



PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

**FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y
CREATIVIDAD**

**MAESTRÍA EN SISTEMAS DE INFORMACIÓN MENCIÓN
CIENCIA DE DATOS**

PROYECTO DE TITULACIÓN

Tema:

Implementación de dashboard para la exploración y gestión eficiente
del talento humano de la Coordinación Zonal 6 del Ministerio de
Salud del Ecuador.

Autor:

Nivicela Arbito Jonnathan Fernando

Director:

MSc. Montero Bermudez Eduardo Jose

Quito DM, marzo 2026

Dedicatoria

A mis padres, Clemencia y Alfonso, por guiar mis pasos en cada etapa académica y por motivarme diariamente a llegar más allá de mis propios límites.

A mis abuelos maternos, “papi” Cornelio y, en especial a “mami” Blanca (†), aunque ya no está físicamente, su legado y ejemplo siguen impulsándome a crecer como profesional y como persona. Así también a mis abuelos paternos, abuelo José (†) y abuelita Mercedes, por su presencia en mi vida.

A mis hermanos maternos: Lucrecia, Luis Alberto, Oscar David y Rosa Guadalupe y a mis hermanos paternos, Alexander y David; por ser mi motivo de superación constante, espero dejar un ejemplo de que pueden alcanzar lo que se propongan, nunca se detengan por menos.

A mis amigos por su apoyo anímico y emocional, en especial a aquellas personas que me han acompañado en este viaje académico y que por diversas situaciones ya no están tan presentes en mi vida, por haber sido motivación y compañía a lo largo de este camino.

Jonnathan F. Nivicela A.

Agradecimiento

Al PhD. Eduardo Montero, por su apoyo oportuno y sin horario para lo que este trabajo de titulación requirió.

A los docentes de la maestría, por compartir su experiencia y conocimientos, permitiéndonos explorar el potencial de herramientas y algoritmos aplicados a casos reales que enriquecieron mi formación.

Al personal de la Coordinación Zonal de Salud 6, en especial a Estefany Aroca y Diana Pozo, por facilitar el acceso a la información y la gestión necesaria para identificar y solventar la problemática abordada en esta investigación.

A mis compañeros y amigos de trabajo en UCuencaTec, especialmente Belen, Williams y Renato, por su confianza, apoyo y por aportar siempre ideas valiosas desde su experiencia profesional y académica.

A las autoridades de UCuencaTec: Ximena, Mónica y Magali, por su paciencia y por el respaldo institucional que me permitió compaginar mis estudios con la labor docente.

Un agradecimiento especial al Tnlgo. Juan Fernando Verdugo, cuya asesoría y apoyo técnico en el área de programación fueron fundamentales para agilizar el desarrollo del producto documentado en este trabajo.

El autor

Resumen

El presente trabajo de titulación aborda la optimización de la gestión de talento humano en la Coordinación Zonal 6 del Ministerio de Salud del Ecuador, mediante el diseño e implementación de un dashboard operativo integral. Este sistema surge como solución a los cuellos de botella derivados del procesamiento manual de registros en hojas de cálculo con operaciones y formulas sujetas al error humano por falta de automatización y alertas. La arquitectura se fundamenta en microframeworks (Flask), contenedorización (Docker) y persistencia mediante formatos columnares (Parquet), integrando inteligencia artificial a través del algoritmo K-means para la clasificación automatizada de eventos, eliminando así el sesgo humano en la depuración de datos. Los resultados en la fase de prueba, realizada con una nómina de 170 colaboradores, demuestran una reducción drástica en el esfuerzo administrativo: operaciones que anteriormente demandaban días-hombre se ejecutan ahora en pocos minutos, reduciendo el tiempo global de procesamiento a apenas el 20% en comparación con el método tradicional. Esta herramienta, diseñada para ser escalable, no solo mejora la eficiencia técnica y el clima laboral, sino que transforma el rol del analista administrativo de un operador de datos a un tomador de decisiones estratégico con capacidad de exploración dinámica.

Tabla de contenidos

Contenido

Dedicatoria	2
Agradecimiento	3
Resumen	4
Índice de tablas.....	7
Índice de figuras.....	8
1. Introducción	9
1.1. Justificación	9
1.2. Formulación del problema	10
1.3. Contextualización del tema	11
1.4. Objetivos.....	12
1.4.1. Objetivo General	12
1.4.2. Objetivos Específicos.....	12
1.5. Preguntas de investigación.....	12
1.6. Marco Conceptual y Marco Teórico	13
1.6.1. Revisión de Literatura	13
1.6.2. Marco Teórico	15
1.6.3. Marco Conceptual	19
2. Metodología de la investigación	22
2.1. Comprensión del negocio y los datos.....	24
2.1.1. Análisis del entorno operativo	24
2.1.2. Caracterización de los datos y línea base de tiempo	24
2.1.3. Análisis Exploratorio de Datos (EDA)	25
2.2. Preparación e ingeniería de datos.....	28
2.2.1. Renombre y eliminación de variables	29
2.2.2. Regularización de los tipos de variables y su contenido.....	29
2.2.3. Creación de nuevas variables.....	29
2.2.4. Creación de nuevas estructuras (dataframes) para apoyo	30
2.3. Modelado de aprendizaje no supervisado	31
2.4. Evaluación del sistema	33
2.4.1. Estructura del sistema.....	34
2.4.2. Módulo para carga de datos sin procesar.....	35

2.4.3.	Módulo para preprocesamiento	36
2.4.4.	Módulo de procesamiento	36
2.5.	Despliegue del sistema	39
2.5.1.	Requerimientos iniciales del sistema web	40
2.5.2.	Ambiente de desarrollo.....	40
2.5.3.	Administración y trazabilidad	42
2.5.4.	Lanzamiento a producción	44
3.	Análisis de resultados	45
3.1.	Análisis de los requerimientos funcionales y de experiencia de usuario.....	45
3.1.1.	Validación de requerimientos funcionales	46
3.1.2.	Evaluación de la Experiencia de Usuario (UX).....	47
3.2.	Análisis de la arquitectura en la centralización y exploración dinámica.....	48
3.2.1.	Centralización de estructuras mensuales y acumulación histórica	49
3.2.2.	Optimización de la exploración dinámica del mes actual.....	49
3.3.	Análisis de mejora en la productividad y reducción de carga laboral.....	50
3.4.	Análisis del nivel de adaptación de los analísticas al dashboard	51
3.4.1.	Evaluación de la aceptabilidad tecnológica	51
3.4.2.	Desarrollo de competencia técnica y autonomía	52
3.4.3.	Impacto en el clima laboral y eficiencia	52
4.	Conclusiones y Recomendaciones	53
4.1.	Conclusiones	53
4.2.	Recomendaciones.....	53
4.3.	Trabajos futuros.....	53
5.	Referencias Bibliográficas	54
Anexos		57

Índice de tablas

Tabla 1: Diccionario de las variables consolidado tras entrevista y EDA.....	26
---	----

Índice de figuras

Figura 1: Resumen informativo del dataframe importado desde archivo de biométrico	25
Figura 2: Extracto de contenido del dataset.....	26
Figura 3: Salida del análisis de los valores únicos en cada variable	27
Figura 4: Esquema de los DataFrames principal, de resumen y de registros diarios.	31
Figura 5: Ejemplo del agrupamiento (clustering) de rutinas horarias individuales.....	32
Figura 6: Módulo de autenticación de usuarios y control de acceso mediante credenciales.....	34
Figura 7: Panel principal de gestión horaria con filtros de búsqueda mensual.	35
Figura 8: Módulo de carga de datos para procesamiento.	35
Figura 9: Módulo de verificación de consistencia y resumen de registros	36
Figura 10: Interfaz de auditoría horaria	37
Figura 11: Módulo de validación de consistencia estructural.....	39
Figura 12: Panel central de administración: Gestión de usuarios, roles y departamentos.....	43
Figura 13: Consola de logs: Trazabilidad de acciones de usuario y eventos del sistema	44
Figura 14: Diagrama de conectividad y despliegue del sistema en contenedores aislados	44
Figura 15: Ejemplo de visualización del resultado de clusterización	47
Figura 16: Interfaz principal de visualización de registros y modificación.....	48
Figura 17: Edición de jornada con interactividad dinámica	50

1. Introducción

1.1. Justificación

La gestión del talento humano constituye un pilar estratégico y fundamental de la gestión administrativa contemporánea, puesto que, el capital intelectual que se expresa en “talento humano”, se concibe como uno de los recursos que se multiplica con el paso del tiempo. Esta progresión ha impulsado la integración de tecnologías de información para la automatización de tareas, el análisis de datos y la optimización del desempeño, elementos que en conjunto facilitan la toma de decisiones informadas y la eficiencia en procesos críticos como la gestión de nómina y el control de asistencia. Al desplazar la carga operativa de las tareas repetitivas, las organizaciones pueden centrar sus esfuerzos en la planificación estratégica y el bienestar integral del colaborador (Paiva de Báez, 2024). En el contexto de Ecuador, esta estructura se formaliza mediante las Unidades de Administración de Talento Humano (UATH), encargadas de operativizar estas dinámicas institucionales.

En el sector salud, la Organización Mundial de la Salud (OMS) ratifica que una gestión eficiente del personal es crucial para garantizar la continuidad y calidad de los servicios sanitarios, siendo un factor determinante para el fortalecimiento de los sistemas de salud pública (Vélez et al., 2025). No obstante, la Coordinación Zonal 6 del Ministerio de Salud Pública (MSP), dada la complejidad geográfica y el volumen de profesionales que administra en las provincias de Azuay, Cañar y Morona Santiago, enfrenta desafíos críticos en el procesamiento manual de registros de asistencia y permisos. Esta dependencia de métodos tradicionales incrementa el riesgo de inconsistencias en la información y genera una alta carga administrativa. Por consiguiente, el desarrollo de un dashboard interactivo se presenta como una solución tecnológica necesaria para centralizar la información; siguiendo los principios de Few (2013), esta herramienta busca reducir la carga cognitiva del administrador mediante una organización visual que permita identificar anomalías y patrones de asistencia de manera inmediata. Esto facilita la transformación de la gestión operativa en una gestión basada en evidencia y una exploración dinámica de datos, conforme a los lineamientos de optimización de procesos internos del MSP (2022).

1.2. Formulación del problema

En el sector público ecuatoriano, regido por la Ley Orgánica del Servicio Público (LOSEP) y el Código de Trabajo, las Unidades de Administración de Talento Humano (UATH) cumplen funciones reglamentadas que se alinean a las tendencias evolutivas de la gestión moderna. Sin embargo, limitaciones presupuestarias y tecnológicas restringen la ejecución eficiente de procesos, afectando incluso el clima laboral de los mismos servidores encargados de dicha gestión. Esta realidad es crítica en el Ministerio de Salud Pública (MSP), institución que mantiene una alta exposición pública y cuya gestión administrativa repercute directamente en la operatividad del sistema sanitario nacional.

En la administración pública, la eficiencia del talento humano es el eje que garantiza el funcionamiento institucional. En el sector salud, donde la continuidad del servicio es vital, la ineficiencia en procesos administrativos y la sobrecarga laboral no solo son problemas internos, sino que pueden comprometer indirectamente la calidad de atención al ciudadano. En este contexto, la Coordinación Zonal 6 del MSP enfrenta desafíos críticos en el manejo de registros de asistencia, lo que genera una brecha entre la disponibilidad de datos y la capacidad de respuesta ante las necesidades del personal.

A nivel operativo, la UATH de la Coordinación Zonal 6 presenta una saturación en el procesamiento de información de ingreso, salida y periodos de alimentación de más de 1700 colaboradores. Esta población se distribuye en 23 unidades operativas (distritos, hospitales y centros de salud), además del componente administrativo central. Actualmente, la gestión fragmentada en hojas de cálculo no solo provoca demoras en la actualización, sino que incrementa la vulnerabilidad de la información ante errores de transcripción, limitando la capacidad de análisis para la planificación estratégica de recursos humanos.

La coexistencia de registros digitales y formularios llenados manualmente agrava la situación, derivando en inconsistencias que deben ser subsanadas por un equipo reducido de apenas 3 o 4 funcionarios. Esta estructura obliga al personal administrativo a dedicar la mayor parte de su jornada a tareas repetitivas de validación, en lugar de enfocarse en funciones de gestión del desempeño o bienestar. La carencia de un sistema integrado que centralice, visualice y permita la exploración dinámica de estos registros consolida una ineficiencia operativa que dificulta la toma de decisiones oportunas y basadas en evidencia

conforme a los lineamientos y pretensiones institucionales (Gobierno Abierto Ecuador, 2024; Ministerio de Salud Pública, 2023; MSP, 2024)

1.3. Contextualización del tema

A nivel global, la administración de las organizaciones ha experimentado un cambio de paradigma impulsado por la transformación digital. En este escenario, el Business Intelligence (BI) se ha consolidado como la disciplina eje para convertir datos crudos en activos estratégicos. Instituciones internacionales subrayan que la capacidad de una entidad para visualizar su información en tiempo real no es solo una ventaja competitiva, sino un requisito de transparencia y eficiencia operativa (Davenport & Harris, 2017). Sin embargo, para los sistemas públicos de salud, este avance enfrenta desafíos globales relacionados con la interoperabilidad de datos y la necesidad de infraestructuras que soporten el análisis a gran escala sin comprometer la seguridad de la información.

En el ámbito nacional, el Estado ecuatoriano ha promovido agendas de simplificación de trámites y eficiencia administrativa, siendo el Ministerio de Salud Pública (MSP) uno de los pilares de esta reforma. No obstante, la ejecución de estos lineamientos se ve limitada por restricciones presupuestarias persistentes (Tapia, 2023), que dificultan la adquisición de software propietario de alta gama. Esta brecha tecnológica genera una dependencia de procesos manuales que, en una institución con la exposición pública del MSP, incrementa el riesgo de inconsistencias. Ante esto, surge la "corresponsabilidad académica", donde la universidad actúa como un motor de innovación, proponiendo soluciones de desarrollo propio que sean escalables, replicables y de bajo costo para el erario nacional.

