Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ciencias Económicas
Escuela Profesional de Economía

Optimización del proceso de detección de errores en encuestas aplicadas por la Defensoría del Pueblo para supervisar a los Centros de Acogida Residencial (CAR) y otras instituciones públicas, y su impacto en la calidad de los indicadores

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de Economista

AUTOR

Balery Yosimar VARGAS NAVARRO

ASESOR

Mg. José Luis ALFARO MENDOZA

Lima, Perú

2026



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Vargas, B. (2026). *Optimización del proceso de detección de errores en encuestas aplicadas por la Defensoría del Pueblo para supervisar a los Centros de Acogida Residencial (CAR) y otras instituciones públicas, y su impacto en la calidad de los indicadores*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Económicas, Escuela Profesional de Economía]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Balery Yosimar Vargas Navarro
Tipo de documento de identidad	DNI
Número de documento de identidad	46240668
URL de ORCID	-------
Datos de asesor	
Nombres y apellidos	José Luis Alfaro Mendoza
Tipo de documento de identidad	DNI
Número de documento de identidad	09017349
URL de ORCID	https://orcid.org/0009-0002-7036-1293
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Eloy Eduardo Avalos Alvarado
Tipo de documento	DNI
Número de documento de identidad	07128775
Miembro del jurado 1	
Nombres y apellidos	David Juan Sanchez Cruz
Tipo de documento	DNI
Número de documento de identidad	08686743
Datos de investigación	
Línea de investigación	D.1.5.2 Gestión y Economía de Empresa

Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Universidad Nacional Mayor de San Marcos y sus coordenadas geográficas. Ejemplo: Edificio: Facultad de Ciencias Económicas de la Universidad Nacional Mayor de San Marcos. País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima. Calle: Av. Universitaria S/N cruce con Av. Venezuela. Latitud: -12.0561578 Longitud: -77.0845196 Se requieren coordenadas, no colocar enlaces. Puedes obtener las **coordenadas GD** de sitios como https://www.google.com.pe/maps/ https://www.coordenadas-gps.com/ https://www.mapsdirections.info/
Año o rango de años en que se realizó la investigación	Abril 2026 - Mayo 2026
URL de disciplinas OCDE	Economía https://purl.org/pe-repo/ocde/ford#5.02.01

ACTA DE LA SUSTENTACIÓN DEL TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO PROFESIONAL DE ECONOMISTA

En Acto Público, en el salón de Grados de la Facultad de Ciencias Económicas de la Universidad Nacional Mayor de San Marcos el día 30 de mayo de 2026 a horas 10.00 a.m. ante el Jurado Presidido por el Catedrático ***Mg. ELOY EDUARDO AVALOS ALVARADO e integrado por los Catedráticos Dr. DAVID JUAN SANCHEZ CRUZ y Mg. JOSÉ LUIS ALFARO MENDOZA,*** *se presenta el* ***Bachiller BALERY YOSIMAR VARGAS NAVARRO*** *para sustentar su Trabajo de Suficiencia Profesional "***OPTIMIZACIÓN DEL PROCESO DE DETECCIÓN DE ERRORES EN ENCUESTAS APLICADAS POR LA DEFENSORIA DEL PUEBLO PARA SUPERVISAR A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR) Y OTRAS INSTITUCIONES PÚBLICAS, Y SU IMPACTO EN LA CALIDAD DE LOS INDICADORES***" para optar el Título Profesional de Economista.***

De acuerdo a las normas legales vigentes el Bachiller después de finalizada la exposición absolvió las preguntas planteadas por los miembros del Jurado. Concluida esta fase se suspendió el Acto Público a fin de que los señores catedráticos en votación secreta procedieran a la calificación correspondiente.

Acto seguido el presidente del Jurado informó el graduando había sido Aprobado *por* unanimidad *con el calificativo de* dieciseis *(* 17 *)*
(Letras y números.)

Para dar fe del acto, suscriben la presente acta el Presidente y los Miembros del Jurado a los 30 días del mes de mayo del 2026.

Mg. ELOY EDUARDO AVALOS ALVARADO
Presidente

Dr. DAVID JUAN SANCHEZ CRUZ
Miembro

Mg. JOSÉ LUIS ALFARO MENDOZA
Asesor

UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, Decana de América

Anexo 1

CERTIFICADO DE SIMILITUD

Yo, **JOSÉ LUIS ALFARO MENDOZA**, en mi condición de asesor acreditado mediante Dictamen de la EP N° 000004-2026-EPE-FCE/UNMSM, del trabajo de suficiencia profesional, cuyo título es "**Optimización del proceso de detección de errores en encuestas aplicadas por la Defensoría del Pueblo para Supervisar a los Centros de Acogida Residencial (CAR) y otras instituciones públicas, y su impacto en la calidad de los Indicadores**" presentado por la bachiller **BALERY YOSIMAR VARGAS NAVARRO**, para optar el Título Profesional de Economista, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **7% de similitud**, nivel PERMITIDO para continuar con los trámites correspondientes y para su publicación en el repositorio institucional.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del Título Profesional de Economista.

Lima, 27 de mayo de 2026.

Firma del Asesor _____

Nombres y Apellidos del asesor: JOSÉ LUIS ALFARO MENDOZA

DNI: 09017349

DEDICATORIA

Dedicado a la madre de mi hijo que está en el cielo, con quien pasé los años más hermosos de mi vida y hoy representa el angelito que me cuida para criar a nuestro hijo. A mis padres por el amor desinteresado, por creer y apostar por mi sin importar sacrificar por completo sus vidas.

Te amo mamá, sueño cumplido.

AGRADECIMIENTO

Agradezco a la Universidad Nacional Mayor de San Marcos por haberme permitido formar parte de ella, a mis compañeros y profesores con quienes compartí momentos de estudio, risas y anécdotas.

A mi familia quienes me apoyaron incondicionalmente y me brindaron apoyo continuo. A cada una de las personas que con sus consejos aportaron con un granito de arena a lograr este sueño.

