsqueda de tendencias en el comportamiento de los consumidores para crear análisis predictivos que faciliten la toma de decisiones acertadas.

Ahondando en las aplicaciones de analítica de datos, tomando referentes mundiales con ejercicios similares a los propuestos en este proyecto, se encontró:

Existen planteamientos que usan la segmentación de la población (diferencias demográficas) para proporcionar una mejor atención médica junto a las diferencias que presentan las personas tratadas [26]:

- a. Personas con buena salud.
- b. En situaciones materno-infantiles.
- c. Con una enfermedad aguda.
- d. Con condiciones crónicas estables.
- e. Con una discapacidad grave pero estable.
- f. Con problemas de salud cercanas a la muerte.
- g. Con falla avanzada del sistema orgánico y con fragilidad a largo plazo.

Con el propósito de lograr una aproximación semejante a la priorización y segmentación del paciente se requerirán cambios en cómo evaluar y monitorear la calidad de la atención médica, cuan dispuestos estamos a pagar por la salud, monitorear la salud y las necesidades de salud, definir una salud óptima y priorizar Necesidades de salud [26]. Esta aproximación nos permite entender que cada persona necesita servicios diferentes para una salud óptima. Por lo tanto, los prestadores de servicios médicos siempre deben personalizar su servicio [26].

Con una aproximación similar, el estudio “Selección inteligente de emergencias frecuentes

departamento de pacientes para manejo de casos: Un marco de aprendizaje automático basado en datos de reclamos”, para el cual se busca optimizar el manejo de casos de emergencias mediante la aplicación de aprendizaje automático para seleccionar a los pacientes prioritarios más allá de los métodos tradicionales que usan variables usuales como la frecuencia de uso de los servicios para enfocar los esfuerzos del departamento. [26]

Así pues, optimizar los esfuerzos en los servicios de salud, los programas necesitan determinar a quién inscribir, especialmente con recursos limitados. Los afiliados ideales son aquellos que lograrán mejores resultados de salud. Si es posible, generar ahorros de costos para la organización [27].

Igualmente, con el objeto de mejorar los modelos de aprendizaje automático que buscan una selección inteligente de pacientes, se evidencia que en estudios previos se ha indicado que la demografía de los usuarios frecuentes, como la edad (mayor) y género (femenino), principales razones de consulta (dolor, lesiones, trastornos de la piel, cardiovascular, antecedente gastrointestinal, tracto urinario, complicaciones y exacerbaciones de enfermedades crónicas), condiciones de salud (abuso de drogas y alcohol y enfermedades mentales), y el uso de sistema de salud general (visitas ambulatorias, visitas de salud mental), puede estar asociado con el uso frecuente de los departamentos de emergencia [28].

En la búsqueda de la solución para este caso de uso, se requiere el empleo de varios estudios que exploran modelos predictivos destinados a usuarios frecuentes de departamentos de emergencia, utilizando métodos tradicionales de clasificación binaria como la regresión logística [29]. Con ello se puede pronosticar futuros usuarios frecuentes basados en datos de registro los departamentos de atención médica. Por ejemplo en [29] estudiaron la clasificación de tres niveles con el fin de agrupar a los pacientes en baja, media y usuarios de alta frecuencia, para los cuales aplicaron varios modelos de clasificación basado en registros de alta. Mostraron que es más fácil de predecir usuarios frecuentes bajos (<1) y altos (>5) que usuarios de frecuencia media (2 – 4 visitas al servicio de urgencias). Todo lo anterior Los estudios señalados se implementan en un conjunto relativamente pequeño de características (es decir, menos de 40), y ninguno de ellos apunta específicamente a seleccionar la población más riesgosa para el manejo de casos [29].

También, resulta primordial identificar beneficiarios de servicios médicos subvencionados para fomentar decisiones informadas de atención médica [30], y comprender las características demográficas del paciente y preferencias del paciente con atributos respectivos a la asistencia sanitaria (como la eficiencia de la atención, la reputación clínica y el entorno hospitalario).

Una población general se divide en grupos basados en la utilización a través de la agrupación de k-medios. El grupo de baja utilización está dirigido para intervenciones preventivas [31].

Para capturar características muy distintas entre una población heterogénea, la clasificación asistida de patrón de contraste para emparejar clasificadores específicos de grupo con segmentos de población que exhibir ciertos patrones [32].

5 Desarrollo del trabajo

5.1. Estructura de los datos

En la compañía E Call You S.A.S. se ha usado la plataforma InConcert de gestión para centros de contacto con el propósito respaldar la atención de los asesores y monitorear la gestión de la campaña. La tarea de extracción, carga y transformación de datos tradicionalmente se realizaba de forma manual, conectándose al aplicativo web y descargando archivos en formatos CSV y PDF por cada necesidad del supervisor a cargo de la campaña, lo que se traduce en un alto volumen de trabajo manual y repetitivo para el analista de datos o *Datamarshall* encargado, generando un alto volumen de archivos independientes que no permiten tener trazabilidad, y que dificulta manejar la información de la campaña con el pasar del tiempo y la llegada de nuevas formas de interacción con que crea datos no estructurados entregando grandes volúmenes de datos.

Ante la necesidad de centralización de información y el acceso inmediato a los datos, se decidió implementar un entorno de “Big data” que permite usar eficientemente los datos de la campaña mediante los recursos tecnológicos de la compañía. El flujo de datos de la compañía se alinearán con 5 etapas, partiendo de una ingesta de información, proveniente de datos transaccionales y estructurados, con ciertas transformaciones y formulaciones matemáticas se pueden almacenar en tablas transaccionales, creando una bodega de datos con bases de datos en MySQL y SQL Server, para un posterior análisis y visualización [figura 5-1].

5-1.



Figura 5-1: Pipeline de datos, Elaboración propia con información de [22].

Este proceso es parte de todo un flujo, en el cual la ingesta de información es un paso vital

para centralizar los datos que son cargados de las fuentes de información en la que se generan los contactos y transacciones con los usuarios. Posteriormente, se realiza un proceso de transformación donde se optimizan los datos disponibles que se almacenan de forma estructurada, limpia, estable y de fácil acceso para etapas avanzadas de análisis de datos con niveles de profundidad descriptivo y predictivo para este caso de estudio.

Los datos de la compañía E Call You S.A.S. provienen de la herramienta de gestión InConcert, la cual permite centralizar en un aplicativo web todos los procesos que debe hacer un asesor del centro de contacto mientras se encuentra suministrando atención a un usuario de la compañía. Esta persona puede brindar atención telefónica, respuesta a correos, chat y mensajería instantánea de redes sociales un solo aplicativo. Esa herramienta nos transmite los datos en un aplicativo web que permite la descarga de datos por periodos de tiempo, permite también, la descarga directa de datos por protocolos SFTP alimentando nuestra bodega de datos desarrollada en 2 niveles:

1. Datos transaccionales provenientes de la gestión, tiempos de interacción y volúmenes de interacciones con usuarios únicos que reposa en una base de datos en SQL Server.
2. Tipificación del formulario de datos de agendamiento de citas médicas con diferentes jerarquías de información, desarrollado en la compañía y que reposa en una base de datos MySQL.

Estas bases reciben actualizaciones de datos diarias con procesos en reposo durante las horas de la madrugada cargando los datos del día anterior. Lo anterior buscando generar eficiencias en el consumo de ancho de la banda de internet de la compañía, evitando generar traumatismos en la operación y permitiendo obtener repositorios estables, seguros y confiables de los datos de la operación de la campaña de salud.

Esa información es visualizada en herramientas de analítica de negocios como Power BI que permite generar hallazgos relevantes frente al comportamiento de la operación siguiendo un proceso de adquisición de datos [figura 5-2].



Figura 5-2: Estructura de flujo de datos en aplicativos para E Call You S.A.S., elaboración propia basado en [22].

En el último año, E Call You S.A.S. ha participado en un proceso de transformación digital que le ha permitido generar un manejo más eficiente de la información, sobre todo para las nuevas campañas y operaciones que han llegado a la compañía en el último año. Gracias a los aprendizajes que tuvo el departamento de tecnología y el equipo de analítica base de negocios, las nuevas campañas de atención al usuario para el agendamiento de citas y procedimientos médicos tienen asignadas estructuras de datos definidas donde se registran los datos en 2 grandes bases de datos transaccionales.

La primera base, implementada en SQL server, se alimenta de la plataforma InConcert, que le permite a los asesores recibir y realizar llamadas, responder contactos de tipo multimedia provenientes aplicaciones como Whatsapp, Facebook Messenger, correos y chat de la página web de las clínicas. En esta base transaccional se captura la información llamada a llamada, contacto a contacto, sobre los datos transaccionales del servicio, indicadores, niveles de atención, tiempo de respuesta y nivel de satisfacción del paciente. Actualmente, la compañía recibe (Inbound) alrededor de 25.000 llamadas mensuales de pacientes de la Clínica Country, junto a otras 10.000 llamadas mensuales salientes (Outbound) que se realizan. En la misma operación se reciben (Inbound) 10.000 llamadas mensuales de pacientes de la Clínica de la Colina, generando alrededor de 4.500 llamadas mensuales salientes (Outbound). Esta información junto a la gestión multimedia, es cargada automáticamente a las bases de datos SQL Server de la compañía en horas de la madrugada, alimentando la información de forma diaria con datos del día anterior.

La segunda base de datos, con alta relevancia para los propósitos de aprendizaje automático frente a la segmentación de los usuarios de la entidad médica de éste proyecto, está desarrollada en MySQL, contiene los registros de la gestión y tipificación, datos de contacto, características y datos personales de cada uno de los usuarios de las clínicas. Esta información se registra para cada interacción que se tiene con el paciente, cada solicitud, procedimiento o cita médica registrada por el paciente en los últimos 16 meses.