A nivel local, la Coordinación Zonal 6 del MSP administra el talento humano de las provincias de Azuay, Cañar y Morona Santiago bajo un modelo descentralizado. Esta jurisdicción implica la gestión de más de 1700 colaboradores distribuidos en 23 unidades operativas, donde la UATH aún procesa registros de asistencia de forma fragmentada en hojas de cálculo. Esta realidad micro-institucional evidencia una saturación administrativa que limita la toma de decisiones oportuna. Por lo tanto, el desarrollo de un dashboard interactivo diseñado específicamente para las necesidades de la Zona 6 se presenta como la

solución técnica necesaria para optimizar el control de personal y permitir una exploración dinámica de los datos que hoy permanecen estáticos.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar un dashboard interactivo que optimice la exploración y gestión del talento humano en la Coordinación Zonal 6 del Ministerio de Salud del Ecuador.

1.4.2. Objetivos Específicos

1. Diseñar un modelo de dashboard que integre los registros de asistencia diarios y facilite la gestión de los permisos del personal.
2. Implementar el dashboard como una solución para el procesamiento, visualización y exploración dinámica de la asistencia y permisos de los colaboradores.
3. Evaluar el impacto del dashboard en la reducción de la carga laboral vista desde el tiempo requerido para el procesamiento de datos.
4. Capacitar al personal administrativo en el uso del dashboard para su correcta implementación.

1.5. Preguntas de investigación

- OE1: Diseñar un modelo de dashboard que integre los registros de asistencia diarios y facilite la gestión de los permisos del personal.
P.I.1. ¿Cuáles son los requerimientos funcionales y de experiencia de usuario (UX) críticos para que el dashboard optimice la gestión de asistencia y permisos?
- OE2: Implementar el dashboard como una solución para el procesamiento, visualización y exploración dinámica de la asistencia y permisos de los colaboradores.
P.I.2. ¿De qué manera la arquitectura del dashboard optimiza la centralización y la exploración dinámica de los registros de asistencia?
- OE3 Evaluar el impacto del dashboard en la reducción de la carga laboral vista desde el tiempo requerido para el procesamiento de datos.

P.I.3. ¿En qué medida la implementación del dashboard optimiza la eficiencia operativa, evaluada a través de la reducción de tiempos de procesamiento y la disminución de inconsistencias o errores manuales en comparación con el manejo tradicional de hojas de cálculo?

- OE4 Capacitar al personal administrativo en el uso del dashboard para su correcta implementación.

P.I.4. ¿Cuál es el nivel de aceptabilidad y competencia técnica alcanzado por el personal tras la capacitación en el uso del dashboard?

1.6. Marco Conceptual y Marco Teórico

1.6.1. Revisión de Literatura

La evolución de las herramientas para la optimización de procesos administrativos ha transitado desde el almacenamiento pasivo en hojas de cálculo hacia sistemas activos que integran Tecnologías de la Información y Comunicación (TIC). Esta transición, enmarcada en la transformación digital, permite que la operatividad institucional alcance niveles superiores de eficiencia mediante la Inteligencia de Negocios (Business Intelligence). Según (Few, 2013), la eficacia de un sistema de información no reside únicamente en la acumulación de datos, sino en la capacidad de presentarlos de forma que el cerebro humano identifique patrones y anomalías de manera inmediata, reduciendo la carga cognitiva del administrador.

El desarrollo de dashboards interactivos se fundamenta en el BI, definido como el conjunto de estrategias que transforman datos brutos en información estructurada para la toma de decisiones. Al respecto, García-Villalón (2024), sostiene que un tablero visual permite a los responsables del área de talento humano obtener una visión macro y micro de la organización, facilitando el monitoreo de indicadores clave de desempeño (KPIs). En el ámbito del talento humano, en algunas bibliografías denominado como recurso humano, esta analítica (People Analytics) es esencial para mitigar la discrecionalidad y el error manual, promoviendo una gobernanza basada en evidencia fáctica (Cho et al., 2023).

En el mercado tecnológico, soluciones orientadas al sector salud, como las propuestas por Conexia (2025), demuestran que los datos administrativos convertidos en un ecosistema de información se transforman en activos institucionales que garantizan

decisiones ágiles y seguras. No obstante, para que un dashboard aporte valor agregado, debe cumplir con factores críticos como la estrategia de diseño basada en la experiencia de usuario (UX), el cumplimiento de estándares de gobernanza de datos, la validación continua y, fundamentalmente, la capacitación del personal para fomentar una cultura de datos orientada al autoservicio informativo.

En el contexto ecuatoriano, el Ministerio de Salud Pública (2023) impulsó la Agenda Digital de Salud 2023-2027, orientada a modernizar la gestión administrativa y hospitalaria. Aunque dicha política enfrentó desafíos de continuidad tras el cambio de Gobierno en 2024, sus lineamientos técnicos siguen siendo un referente sobre cómo la digitalización optimiza la gobernanza interna y la eficiencia en el uso de los recursos públicos. Esta iniciativa, que involucró a actores como CEDIA, FUND-API y el MSP (Gobierno Abierto Ecuador, 2024)., subraya que la colaboración entre el sector académico y público es la vía idónea para desarrollar herramientas que, además de ser técnicamente robustas, respondan a la realidad presupuestaria y operativa de las instituciones nacionales.

En el ámbito del desarrollo de software para la gestión pública, la arquitectura de sistemas basados en entorno web ha demostrado ser la solución más eficiente para instituciones con estructuras geográficamente distribuidas. Según Pressman & Maxim (2020), las aplicaciones web facilitan la centralización de datos en servidores únicos, permitiendo que la información de múltiples unidades operativas, como las 23 sedes de la Coordinación Zonal 6, sea procesada y visualizada de forma síncrona sin los riesgos de duplicidad inherentes a las hojas de cálculo locales. Esta arquitectura no solo optimiza el acceso, sino que garantiza la integridad de los registros de asistencia al eliminar la fragmentación del dato.

De manera complementaria, la efectividad de un dashboard interactivo depende de la aplicación de principios de visualización de datos que reduzcan el ruido informativo. Knaflitz (2015) sostiene que una comunicación visual efectiva requiere eliminar elementos innecesarios para que el usuario pueda enfocarse en los hallazgos críticos. En el caso del talento humano del sector salud, donde la carga administrativa es elevada, el uso de técnicas de diseño que resalten automáticamente las inconsistencias en permisos o faltas de asistencia permite una “gestión por excepción”. Este enfoque técnico asegura que el

dashboard no sea solo una herramienta de consulta, sino un motor de eficiencia operativa que impacta directamente en la reducción de los tiempos de respuesta institucional.

1.6.2. Marco Teórico

Sistemas de Información (SI)

Un Sistema de Información (SI) se define técnicamente como un conjunto de componentes interrelacionados que recolectan, procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control en una organización. Según Laudon & Laudon (2020), la arquitectura de un SI se fundamenta en un ciclo de retroalimentación compuesto por tres actividades básicas: la entrada, que consiste en la captura de datos brutos (como los registros biométricos de la Coordinación Zonal 6); el procesamiento, que transforma estos datos en una forma con sentido; y la salida, que transfiere la información procesada a las personas o actividades que la utilizarán. En este contexto, la retroalimentación actúa como la salida que se devuelve a los miembros adecuados de la organización para ayudarles a evaluar o corregir la etapa de entrada o el procesamiento, garantizando la integridad del sistema administrativo.

Dentro de la taxonomía de los SI, el presente proyecto se categoriza como un Sistema de Soporte a la Decisión (DSS por sus siglas en inglés). Los DSS son sistemas informáticos interactivos que ayudan a los encargados de la toma de decisiones a utilizar datos y modelos para resolver problemas no estructurados. De acuerdo con Turban et al. (2020), un DSS en la gestión pública permite reducir la ambigüedad operativa al proporcionar herramientas analíticas que transforman grandes volúmenes de datos transaccionales en conocimiento estratégico. En el ámbito del talento humano, un DSS actúa como un motor de eficiencia que permite a los analistas de la UATH trascender el procesamiento manual de hojas de cálculo, enfocándose en la interpretación de patrones de asistencia y la optimización de la jornada laboral bajo criterios técnicos y normativos. Entre sus características se destacan la interactividad, facilidad de uso, el análisis de escenarios potenciales y la integración orientada al apoyo a decisiones para la reducción de errores.

People Analytics y toma de decisiones

La Analítica de Personas, comúnmente denominada People Analytics, se define como el uso de métodos estadísticos y de investigación para comprender y optimizar el talento humano dentro de las organizaciones. De acuerdo con Marler & Boudreau (2017), esta disciplina no se limita a la acumulación de métricas, sino que busca integrar datos de diversas fuentes para generar insights que impacten directamente en la estrategia institucional. En el sector público, donde la normativa y el presupuesto imponen restricciones estrictas, People Analytics surge como el mecanismo para transitar de una gestión basada en la intuición o la costumbre administrativa, hacia una toma de decisiones basada en evidencia fáctica y documentada.

El desarrollo de esta capacidad analítica suele estructurarse a través de modelos de madurez. Según el modelo de Josh Bersin (atwork, 2026) y de acuerdo a lo destacado por Król & Zdonek (2020), las organizaciones atraviesan cuatro niveles:

- Reportes operativos: Descripción básica de qué sucedió, como puede ser el registro fuera de tiempo o con inconsistencia.
- Reportes avanzados: Identificación de tendencias y comparativas proactivas.
- Análisis estratégico: Uso de modelos estadísticos para entender las causas de los comportamientos.
- Análisis predictivo: Uso de algoritmos para anticipar escenarios futuros.

El presente sistema de información permite a la Coordinación Zonal 6 superar el primer nivel de madurez, donde el procesamiento manual de hojas de cálculo consume la capacidad analítica del personal, para posicionarse en un nivel donde la limpieza y segmentación automatizada de datos mediante técnicas como el clustering, facilita un análisis diagnóstico preciso. Este salto tecnológico es lo que Pfeffer & Sutton (2006) denominan "gestión basada en la evidencia", la cual reduce los sesgos cognitivos y los errores de interpretación en la administración del talento humano, garantizando que cada registro de asistencia sea una unidad de información verificable y útil para la planificación institucional.

Gestión estratégica en el sector de la salud

La gestión de las instituciones de salud pública contemporáneas se enmarca en la doctrina de la Nueva Gerencia Pública (NGP), la cual promueve la adopción de prácticas del sector privado para mejorar la eficiencia, la transparencia y el rendimiento en el servicio público. De acuerdo con Barzelay (2001), la NGP enfatiza el control de resultados y la desagregación de unidades operativas para facilitar una administración más ágil. En el caso del Ministerio de Salud Pública de Ecuador, la descentralización en Coordinaciones Zonales exige que la gestión administrativa no sea un obstáculo, sino un habilitador de la continuidad del servicio. Un sistema de información que optimiza el control del talento humano reduce la "carga burocrática" y permite que los recursos institucionales se enfoquen en la misión primaria: la atención al ciudadano (Frenk & Moon, 2013).

La eficiencia operativa en este sector depende críticamente de la integridad de los datos. Según Pollitt & Bouckaert (2017), la gestión pública moderna se basa en el desempeño, donde la precisión de la información administrativa, como el cumplimiento de jornadas laborales, es el insumo básico para la planificación financiera y la auditoría de procesos. En la Coordinación Zonal 6, la integridad del dato no solo asegura la transparencia en el uso de fondos públicos destinados a nómina, sino que garantiza que el personal operativo y administrativo se encuentre en sus puestos de trabajo según la planificación de turnos. Un fallo en la veracidad de estos registros impacta directamente en la capacidad de respuesta institucional, por lo que la transición de hojas de cálculo manuales a sistemas automatizados robustos representa una salvaguarda contra la discrecionalidad y el error humano en la administración pública.

Ciencia de la visualización de datos y la percepción

La visualización de datos no se limita a la representación gráfica de cifras, sino que constituye una disciplina técnica orientada a ampliar la capacidad cognitiva del usuario mediante el uso de representaciones visuales abstractas. Según Tufte (2001), la excelencia en un sistema de información visual radica en comunicar ideas complejas con claridad, precisión y, sobre todo, eficiencia. Para lograrlo, el diseño debe maximizar la "proporción dato-tinta", eliminando cualquier elemento decorativo que no aporte información relevante. En el dashboard desarrollado para la Coordinación Zonal 6, este principio se traduce en una

interfaz minimalista donde los indicadores de tiempo descontado y los estados de validación se resaltan sobre el fondo, permitiendo que el analista de la UATH identifique anomalías sin distracciones visuales.

La efectividad de un tablero de control (dashboard) depende de la comprensión de los principios de la Gestalt y la gestión de la carga cognitiva. De acuerdo con Ware (2013), el cerebro humano procesa ciertas características visuales, como el color y la posición, de forma "pre-atentiva", es decir, antes de que el usuario realice un esfuerzo consciente. Al aplicar leyes de la Gestalt como la proximidad (agrupar registros por semanas) y la semejanza (usar códigos de colores para estados "Válido" o "Atraso"), el sistema facilita la detección inmediata de patrones.

Es así que, una arquitectura visual del sistema debe alinearse con los principios de Few (2013), quien sostiene que un dashboard debe permitir el monitoreo de un vistazo. Al integrar notificaciones de tiempo real y comparativas de jornada (entrada/salida), el sistema reduce la carga cognitiva intrínseca del proceso de auditoría de asistencia. Esto transforma una tarea que anteriormente requería un alto esfuerzo de memoria y atención en las hojas de cálculo manuales, en un proceso de reconocimiento visual rápido y preciso, sustentado en la jerarquía visual y la organización lógica de los componentes de la interfaz.

Interactividad en dashboards para la gestión de datos

Un dashboard se define como una interfaz visual que integra y organiza la información crítica de una organización en una sola pantalla, permitiendo su monitoreo y análisis de un vistazo. Según Eckerson (2011), los dashboards se clasifican en tres categorías principales según su propósito: estratégicos, enfocados en KPIs a largo plazo; analíticos, orientados a la exploración de tendencias y causas; y operativos, diseñados para el seguimiento de procesos en tiempo real y la intervención inmediata. El sistema propuesto para la Coordinación Zonal 6 se alinea con la categoría de dashboard operativo, ya que su fin primordial es la validación diaria de registros y la corrección de anomalías en las jornadas laborales.