ÍNDICE GENERAL

DEDICATORIA....ii
AGRADECIMIENTO.... iii
ÍNDICE GENERAL....iv
ÍNDICE DE TABLAS vi
ÍNDICE DE FIGURAS....vii
ÍNDICE DE ANEXOS.... viii
RESUMEN....ix
ABSTRACT.... viii
INTRODUCCIÓN 1
CAPITULO I: INSTITUCIÓN DONDE SE REALIZA EL TSP....4
1.1 Razón social de la institución....4
1.2 Actividad económica a la que pertenece 4
1.3 Organigrama de la institución e identificación del área a la que pertenece el participante del TSP....5
1.4 Área de la institución donde se desarrolla la experiencia profesional....6
1.5. Cargo que desempeña....6
1.6. Los objetivos del cargo....6
1.7. Descripción del cargo que desempeña....6
1.8. Funciones que realiza 6
1.9. Problemas importantes identificados en el cumplimiento de su función 8
1.10. Identificación y presentación del problema, razón del presente TSP....9
CAPITULO II: PLANTEAMIENTO DEL PROBLEMA 10
2.1. Situación problemática 10
2.2. Formulación del problema....12
2.2.1. Problema Principal 12
2.2.2. Problemas Específicos....12
2.3. Objetivo del TSP 13
2.3.1. Objetivo General....13
2.3.2. Objetivos Específicos 13
2.4. Justificación teórica y práctica del TSP.....13

2.4.1. Justificación Teórica 13
2.4.2. Justificación práctica 15
CAPITULO III: BASES TEÓRICAS 17
3.1. Antecedentes de la investigación 17
3.2. Marco teórico en la que se sustenta para resolver TSP 19
CAPITULO IV: RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS OBTENIDOS 25
4.1 Presentación del caso Supervisión Nacional a Centros de Acogida Residencial (CAR). 25
4.2. Participación en la solución del problema: aportes y resultados (evidencias). 28
4.2.1 Procedimiento para la implementación del Reporte de Errores 29
4.2.2 Impacto de la implementación del Reporte de Errores 40
4.3 Conclusiones 43
4.4 Recomendaciones 44
4.5 Referencias 45
4.6 Anexos 46

ÍNDICE DE TABLAS

Tabla 1. IMPACTO ECONÓMICO Y OPERATIVO DE LA IMPLEMENTACIÓN DEL REPORTE DE ERRORES 43

ÍNDICE DE FIGURAS

Figura 1. ORGANIGRAMA ESTRUCTURAL DE LA DEFENSORÍA DEL PUEBLO..... 5
Figura 2. SECCIONES DE LA ENCUESTA PARA LA SUPERVISIÓN NACIONAL A CENTROS DE ACOGIDA RESIDENCIAL (CAR)..... 25
Figura 3. FLUJO DE TRABAJO PARA LA IMPLEMENTACIÓN DEL REPORTE DE ERRORES EN LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR) 29
Figura 4. VISUALIZACIÓN DE TABLAS DE LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR) EN EL PROGRAMA ORACLE SQL DEVELOPER 30
Figura 5. CONEXIÓN PYTHON A LAS TABLAS DE LA BASE DE DATOS ORACLE DE LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR)..... 31
Figura 6. PREGUNTA 1 CON ERROR EN LA ENCUESTA VIRTUAL 32
Figura 7. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 1 CON ERROR (OBS1)..... 32
Figura 8. PREGUNTA 2 CON ERROR EN LA ENCUESTA VIRTUAL 33
Figura 9. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 2 CON ERROR (OBS2)..... 33
Figura 10. PREGUNTA 3 CON ERROR EN LA ENCUESTA VIRTUAL 34
Figura 11. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 3 CON ERROR (OBS3)..... 35
Figura 12. PROCESO DE EXPORTACIÓN DEL REPORTE DE ERRORES DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL (CAR)..... 36
Figura 13. PRIMER REPORTE DE ERRORES DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA “ERRORES POR CAR” 37
Figura 14. PRIMER REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA “RESUMEN” 38
Figura 15. ÚLTIMO REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA: ERRORES POR CAR..... 39
Figura 16. ÚLTIMO REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA: RESUMEN..... 39

ÍNDICE DE ANEXOS

Anexo 1. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN: INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL 46
Anexo 2. ENCUESTA VIRTUAL DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN I: POBLACIÓN DEL CAR 47
Anexo 3. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN II: DOCUMENTACIÓN DEL CAR 47
Anexo 4.ENCUESTA VIRTUAL DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN III: PERSONAL DEL CAR 48
Anexo 5. ENCUESTA DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN IV: IDENTIDAD 48
Anexo 6. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN V: EDUCACIÓN 49
Anexo 7. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VI: SALUD 49
Anexo 8. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VII: REGLAMENTO INTERNO 50
Anexo 9. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VIII: NORMAS DE CONVIVENCIA 50
Anexo 10. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN IX: INSTANCIAS EXTERNAS 51
Anexo 11. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN X: CONDICIONES MATERIALES 51

RESUMEN

El presente Trabajo de Suficiencia Profesional aborda la problemática asociada a la gestión manual y no estructurada del proceso de validación de encuestas de supervisión en la Defensoría del Pueblo, situación que generaba información de bajo valor económico a un alto costo de transacción organizacional.

El bajo valor económico de la información se debía a la persistencia de valores faltantes e inconsistencias, así como a la disponibilidad tardía de los datos utilizados para la elaboración de Informes Defensoriales y la generación de indicadores institucionales. Asimismo, el alto costo de transacción organizacional estaba asociado a los elevados tiempos operativos y el uso de recursos humanos relacionados a las actividades repetitivas de detección, revisión, seguimiento y subsanación de errores realizadas durante el proceso manual de validación.

El objetivo del trabajo fue optimizar el proceso de detección continua de errores mediante la implementación de un sistema automatizado desarrollado en Python, basado en la aplicación de reglas de validación sobre las bases de datos generadas por las encuestas de supervisión. El sistema permitió identificar automáticamente valores faltantes e inconsistencias, generando reportes diarios en formato Excel para facilitar el seguimiento continuo de la subsanación de errores.