Mientras que la información transaccional de la Clínica de la Colina, referente al periodo de tiempo comprendido entre los meses de Enero de 2019 al mes de Junio de 2020, contiene el registro de más de 135.290 tipificaciones de llamadas con cédulas de pacientes, con más de 56.530 agendamientos para la Clínica. Esta información contiene el 58.3% de pacientes recurrentes, pertenecientes en su inmensa mayoría (más de 130.000 contactos) a servicios de medicina prepagada. Estos usuarios son personas de todas las edades, desde neonatos hasta usuarios que cuentan con más de 106 años, como se puede observar en la [figura 5-3].

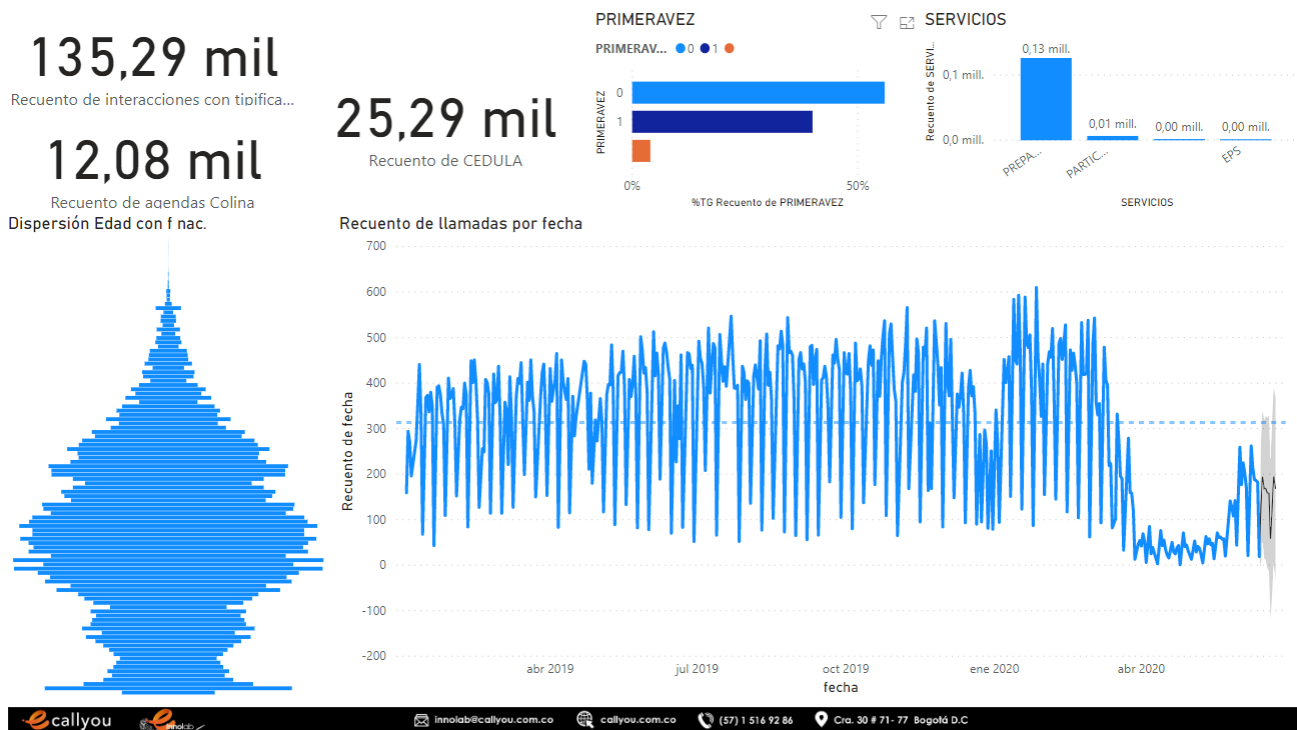


Figura 5-3: Datos relevantes gestión Clínica de la Colina, archivo personal elaborado para E Call You S.A.S.

El entendimiento y organización de la captura de datos se da a partir del diccionario de datos de las 2 grandes tablas que conforman la bodega de datos correspondiente a esta operación de atención al usuario y agendamiento de citas médicas.

Los campos y variables definidos en las tablas transaccionales contienen un orden, definición y utilidad, la estructura básica para entender el funcionamiento de los datos de agendamiento se pueden distinguir en la tabla [5-1].

En el mismo sentido, los campos, variables e información referente a la operación transaccional y de gestión de interacciones se puede comprender en su estructura en la tabla [5-2].

Tabla 5-1: Agendamientos E Call You S. A. S. Diccionario de datos, Archivo personal elaborado para E Call You S.A.S.

Fecha captura de información	1 de enero 2019 a al 15 de Junio 2020, con actualización diaria
Muestra	Tabla de agendamientos en MySQL con agendamiento de citas médicas en clínicas privadas.
	Transacciones de más de 25.000 pacientes mensuales de la Clínica de la Colina en la ciudad de Bogotá.
	6 especialidades Clínica de la Colina
	Operación mensual de 21 asesores de servicio al usuario para agendamiento de citas médicas
Descripción de los campos de la tabla	La tabla cuenta con campos de diferentes segmentos de información indagando por diferentes características que perfilan al paciente del servicio y el rendimiento de la campaña:
	· Caracterización sociodemográfica de los pacientes.
	· Información de servicios médicos que posee el paciente.
	· Tipos de consultas médicas solicitadas y recurrencia de pacientes.
	· Preferencia de ubicaciones y médicos.
	· Necesidades de los pacientes.
	· Agenda disponible por parte de las clínicas.
· Asignación óptima de citas.	

Tabla 5-2: Transacciones E Call You S. A. S. Diccionario de datos, Archivo personal elaborado para E Call You S.A.S.

Fecha captura de información	1 de enero 2019 al 15 de Junio 2020, con actualización diaria.
Muestra	Tabla de transacciones en SQL Server con las interacciones de los asesores por múltiples canales de contacto con los usuarios de la clínica.
	transacciones de más de 25.000 pacientes mensuales de la Clínica de la Colina en la ciudad de Bogotá
	6 especialidades Clínica de la Colina
	Operación mensual de 21 asesores de servicio al usuario para agendamiento de citas médicas
Descripción de los campos de la tabla	La tabla cuenta con campos de diferentes segmentos de información indagando por diferentes características que perfilan al paciente del servicio y el rendimiento de la campaña:
	· Caracterización de los pacientes con ID.
	· Información de la gestión telefónica y niveles de atención.
	· Tiempos de respuesta, clasificación y codificación de las llamadas.
	· Indicadores de calidad de la atención de las llamadas.
	· Datos de identificación del asesor que responde a las necesidades del usuario.
	· Clasificación de las horas de atención y estados de las interacciones con los usuarios.

5.2. Inteligencia de negocios

Basado en la gestión histórica de la campaña se realiza un análisis exploratorio para determinar qué interacciones de la labor son útiles para optimizar el agendamiento de salud. Para acometer esta tarea, se plantearon una serie de preguntas para ser respondidas por la analítica de datos:

- a. ¿La efectividad del agendamiento depende de la hora de contacto del paciente?
- b. ¿Se puede identificar los momentos con mayor tráfico de llamadas y cómo podemos tomar medidas correctivas para mejorar el desempeño de los indicadores clave de atención?

c. ¿Es posible conocer la oportunidad de la asignación de citas médicas y su comportamiento histórico?

d. ¿Existen diferencias en la gestión y agendamiento de citas médicas?

Con estas preguntas de ejemplo, se explora el tipo de conocimientos que se pueden observar para realizar un mejor entendimiento del negocio. Estos análisis se realizan con los datos de la gestión de la campaña Salud, desde Enero de 2019 hasta abril de 2020, encontrando lo siguiente:

Se evidencia un comportamiento estacional en el flujo de interacciones de los usuarios que se comunican con la compañía, se precisan días con mayor nivel de contacto, como lo es el inicio de la semana, los usuarios intentan agendar y solucionar sus inconvenientes de salud iniciando la semana [figura 5-4], de igual forma existen 4 intervalos del día donde los pacientes intentan comunicarse con mayor intensidad, los minutos cercanos a las 9AM, minutos previos a las 12MM, antes de las 3PM, los minutos anteriores a las 5PM y 6PM [figura 5-4].

5-4.



Figura 5-4: Gestión histórica de interacciones por intervalos de 15 minutos, archivo personal elaborado para E Call You S.A.S.

De igual forma, es posible reconocer los momentos en los que más se está abandonando interacciones por parte de los asesores de la compañía. En este sentido, el intervalo de 15 minutos de las 9:30AM, el intervalo de minutos posteriores a las 12:15MM y el intervalo de las 14:45PM son los momentos más críticos para la operación [figura 5-4], implicando una alerta para los supervisores que deben prever una cantidad óptima de asesores para la atención al público.

Así mismo, se logró identificar el comportamiento de un indicador diferencial en la gestión de agendamiento de consultas, exámenes y citas médicas en el sector salud. Usando los campos de “fecha interacción” que permite conocer la fecha exacta en la que el paciente se comunicó con las líneas de la clínica y “fecha paciente” que se asigna cuando se agenda un servicio médico, se crea la oportunidad en la asignación de citas médicas, un indicador solicitado por la superintendencia de salud a las empresas prestadoras de servicios de salud con el fin de mantener en intervalos de tiempo aceptables la asignación de citas médicas para los usuarios [figura 5-5].

5-5.