La evolución de estas herramientas ha permitido transitar de tableros estáticos a dashboards interactivos con capacidad de edición. De acuerdo con Shneiderman (1996), la interacción efectiva en sistemas de información sigue el mantra: "primero una visión

general, luego filtrar y hacer zoom, y finalmente obtener detalles bajo demanda". En este sistema propuesto, la interactividad no se limitaría a la filtración de datos, sino que incorpora la capacidad de corrección y validación de registros directamente desde la interfaz. Esta funcionalidad permitiría que el analista de la UATH no solo visualice las inconsistencias detectadas y apoyadas por algoritmos, sino que actúe sobre ellas, modificando o confirmando marcas horarias para asegurar la integridad de la nómina. Esta transición hacia dashboards transaccionales reduce la brecha entre el análisis de datos y la ejecución administrativa, convirtiendo al sistema en una herramienta de gestión activa y no solo de consulta pasiva. Cairo (2013) sostiene que la visualización es una "herramienta tecnológica" antes que una "obra de arte".

1.6.3. Marco Conceptual

La arquitectura del sistema de información propuesto se fundamenta en un modelo de capas que separa la lógica del servidor, la persistencia de datos y la interfaz de usuario, garantizando un entorno escalable y de fácil mantenimiento.

Arquitectura de software y ecosistema de desarrollo

En la arquitectura del lado del servidor (backend), se implementa Flask, un microframework basado en Python diseñado bajo una filosofía de desarrollo ágil y modular. A diferencia de los marcos de trabajo monolíticos que imponen una estructura rígida, Flask se define como una infraestructura que proporciona únicamente las herramientas esenciales para la gestión de rutas y peticiones HTTP, eliminando la sobrecarga de componentes innecesarios que podrían comprometer el rendimiento del sistema.

En esta propuesta, Flask facilita la función de un simple servidor web para actuar como el núcleo de orquestación técnica. Su diseño permite una integración nativa y directa con librerías de análisis de datos y motores de procesamiento, facilitando que el sistema reciba archivos de marcaciones biométricas y administre funciones complejas directamente desde la lógica de Python. Según Grinberg (2018), la naturaleza extensible es ideal para proyectos que requieren un alto grado de personalización en la lógica de negocio. Esta flexibilidad es la que permite el despliegue eficiente del motor de procesamiento de marcaciones

biométricas, garantizando que el procesamiento de registros no sea un proceso aislado sino una extensión fluida de la infraestructura de comunicación del sistema.

De acuerdo con Merkel (2014), la contenedorización elimina el conflicto de "dependencias rotas" y facilita proceso de despliegue (deployment) continuo. Para este proyecto en particular, el uso de contenedores garantiza la réplica absoluta del entorno, asegurando que la arquitectura de la base de datos y el entorno de ejecución de Python sean idénticos tanto en la fase de desarrollo local como en los servidores de producción de la Coordinación Zonal 6, este enfoque mitiga los errores de compatibilidad tradicionales y establece una arquitectura de microservicios aislada, donde el dashboard y el motor de procesamiento funcionan de manera estable y segura.

Para asegurar la portabilidad y consistencia del sistema en servidores institucionales se ha seleccionado a Docker, el que se define como una tecnología de aislamiento que permite encapsular el software, sus bibliotecas y la totalidad de sus dependencias en una unidad estándar, ligera e independiente del sistema operativo anfitrión. En el desarrollo del sistema de gestión horaria, esta tecnología es fundamental para asegurar la portabilidad y la consistencia operativa en los servidores institucionales, permitiendo empaquetar no solo el código fuente, sino también las variables de entorno y configuraciones específicas del sistema.

La interactividad y la respuesta inmediata de la interfaz de usuario (frontend) se fundamentan en la dinámica del cliente, gestionada a través del lenguaje de programación JavaScript. Esta tecnología se refiere a la capacidad de procesamiento directamente en el navegador, lo que permite mejorar la interactividad sin depender de constantes peticiones al servidor que podrían ralentizar el flujo de trabajo. Mediante el uso de objetos JSON, se establece un puente de comunicación reactiva que facilita al usuario la edición de registros y la validación de estados de forma instantánea, un aspecto esencial para la auditoría técnica de registros de asistencia.

Según Flanagan (2020), el uso de JavaScript en el lado del cliente optimiza la experiencia del usuario al reducir la latencia de respuesta. En el contexto de este sistema, dicha optimización resulta ser un factor crítico para la eficiencia administrativa; asegura que el analista de la UATH mantenga un ritmo de trabajo fluido y sin interrupciones cognitivas al validar y auditar miles de registros de forma consecutiva.

Ingeniería de datos y persistencia

La ingeniería de datos constituye la base operativa de cualquier ecosistema analítico, definiéndose como el conjunto de procesos para la recolección, transformación y entrega de datos para su posterior análisis (DAMA, 2017). En el contexto de la gestión del talento humano, este componente asegura que la información administrativa sea tratada bajo estándares de integridad y disponibilidad. Para la integración de las fuentes de datos institucionales, se implementa un flujo de Extracción, Transformación y Carga (ETL). Según Kimball y Ross (2013), la fase de transformación es crítica, ya que permite la limpieza de datos, el manejo de valores nulos y la estandarización de atributos, garantizando que el dashboard final refleje una "versión única de la verdad". En proyectos de administración pública, esto reduce significativamente los errores manuales y la duplicidad de registros.

La persistencia se refiere al almacenamiento no volátil de la información procesada. Para este proyecto, se opta por una arquitectura de almacenamiento relacional, la cual permite mantener la integridad referencial de los datos de los funcionarios mediante el cumplimiento de las propiedades ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad) (Date, 2004). Esta estructura facilita consultas complejas y asegura que la trazabilidad de los movimientos de personal se mantenga histórica y técnicamente auditable.

Ciclo de Salida y Usabilidad

La usabilidad en un entorno visual, dashboard, a través de un sistema de información busca que sea diseñado para sintetizar información compleja en un formato de lectura simple y directo, un tablero efectivo debe permitir la asimilación de datos críticos "de un solo vistazo" para que el usuario actúe rápidamente con un mínimo esfuerzo cognitivo.

En este sistema, el tablero interactivo facilita la supervisión humana del algoritmo, permitiendo que el auditor identifique de forma inmediata registros marcados bajo los estados de "Inconsistentes" o "Por Validar", enfocando la gestión de excepciones,

garantizando que la validación final siempre cuente con el criterio y respaldo del auditor administrativo antes de su almacenamiento definitivo.

La serialización y reportabilidad se define como el proceso de codificación de estructuras de datos internas hacia formatos de intercambio estándar y persistentes. Según el estándar IEEE 610.12, la interoperabilidad es la capacidad de dos o más sistemas o componentes para intercambiar información y utilizar de forma efectiva los datos que han sido intercambiados.

Dentro del sistema esta capacidad se traduce en la exportación de los resultados del motor analítico a archivos .xlsx, garantizando la interoperabilidad administrativa. Este estándar asegura que la información generada no sea un dato aislado, sino que pueda integrarse con otros sistemas manteniendo una trazabilidad íntegra desde la captura del registro en el dispositivo físico hasta la generación del reporte final consolidado.

Inteligencia Artificial en el procesamiento de información

La integración de técnicas de Inteligencia Artificial permite transitar de un análisis descriptivo básico hacia un análisis de diagnóstico y descubrimiento de patrones. En el sistema propuesto, la IA actúa como el motor analítico para la clasificación y el tratamiento avanzado de registros administrativos. Uno de sus áreas es el Machine Learning (ML) donde se cuenta con algoritmos de clasificación de aprendizaje no supervisado como k-means

El algoritmo K-means es un método de particionamiento heurístico que busca dividir un conjunto de n observaciones en k grupos, donde cada observación pertenece al clúster con la media más cercana. Según MacQueen (1967), este proceso es iterativo y se fundamenta en la actualización constante de los centroides hasta alcanzar la convergencia, minimizando la varianza intra-clúster. Matemáticamente, el objetivo es minimizar la suma de los cuadrados dentro de cada clúster (WCSS).

2. Metodología de la investigación

Para el desarrollo de la presente investigación, se adopta el estándar CRISP-DM (Cross-Industry Standard Process for Data Mining), el cual proporciona un marco estructurado y cíclico ideal para la creación de sistemas de información orientados a la analítica. En las fases de comprensión del negocio y de los datos, se analizan los aspectos

relacionados con las jornadas laborales y la estructura operativa de la UATH de la Coordinación Zonal 6, identificando inconsistencias en los registros de asistencia en una muestra de alrededor de un centenar de colaboradores que luego permitirá escalar a los más de 1700 colaboradores distribuidos en diversas unidades operativas. Estos registros mensuales, obtenidos desde dispositivos biométricos y almacenados originalmente en hojas de cálculo, presentan desafíos de integridad y volumetría que justifican una intervención tecnológica avanzada.

La fase de preparación (ingeniería) de datos consiste en la limpieza, normalización y transformación de los registros crudos para construir un repositorio estructurado. En esta etapa, se emplean DataFrames para la manipulación en memoria y el formato de almacenamiento Parquet para optimizar la velocidad de procesamiento y asegurar la persistencia eficiente de la información. Posteriormente, en el modelado, se implementa un algoritmo de aprendizaje no supervisado (K-means) sobre las marcas temporales. Esta técnica permite la clasificación automatizada de eventos en categorías de entrada, salida y periodos de alimentación, reduciendo drásticamente el error humano asociado a la selección y etiquetado manual de datos.

La fase de evaluación se centra en la mejora continua del desarrollo para optimizar la visualización e interacción con los registros, garantizando la integridad en el formato de fechas, horas y la generación de notificaciones en tiempo real. Esta etapa se orienta en alcanzar la reducción de los tiempos de respuesta frente al procesamiento tradicional (línea base mediante hojas de cálculo), mediante implementaciones para el procesamiento automatizado ofertado por el sistema. Por último, en la fase de despliegue, se integra este motor analítico en una aplicación web desarrollada con Flask y contenedorizada mediante Docker. El resultado propuesto es un dashboard interactivo que permite la exploración dinámica de los datos y la exportación de reportes validados, asegurando una toma de decisiones institucional oportuna y una emisión de informes basada estrictamente en evidencia técnica.

2.1. Comprensión del negocio y los datos

2.1.1. Análisis del entorno operativo

El proceso de gestión de asistencia en la Coordinación Zonal 6 se origina con la captura de eventos mediante relojes biométricos distribuidos en las unidades operativas. La muestra seleccionada para este estudio comprende los registros de más de un centenar de colaboradores del área administrativa, cuya gestión se considera el modelo base para escalar el sistema a las demás dependencias. Mensualmente, el departamento de Tecnologías de la Información y Comunicación (TIC) extrae las marcaciones crudas en formato .xls (hojas de cálculo), las cuales carecen de preprocesamiento o validación lógica previa.

Cada mes, el personal de la Unidad de Administración de Talento Humano (UATH) ejecuta una labor de cotejamiento manual exhaustiva. Este procedimiento implica el uso de herramientas ofimáticas tradicionales y calendarios físicos para verificar la puntualidad y la integridad de las marcas. Se ha identificado que la eficiencia de este proceso se ve severamente degradada por dos factores críticos: configuraciones ambiguas en la lógica de los dispositivos biométricos y una cultura de registro inconsistente por parte de los colaboradores. Estas condiciones derivan en tres tipologías de errores persistentes:

- Redundancia de registros: Marcaciones duplicadas con intervalos de pocos minutos que generan ruido en el cálculo de horas laboradas.
- Inconsistencia de etiquetas: Registros horarios cuya etiqueta de sistema (ej. "Entrada") no corresponde con la ventana temporal real de la marcación.
- Fragmentación de la jornada: Secuencias incompletas que rompen el esquema institucional de cuatro marcas diarias (Entrada, Inicio de Almuerzo, Fin de Almuerzo y Salida).

2.1.2. Caracterización de los datos y línea base de tiempo

La diversidad de jornadas laborales y la carencia de una base de datos unificada exacerbaban la latencia en el procesamiento de la información. Según el diagnóstico realizado con el personal de la UATH, se estima que la consolidación manual de los registros de un solo dispositivo biométrico demanda aproximadamente 10 días laborales (80 horas hombre) para un analista. Para mitigar este cuello de botella, la institución distribuye la carga entre

cuatro analistas, quienes operan en paralelo para intentar cumplir con los plazos de cierre de nómina, lo que evidencia una alta dependencia de procesos manuales repetitivos y un elevado riesgo de error humano.

2.1.3. Análisis Exploratorio de Datos (EDA)

Mediante el uso de Google Colab y librerías especializadas en Python, se realizó un Análisis Exploratorio de Datos sobre la muestra inicial para cuantificar los hallazgos previos. Este análisis permitió identificar patrones de distribución temporal y la dispersión de las marcaciones, sentando las bases para la lógica de clustering posterior. Partiendo de un registro exportado del mes de septiembre 2024, se destacan los siguientes hallazgos principales.