Como resultado, se logró reducir el tiempo del proceso de validación de aproximadamente 7.5 meses a 3.5 meses, disminuyendo significativamente la carga operativa asociada a las

actividades manuales de revisión y seguimiento. Asimismo, la automatización permitió liberar aproximadamente 1,920 horas-hombre y generar un ahorro operativo estimado de S/ 52,000 en costos laborales asociados al proceso manual de validación, evidenciando una relación beneficio/costo favorable.

En conclusión, la automatización del proceso de validación permitió reducir los costos de transacción organizacionales asociados al proceso manual de depuración de datos y fortalecer la disponibilidad oportuna y confiable de información para la generación de indicadores y la toma de decisiones institucionales.

Palabras clave: Costo de transacción institucional, valor económico de la información, Automatización, validación, detección de errores, valores nulos, inconsistencias, calidad de datos.

ABSTRACT

This Professional Sufficiency Work addresses the problems associated with the manual and unstructured management of the validation process for supervision surveys conducted by the Ombudsman's Office of Peru, a situation that generated information of low economic value at a high organizational transaction cost.

The low economic value of the information was due to the persistence of missing values and inconsistencies, as well as the delayed availability of data used for the preparation of Ombudsman Reports and the generation of institutional indicators. Likewise, the high organizational transaction cost was associated with long operational times and repetitive activities of error detection, review, monitoring, and correction carried out during the manual validation process.

The objective of this work was to optimize the continuous error detection process through the implementation of an automated system developed in Python, based on the application of validation rules to the databases generated from supervision surveys. The system automatically identified missing values and inconsistencies, generating daily Excel reports to facilitate the continuous monitoring of error correction.

As a result, the validation process time was reduced from approximately 7.5 months to 3.5 months, significantly decreasing the operational workload associated with manual review and monitoring activities. In addition, the automation made it possible to free

approximately 1,920 man-hours and generate an estimated operational saving of S/ 52,000 in labor costs associated with the manual validation process, evidencing a favorable cost-benefit relationship.

In conclusion, the automation of the validation process made it possible to reduce the organizational transaction costs associated with the manual data cleansing process and strengthen the timely and reliable availability of information for indicator generation and institutional decision-making.

Keywords: Economic value of information, organizational transaction costs, automation, data validation, error detection, missing values, inconsistencies, data quality.

INTRODUCCIÓN

La producción de información confiable y oportuna constituye un elemento fundamental para la toma de decisiones en las instituciones públicas, especialmente en contextos donde las problemáticas sociales requieren respuestas rápidas. Sin embargo, los procesos manuales y no estructurados de validación de información suelen generar dificultades para garantizar la calidad y consistencia de los datos, prolongar los tiempos operativos y elevar los costos de transacción organizacionales asociados a las actividades de revisión, seguimiento y subsanación de errores.

En la Defensoría del Pueblo, esta problemática se evidenciaba en las supervisiones realizadas a nivel nacional, las cuales implicaban el levantamiento de información mediante encuestas aplicadas en campo para la construcción de indicadores utilizados en la elaboración de Informes Defensoriales. La validación y depuración de dicha información se realizaba de manera manual y no estructurada, generando múltiples ciclos de revisión, seguimiento y subsanación de errores entre analistas, Adjuntías y equipos de campo.

Esta dinámica generaba elevados tiempos operativos y altos costos de transacción organizacionales asociados al proceso de validación, debido a la utilización intensiva de recursos humanos en actividades repetitivas de revisión, seguimiento y subsanación de errores, además de permitir la persistencia de valores faltantes e inconsistencias dentro de las bases de datos institucionales.

El tiempo de validación podía extenderse entre siete y ocho meses, ocasionando que la información para las actividades de supervisión institucional estuviera disponible de manera tardía. Esta situación limitaba la capacidad de respuesta frente a problemáticas

sociales y retrasaba la formulación de políticas públicas y la ejecución de acciones de protección orientadas a poblaciones vulnerables.

En este contexto, la relevancia del presente Trabajo de Suficiencia Profesional radica en la necesidad de optimizar el proceso de detección continua de errores mediante la implementación de mecanismos automatizados de validación que permitan fortalecer la completitud y consistencia de la información, reducir los costos de transacción organizacionales asociados al proceso manual de validación y garantizar la disponibilidad de información en el momento requerido para las actividades de supervisión y toma de decisiones institucionales.

El objetivo general del presente trabajo es optimizar el proceso de detección continua de errores en los datos recolectados mediante encuestas de supervisión de la Defensoría del Pueblo, mediante la implementación de mecanismos automatizados de validación, a fin de reducir los costos de transacción organizacionales asociados al proceso de validación y fortalecer la disponibilidad oportuna y confiable de la información para la formulación de políticas públicas.

De manera específica, se busca implementar mecanismos automatizados que permitan detectar valores faltantes e inconsistencias en los datos recolectados, fortaleciendo la completitud, consistencia y confiabilidad de la información utilizada para la generación de indicadores.

En el presente trabajo, la optimización se entiende como la mejora de la eficiencia operativa del proceso institucional de detección continua de errores y validación de información, mediante la reducción de tiempos operativos, costos de transacción organizacionales y utilización intensiva de recursos humanos. En ese sentido, la

automatización desarrollada en Python constituye el mecanismo técnico utilizado para alcanzar dicha optimización.

El alcance del estudio se centra en la supervisión nacional a Centros de Acogida Residencial (CAR), considerando el análisis de las bases de datos generadas a partir de las encuestas aplicadas y la implementación de reglas automatizadas de validación sobre dichas estructuras. No se aborda el rediseño del instrumento de recolección ni la incorporación de validaciones en la etapa de captura de datos.