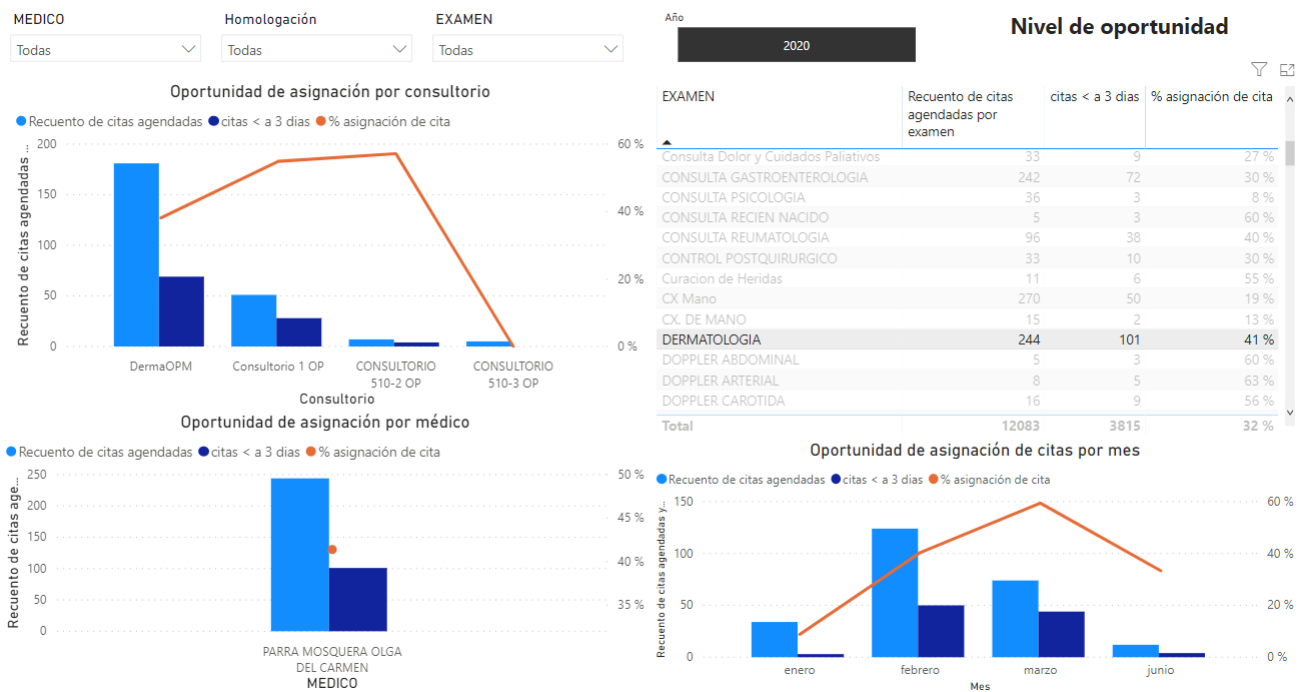


Figura 5-5: Oportunidad en la asignación de citas médicas, archivo personal elaborado para E Call You S.A.S.

Este indicador permite conocer cuáles agendamientos de servicios médicos se realizaron en los 3 días posteriores a la comunicación del paciente con la línea, permitiendo conocer qué porcentaje de citas están teniendo una pronta respuesta y utilización de los servicios médicos con la agenda disponible.

De mismo modo, es posible conocer el comportamiento de los agendamientos por tipo de examen médico, por profesional de la salud que atiende los requerimientos, e incluso podemos generar un ranking de agendamientos de citas médicas por nombre de asesor que atiende las interacciones [figura 5-6].

5-6.

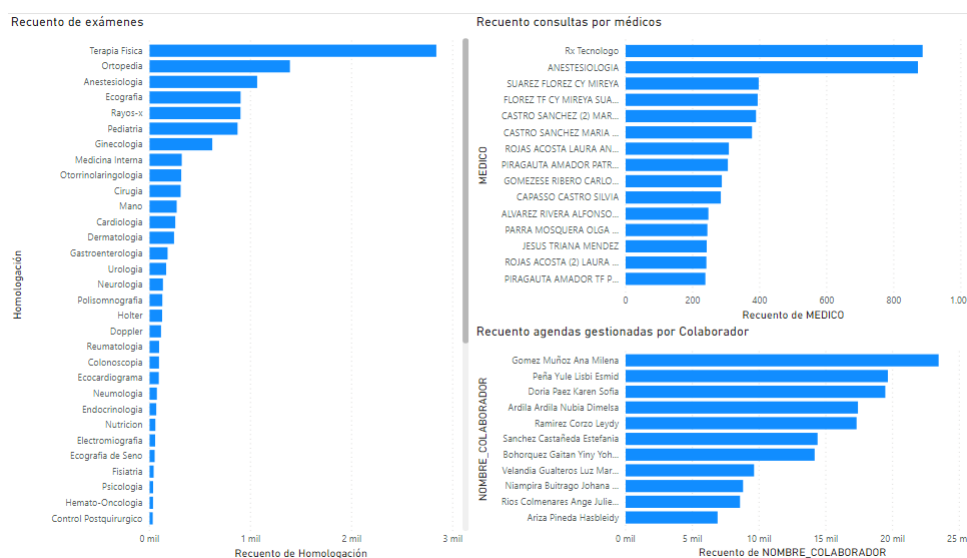


Figura 5-6: Volumen de agendamiento de citas médicas, archivo personal elaborado para E Call You S.A.S.

Con esta información es posible conocer cuáles servicios médicos tienen una mayor demanda para la clínica, identificando las variaciones de estos indicadores día a día. También podemos generar incentivos para el asesor basado en su gestión y nivel de respuesta, en caso de que se promuevan metas de cumplimiento para otorgar comisiones, pagos variables, etc.

5.3. Analítica predictiva y aprendizaje automático

Con los datos recopilados en la campaña se logra desarrollar modelos de aprendizaje automático, útiles para mejorar la atención a los usuarios, brindando un servicio personalizado de acuerdo con las características de los pacientes, así como mejorar la labor del agendamiento de citas y procedimientos médicos. Para satisfacer esta necesidad se realizaron 2 modelos.

El primero utiliza métodos de aprendizaje no supervisado y sirve para hacer la segmentación de los usuarios y determinar las características específicas de cada segmento con el fin de usarlas como herramienta para determinar mejores estrategias de contacto y entendimiento del paciente. Mediante un modelo de Clustering que usa el método de K-medios encontrando un número óptimo de 8 agrupamientos o clusters. Se eligió K-medios, porque es el que brinda una mayor interpretación de los segmentos.

Con el ánimo de facilitar estos agrupamientos de pacientes, se utilizaron diferentes variables comportamentales y de descripción de los usuarios únicos que se comunican con las líneas e interactúan con los asesores. En una primera fase analítica se generó un mapa de calor de correlaciones entre las variables respuesta de las bases de datos, las cuales en muchos casos eran dicotómicas, categóricas o de clasificación. Lo anterior no permitía identificar de forma sencilla la relación entre las variables, pero permitió identificar cuáles características serían útiles para segmentar a los usuarios de la Clínica [figura 5-7]. 5-7.

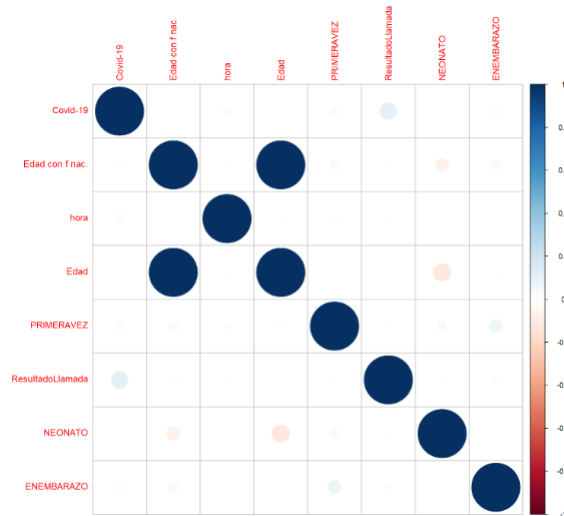


Figura 5-7: Correlación de variables con mapa de calor, archivo personal elaborado para E Call You S.A.S.

Con lo anterior, se decidió realizar un escalado de variables (aplicar el código “scale” al set de datos) con el fin de utilizar las columnas con valores nominales interrelacionales entre ellos,

permitiendo encontrar valores medios cercanos entre los agrupamientos de los segmentos de los pacientes.

Frente a lo experimentado con la correlación de variables, así como lo visto en la literatura existente frente al agrupamiento de variables para la segmentación de usuarios de servicios médicos, se utilizaron los siguientes campos de la base de datos como se muestra en la tabla [5-3]:

Tabla 5-3: Variables seleccionadas para el agrupamiento de segmentos, Archivo personal elaborado para E Call You S.A.S.

Variables para la segmentación de pacientes	
Edad con fecha de nacimiento	Se generó una variable dinámica en el dataset que calcula la edad actual de los pacientes basado en su fecha de nacimiento.
Rango de edad	Se generó una variable que agrupa las edades con fecha de nacimiento con intervalos por grupos etarios.
Sexo	Se generó una variable a partir del nombre del paciente.
Primera vez	Variable Dummy para identificar usuarios recurrentes.
Examen Homologado	Se homologaron los servicios médicos prestados.
Resultado llamada (tipificación)	Se utilizó la categorización del motivo por el cual el paciente realizó la llamada.
En embarazo	Se tuvo en cuenta el estado del paciente y su prioridad.
Neonato	Se tuvo en cuenta la edad del paciente y su prioridad.
Tipo de cita	Hace referencia a la recurrencia en la solicitud de procedimientos médicos.

En la búsqueda del número óptimo de grupos o clústeres en los cuales se especificarían las características clave de los usuarios tipo de la clínica, se implementó una función que tendría en cuenta las variables disponibles en el conjunto de datos y evaluaría el número de agrupamientos que pueden darse en función de los datos. Con esto, no sólo se utilizó el método k-medios para encontrar grupos en la segmentación de datos, sino que se utilizaron otras metodologías para encontrar y definir el número óptimo de grupos para la segmentación de pacientes.

