



PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE DE INGENIERÍA

Trabajo de Titulación presentado como requisito para la obtención del título
de Magíster en Sistemas de Información mención en Data Science

**Análisis de la contratación y asignación de ingenieros de campo
en una empresa de tecnología médica**

David de Jesús Ponce Ortiz

ddponce@puce.edu.ec

DIRECTOR: Christian José Tipantuña Tenelema

christian.tipantuna@epn.edu.ec

DMQ, agosto 2024

DECLARACIÓN Y AUTORIZACIÓN

Yo, David de Jesús Ponce Ortiz, con cédula de identidad 1723467831, autor del trabajo de graduación titulado: **“Análisis de la contratación y asignación de ingenieros de campo en una empresa de tecnología médica”**, previo a la obtención del título de Magíster en Sistemas de Información mención en Data Science, de la Facultad de Ingeniería:

1. Declaro tener pleno conocimiento de la obligación que tiene la Pontificia Universidad Católica del Ecuador, de conformidad con el artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de graduación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.
2. Autorizo a la Pontificia Universidad Católica del Ecuador a difundir a través de sitio web de la Biblioteca de la PUCE, el referido trabajo de graduación, respetando las políticas de propiedad intelectual de Universidad.

Quito, 28 de agosto de 2024

Atentamente,



David de Jesús Ponce Ortiz
Autor Trabajo de Titulación

APROBACIÓN DEL TUTOR

En mi carácter de Director – Tutor del Trabajo de Posgrado Titulado: **“Análisis de la contratación y asignación de ingenieros de campo en una empresa de tecnología médica”**, presentado por el maestrante DAVID DE JESÚS PONCE ORTIZ, titular de la Cédula de Identidad N° 1723467831 para optar al Grado de Magíster en Magíster en Sistemas de Información mención en Data Science, considero que dicho Trabajo de Investigación reúne los requisitos y méritos suficientes para ser sometido a la evaluación por parte de los Lectores – Evaluadores que se designen para tal fin por parte de las autoridades de la Facultad de Ciencias de la Educación.

En la ciudad de Quito, 28 de agosto de 2024



CHRISTIAN JOSÉ TIPANTUÑA TENELEMA, Ph. D.

DIRECTOR DE PROYECTO

C.I. 1720979093

Telf: 0995719229

Nota:

Se comunica que en el servicio de análisis Turnitin, el referido trabajo de titulación alcanzó el siguiente resultado: 7% índice de similitud con otras fuentes.

Turnitin Originality Report	
Processed on: 25-Aug-2024 2:58 PM -05 ID: 2437311813 Word Count: 19192 Submitted: 2	
Tesis master TI PUCE DP By DAVID DE JESÚS PONCE ORTIZ	
Similarity Index 7%	Similarity by Source Internet Sources: 6% Publications: 1% Student Papers: 3%

1% match (Internet from 28-Aug-2023) http://repositorio.puce.edu.ec/bitstream/handle/22000/21723/Sinailin%20Sotalin%20Vilma%20Lisbeth.pdf?isAllowed=y&sequence=1
1% match (Internet from 30-Dec-2023) https://dspace.pucesi.edu.ec/bitstream/11010/1035/3/TESIS%20MAURICIO%20PINTO.pdf
< 1% match (Internet from 26-Oct-2023) http://repositorio.puce.edu.ec/bitstream/handle/22000/21782/Borja%20Sarango%20Sof%20c3%ada%20Paulina.pdf?isAllowed=y&sequence=1
< 1% match (Internet from 30-Sep-2023) http://repositorio.puce.edu.ec/bitstream/handle/22000/21907/Zapata%20Rodriguez%20Necser%20Alejandro.pdf?isAllowed=y&sequence=1
< 1% match (Internet from 23-Oct-2022) https://www.conacyt.gov.py/sites/default/files/Tesis-Osvaldo_Gonzalez.pdf
< 1% match (student papers from 31-Jul-2024) Submitted to National University College - Online on 2024-07-31
< 1% match (student papers from 18-Aug-2024) Submitted to National University College - Online on 2024-08-18
< 1% match (student papers from 16-Aug-2024) Submitted to UPAP: Universidad Popular Autónoma del Estado de Puebla on 2024-08-16
< 1% match () Espinosa Muñoz, Javier Isaac. "Una aproximación a la predicción del valor de acciones en la bolsa de valores aplicando técnicas de Data Mining", E.T.S. de Ingenieros Informáticos (UPM), 2015
< 1% match (Internet from 20-Aug-2022) https://oa.upm.es/1088/1/DIEGO_GUILLELMO_MANZANAL.pdf
< 1% match (Internet from 09-Jun-2024) https://repositorio.upse.edu.ec/bitstream/46000/11213/1/UPSE-MTI-2024-0011.pdf
< 1% match (student papers from 16-May-2024) Submitted to Universidad Francisco de Vitoria on 2024-05-16
< 1% match (Internet from 28-Jun-2024) https://cathi.uacj.mx/bitstream/handle/20_500.11961/26475/Tomo%2b07%2b-%2bIngenier%2bc3%adas%2b-%2bArt%2bc3%aduculos%2bde%2bCongreso%2bAcademia%2bJournals%2bChetumal%2b2023%20p%20c3%a1gina%20de%20la%20147%20isAllowed=y&sequence=1
< 1% match (Internet from 06-Oct-2021) https://www.zetes.com/es/soluciones-end-to-end/soluciones-end-to-end/identificar-renovar-optimizar-maneras-efectivas-de
< 1% match (Internet from 19-Aug-2024) https://repositorio.utm.edu.ec/bitstream/123456789/16332/2/04%20SOF%20056%20TRABAJO%20GRADO.pdf
< 1% match () Álvarez Osorio, Francisco Javier. "Estudio de técnicas avanzadas de análisis de datos para el proceso de auditoría informática en el módulo del sistema académico de la Universidad Técnica del Norte (UTN)", 2023
< 1% match (Internet from 11-Nov-2020) https://www.juanbarrios.com/que-paises-estan-perdiendo-la-guerra-contra-el-covid19/
< 1% match (student papers from 16-Jul-2024) Submitted to Universidad Mariano Gálvez de Guatemala on 2024-07-16
< 1% match (student papers from 28-Jul-2024) Submitted to Universidad Mariano Gálvez de Guatemala on 2024-07-28
< 1% match (Internet from 22-Feb-2023) https://www.researchgate.net/publication/344090659_Athlete's_philanthropy_and_social_responsibility_communication_on_social_media_19
< 1% match (Internet from 02-Feb-2023) https://www.researchgate.net/publication/46533917_La_distribucion_funcional_de_la_renta_en_Espana_entre_1914_y_1959_una_propue
< 1% match (Internet from 20-Dec-2022) http://repositorio.ug.edu.ec/bitstream/redug/63193/1/BMAT-MIS%2006-2022-Ing.%20CIVIL%20-%20CUZCO%20CHOA%20CINDY%20DENNY.pdf

DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, David de Jesús Ponce Ortiz, con cédula de ciudadanía No. 1723467831, declaro que los resultados obtenidos en la investigación **“Análisis de la contratación y asignación de ingenieros de campo en una empresa de tecnología médica”**, como requisito previo para lo obtención del Grado Académico de Magíster en Sistemas de Información mención Data Science son absolutamente originales, auténticos y personales.

En tal virtud, declaro que el contenido, las conclusiones y los efectos legales y académicos, que se desprenden del trabajo de investigación, y luego de la redacción de este documento, son y serán de mi sola y exclusiva responsabilidad legal y académica.

En la ciudad de Quito, 28 de agosto de 2024

Atentamente,



David de Jesús Ponce Ortiz
Autor Trabajo de Titulación

AGRADECIMIENTO

Agradezco a todas las personas que me acompañaron a lo largo de la maestría y me ayudaron a crecer personal y profesionalmente; principalmente agradecer a Dios, a mis padres, a mis hermanos y mis perritos.

También agradecer al ingeniero Davy Rueda y al equipo de Customer Services en Ecuador por todo el acompañamiento desde inicio y fin en el proyecto y por su feedback de mejora continua y al ingeniero Christian Tipantuña por su soporte dentro del trabajo.

Gracias Nubia por todo tu apoyo y ayuda.

Gracias Siemens Healthineers Ecuador.

#ProudToBeAHealthineer

ÍNDICE DE CONTENIDO

1	INTRODUCCIÓN	1
1.1	Antecedentes	1
1.2	Objetivo general	1
1.3	Objetivos específicos	1
1.4	Justificación	2
2	MARCO TEÓRICO	3
2.1	CRISP-DM	3
2.1.1	Ventajas CRISP-DM	4
2.2	Algoritmo K-Means	5
2.2.1	Funcionamiento de K-Means	6
2.2.2	Ventajas de K-Means	7
2.2.3	Desventajas de K-Means	7
2.3	Algoritmo Agrupamiento Aglomerativo	8
2.4	Algoritmo Gaussian Mixture Model	9
2.4.1	Parámetros de Entrada	9
2.4.2	Ventajas de GMM	9
2.4.3	Desventajas de GMM	9
2.5	Herramientas utilizadas en Data Science	10
2.5.1	Python	10
2.6	La empresa: Siemens Healthineers	11
2.6.1	Misión	12
2.6.2	Visión	12
2.6.3	Principales áreas de negocio y productos	12
2.6.4	Siemens Healthineers en Latinoamérica	13
2.6.5	Siemens Healthineers en Ecuador	14
2.6.6	SAP CRM	16
3	METODOLOGÍA	18
3.1	Caso de estudio I: Contratación de nuevo personal	18
3.2	Fase 1: Comprensión del negocio	19
3.2.1	Objetivos de negocio	19

3.2.2	Evaluación de la situación actual	19
3.3	Fase 2: Comprensión de los datos	20
3.3.1	Captura de datos	21
3.3.2	Descripción de datos	21
3.3.3	Exploración de datos	22
3.4	Fase 3: Preparación de los datos	27
3.4.1	Filtrado y limpieza de datos	27
3.4.2	Anonimización de datos	27
3.5	Fase 4: Modelado	28
3.5.1	Análisis previo de horas para determinar la cantidad de ingenieros de campo	29
3.5.2	Código base	31
3.5.3	K-Means	32
3.5.4	Agglomerative Clustering	33
3.5.5	Gaussian Mixture Model	34
3.6	Fase 5: Evaluación	35
3.6.1	Análisis teórico	35
3.6.2	Análisis experimental	38
3.6.3	Justificación	48
3.7	Fase 6: Despliegue	48
3.7.1	Diseño de la interfaz gráfica en Python	48
3.7.2	Ejecución del modelo y generación del mapa	49
3.8	Caso de estudio II: Asignación en los servicios de campo	53
3.8.1	Análisis previo	53
3.8.2	Implementación asignación de CSEs	55
3.9	Future Work	59
4	CONCLUSIONES Y RECOMENDACIONES	61
4.1	Conclusiones	61
4.2	Recomendaciones	62
5	REFERENCIAS BIBLIOGRÁFICAS	64
6	ANEXOS	I

ÍNDICE DE FIGURAS

2.1	Esquema estándar del ciclo CRISP-DM (Haya, 2021).	3
2.2	Diagrama de flujo del algoritmo K-Means (Soni & Patel, 2017).	6
2.3	Diagrama de flujo agrupamiento aglomerativo (Kuchiki, 2008).	8
2.4	KPIs LAM (Quintero, 2023).	13
2.5	Línea de negocio de equipos de Imágenes (Quintero, 2023).	15
2.6	Módulos más importantes en SAP (Ejecutiva, 2021).	16
3.1	BPMN of CRISP-DM framework (Riess, 2022).	18
3.2	Histogramas de las variables numéricas.	24
3.3	Diagrama de cajas de las variables numéricas.	24
3.4	Heatmap de la cantidad de equipos por ciudad.	25
3.5	Representación gráfica coeficientes de Silueta y Davies-Bouldin.	37
3.6	Representación gráfica del coeficiente Calinski-Harabas.	37
3.7	Ubicación geográfica de 2 CSEs con K-Means.	39
3.8	Plano cartesiano: ubicación geográfica de 2 CSEs con K-Means.	40
3.9	Ubicación geográfica de 2 CSEs con Agrupamiento Aglomerativo.	41
3.10	Plano cartesiano: ubicación geográfica de 2 CSEs con Agrupamiento Aglomerativo.	42
3.11	Ubicación geográfica de 2 CSEs con GMM.	43
3.12	Plano cartesiano: ubicación geográfica de 2 CSEs con GMM.	44
3.13	Ubicación geográfica de 3 CSEs con K-Means.	45
3.14	Plano cartesiano: ubicación geográfica de 3 CSEs con K-Means.	46
3.15	Ubicación geográfica de 3 CSEs con Agrupamiento Aglomerativo.	47
3.16	Ubicación geográfica de 3 CSEs con GMM.	47
3.17	Diagrama de flujo de la interfaz GUI implementada.	49
3.18	Pantalla de inicio interfaz implementada.	51
3.19	Ventana/Pop Up emergente.	51
3.20	Simulación final presentadas en un ejemplo con 3 centroides.	52
3.21	Visualización de CSEs por ciudad vs demanda de horas anuales.	54
3.22	Simulación asignación de CSEs.	58

ÍNDICE DE TABLAS

3.1	Descripción atributos del dataset.	22
3.2	Análisis descriptivo I.	23
3.3	Análisis descriptivo II.	23
3.4	Cantidad de equipos por división.	24
3.5	Cantidad de equipos activos por división.	25
3.6	Cantidad de equipos por ciudad y división.	26
3.7	Cantidad de horas por modalidad de equipos con contrato activo.	29
3.8	Resultados evaluación modelos predictivos.	36
3.9	Experiencia CSE por modalidad.	55

ÍNDICE DE CÓDIGOS FUENTE

3.1	EDA.	22
3.2	Limpieza previa de datos.	27
3.3	Aplicación técnica de anonimización de datos por seudonimización.	28
3.4	Cálculo de horas anuales.	29
3.5	Código base modelos predictivos.	31
3.6	Modelo predictivo con K-Means.	32
3.7	Código base Agrupamiento Aglomerativo.	33
3.8	Código base GMM.	34
3.9	Evaluación algoritmos implementados.	36
3.10	Implementación GUI con Tkinter.	50
3.11	Implementación GUI con TKinter II.	53
3.12	Implementación GUI asignación de CSE I.	56

RESUMEN

Los requerimientos de automatización, optimización y digitalización de procesos se han vuelto clave en la mejora de productividad y ganancias dentro de una empresa. En este contexto, la optimización de la asignación de recursos, como los ingenieros de campo, es fundamental para garantizar un servicio técnico eficiente y de calidad en cualquier área, sea de salud, automotriz, telecomunicaciones, tecnologías de la información, entre otras. El trabajo de titulación presenta un modelo predictivo para la asignación y contratación de CSEs en Siemens Healthineers Ecuador, considerando factores geográficos, experiencia del técnico y la demanda de servicio. Se implementa la metodología CRISP-DM para guiar el proceso de minería de datos, desde la comprensión del negocio hasta el despliegue de la solución. Se exploraron diversos modelos de agrupamiento, incluyendo K-Means, Modelo Gausiano Mixto y Agrupamiento Aglomerativo, para determinar la ubicación óptima de los CSEs en función de la demanda de servicio y la distribución geográfica de los equipos.

La evaluación de los modelos se realizó tanto teóricamente, utilizando métricas como el coeficiente de silueta y experimentalmente, considerando factores como la distancia a los equipos y el equilibrio de la carga de trabajo. K-Means se destacó por su capacidad para incorporar criterios específicos del negocio a través de un sistema de pesos, lo que permitió una asignación más estratégica y eficiente de los CSEs. El resultado final fue un modelo predictivo que, al ser integrado en una interfaz de usuario intuitiva, facilita la toma de decisiones informadas y contribuye a mejorar la calidad del servicio técnico y la satisfacción del cliente.

PALABRAS CLAVE: Minería de datos, CRISP-DM, K-Means, tecnología médica, Siemens Healthineers.

ABSTRACT

The requirements for automation, optimization and digitalization of processes have become key in improving productivity and profits within a company. In this context, the optimization of the allocation of resources, such as field engineers, is essential to guarantee an efficient and quality technical service in any area, be it health, automotive, telecommunications, information technology, among others. The thesis presents a predictive model for the assignment and hiring of CSEs at Siemens Healthineers Ecuador, considering geographic factors, technician experience and service demand. The CRISP-DM methodology is implemented to guide the data mining process, from understanding the business to deploying the solution. Various clustering models were explored, including K-Means, Mixed Gaussian Model and Agglomerative Clustering, to determine the optimal location of the CSEs based on service demand and geographic distribution of the equipment.

The evaluation of the models was conducted both theoretically, using metrics such as the silhouette coefficient, and experimentally, considering factors such as the distance to equipment and workload balance. K-Means stood out for its ability to incorporate specific business criteria through a weighting system, allowing for a more strategic and efficient allocation of CSEs. The final result was a predictive model that, when integrated into an intuitive user interface, facilitates informed decision-making and contributes to improving the quality of technical service and customer satisfaction.

KEYWORDS: Data Mining, CRISP-DM, K-Means, Siemens Healthineers, Medical Technology.

1 INTRODUCCIÓN

1.1 ANTECEDENTES

El campo de la ciencia de datos y la minería de datos han experimentado un crecimiento significativo en los últimos años. La aplicación de técnicas de análisis de datos para la toma de decisiones empresariales se ha convertido en un factor clave para la optimización y eficiencia de las operaciones. En este contexto, el proceso CRISP-DM (Cross Industry Standard Process for Data Mining) ha emergido como un marco estándar que guía la implementación exitosa de proyectos de minería de datos y modelado predictivo (Espinosa, 2020). También, en el ámbito empresarial, la gestión eficiente de recursos humanos y la asignación óptima de personal son cruciales para la eficiencia operativa. Empresas de diversos sectores han recurrido al uso de modelos predictivos para optimizar la asignación de recursos y mejorar la QoS (del inglés, Quality of Service).

1.2 OBJETIVO GENERAL

Desarrollar un modelo predictivo para la asignación óptima de Ingenieros de Campo (CSEs) en el mantenimiento de equipos médicos en Ecuador, mediante la integración de factores geográficos, demanda por ciudad y la experiencia específica de los CSEs, con el fin de mejorar la eficiencia operativa y la calidad del servicio en instituciones de salud.

1.3 OBJETIVOS ESPECÍFICOS

1. Describir los fundamentos teóricos asociados conceptos y fases del proceso CRISP-DM (Cross Industry Standard Process for Data Mining) para el desarrollo del modelo predictivo en la asignación de Ingenieros de Campo (CSEs) en mantenimiento de equipos médicos.
2. Recopilar información acerca de considerando factores geográficos de los clientes y CSEs y la demanda de horas anuales por cada modalidad en cada ciudad.
3. Procesar la información recopilada afín de conocer la situación actual de la empresa, empleando herramientas de visualización de datos para identificar patrones, tendencias y necesidades específicas en la asignación de CSEs.

4. Desarrollar un modelo predictivo para la contratación de CSEs, integrando variables como la ubicación geográfica de los equipos médicos, la demanda de mantenimiento por provincia y la disponibilidad de ingenieros.
5. Analizar los resultados derivados del modelo predictivo, comparando la asignación de CSEs antes y después de la implementación del modelo, extrayendo información relevante para comprender las mejoras y desafíos en la optimización de tiempos y distancias en la asignación de recursos para el mantenimiento de equipos médicos

1.4 JUSTIFICACIÓN

La presente propuesta de trabajo de titulación, centrada en la creación de un modelo predictivo para apoyar el proceso de contratación de ingenieros de campo, surge como respuesta a desafíos cruciales en la gestión eficiente de recursos y la mejora continua en el servicio. La originalidad de este proyecto radica en que no solo busca abordar la asignación geográfica de los CSEs, sino que también considera factores como la cantidad de horas anuales que demanda los mantenimientos preventivos y correctivos de la empresa junto con necesidades específicas de cada ciudad. Este enfoque integral distingue esta propuesta, proporcionando una solución más completa y adaptable a la dinámica cambiante de la demanda de servicios, como también la automatización y digitalización de la asignación de CSEs a los servicios de campo, dado que hoy en día es un proceso manual. Desde una perspectiva teórica, el trabajo contribuirá al desarrollo de modelos predictivos en el ámbito de asignación de recursos en mantenimiento de equipos médicos, aportando a la creación de conocimiento y prácticas innovadoras en el ámbito de la salud.

Los resultados obtenidos influirán de manera directa en la optimización de los procesos de asignación de CSEs, permitiendo a la empresa anticiparse a las demandas específicas de cada ciudad y modalidad de mantenimiento. Esto se traducirá en una respuesta más ágil y efectiva ante las necesidades de los clientes, fortaleciendo la ventaja competitiva de la compañía en la industria de servicios médicos. Los hallazgos de esta investigación podrían sentar las bases para futuros desarrollos en la gestión de servicios técnicos en otras industrias y sectores.

Esta propuesta de trabajo de titulación se presenta como una iniciativa integral y aplicativa, con el potencial de generar conocimientos valiosos, contribuir a prácticas innovadoras y brindar soluciones prácticas a desafíos significativos en el ámbito de asignación de recursos en mantenimiento de equipos médicos.

2 MARCO TEÓRICO

Este capítulo presenta los conceptos asociados a las fases del proceso CRISP-DM así como los fundamentos teóricos asociados de los 3 algoritmos a implementarse: i) K-Means, ii) Agglomerative Clustering y iii) Gaussian Mixture Model (GMM). Además, se presenta información sobre la empresa Siemens Healthineers y una introducción a su sistema CRM, lo cual es fundamental para contextualizar el estudio y entender el entorno en el que se aplicarán estos algoritmos.

2.1 CRISP-DM

El modelo de proceso actual para la minería de datos (CRISP-DM) se divide en seis etapas, lo cual permite estructurar el proceso según se requiera. Este modelo es adaptable e iterativo, ya que facilita la creación de un modelo de minería de datos ajustado a las necesidades, debido a que en algunos casos se avanza o retrocede entre fases si se requiere. Por lo tanto, la secuencia de fases no es necesario cumplirla estrictamente (Chapman, 2019) (Hernández, 2024). La Figura 2.1 representa las fases del modelo CRISP-DM llevadas a cabo en proyectos de minería de datos.