Estructura de las columnas

Como se puede observar en la Figura 1 la hoja de cálculo contiene 10 variables para 11579 registros, cada uno corresponde a una marcación de un horario; al importar el archivo a un dataframe, algunas columnas se interpretan como tipo objeto y otras como enteros o flotantes, lo cual no corresponde correctamente en algunos casos, además se puede notar que ‘CardNo’ está completamente llena de datos nulos mientras que ‘No. Cédula’ tiene parcialmente una cantidad de datos nulos.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11579 entries, 0 to 11578
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Dpto.           11579 non-null  object
1   Nombre          11579 non-null  object
2   AC_No           11579 non-null  int64
3   Fecha/Hora      11579 non-null  object
4   Marc-Ent/Sal    11579 non-null  object
5   Reloj ID        11579 non-null  int64
6   No. Cédula      9914 non-null   float64
7   Incidencia      11579 non-null  int64
8   Verificación    11579 non-null  object
9   CardNo          0 non-null      float64
dtypes: float64(2), int64(3), object(5)
memory usage: 904.7+ KB
```

Figura 1: Resumen informativo del dataframe importado desde archivo de biométrico

Cabeceras de las columnas

Cada variable tiene su etiqueta en la cabecera como se observa en la Figura 2, sin embargo, las mismas tienen caracteres especiales (punto, tildes, espacios, slash) que pueden generar conflicto en la preparación y procesamiento de los datos, por lo cual se requerirá cambiar la denominación de las columnas.

df.head()

	Dpto.	Nombre	AC_No	Fecha/Hora	Marc-Ent/Sal	Reloj ID	No. Cédula	Incidencia	Verificación	CardNo
0	REPCONTVER	MINGA DURAN DIEGO	233	03/09/2024 14:26:25	SalFuera	1	NaN	0	FP	NaN
1	REPCONTVER		317	02/09/2024 10:18:09	M/Ent	1	NaN	0	FP	NaN
2	REPCONTVER		317	02/09/2024 12:32:48	Sal	1	NaN	0	FP	NaN
3	REPCONTVER		317	02/09/2024 12:56:41	Sal	1	NaN	0	FP	NaN
4	REPCONTVER		317	02/09/2024 17:04:25	M/Sal	1	NaN	0	FP	NaN

Figura 2: Extracto de contenido del dataset

Diccionario de variables

La hoja de cálculo no cuenta con un diccionario que especifique el contenido de cada columna o el tipo de dato que se almacena, sin embargo, luego de entrevistas con el personal de UATH y de TICs se ha comprendido lo que se detalla en la siguiente tabla. Además, se propone el nuevo nombre que se dará a la variable para su procesamiento siguiendo la estructura camelCase.

Tabla 1: Diccionario de las variables consolidado tras entrevista y EDA.

Variable	Descripción	Nuevo nombre
Dpto.	Una cadena de texto que corresponde al departamento al cual pertenece el colaborador registrado en el campo 'Nombre'.	departamento
Nombre	Una cadena de texto que contiene el identificador del colaborador.	nombre
AC_No	Un código numérico que puede ser considerado como cadena o entero asignado a cada colaborador.	numAcceso
Fecha/Hora	Una marca de tiempo bajo la estructura dd/mm/yyyy hh:mm:ss correspondiente a la marcación realizada.	fecha_hora
Marc-Ent/Sal	Una cadena de texto en redacción contraída que refiere al evento al que corresponde la marca de tiempo del campo anterior.	tipoMarcacion
Reloj ID	Un identificador del reloj biométrico.	relojID
No. Cédula	Identificador de número de cédula de ciudadanía bajo el formato de Ecuador.	cedula
Incidencia	Una bandera que indica si hubo algún inconveniente o particularidad en la marcación.	incidencia
Verificación	Una cadena de texto de palabra única que refiere a la forma	verificacion

	como fue registrada la marcación, se ha identificado que puede ser por huella digital, reconocimiento facial o tarjeta.	
CardNo	Código que se almacena en el reloj biométrico en el caso de que la marcación se haya realizado mediante el método de verificación empleando una tarjeta.	numTarjeta

Cardinalidad de las variables

Con los nuevos nombres de columnas, como preparación para procesamiento e ingeniería de datos, se ha revisado la cardinalidad en cada variable, visualizando la salida de la **Figura 3**, donde se puede notar la existencia de 15 departamentos, 174 nombres de colaboradores que corresponde uno a uno a los 174 valores que toma numAcceso, se infiere que a cada colaborador se le asigna un número, sin embargo solo hay 146 cédulas lo que significa que algunos colaboradores no tienen asignado un valor en la variable cédula. Aparecen 4 tipo de marcaciones, siendo lo esperado una de entrada, una de salida y dos correspondientes al almuerzo o receso. El método de verificación registra 3 valores diferentes mientras que las variables incidencia e ID de reloj tienen un valor único cada uno, y el número de tarjeta no tiene valores; estas tres últimas variables se considerarían para descarte.

```
for column in df.columns:
    unique_count = df[column].nunique()
    print(f"Column '{column}': {unique_count} unique values")