Resultado función en R para calcular el número de Clústeres:

```
-----
#the index to be calculated. This should be one of: "kl", "ch", "hartigan", "gcc",
"scott",
#"marriot", "trcovw", "tracew", "friedman", rubin", "gindex", "db", "silhouet-
te", "duda",
#"pseudot2", "beale", ratkowsky", "ball", "ptbiserial", "gap", "frey", "mcclain",
"gamma",
#"gplus", "tau", "dunn", "hubert", "sdindex", "dindex", "sdbw", .all"(all in-
dices except GAP,
#Gamma, Gplus and Tau), .alllong"(all indices with Gap, Gamma, Gplus and Tau in-
cluded).
resnumclust<-NbClust(df, distance = .euclidean", min.nc=2, max.nc=10, method =
"kmeans", index = .alllong")
-----
```

The Hubert index is a graphical method of determining the number of clusters.

In the plot of Hubert index, we seek a significant knee that corresponds to a significant increase of the value of the measure i.e the significant peak in Hubert index second differences plot.

The D index is a graphical method of determining the number of clusters.

In the plot of D index, we seek a significant knee (the significant peak in Din-
dex

second differences plot) that corresponds to a significant increase of the va-
lue of the measure.

```
-----
* Among all indices:
* 13 proposed 8 as the best number of clusters
* 2 proposed 7 as the best number of clusters
* 1 proposed 4 as the best number of clusters
* 1 proposed 5 as the best number of clusters
* 7 proposed 6 as the best number of clusters
* 1 proposed 9 as the best number of clusters
* 3 proposed 10 as the best number of clusters
```

***** Conclusion *****

* According to the majority rule, the best number of clusters is 8

Usando las variables mencionadas anteriormente, se aplicó el método gráfico para realizar agrupamiento de variables de acuerdo a sus características tuvo en cuenta la frecuencia de llamadas de los pacientes que se comunicaban con la clínica, así como sus edades, encontrando 8 segmentos con características diferentes entre sí [Figura 5-8].

5-8.

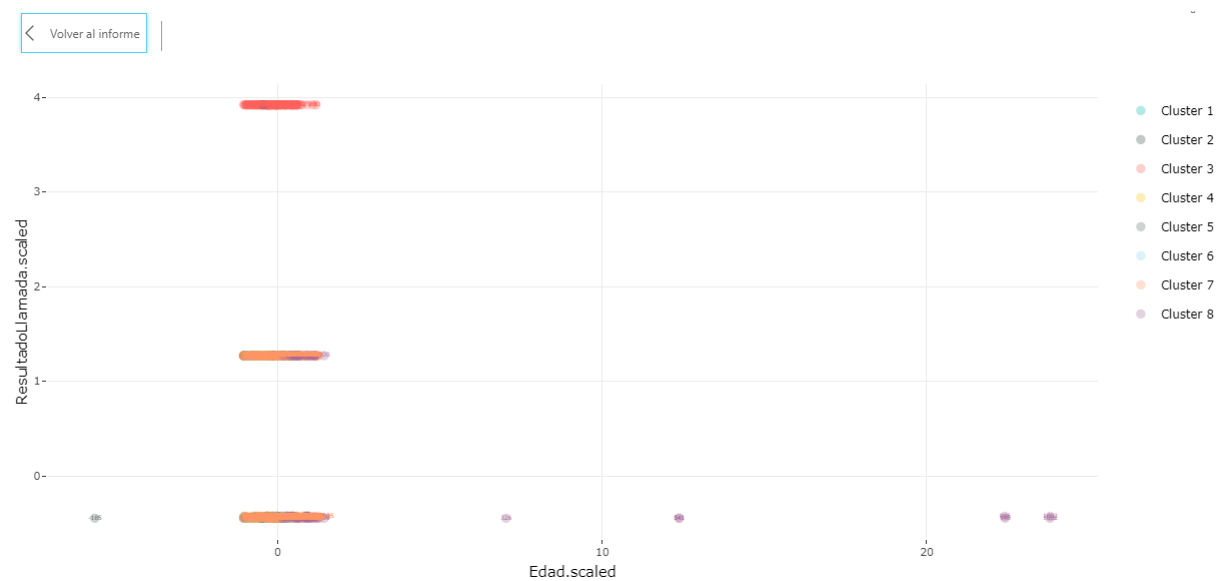


Figura 5-8: Clúster de variables usando K-medios con gráfica de R en Power BI, archivo personal elaborado para E Call You S.A.S.

Con esta segmentación generada por el clúster de k-medios se incluyeron mediante código una variable en el conjunto de datos que asigna un número a los pacientes de acuerdo con el segmento al que corresponden:

```
TipificacioncymColina% >%
mutate(Cluster_k8 $ cluster)% >%
group_by(Cluster)% >%
summarise_all("mean")
df <- TipificacioncymColina
```

```
df
df$clus <- as.factor(k8$cluster)
df
```

```
df <- TipificacioncymColina
df <- scale(df)
df<- as.data.frame(df)
df$clus <- as.factor(k8$cluster)
df
```

Basado en lo anterior, los pacientes segmentados por el método de k-medios comparten características comportamentales como la frecuencia de comunicación con los canales de interacción de la clínica, grupo etario, exámenes médicos agendados en el periodo de tiempo seleccionado para este estudio, primera vez, en embarazo, tipo de cita y resultado de llamada que hacen referencia al tipo de contacto y paciente. Esta clasificación se puede develar en la tabla [5-4]:

Tabla 5-4: Segmentos de pacientes generados, Archivo personal elaborado para E Call You S.A.S.

Segmentos de población con ejemplos de pacientes típicos	
Clúster 1. Primera infancia	Son los pacientes más jóvenes de la clínica 0-5 años, que consultan por pediatría, rayos x y ecografía principalmente.
Clúster 2. Niños y jóvenes	Comprenden edades entre 6 y 15 años, afiliados a medicina prepagada, consultan por pediatría, ortopedia y terapia física. Los accidentes producto del juego, son recurrentes a esta edad.
Clúster 3. Jóvenes adultos	Con edades entre 16 y 20 años, consultan por terapia física, ortopedia y aun por pediatría.
Clúster 4. Adultos	Pertenecen al grupo de los “20 añeros”, aparecen las consultas con todo tipo de especialistas, es importante brindar asesoría personalizada.
Clúster 5. Adultos recurrentes	Los pacientes con edades de 30 a 40 años, basado en la dispersión de edades son el grupo con mayor cantidad de interacciones con la clínica, también requieren un trato especial, empiezan a padecer enfermedades que los acompañaran por el resto de sus vidas.
Clúster 6. Generación X	Comprenden edades de 40 a 60 años, saben cómo relacionarse con sus servicios médicos, son pacientes conocedores y exigentes.
Clúster 7. Baby boomers	Pacientes entre 60 y 70 años, tienen visitas recurrentes durante el año a los servicios de salud, exigen un trato especial.
Clúster 8. sobrevivientes	Son los mayores de 80 años, aparecen servicios como cardiología, oncología entre otros, la comunicación se da con familiares.

Con la clasificación y segmentación asignada por el clúster al que pertenecen los pacientes, se genera una herramienta visual que permite conocer las características especiales de cada grupo de pacientes, brindando al cliente un entendimiento panorámico de sus usuarios. Adicionalmente se genera para el asesor, que recibe las interacciones con los usuarios, un apoyo visual con el que puede brindar una atención personalizada a los pacientes que se comunican. Al introducir la cédula o ID del usuario se puede evidenciar a que grupo pertenece [figura 5-9].

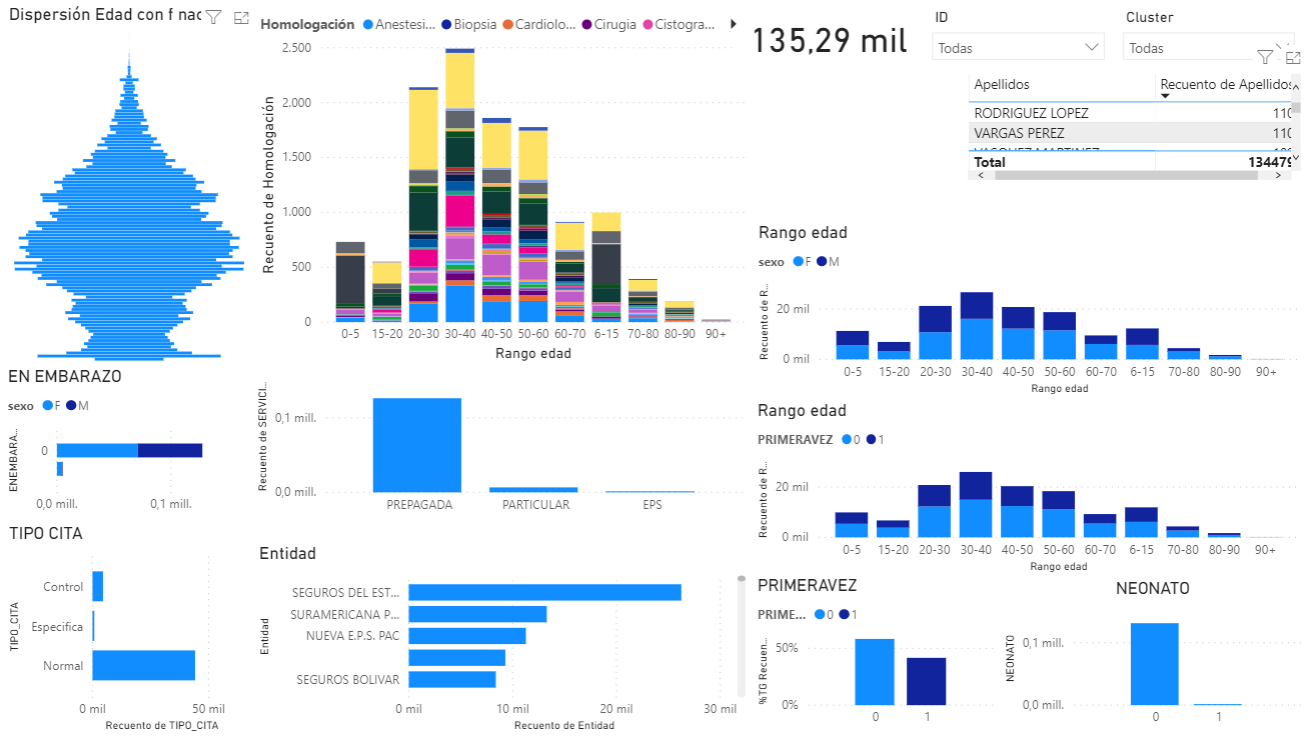


Figura 5-9: Ficha de segmentación de usuarios de acuerdo con agrupamiento por clústeres, archivo personal elaborado para E Call You S.A.S.