Como resultado de la implementación del sistema automatizado, se logró reducir el tiempo del proceso de validación de aproximadamente 7.5 meses a 3.5 meses, liberando alrededor de 1,920 horas-hombre y generando un ahorro operativo estimado de S/ 52,000 en costos laborales asociados a actividades manuales de validación y subsanación de errores. Asimismo, se implementaron 164 reglas automatizadas de validación, permitiendo fortalecer la completitud y consistencia de la información utilizada para la elaboración de Informes Defensoriales y la formulación de políticas públicas.

El presente documento se estructura en cuatro capítulos. En el Capítulo I se describe la institución donde se desarrolla la experiencia profesional, así como las funciones desempeñadas y la problemática identificada. En el Capítulo II se presenta el planteamiento del problema, los objetivos y la justificación del trabajo. En el Capítulo III se desarrollan los antecedentes y el marco teórico que sustentan la propuesta. Finalmente, en el Capítulo IV se exponen los resultados obtenidos a partir de la implementación del sistema automatizado, así como las conclusiones y recomendaciones derivadas del estudio.

CAPITULO I: INSTITUCIÓN DONDE SE REALIZA EL TSP

1.1 Razón social de la institución

DEFENSORÍA DEL PUEBLO, RUC 20304117142

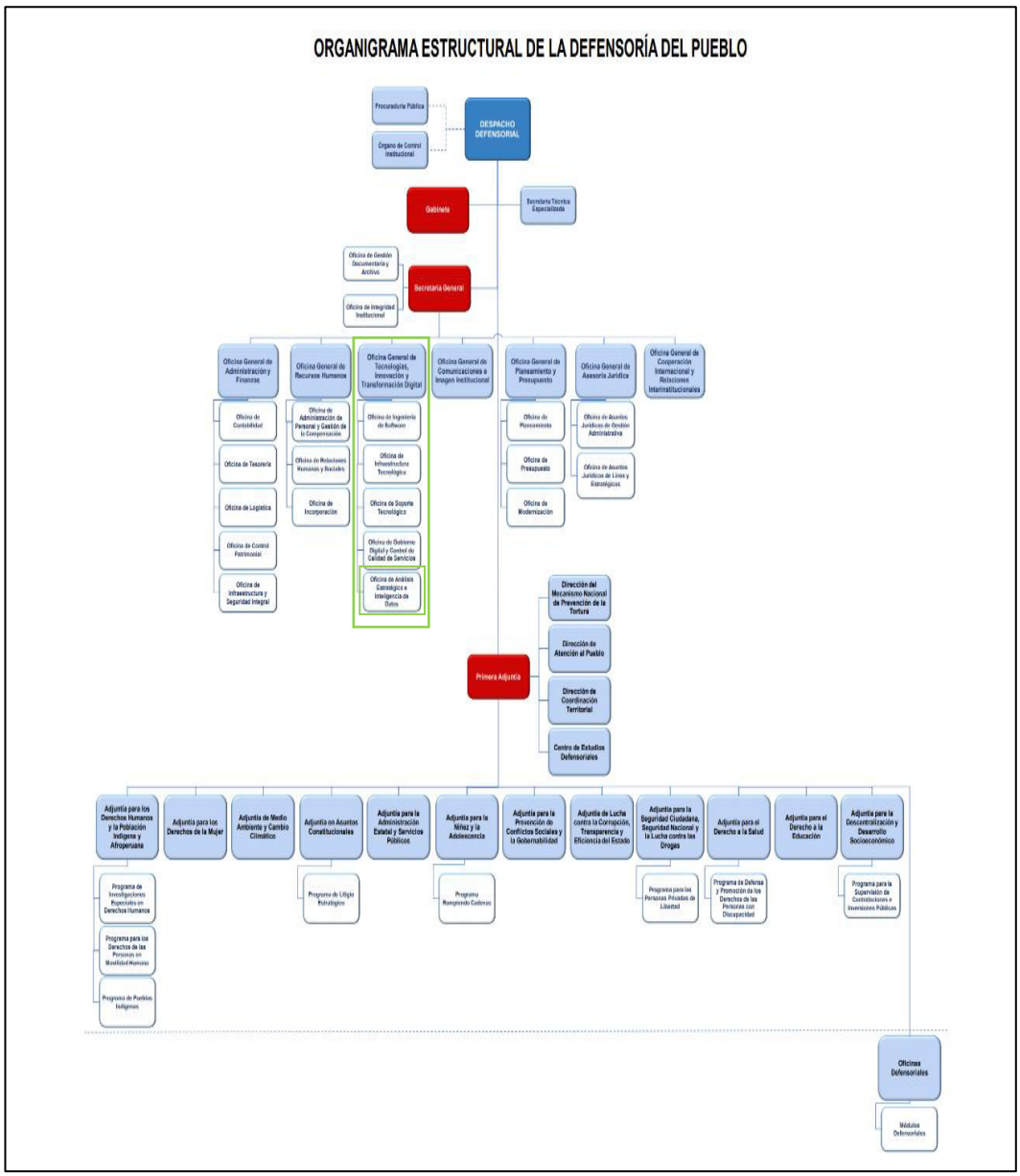
1.2 Actividad económica a la que pertenece

Principal - 8411 – Actividades de la Administración Pública en General.

La Defensoría del Pueblo es una entidad del Estado encargada de defender y promover los derechos de las personas, supervisar la eficacia de la actuación de la administración estatal y garantizar la adecuada prestación de los servicios públicos, a fin de que la población, especialmente aquellos en situación de vulnerabilidad, puedan desarrollar una vida digna y plena.

1.3 Organigrama de la institución e identificación del área a la que pertenece el participante del TSP

Figura 1. ORGANIGRAMA ESTRUCTURAL DE LA DEFENSORÍA DEL PUEBLO



Nota. Organigrama de la Defensoría del Pueblo que muestra la estructura jerárquica institucional, incluyendo el Despacho Defensorial, órganos de apoyo, direcciones, adjuntías especializadas y oficinas defensoriales a nivel nacional. En la Oficina General de Tecnologías, Innovación y Transformación Digital (OGTITD), en específico La Oficina de Análisis Estratégico de Inteligencia de Datos se procesan las encuestas correspondientes a las supervisiones Defensoriales. Fuente: Defensoría del Pueblo.