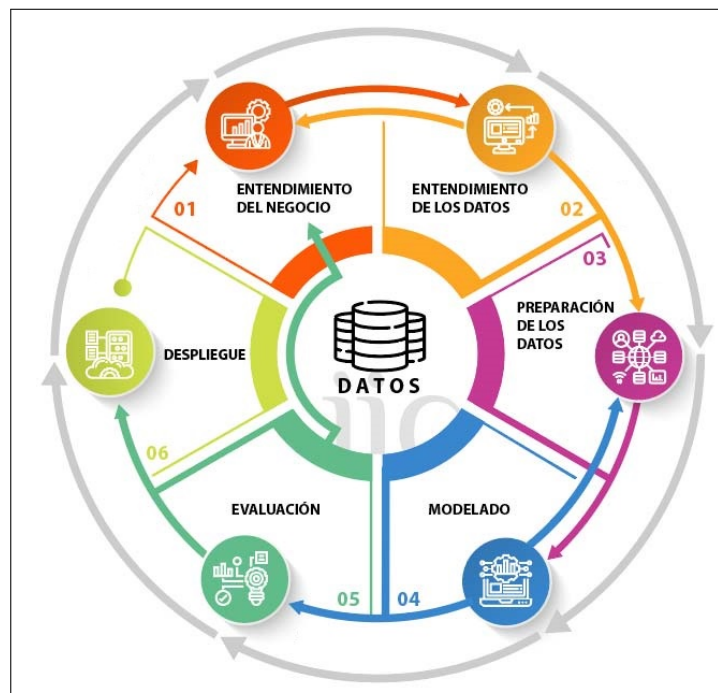


Figura 2.1: Esquema estándar del ciclo CRISP-DM (Haya, 2021).

1. **Comprensión empresarial:** Esta fase implica conocer los datos, es decir, se realiza un análisis de datos descriptivos y exploratorios e identificación de problemas con la calidad de los datos. Además, es una investigación exploratoria a medida que se inicia a comprender los matices y el potencial de los datos (Arias et al., 2021).
2. **Entendimiento del negocio:** En esta etapa, el analista de datos debe comprender a fondo el contexto comercial de la solicitud: i) dónde está el punto débil que estamos tratando de “curar”, ii) qué hemos hecho en el pasado, iii) quiénes son los actores, identificar los riesgos, recursos, etc. y los criterios adecuados para el éxito del proyecto (Arias et al., 2021).
3. **Preparación de datos:** La esencia de esta fase es seleccionar los datos con los que desea trabajar, justificar su inclusión/exclusión, limpiar y transformar los datos a un formato apropiado para el respectivo análisis. Todo esto es muy importante para obtener un buen resultado (Wirth & Hipp, 2000).
4. **Modelado:** Consiste en crear un cierto umbral para la métrica “verdadero o falso” (por ejemplo, si el proyecto es comprender o determinar las razones de la salida de clientes). En esta etapa, se aplican varias técnicas de modelado a los datos preparados para que pueda compararlos entre sí y comprender cuáles tienen más éxito (Wirth & Hipp, 2000).
5. **Evaluación:** Implica evaluar críticamente en función de los objetivos comerciales y los criterios de éxito establecidos en la etapa uno. Esto proporciona información que se puede utilizar para retroceder y redefinir la comprensión del negocio (Arias et al., 2021).
6. **Despliegue:** La etapa final implica aplicar el modelo a los datos y situaciones del mundo real, poner en práctica de manera efectiva los resultados del análisis y comenzar a utilizar lo aprendido para optimizar el desempeño del equipo (Wirth & Hipp, 2000).

La metodología CRISP DM aumenta la probabilidad de que el análisis sea más confiable porque obliga a seguir todos estos pasos, dejando espacio para la creatividad. A continuación se presentan las ventajas de esta metodología.

2.1.1 Ventajas CRISP-DM

Las ventajas de CRISP-DM son numerosas y la aplicación del enfoque dentro de una empresa proporciona (García, 201):

- ❑ Respuesta a desafíos clave para los equipos de ciencia de datos en términos de comprensión de datos e implementación.

- ❑ Es simple y accesible que permite participar activamente a los miembros del equipo sin conocimientos de ciencia de datos.

- ❑ Permite comprender las especificaciones de cada proyecto, es adaptable y además iterativo que fomenta el compromiso de los miembros del equipo.

A continuación se presenta el marco teórico referente a los 3 modelos predictivos implementados. Estos métodos fueron seleccionados debido a su capacidad para agrupar los datos en un número pre-especificado de clusters, lo cual es fundamental para establecer el número exacto de CSEs necesarios en cada zona. A diferencia de DBSCAN, que identifica clusters de forma automática sin un número fijo de grupos (Dharni & Bnasal, 2013), estos algoritmos permiten un mayor control sobre la solución, adaptándose mejor a los requerimientos del trabajo presentado.

2.2 ALGORITMO K-MEANS

K-Means clustering es un algoritmo de aprendizaje no supervisado, de todos los algoritmos de aprendizaje no supervisados K-Means puede ser el más utilizado debido a su potencia y simplicidad (Amat, 2020).

K-Means permite agrupar los datos ágil y eficazmente en pocas iteraciones. Funciona etiquetando todas las instancias de un clúster con el centroide más cercano. Cuando las instancias se concentran alrededor de un punto específico, ese punto se llama centroide (Xu & Tian, 2015).

Si se obtienen las etiquetas de las instancias, K-Means puede encontrar fácilmente todos los elementos promediando todas las instancias de cada grupo. Aquí no se da una etiqueta ni centroides, por lo que se inicia colocando los centroides aleatoriamente seleccionando k instancias aleatorias y usando sus posiciones como centroides (Kang, 2016).

Se emplea en una amplia gama de aplicaciones, tales como: i) segmentación de clientes, ii) análisis de los datos, iii) reducción de dimensiones, iv) detección de anomalías, v) aprendizaje semisupervisado, vi) búsqueda de imágenes y vii) segmentación de imagen.

2.2.1 Funcionamiento de K-Means

La Figura 2.2, presenta el diagrama de flujo del funcionamiento K-Means (Soni & Patel, 2017).

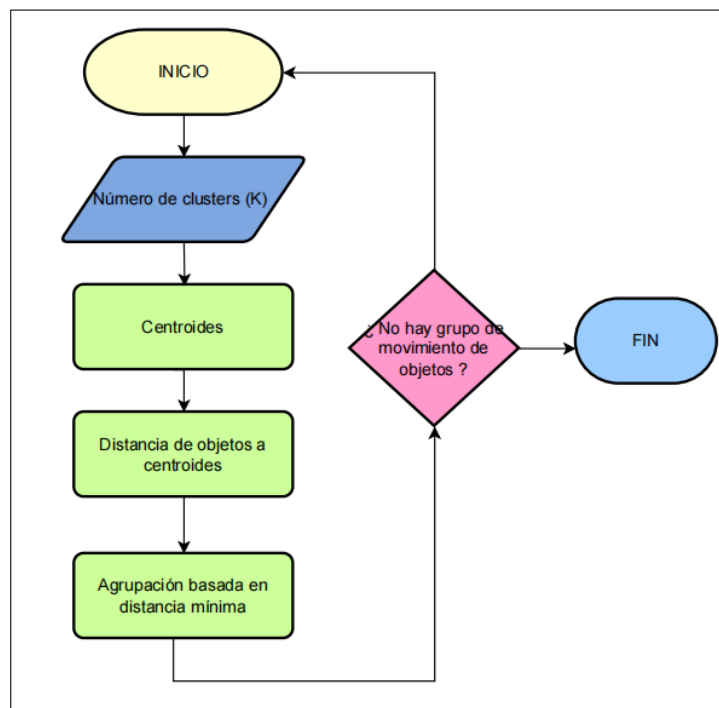


Figura 2.2: Diagrama de flujo del algoritmo K-Means (Soni & Patel, 2017).

1. **Selecciona el número de clusters (K):** Determinar la cantidad de clusters en los que se dividirán los datos.
2. **Inicializa de centroides:** K puntos de los datos se seleccionan aleatoriamente para que sean los centroides iniciales de los grupos.
3. **Asigna de puntos a conglomerados:** Cada punto de datos se asigna al conglomerado cuyo centroide está más cercano a ese punto.
4. **Actualiza el centroide:** Para cada grupo, se calcula un nuevo centroide como el promedio de todos los puntos que pertenecen a ese grupo.
5. **Repetir el paso 3 y paso 4:** Estos pasos se repiten hasta que los centroides dejen de cambiar o se alcance el número máximo de iteraciones.

El proceso K-Means puede considerarse como un proceso iterativo que mejora gradualmente la división de datos en grupos. En cada paso, el algoritmo recalcula los centroides y

reassigna los puntos de datos a los centroides más cercanos. Este proceso continúa hasta que los cambios en los centroides se vuelven insignificantes o se alcanza el número máximo de iteraciones (Kang, 2016).

2.2.2 Ventajas de K-Means

- ❑ **Simplicidad y velocidad:** K-Means es fácil de implementar y funciona rápidamente incluso en grandes conjuntos de datos (Datapeaker, 2023).
- ❑ **Interpretabilidad:** Los resultados del método son fáciles de interpretar y visualizar (Datapeaker, 2023).
- ❑ **Flexibilidad:** K-Means se aplica a diferentes tipos de datos y problemas (Datapeaker, 2023).

K-Means es un algoritmo que lo caracteriza por la simplicidad y velocidad, el cual permite que su uso sea atractivo para muchas aplicaciones. Además, es fácil de implementar y usarse para procesar grandes cantidades de datos. La interpretabilidad de los resultados facilita la comprensión de cómo se dividieron los datos en grupos, lo cual es especialmente importante para tomar decisiones basadas en datos (Kang, 2016).

2.2.3 Desventajas de K-Means

- ❑ **Incertidumbre en el número de conglomerados:** El número de conglomerados K debe especificarse de antemano, lo que no siempre es obvio (Datapeaker, 2023).
- ❑ **Sensibilidad a los valores atípicos:** El método es susceptible a los valores fuera de lo común y al ruido en los datos. (Pills, 2023).

La dependencia de las condiciones iniciales puede conducir a resultados diferentes cada vez que se ejecuta el algoritmo. Esto se puede solucionar ejecutando el algoritmo varias veces con diferentes condiciones iniciales y eligiendo el mejor resultado. La incertidumbre en el número de conglomerados requiere el uso de métodos adicionales para determinar el valor óptimo de K. La sensibilidad a los valores atípicos se puede reducir mediante el preprocesamiento de datos (Kang, 2016).

2.3 ALGORITMO AGRUPAMIENTO AGLOMERATIVO

La Figura 2.3 presenta el diagrama de flujo del funcionamiento de agrupamiento jerárquico.

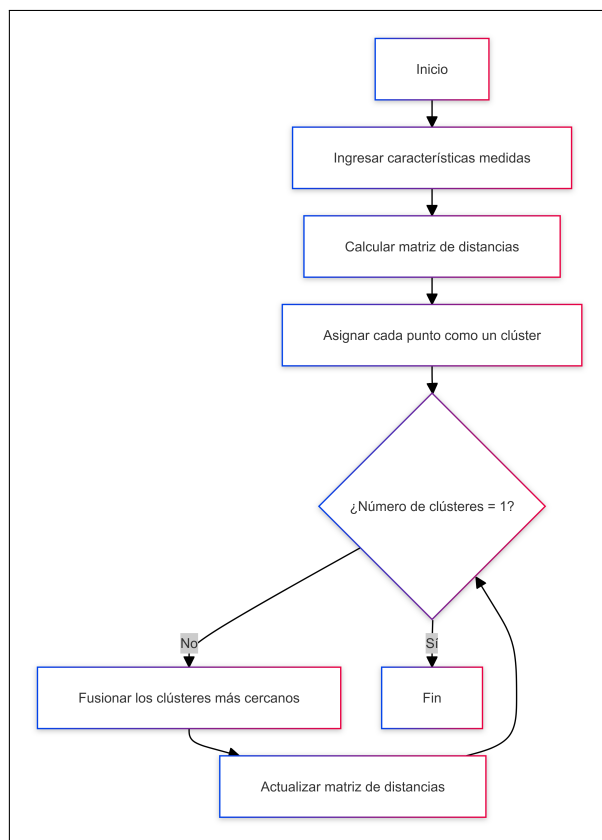


Figura 2.3: Diagrama de flujo agrupamiento aglomerativo (Kuchiki, 2008).

El Agrupamiento jerárquico es una familia de algoritmos de agrupamiento que construyen clústeres anidados mediante la fusión sucesiva de observaciones individuales o grupos de observaciones. Este proceso ascendente, también conocido como enfoque “bottom-up”, inicia tratando cada observación como un clúster separado y, en cada iteración, combina los dos clústeres más similares hasta que todos los puntos pertenecen a un único clúster. La similitud entre clústeres se mide mediante un criterio de enlace (linkage), que determina la estrategia de fusión. El resultado final es una jerarquía de clústeres representada como un dendrograma, donde la raíz representa el clúster que engloba todos los datos y las hojas representan los clústeres individuales (Ackermann et al., 2014).

2.4 ALGORITMO GAUSSIAN MIXTURE MODEL

El modelo de mezcla Gaussiana (GMM: Gaussian Mixture Model) es un método de aprendizaje automático no supervisado que se utiliza para representar la distribución de probabilidad de un conjunto de datos. La idea central detrás de GMM es que los datos se generan a partir de una mezcla de varias distribuciones gaussianas (normales), cada una con sus propios parámetros (media, covarianza y peso). El algoritmo busca estimar estos parámetros para cada componente gaussiano, de manera que la mezcla resultante se ajuste lo mejor posible a los datos observados (Reynolds et al., 2009).

2.4.1 Parámetros de Entrada

- ❑ **n_components:** El número de componentes gaussianos en la mezcla. Este parámetro define cuántos clústeres o grupos se intentarán identificar en los datos.
- ❑ **covariance_type:** El tipo de matriz de covarianza que se utilizará para modelar cada componente gaussiano.

2.4.2 Ventajas de GMM

- ❑ **Flexibilidad:** GMM puede modelar clústeres de diferentes formas y tamaños, lo que lo hace más versátil que algoritmos como K-Means, que asumen clústeres esféricos.
- ❑ **Asignación probabilística:** En lugar de asignar cada punto a un solo clúster, GMM calcula la probabilidad de que un punto sea parte de cada elemento de la mezcla. Esto permite una asignación más suave y una mejor representación de la incertidumbre en los datos.
- ❑ **Estimación de densidad:** GMM puede utilizarse para estimar la densidad de probabilidad de los datos, aplicable por ejemplo en el caso de estudio de detección de anomalías.

2.4.3 Desventajas de GMM

- ❑ **Sensibilidad a la inicialización:** El algoritmo EM utilizado para entrenar GMM puede converger a óptimos locales, por lo que la elección de la inicialización puede influir en los resultados.
- ❑ **Complejidad computacional:** El entrenamiento de GMM puede ser más costoso

computacionalmente que otros algoritmos de agrupamiento, especialmente para grandes conjuntos de datos o un gran número de componentes.

- ❑ **Número de componentes:** La elección del número de componentes puede ser un desafío y requiere técnicas adicionales como el Criterio de Información Bayesiano (BIC) o la validación cruzada.

2.5 HERRAMIENTAS UTILIZADAS EN DATA SCIENCE

Un científico de datos emplea diversas herramientas y lenguajes de programación que permiten el análisis, visualización y modelado de datos de manera efectiva. Entre los lenguajes más utilizados se encuentran Python, R y Matlab. Python es ampliamente valorado por su simplicidad y la amplia variedad de bibliotecas especializadas en el análisis de datos, como Scikit-learn, Pandas y Matplotlib. Por otro lado, es conocido por su fuerte enfoque en la estadística y su capacidad para realizar análisis complejos con facilidad. Matlab se utiliza frecuentemente en entornos académicos y de ingeniería debido a su potente capacidad de procesamiento numérico y su facilidad para trabajar con matrices. Estas herramientas son esenciales para los científicos de datos, ya que permiten convertir datos originales sin procesamiento y análisis en información útil, generar modelos predictivos y presentar resultados de manera comprensible (Lowndes et al., 2017).

2.5.1 Python

Python es un lenguaje de programación, popular y versátil en el campo de los sistemas de información. Su sintaxis sencilla y la extensa colección de bibliotecas hacen que sea ideal tanto para principiantes como para expertos. En el presente trabajo de titulación, se utiliza Python para desarrollar modelos predictivos y realizar análisis geospaciales, además de construir interfaces de usuario intuitivas a través de las siguientes bibliotecas (Van Rossum & Drake Jr, 1995).

2.5.1.1 Scikit-learn

Scikit-learn es una biblioteca de Python que proporciona una gama completa de herramientas para la modelización predictiva, incluyendo algoritmos para clasificación, regresión, clustering y reducción de dimensionalidad (Bloice & Holzinger, 2016). En particular, se utilizaron modelos de clustering como K-Means, GMM y Agrupamiento Aglomerativo para optimizar la contratación de CSEs según la demanda de servicio y la distribución geográfica de los equipos.

2.5.1.2 Folium

Para la visualización de datos geográficos, se empleó la biblioteca Folium, que permite generar mapas interactivos fácilmente integrables en aplicaciones web. Folium se basa en Leaflet.js y facilita la creación de mapas con marcadores, líneas y capas personalizadas (Bloice & Holzinger, 2016). En este contexto, Folium se utilizó para visualizar la ubicación óptima de los CSEs y analizar la distribución geográfica de la demanda de servicios técnicos.

2.5.1.3 Tkinter

Tkinter es la biblioteca predeterminada en Python para el desarrollo de interfaces gráficas de usuario. Permite desarrollar aplicaciones de escritorio con componentes interactivos como botones, cuadros de texto y menús (Bloice & Holzinger, 2016). A través de esta GUI, los usuarios pueden interactuar con el modelo predictivo y visualizar los resultados de manera clara y eficiente.

Ya establecido el marco teórico de la ciencia de datos, la siguiente sección se adentra en la aplicación práctica de estos conceptos aplicados al entendimiento del negocio.

2.6 LA EMPRESA: SIEMENS HEALTHINEERS

Siemens Healthineers (abreviado como SHS) es una empresa mundial líder en el campo de tecnología médica, con más de 120 años de trayectoria en el sector. La compañía se dedica a proporcionar soluciones innovadoras en las áreas de diagnóstico de laboratorio (DX) y diagnóstico por imagen (DI) lo cual le permite tener un impacto importante en el cuidado de la salud mundial (Siemens, 2024a).

En ese contexto, los CSEs de SHS son profesionales técnicos altamente certificados y calificados responsables de realizar mantenimientos preventivos y correctivos en equipos médicos. Estos ingenieros juegan un papel esencial en el mantenimiento funcional/operativo de los sistemas de diagnóstico y tratamiento, asegurando que operen dentro de las especificaciones y estándares establecidos por la empresa.

Los CSEs están entrenados para cumplir con estas directrices y para intervenir de manera efectiva cuando se identifican riesgos potenciales, garantizando así la seguridad del usuario y personal médico que opera los equipos.

Además, los CSEs están equipados para responder a las necesidades de asistencia técnica, ya sea de manera presencial o a través del soporte remoto, y están disponibles para contactar a

través de los canales de asistencia técnica de SHS. Su trabajo no solo se limita a la resolución de problemas técnicos, sino que también incluye la actualización y mejora continua de los equipos, en lineamiento con casa matriz (ubicada en Alemania) de mejorar constantemente sus productos y servicios.

Los CSEs también están al tanto de los cambios en las especificaciones, el equipamiento y los procedimientos de mantenimiento, puesto que SHS se reserva el derecho de efectuar modificaciones en cualquier momento y sin aviso previo. Así los CSEs están en constante aprendizaje y adaptación para mantenerse actualizados con las últimas tecnologías y prácticas de mantenimiento.

2.6.1 Misión

La misión de SHS es proporcionar soluciones innovadoras en el ámbito de la tecnología médica y enfocarse en desarrollar productos y servicios que aumenten la eficiencia, reduzcan los costos y, sobre todo, mejoren los resultados clínicos para los pacientes (Quintero, 2023).

2.6.2 Visión

SHS transformar la atención médica mediante el uso de tecnologías avanzadas y soluciones digitales que permitan un diagnóstico y tratamiento más eficientes y efectivos, ayudando a que la atención sanitaria sea más precisa, personalizada y accesible (Quintero, 2023).

2.6.3 Principales áreas de negocio y productos

SHS se enfoca en áreas clave dentro del sector de la salud, donde sus principales áreas de negocio y productos incluyen:

1. **Diagnóstico por Imagen:** SHS presenta variedad de equipos de diagnóstico por imagen, incluyendo resonancia magnética (MR), tomografía computarizada (CT), ultrasonido, mamografía digital, equipos de rayos X y sistemas de medicina nuclear (Siemens, 2024e).
2. **Laboratorio Clínico:** La empresa proporciona sistemas de diagnóstico de laboratorio clínico que incluyen analizadores automatizados para pruebas de química clínica, hematología, inmunología, coagulación y pruebas de diagnóstico molecular (Siemens, 2024i).
3. **Tecnología Avanzada en Salud:** SHS también se dedica al desarrollo de tecnologías avanzadas en salud, como sistemas de terapia y diagnóstico guiados por imagen para

procedimientos quirúrgicos mínimamente invasivos y terapia oncológica (Siemens, 2024b).

4. **Gestión de Datos e Informática:** Ofrecen soluciones de gestión de datos e informática para hospitales y centros de salud, incluyendo sistemas gestores de imágenes (abreviado como PACS) y de información hospitalaria (abreviado como HIS), (Siemens, 2024j).

SHS continúa innovando en tecnologías médicas para mejorar el diagnóstico, tratamiento y gestión de enfermedades, contribuyendo así al avance de la medicina a nivel global.