El segundo modelo de aprendizaje automático determina el pronóstico del volumen de llamadas y agendamientos de los usuarios de la clínica en un período de tiempo. En este caso se hizo una comparación entre métodos de series de tiempo: Auto Regresivo de orden 10, ARIMA de orden 10, y el método sin estacionalidades TBATS de orden 10.

El objetivo de este modelo generado es conocer el pronóstico o previsión del volumen de llamadas y agendamiento de servicios médicos permitiendo a la operación prever el dimensionamiento de las personas dispuestas a atender las interacciones de los usuarios de la clínica. Con esto se usaron la cantidad de interacciones por día, desde el 2 de enero de 2019 hasta el 8 de mayo de 2020. Con esas interacciones se decide realizar modelos de series de tiempo que usan lo ocurrido en periodos anteriores para pronosticar qué puede ocurrir en los siguientes periodos del horizonte temporal.

A priori, este ejercicio parece una tarea fácil, teniendo en cuenta que Power BI incluye un modelo gráfico de previsiones para entender el comportamiento y pronosticar lo que puede acontecer con una variable en periodos futuros [figura 12]. Este ejercicio sencillo fue útil para comprender las particularidades de la gestión de interacciones y agendamientos de la operación.

El modelo de series de tiempo proporcionado por Power BI responde a una serie de tiempo Auto Regresiva con un intervalo de confianza de 95 % y prediciendo los siguientes 10 puntos. El punto clave es la detección de estacionalidades, la cual se hace con un proceso automático que no tiene en cuenta el promedio de los datos. Con estas características, el modelo inicial pronosticaba interacciones negativas en los periodos siguientes al 18 de enero [figura 5-10].

5-10.

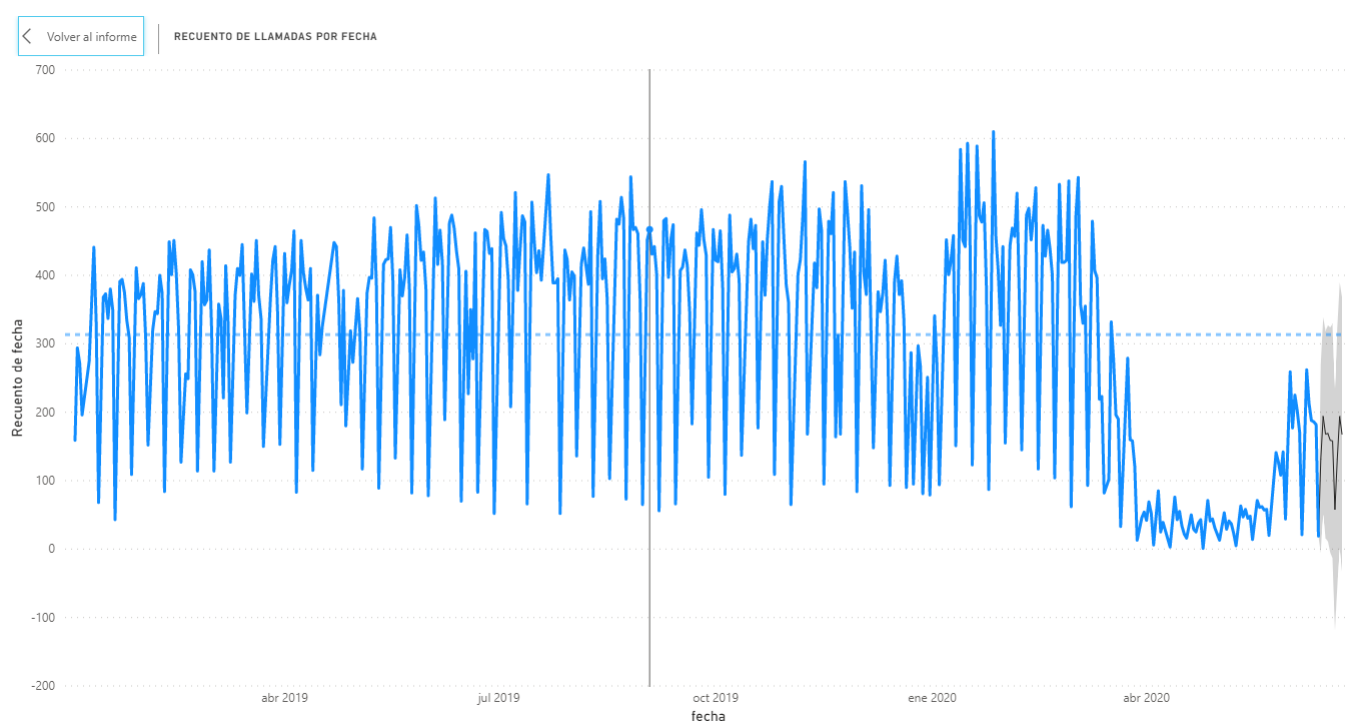


Figura 5-10: Pronóstico de llamadas por fecha, modelo Auto Regresivo de orden 10, archivo personal elaborado para E Call You S.A.S.

Se evidencia el impacto que tienen los periodos estacionales de la temporada decembrina, mitad de año y la caída típica de contactos durante los días sábado. Adicionalmente, se debe aclarar que el periodo se limitó a las fechas previas al 18 de marzo de 2020, pues a partir de ese momento se presenta una distorsión en la operación por el efecto que genera la cuarentena decretada por el gobierno para proteger a la población de la amenaza de contagio por el COVID-19.

Este acontecimiento tiene un fuerte efecto sobre la gestión de agendamiento de la clínica, que decreta estado de emergencia, cancela la atención al público con algunas excepciones para atender urgencias, emergencias o consultas por fuerza mayor.

Buscando corregir el error que generan las estacionalidades de los datos, se realizó el mismo proceso con un modelo diferente, se seleccionó una visualización de R para generar un modelo ARIMA (Modelo auto regresivo integrado de media móvil), el cual utiliza variaciones y regresiones de los datos para pronosticar valores futuros [figura 5-11].

Este modelo presenta un intervalo de confianza en primer nivel de 95 % y de 90 % en segundo nivel, fue limitado para predecir los siguientes 10 puntos, se incluyó un factor objetivo de estacionalidad semanal, permitiendo pronosticar valores máximos y mínimos en la predicción.

Con esta visualización la precisión del modelo mejora, continúa presentándose falta de precisión, especialmente en los rangos de valores mínimos y máximos de pronóstico generados con el modelo [figura 5-11].

5-11.

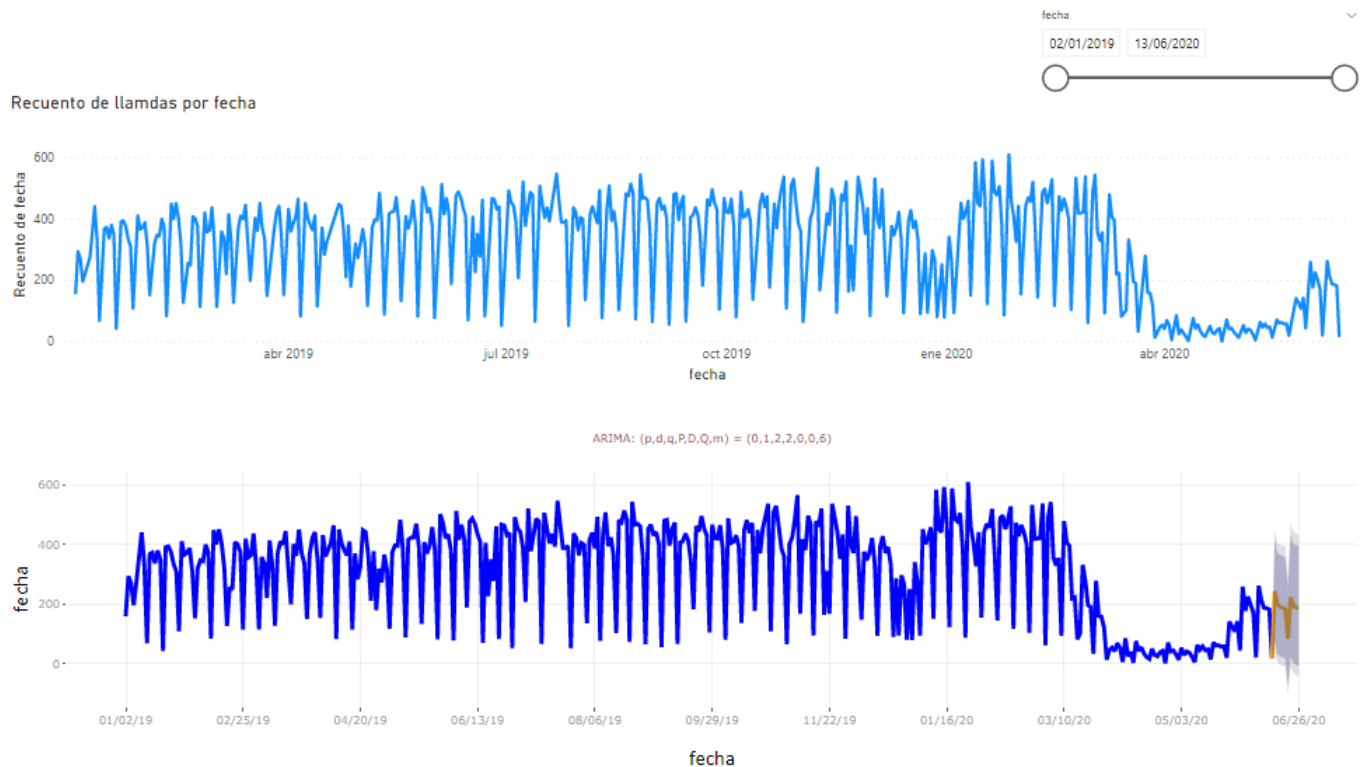


Figura 5-11: Pronóstico de llamadas por fecha, modelo Auto Regresivo con promedio móvil de orden 10, archivo personal elaborado para E Call You S.A.S.