1.4 Área de la institución donde se desarrolla la experiencia profesional

Oficina de Análisis Estratégico e Inteligencia de Datos, que a su vez pertenece a la Oficina General de Tecnologías, Innovación y Transformación Digital (OGTITD).

1.5. Cargo que desempeña

Analista de datos.

1.6. Los objetivos del cargo

El objetivo del puesto es el análisis de datos y construcción de indicadores, a solicitud de las Adjuntías, los cuales sirven de insumo para la elaboración de Informes Defensoriales. Asimismo, se elaboran tableros de indicadores en POWER BI.

1.7. Descripción del cargo que desempeña

El analista de datos de la Oficina de Análisis Estratégico e Inteligencia de Datos de la Defensoría del Pueblo realiza las labores de limpieza y transformación de datos de encuestas, las cuales se recopilan en campo a través de aplicativos webs y se alojan en bases de datos SQL en servidores, para finalmente construir indicadores que sirven de insumo para que la Adjuntías elaboren sus Informes Defensoriales (en formato Word) los cuales son publicados en el portal de la Defensoría. Asimismo, se elaboran tableros de indicadores en POWER BI.

1.8. Funciones que realiza

Las funciones que realizo actualmente en la Defensoría del Pueblo son:

- Extraer, transformar y cargar datos desde múltiples fuentes de datos para consolidar información institucional en plataformas de inteligencia de negocios a fin de habilitar

la toma de decisiones basada en datos proporcionando información limpia, validada y accesible para todos los niveles organizacionales.

- Elaborar reportes tipo tablero de indicadores con Power BI, dónde se expongan estadísticas detalladas que reflejen el desempeño de los programas sociales y el impacto de las iniciativas de apoyo al ciudadano.

- Crear modelos de datos, a través de la construcción de algoritmos y modelos matemáticos, con la finalidad de analizar y predecir el comportamiento de la información, utilizando herramientas especializadas para tal fin.

- Elaborar y evaluar, las proyecciones, edición, control estadístico, estandarización y validación de una Base de Datos que permitan asegurar datos precisos y útiles para la toma de decisiones y elaboración de documentos institucionales.

- Revisar y aprobar metodologías para las diferentes actividades estadísticas de la entidad. - Brindar asesoramiento a las dependencias de la entidad en temas estadísticos para el desarrollo de sus investigaciones y/o supervisiones; así como coordinar con las dependencias los pedidos de información que se requieran al INEI.

- Recopilar y limpiar los datos obtenidos a través de los sistemas de información, con la finalidad de garantizar su exactitud y análisis. - Participar en las supervisiones defensoriales de alcance nacional que el/la directora(a) General de la Oficina General requiera, para el cumplimiento de las funciones institucionales.

- Otras funciones asignadas por la jefatura inmediata, relacionadas a la misión del puesto.

1.9. Problemas importantes identificados en el cumplimiento de su función

En los procesos de supervisión institucional basados en encuestas, la información constituye un recurso fundamental para la generación de indicadores y la toma de decisiones. Sin embargo, durante el desarrollo de las funciones como analista de datos en la Defensoría del Pueblo, se identificó que el proceso de validación y depuración de la información primaria recolectada se realizaba de manera manual y no estructurada, generando elevados tiempos operativos y altos costos de transacción organizacionales asociados a las actividades de revisión, seguimiento y subsanación de errores.

Asimismo, este proceso prolongado generaba la persistencia de valores faltantes e inconsistencias, que, junto con la disponibilidad tardía de la información, ocasionaba que los datos utilizados para la elaboración de Informes Defensoriales y actividades de supervisión institucional presentaran un bajo valor económico, debido a problemas de completitud, consistencia y oportunidad en la disponibilidad de la información.

Como consecuencia, se incrementaba el riesgo de formular diagnósticos imprecisos y tardíos, limitando la capacidad de respuesta institucional y afectando la formulación oportuna de políticas públicas orientadas a poblaciones vulnerables.

1.10. Identificación y presentación del problema, razón del presente TSP

Durante el desarrollo de las funciones como analista de datos en la Defensoría del Pueblo, se identificó una problemática recurrente relacionada con la gestión manual y no estructurada del proceso de validación de encuestas de supervisión.

Si bien la presencia de valores faltantes e inconsistencias constituye una situación inherente a los procesos de levantamiento de información en campo, el principal problema radicaba en la ausencia de mecanismos automatizados que permitieran detectar y gestionar los errores de manera continua y eficiente. En la práctica, el proceso de validación se realizaba mediante múltiples ciclos de revisión, seguimiento y subsanación de errores entre analistas, Adjuntías y equipos de campo, generando altos costos de transacción organizacionales asociados al proceso manual de validación. Asimismo, la dilatación en el tiempo de validación y la persistencia de errores en los datos restaba valor económico a la información generada.

En este contexto, se identificó la necesidad de optimizar el proceso de detección continua de errores mediante la implementación de mecanismos automatizados de validación que permitieran reducir los costos de transacción organizacionales asociados al proceso manual de validación, fortalecer la completitud y consistencia de la información y garantizar su disponibilidad en el momento requerido para fortalecer la capacidad institucional para la formulación de políticas públicas.

CAPITULO II: PLANTEAMIENTO DEL PROBLEMA

2.1. Situación problemática

La Defensoría del Pueblo es una entidad del Estado encargada de defender y promover los derechos de las personas, supervisar la eficacia de la actuación de la administración estatal y garantizar la adecuada prestación de los servicios públicos, especialmente en favor de poblaciones en situación de vulnerabilidad. Para el cumplimiento de estas funciones, la institución desarrolla supervisiones a nivel nacional mediante encuestas aplicadas en campo, las cuales constituyen una fuente fundamental para la generación de indicadores y la elaboración de Informes Defensoriales.