Column 'departamento': 15 unique values
Column 'nombre': 174 unique values
Column 'numAcceso': 174 unique values
Column 'fecha_hora': 11579 unique values
Column 'tipoMarcacion': 4 unique values
Column 'relojID': 1 unique values
Column 'cedula': 146 unique values
Column 'incidencia': 1 unique values
Column 'verificacion': 3 unique values
Column 'numTarjeta': 0 unique values
```

Figura 3: Salida del análisis de los valores únicos en cada variable

Particularidades en las variables de interés

Al continuar con el EDA, se ha revisado los valores únicos en cada variable, especialmente en las variables de interés que serán gestionadas para los registros de asistencia y marcaciones, de lo cual:

- La variable “fecha_hora” tiene un formato de objeto, similar a una cadena, lo que deberá ser considerado en la ingeniería de datos, además resulta pertinente separar en dos variables, fecha y hora para facilitar el filtrado.
- La variable “verificación”, al revisar sus valores únicos, se obtiene la lista ['FP', 'Face', 'FACE'], ante lo cual surge la necesidad de llevar a mayúsculas para homogenizar los valores, armonizando con la característica que las otras variables llevan sus valores en mayúsculas.
- La variable de mayor interés y principal tema de estudio en esta investigación es “tipoMarcacion”, de la cual, los valores únicos que aparecen en el dataset inicialmente analizado son ['M/Sal', 'M/Ent', 'SalFuera', 'Sal'], a ello se le agrega un quinto valor encontrado en registros de otro mes ['Sal Hrs Ext'], ante lo que, no se puede dar una interpretación coherente ni correcta. En entrevista, los analistas de la UATH y de TICs difieren en indicar que ['M/Sal', 'M/Ent'] refieren a la salida y entrada del mediodía por un lado y por otro que pueden indicar Salida y Entrada de la jornada laboral, lo que se contrapone con los valores ['SalFuera', 'Sal'] que parecen sinónimos. Es en este punto donde los analistas fundamentan la necesidad de la analítica de esta investigación, bajo la premisa que, tales inconsistencias generan el trabajo exhaustivo y demorado para consolidar las nóminas de asistencia.
- El resto de variables cumplen con características informativas e identificativas que permitirán estructurar y generar un reporte de salida útil para otros sistemas o procesos, este sistema no se encargará de corregir inconsistencia como nombres incompletos o nombres que no siguen el formato de dos nombres y dos apellidos, no realizará validaciones de cédulas en cuanto a números de dígitos o lógica de los mismos, puestos que el sistema solo se apoya de las variables para filtros y estructura.

2.2. Preparación e ingeniería de datos

El proceso de ingeniería de datos constituye la etapa crítica donde los registros iniciales provenientes de los dispositivos biométricos se transforman en información estructurada y apta para el análisis algorítmico. Esta fase asegura que el motor de

procesamiento trabaje sobre datos limpios, consistentes y normalizados, permitiendo una transición fluida desde el archivo.xlsx crudo hacia un ecosistema de persistencia en formato parquet.

2.2.1. Renombre y eliminación de variables

El primer paso consiste en la reestructuración de la cabecera del conjunto de datos original. Se implementa un mapeo de etiquetas para estandarizar los nombres de las columnas (ej. de Dpto. a departamento, AC_No a numAcceso), facilitando la legibilidad en las capas posteriores del sistema. Simultáneamente se realiza una limpieza de variables que no aportan valor analítico al cálculo de jornadas como relojID, incidencia y numTarjeta. Esta reducción de dimensionalidad optimiza el uso de memoria RAM y agiliza la persistencia en formato binario.

2.2.2. Regularización de los tipos de variables y su contenido

Para garantizar la integridad en las comparaciones de texto, se aplica una homogenización de caracteres, convirtiendo la columna de verificación a mayúsculas sostenidas. Esto evita duplicidad de categorías por errores de digitación (como "manual" con "MANUAL"). Asimismo, se realiza una limpieza en los campos de identificación (cédula), asegurando que los valores numéricos se presenten en un formato estándar de 10 dígitos, eliminando sufijos decimales generados incorrectamente por la lectura inicial del archivo.

2.2.3. Creación de nuevas variables

La temporalidad es el eje central del sistema partiendo del campo original de fecha y hora, se realiza una descomposición en dos nuevas variables: fecha y hora, lo cual permite agrupamientos por jornadas diarias sin perder el ancla cronológica original. Adicionalmente, se introduce la variable hora_decimal, que transforma el formato HH:MM:SS en una escala numérica continua (ej. 08:30:00 se convierte en 8.5). Esta conversión es un requisito técnico indispensable para que el algoritmo K-means pueda calcular distancias y determinar centroides con precisión matemática.

2.2.4. Creación de nuevas estructuras (dataframes) para apoyo

El sistema no solo procesa los registros individuales, sino que genera tres estructuras de datos independientes y complementarias para el soporte de la auditoría:

- DataFrame principal (df): Actúa como el libro mayor de marcaciones contiene la trazabilidad completa de cada registro individual. Además de los datos originales depurados se inyectan columnas de metadatos inicializadas como objetos nulos (NaN):
 - Columnas: departamento, nombre, numAcceso, fecha_hora, tipoMarcacion, cedula, verificacion, fecha, hora, marcacionCluster, marcacionReal, tipoValidacion.
 - Propósito: La columna marcacionCluster guarda el etiquetado automático del motor, mientras que marcacionReal y tipoValidacion están diseñadas para capturar la intervención humana permitiendo que el sistema priorice las correcciones del auditor sobre el cálculo algorítmico.
- DataFrame de para resumen (df_resumen): Centraliza los perfiles analíticos de los funcionarios para el mes procesado.
 - Columnas: funcionario, díasAsistidos, tendEntrada, tendSalidaAl, tendEntradaAl, tendSalida, entradaOficial, salidaOficial, recesoOficial, atrasoTotal, grafica.
 - Propósito: Facilita la carga de perfiles en el dashboard y aloja las tendencias calculadas y la representación visual de la jornada (codificada en Base64), optimizandopara tener un registro inicial mediante el grafico.
- DataFrame para registros diarios (df_diario): Estructura orientada a la métrica por jornada laboral.
 - Columnas: funcionario, fecha, numMarcacionesInicial, numMarcacionesFinal, atrasoEntrada, atrasoAlmuerzo, atrasoSalida.
 - Propósito: Esencial para detectar inconsistencias compara el número de marcas detectadas originalmente frente a las validadas tras la auditoría, y registra las penalizaciones de tiempo específicas de cada día.

La Figura 4 representa gráficamente el esquema detallado de estas tres estructuras de datos y la organización de sus columnas respectivas, sirviendo como mapa visual para la implementación técnica.

DataFrame principal	DataFrame resumen	DataFrame registros diarios
departamento	funcionario	funcionario
nombre	diasAsistidos	fecha
numAcceso	tendEntrada	numMarcacionesInicial
fecha_hora	tendSalidaAl	numMarcacionesFinal
tipoMarcacion	tendEntradaAl	atrasoEntrada
cedula	tendSalida	atrasoAlmuerzo
verificacion	entradaOficial	atrasoSalida
fecha	salidaOficial	
hora	recesoOficial	
marcacionCluster	atrasoTotal	
marcacionReal	grafica	
tipoValidacion		

Figura 4: Esquema de los DataFrames principal, de resumen y de registros diarios.

2.3. Modelado de aprendizaje no supervisado

Para la interpretación de los patrones de asistencia, el sistema emplea un motor de Machine Learning basado en el algoritmo de agrupamiento K-means. Este modelo tiene la capacidad de aprender la rutina de cada funcionario de manera individual, adaptándose a sus horarios habituales de entrada y salida sin necesidad de reglas predefinidas.

Procesamiento de clústeres y visualización

La única variable de entrada para el modelado es la hora_decimal. A diferencia de modelos estáticos el número de clústeres se determina de forma dinámica para cada funcionario utilizando la moda de las marcaciones diarias. Esto permite que el algoritmo reconozca automáticamente si un empleado cumple una jornada simple (2 marcas) o completa (4 marcas).

Para la supervisión humana el sistema genera diagramas de dispersión donde el eje horizontal representa la fecha y el vertical la hora decimal. En estos gráficos, los centroides resultantes se trazan como líneas de tendencia, permitiendo al auditor visualizar

rápidamente la "franja horaria" de mayor concentración de registros de un colaborador. Estas imágenes se codifican en cadenas Base64 para su renderizado directo en la interfaz web.

La **Figura 5** ilustra un ejemplo de esta visualización, mostrando cómo el algoritmo agrupa los registros diarios y proyecta las líneas de tendencia central (centroides) para un funcionario con jornada de cuatro marcaciones.

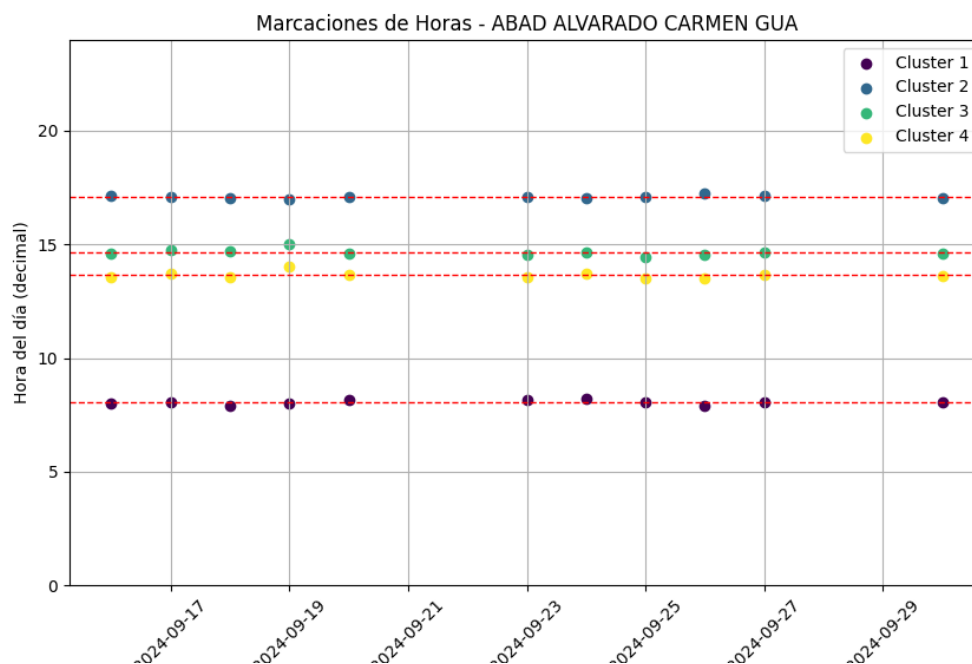


Figura 5: Ejemplo del agrupamiento (clustering) de rutinas horarias individuales.

Etiquetado condicional y asignación de tendencias

Una vez que el algoritmo ha agrupado los registros, el sistema ejecuta una lógica de interpretación cronológica sobre los centroides resultantes. Este proceso transforma las coordenadas numéricas en categorías administrativas funcionales:

- Asignación de etiquetas por jerarquía temporal: Los centroides se ordenan de menor a mayor en la línea del tiempo. El valor mínimo se etiqueta automáticamente como "Entrada" y el máximo como "Salida". En escenarios donde el clúster dinámico detecta cuatro grupos, los centroides intermedios se asignan como "Salida Almuerzo" y "Regreso Almuerzo" respectivamente. En casos excepcionales de una

sola marcación, el sistema aplica un umbral (12:00 PM) para discriminar el tipo de evento.

- Ajuste a la matriz institucional: El sistema extrae la "hora de tendencia" (el centro exacto del clúster) y realiza una búsqueda de proximidad mediante el cálculo del error mínimo dentro de una matriz de horarios institucionales permitidos (ej. 07:00, 07:30, 08:00). Este proceso genera una Hora Oficial Sugerida, proporcionando al auditor una base normativa para comparar el comportamiento histórico del funcionario frente a su contrato laboral.

Esta etapa final permite que el sistema no solo calcule, sino que proponga estados de validación de forma autónoma. La arquitectura de visualización en el dashboard se rige por una jerarquía de prioridades que garantiza la integridad de la auditoría:

- Prioridad de la intervención humana (marcacionReal): Si existe un registro en esta columna, el sistema asume que un auditor administrativo ha revisado y corregido el dato. Esta etiqueta tiene precedencia absoluta sobre cualquier cálculo algorítmico, mostrándose en el dashboard como un registro "Válido" o "Actualizado".
- Sustento del motor analítico (marcacionCluster): En ausencia de correcciones manuales el sistema proyecta la etiqueta generada por el algoritmo. Si el comportamiento diario del funcionario se alinea con el patrón de clústeres predominante del mes el registro se marca como "Por Validar", simplificando la tarea del auditor al confirmar comportamientos regulares.
- Detección autónoma de anomalías ("Inconsistentes"): El sistema identifica automáticamente como registros "Inconsistentes" aquellos días donde el número de marcas es impar o se detectan tipos de eventos duplicados. Esta lógica obliga a la intervención humana, asegurando que cualquier desviación del patrón habitual sea analizada críticamente antes de su almacenamiento definitivo en los archivos parquet.

2.4. Evaluación del sistema

La evaluación del sistema parte del cumplimiento y requerimientos funcionales establecidos en el punto 2.2, enfocándose en la creación de una interfaz interactiva

haciendo uso de una paleta de colores acordes a logos institucionales, transformando la transición y manejo del sistema con diseños estables y escalables. El objetivo principal es proporcionar al usuario un entorno de automatización que optimice la gestión de marcaciones.

2.4.1. Estructura del sistema

Como primer pilar de seguridad y control de acceso, se implementó un módulo de autenticación (Login) que garantiza la integridad de la información y restringe el acceso únicamente a personal autorizado del MSP visto en la Figura 6. Una vez superada esta barrera de seguridad, el sistema despliega un panel principal Dashboard diseñado bajo principios de usabilidad, el cual articula las secciones estratégicas del proyecto:

- Registros de funcionarios: Permite la administración y consulta del historial mensual de marcaciones de forma individual.
- Carga de registros: Facilita la abstracción masiva de datos crudos mediante una interfaz asistida y automatizada.
- Exportación de resultados: Permite obtener la información consolidada y procesada en formatos compatibles (xlsx) para su análisis externo.

Este diseño estructural establece el entorno de desarrollo necesario para que los procesos analíticos descritos en la Figura 7, operen de manera transparente para el usuario final, transformando tareas manuales complejas en un flujo de trabajo automatizado y eficiente.

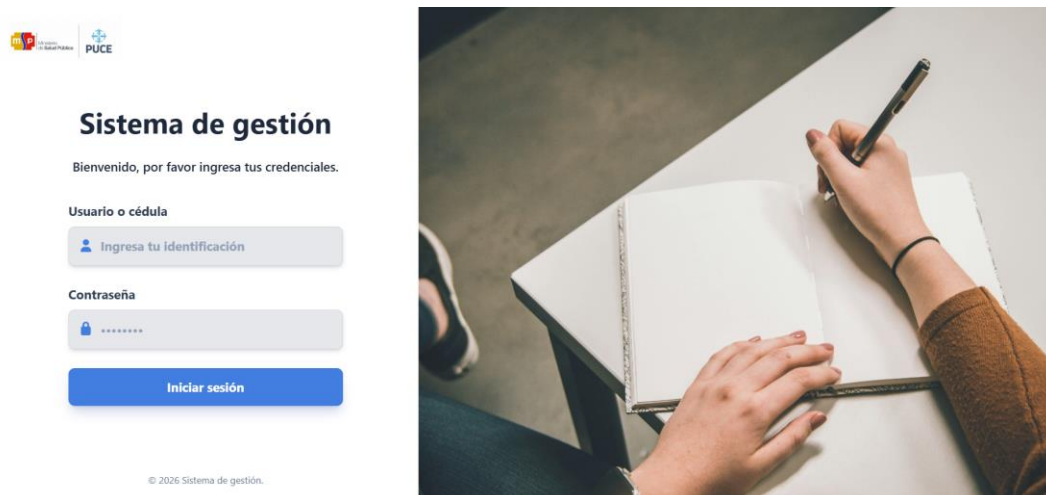


Figura 6: Módulo de autenticación de usuarios y control de acceso mediante credenciales.

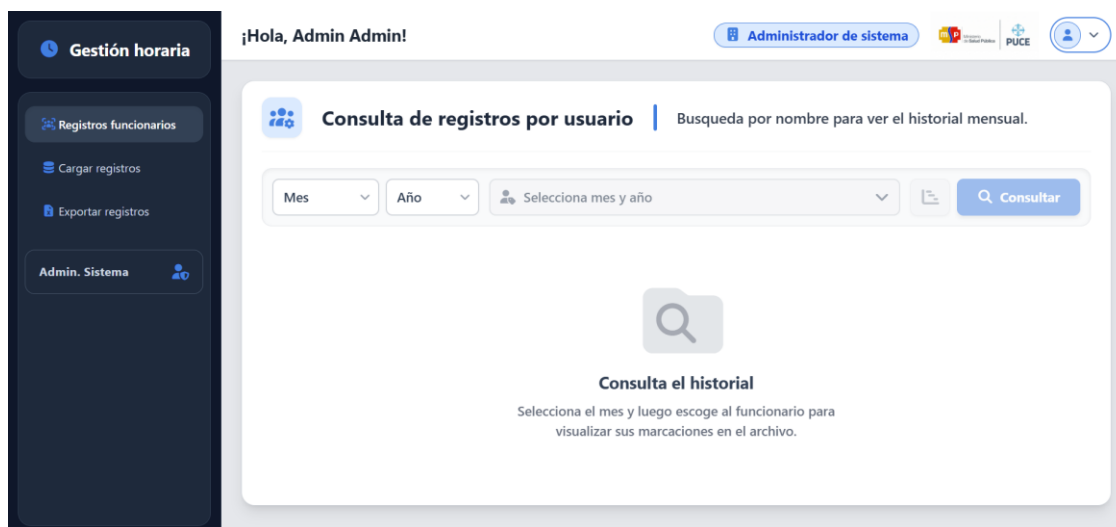


Figura 7: Panel principal de gestión horaria con filtros de búsqueda mensual.