En el proceso de análisis se decide implementar un último modelo de series de tiempo que permite afianzar los resultados del pronóstico mediante series de tiempo. Utilizando el modelo TBATS (acrónimo de Trigonometric seasonality, Box-Cox transformation, ARMA errors,

Trend and Seasonal components), este modelo permite entender que cada estacionalidad está modelada por una representación trigonométrica basada en la serie de Fourier. Una de las principales ventajas de este enfoque es que solo requiere 2 estados semilla, independientemente de la duración del período [13].

El modelo implementado tiene en cuenta el recuento de interacciones por fecha para pronosticar el resultado de volumen de interacciones en los siguientes 10 periodos. Para esto el modelo tiene en cuenta un intervalo de confianza en primer nivel de 95 % y de 90 % en segundo nivel, se incluyó un primer factor objetivo de estacionalidad semanal, y un segundo factor mensual, lo que permite tener un pronóstico más certero, incluyendo los valores máximos y mínimos en la predicción. Con esto el modelo replica el comportamiento de las estacionalidades propias del volumen de interacciones y el agendamiento de citas médicas. Lo que logra un panorama futuro ajustado al comportamiento real para la toma de decisiones frente al dimensionamiento de la atención en la operación [figura 5-12].

5-12.

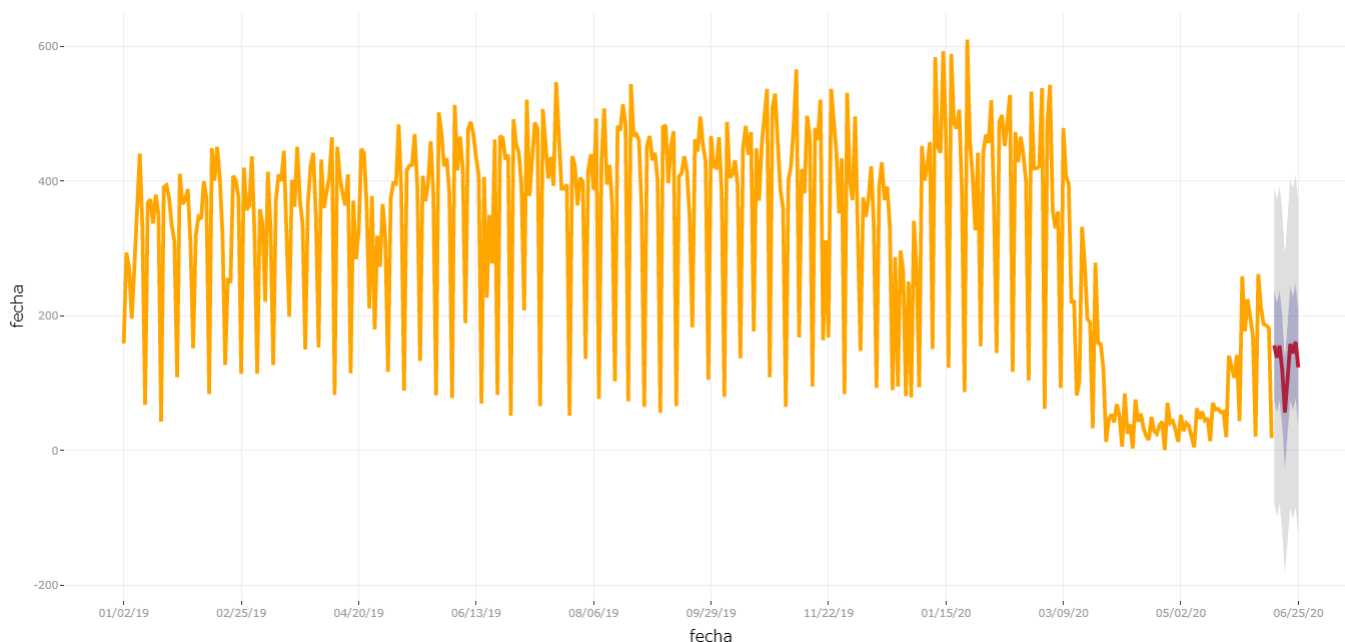


Figura 5-12: Pronóstico de llamadas por fecha, modelo TBATS de orden 10, archivo personal elaborado para E Call You S.A.S.

Actualmente se han desplegado los modelos, brindando acceso a los usuarios interesados a la información. Se necesitarán algunos meses para evaluar el impacto de estos desarrollos frente a qué tanto ha mejorado la gestión, operación y relacionamiento con el paciente desde

la implementación de estos métodos.

5.4. Desarrollo de producto

El proyecto realizado potencializa la operación del departamento de analítica base de negocios de la compañía, que tradicionalmente contaba con un equipo de Workforce compuesto por 4 datamarshall, que no tenían el tiempo (debido al volumen de trabajo y realización de tareas manuales) ni con el conocimiento para desarrollar tareas que mejoraran y dieran visibilidad a la gestión de la analítica de la información.

El producto mínimo viable avanzado con este proyecto gestionó la unificación y organización de los datos en bodegas de datos alojada en los servidores de la compañía en bases SQL Server para la información transaccional de la gestión de la operación y en una base en MySQL para alojar la información relacionada con el agendamiento y tipificación de los servicios médicos de la clínica.

Se empezó a generar más valor con los datos almacenados en la compañía mediante Inteligencia de negocios, se producen modelos predictivos y de segmentación que usan aprendizaje automático para la toma de decisiones acertadas en la operación de servicio al usuario, se crean visualizaciones en Power BI con formularios con inicio sesión en una aplicación web embebida en la página web de E Call You S.A.S. [figura 5-13].

Con estos reportes encriptados se permite dar visibilidad a la operación y al cliente externo de la gestión del equipo de trabajo quitando los reportes manuales y generando actualizaciones diarias de reportes automatizados [figura 5-14].

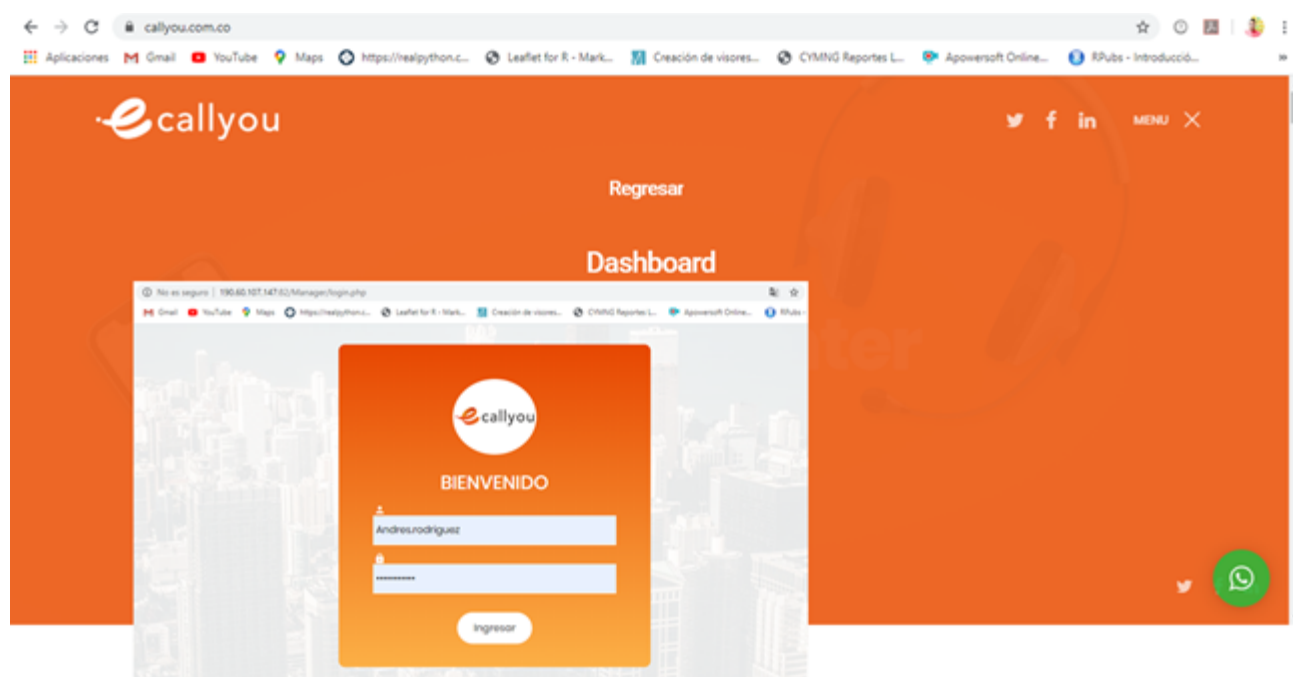


Figura 5-13: Formulario de registro usuario y contraseña para cliente interno y externo, archivo personal elaborado para E Call You S.A.S.