El proceso de supervisión involucra distintas etapas institucionales. Las Adjuntías diseñan los instrumentos de recolección de información; posteriormente, estos son implementados en aplicativos webs por la Oficina General de Tecnologías, Innovación y Transformación Digital (OGTITD) y aplicados en campo mediante dispositivos digitales por parte del personal operativo. Finalmente, la información recopilada es almacenada en bases de datos institucionales y procesada por los analistas de datos de la Oficina de Análisis Estratégico e Inteligencia de Datos, quienes elaboran indicadores y tableros de información utilizados como insumo para la toma de decisiones institucionales y la elaboración de Informes Defensoriales.

No obstante, durante el desarrollo de este proceso se enfrentaba una problemática asociada a la gestión de los errores presentes en las encuestas de supervisión. En la práctica, las bases de datos contenían preguntas sin responder (valores nulos o vacíos) e inconsistencias entre respuestas relacionadas, originadas por errores humanos durante la aplicación de las encuestas, dificultades en la interpretación de las preguntas o ausencia de validaciones adecuadas en el aplicativo de captura.

Sin embargo, la existencia de estos errores no constituía en sí misma la principal problemática, debido a que este tipo de deficiencias son inherentes a los procesos de levantamiento de información en campo. El principal inconveniente radicaba en la forma en que dichos errores eran gestionados dentro de la institución. El proceso de validación y depuración de datos se realizaba de manera manual y no estructurada, demandando múltiples ciclos de revisión, seguimiento y corrección de errores entre analistas, Adjuntías y equipos de campo.

En la práctica, los analistas debían identificar los errores, reportarlos, esperar su corrección y posteriormente volver a verificar manualmente la información en la base de datos, generándose reiterados ciclos de coordinación y revisión que podían extenderse entre siete y ocho meses. Esta dinámica generaba altos costos asociados a la utilización intensiva de recursos humanos para el proceso de validación, debido a las constantes actividades de revisión, seguimiento y subsanación de errores, permitiendo además la persistencia de información de baja calidad dentro de la base de datos institucional.

En conclusión, la institución generaba información de bajo valor económico a un alto costo. El bajo valor económico de la información se debía a que los datos presentaban problemas de calidad (completitud y consistencia), además de encontrarse disponibles de manera tardía para la formulación de políticas públicas. Por su parte, los altos costos de transacción organizacionales resultaban de los largos tiempos de validación y de la utilización intensiva de recursos humanos en actividades manuales de revisión y seguimiento.

En este contexto, se identificó la necesidad de optimizar el proceso de detección continua de errores mediante la implementación de mecanismos automatizados de validación que permitieran reducir los tiempos operativos y los costos de transacción organizacionales asociados al proceso de depuración, fortalecer la completitud y consistencia de la información y garantizar su disponibilidad en el momento requerido para la formulación de políticas públicas.

Como parte de esta solución, se desarrolló un reporte automatizado de errores que permitió realizar el seguimiento continuo de valores faltantes e inconsistencias en las encuestas, facilitando una depuración más rápida y estructurada de la información. Ello permitió producir información confiable de manera oportuna y a menor costo para las actividades de supervisión institucional y la formulación de políticas públicas.

2.2. Formulación del problema

2.2.1. Problema Principal

¿De qué manera optimizar el proceso de detección continua de errores en los datos recolectados mediante encuestas de supervisión de la Defensoría del Pueblo, a fin de reducir los costos de transacción asociados al proceso de validación y fortalecer la disponibilidad oportuna y confiable de la información para la formulación de políticas públicas?

2.2.2. Problemas Específicos

a) ¿De qué manera optimizar el proceso de detección continua de valores faltantes (nulos o vacíos) en los datos recolectados, a fin de fortalecer la completitud y disponibilidad oportuna de la información?

b) ¿De qué manera optimizar el proceso de detección continua de inconsistencias en los datos recolectados, a fin de fortalecer la consistencia y disponibilidad oportuna de la información?

2.3. Objetivo del TSP

2.3.1. Objetivo General

Optimizar el proceso de detección continua de errores en los datos recolectados mediante encuestas de supervisión de la Defensoría del Pueblo, a través de la implementación de mecanismos automatizados de validación, a fin de reducir los costos de transacción asociados al proceso de validación y fortalecer la disponibilidad oportuna y confiable de la información para la formulación de políticas públicas.

2.3.2. Objetivos Específicos

a) Diseñar e implementar un sistema automatizado de validación que permita la detección continua de valores faltantes (nulos o vacíos) en los datos recolectados, a fin de fortalecer la completitud y disponibilidad oportuna de la información.
b) Diseñar e implementar un sistema automatizado de validación que permita la detección continua de inconsistencias en los datos recolectados, a fin de fortalecer la consistencia y disponibilidad oportuna de la información.

2.4. Justificación teórica y práctica del TSP.

2.4.1. Justificación Teórica

El presente trabajo se sustenta en los enfoques de Economía de la Información, Nueva Economía Institucional y calidad de datos, los cuales establecen que la disponibilidad

oportuna y confiable de la información constituye un elemento fundamental para reducir la incertidumbre y fortalecer la toma de decisiones dentro de las organizaciones.

Desde la Economía de la Información, autores como Arrow (1963) y Stiglitz (2000) sostienen que la información posee valor económico debido a su capacidad para reducir incertidumbre y mejorar los procesos de decisión. En ese sentido, la existencia de información incompleta, inconsistente o disponible de manera tardía limita la capacidad institucional para realizar diagnósticos precisos y formular decisiones públicas eficientes.

Asimismo, desde la Nueva Economía Institucional, Williamson (1985) plantea que los costos de transacción corresponden a los recursos destinados a coordinar, supervisar, verificar y corregir actividades dentro de una organización. En el ámbito público, los procesos manuales y no estructurados de validación de información incrementan dichos costos debido a la existencia de múltiples ciclos de revisión, seguimiento y coordinación entre áreas institucionales.

Por otro lado, la literatura sobre calidad de datos, especialmente el modelo propuesto por ISO/IEC 25012, establece que dimensiones como la completitud y consistencia permiten evaluar la presencia de valores faltantes e inconsistencias lógicas en los datos, problemas que afectan directamente la confiabilidad de la información utilizada en procesos de análisis y generación de indicadores.