2.4.2. Módulo para carga de datos sin procesar

Este módulo fue evaluado bajo protocolos de integridad de archivos. Se propone el manejo de datos mediante archivos en formato .xlsx por ser un estándar moderno, estable y compatible con diversas herramientas de análisis. Se implementó una lógica de validación que actúa como filtro de seguridad, el sistema analiza las cabeceras del archivo cargado y las contrasta con el diccionario de datos definido en la sección 2.1.3. A partir de este punto, los datos se diversifican en columnas para permitir su procesamiento posterior para que el usuario administre únicamente la carga del archivo siguiendo la estructura planteada como se aprecia en la **Figura 8**.



Figura 8: Módulo de carga de datos para procesamiento.

2.4.3. Módulo para preprocesamiento

La evaluación de este módulo confirmó la correcta ejecución de las reglas de limpieza programadas. Se validó que el motor de preprocesamiento elimine de forma automática los caracteres especiales detectados durante el EDA. Este proceso garantiza que los datos que alimentan al Dashboard sean consistentes. Asimismo, permite que el usuario habilitado verifique dicha consistencia antes de proceder a la validación, guardado, clusterización y reestructuración de inconsistencias, incluyendo la agregación de las nuevas cabeceras detalladas en la sección 2.2 como se puede ver en la Figura 9, posterior a la carga del archivo de prueba.



Figura 9: Módulo de verificación de consistencia y resumen de registros

2.4.4. Módulo de procesamiento

Partiendo de las etapas previas se establece el punto de ejecución principal del motor del sistema el cual opera con datos ya filtrados y supervisados. Esta etapa consolida los procesos internos y define el guardado dentro de los tres dataframes utilizados para el procesamiento en memoria aquí se ejecuta la predicción de datos y el algoritmo K-means, agrupando la información en columnas estructuradas para facilitar tareas de edición como la agregación y modificación de registros en el modelado de tipos de marcaciones.

Dentro de este proceso se contempla el guardado de estados de validación para el sistema interactivo mediante nuevos encabezados, se asigna a cada registro el estado de "Válido" o "Actualizado". El manejo de la información se rige a estándares de integridad para asegurar que no se alteren los registros existentes. En caso de nuevas entradas, estas se marcan como "Manual"; si se modifica una predicción, el registro transita de "Por validar" a "Válido" o "Actualizado".

Es fundamental señalar que bajo ninguna circunstancia se altera el registro horario de los métodos de autenticación físicos (FACE o FP). Finalmente, siguiendo las normas de manejo de datos los registros nunca son eliminados por completo en su lugar, se emplea un método de descarte para ignorar la duplicidad de registros marcada por los propios funcionarios, como se puede apreciar en la Figura 10, donde al hacer una consulta dentro de la base de datos alojada dentro de los archivos parquet se puede consultar y validar los registros.

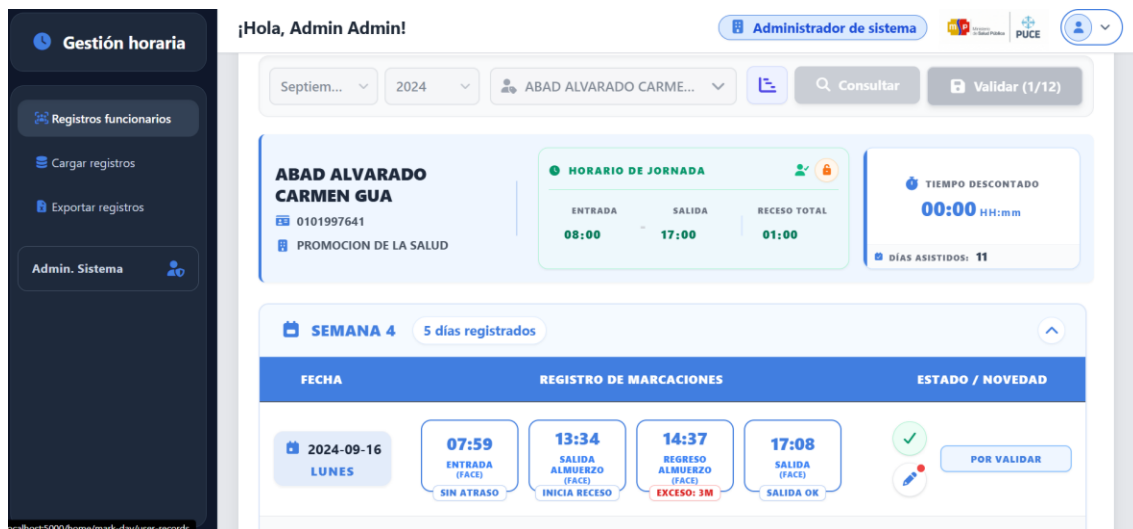


Figura 10: Interfaz de auditoría horaria

Como etapa final de este módulo, se implementó una interfaz de gestión y auditoría reactiva que permite al usuario interactuar con los resultados del motor de análisis. Esta fase es crucial ya que permite la transición de datos "predichos" por el algoritmo a datos validados por el criterio humano, asegurando la máxima consistencia antes del almacenamiento definitivo.

Validación y edición interactiva de registros

El sistema dispone de un entorno de edición seguro que opera bajo una lógica de aislamiento de memoria. Cuando un auditor accede a la jornada de un funcionario el sistema genera una copia temporal de los registros esto garantiza que cualquier modificación solo se aplique a la base de datos tras superar un estricto protocolo de validaciones lógicas, evitando la corrupción de la información original durante el proceso de edición.

Dentro de este entorno, se destacan los siguientes componentes funcionales:

- **Cálculo automatizado de tiempos:** El sistema ejecuta algoritmos en tiempo real para convertir las marcaciones físicas en métricas de cumplimiento mediante funciones matemáticas de sustracción de tiempos (formato HH:mm) se calcula automáticamente el atraso en entrada, exceso en receso y salidas anticipadas. Estos cálculos se contrastan con el horario oficial del MSP cargado en el sistema por

usuario dentro de sus jornadas laborales, eliminando el error humano en la contabilidad de minutos descontados.

- Validación de consistencia estructural: Para que un registro pase al estado de "Guardado", el sistema obliga al cumplimiento de reglas de negocio automatizadas:
 - Paridad de registros: Una jornada debe poseer un número par de marcaciones (parejas de entrada y salida).
 - Orden cronológico: Se bloquea el guardado si se detectan incoherencias temporales, como una salida registrada antes de una entrada o un regreso de almuerzo previo a la salida del mismo.
 - Unicidad: Se prohíbe la existencia de registros duplicados en hora o tipo dentro de la misma jornada.
- Auditoría de estados: El usuario tiene la facultad de validar los procesos ya predichos por la computadora. Los registros transitan de un estado de "Por Validar" (estado inicial tras el proceso de clustering) a "Válido" o "Actualizado". Además, se implementó un método de descarte lógico (papelera), donde los registros inconsistentes no se eliminan físicamente, sino que se marcan como "Descartados" para mantener la trazabilidad histórica de la información recolectada por los dispositivos físicos.

HORA	TIPO DE MARCACIÓN	MÉTODO	ESTADO	ACCIÓN
07:59	Entrada	FACE	POR VALIDAR	✓ ✖
13:34	Salida Almuerzo	FACE	POR VALIDAR	✓ ✖
14:37	Regreso Almuerzo	FACE	POR VALIDAR	✓ ✖
17:08	Salida	FACE	POR VALIDAR	✓ ✖

Cancelar Guardar en memoria

Figura 11: Módulo de validación de consistencia estructural

Consolidación y almacenamiento en memoria

Una vez que el auditor confirma la validez de la jornada, el sistema consolida la información en tres dataframes específicos que actúan como base de datos para futuras consultas. Este método de guardado permite que el sistema interactivo sea altamente eficiente, ya que las consultas posteriores no requieren volver a ejecutar el motor analítico completo, sino que acceden directamente a los datos ya supervisados y estructurados. Finalmente, esta información se empaqueta y se envía al servidor para actualizar los archivos de persistencia alojados dentro de los parquet " (mes) - (año) .parquet", "resumen- (mes) - (año) .parquet" y "diario- (mes) - (año) .parquet", garantizando que el reporte mensual refleje con exactitud la realidad laboral del funcionario.

2.5. Despliegue del sistema

El despliegue del sistema se ejecuta bajo un modelo de contenedorización, con el objetivo de garantizar la portabilidad y facilitar la ejecución tanto en servidores locales como en la nube o en estaciones de trabajo distintas sin conflictos de dependencias. Para lograr esto se utilizó Docker y Docker Compose, permitiendo orquestar los servicios de manera aislada y reproducible que separa la lógica del sistema con la de gestión de datos.

2.5.1. Requerimientos iniciales del sistema web

Para la base estructural del sistema, se optó por un entorno multi-contenedor que separa la lógica de negocio de la persistencia de datos segmentando las responsabilidades del sistema:

- **Motor de base de datos:** Se implementó una instancia de MySQL 8.0.36 gestionada mediante un contenedor dedicado (db-sistema). Este componente está pensado para escalabilidad y gestiona el esquema de sesiones, roles y departamentos. Se estableció un sistema de seguridad basado en el cifrado de credenciales, asegurando que las contraseñas de los usuarios no se almacenen en texto plano, cumpliendo con estándares de integridad.
- **Gestión de privilegios:** El sistema establece una jerarquía de acceso donde el usuario con rol de Administrador, vinculado al departamento de "Administrador de Sistema" posee facultades totales para la creación de nuevos departamentos, gestión

de funcionarios y reseteo de credenciales. Se incluyó un mecanismo de protección para el perfil administrador principal garantizando que los datos base de seguridad no sean alterados accidentalmente.

- Persistencia de datos: Se definieron volúmenes gestionados (`vol-app`) y (`vol-data`) que vinculan el almacenamiento físico del host con el contenedor. Esto garantiza que la información de la base de datos y los archivos generados permanezcan intactos incluso si los contenedores son reiniciados o eliminados.

2.5.2. Ambiente de desarrollo

El entorno de desarrollo se construyó sobre Python 3.11- slim-bookworm, utilizando una imagen optimizada y herramientas como Visual Studio Code, Docker Desktop y MySQL Workbench, la arquitectura de ficheros sigue un diseño modular que facilita la escalabilidad y el mantenimiento:

- Organización de archivos: El proyecto se divide en carpetas especializadas: `controllers` para la lógica, `db` para la gestión de la base de datos, `routers` para las rutas de la API y `static` (CSS, JS, imágenes) junto a `templates` para el frontend. Esta separación de responsabilidades permite integrar nuevas funcionalidades sin comprometer la estabilidad del sistema existente.
- Gestión de dependencias: El archivo `Dockerfile` automatiza la instalación de las librerías establecidas en el archivo `requirements.txt`, garantizando un entorno idéntico en cualquier máquina de despliegue.
- Infraestructura y motores web: El núcleo del servidor se apoya en el ecosistema de Flask para la orquestación de la plataforma:
 - Flask: Actúa como el micro-framework principal, proporcionando el motor de enrutamiento y la gestión de peticiones HTTP. Durante el desarrollo se utilizó su servidor nativo con modo `debug` activo para la depuración en tiempo real.
 - Jinja2: Utilizado como motor de plantillas para el renderizado dinámico de HTML. Permite la inyección directa de variables de Python en la interfaz, facilitando la creación de vistas personalizadas según el rol del usuario.

- Serialización JSON (Python Standard Library): Pieza crítica para la comunicación asíncrona dado que Python opera internamente con diccionarios y listas, se emplea el módulo json para la conversión de estas estructuras de datos en objetos JSON. Esto permite que el backend "hable" el mismo lenguaje que el frontend (Alpine.js), asegurando una transferencia de información íntegra y ligera.
- Librerías de Procesamiento de datos: El núcleo analítico del sistema se apoya en librerías de alto rendimiento:
 - Pandas y Openpyxl: Utilizadas para la abstracción, limpieza y transformación de datos provenientes de archivos .xlsx.
 - Numpy: Fundamental para la ejecución del algoritmo K-means, el cual procesa las inconsistencias en las marcaciones mediante agrupamiento estadístico
 - Pyarrow: Empleada para el manejo eficiente de archivos en formato Parquet, optimizando el almacenamiento y la velocidad de lectura.
- Interfaz Interactiva y Frontend: El desarrollo visual se basó en Tailwind CSS para un diseño moderno y profesional. La interactividad de los modales, las animaciones de carga y los componentes dinámicos se gestionaron con Alpine.js, lo que permite una comunicación reactiva mediante los objetos JSON previamente serializados en el servidor, evitando recargas innecesarias de la página y optimizando la experiencia del usuario.
- Iconografía: Se integró la biblioteca Font Awesome (versión gratuita) para proporcionar una guía visual intuitiva al usuario mediante iconos estandarizados en botones, menús y alertas de estado.
- Gestión de Activos (Assets): Es relevante destacar que la implementación de estas herramientas se realizó prescindiendo de gestores de paquetes externos como npm o procesos de compilación complejos. En su lugar, se optó por la descarga y gestión local de los archivos .css y .js optimizados desde sus respectivos CDNs. Esta decisión técnica garantiza que el sistema sea completamente autónomo y funcional en entornos de red local o servidores sin acceso a internet, reforzando el principio de portabilidad y facilidad de despliegue dentro del contenedor Docker.

Nota sobre el aprovisionamiento del entorno

Es importante precisar que se requiere de una conexión activa a internet de manera exclusiva durante la fase inicial de construcción (build) y aprovisionamiento del sistema. Este requisito es indispensable para la descarga de las imágenes base desde Docker Hub (Python y MySQL), la instalación de las dependencias de sistema mediante el gestor de paquetes de Linux y la obtención de las librerías especializadas definidas en el archivo requirements.txt.

Una vez completado este proceso de orquestación e instalación el sistema queda configurado dentro del contenedor permitiendo su ejecución y operación en entornos de red local completamente aislados (offline).

2.5.3. Administración y trazabilidad

La estrategia de lanzamiento y mantenimiento a largo plazo asegura que el sistema sea apto para un entorno de producción real:

- Panel de administración y auditoría: Se implementó una sección exclusiva para el perfil Administrador dentro del departamento Administrador de sistema que permite gestionar departamentos y usuarios como se demuestra en la Figura 12. Un punto importante es el visor de logs, una herramienta de auditoría que ofrece trazabilidad total. Este módulo diferencia entre logs de sistema (errores técnicos) y logs de auditoría (acciones de usuarios), permitiendo al administrador visualizar hasta las últimas 200 líneas de actividad mediante una simulación de consola en tiempo real como se ve en el ejemplo de la Figura 13, estos registros se mantienen persistentes en vol-app (volumen destinado para la aplicación) desde la creación del contenedor.
- Seguridad y control de acceso: El sistema protege perfiles críticos, como el administrador principal, impidiendo la modificación accidental de sus datos base. Además, incluye funciones para el reseteo forzado de credenciales, estableciendo claves temporales que el usuario debe cambiar en su primer inicio de sesión para cumplir con protocolos de seguridad informática.

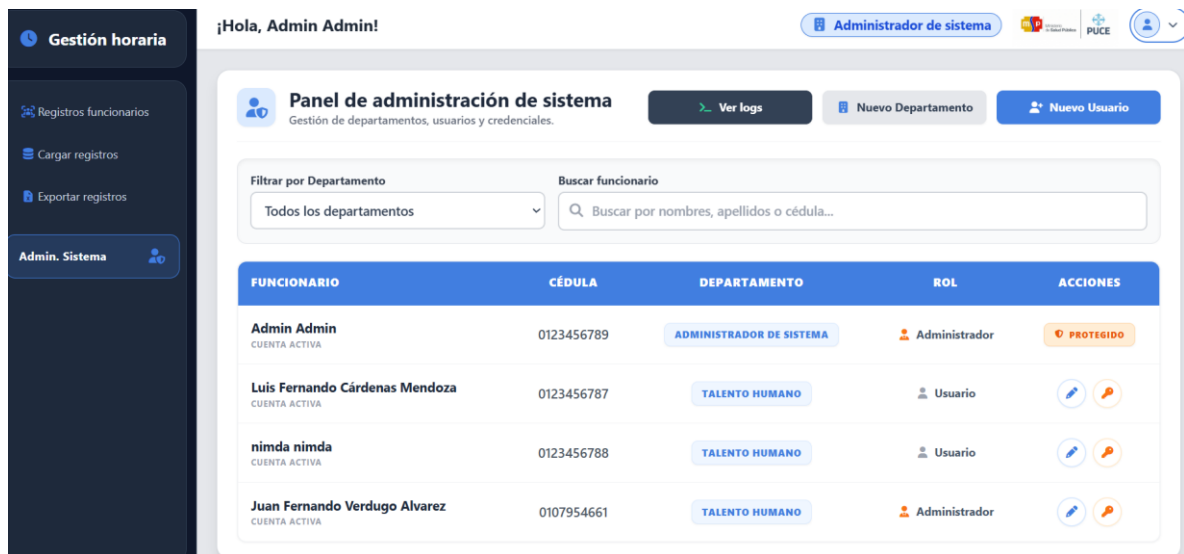


Figura 12: Panel central de administración: Gestión de usuarios, roles y departamentos.

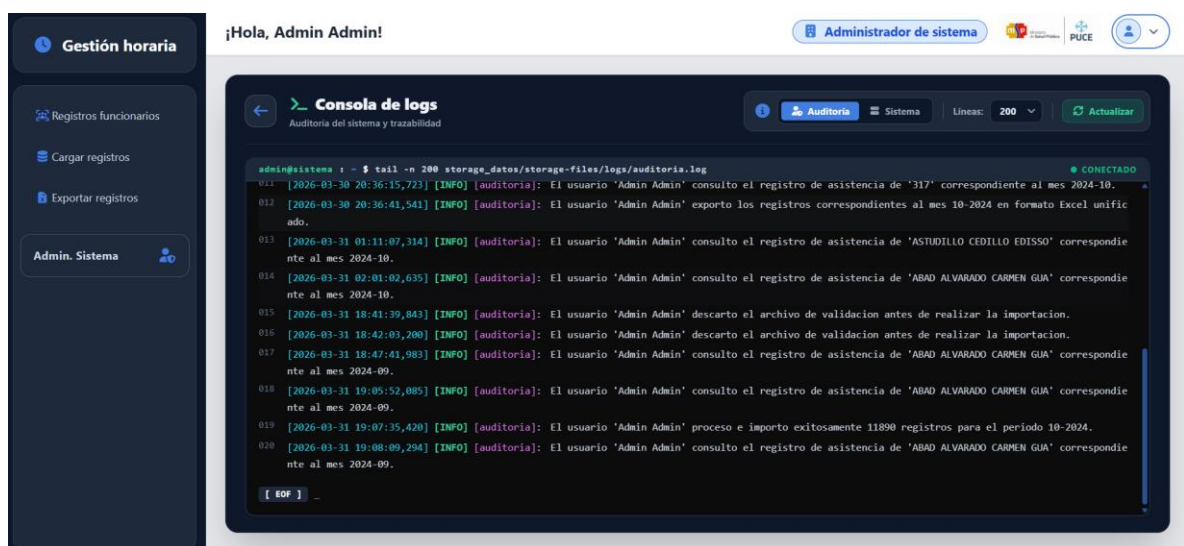


Figura 13: Consola de logs: Trazabilidad de acciones de usuario y eventos del sistema

2.5.4. Lanzamiento a producción

La estrategia de puesta en marcha y mantenimiento a largo plazo del sistema se fundamenta en una arquitectura robusta que garantiza la transición de un entorno de desarrollo a uno de producción sin comprometer la integridad de los datos ni la disponibilidad del servicio. Este proceso se articula bajo tres ejes estratégicos:

Optimización del servidor y seguridad de comunicación

Para el despliegue final se sustituye el servidor de desarrollo nativo de Flask por un servidor WSGI (Web Server Gateway Interface) de grado industrial como Gunicorn. Este ajuste técnico es crítico para manejar múltiples workers y sesiones simultáneas, garantizando la estabilidad operativa ante peticiones concurrentes de diversos funcionarios.

Esta capa de aplicación se complementa con la implementación de Nginx como Proxy Inverso, encargado de la gestión del protocolo HTTPS. Como se detalla en la Figura 14, Nginx actúa como el único punto de entrada al sistema realizando la terminación SSL y protegiendo el servidor de aplicaciones de exposiciones directas a la red, cumpliendo así con los estándares modernos de seguridad informática.



Figura 14: Diagrama de conectividad y despliegue del sistema en contenedores aislados

Persistencia de datos e integridad de archivos

Dada la naturaleza del sistema (procesamiento de registros históricos), la gestión de la información no depende del ciclo de vida del contenedor. Mediante el uso de volúmenes gestionados (vol-app y vol-data) el sistema redirige el almacenamiento de documentos procesados, archivos Parquet y registros de auditoría a una ruta persistente en el host (/storage_datos). Esta configuración asegura que las copias de seguridad y la migración de archivos entre servidores sean procesos ágiles y seguros evitando la pérdida de información ante reinicios o actualizaciones del motor de software.

Portabilidad y flujo de trabajo (CI/CD)

El sistema ha sido diseñado bajo una portabilidad absoluta mediante la integración con GitHub establece una base sólida para procesos de integración continua, permitiendo el control de versiones y la trazabilidad del código. Al estar el entorno completamente dockerizado, el despliegue en cualquier servidor con soporte para Docker (Linux o Windows) se reduce a un flujo de trabajo simplificado:

1. Clonación del repositorio central.
2. Ejecución del comando de orquestación: `docker-compose up -d --build`.

Este diseño permite que el sistema pase de un estado de prueba a uno de producción en cuestión de minutos ofreciendo una escalabilidad inmediata y la capacidad de replicar el entorno de gestión en diferentes infraestructuras sin conflictos de dependencias.

3. Análisis de resultados

A continuación, se presentan los resultados tras el análisis de la implementación del dashboard al ponerlo a prueba en una etapa de despliegue, para lo cual se aborda y se da respuesta a las preguntas de investigación:

3.1. Análisis de los requerimientos funcionales y de experiencia de usuario

- **P.I.1. ¿Cuáles son los requerimientos funcionales y de experiencia de usuario (UX) críticos para que el dashboard optimice la gestión de asistencia y permisos?**

El análisis de los resultados obtenidos tras la implementación del sistema permite determinar que la optimización de la gestión de asistencia no depende únicamente de la captura de datos, sino de la eficiencia con la que estos son procesados y presentados.

3.1.1. Validación de requerimientos funcionales

A partir de las pruebas de carga y procesamiento realizadas sobre los registros administrativos, se identificaron tres pilares funcionales críticos:

- **Eficiencia en el procesamiento masivo:** La implementación de dataframes y el almacenamiento en formato parquet, así como los algoritmos de pre procesamiento y procesamiento de los datos ha evidenciado la capacidad del sistema para manejar registros concurrentes sin pérdida de integridad, es decir que se ha podido procesar lotes de alrededor de 170 colaboradores con más de 11000 registros en cada hoja de cálculo, habiendo probado con diversos archivos de entrada (diferentes meses) que siguen una estructura similar en sus variables pero que difieren en la consistencia de su contenido.
- **Automatización de la clasificación:** El uso del algoritmo K-means resolvió la ambigüedad en los registros manuales. Al contrastar los resultados, se determinó que la lógica de IA categorizó correctamente al menos 3 de 4 clusters

correspondientes a los eventos de asistencia, cumpliendo con el requerimiento de reducir la intervención humana en la depuración de datos. Según la revisión manual luego efectuada por el analista de talento humano, 1 de cada 10 colaboradores necesita una revisión detenida, en la Figura 15 se puede ver el resultado gráfico que también puede visualizar el analista.

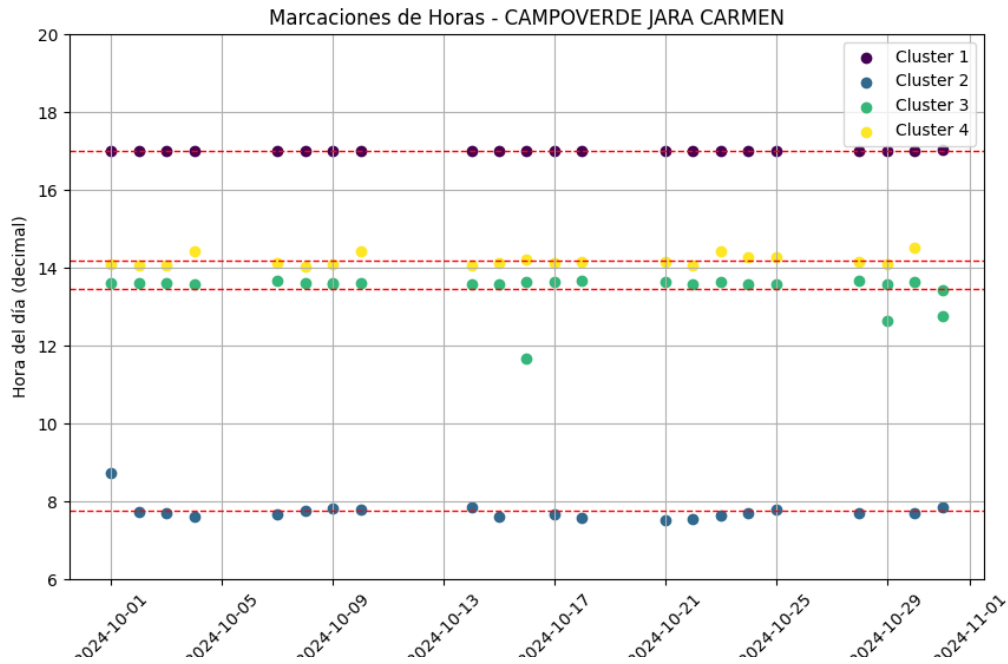


Figura 15: Ejemplo de visualización del resultado de clusterización

- Portabilidad y despliegue: Gracias a la contenedorización con Docker, se validó que el entorno de ejecución es idéntico tanto en desarrollo como en el servidor local institucional, eliminando errores de compatibilidad de librerías.

3.1.2. Evaluación de la Experiencia de Usuario (UX)

Para medir la efectividad del dashboard, se aplicaron criterios de usabilidad basados en el modelo SUS (System Usability Scale) o métricas de observación según la Figura 16:

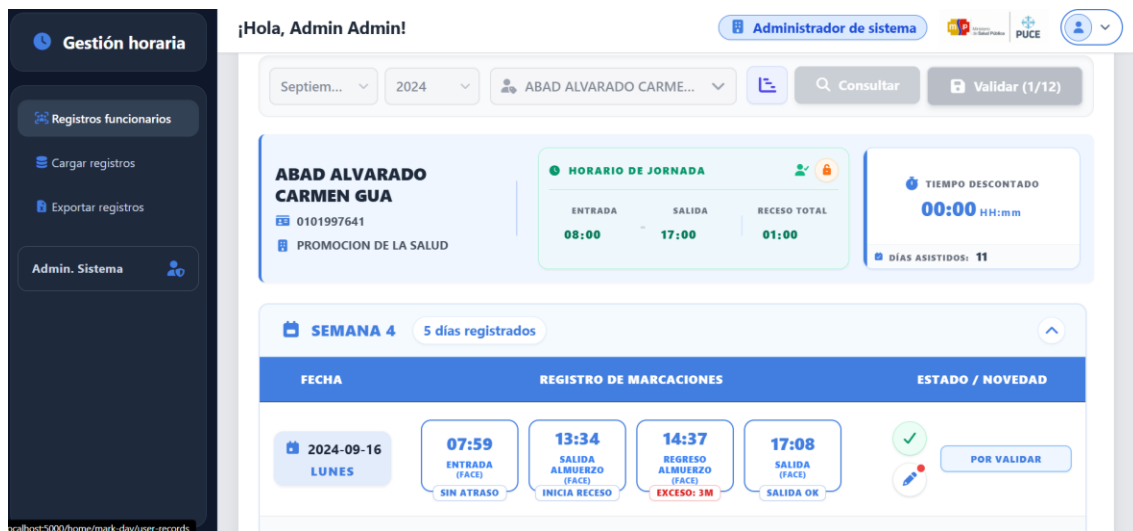


Figura 16: Interfaz principal de visualización de registros y modificación

- Claridad visual y jerarquía: La interfaz desarrollada permitió una navegación intuitiva. Los controladores mostrados en la Figura 16 evidencia el uso de filtros dinámicos en la parte superior para seleccionar el mes de análisis y el colaboradores, lo cual permite a los analistas de talento humano pasar rápidamente de revisar un colaborador a poder revisar de otro en menos de tres interacciones.
- Velocidad de percepción: El tiempo de carga de las visualizaciones interactivas se mantuvo por debajo de los 2 segundos, con un equipo de escritorio o laptop de recursos mínimo que se emplea en oficina, lo cual es crítico para evitar el abandono del usuario y asegurar que el dashboard así como el sistema sea una herramienta de trabajo diaria y no una carga administrativa adicional.

3.2. Análisis de la arquitectura en la centralización y exploración dinámica

- **P.I.2. ¿De qué manera la arquitectura del dashboard optimiza la centralización y la exploración dinámica de los registros de asistencia?**

La arquitectura del sistema no solo funciona como un contenedor de datos, sino como un facilitador de agilidad administrativa. El apartado de análisis permite identificar cómo la combinación de componentes optimizó la gestión de los registros de asistencia.

3.2.1. Centralización de estructuras mensuales y acumulación histórica

Tradicionalmente, la información de asistencia se gestiona en hojas de cálculo que contienen la información de un mes calendario. En el flujo manual previo, acumular varios meses para un análisis de tendencias implicaba procesos de copiado y pegado propensos al error humano.

- **Ingesta modular:** El sistema permite cargar las hojas de cálculo mensuales de manera independiente. El motor basado en Flask y Pandas extrae la información y la unifica en una estructura de DataFrame común, eliminando la fragmentación de la información.
- **Persistencia eficiente:** Al utilizar el formato Parquet, el sistema no solo guarda los datos del mes actual, sino que permite la acumulación de registros históricos en archivos individuales sin degradar el rendimiento y de manera segmentada para que se exporte a herramientas externas o complementarias. Esto permite que el sistema almacena y permita regresar a revisar meses anteriores, facilitando el acceso inmediato sin tener que abrir un archivo sino a través de un filtro, lo que antes eran técnicamente imposibles para los analistas en tiempos cortos, además que se asegura el almacenamiento y exportación.

3.2.2. Optimización de la exploración dinámica del mes actual

El análisis de resultados confirma que la arquitectura está diseñada para priorizar el mes de interés del analista, permitiendo una exploración dinámica que el Excel no ofrece por sí solo:

- **Segmentación selectiva (Slicing):** Aunque el sistema puede acumular grandes volúmenes de datos, la lógica de programación permite al analista aislar instantáneamente el mes que desea auditar. Esto significa que el procesamiento se enfoca en los 170 registros de prueba del mes vigente, garantizando que el dashboard sea ágil y no se sature con información histórica irrelevante para ese momento.
- **Interactividad dinámica:** A través de la interfaz a modo de dashboard, el usuario puede cambiar entre el mes de interés a un mes anteriormente procesado sin necesidad de cargar nuevos archivos. Esta capacidad histórica

permite indagar o comparar manualmente tendencias o patrones, por ejemplo, si un patrón de atrasos identificado en el mes actual es una anomalía puntual o una conducta recurrente de meses pasados. Además, esta interactividad dinámica permite acceder a los registros para cada día donde se puede agregar, eliminar o corregir aquellas marcaciones que por la clusterización no corresponden o que por inconsistencias no detectadas computacionalmente necesitan la revisión y asistencia del analista, tal como se puede ver en la Figura 17.

HORA	TIPO DE MARCACIÓN	MÉTODO	ESTADO	ACCIÓN
07:59	Entrada	FACE	POR VALIDAR	✓
13:34	Salida Almuerzo	FACE	POR VALIDAR	✓
14:37	Regreso Almuerzo	FACE	POR VALIDAR	✓
17:08	Salida	FACE	POR VALIDAR	✓

Figura 17: Edición de jornada con interactividad dinámica

3.3. Análisis de mejora en la productividad y reducción de carga laboral

P.I.3. ¿En qué medida la implementación del dashboard optimiza la eficiencia operativa, evaluada a través de la reducción de tiempos de procesamiento y la disminución de inconsistencias o errores manuales en comparación con el manejo tradicional de hojas de cálculo?

Durante las pruebas, se procesó varios conjuntos de datos reales, cada uno con alrededor de 170 colaboradores dando seguimiento a la validación y comprobación de los resultados de acuerdo a la ingeniería de datos y observando los resultados del motor de clusterización, ante lo cual, los resultados, vistos desde los tiempos labores para su procesamiento, muestran:

- Reducción de tiempos: Mientras que el método manual para 170 registros requeriría proporcionalmente 6.8 días-hombre, considerando que la dedicación y concentración se interrumpía por otras actividades propias del analista, así como jornadas de receso. En contraste, el sistema actual completó la ingesta, clasificación y generación de visualizaciones en un tiempo de ejecución técnico inferior a los 15 minutos, mientras que la corrección de marcaciones con inconsistencias o que la clusterización asignó en un grupo incorrecto, no toma más allá de 1 día-hombre.
- Optimización del talento humano: La automatización permite que los analistas puedan confiar en su toma de decisiones, basado en evidencias, eliminando el error humano en la fase de clasificación de eventos o perceptiva, con ayudas visuales sobre la clasificación de los eventos de marcaciones. La implementación del dashboard dentro de un sistema de información permitió una transición desde la lectura de tablas o archivos planos con operaciones manuales, hacia la percepción visual de patrones, como inconsistencias, atrasos y notificaciones para asegurar integridad y almacenamiento de datos.

3.4. Análisis del nivel de adaptación de los analísticas al dashboard

- **P.I.4. ¿Cuál es el nivel de aceptabilidad y competencia técnica alcanzado por el personal tras la capacitación en el uso del dashboard?**

El éxito de una implementación tecnológica no reside únicamente en su robustez técnica, sino en la adopción por parte de sus usuarios finales. El análisis de los resultados de este apartado surge luego de una continua capacitación al personal, lo cual ha sido continuo más no una jornada, pues ha sido una construcción de conocimiento al receptar sus requerimientos y a la vez ajustar el dashboard con el sistema para facilitar la adaptabilidad y lograr una alta competencia técnica.