Estos reportes están hechos a la medida de las necesidades que expresa el cliente, con actualizaciones en tiempo real en algunas tablas, también incluye la segmentación de pacientes, especialidades y exámenes con solicitudes recurrentes, identificación de niveles de satisfacción y cambios en su percepción. Esta información se encuentra disponible con accesos web para las personas interesadas, por ejemplo, podrían conocer los volúmenes de interacciones en la gestión transaccional, como se muestra a continuación [figura 5-14]. De igual forma, se dispone un reporte en línea para conocer los indicadores de efectividad y los desarrollos frente inteligencia de negocios [figura 5-15].

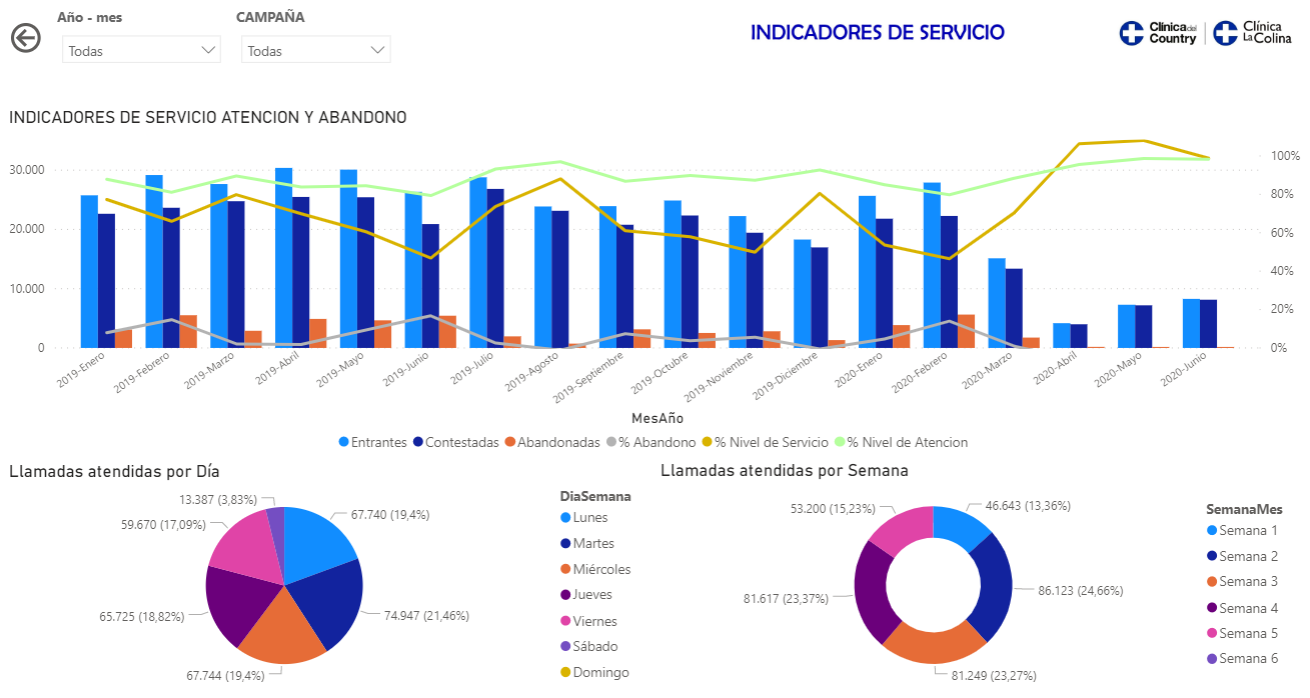


Figura 5-14: Gestión transaccional, indicadores de atención y abandono, atención de llamadas por intervalos, archivo personal elaborado para E Call You S.A.S.

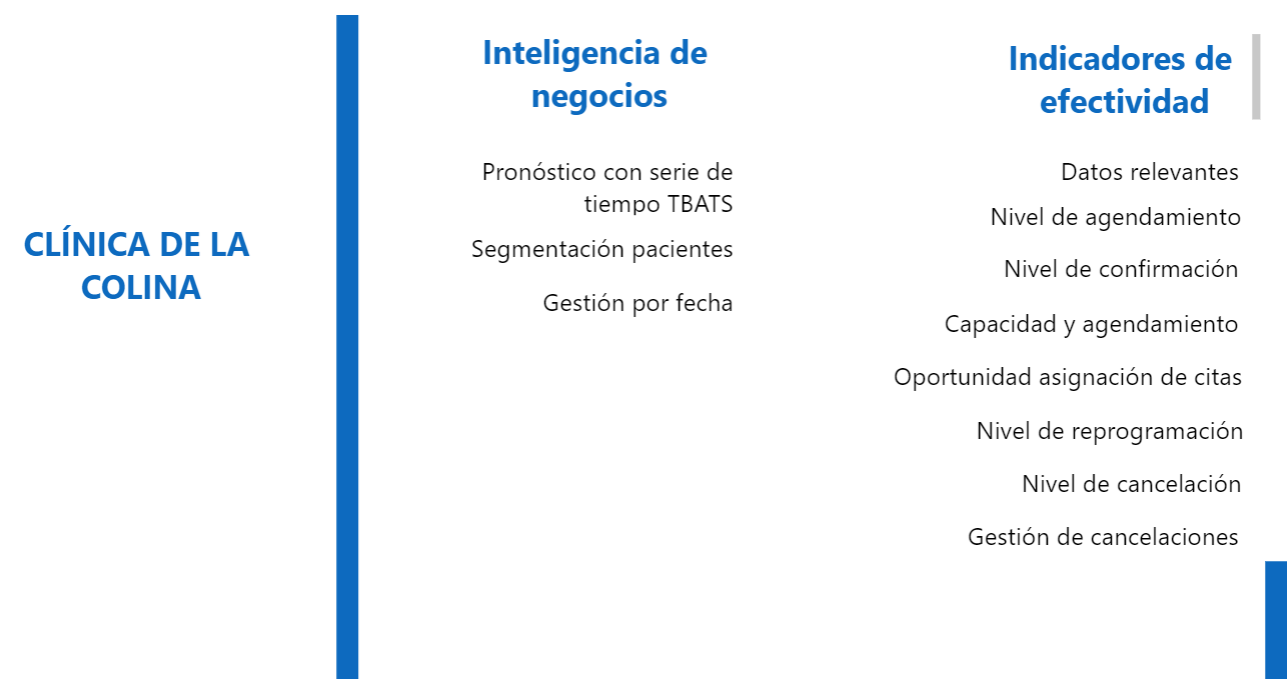


Figura 5-15: Inteligencia de negocios e indicadores de efectividad Clínica de la Colina, archivo personal elaborado para E Call You S.A.S.

Con el interés de agrupar los resultados del desarrollo del proyecto, se diseña la tabla [5-5] que resume de los documentos generados.

Tabla 5-5: Resumen de documentos generados en el desarrollo del producto, Archivo personal elaborado para E Call You S.A.S.

Tipo	Nombre	Descripción	Frecuencia de uso
Código	Agendamiento campaña de salud	Bodega de datos en MySQL con el esquema indicado para la gestión de citas de cada campaña.	Continuo
Código	Transaccional de interacciones	Bodega de datos en SQL Server con el esquema indicado para la gestión de citas de cada campaña.	Continuo
Código	Trigger	Carga y actualiza los datos de las tablas SQL, los transforma para que queden en formato adecuado (dataframes) y consistente.	Diario
Código	Análisis R	Hace un análisis de las variables más relevantes y permite entender cuáles son las tendencias y patrones comportamentales de los pacientes y cómo evolucionan en el tiempo.	Quincenal
Código	Clustering	Permite hacer una segmentación de los pacientes en grupos que comparten características específicas, útiles para entender mejor su comportamiento, variables utilizadas para mejorar el acercamiento de los asesores en la gestión de la campaña.	Quincenal
Código	Forecast	Predice el volumen de llamadas y de agendamientos. Se actualiza diariamente según los resultados la gestión.	Diario
Aplicación web	Informes internos	Aplicación web desarrollada en Power BI que permite a los gerentes operativos conocer diariamente el desarrollo de la gestión, segmentación de pacientes y pronóstico de agendamiento de citas médicas por especialidad.	Continuo
Aplicación web	Informes externos	Aplicación web desarrollada en Power BI que permite a los interesados administrativos de la clínica Colina conocer diariamente el desarrollo de la gestión, segmentación de pacientes y pronóstico de agendamiento de citas médicas por especialidad.	Continuo

5.5. Progreso

Se diseña la tabla [5-6] con el interés de compartir un resumen del avance del proyecto, teniendo en cuenta que las labores que se encuentran en curso hacen referencia a la implementación de la herramienta como solución y apoyo para los asesores del centro de contactos. Por esto, las tareas llevarán un proceso de socialización, adaptación de nuevas tecnologías y capacitaciones a clientes, directores operativos y asesores para implementar de forma completa los desarrollos.

Tabla 5-6: Resumen avance del proyecto, Archivo personal elaborado para E Call You S.A.S.