2.6.4 Siemens Healthineers en Latinoamérica

La Figura 2.4 presenta algunos de los KPIs de SHS obtenidos desde octubre del 2022 hasta septiembre del 2023. El recuadro 1 al 4 presenta la situación de la empresa a nivel operativa. SHS trabaja directamente en 10 países divididos en 4 grupos, cuenta con aproximadamente 700 empleados. Además, tiene más de 15000 equipos de imágenes instalados (Quintero, 2023).



Figura 2.4: KPIs LAM (Quintero, 2023).

Su operación se encuentra estructurada de la siguiente manera:

- ❑ **GoC North:** Colombia, Ecuador y Perú
- ❑ **GoC South:** Argentina, Chile y Uruguay
- ❑ **GoC MAM:** México, Costa Rica y El Salvador
- ❑ **Brasil:** Brasil

En recuadros 5 y 6 presentados en la Figura 2.4, se encuentra información clave sobre la estructura de servicio y el alcance de las operaciones. Los 16 “ESP’s” (del inglés Extended Services Providers) o proveedores de servicios extendidos. Estos ESPs juegan un papel vital en la estrategia de servicio de la empresa, extendiendo su alcance y capacidad para brindar soporte y mantenimiento a los sistemas instalados. El “Revenue” de €250 millones representa los ingresos totales generados por la empresa, ofreciendo una perspectiva de la magnitud de sus operaciones (Quintero, 2023).

Finalmente los últimos dos recuadros se centran en métricas financieras y de eficiencia. La “Productividad” de €13 millones (IVV: In-Vivo) es el valor generado por la productividad en el área de imágenes, relacionado con la eficiencia en la instalación, mantenimiento de equipos de la línea de negocios de imágenes. El porcentaje del 44.5% correspondiente al ‘GP’ representa el ‘Gross Profitó ‘Margen Bruto’, el cual muestra el porcentaje de ingresos que permanece tras restar los costos directos asociados con la prestación y/o producción del servicio. Esto representa un KPI clave de rentabilidad de SHS y su capacidad para generar ganancias (Quintero, 2023).

2.6.5 Siemens Healthineers en Ecuador

Actualmente en Ecuador, SHS trabaja únicamente en la línea de negocio de imágenes donde los servicios que ofrecen son: i) mantenimientos correctivos, ii) mantenimientos preventivos y iii) Start-Ups (puesta en marcha de funcionamiento de equipos). La Figura 2.5 presenta las principales modalidades dentro de esta línea de negocio.

Dónde:

- ❑ **AX (Angiography):** La angiografía es una técnica de imagen utilizada para visualizar el interior de los vasos sanguíneos y órganos del cuerpo, especialmente arterias, venas y el corazón, que permite detectar bloqueos o estrechamientos en los vasos sanguíneos que pueden conducir a enfermedades como la enfermedad arterial coronaria (National Health Service, 2024) (Siemens, 2024c).
- ❑ **MI (Molecular Imaging):** La imagen molecular es una técnica que permite visualizar y caracterizar procesos biológicos a nivel celular y molecular dentro de los organismos vivos. En el campo de la medicina nuclear, se emplea para el trato y diagnóstico de enfermedades como el cáncer y las enfermedades cardíacas (Stanford University School, 2024) (Siemens, 2024f).

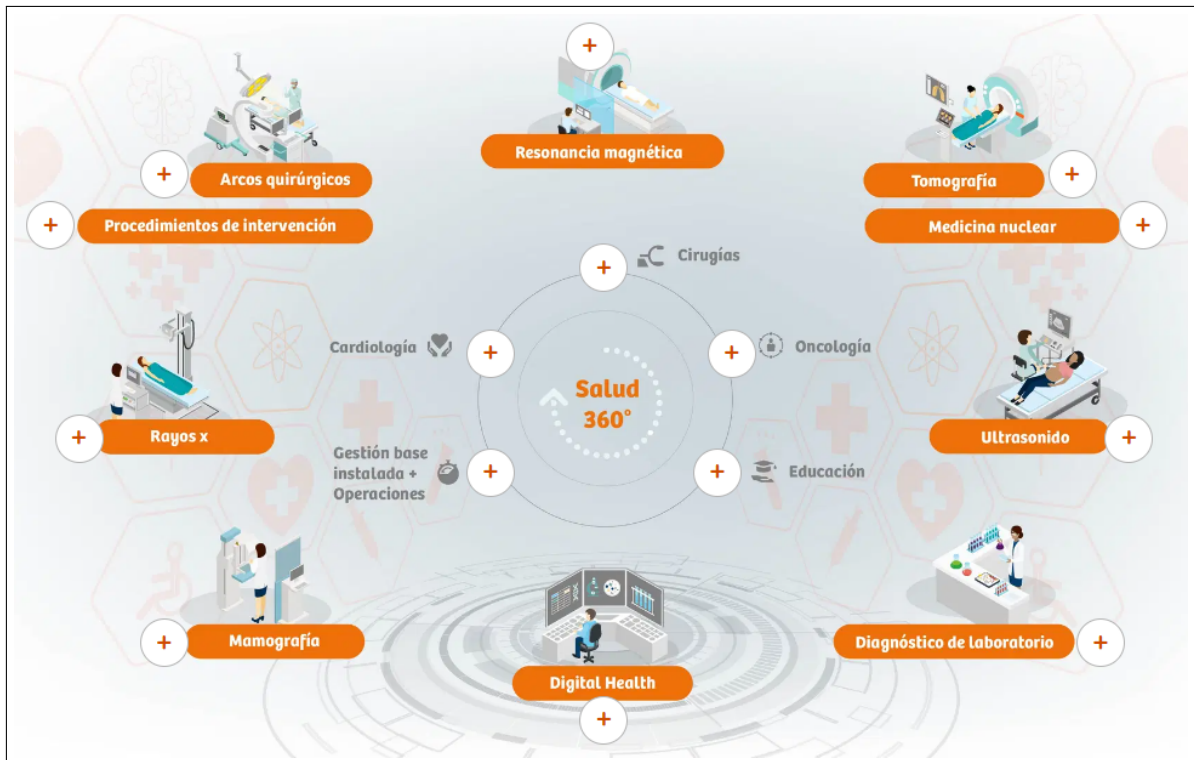


Figura 2.5: Línea de negocio de equipos de Imágenes (Quintero, 2023).

- ❑ **CT (Computed Tomography):** Una técnica de imagen que emplea rayos X y un sistema informático para generar imágenes detalladas del interior del cuerpo es la tomografía computarizada (Siemens, 2024k). Estas imágenes pueden proporcionar más información que los rayos X convencionales, donde su propósito es diagnosticar una cantidad considerable de enfermedades (National Institutes of Health, 2022).
- ❑ **MR (Magnetic Resonance):** Generar imágenes detalladas de los órganos y tejidos del cuerpo es el propósito de la resonancia magnética, una técnica de imagen que emplea campos magnéticos (Siemens, 2024g). Esta técnica resulta especialmente útil para observar estructuras blandas como los ligamentos, los músculos y principalmente el cerebro (National Institutes of Health, 2023).
- ❑ **US (Ultrasound):** Una técnica de imagen que emplea ondas sonoras de alta frecuencia para generar imágenes de órganos y estructuras internas del cuerpo es el ultrasonido, también conocido como ecografía. Como una herramienta de diagnóstico no invasiva y segura, se usa frecuentemente para examinar los órganos abdominales, el corazón y los vasos sanguíneos, como también es utilizado durante el embarazo. (Radiopaedia.org, 2023).
- ❑ **XP (X-Ray Products):** Los productos de rayos X incluyen una gama de equipos y tecnologías utilizados para generar imágenes utilizando rayos X. Estos productos se utilizan

para diagnosticar y tratar una variedad de condiciones médicas y pueden incluir sistemas de radiografía convencional, fluoroscopia y mamografía (Siemens, 2024h).

2.6.6 SAP CRM

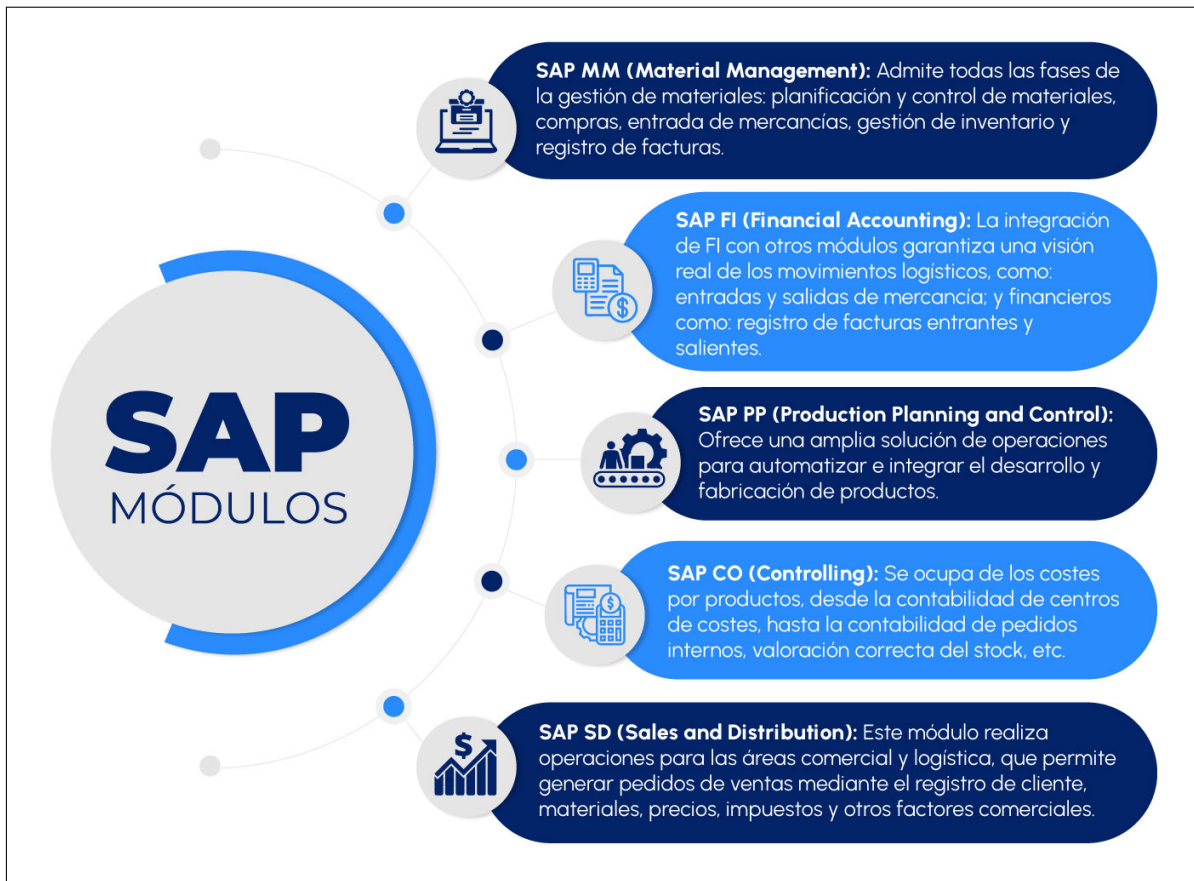


Figura 2.6: Módulos más importantes en SAP (Ejecutiva, 2021).

SAP (del inglés, Systems, Applications, and Products), traducido como “Sistemas, aplicaciones y productos para el procesamiento de datos” es una plataforma de software empresarial que proporciona soluciones integradas para la gestión de diferentes aspectos de las operaciones de una organización. La Figura 2.6 presenta algunos módulos importantes dentro del sistema SAP. El ERP (Enterprise Resource Planning) de SAP es un sistema diseñado para gestionar las relaciones y las interacciones con los clientes, ayudando a optimizar las operaciones comerciales (SAP, 2024).

SAP en SHS proporciona información completa e integral sobre los clientes y tiene como objetivo resolver problemas relevantes para la industria (SAP, 2024). Este sistema está dividido en módulos, cada uno diseñado para abordar diferentes aspectos de la administración de

hospitales, equipos médicos y clientes. Para el estudio en cuestión, se utilizó específicamente el módulo de Customer Services, el cual es fundamental para la consulta de información sobre la base instalada y los contratos de servicio en todo el país. Este módulo facilita la recopilación y análisis de datos relacionados con los servicios ofrecidos, garantizando una gestión eficiente de los contratos y un seguimiento exhaustivo de los equipos (Himavanth, 2019).

Además, el módulo de Customer Services permite realizar comprobaciones de disponibilidad de recursos en tiempo real, así como gestionar contratos y procesos de facturación. Estas capacidades son esenciales para mantener un alto grado de transparencia en el cumplimiento de los pedidos y en el seguimiento del estado de los servicios prestados. La solución SAP CRM no solo optimiza la gestión interna de los servicios, sino que también proporciona un soporte integral al cliente a través de diversos canales, incluyendo centros de interacción con el cliente y aplicaciones de autoservicio en línea (Himavanth, 2019). Gracias a estas herramientas, SHS puede ofrecer un servicio al cliente más eficiente y adaptado a las necesidades específicas de sus usuarios.

Así concluye este primer capítulo, que ha proporcionado una visión general del contexto y los conceptos clave necesarios para comprender el estudio. En el siguiente capítulo, se abordará la metodología aplicada en este trabajo, con un enfoque particular en las seis fases del proceso CRISP-DM.

3 METODOLOGÍA

En este capítulo, se detalla la metodología empleada para el desarrollo del modelo predictivo, centrándose en la aplicación del proceso CRISP-DM. Se abordan las fases desde la comprensión del negocio hasta la ejecución final.

3.1 CASO DE ESTUDIO I: CONTRATACIÓN DE NUEVO PERSONAL

La Figura 3.1 presenta el diagrama BPMN (BPMN: Business Process Modeling Notation¹) del flujo para este caso de estudio de contratación de nuevos técnicos, a través de CRISP-DM.

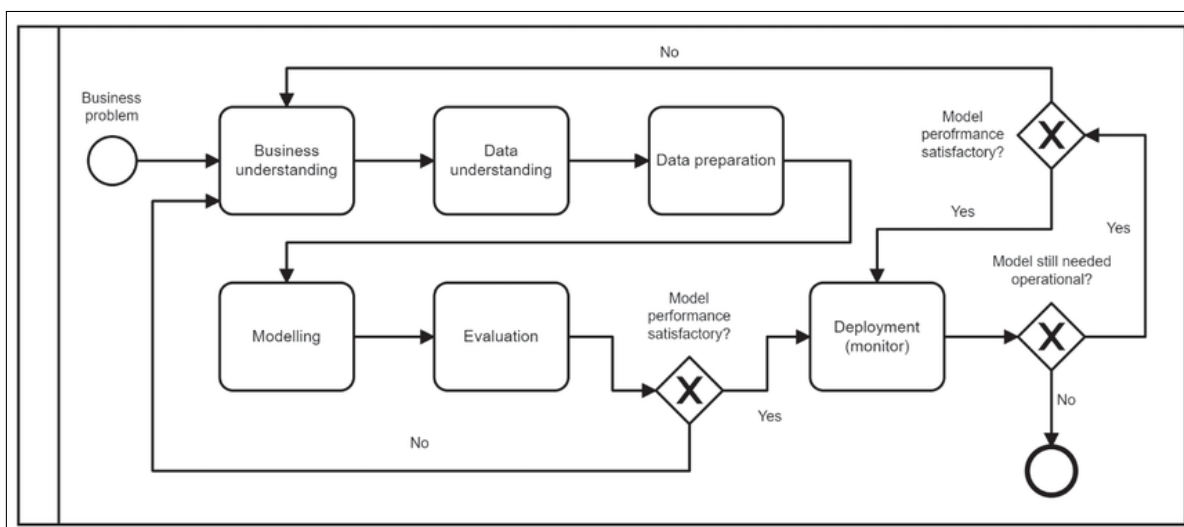


Figura 3.1: BPMN of CRISP-DM framework (Riess, 2022).

El inicio del proceso se da con la fase de “Comprensión del Negocio”. Se definen los objetivos del proyecto y se evalúa la situación actual de la asignación de CSEs. Posteriormente, en la fase de “Comprensión de los Datos”, se recopila datos relevantes del sistema CRM SAP y otras fuentes internas, se exploraron y se evaluó su calidad.

La “Preparación de los Datos” involucra la limpieza, transformación y selección de

¹BPMN: Un sistema gráfico diseñado para estandarizar la representación de procesos de negocio mediante un lenguaje uniforme, con el objetivo de simplificar la comprensión del rendimiento y la gestión de una o más actividades (Ponce et al., 2023).

variables pertinentes para el análisis. En la fase de “Modelado”, se implementan y evalúan tres algoritmos de agrupamiento: K-Means, GMM y Agrupamiento Aglomerativo.

Finalmente, en la fase de “Despliegue”, se propuso la implementación de una interfaz en Python para facilitar el uso del modelo y la visualización de los resultados.

El diagrama BPMN refleja la iteración entre las fases de “Modelado” y “Evaluación”, resaltando la importancia de refinar y ajustar los modelos hasta alcanzar resultados satisfactorios. Además, la flecha de retroalimentación desde el “Despliegue” hacia la “Evaluación” indica la necesidad de un monitoreo continuo y mejoras en el modelo una vez implementado, asegurando su adaptación a los cambios en el entorno y la demanda de servicios.

3.2 FASE 1: COMPRENSIÓN DEL NEGOCIO

3.2.1 Objetivos de negocio

Se detallan los siguientes objetivos:

- ❑ Identificar las provincias con alta demanda de horas de servicio con contrato vigente.
- ❑ Identificar la cantidad de horas anuales que un ingeniero de servicio debe trabajar con una productividad al 80%.
- ❑ Implementar un algoritmo de clustering que sugiera una ubicación geográfica adecuada para la contratación de nuevo personal de campo, de manera que se minimicen los tiempos de desplazamiento y se garantice una cobertura adecuada del territorio.

3.2.2 Evaluación de la situación actual

SHS trabaja con CSEs cualificados que se encargan de la instalación, el mantenimiento y la reparación de sus equipos en todo el mundo.

La calidad en el servicio técnico es un pilar fundamental de la propuesta de valor, asegurando que los equipos médicos esenciales funcionen de manera óptima y confiable. El mantenimiento preventivo es una práctica crítica en este contexto, ya que previene fallos inesperados y mantiene la operatividad de los sistemas (Siemens, 2024d).

La importancia de este mantenimiento es doble: garantiza la seguridad del paciente y la persistencia de la atención sanitaria, y al mismo tiempo, protege los equipos de posibles daños que pueden surgir de situaciones como interrupciones de energía o problemas eléctricos.

Sin embargo, la empresa enfrenta desafíos significativos en la asignación de los CSEs para llevar a cabo estas tareas preventivas. La asignación ineficaz de CSEs puede llevar a tiempos de respuesta lentos, especialmente en situaciones de urgencia donde el sistema está fuera de servicio o su funcionalidad se ve comprometida. Esto no solo afecta la operación de los equipos médicos, sino que también puede tener un efecto adverso en la opinión del cliente acerca de la calidad y la confiabilidad de los servicios.

El mantenimiento preventivo es una parte esencial de la estrategia de gestión de activos de SHS. Se trata de un conjunto de actividades planeadas y programadas que se realizan con el objetivo de prevenir fallas en los equipos y prolongar su vida útil. Un mantenimiento preventivo efectivo puede reducir significativamente los costos de reparación, mejorar la confiabilidad de los equipos y aumentar la satisfacción del cliente.

3.2.2.1 Fijación de los objetivos a nivel de minería de datos

Se presenta los siguientes objetivos específicos para los modelos de minería de datos:

- ❑ Desarrollar varios modelos de aprendizaje no supervisado basado en técnicas de clustering como K-Means, Agglomerative Clustering y GMM que ubique a los CSEs (por contratar) a los mantenimientos preventivos/correctivos de manera eficiente, teniendo en cuenta la ubicación geográfica de los equipos y la demanda de servicios por provincia.

- ❑ Evaluar y comparar los diferentes modelos para determinar cuál es el mejor modelo que se adapte a las necesidades del negocio.

A continuación se presenta la fase 2 de CRISP-DM aplicado al caso de estudio.

3.3 FASE 2: COMPRENSIÓN DE LOS DATOS

A continuación, esta fase profundiza en la naturaleza y características la data recopilados. Se examinaron las diferentes fuentes de datos, se exploró la estructura y contenido dataset, y se realizaron análisis preliminares para identificar áreas de mejora en la calidad de los datos. Este entendimiento exhaustivo sentó las bases para las etapas posteriores de preparación y modelado, asegurando que los datos utilizados fueran relevantes, confiables y adecuados para abordar los objetivos del trabajo de titulación.

3.3.1 Captura de datos

La principal fuente de datos fue el sistema SAP CRM de SHS, específicamente el módulo de Customer Services. Este sistema almacena información detallada sobre los equipos médicos, los clientes y los contratos de servicio. Se extrajeron los datos relevantes para el análisis, incluyendo:

- ❑ **Información de equipos:** Identificación del equipo, nombre, modalidad, horas de mantenimiento preventivo y correctivo, entre otros.
- ❑ **Información de clientes:** Nombre del cliente y su ubicación.
- ❑ **Información de contratos:** Fechas de inicio y fin de los contratos, estado de vigencia, y número de visitas anuales programadas.

Además de SAP, se utilizaron fuentes internas adicionales de la empresa para complementar la información, como tablas fijas emitidas desde casa matriz acerca de la base instalada y también se generó una tabla fija de datos geográficos que permitió determinar los valores de latitud y longitud de cada ciudad, la presencia de aeropuertos cercanos a las ubicaciones de los equipos entre otros atributos.

3.3.2 Descripción de datos

La Tabla 3.1 presenta el conjunto de datos resultante de la consolidación de las fuentes de datos (CRM SAP y fuentes internas de la empresa), la cual contiene las siguientes variables relevantes para el análisis:

Este conjunto de datos proporciona una visión completa de los equipos, clientes y contratos de servicio en la empresa, lo que permite realizar un análisis detallado para las siguientes fases.

Tabla 3.1: Descripción atributos del dataset.