En este contexto, los enfoques de automatización y validación de datos señalan la importancia de incorporar mecanismos automatizados de detección continua de errores para reducir la intervención manual, fortalecer la calidad de la información y mejorar

la eficiencia operativa en el procesamiento de grandes volúmenes de datos. Bajo esta perspectiva, la implementación de mecanismos automatizados de validación contribuye a reducir los costos de transacción organizacionales asociados al proceso de depuración y fortalecer la disponibilidad oportuna y confiable de información para las actividades de supervisión y formulación de políticas públicas.

2.4.2. Justificación práctica

Desde un enfoque práctico, el presente Trabajo de Suficiencia Profesional aborda una problemática real de la Defensoría del Pueblo relacionada con la gestión manual y no estructurada del proceso de validación y subsanación de datos recolectados mediante encuestas de supervisión.

Esta situación permitía la persistencia de valores faltantes e inconsistencias dentro de las bases de datos institucionales. Asimismo, generaba elevados tiempos operativos y altos costos de transacción organizacionales asociados al uso intensivo de recursos humanos en actividades de revisión, seguimiento y subsanación de errores.

En este contexto, la implementación de un sistema automatizado de detección continua de errores permitió fortalecer la completitud y consistencia de la información mediante la identificación automatizada de valores faltantes e inconsistencias en un menor tiempo, garantizando la disponibilidad de información confiable cuando era requerida para las actividades de supervisión institucional y la formulación de políticas públicas orientadas a poblaciones vulnerables.

Asimismo, la automatización permitió reducir significativamente las actividades manuales de validación y seguimiento, contribuyendo a disminuir los costos de

transacción organizacionales asociados al uso intensivo de recursos humanos dentro del proceso de validación y subsanación de errores

CAPITULO III: BASES TEÓRICAS

3.1. Antecedentes de la investigación

Diversos estudios han abordado la problemática de la calidad de los datos y la necesidad de implementar mecanismos automatizados de validación para reducir errores y mejorar la confiabilidad de la información utilizada en procesos de análisis y toma de decisiones institucionales.

Por ejemplo, en el estudio "Evaluación de la calidad de datos: una aproximación criptográfica", Yanes Pavón et al. (2021) señalan que la calidad de los datos se sustenta en dimensiones como exactitud, completitud y consistencia, destacando además que la integridad de la información puede verse afectada por errores o modificaciones no autorizadas. Los autores sostienen que la presencia de datos incompletos o inconsistentes reduce la confiabilidad de la información y limita su utilidad para los procesos organizacionales. Este enfoque guarda relación con el presente trabajo, debido a que la detección automatizada de valores faltantes e inconsistencias busca fortalecer la completitud y consistencia de la información utilizada para las actividades de supervisión institucional.

De manera similar, Azeroual y Abuosba (2019) analizan problemas de calidad de datos en sistemas organizacionales, identificando la presencia de valores faltantes, duplicidad e inconsistencias entre registros. A través de la aplicación de métricas de calidad y reglas automatizadas de validación, los autores concluyen que los procesos manuales de control resultan insuficientes frente al creciente volumen de información, debido a los elevados tiempos operativos y limitaciones en la detección oportuna de errores. Este planteamiento se relaciona con el presente trabajo, donde la automatización del proceso de validación permitió reducir la dependencia de

revisiones manuales y mejorar la detección continua de errores en las encuestas de supervisión.

Por otro lado, Peralta Ascue (2021) propone un modelo de evaluación basado en el estándar ISO/IEC 25012 aplicado a una base de datos institucional con más de 300 mil registros, identificando deficiencias en dimensiones como consistencia, conformidad y actualidad. Como resultado, se implementó un proceso automatizado de limpieza y validación de datos mediante consultas SQL y herramientas de depuración, logrando mejorar significativamente la calidad de la información. Este antecedente se vincula con el presente trabajo debido a que ambos enfoques utilizan mecanismos automatizados de validación para fortalecer la confiabilidad de la información y optimizar el procesamiento de grandes volúmenes de datos.

Asimismo, Zúñiga Segura y Sánchez Godínez (2012) desarrollan una metodología de evaluación de calidad de datos sobre una base institucional, identificando problemas relacionados con registros incompletos e inconsistencias entre campos. Los autores evidencian que la ausencia de controles sistemáticos de validación incrementa la probabilidad de errores y limita la confiabilidad de la información utilizada en los procesos organizacionales. En este sentido, el estudio resalta la necesidad de implementar mecanismos continuos de control y validación desde las etapas de captura y procesamiento de datos, enfoque que se alinea con la propuesta desarrollada en el presente trabajo.

En conjunto, los antecedentes revisados coinciden en que la automatización de los procesos de validación y depuración constituye un mecanismo fundamental para fortalecer la calidad y confiabilidad de la información, reducir la dependencia de procesos manuales y mejorar la eficiencia operativa en el procesamiento de datos.

Asimismo, evidencian que la disponibilidad de información consistente y oportuna resulta clave para reducir incertidumbre y fortalecer los procesos de análisis y toma de decisiones dentro de las organizaciones. En este contexto, el presente trabajo se enmarca en dichos enfoques, proponiendo una solución automatizada orientada a reducir los costos de transacción organizacionales asociados al proceso manual de validación y fortalecer la disponibilidad oportuna y confiable de información para las actividades de supervisión institucional y formulación de políticas públicas.

3.2. Marco teórico en la que se sustenta para resolver TSP

a) Economía de la Información:

La Economía de la Información analiza cómo la disponibilidad, calidad y distribución de la información afectan la toma de decisiones económicas e institucionales. Esta corriente surge como una crítica a los modelos económicos tradicionales que asumían información perfecta entre los agentes.

Kenneth Arrow (1963) sostiene que la información posee valor económico debido a su capacidad para reducir incertidumbre y mejorar la toma de decisiones. En ese sentido, la información no es gratuita ni perfecta, ya que su obtención, procesamiento y validación implican costos económicos y operativos para las organizaciones.