Etapa	Tarea	Progreso
I	Recuento histórico de los avances que llevaron al desarrollo de las nuevas metodologías que se aplicaran en el proyecto.	Finalizado
	Análisis de conceptos básicos relacionados con Big Data y aprendizaje automático.	Finalizado
	Estudio de los conceptos básicos usados en el entendimiento del paciente y agendamiento en salud.	Finalizado
	Revisión de los métodos usados por otras compañías del mismo sector para solucionar estas necesidades.	Finalizado
II	Diseño de una base de datos.	Finalizado
	Paso a un almacenamiento en la nube o servidor.	Finalizado
	Organización de datos en servidor.	Finalizado
III	Exploración profunda de los datos.	Finalizado
	Generación de un diccionario de datos.	Finalizado
	Analizar incongruencias en la información.	Finalizado
IV	Entendimiento profundo de los datos existentes teniendo en cuenta los patrones de solicitud de cita, características demográficas de los pacientes, necesidades urgentes y las características del entorno que afectan la operación de la campaña.	Finalizado
	Determinar mejoras tempranas en procesos y asignación de cita.	En curso
	Creación de tableros interactivos que permitan evaluar día a día el rendimiento y crear alarmas sobre posibles fallas en el servicio.	Finalizado
V	Perfilación del paciente tipo (Fichas de consumidor).	Finalizado
	Optimización de procesos internos de contacto y comunicación.	En curso
	Definir las variables para un estudio óptimo frente a la implementación predictores de agendamiento de citas con Aprendizaje automático.	Finalizado
	Validación de los modelos para garantizar la predictibilidad	Finalizado

6 Conclusiones

Objetivo 1. Implementar y dar seguimiento a un entorno de “Big Data” que permita el almacenamiento de la información en Bodega de datos de SQL Server, accediendo de forma eficiente a los datos.

Se implementó una bodega de datos que consta de dos esquemas principales: gestión de datos transaccionales provenientes de la herramienta para el centro de contacto y recolección de la información de tipificación y agendamiento de citas médicas, guardando información de la campaña de servicio al usuario y atención a los pacientes de la Clínica de la Colina, lo que disminuyó el uso de horas hombre en la generación de reportes, centralizó y democratizó el acceso a los datos clave de la campaña.

Objetivo 2. Desarrollar dos modelos de Aprendizaje automático para incrementar y mejorar el entendimiento y producción del agendamiento de citas en salud.

Se generaron dos modelos de aprendizaje automático que permiten, por un lado, determinar un pronóstico del agendamiento de citas médicas con precisión. De igual forma conocer el segmento al que el paciente pertenece y los patrones de comportamiento. Los modelos de aprendizaje automático se encuentran en evaluación frente a sus resultados de los próximos meses.

Objetivo 3. Integrar en una aplicación web estos desarrollos para dar valor agregado a la compañía permitiendo a los clientes monitorear la gestión en agendamiento de citas en el sector salud.

Se creó una visualización en Power BI contenida en aplicación web en la página de la compañía que genera un valor agregado, ya que muestra a las clínicas qué se está haciendo diariamente y cómo se ha avanzado. Anteriormente, los clientes tenían que esperar al final del mes para recibir un reporte, no tenían conocimiento de la gestión diaria de la campaña. Este aplicativo reporta los indicadores claves que permiten evaluar diariamente la labor de los asesores por parte de los directivos de la campaña.

Objetivo general. Optimizar la productividad de la compañía E Call You S.A.S. en el agendamiento de citas médicas mediante la implementación de un modelo de analítica de datos

para segmentar y entender el comportamiento de los pacientes, así como pronosticar el volumen de agendamiento.

Este proyecto hace parte de un proceso de innovación y transformación digital con modelos analíticos que generan tiempos óptimos de trabajo para los implicados, construye un panorama amplio para la toma de decisiones de las partes involucradas y la apertura hacia la generación de estrategias desde los datos.

Se debe dar especial seguimiento en los meses venideros a la evaluación de los modelos que pronostican volúmenes de llamadas y al refinamiento de la segmentación de los pacientes.

7 Referencias bibliográficas

[1] Joyanes Aguilar, L. Inteligencia de negocios y analítica de datos: una visión global de business intelligence & analytics, cap. 1 Inteligencia de negocios. Una panorámica global. Marcombo, Ediciones Técnicas, Alfa omega. 2019.

[2] Devops and Agile Methodologies in your Software Development Life Cycle [en línea]. Disponible en: <https://medium.com/@DENKENSOLUTIONS/devops-and-agile-methodologies-in-your-development-life-cycle-52a3f6478e0b> 2020.

[3] RAE: Definición “paciente” [en línea]. Disponible en <https://dle.rae.es/paciente> 2020.

[4] Varían, H. Análisis Microeconómico. Antoni Bosch, tercera ed. 1998.

[5] Diamond, P., and H. Vartiainen: Behavioral Economics and its Applications. Princeton University Press, 2007.

[6] Koenker R, Hallock K, ”Quantile Regression”. Journal of Economic Perspectives, v. 15, n. 4, pp. 143-156. 2001.

[7] Finlay, S. Predictive Analytics, Data Mining and Big Data Myths, Misconceptions and Methods, 2014.

[8] Pindyck, R, and Rubinfeld, D. Econometric Methods and Economic Forecasts, McGraw-Hill. 1998.

[9] Alpaydin, Ethem. Introduction to Machine Learning. MIT Press. p. 9. 2010.

[10] Venu Govindaraju. Introduction to Hidden Markov Models [en línea]. Disponible en: https://www.researchgate.net/profile/Venu_Govindaraju/publication/7252786_Hidden_Markov_models_combining_discrete_symbols_and_continuous_attributes_in_handwriting_recognition/links/02bfe5117e059ee6cc000000/Hidden-Markov-models-combining-discrete-symbols-and-continuous-attributes-in-handwriting-recognition.pdf 2019.

- [11] Time Series. The R project for statistical computing [en línea]. Disponible en: <https://rpubs.com/palominom/series> 2020.
- [12] Botero, S. and Cano, J. Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia. Disponible en: <http://www.scielo.org.co/pdf/ceco/v27n48/v27n48a07.pdf> 2008.
- [13] Skorupa, G. Forecasting Time Series with Multiple Seasonalities using TBATS in Python, Disponible en: <https://medium.com/intive-developers/forecasting-time-series-with-multiple-seasonalities-using-tbats-in-python-398a00ac0e8a> 2019.
- [14] MathWorks. Machine Learning. Tres cosas que es necesario saber [en línea] Disponible en: <https://es.mathworks.com/discovery/machine-learning.html> 2020.
- [15] Nadim, J. Top 5 best programming languages for artificial intelligence field [en línea]. Geeks for Geeks. Disponible en: <https://www.geeksforgeeks.org/top-5-best-programming-languages-for-artificial-intelligence-field/> 2019.
- [16] The R Foundation. The R project for statistical computing [en línea]. Disponible en: <https://www.r-project.org/> 2020.
- [17] Python Software Foundation. The Python Language Reference [en línea]. Disponible en: <https://www.python.org/> 2020.
- [18] Shiny from Rstudio. Interact. Analyze. Communicate [en línea]. Disponible en: <https://shiny.rstudio.com/> 2020.
- [19] Microsoft Power BI. Obtenga claridad cuando más la necesita [en línea]. Disponible en: <https://powerbi.microsoft.com/es-es/> 2020.
- [20] Microsoft: Access SQL: conceptos básicos, vocabulario y sintaxis [en línea]. Disponible en: <https://support.office.com/es-es/article/access-sql-conceptos-b%C3%A1sicos-vocabulario-y-sintaxis-444d0303-cde1-424e-9a74-e8dc3e460671#bm1> 2020
- [21] Croll, A. and Yoskovitz, B. Lean Analytics, Use Data to Build a Better startup Faster, 2013.
- [22] Transformación digital en la nube, Servcloud [en línea]. Disponible en: <https://servinformacion.com/servicloud/> 2020.

-
- [23] Morales, P. Inteligencia artificial en procesos de cobranza de grandes empresas: más allá del Bot [e-Book]. Providencia, Chile: Digevo Corp. Disponible por descarga en: <http://soluciones.digevo.com/ebook-inteligenciaartificial-en-procesos-de-cobranza-2> 2019.
- [24] Marbán, O. Mariscal, G and Segovia, J. A Data Mining and Knowledge Discovery Process Model, Data Mining and Knowledge Discovery in Real Life Applications, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-53-0, InTech, Disponible en: http://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/a_data_mining_amp_knowledge_discovery_process_model 2019.
- [25] Cook, K. Understand the machine learning from scratch for beginners [en línea], 24 de septiembre]. House of Bots. Disponible en es.mathworks.com/discovery/machine-learning.html 2018.
- [26] Lynn, J., Straube, B., Bell, K., Jencks, S. and Kambic, R. Using Population Segmentation to Provide Better Health Care for All: The “Bridges to Health” Model, pp. 185-208. 2007.
- [27] Anderson, D., and Bjarnadóttir, M. When is an ounce of prevention worth a pound of cure? Identifying high-risk candidates for case management. IIE Transactions on Healthcare Systems Engineering, 6(1), 22–32. 2016
- [28] Hu, X., Sean Barnes, S. Bjarnadóttir, M. and Golden. B. Intelligent selection of frequent emergency department patients for case management: A machine learning framework based on claims data, ISSN: 2472-5579 (Print) 2472-5587 En línea en: Journal homepage: <https://doi.org/10.1080/24725579.2017.1351502> 2017.
- [29] Pereira, M., Singh, V., Hon, C. P., McKelvey, T. G., Sushmita, S., and De Cock, M. Predicting future frequent users of Emergency Departments in California State. Proceeding of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB ‘16), Seattle, WA, USA, 603.610. ACM, New York, NY, USA. 2016.
- [30] Williams, S. and Heller, A. Patient activation among Medicare beneficiaries: Segmentation to promote informed health care decision making. International Journal of Pharmaceutical and Healthcare Marketing, 1(2), 199–213. 2007.
- [31] Morrison, M., Murphy, T., Nalder, C. Consumer preferences for general practitioner services. Health Marketing Quarterly, 20(3), 3–19. 2003
- [32] Dong, G., Taslimitehrani, V. Pattern aided classification. Proceedings of the 2016 SIAM International Conference on Data Mining, 225–233. 2016.