Equipment	Identificador único del equipo
City	Ciudad donde se encuentra el equipo
Nombre EQ	Nombre del equipo
Material	Número de material del equipo
Serial No.	Número de serie del equipo
Funcional Loc.	Ubicación funcional del equipo
Nombre CLIENTE	Nombre del cliente al que pertenece el equipo
LATITUD	Latitud geográfica del equipo
LONGITUD	Longitud geográfica del equipo
REGIÓN	Región geográfica donde se encuentra el equipo
HORAS_AÑO	Horas totales de mantenimiento preventivo y correctivo programadas para el equipo en un año
MODALIDAD_SAP	Modalidad del equipo según la clasificación en SAP
MODALIDAD_TRADUCIDA	Modalidad del equipo traducida al español
NOM_MODALIDAD	Nombre abreviado de la modalidad del equipo
BUSINESS_AREA	Área de negocio a la que pertenece el equipo
INICIO_CONTR	Fecha de inicio del contrato de servicio del equipo
FIN_CONTR	Fecha de finalización del contrato de servicio del equipo
CONTRATO_VIGE	Indicador de si el contrato de servicio del equipo está vigente (1) o no (0)
VISITAS_AÑO	Número de visitas de mantenimiento programadas para el equipo en un año
CANTIDAD_HORAS_AÑO_POR_VISITA	Cantidad promedio de horas por visita de mantenimiento
HORAS_PREVENTIVAS	Horas anuales de mantenimiento preventivo programadas para el equipo
HORAS_CORRECTIVAS	Horas anuales de mantenimiento correctivo estimadas para el equipo
HORAS_TOTALES_ANUALES	Horas totales anuales de mantenimiento (preventivo + correctivo)
AEROPUERTO?	Indicador de si hay un aeropuerto cerca de la ubicación del equipo (1) o no (0)

3.3.3 Exploración de datos

Se llevó a cabo un análisis preliminar de datos (EDA: Exploratory Data Analysis) para obtener una mejor comprensión de las características y las conexiones dentro dataset. El EDA se dividió en dos partes principales:

3.3.3.1 Parte 1: Análisis Exploratorio General

En esta primera parte, se examinaron las características generales de los datos. El Código 3.1 presenta el análisis de las variables numéricas.

Código 3.1: EDA.

```

1 # Variables relevantes
2 variables_relevantes = ['City', 'Nombre_CLIENTE', 'REGION', 'HORAS_AÑO', '
    MODALIDAD_TRADUCIDA', 'NOM_MODALIDAD',
3     'BUSINESS_AREA', 'VISITAS_AÑO', 'CANTIDAD_HORAS_AÑO_POR_VISITA
    ',
4     'HORAS_PREVENTIVAS', 'HORAS_CORRECTIVAS', '
    HORAS_TOTALES_ANUALES']
5 df = df[variables_relevantes]
6 print(df.describe()) # Resumen estadístico
7 # Histogramas
8 axes = df.hist(bins=15, figsize=(15, 10), layout=(3, 3), edgecolor='red', color='green
    ')
9 for ax_row in axes:
10     for ax in ax_row:
11         ax.grid(False)
12 plt.show()
13 plt.grid(False)
14 plt.show()

```

❑ **Resumen estadístico:** Se calcularon estadísticas descriptivas (media, desviación estándar,

valores mínimo, máximo y percentiles) para las variables numéricas, como HORAS_AÑO, HORAS_PREVENTIVAS y HORAS_CORRECTIVAS. Esto proporcionó una perspectiva global sobre la dispersión y variabilidad del dataset. La Tabla 3.2 y Tabla 3.3 presentan el resumen estadístico de las variables numéricas.

Tabla 3.2: Análisis descriptivo I.

	HORAS_AÑO	VISITAS_AÑO	CANTIDAD_HORAS_AÑO_POR_VISITA
count	628	535	535
mean	4,19	2,44	2,32
std	3,85	0,92	2,6
min	2	1	1
25%	2	2	1
50%	3	2	2
75%	4	3	3
max	21	6	21

Tabla 3.3: Análisis descriptivo II.

	HORAS_PREVENTIVAS	HORAS_CORRECTIVAS	HORAS_TOTALES_ANUALES
count	628	628	628
mean	4,19	1,68	5,86
std	3,85	1,54	5,39
min	2	0,8	2,8
25%	2	0,8	2,8
50%	3	1,2	4,2
75%	4	1,6	5,6
max	21	8,4	29,4

- ❑ **Verificación de valores nulos:** Se examinó la presencia de valores nulos en el conjunto de datos, ya que estos pueden afectar el análisis y modelado posterior.
- ❑ **Histogramas y diagramas de caja:** La Figura 3.2 y Figura 3.3 presentan histogramas y diagramas de caja para las variables numéricas. Esto facilita la visualización de la distribución de los datos, la detección de valores atípicos y la evaluación de la dispersión y simetría de las variables.

3.3.3.2 Situación previa

Con fecha de corte al 31 de diciembre del 2023, la Figura 3.4, la Tabla 3.4 y la Tabla 3.5 presentan la cantidad de equipos por ciudad. En Ecuador se tienen 55 empleados que trabajan directamente en la empresa, de los cuáles 11 son CSEs. La cantidad de equipos instalados son de

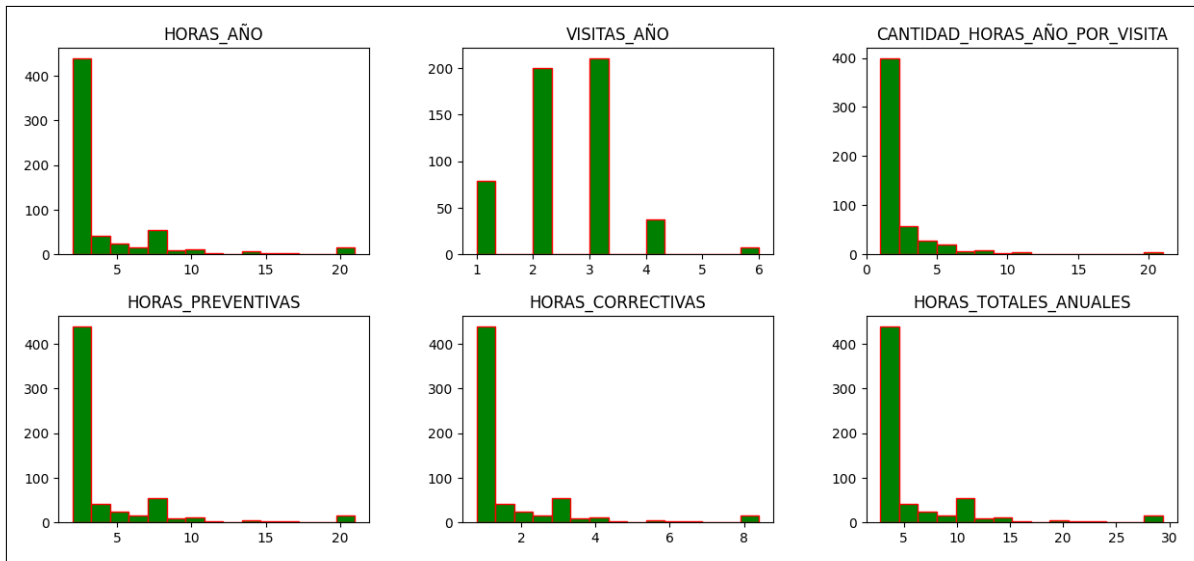


Figura 3.2: Histogramas de las variables numéricas.

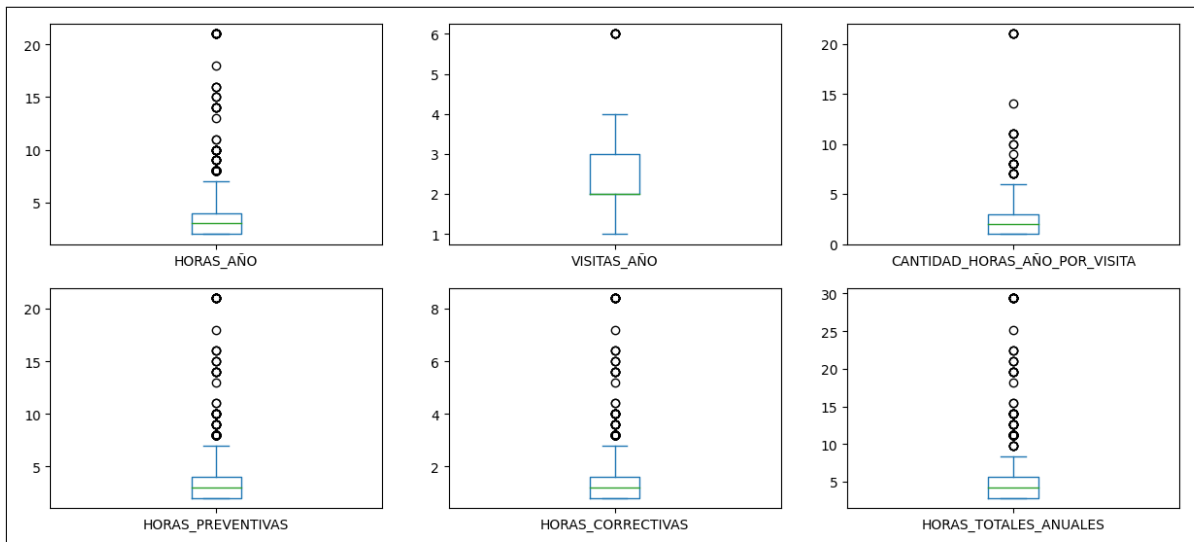


Figura 3.3: Diagrama de cajas de las variables numéricas.

628, de los cuales 387 tienen un contrato vigente. Actualmente en el país la base instalada se encuentran distribuidos de la siguiente manera:

Tabla 3.4: Cantidad de equipos por división.

Abreviación	Descripción	Cantidad
AX	Angiography	119
CT	Computed Tomography	104
MI	Molecular Imaging	11
MR	Magnetic Resonance	39
SY	Syngo	51
US	Ultrasoind	130
XP	X-Ray Products	174
Total		628

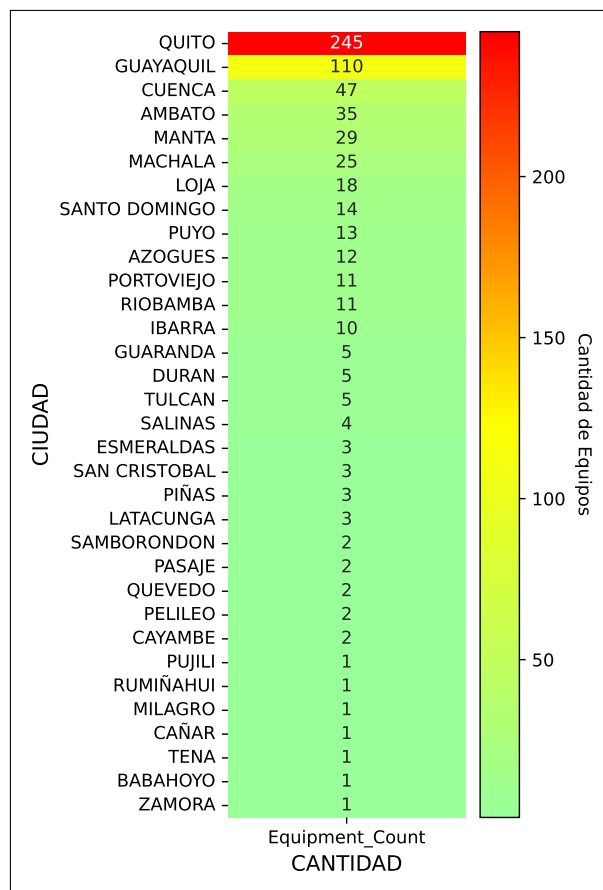


Figura 3.4: Heatmap de la cantidad de equipos por ciudad.

Tabla 3.5: Cantidad de equipos activos por división.

Abreviación	Descripción	Cantidad
AX	Angiography	85
CT	Computed Tomography	75
MI	Molecular Imaging	7
MR	Magnetic Resonance	27
SY	Syngo	38
US	Ultrasoind	64
XP	X-Ray Products	91
Total		387

La distribución desigual de equipos por división y ciudad agrava aún más este desafío. Algunas regiones y modalidades de equipos presentan una mayor concentración de demanda, lo que podría generar una sobrecarga de trabajo para los CSEs asignados a esas áreas, mientras que otras zonas podrían estar subatendidas. Esta situación puede resultar en tiempos de respuesta prolongados y afectar la complacencia del cliente, impactando negativamente en la percepción de la marca y la calidad del servicio (del inglés, QoS: Quality of Service).

3.3.3.3 Parte 2: Análisis de Horas Totales Anuales

En esta segunda parte, la Tabla 3.6 presenta un análisis más específico de las horas totales anuales que demanda cada ciudad por línea de negocio.

Tabla 3.6: Cantidad de equipos por ciudad y división.

CIUDAD / MODALIDAD	AX	CT	MI	MR	SY	US	XP	Total
AMBATO	4	8	0	2	2	3	16	35
AZOGUES	2	2	0	1	1	3	3	12
BABAHOYO	0	1	0	0	0	0	0	1
CAYAMBE	0	1	0	0	0	0	1	2
CUENCA	18	4	1	4	4	6	10	47
DURAN	0	0	0	0	1	3	1	5
ESMERALDAS	0	2	0	0	0	0	1	3
GUARANDA	1	3	0	0	0	0	1	5
GUAYAQUIL	22	11	2	9	9	26	31	110
IBARRA	3	2	0	0	0	4	1	10
LATACUNGA	0	1	0	0	0	2	0	3
LOJA	2	2	1	3	2	4	4	18
MACHALA	5	3	0	0	2	7	8	25
MANTA	5	2	0	1	3	9	9	29
MILAGRO	0	1	0	0	0	0	0	1
PASAJE	1	1	0	0	0	0	0	2
PELILEO	0	0	0	0	0	1	1	2
PIÑAS	0	0	0	0	0	1	2	3
PORTOVIEJO	3	2	0	1	0	3	2	11
PUJILI	1	0	0	0	0	0	0	1
PUYO	1	1	0	1	0	5	5	13
QUEVEDO	0	2	0	0	0	0	0	2
QUITO	47	41	7	16	25	43	66	245
RIOBAMBA	1	3	0	1	1	1	4	11
RUMIÑAHUI	0	1	0	0	0	0	0	1
SALINAS	2	2	0	0	0	0	0	4
SAMBORONDON	0	0	0	0	0	1	1	2
SAN CRISTOBAL	0	0	0	0	1	1	1	3
SANTO DOMINGO	1	4	0	0	0	4	5	14
TENA	0	1	0	0	0	0	0	1
TULCAN	0	2	0	0	0	2	1	5
ZAMORA	0	0	0	0	0	1	0	1
Total	119	104	11	39	51	130	174	628

Este análisis exploratorio permitió obtener una comprensión profunda de los datos, identificar las variables más relevantes y detectar posibles problemas de calidad. Además, se obtuvieron insights valiosos acerca de la distribución territorial de la demanda de servicios y la carga laboral según modalidad de equipo, lo cual será fundamental para el desarrollo de los modelos de optimización en fases posteriores.

3.4 FASE 3: PREPARACIÓN DE LOS DATOS

En esta fase, se llevaron a cabo una serie de transformaciones y adecuaciones sobre el conjunto de datos recopilado en la fase anterior. El objetivo principal fue garantizar la calidad, consistencia y relevancia de los datos para su posterior utilización en el modelado y análisis. Se realizaron tareas de limpieza, selección de variables y transformación de datos con el fin de preparar un conjunto de datos óptimo para el entrenamiento de los modelos de optimización de contratación de CSEs. Véase el *Anexo 6*) para visualizar el dataset procesado.

3.4.1 Filtrado y limpieza de datos

El Código 3.2 presenta la eliminación de varias columnas del dataset, irrelevantes para el diseño del modelo predictivo ya que la información proviene desde el sistema CRM y contiene códigos que son propios del sistema SAP que no son relevantes para el estudio.

Código 3.2: Limpieza previa de datos.

```
1 df = pd.read_excel('DataSET_V1.xlsx', sheet_name='Data') # Cargar los datos
2 columnas_no_relevantes = ['Equipment', 'Material', 'Serial No.', 'Functional Loc.', '
   MODALIDAD_SAP', 'BUSINESS_AREA', 'CANTIDAD_HORAS_ANO_POR_VISITA']
3 df_procesado = df.drop(columns=columnas_no_relevantes) # Eliminar las columnas no
   relevantes
4 df_procesado.to_excel('DataSET_PROCESADO.xlsx', index=False) # Guardar el nuevo
   dataset en un archivo Excel
```

3.4.2 Anonimización de datos

El Código 3.3 presenta el proceso de anonimización de los atributos que contienen información confidencial. La técnica utilizada es la seudonimización. Esta técnica implica reemplazar valores identificables, como nombres de clientes o equipos, con versiones cifradas o tokens. En este caso, los nombres de los clientes fueron sustituidos por identificadores anónimos (CLIENTE_1, CLIENTE_2, etc.), y los nombres de los equipos por códigos generados a partir de la modalidad correspondiente (PDPC, 2022).

La seudonimización permite proteger la identidad de los individuos al hacer que los datos se vuelvan no identificables de forma directa. Sin embargo, la información original puede potencialmente ser reconstruida si se tiene acceso a la tabla de correspondencias (en este caso, el diccionario que asocia los nombres originales con los identificadores anónimos) (PDPC, 2022).

Código 3.3: Aplicación técnica de anonimización de datos por seudonimización.

```
1 #ANONIMIZACION DE CLIENTES
2 #Diccionario de clientes
3 clientes_unicos = df['Nombre_CLIENTE'].unique()
4 cliente_anonimo_map = {cliente: f'CLIENTE_{i+1}' for i, cliente in enumerate(
5     clientes_unicos)}
6 df['nombre_CLIENTE_anonimo'] = df['Nombre_CLIENTE'].map(cliente_anonimo_map) # Aplicar
7     el mapeo al DataFrame
8
9 #ANONIMIZACION DE EQUIPOS POR MODALIDAD
10 contadores_modalidad = {} # Crear un diccionario para almacenar contadores por
11     modalidad
12 nombres_unicos = df['Nombre_EQ'].unique() # Obtener los nombres de equipos únicos
13 df_unicos = df.drop_duplicates(subset=['Nombre_EQ']).copy() # Crear un DataFrame para
14     los nombres únicos
15
16 # Función para generar nombres de equipos anónimos
17 def generar_nombre_equipo(fila):
18     modalidad = fila['MODALIDAD_TRADUCIDA']
19     if modalidad not in contadores_modalidad:
20         contadores_modalidad[modalidad] = 1
21     else:
22         contadores_modalidad[modalidad] += 1
23     return f"{modalidad.split()[0]}_EQ_{contadores_modalidad[modalidad]}"
24
25 # Aplicar la función para generar nombres de equipos anónimos
26 df_unicos['Nombre_EQ_anonimo'] = df_unicos.apply(generar_nombre_equipo, axis=1)
27 # Crear un diccionario
28 mapeo_nombres = dict(zip(df_unicos['Nombre_EQ'], df_unicos['Nombre_EQ_anonimo']))
29
30 # Aplicar el mapeo al DataFrame original
31 df['Nombre_EQ_anonimo'] = df['Nombre_EQ'].map(mapeo_nombres)
```

Se reemplazan los nombres de los clientes por identificadores genéricos como “CLIENTE_1”, “CLIENTE_2”, etc. Esto asegura que la identidad de los clientes no pueda ser revelada a partir de los datos.

Luego, se realiza una anonimización más sofisticada de los nombres de los equipos. En lugar de asignar identificadores aleatorios, el proceso tiene en cuenta la modalidad a la que pertenece cada equipo. Por ejemplo, un equipo de MR es renombrado como “MR_EQ_1”, mientras que un equipo de CT es renombrado a “CT_EQ_2”. Este enfoque permite preservar información relevante sobre los equipos (su modalidad) mientras se oculta su identidad específica.

La seudonimización es crucial en muchos escenarios porque proporciona un equilibrio entre privacidad y utilidad. A diferencia de la aleatorización, que introduce ruido o cambia los datos de manera que pueden volverse menos útiles para ciertos análisis, la seudonimización permite conservar la estructura y las relaciones dentro dataset, importante cuando la data necesita ser procesada o analizada sin comprometer la identidad de los individuos (PDPC, 2022).

3.5 FASE 4: MODELADO

Esta fase presenta la implementación y evaluación tres algoritmos de aprendizaje no supervisado de clustering: K-Means, Agglomerative Clustering y GMM. Estos modelos se

seleccionaron debido a su capacidad para descubrir patrones y estructuras subyacentes en los datos sin requerir etiquetas previas. El objetivo es agrupar los datos en clusters significativos que faciliten la interpretación y análisis.

3.5.1 Análisis previo de horas para determinar la cantidad de ingenieros de campo

Antes de abordar el modelo predictivo para la contratación de CSEs (clusters a ubicar), es fundamental realizar un análisis previo de las horas de trabajo requeridas para cubrir la demanda de mantenimiento preventivo en la base instalada de equipos. Este análisis permitirá determinar la cantidad de CSEs necesarios, representada por el valor “n” en el modelo predictivo.

❑ Demanda de la Base Instalada

El Código 3.4 presenta una agrupación por modalidad dónde si el equipo tiene un contrato vigente se toma en cuenta para calcular el total de horas anuales que demanda la base instalada.

Código 3.4: Cálculo de horas anuales.

```

1 # Contratos vigentes se quedan
2 filtered_df = df[df['CONTRATO_VIGE'] == 1]
3 pivot_table = filtered_df.groupby('NOM_MODALIDAD')['HORAS_TOTALES_ANUALES'].sum().
   reset_index() #Tabla dinámica
4 total = pivot_table['HORAS_TOTALES_ANUALES'].sum() # Calcular el total
5 print(pivot_table)
6 print(f"\nTotal de HORAS_TOTALES_ANUALES: {total}")

```

La Tabla 3.7 se presenta la agrupación por modalidad y la demanda de horas totales anuales de todos los equipos con contrato vigente con fecha de corte al 31 de diciembre del 2023.