3.4.1. Evaluación de la aceptabilidad tecnológica

La aceptabilidad se midió bajo dos dimensiones críticas:

- Utilidad percibida: Los analistas manifestaron un nivel de aceptabilidad "Muy Alto", fundamentado en la eliminación de tareas repetitivas. Al contrastar que el sistema procesa en minutos lo que antes tomaba 10 días laborables a 4 personas, el

personal percibe al dashboard no como una carga adicional, sino como un aliado estratégico que libera tiempo para el análisis de casos críticos de personal.

- **Confianza en el dato:** Un punto clave de aceptabilidad fue la validación de la lógica de clusterización mediante k-means. Inicialmente, existía escepticismo sobre la clasificación automatizada; sin embargo, al verificar algunos casos, evidenciaron que los resultados coincidían con la realidad de las marcaciones que deben corresponder en la mayoría de los 170 registros de prueba, la confianza en el sistema se consolidó.

3.4.2. Desarrollo de competencia técnica y autonomía

La capacitación no solo se enfocó en "qué botones presionar", sino en elevar el nivel de competencia técnica de los funcionarios.

- **Desarrollo de destrezas técnicas:** Se observó un incremento en la competencia técnica al pasar de una manipulación básica de celdas en hojas de cálculo a la interpretación de visualizaciones dinámicas. El personal demostró capacidad para navegar entre los registros por colaborador y sus marcaciones en el mes de análisis la revisión y comprensión de gráficas generadas, comprendiendo la lógica de los filtros y la exportación de reportes serializados.
- **Curva de aprendizaje:** Gracias a la arquitectura intuitiva y la estandarización mediante Docker (que garantiza que el sistema siempre funcione igual), la curva de aprendizaje fue breve. Los analistas lograron autonomía total en la carga de nuevas hojas de cálculo mensuales y en la generación de reportes XLSX sin asistencia técnica externa, siendo que el sistema se muestra intuitivo y se han integrado validaciones, así como notificaciones que guía el actuar del analista.

3.4.3. Impacto en el clima laboral y eficiencia

Según expresiones de la UATH, se espera que la competencia técnica alcanzada permita un impacto en la reducir el estrés operativo asociado al cierre de mes. La capacidad de exploración dinámica y automatización de tarea aportaría que se pueda dar una respuesta inmediata ante pedidos de información de las autoridades, validando que el dashboard optimiza la capacidad de respuesta de la unidad de Talento Humano.

La capacidad de filtrar por colaborador y visualizar de manera estructurada por semanas, días y el ajuste de marcaciones, permitió receptar opiniones que se comprenden como reducción de la carga cognitiva, el analista menciona que le permite enfocarse en la gestión de las marcaciones como talento humano y no tiene que preocuparse de las operaciones o del uso de fórmulas para efectuar el ordenamiento de registros y revisión total de los datos. Como resultado, la exploración dinámica permite que el analista dedique la mayoría de su tiempo a interpretar los resultados del mes de análisis gracias a la centralización y reorganización eficientemente de la información administrativa dispersa en la hoja de cálculo.

4. Conclusiones y Recomendaciones

4.1. Conclusiones

La implementación del dashboard operativo permitió reducir el tiempo de procesamiento global al 20% respecto al método tradicional. La transición de un flujo manual a uno automatizado basado en arquitecturas de alto rendimiento (Parquet y Flask) convirtió procesos de días-hombre en tareas de pocos minutos.

El uso del algoritmo k-means para la clasificación de eventos de asistencia garantizó una depuración de datos técnica y objetiva. Esto permitió procesar con éxito los registros de la fase de prueba (170 colaboradores), asegurando una base de datos con mayor limpieza para la toma de decisiones. Si bien, para garantizar la precisión en la clasificación mediante k-means se emplea el Método del Codo, sin embargo, en el sistema desarrollado para la visualización en el dashboard, no se utiliza tal técnica puesto que se tiene el valor deseado o buscado de los k grupos, correspondiente al número de marcaciones frecuentes que realiza cada colaborador, lo que se obtuvo desde la estadística descriptiva como la moda, con algunos condicionales para mejorar el criterio que define el valor de k.

Se validó que la arquitectura contenedorizada con Docker facilita la centralización de registros que antes estaban dispersos en silos mensuales. El sistema demostró ser técnicamente capaz de escalar desde la muestra de prueba hacia la nómina total de la institución sin pérdida de rendimiento.

La aceptabilidad del sistema por parte del personal confirma que la herramienta reduce la carga cognitiva y el estrés operativo, permitiendo que los analistas de talento humano abandonen las tareas mecánicas de limpieza para enfocarse en la interpretación de indicadores y la gestión estratégica.

4.2. Recomendaciones

Se recomienda institucionalizar el uso del dashboard como la interfaz primaria de gestión, sustituyendo definitivamente el uso de hojas de cálculo intermedias para evitar la duplicidad de esfuerzos y la fragmentación de la información.

Dado que el sistema utiliza Docker, se recomienda establecer un protocolo de actualizaciones periódicas para las imágenes y contenedores, garantizando la seguridad y estabilidad del sistema ante futuros cambios en los servidores institucionales.

Extender las sesiones de formación hacia otros niveles directivos para que el dashboard sea utilizado no solo por analistas, sino como una herramienta de consulta para la planificación estratégica zonal.

4.3. Trabajos futuros

Tras la evaluación del sistema y la validación de sus módulos, se proponen las siguientes recomendaciones prácticas para potenciar la escalabilidad, usabilidad y seguridad de la plataforma en un entorno de producción extendido:

Si bien el sistema utiliza un diseño adaptado al entorno de trabajo, para un diseño moderno y escalable se recomienda profundizar en la optimización de los componentes para dispositivos con resoluciones distintas, permitiendo que los administradores supervisen las inconsistencias desde dispositivos móviles o tabletas con la misma fluidez que en una estación de trabajo, mejorando la calidad de la gestión.

Se recomienda la mejora de implementación del nivel de acceso “Usuario” permitiendo que cada funcionario mediante sus propias credenciales acceda exclusivamente a su historial de marcaciones. Esto reduciría la carga operativa del Administrador de Sistema, permitiendo que el empleado visualice sus estados de "Válido" o "Por Validar" en tiempo real, fomentando la transparencia en la gestión de la jornada laboral.

Se sugiere integrar un sistema de notificaciones vía correo electrónico o alertas internas. El objetivo es que el sistema notifique automáticamente al administrador o al funcionario cuando el algoritmo detecte una inconsistencia crítica (como una falta de paridad en los registros o una jornada incompleta) permitiendo una corrección inmediata.

Se recomienda la implementación de certificados SSL/TLS (HTTPS) de nivel institucional para cifrar el tráfico de datos. Asimismo, se sugiere establecer una programación de backups automáticos de los archivos parquet, archivos logs y la base de datos MySQL en un servidor de almacenamiento externo o en la nube para garantizar la recuperación ante desastres.

Dado que el módulo de carga de registros es el punto de entrada principal, se recomienda explorar la integración de tecnologías de reconocimiento óptico para digitalizar aquellas marcaciones que por fallos técnicos en los relojes físicos o aún se manejen en formatos físicos o manuales. Esto consolidaría una base de datos escalable y modularizada en diferentes esquemas dentro de la misma base.

5. Referencias Bibliográficas

- atwork. (2026, abril). *Guía de Josh Bersin: Revoluciona los RRHH con Datos de Empleados*.
<https://atwork.ai/es/resources/articles/josh-bersin-revolucion-rrhh-datos-empleados/>
- Barzelay, M. (2001). La nueva gerencia pública. Un ensayo bibliográfico para estudiosos latinoamericanos (y otros). *Revista del CLAD Reforma y Democracia*, (19).
<https://www.preparatic.org/material/20080712/gestion-publica.pdf>
- Cairo, A. (2013). *The functional art: an introduction to information graphics and visualization*. New Riders.
- Cho, W., Choi, S., & Choi, H. (2023). Human Resources Analytics for Public Personnel Management: Concepts, Cases, and Caveats. *Administrative Sciences*, 13(2).
<https://doi.org/10.3390/admsci13020041>
- Conexia. (2025, octubre 2). *Dashboards en salud: una inversión estratégica*.
<https://conexia.com/blog/dashboards-en-salud-una-inversion-estrategica/>
- Davenport, T., & Harris, J. (2017). *Competing on analytics: Updated, with a new introduction: The new science of winning*. Harvard Business Press.
- Eckerson, W. W. (2011). *Performance dashboards: measuring, monitoring, and managing your business*. Wiley.
- Few, S. (2013). *Information dashboard design : displaying data for at-a-glance monitoring / Stephen Few* (2nd. ed.). Analytics Press.
- Frenk, J., & Moon, S. (2013). Governance Challenges in Global Health. *New England Journal of Medicine*, 368(10), 936–942. <https://doi.org/10.1056/NEJMr1109339>
- García-Villalón, S. (2024). *Dashboard para la gestión de recursos humanos mediante el método de los perfiles*. <https://uvadoc.uva.es/bitstream/handle/10324/67754/TFG-I-2931.pdf?sequence=1>
- Gobierno Abierto Ecuador. (2024, diciembre 16). *Ecuador avanza hacia una salud digital al alcance de todos – Gobierno Abierto Ecuador*. MSP presentó Política Nacional de Transformación Digital del Sector Salud. <https://www.gobiernoabierto.ec/ecuador-avanza-hacia-una-salud-digital-al-alcance-de-todos/>
- Knaflic, C. Nussbaumer. (2015). *Storytelling with data: a data visualization guide for business professionals*. Wiley. https://luutruvn.com/wp-content/uploads/2025/03/Storytelling-with-Data_-A-Data-Visualization-Guide-for-Business-Professionals-PDFDrive-.pdf
- Król, K., & Zdonek, D. (2020). Analytics Maturity Models: An Overview. *Information*, 11(3), 142.
<https://doi.org/10.3390/info11030142>
- Laudon, K. C. ., & Laudon, J. P. . (2020). *Management information systems: managing the digital firm* (16a ed.). Pearson Education.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, 28(1), 3–26.
<https://doi.org/10.1080/09585192.2016.1244699>

- Ministerio de Salud Pública. (2023). *Agenda Digital de Salud 2023 - 2027*. https://www.salud.gob.ec/wp-content/uploads/2023/06/Manual_Agenda_Digital_2023_Seg.pdf
- MSP. (2022). *Modelo de Gestión Institucional*. <https://www.salud.gob.ec/wp-content/uploads/2023/09/Modelo-de-Gesti%00%B8%82n-Institucional-MSP-2022.pdf>
- MSP. (2024, julio 4). *MSP realizó el lanzamiento de la plataforma de datos abiertos sobre los profesionales de la salud a escala nacional*. <https://www.salud.gob.ec/msp-realizo-el-lanzamiento-de-la-plataforma-de-datos-abiertos-sobre-los-profesionales-de-la-salud-a-escala-nacional/>
- Paiva de Báez, G. I. (2024). Evolución histórica de la gestión del talento humano, año 2023. *Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(1). <https://doi.org/10.56712/latam.v5i1.1653>
- Pfeffer, J., & Sutton, R. I. (2006). Evidence-Based Management. *Harvard Business Review*, 63–74. <https://cebma.org/assets/Uploads/Pfeffer-Sutton-Evidence-Based-Management-v2.pdf>
- Pollitt, Christopher., & Bouckaert, Geert. (2017). *Public management reform: a comparative analysis - into the age of austerity* (4a ed.). Oxford University Press.
- Pressman, R. S. ., & Maxim, B. R. . (2020). *Software engineering : a practitioner's approach* (9a ed.). McGraw-Hill Education.
- Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. <http://www.cs.umd.edu/projects/hcil/>
- Tapia, E. (2023, septiembre 12). Salud y educación: Estado nunca ha cumplido la meta de la Constitución. *Primicias*. <https://www.primicias.ec/noticias/economia/gasto-salud-educacion-constitucion-pib/>
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press. <https://kyl.neocities.org/books/%5BTEC%20TUF%5D%20the%20visual%20display%20of%20quantitative%20information.pdf>
- Turban, Efraim., Pollard, Carol., & Wood, G. R. . (2020). *Information technology for management: on-demand strategies for performance, growth and sustainability* (11a ed.). Wiley.
- Vélez, J. C. C., Vélez, J. C. C., García, T. N. F., Flores, M. A. D., & Robles, M. del C. A. (2025). Gestión del talento humano en instituciones de salud: clave para la calidad y la continuidad del servicio. *Polo del Conocimiento*, 10(8), 2456–2474. <https://doi.org/10.23857/pc.v10i8.10332>
- Ware, Colin. (2013). *Information visualization: perception for design* (3a ed.). Morgan Kaufmann.

Anexos

El sistema desarrollado y documentación adicional ha sido respaldada en el siguiente repositorio:

<https://github.com/JoFNAr/sistema-gestion>

Una prueba realizada del funcionamiento del sistema, antes de su versión final, se encuentra en el siguiente enlace

<https://drive.google.com/file/d/1ATtHFQdkRurS2fYEqRHi2WXQpc6TkUWu/view?usp=sharing>

Archivos utilizados durante las pruebas, así como otros recursos que se utilizaron en el proceso CRISP-DM se encuentran en la siguiente carpeta compartida.

https://drive.google.com/drive/folders/15jaHxxhD2n0TykHF5wA2_bQzLdsdUE6N?usp=drive_link