Tabla 3.7: Cantidad de horas por modalidad de equipos con contrato activo.

	Modalidad	Horas totales anuales
0	Angiography	666.4
1	Computed Tomography	406.0
2	Magnetic Resonance	323.4
3	Molecular Imaging	121.8
4	Syngo	106.4
5	Ultrasound	225.4
6	X-Ray Products	424.2

Y el valor que demanda la base instalada anual es de: 2273.60 horas.

- ❑ **Horas de Trabajo de un CSE:** Un factor crucial a considerar es la cantidad de horas que un CSE debe cumplir anualmente. Legalmente, se establece un mínimo de 1600 horas de trabajo al año. Para llegar a esta cifra, se parte de un cálculo base de 2080 horas (ocho horas diarias, cinco días semanales, cincuenta y dos semanas anuales), del cual se restan 480 horas correspondientes a vacaciones, enfermedad, cursos y feriados (15 días cada uno, a 8 horas por día).

- ❑ **Productividad:** Además de las horas obligatorias, es importante establecer un objetivo de productividad para los CSEs. En este caso, por disposición de casa matriz se define una productividad del 80%, lo que implica que, de las 1600 horas disponibles para trabajo efectivo, se espera que cada CSE realice al menos 1280 horas de mantenimiento preventivo, correctivo y start-ups.

- ❑ **Cálculo de la Cantidad de CSEs (n):** Conociendo la demanda total de horas de mantenimiento preventivo y las horas que cada CSE puede realizar de manera efectiva, podemos calcular la cantidad de CSEs necesarios (n) dividiendo la demanda total entre las horas por CSE:

Finalmente:

$$n = \frac{\text{Demanda total de horas}}{\text{Horas por CSE}} \quad (3.1)$$

$$n = \frac{2260.3}{1280}$$

$$n = 1.765$$

$$n \rightarrow 2$$

La Ecuación 3.1 presenta el valor de n que es el argumento de entrada para el modelo predictivo, ya que determinará el número de clústeres a formar, cada uno representando la ubicación óptima para un nuevo CSE.

3.5.2 Código base

El Código 3.5 presenta la base que se utilizó en los diferentes modelos predictivos. Estos códigos sirven como punto de partida para todos los modelos predictivos y garantizan la reproducibilidad y comparabilidad de los resultados obtenidos.

Código 3.5: Código base modelos predictivos.

```
1
2 **MODELO_PREDICTIVO**
3
4 #Plot 1
5 df_filtered.groupby('PROVINCIA')['HORAS_TOTALES_ANUALES'].sum().reset_index()
6 provincias = gpd.read_file('ecuador.shp') # Shapefile
7 provincias = provincias.merge(horas_provincia, how='left', left_on='DPA_DESPRO',
8     right_on='PROVINCIA') # JOIN 2 ARCHIVOS
9 provincias['HORAS_TOTALES_ANUALES'] = provincias['HORAS_TOTALES_ANUALES'].fillna(0)
10 m = folium.Map(location=[-1.8312, -78.1834], zoom_start=7)
11
12 #Colormap
13 colormap = cm.linear.RdYlGn_11.scale(provincias['HORAS_TOTALES_ANUALES'].min(),
14     provincias['HORAS_TOTALES_ANUALES'].max())
15
16 # Estilo GeoJson
17 def style_function(feature):
18     horas = feature['properties']['HORAS_TOTALES_ANUALES']
19     return {
20         'fillOpacity': 0.7 if horas > 0 else 0,
21         'weight': 0.5,
22         'fillColor': colormap(horas) if horas > 0 else 'transparent',
23     }
24
25 geojson = folium.GeoJson(
26     provincias,
27     style_function=style_function,
28     tooltip=GeoJsonTooltip(
29         fields=['DPA_DESPRO', 'HORAS_TOTALES_ANUALES'],
30         aliases=['Provincia:', 'Horas Totales Anuales:'],
31         localize=True
32     )
33 ).add_to(m)
34 colormap.add_to(m)
35
36 for centroid in centroids:
37     folium.Marker(
38         location=[centroid[0], centroid[1]],
39         icon=folium.Icon(color='blue', icon='info-sign')
40     ).add_to(m)
41 display(m)
42 #Plt 2 -> Plano Cart
43 plt.figure(figsize=(10, 8))
44 plt.scatter(df_filtered['LONGITUD'], df_filtered['LATITUD'], marker='o', c='blue', s
45     =50, label='Ciudades')
46 # Añadir etiquetas a las ciudades (usando el nombre de la ciudad)
47 for idx, row in df_filtered.iterrows():
48     plt.annotate(row['City'], (row['LONGITUD'], row['LATITUD']), fontsize=10, ha='
49         center', va='bottom')
50
51 #Centroides
52 plt.scatter(centroids[:, 1], centroids[:, 0], marker='o', c='red', s=100, label='
53     Centroides')
54
55 #Labels
56 for i, (lat, lon) in enumerate(centroids):
57     plt.annotate(f'CSE {i+1}', (centroids[i, 1], centroids[i, 0] + 0.2), fontsize=10,
58         ha='center')
```

Esta codificación 3.5 realiza los siguientes pasos:

1. **Modelo Predictivo:** La variable *Modelo_Predictivo* reemplazará al modelo predictivo a utilizarse. En particular para nuestro caso de estudio serán: K-Means, Agglomerative Clustering y GMM.
2. **Análisis gráficas:** Las gráficas resultantes del modelo predictivo se comprenden de la siguiente forma:
 - ❑ **Centroides:** Las coordenadas de los centroides de los clústeres representan las ubicaciones óptimas sugeridas para los CSEs.
 - ❑ **Asignación de ciudades:** Cada ciudad se asigna al clúster cuyo centroide está más cerca, lo que indica a qué CSE debería ser asignada para el mantenimiento preventivo/correctivo.
3. **Visualización de Resultados:** El código genera dos visualizaciones para interpretar los resultados del modelo. La primera es un mapa interactivo que muestra la ubicación de las ciudades consideradas y los centroides de los clústeres, lo que permite visualizar la distribución geográfica de los equipos y las ubicaciones propuestas para los CSEs. La segunda visualización es un gráfico de dispersión que muestra las ciudades y los centroides en un plano cartesiano, con etiquetas que indican las coordenadas de cada punto. Estas visualizaciones facilitan la comprensión de los resultados del modelo y la toma de decisiones informadas sobre la ubicación de los CSEs.

3.5.3 K-Means

El Código 3.6 presenta la implementación del modelo predictivo basado en el algoritmo K-Means. Esta elección se fundamenta en la necesidad de identificar agrupamientos espaciales y en la capacidad de K-Means para manejar de manera eficiente grandes volúmenes de datos geográficos, lo cual es fundamental para este análisis.

Código 3.6: Modelo predictivo con K-Means.

```
1 df['weight_aeropuerto'] = df['AEROPUERTO?'].apply(lambda x: 2 if x == 1 else 1)
2 df['weight Equipos'] = df.groupby('City')['Nombre_EQ'].transform('size').apply(lambda
   x: 2 if x >= 5 else 1)
3 df['weight_contratos'] = df['CONTRATO_VIGE'].apply(lambda x: 2 if x == 1 else 0)
4 df['weight'] = df['weight_aeropuerto'] * df['weight Equipos'] * df['weight_contratos']
5 n = int(input("Cantidad de CSEs a ubicar (n): ")) # Input
6 kmeans = K-Means(n_clusters=n, random_state=0)
7 df['cluster'] = kmeans.fit_predict(df[['LATITUD', 'LONGITUD']], sample_weight=df['
   weight'])
8 centroids = kmeans.cluster_centers_ #Coordenadas centroides
```

Esta codificación 3.6 implementa un modelo predictivo con K-Means. El algoritmo es aplicado a las coordenadas geográficas de las ciudades que cumplen con ciertos parámetros de entrada que son variables analizadas desde una perspectiva operativa como: tener un contrato vigente, tener al menos 5 equipos y considerar la presencia de aeropuertos cercanos. Estos criterios buscan garantizar que las ubicaciones seleccionadas tengan una demanda de mantenimiento suficiente para justificar la contratación de un CSE. El algoritmo agrupa las ciudades en n clústeres, y las coordenadas de los centroides de estos clústeres representan las ubicaciones sugeridas para los CSEs.

Para las condiciones iniciales planteadas se presentaron como la multiplicación de los pesos. Al multiplicar los pesos, se da mayor importancia a las ciudades que cumplen con **TODOS** los criterios, similar al comportamiento de una compuerta lógica AND. Si una ciudad tiene aeropuerto, contrato vigente y más de 5 equipos su peso total será mayor a la de un registro que no cumpla con ninguno de esos criterios, lo que la hace mucho más influyente en el agrupamiento que una ciudad que solo cumple con uno o ninguno de los criterios. La multiplicación de los pesos permite enfatizar en que se deben cumplir todos los criterios establecidos simultáneamente (Sankalana, 2023).

3.5.4 Agglomerative Clustering

El Código 3.7 presenta la implementación del algoritmo Agrupamiento Aglomerativo, una técnica de agrupamiento jerárquico. Este método comienza asignando cada dato a su propio grupo y, posteriormente, va combinando los grupos más similares entre sí de manera iterativa. El proceso de fusión se basa en un criterio de enlace predefinido. Esta aproximación es particularmente valiosa para descubrir la organización jerárquica inherente en los datos y para detectar patrones en múltiples niveles de detalle.

Código 3.7: Código base Agrupamiento Aglomerativo.

```
1 m = int(input("Cantidad de CSEs a ubicar (n): ")) # Input
2 ac = AgglomerativeClustering(n_clusters=m, affinity='euclidean', linkage='ward')
3 df['centroid'] = ac.fit_predict(df[['LATITUD', 'LONGITUD']])
4 centroids = df.groupby('cluster')[['LATITUD', 'LONGITUD']].mean().values
```

Similar a K-Means, esta codificación 3.7 implementa el modelo de Agrupamiento Aglomerativo. Se aplica el algoritmo de Agrupamiento Aglomerativo a la columna de latitud y columna de longitud de las ciudades, utilizando el método de enlace 'ward' y la distancia euclidiana como medida de similitud. El resultado es la asignación de cada ciudad a uno de los n clústeres.

La elección de estos hiperparámetros se justifica por la naturaleza geográfica de los datos. La distancia euclidiana es ideal para representar distancias en un mapa, mientras que el método de enlace *ward* favorece la creación de clusters compactos al minimizar la varianza interna. Esto es particularmente útil para agrupar ciudades de manera eficiente para la ubicación de CSEs.

Esta configuración permite una agrupación flexible y adaptable de las ciudades, creando clusters geográficamente cohesivos. Esto facilita la planificación estratégica de la distribución de CSEs, permitiendo ajustar el número de clusters según las necesidades específicas del modelo y garantizando una cobertura eficiente del territorio.

Finalmente, dado que el Agrupamiento Aglomerativo no calcula centroides de manera directa como K-Means, se obtiene la ubicación representativa de cada clúster calculando el promedio de las latitudes y longitudes de las ciudades que pertenecen a cada grupo. Estos promedios se almacenan en la variable `centroids` y se utilizarán posteriormente para visualizar la ubicación sugerida de los CSEs en el mapa.

3.5.5 Gaussian Mixture Model

El Código 3.8 implementa el algoritmo de GMM, el cual fue utilizado para agrupar los datos de forma automática. Este modelo, al igual que un sistema de clasificación, identificó de manera eficiente grupos de datos con propiedades comunes, contribuyendo a revelar patrones ocultos en el conjunto de datos.

Código 3.8: Código base GMM.

```
1 p = int(input("Cantidad de CSEs a ubicar (n): ")) # Input
2 gmm = GaussianMixture(n_components=p, random_state=0) #GMM
3 df['centroid'] = gmm.fit_predict(df[['LATITUD', 'LONGITUD']])
4 centroids = gmm.means_
```

Similar, esta codificación 3.8 implementa el modelo de GMM. Se aplica el algoritmo GMM a las ubicaciones geográficas (lat, lon) de las ciudades. El modelo asume que los datos se generan a partir de una mezcla de distribuciones gaussianas, lo que le permite adaptarse a clústeres de diferentes formas y tamaños. Una vez entrenado el modelo, se utiliza el método `fit_predict` para asignar cada ciudad a uno de los clústeres identificados. Finalmente, se obtienen las coordenadas de los centros de los clústeres utilizando el atributo `means_` del modelo GMM. Estos centros representan las ubicaciones sugeridas para los CSEs, teniendo en cuenta la distribución probabilística de los datos en cada clúster.

La siguiente sección presenta la fase de evaluación de los 3 diferentes modelos predictivos

implementados.

3.6 FASE 5: EVALUACIÓN

Para analizar la efectividad de los modelos de clustering implementados, se utilizaron tres índices principales: Silhouette Coefficient, Calinski-Harabasz Index y Davies-Bouldin Index. Estos índices proporcionan una evaluación numérica de la cohesión y separación de los clusters, permitiendo una comparación objetiva entre los diferentes modelos. Adicional se presenta un análisis experimental para determinar bajo las condiciones iniciales presentadas en la fase previa.

3.6.1 Análisis teórico

3.6.1.1 Coeficiente de Silueta

Métrica que mide el grado de similitud entre los objetos dentro de un cluster en comparación con otros clusters. (Sai et al., 2017). Se calcula para cada punto y evalúa la distancia promedio entre dicho punto y todos los demás puntos dentro de su mismo cluster, en comparación con la distancia promedio entre ese punto y los puntos del cluster más cercano al que no pertenece. Un coeficiente elevado sugiere que los puntos están agrupados de manera efectiva, mientras que un valor bajo o negativo sugiere que los puntos podrían estar mal agrupados (Sai et al., 2017).

3.6.1.2 Índice de Calinski-Harabasz

Denominado “razón de varianza”, se calcula como la relación entre la suma de la dispersión entre clusters y la dispersión dentro de los clusters. Este índice favorece configuraciones con un alto número de clusters si la dispersión entre los clusters es grande. Un valor alto del índice indica una mejor definición de los clusters, con una buena separación entre los grupos y una alta cohesión dentro de ellos (Ashari et al., 2023).

3.6.1.3 Índice de Davies-Bouldin

Evalúa la relación entre la dispersión interna de los clusters y la distancia entre los clusters. Se determina mediante la media de las razones máximas de dispersión dentro del cluster a la separación entre los clusters, para cada cluster. Un valor más bajo del índice indica una mejor formación de clusters, con poca dispersión interna y buena separación entre los clusters

(Ashari et al., 2023).

El Código 3.9 presenta la codificación base para la evaluación de los diferentes modelos predictivos, dónde se fue variando el modelo predictivo a implementar.

Código 3.9: Evaluación algoritmos implementados.

```
1 # Evaluación K-Means
2 silhouetteAvg = silhouette_score(df[['LATITUD', 'LONGITUD']], df['cluster'])
3 calinski harabasz = calinski harabasz_score(df[['LATITUD', 'LONGITUD']], df['cluster'
4 ])
5 daviesBouldin = daviesBouldin_score(df[['LATITUD', 'LONGITUD']], df['cluster'])
```

El valor de cada 'cluster' se obtiene según el modelo predictivo utilizado, detallado en el Código 3.6 para K-Means, el Código 3.7 para Agrupamiento Aglomerativo y el Código 3.8 para GMM

3.6.1.4 Resumen evaluación de índices

La Tabla 3.8, la Figura 3.5 y la Figura 3.6 presentan los resultados de las métricas de evaluación para tres modelos de agrupamiento diferentes: K-Means, Agrupamiento Aglomerativo y GMM.

Tabla 3.8: Resultados evaluación modelos predictivos.

Modelo	Índice		
	Silueta	Calinski-Harabas	Davies-Bouldin
K-Means	0.6565	845.112	0.527
Agglomerative Clustering	0.6599	845.876	0.552
Gaussian Mixture Model	0.5088	400.084	0.602

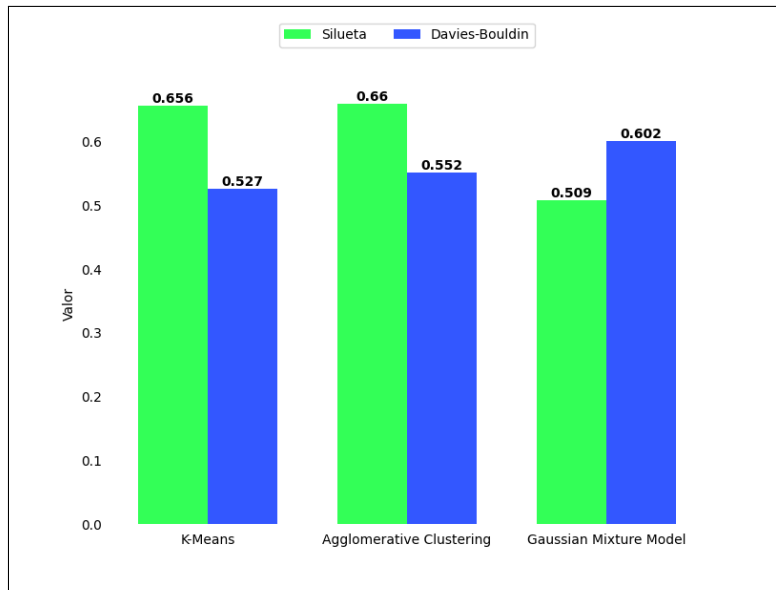


Figura 3.5: Representación gráfica coeficientes de Silueta y Davies-Bouldin.

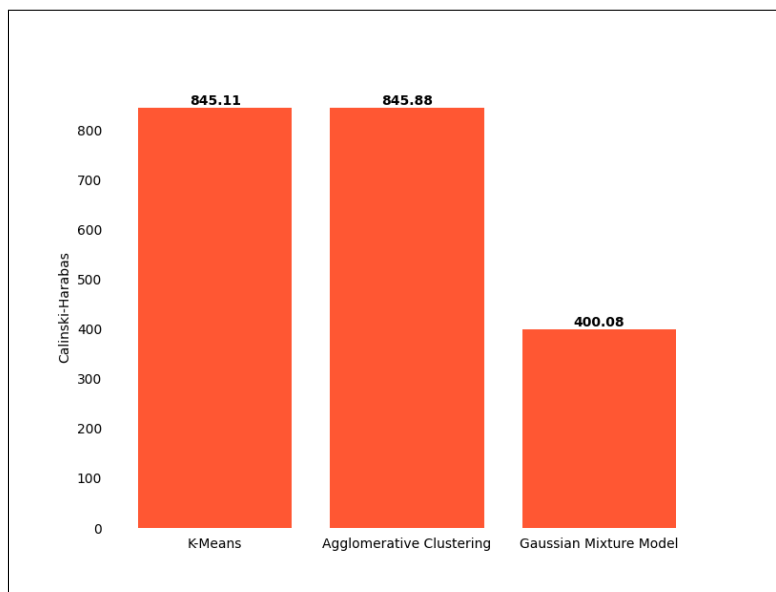


Figura 3.6: Representación gráfica del coeficiente Calinski-Harabasz.

1. Silhouette coefficient:

- ❑ *Interpretación:* Un valor más alto indica que los puntos están bien agrupados dentro de sus clústeres y separados de otros clústeres.
- ❑ *Resultados:* El Agrupamiento Aglomerativo tiene el coeficiente de silueta más alto (0.6599), seguido de K-Means (0.6565) y GMM (0.5088). El algoritmo por Agrupamiento Aglomerativo produce clústeres ligeramente más definidos que los otros dos modelos.

2. Calinski-Harabasz index:

- ❑ *Interpretación:* Un valor superior sugiere que los clústeres son más compactos y están mejor separados.
- ❑ *Resultados:* K-Means y Agrupamiento Aglomerativo tienen índices muy similares (845.112 y 845.876, respectivamente), lo que sugiere una calidad de agrupamiento similar en términos de densidad y separación. GMM tiene un índice mucho más bajo (400.084), lo que indica que sus clústeres son menos densos o están menos separados.

3. Davies-Bouldin index:

- ❑ *Interpretación:* Un valor inferior sugiere una mejor separación entre clústeres.
- ❑ *Resultados:* K-Means tiene el índice de Davies-Bouldin más bajo (0.527), seguido de Agrupamiento Aglomerativo (0.552) y GMM (0.602). Esto sugiere que K-Means produce clústeres mejor separados que los otros dos modelos.

En general, K-Means y Agrupamiento Aglomerativo demuestran un desempeño comparable en cuanto a la calidad de la agrupación, con una leve superioridad para el Agrupamiento Aglomerativo en términos de definición de clústeres (coeficiente de silueta). K-Means tiene una ligera ventaja en términos de separación de clústeres (índice Davies-Bouldin). GMM, por otro lado, muestra un rendimiento inferior en las tres métricas, lo que indica que sus clústeres son menos densos y están menos separados que los de los otros dos modelos.

3.6.2 Análisis experimental

Se llevó a cabo un análisis desde una perspectiva del negocio. Esto se realizó a través de la ubicación de los diferentes clusters asignados por cada modelo predictivo. Se tomó como referencia el Código 3.5 que presenta la codificación para evaluar gráficamente los distintos modelos predictivos implementados. Se evaluó la contratación de CSEs a través de una visualización geográfica que integra la demanda de servicios y las ubicaciones propuestas. Se filtraron los datos de equipos con contratos vigentes, ya que estos representan la demanda real de mantenimiento. Luego, se agrupó la información por provincia, sumando las horas totales anuales de servicio requeridas en cada una.

Para evaluar la ubicación propuesta de los CSEs, se superpusieron en el mapa los centroides de los clústeres obtenidos con los diferentes modelos. Estos centroides, representados por marcadores azules, indican las ubicaciones óptimas para los CSEs según el modelo.

Finalmente, se agregaron etiquetas en el centro de cada provincia con demanda, mostrando las horas totales anuales de servicio requeridas. Así, se facilita la comparación visual entre la demanda de servicio y la ubicación de los CSEs propuestos, permitiendo evaluar si el modelo asigna los recursos de manera eficiente y adecuada según los requisitos de cada zona geográfica. Este análisis visual proporciona una primera aproximación a la evaluación del modelo, permitiendo identificar posibles áreas de mejora en la contratación de CSEs.

3.6.2.1 Análisis con 2 centroides

La Figura 3.7 y la Figura 3.8 presentan el resultado de las gráficas con $n = 2$ con K-Means.

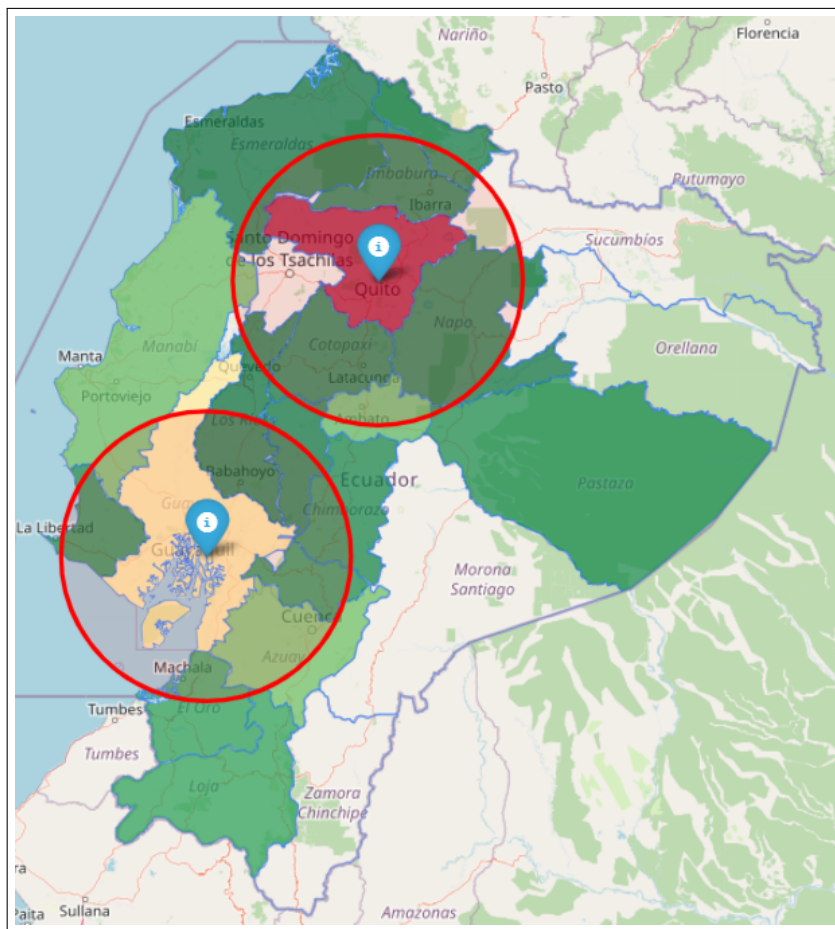


Figura 3.7: Ubicación geográfica de 2 CSEs con K-Means.

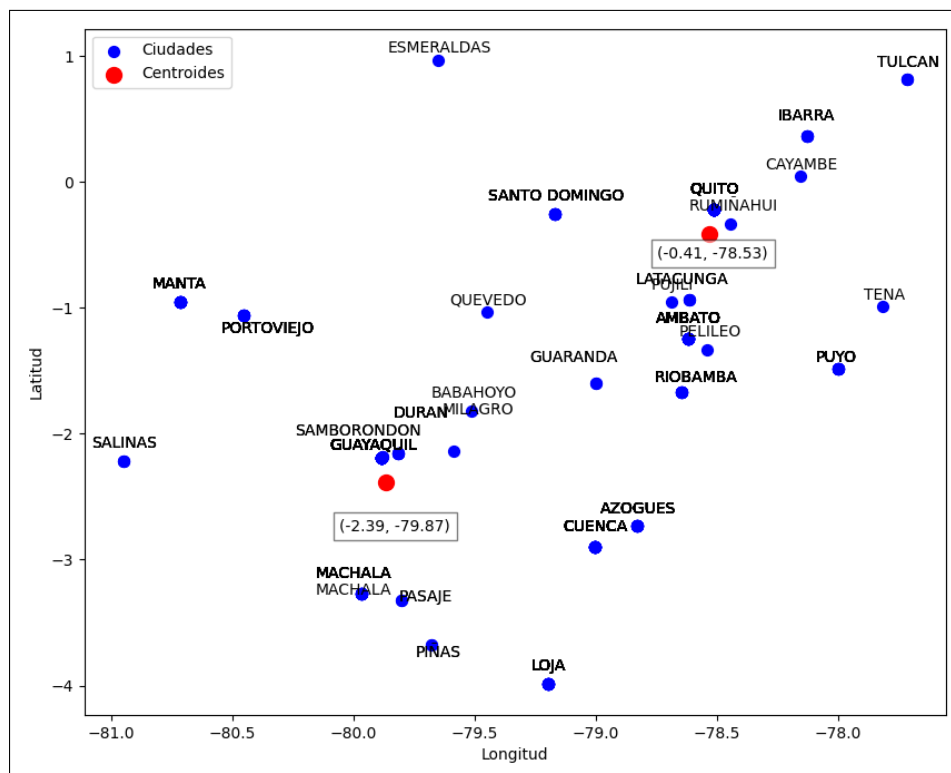


Figura 3.10: Plano cartesiano: ubicación geográfica de 2 CSEs con Agrupamiento Aglomerativo.

Al igual que en el caso de K-Means, los resultados parecen coincidir con la alta demanda de servicios en Quito y Guayaquil. Un centroide se encuentra cerca de Quito, mientras que el otro está ubicado en la región costera, próximo a Guayaquil. Esto sugiere que el Agrupamiento Aglomerativo también ha logrado identificar estas dos ciudades como áreas clave para la contratación de CSEs, lo que es fundamental para asegurar una respuesta rápida y eficaz a los requerimientos de mantenimiento en estas zonas de alta demanda.

La división del país en dos clústeres principales refleja una clara separación geográfica entre la región norte/centro, representada por el clúster de Quito, y la región sur/costa, representada por el clúster cercano a Guayaquil. Esta división geográfica se alinea con la disposición de la población y los recursos de infraestructura en Ecuador, lo que respalda la validez del modelo de agrupamiento.

La Figura 3.11 y la Figura 3.12 presentan el resultado en el plano cartesiano con $n = 2$ con GMM.

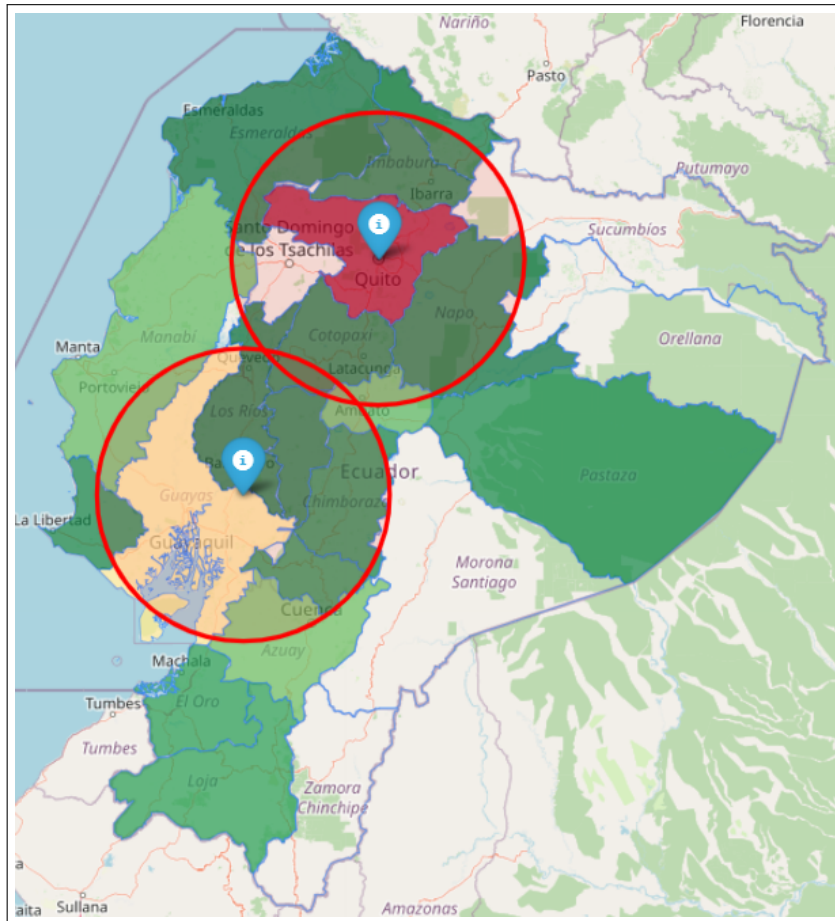


Figura 3.11: Ubicación geográfica de 2 CSEs con GMM.

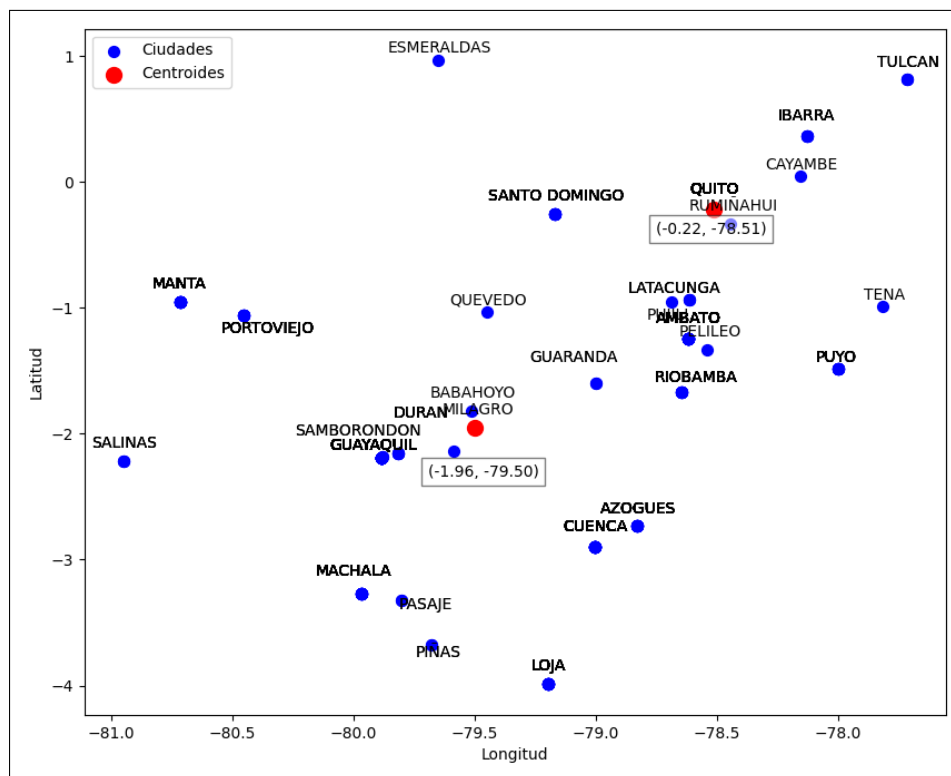


Figura 3.12: Plano cartesiano: ubicación geográfica de 2 CSEs con GMM.

Al igual que en los casos anteriores (K-Means y Agrupamiento Jerárquico), los resultados de GMM también parecen coincidir con la alta demanda de servicios en Quito y Guayaquil. Un centroide se encuentra cerca de Quito, mientras que el otro está posicionado más al sur, en la región costera, abarcando potencialmente tanto Guayaquil como otras ciudades importantes de la costa. Esto sugiere que GMM también ha logrado identificar estas dos regiones como áreas clave para la contratación de CSEs.

Dado que los resultados experimentales con dos centroides ($k=2$) para los tres modelos de agrupamiento (K-Means, Agrupamiento Jerárquico y GMM) mostraron similitudes en cuanto a la distribución geográfica de los clústeres y la ubicación de los CSEs, se procederá a realizar una segunda evaluación con un mayor número de clústeres ($k=3$). Esto permitirá explorar si un modelo en particular ofrece una mejor solución en términos de eficiencia y cobertura de servicio.

3.6.2.2 Análisis con 3 centroides

La Figura 3.13 y la Figura 3.14 presentan el resultado de la ubicación de $n = 3$ centroides con K-Means.

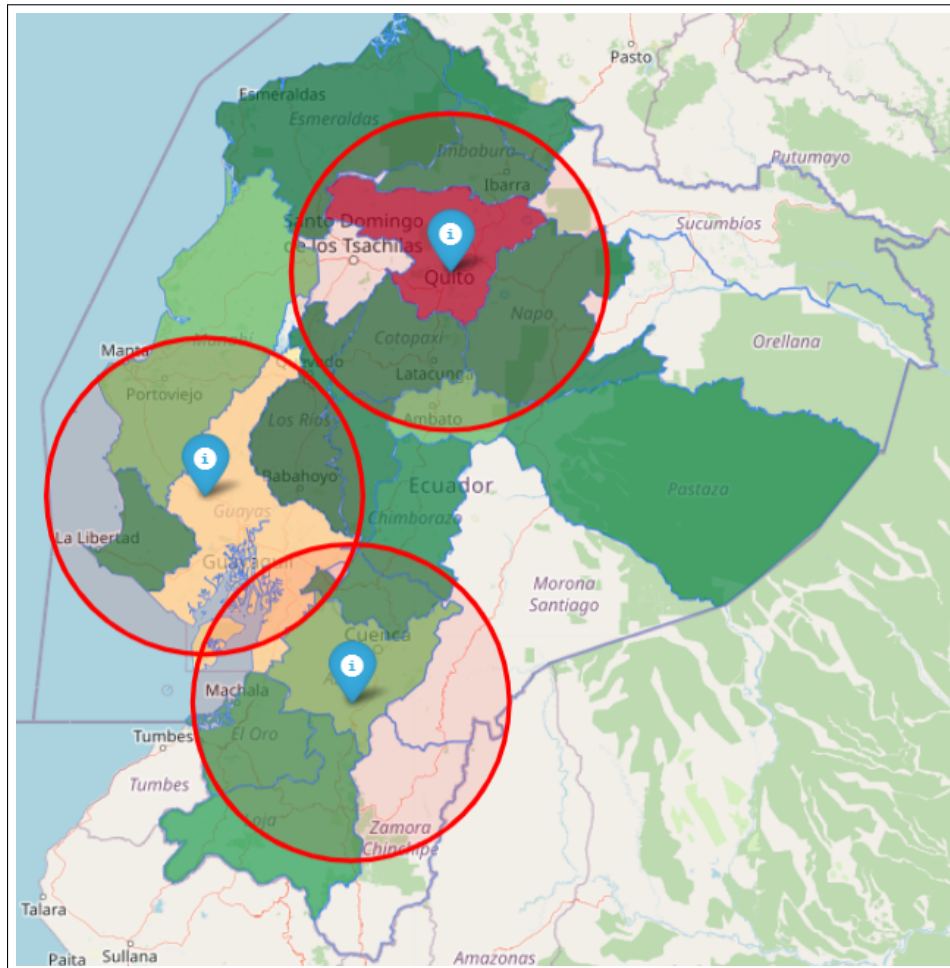


Figura 3.13: Ubicación geográfica de 3 CSEs con K-Means.

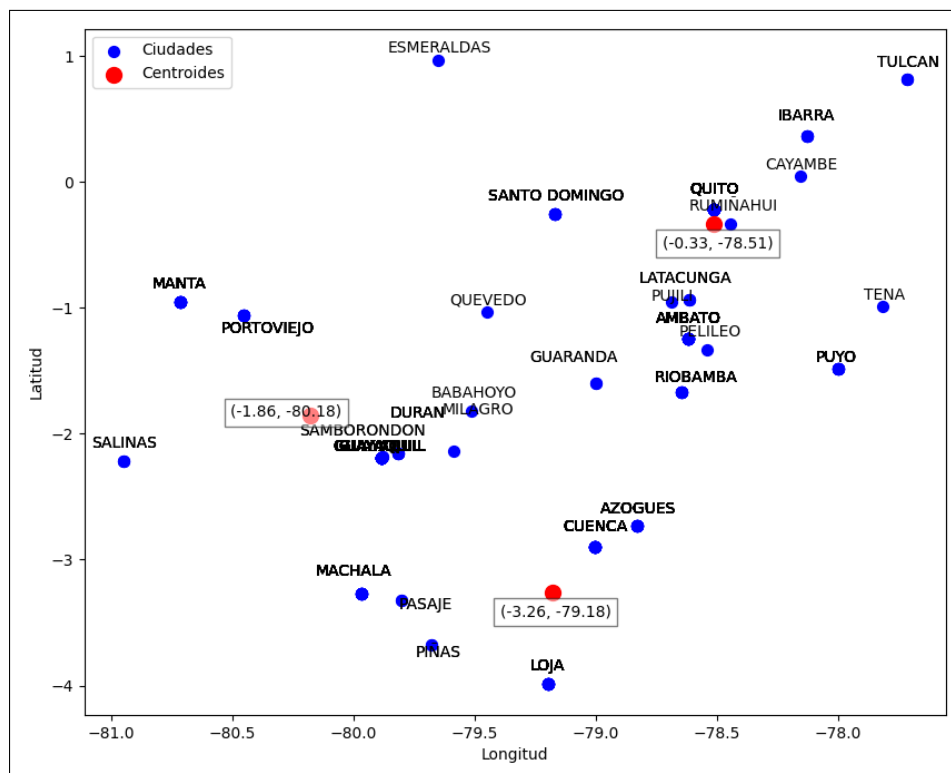


Figura 3.14: Plano cartesiano: ubicación geográfica de 3 CSEs con K-Means.

Como era de esperar, dada la alta demanda de servicios en Quito, Guayaquil y Cuenca, se observa centroides estratégicamente posicionados cerca de estas ciudades.

El centroide en $(-0.33, -78.51)$ corresponde claramente al área metropolitana de Quito, capturando la considerable demanda de servicios en la capital y sus alrededores. Otro centroide se ubica en $(-1.86, -80.18)$, alineándose estrechamente con la región de Guayaquil, abordando las altas necesidades de servicio en esta importante ciudad costera.

El tercer centroide, situado en $(-3.26, -79.18)$, está estratégicamente ubicado en la parte sur del país, abarcando ciudades como Cuenca y Loja. Esto sugiere que el modelo reconoce una concentración significativa de demanda de servicios en esta región también.

La Figura 3.15 presenta el resultado de la ubicación de $n = 3$ centroides con Agrupamiento Aglomerativo.

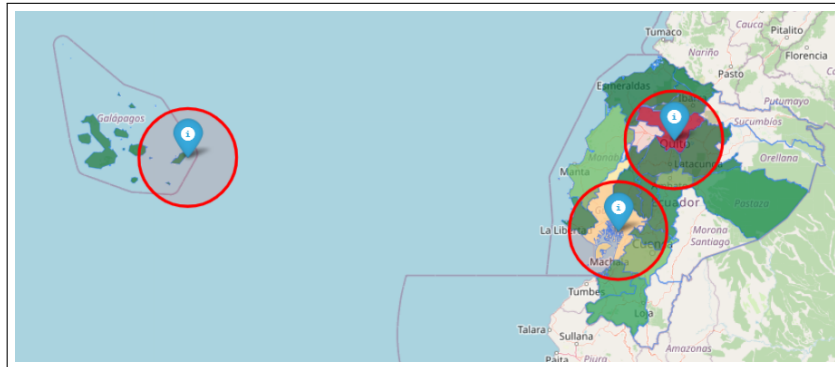


Figura 3.15: Ubicación geográfica de 3 CSEs con Agrupamiento Aglomerativo.

El mapa muestra la distribución geográfica de los tres clústeres generados por el algoritmo de Agrupamiento Aglomerativo. Dos de los centroides se ubican de manera similar a los resultados de K-Means, cerca de Quito y en la región costera, lo que sugiere una adecuada identificación de las principales áreas de demanda de servicio.

Sin embargo, el tercer centroide se encuentra en la región Insular (Galápagos), lo cual plantea un desafío logístico y operativo. A pesar de que el modelo ha identificado un clúster en esta región, la contratación de un CSE a las Islas Galápagos sería ineficiente debido a la lejanía y los altos costos asociados al transporte y viáticos. Además, es probable que la demanda de servicios en esta región sea significativamente menor en comparación con las zonas continentales, lo que resultaría en una subutilización del CSE asignado.

La Figura 3.16 presenta el resultado de la ubicación de $n = 3$ centroides con GMM.

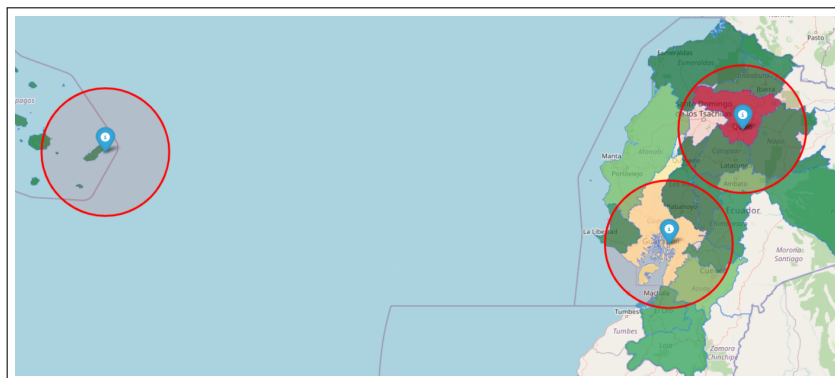


Figura 3.16: Ubicación geográfica de 3 CSEs con GMM.

La imagen muestra la distribución geográfica de los tres clústeres generados por el

algoritmo GMM. Al igual que en los modelos anteriores, se observa un centroide cerca de Quito y otro en la región costera, lo que sugiere una adecuada identificación de estas áreas de alta demanda de servicio.

Sin embargo, al igual que con el Agrupamiento Jerárquico, el tercer centroide se encuentra en la región Insular (Galápagos).

3.6.3 Justificación

A pesar de que los resultados teóricos del Agrupamiento Aglomerativo y GMM mostraron métricas ligeramente superiores en algunos aspectos, como la definición de clústeres en el caso de Agrupamiento Aglomerativo, la elección final del modelo predictivo fue K-Means. Esta decisión se fundamenta en la capacidad de K-Means para incorporar condiciones iniciales relevantes al contexto del problema a través de un sistema de pesos.

La flexibilidad de K-Means para ajustar la importancia de ciertas características, como la presencia de aeropuertos o la cantidad de equipos por ciudad, permite una adaptación más precisa del modelo a la realidad operativa de SHS en Ecuador. Esto se traduce en una contratación de CSEs más estratégica y alineada con los objetivos de la empresa, priorizando no solo la cercanía geográfica, sino también factores críticos como la accesibilidad y la carga de trabajo.

Aunque otros modelos pueden ofrecer un rendimiento teórico ligeramente superior en métricas abstractas, K-Means demuestra una mayor capacidad para adaptarse a las necesidades específicas del negocio y proporcionar soluciones eficientes y sencillas para mejorar la contratación de CSEs.

3.7 FASE 6: DESPLIEGUE

Esta última fase busca facilitar el uso del modelo K-Means desarrollado, permitiendo a los usuarios finales determinar la ubicación óptima de los CSEs de manera sencilla e intuitiva. Para ello, se implementará una interfaz en Python que permitirá ingresar la cantidad deseada de CSEs (n) y visualizar los resultados en un mapa interactivo.

3.7.1 Diseño de la interfaz gráfica en Python

La interfaz se diseñó utilizando bibliotecas de Python como Tkinter, que permiten crear interfaces gráficas de usuario (GUI) de manera sencilla. La interfaz contiene los siguientes elementos:

- ❑ **Entrada de datos:** Un campo de entrada donde el usuario podrá ingresar el número de CSEs a ubicar (n).
- ❑ **Botón de ejecución:** Un botón que, al ser presionado, ejecutará el modelo K-Means con el valor de n ingresado.
- ❑ **Visualización de resultados:** Un mapa interactivo generado con Folium que mostrará la ubicación geográfica de los equipos, los clústeres identificados por K-Means y los centroides de los clústeres (ubicaciones sugeridas para los CSEs).

3.7.2 Ejecución del modelo y generación del mapa

La Figura 3.17 representa el diagrama de flujo que simula el evento de presionar el botón de ejecución.

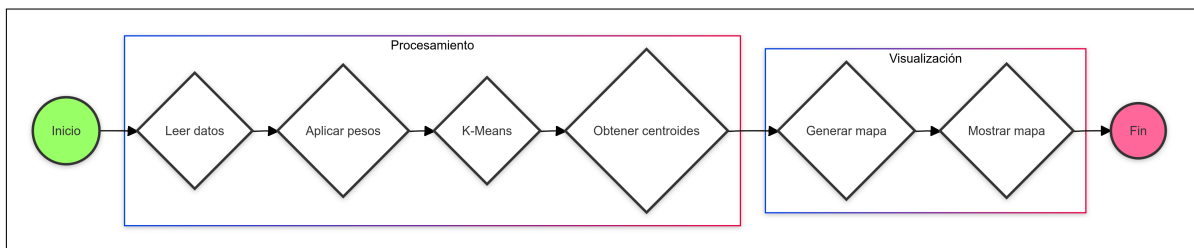


Figura 3.17: Diagrama de flujo de la interfaz GUI implementada.

Dónde la interfaz presentada en la Figura 3.17 realiza las siguientes acciones:

1. **Lectura de datos:** Tomará toda la información de las diferentes tablas dinámicas que se actualizan a tiempo real día a día junto las tablas fijas y generará una sola tabla con toda la información del modelo predictivo.
2. **Aplicación de pesos** Aplicará la ponderación de pesos a los siguientes criterios: a las ciudades con aeropuerto, ciudades que tengan equipos con contratos vigentes y equipos que dispongan almenos 5 equipos instalados.
3. **Aplicación de K-Means:** Entrenará el modelo K-Means con el valor de n ingresado por el usuario y las coordenadas geográficas de las ciudades filtradas.
4. **Obtención de centroides:** Calculará las coordenadas de los centroides de los clústeres resultantes.

5. **Generación y visualización del mapa:** Creará un mapa de Folium que muestre la ubicación de los equipos, los clústeres y los centroides, permitiendo al usuario explorar los resultados.

El Código 3.10 presenta la codificación de la interfaz gráfica de la fase de despliegue.

Código 3.10: Implementación GUI con Tkinter.

```
1 horas_bi=cargar_y_procesar_datos()
2
3 def solo_numeros(char):
4     return char.isdigit()
5 ventana = tk.Tk()
6 vcmd = (ventana.register(solo_numeros), '%S')
7 ventana.title("App de Simulación de contratación de CSEs")
8 ventana.configure(bg='orange')
9
10 #Responsive
11 ventana.columnconfigure(0, weight=1) # Columna principal expandible
12 ventana.rowconfigure(0, weight=0)    # Fila del logo (no expandible)
13 ventana.rowconfigure(1, weight=1)   # Fila central expandible
14 ventana.rowconfigure(2, weight=0)   # Fila de los botones (no expandible)
15 original_logo_image = tk.PhotoImage(file='T:\\Logo.png')
16 logo_image = original_logo_image.subsample(10, 10)
17
18 ventana.minsize(width=1000, height=500)
19 mensaje_bienvenida = tk.Label(
20     ventana,
21     text=f"Bienvenido a la App de Simulación de contratación y contratación de CSEs",
22     font=("Arial", 16, "bold"),
23     bg='orange',
24     wraplength=400,
25     justify='center'
26 )
27 mensaje_bienvenida.grid(row=0, column=0, pady=(logo_image.height() + 10, 10))
28 # Texto explicativo
29 cantidad_CSEs = round(horas_bi/1280)
30 texto_explicativo = tk.Label(
31     ventana,
32     text=f"Acorde a la demanda de horas de la base instalada ({horas_bi} hrs) y
33         asumiendo una productividad del 80% ANUAL (1280 hrs) por cada CSE se
34         recomienda realizar una simulación con {cantidad_CSEs} CSEs",
35     font=("Arial", 14),
36     bg='orange',
37     wraplength=400, #Ancho
38     justify='center'
39 )
40 label_cantidad_cses = tk.Label(ventana, text="Ingrese la cantidad de CSEs a simular:",
41     font=("Arial", 12), bg='orange')
42 label_cantidad_cses.grid(row=2, column=0, pady=10) # Centrar en la fila central
43 ventana.mainloop()
```

Esta codificación 3.10 construye una interfaz gráfica de usuario (GUI) utilizando la biblioteca Tkinter en Python. La GUI tiene como propósito principal servir como punto de entrada para una simulación de contratación y asignación de ingenieros de servicio.

La interfaz se inicia con la definición de una función *solo_numeros* que restringe la entrada del usuario en un campo de texto a solo números. Luego, se crea la ventana principal, se configura su título, color de fondo y se ajusta su tamaño mínimo. Se carga y se muestra el logo de la compañía.

Finalmente se muestra un mensaje de bienvenida y un texto explicativo que incluye la cantidad recomendada de CSEs a contratar, basada en la demanda de horas de servicio y la productividad esperada. Se proporciona un campo de entrada donde el usuario puede ingresar la cantidad de CSEs que desea simular. Finalmente, se presentan dos botones: "Iniciar simulación" para ejecutar la simulación con la cantidad de CSEs ingresada y "Salir" para cerrar la aplicación. La Figura 3.18, la Figura 3.19 y la Figura 3.20 presentan la GUI implementada con Python.

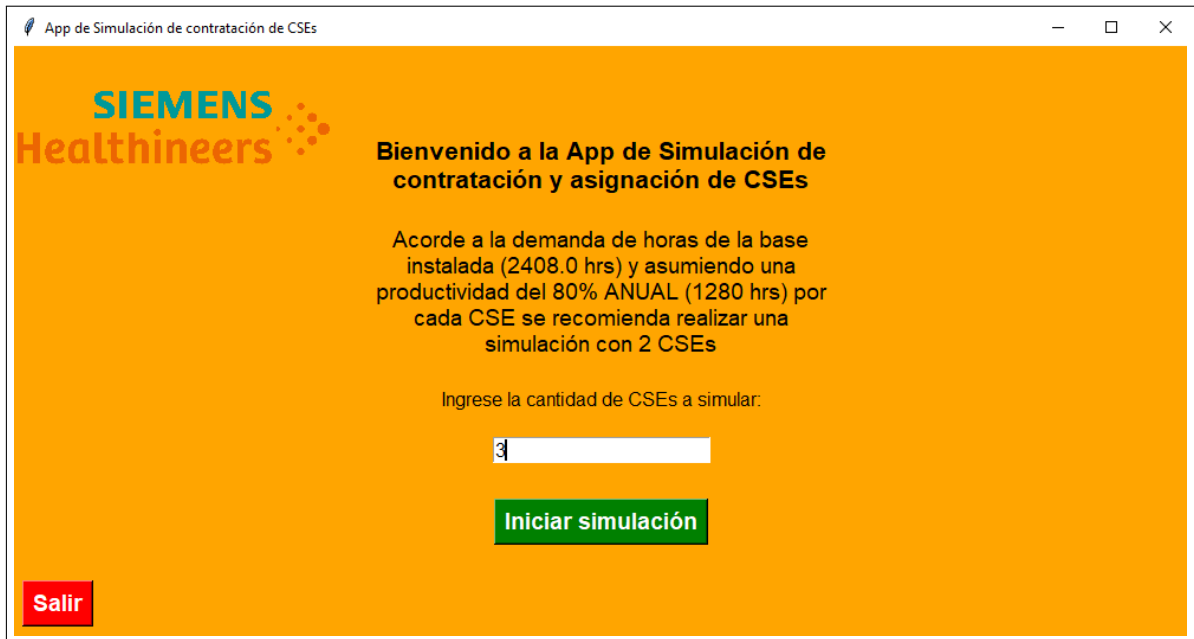


Figura 3.18: Pantalla de inicio interfaz implementada.

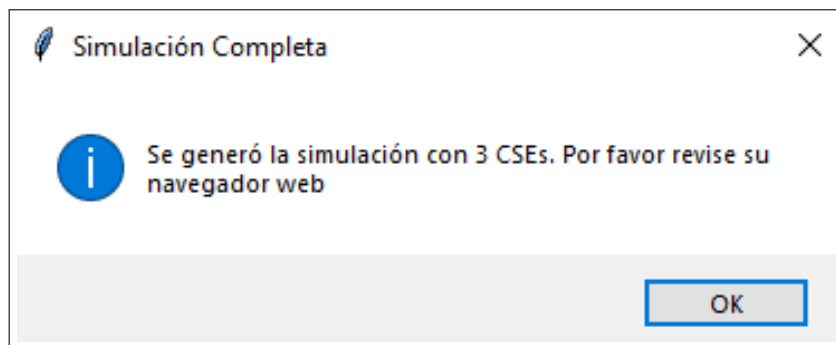


Figura 3.19: Ventana/Pop Up emergente.

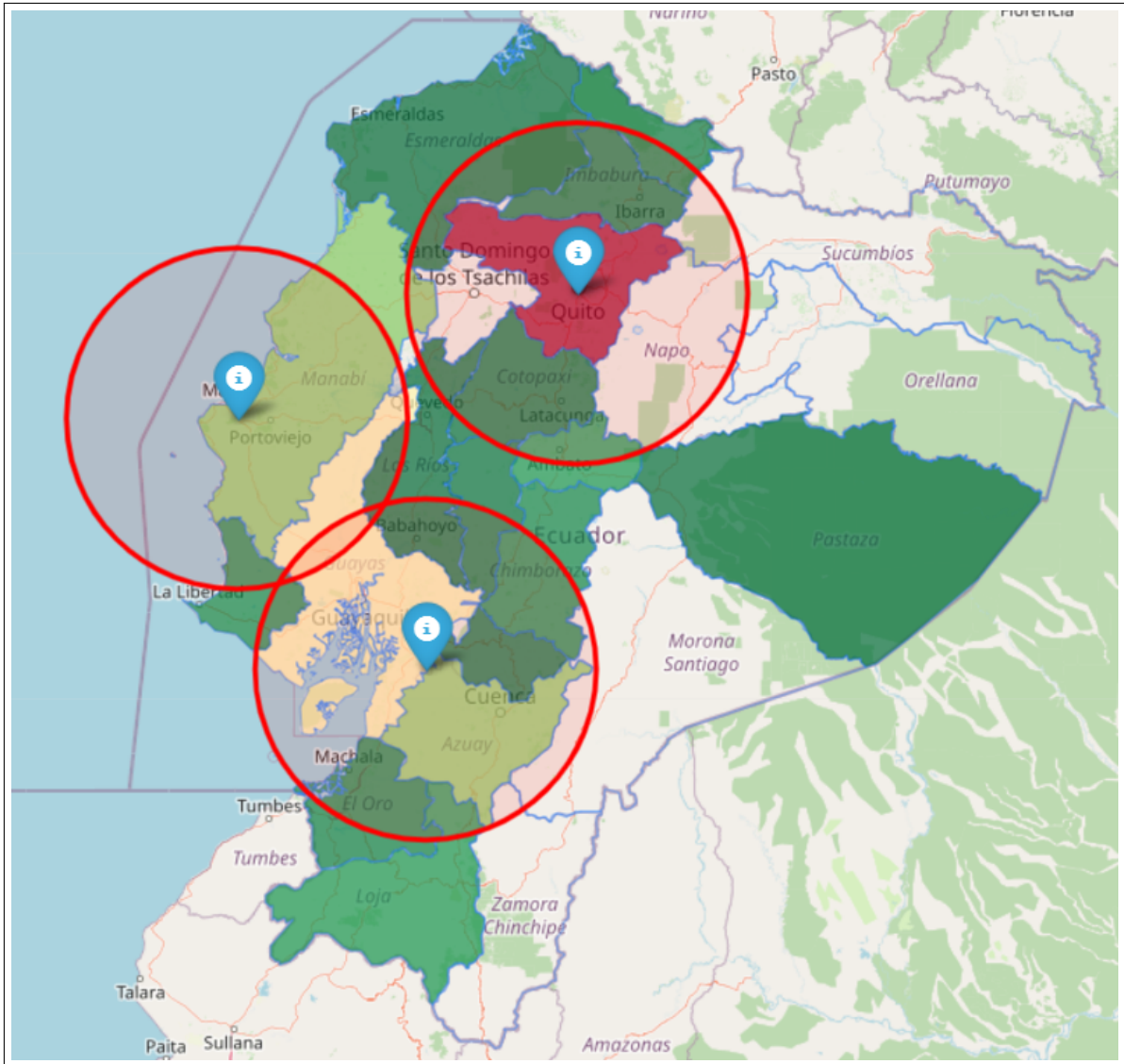


Figura 3.20: Simulación final presentadas en un ejemplo con 3 centroides.

Una vez que la interfaz esté en funcionamiento, es fundamental establecer un plan de monitoreo y mantenimiento para garantizar su correcto funcionamiento a largo plazo. Esto incluye la actualización periódica de los datos de entrada, la revisión y ajuste de los parámetros del modelo, y la detección y corrección de posibles errores o problemas técnicos. El monitoreo continuo permitirá identificar áreas de mejora y adaptar el modelo a las variaciones en la distribución territorial de los equipos o en la demanda de servicios.

Así concluye el caso de estudio I, donde se abordó exitosamente la problemática de determinar la ubicación óptima para la contratación de nuevos CSEs a través de modelos de agrupamiento espacial, ahora nos adentramos en el caso de estudio II. En esta etapa, aprovecharemos los resultados obtenidos previamente para abordar el desafío de asignar

eficientemente los CSEs existentes a las solicitudes de mantenimientos preventivos y correctivos. Esta etapa es fundamental para asegurar una respuesta ágil y eficiente a las demandas de los clientes, optimizando distancias a través de la ubicación y experiencia que tienen los CSEs.

3.8 CASO DE ESTUDIO II: ASIGNACIÓN EN LOS SERVICIOS DE CAMPO

Este caso de estudio hace referencia a la asignación de CSEs a las solicitudes de mantenimientos preventivos y correctivos. A partir de la cantidad de horas anuales que demanda la base instalada y considerando factores como la experiencia de cada técnico, la distancia a recorrer y la disponibilidad de transporte aéreo, se busca optimizar la asignación de recursos para garantizar una respuesta rápida y eficiente a las necesidades de los clientes. Este caso de estudio se centrará en escenarios donde no existe un CSE en la ciudad donde se requiere el servicio, explorando cómo seleccionar el técnico más idóneo para cada tarea, minimizando los tiempos de desplazamiento y maximizando la productividad.

3.8.1 Análisis previo

Para obtener la información y experiencia de los CSEs se empleó el sistema SAP como fuente de datos, la transacción IW67 que permite obtener todas las tareas reportadas por los CSEs y en particular se hizo énfasis en los trabajos realizados en campo durante el último año y adicional se utilizó la tabla fija geográfica por provincias. El Código 3.11 presenta el procesamiento que se le dió a la tabla de CSEs.

Código 3.11: Implementación GUI con TKinter II.

```
1 cse_unicos = df_procesado['CSE'].unique() # Identificar valores únicos
2 cse_map = {cse: f'CSE_{i+1}' for i, cse in enumerate(cse_unicos)} # Crear diccionario
  de mapeo
3 df_procesado['CSE_ID'] = df_procesado['CSE'].map(cse_map) # Reemplazar con
  identificadores anónimos
4 df_procesado.to_excel('DataCSE.xlsx', index=False)
5 df_encoded = pd.read_excel('DataCSE.xlsx', sheet_name='Sheet1')
6 df_encoded = pd.get_dummies(df_encoded, columns=['ModSAP'], prefix='', prefix_sep='')
7 co2 = ['DescriMod']
8 df_encoded = df_encoded.drop(columns=co2)
9 df_encoded.to_csv('DataCSE.txt', sep='\t', index=False)
10 df_cse=df_encoded
11
12 #Cantidad CSEs por ciudad
13 cses_por_ciudad = df_cse.groupby(['CIUDAD', 'LATITUD', 'LONGITUD'])['CSE_ID'].nunique
  ().reset_index(name='NUM_CSEs')
```

Esta codificación 3.11 crea nuevas columnas binarias (0 o 1) para cada modalidad, indicando si el CSE tiene experiencia en esa modalidad o no. La siguiente parte del código se

La Figura 3.21 destaca la ubicación de los CSEs en 3 ciudades clave. Los números en círculos en el mapa corresponden a la cantidad de CSE presentes en cada ciudad. De esto, Quito tiene la mayor concentración de CSE con 7, seguido de Guayaquil con 3. Cuenca, marcada con un solo número en un círculo, tiene 1 CSE disponible.

La Tabla 3.9 muestra la distribución de las áreas de experiencia de los CSEs en Ecuador. Cada fila representa un CSE, identificado por su ID, y cada columna representa una modalidad de servicio específica. Un valor de “1” indica que el CSE tiene experiencia en esa modalidad, mientras que “0” indica que no la tiene.

Tabla 3.9: Experiencia CSE por modalidad.

CSE_ID	AX	CT	DS	MI	MR	SY	US	XP
1	1	1	0	1	1	1	1	1
2	1	1	0	1	1	1	1	1
3	1	1	0	1	1	1	1	1
4	1	1	0	1	1	1	1	1
5	1	1	0	1	1	1	1	1
6	1	1	0	1	1	1	1	1
7	1	1	1	1	1	1	1	1
8	1	1	0	1	1	1	1	1
9	1	1	0	1	1	1	1	1
10	1	1	0	0	1	0	1	0
11	1	1	1	0	1	1	1	1

Un trabajo futuro propone evaluar la calidad de los ingenieros para garantizar que el equipo de CSEs posea las habilidades y conocimientos necesarios para brindar un servicio técnico de alta calidad. Esta evaluación facilita la identificación de las zonas que necesitan capacitación adicional, tanto para los nuevos CSEs como para los que ya están en el campo.

Según la evaluación de calidad de los CSEs, se puede desarrollar un plan de capacitación integral que aborde las necesidades específicas de cada CSE y de la empresa en general. Este plan debe considerar tanto las áreas de fortaleza como las áreas de mejora, y debe incluir estrategias de capacitación efectivas, como cursos presenciales/virtuales, talleres y seminarios.

La siguiente sección presenta el código implementado para la asignación de CSEs en Ecuador.

3.8.2 Implementación asignación de CSEs

El Código 3.12 presenta una parte de la codificación para la resolución del caso de estudio de asignación de CSEs. Véase el *Anexo 6)* para visualizar el código completo.

Código 3.12: Implementación GUI asignación de CSE I.

```
1 ventana = tk.Tk()
2 ventana.title("Asignación de CSEs")
3 # Listado de ciudades
4 ciudades = df_MAIN['City'].unique().tolist()
5 ciudad_aeropuerto_map = df_MAIN.set_index('City')['AEROPUERTO?'].to_dict()
6 ciudades = sorted(ciudades)
7
8 #cbxCIUDDAD
9 label_ciudad = tk.Label(ventana, text="Seleccione una ciudad:")
10 cbxDP_ciudad.bind("<<ComboboxSelected>>", actualizar_cbxDP_y_label)
11 #cbxCLIENTE
12 label_cliente = tk.Label(ventana, text="Seleccione un cliente:")
13 cbxDP_ciudad.bind("<<ComboboxSelected>>", actualizar_cbxDP_y_label)
14
15 #cbxMODALIDAD
16 label_modalidad = tk.Label(ventana,
17 cbxDP_modalidad.bind("<<ComboboxSelected>>", actualizar_cbxMODALIDAD)
18
19 #cbxCLIENTE
20 label_cliente = tk.Label(ventana, text="Seleccione un cliente:")
21 cbxDP_cliente.bind("<<ComboboxSelected>>", actualizar_cbxCLIENTE)
22 label_aeropuerto = tk.Label(ventana, text="", font=("Arial", 12))
23 #Botón
24 boton_asignar = tk.Button(ventana, text="Asignar CSE", command=asignar_cse, state='
    disabled')
25
26 #Treeview
27 treeview = ttk.Treeview(ventana, columns=('N', 'CSE', 'Ciudad CSE', 'Distancia [Km]', '
    Tiempo de respuesta [hrs]'), show='headings')
28
29 def on_cbxDP_select(event):
30     limpiar_treeview()
31
32 def limpiar_treeview():
33     for item in treeview.get_children():
34         treeview.delete(item)
35
36 treeview.heading('N', text='N')
37 treeview.heading('CSE', text='CSE')
38 treeview.heading('Ciudad CSE', text='Ciudad CSE')
39 treeview.heading('Distancia [Km]', text='Distancia [Km]')
40 treeview.heading('Tiempo de respuesta [hrs]', text='Tiempo de respuesta [hrs]')
41 treeview.column('N', width=30)
42 treeview.column('CSE', width=100)
43 treeview.column('Ciudad CSE', width=150)
44 treeview.column('Distancia [Km]', width=160)
45 treeview.column('Tiempo de respuesta [hrs]', width=250)
46 ventana.mainloop()
```

Para la asignación de CSEs se creó una interfaz en Tkinter considerando su ubicación, experiencia y la necesidad de transporte aéreo o terrestre. La interfaz permite al usuario seleccionar una ciudad, una modalidad de servicio y un cliente específico. Posteriormente, al seleccionar el botón “Asignar CSE”, se calcula las distancias y tiempos de viaje para los CSEs disponibles que tienen la experiencia necesaria, y presenta una tabla ordenada con los CSEs más adecuados para realizar el mantenimiento.

Adicional, se implementó un sistema de filtros dinámicos en forma de combobox (listas desplegables) para guiar la selección del usuario. El primer combobox muestra las ciudades disponibles. Una vez que se selecciona una ciudad, el segundo combobox se actualiza para mostrar solo las modalidades de servicio presentes en esa ciudad. Finalmente, al seleccionar una

modalidad, el tercer combobox muestra los clientes asociados a esa ciudad y modalidad. Este enfoque de filtrado dinámico asegura que las opciones presentadas al usuario sean relevantes y coherentes con sus elecciones anteriores, mejorando la usabilidad de la interfaz.

Cálculo de distancias y tiempos de viaje

Dentro del programa se creó una función para el cálculo de distancia entre las dos ciudades, clave para determinar la distancia entre la ubicación del equipo que requiere mantenimiento y la ubicación de cada CSE disponible. Si la ciudad del equipo tiene aeropuerto y el CSE también está en una ciudad con aeropuerto, se utiliza la fórmula de Haversine² para calcular la distancia geográfica en línea recta, asumiendo un viaje en avión. En caso contrario, se consulta una tabla de distancias terrestres (Asesor de Planificación-Estadística, 2024).

Una vez calculada esta distancia, se calculó el tiempo de viaje en función de la distancia y el medio de transporte. Para viajes en avión, se asume un tiempo fijo de 3 horas, que incluye el vuelo y otros tiempos asociados como migración. Para viajes terrestres, se divide la distancia por una velocidad promedio de 60km/h . Si el CSE se encuentra en la misma ciudad que el equipo, se asume un tiempo de viaje referencial de 2 horas.

Asignación de CSEs y presentación de resultados

La asignación final del CSE se realiza en dos etapas. Primero, se busca un CSE ubicado en la misma ciudad que el equipo, asignándolo directamente si se encuentra disponible. En caso contrario, se filtra a los CSEs que poseen la experiencia requerida para la modalidad de servicio y se calcula la distancia y el tiempo de viaje estimado para cada uno. Finalmente, se presenta al usuario una tabla ordenada con los CSEs más idóneos, priorizando aquellos con menor distancia y tiempo de respuesta. Es importante destacar que este programa actúa como un recurso de apoyo para la elección de decisiones, sugiriendo al CSE más adecuado para cada tarea, pero no tiene en cuenta su disponibilidad real en la agenda.

La Figura 3.22 presenta la interfaz implementada en Tkinter de este caso de estudio.

²**Fórmula de Haversine:** Calcula la distancia mínima entre dos coordenadas en una esfera basándose en latitudes y longitudes, considerando la curvatura de la Tierra (Amiruzzaman et al., 2021).

Asignación de CSEs

SIEMENS Healthineers

Simulación de asignación de CSEs

Por favor seleccione las opciones para visualizar cuáles son los CSEs idóneos para el trabajo en Campo

Seleccione una ciudad: AZOGUES

Seleccione una modalidad: SY

Seleccione un cliente: CLIENTE_175

La ciudad NO tiene aeropuerto, se enviará un CSE vía terrestre

Asignar CSE

N°	CSE	Ciudad CSE	Distancia [Km]	Tiempo de respuesta [hrs]
1	1	CUENCA	41	0h41
2	8	GUAYAQUIL	209	3h29
3	9	GUAYAQUIL	209	3h29
4	11	QUITO	391	6h31
5	2	QUITO	391	6h31
6	3	QUITO	391	6h31
7	4	QUITO	391	6h31
8	5	QUITO	391	6h31
9	6	QUITO	391	6h31
10	7	QUITO	391	6h31

Salir

Figura 3.22: Simulación asignación de CSEs.

La interfaz de usuario desarrollada permite una asignación eficiente y transparente de los CSEs, considerando tanto la demanda de servicio como la ubicación y experiencia de los técnicos. Al seleccionar una ciudad, modalidad y cliente específicos, el sistema proporciona una lista ordenada de los CSEs más adecuados, erradicando el trabajo manual de búsqueda. La visualización del heatmap, muestra distribución geográfica de la demanda de servicio, complementa la información tabular y permite una comprensión más completa del contexto en el que se realiza la asignación.

El particular, los resultados mostrados en la interfaz para la ciudad de Azogues y la modalidad SY ilustran la utilidad del sistema. A pesar de que Azogues no cuenta con un aeropuerto, el sistema identifica al *CSE_1* ubicado en Cuenca como la opción más cercana, con un tiempo de respuesta estimado de 41 minutos. Adicionalmente, se presentan otros CSEs ubicados en Guayaquil y Quito, aunque con tiempos de respuesta considerablemente mayores

debido a la distancia y la necesidad de transporte terrestre.

La implementación de esta interfaz de asignación de CSEs representa un avance significativo en la optimización del servicio técnico de Siemens Healthineers en Ecuador. Al combinar información geográfica, datos de demanda y criterios de experiencia, el sistema proporciona recomendaciones prácticas y eficientes para la asignación de recursos, lo que resulta una reducción en los tiempos de respuesta, una mayor satisfacción del cliente y una gestión más efectiva de los técnicos de campo. Este enfoque basado en datos tiene el potencial de generar un efecto favorable en la eficiencia y los beneficios de SHS, al tiempo que garantiza un servicio de alta calidad para las instituciones de salud en todo el país.

A lo largo de este trabajo, se ha abordado el desafío de optimizar la gestión de ingenieros de campo (CSEs) en Siemens Healthineers Ecuador, desde dos perspectivas fundamentales: la contratación estratégica de nuevo personal y la asignación eficiente de los CSEs a las mantenimientos preventivos y/o correctivos. Mediante la aplicación de CRISP-DM se desarrollaron modelos predictivos que permiten identificar ubicaciones óptimas para la contratación de nuevos CSEs y asignar de manera inteligente los recursos existentes, considerando factores como la demanda de servicio, la ubicación geográfica y la experiencia de los técnicos.

Los hallazgos revelan la capacidad de la ciencia de datos para optimizar la QoS y la eficiencia operativa en la industria de la medicina por imagen y medicina por laboratorio. La implementación de estos modelos en un entorno real permitiría a la empresa optimizar sus recursos. Sin embargo, este trabajo también abre nuevas líneas de investigación y desarrollo, que serán exploradas en la siguiente sección de Trabajo Futuro.

3.9 FUTURE WORK

Si bien este trabajo de titulación ha sentado las bases para una asignación más eficiente de los CSEs, existen oportunidades de mejora y expansión. Un aspecto crucial a considerar en trabajos futuros es la incorporación del **factor económico** en el modelo de optimización. La inclusión de costos asociados a viáticos, transporte, costo de boletos de aviones, costo por Km terrestre, entre otros permitiría evaluar el impacto financiero de las decisiones de asignación y buscar soluciones que maximicen no solo la eficiencia operativa, sino también la rentabilidad de la empresa. Esto es especialmente relevante en el contexto actual, donde la productividad y el margen bruto son indicadores clave de éxito para las empresas.

Asimismo, se sugiere ampliar el estudio de las habilidades y competencias de los CSEs.

En lugar de basarse únicamente en la modalidad general del equipo, se podría realizar un análisis más detallado de los cursos y certificaciones obtenidos por cada técnico. Esto permitiría una asignación más precisa y personalizada, garantizando que los CSEs cuenten con la experiencia y los conocimientos específicos necesarios para cada tarea de mantenimiento. La integración de esta información en el modelo de asignación podría optimizar aún más la eficacia del servicio técnico.

4 CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

La implementación de modelos de agrupamiento (clustering) para la asignación óptima de Ingenieros de Campo (CSEs) en Siemens Healthineers Ecuador ha demostrado ser una estrategia eficaz para mejorar la eficiencia operativa y la calidad del servicio técnico. Los resultados obtenidos al aplicar los modelos de agrupamiento K-Means y Aglomerativo a la asignación de CSEs son prometedores. El coeficiente de silueta, que oscila entre -1 y 1, alcanzó valores de 0.6565 y 0.6599 para K-Means y Agrupamiento Aglomerativo, respectivamente. Estos valores, situados en la parte superior del rango, indican una alta cohesión interna de los clusters y una buena separación entre ellos, lo cual sugiere que ambos modelos han logrado identificar grupos geográficos significativos.

En el ámbito del servicio técnico, el uso de técnicas de machine learning para la asignación de técnicos especialistas ofrece beneficios tangibles. Al considerar factores geográficos y la demanda de servicios, estos modelos permiten una distribución más inteligente de los recursos, minimizando los tiempos de desplazamiento y los costos asociados. Esto se traduce en una mayor rapidez y eficiencia en la atención a los clientes, lo que a su vez mejora su satisfacción y fortalece la imagen de la empresa. Además, la optimización de la asignación de personal contribuye a equilibrar la carga de trabajo y a maximizar la productividad de los técnicos, lo que impacta positivamente en la rentabilidad y el crecimiento del negocio.

Los modelos predictivos, como el método de agrupamiento utilizado en este proyecto, se han vuelto esenciales para aumentar la eficiencia y apoyar la toma de decisiones estratégicas en diversas industrias. Al analizar grandes cantidades de datos y descubrir patrones clave, las empresas pueden mejorar sus operaciones, disminuir gastos y elevar la calidad de sus servicios.

La ciencia de datos transforman la gestión de operativa en empresas de cualquier índole. La implementación de modelos predictivos como K-Means permite tomar decisiones más informadas y eficientes, lo que resulta en ventajas concretas tanto para la empresa como para sus clientes. La integración de estos modelos en sistemas de información dinámicos y accesibles facilita su adopción y promueve una cultura de mejora continua basada en datos.

Además, la combinación de análisis espacial, modelos de agrupamiento y herramientas de

visualización interactiva ofrece un enfoque poderoso para abordar problemas de asignación de recursos en contextos geográficamente distribuidos. La capacidad de identificar patrones en la cantidad de servicios y de determinar la ubicación ideal de los técnicos permite a las empresas tomar decisiones estratégicas que maximizan la eficiencia y minimizan los costos operativos. Este enfoque puede ser aplicado no solo en el sector de la tecnología médica, sino también en otras industrias que requieran una gestión eficiente de personal de campo, como telecomunicaciones, energía y servicios públicos.

La ciencia de datos no solo impulsa la automatización y optimización de procesos, sino que también introduce una transparencia sin precedentes en la toma de decisiones dentro de los stakeholders. Al basarse en datos concretos y modelos predictivos, las decisiones dejan de ser intuitivas o basadas en experiencias pasadas, y se convierten en acciones estratégicas respaldadas por evidencia. Esto permite a las empresas justificar sus decisiones de manera objetiva, mejorando la confianza tanto interna como externa, y promoviendo una cultura de perfeccionamiento constante fundamentada en el análisis de datos.

4.2 RECOMENDACIONES

Antes de implementar el proyecto desde una perspectiva operativa, se recomienda involucrar a los stakeholders sobre el proyecto. Su conocimiento práctico y experiencia en el día a día de la operación son invaluable, ya que pueden aportar perspectivas cruciales sobre la realidad del país y cómo ésta podría diferir del proceso teórico. La participación activa de los stakeholders en CRISP-DM asegurará que la investigación se adapte a las necesidades y desafíos específicos del contexto local, maximizando así su impacto y efectividad.

Se recomienda ampliar el alcance de este proyecto para considerar no solo la optimización de los tiempos de viaje, sino también otros factores clave para la empresa, como la reducción de costos asociados en el proceso completo de la realización de los mantenimientos por ejemplo viáticos, pasajes de aviones/terrestres, hotel, alimentación entre otros. Esta visión más completa permitirá desarrollar un modelo de asignación de CSEs que se ajuste de manera más precisa a las necesidades y prioridades de la empresa, maximizando así el impacto positivo en la eficiencia y la calidad del servicio.

Integrar modelos junto con los análisis desarrollados dentro de un sistema de información dinámico y accesible en tiempo real. Esta plataforma permitiría a los usuarios finales, como los gerentes de servicio técnico y/o al área de ventas, interactuar con los resultados de manera

intuitiva y hacer elecciones basadas en datos para la distribución de recursos.

Es importante destacar que la implementación de este modelo de optimización es solo el primer paso en dirección a una administración más efectiva de los recursos de mantenimiento en Siemens Healthineers Ecuador. Se sugiere seguir investigando y adoptando métodos y tecnologías emergentes basados en Data Science, para mejorar aún más la asignación de CSEs y garantizar un servicio técnico de alta calidad para los clientes.

Se recomienda considerar la escalabilidad del sistema implementado, anticipando posibles aumentos en la cantidad de equipos o CSEs. Esto podría implicar la implementación de una infraestructura más robusta o la optimización del código para manejar grandes volúmenes de datos de manera eficiente. Además, se podría explorar la integración del modelo con otros sistemas de gestión de la empresa para automatizar aún más el proceso de asignación de CSEs y mejorar la eficiencia operativa.

Un elemento adicional a tener en cuenta es la opción de integrar comentarios de los usuarios finales en el proceso de perfeccionamiento constante del modelo. Esto podría incluir: i) comentarios de los stakeholders, ii) la recopilación de datos sobre la evaluación del desempeño de los CSEs en las diferentes ubicaciones y iii) la identificación de nuevas variables relevantes para la asignación. Al incorporar esta retroalimentación, el modelo se volverá más preciso y útil con el tiempo, adaptándose a las necesidades cambiantes del negocio.

5 REFERENCIAS BIBLIOGRÁFICAS

- Ackermann, M. R., Blömer, J., Kuntze, D., & Sohler, C. (2014). Analysis of agglomerative clustering. *Algorithmica*, 69, 184-215.
- Amat, R. (2020). Clustering con Python.
<https://cienciadedatos.net/documentos/py20-clustering-con-python>
- Amiruzzaman, M., Rahman, R., Islam, M. R., & Nor, R. M. (2021). Evaluation of DBSCAN algorithm on different programming languages: An exploratory study. *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 1-6.
- Arias, E. B. N., Nuñez, B. M. G., Fernández, L. N., & Pupo, J. M. R. (2021). CRISP-DM y K-means neutrosofía en el análisis de factores de riesgo de pérdida de audición en niños. *Revista Asociación Latinoamericana de Ciencias Neutrosóficas*.
- Asesor de Planificación-Estadística. (2024). Cuadro de Distancias Terrestres en Kilómetros. *Ministerio de Transporte y Obras Públicas*. <https://www.obraspublicas.gob.ec/>
- Ashari, I., Nugroho, E., Baraku, R., Yanda, I., Liwardana, R., et al. (2023). Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta. *Journal of Applied Informatics and Computing*.
- Bloice, M., & Holzinger, A. (2016). A tutorial on machine learning and data science tools with Python. *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*.
- Chapman, P. (2019). The CRISP-DM Process Model. *no. C*.
- Datapeaker. (2023). Una explicación simple de la agrupación en clústeres de K-medias y sus ventajas. <https://datapeaker.com/big-data/una-explicacion-simple-de-la-agrupacion-en-clusteres-de-k-medias-y-sus-ventajas/>
- Dharni, C., & Bnasal, M. (2013). An improvement of DBSCAN Algorithm to analyze cluster for large datasets. *2013 IEEE international conference in MOOC, innovation and technology in education (MITE)*, 42-46.
- Ejecutiva, E. (2021). El sistema ERP por excelencia de las empresas: SAP. *We-Education*.
<https://we-educacion.com/blog-sap>

- Espinosa, J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, investigación y tecnología*.
- García, G. (201). Aplicación de la metodología CRISP-DM a la recolección y análisis de datos georreferenciados desde Twitter. https://doi.org/10.48713/10336_42568
- Haya, P. (2021). Metodología CRISP-DM - Ciencia de Datos. <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>
- Hernández, J. (2024). Modelo de segmentación para crear estrategias de publicidad y venta en el sector inmobiliario empleando análisis de datos. https://doi.org/10.48713/10336_42568
- Himavanth, S. (2019, 30 de septiembre). *Service management in SAP with customer service module (CS)*. <https://community.sap.com/t5/enterprise-resource-planning-blogs-by-sap/service-management-in-sap-with-customer-service-module-cs/ba-p/13418539>
- Kang, M. (2016). Understanding K-Means clustering algorithm. *Journal of Clustering and Data Mining*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c0dcb61a41a7f9aeb8f5bf85eed7583ea6d2c77d>
- Kuchiki, A. (2008). The flowchart model of cluster policy: the automobile industry clusters in China. *International Journal of Human Resources Development and Management*, 8(1-2), 63-95.
- Lowndes, J., Best, B., Scarborough, C., Afflerbach, J., Frazier, M., O'Hara, C., Jiang, N., & Halpern, B. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*. <https://doi.org/10.1038/s41559-017-0160>
- National Health Service. (2024). Angiography. <https://www.nhs.uk/conditions/angiography/>
- National Institutes of Health, N. (2022). Computed Tomography (CT). <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>
- National Institutes of Health, N. (2023). Magnetic Resonance Imaging (MRI). <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>
- PDPC. (2022). Guía básica de anonimización. <https://www.aepd.es/documento/guia-basica-anonimizacion.pdf>
- Pills, M. (2023). Clustering K-Means. <https://mlpills.dev/kmeans-clustering>
- Ponce, D., Tipantuña, C., & Espinosa, C. (2023). Analysis of Internet Traffic in Ecuador. *IEEE*. <https://doi.org/10.1109/ACCESS.2023.10313312>
- Quintero, I. (2023). Overview of our countries and business. Customer Services Siemens Healthineers.
- Radiopaedia.org. (2023). Ultrasound. <https://radiopaedia.org/articles/ultrasound-introduction>

- Reynolds, D. A., et al. (2009). Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663).
- Riess, M. (2022). Automating model management: A survey on metaheuristics for concept-drift adaptation. *Journal of Data, Information and Management*.
<https://doi.org/10.1007/s42488-022-00075-5>
- Sai, L. N., Shreya, M. S., Subudhi, A. A., Lakshmi, B. J., & Madhuri, K. B. (2017). Optimal k-means clustering method using silhouette coefficient. *IndianJournals.com*.
- Sankalana, N. (2023). K-means Clustering: Choosing Optimal K, Process, and Evaluation Methods. *Medium*. <https://medium.com/@nirmalsankalana/k-means-clustering-choosing-optimal-k-process-and-evaluation-methods-2c69377a7ee4>
- SAP. (2024). CRM. <https://www.sap.com/latinamerica/products/crm/what-is-crm.html>
- Siemens, H. (2024a). *Acerca de Siemens Healthineers*.
<https://www.siemens-healthineers.com/ec/about>
- Siemens, H. (2024b). Advanced Therapies.
<https://healthineers.sharepoint.com/sites/Advanced-Therapies>
- Siemens, H. (2024c). Angiography systems. <https://www.siemens-healthineers.com/angio>
- Siemens, H. (2024d). Atellica Sample Handler: Productivity and precision. <https://www.siemens-healthineers.com/en-us/laboratory-diagnostics/assays-by-diseases-conditions/infectious-disease-assays/infectious-disease-on-atellica/atellica-sample-handler>
- Siemens, H. (2024e). Diagnostic Imaging.
<https://healthineers.sharepoint.com/sites/Diagnostic-Imaging>
- Siemens, H. (2024f). Imaging. <https://www.siemens-healthineers.com/molecular-imaging>
- Siemens, H. (2024g). Magnetic Resonance Imaging.
<https://www.siemens-healthineers.com/magnetic-resonance-imaging>
- Siemens, H. (2024h). Mobilett XP Family.
<https://www.siemens-healthineers.com/es/radiography/mobile-x-ray/mobilett-xp-family>
- Siemens, H. (2024i). SHS Diagnostics.
<https://healthineersnam.sharepoint.com/sites/shs-diagnostics>
- Siemens, H. (2024j). TE SharePoint Site. <https://healthineers.sharepoint.com/sites/TE>
- Siemens, H. (2024k). Tomography.
<https://www.siemens-healthineers.com/computed-tomography>
- Soni, K. G., & Patel, A. (2017). Comparative analysis of K-means and K-medoids algorithm on IRIS data. *International Journal of Computational Intelligence Research*.

- Stanford University School, M. (2024). Molecular Imaging.
<https://med.stanford.edu/mips/research.html>
- Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial* (Vol. 620). Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.
- Xu, D., & Tian, Y. (2015). Comprehensive survey of clustering algorithms. *Annals of Data Science*. <http://dx.doi.org/10.1007/s40745-015-0040-1>

6 ANEXOS

En el siguiente repositorio se encuentra todos los archivos generados en el presente trabajo de titulación:

https://github.com/davidjponce/Asignacion_CSEs_MP