Por su parte, Joseph Stiglitz (2000) señala que las asimetrías de información generan ineficiencias institucionales, debido a que las decisiones pueden ser adoptadas sobre información incompleta, inconsistente o tardía. En el ámbito público, ello puede afectar la capacidad de respuesta de las instituciones y limitar la adecuada formulación de políticas públicas orientadas a poblaciones vulnerables.

En este contexto, la generación de información confiable y disponible en el momento requerido adquiere relevancia económica e institucional, debido a que permite fortalecer la calidad de los indicadores y reducir la incertidumbre en los procesos de supervisión y toma de decisiones.

En el ámbito público, los indicadores constituyen herramientas fundamentales para el diagnóstico y seguimiento de problemáticas sociales. Por ello, cuando la información utilizada presenta inconsistencias, vacíos o retrasos, puede afectar la confiabilidad de los diagnósticos y limitar la efectividad de las decisiones y políticas públicas dirigidas a poblaciones vulnerables.

b) Calidad de datos:

La calidad de datos se define como el grado en que los datos cumplen con los requisitos necesarios para su utilización en procesos de análisis, generación de información y toma de decisiones (ISO/IEC 25012, 2008). En ese sentido, la calidad de los datos constituye un elemento fundamental para garantizar la confiabilidad de la información utilizada en sistemas organizacionales e institucionales.

De acuerdo con el estándar ISO/IEC 25012, la calidad de datos está compuesta por múltiples dimensiones, entre las cuales destacan la exactitud, completitud, consistencia y oportunidad. Estas dimensiones permiten evaluar la confiabilidad y utilidad de la información dentro de los sistemas de información.

Desde una perspectiva institucional, la existencia de datos incompletos o inconsistentes puede generar diagnósticos erróneos, afectar la generación de indicadores y limitar la capacidad de respuesta de las organizaciones públicas.

En el presente trabajo, la calidad de los datos se aborda mediante la implementación de mecanismos automatizados de validación orientados a fortalecer la completitud y consistencia de la información recolectada en las encuestas de supervisión.

c) Dimensiones de la calidad de datos:

De acuerdo con el estándar ISO/IEC 25012 y Wang y Strong (1996), las principales dimensiones de calidad relacionadas con el presente trabajo son la exactitud, completitud, consistencia y oportunidad de la información.

• Exactitud: grado en que los datos representan correctamente la realidad observada.

• Completitud: presencia de todos los datos requeridos sin omisiones.

• Consistencia: coherencia lógica entre datos relacionados dentro de un sistema.

• Oportunidad: disponibilidad de la información en el momento requerido para la toma de decisiones.

Wang y Strong (1996) señalan que la utilidad de la información no depende únicamente de su nivel de calidad, sino también de su disponibilidad oportuna, debido a que la información pierde valor cuando no se encuentra disponible en el momento necesario para la toma de decisiones institucionales.

En este contexto, el proceso automatizado de detección de errores desarrollado en el presente trabajo permite no solo fortalecer la completitud de los datos mediante la identificación de valores faltantes y la consistencia de la información a través de reglas lógicas de validación, sino también reducir el tiempo requerido para su detección y subsanación, fortaleciendo la disponibilidad oportuna de información para elevar la capacidad de respuesta de las políticas públicas.

d) Nueva economía institucional y costos de transacción:

Williamson (1985) plantea que los costos de transacción corresponden a los recursos destinados a coordinar, supervisar, verificar y corregir actividades dentro de una organización. En el ámbito público, los procesos manuales y burocráticos pueden incrementar estos costos debido a la duplicidad de tareas, tiempos excesivos de revisión y utilización intensiva de recursos humanos.

Asimismo, Niskanen (1971) señala que las organizaciones burocráticas tienden a generar ineficiencias operativas derivadas de procedimientos administrativos complejos y actividades repetitivas que incrementan el gasto corriente institucional.

En el presente trabajo, el proceso manual de validación de encuestas implicaba múltiples ciclos de revisión, seguimiento y corrección de errores, generando elevados tiempos operativos y altos costos asociados al proceso de validación. En este contexto, la automatización permitió reducir actividades manuales repetitivas y los tiempos de subsanación de errores, contribuyendo a la reducción de los costos de transacción organizacionales y al fortalecimiento de la eficiencia operativa institucional.

e) Validación y depuración de datos:

La validación de datos es el proceso mediante el cual se verifican reglas orientadas a asegurar la consistencia y completitud de la información antes de su utilización en procesos de análisis y toma de decisiones. Por su parte, la depuración de datos comprende el conjunto de técnicas destinadas a detectar y corregir errores como valores nulos, inconsistencias, duplicidad o registros incorrectos dentro de una base de datos.

Según Kimball y Caserta (2011), la validación de datos constituye una etapa fundamental para garantizar la confiabilidad de la información utilizada en procesos de análisis y toma de decisiones, debido a que la calidad de los datos impacta directamente en la confiabilidad de los resultados obtenidos. En el ámbito de la gestión pública, la existencia de información incompleta o inconsistente puede afectar la generación de indicadores y limitar la confiabilidad de la información utilizada para actividades de supervisión y toma de decisiones.

En este contexto, la completitud se aborda mediante la detección de valores faltantes, mientras que la consistencia se evalúa a través de reglas lógicas de validación entre datos relacionados dentro de un sistema de información.

-Valores faltantes:

Los valores faltantes representan la ausencia de información en campos requeridos dentro de una base de datos. Little y Rubin (2019) señalan que este tipo de problemas puede introducir sesgos en los análisis y afectar la confiabilidad de la información utilizada para la generación de indicadores y toma de decisiones.

-Inconsistencias de datos:

Las inconsistencias de datos se producen cuando existen contradicciones lógicas entre valores relacionados dentro de un conjunto de datos. Rahm y Do (2000) sostienen que este tipo de problemas afecta la confiabilidad y utilidad de la información en los sistemas organizacionales.

f) Automatización y optimización de procesos en la gestión pública:

En el ámbito de la gestión pública, la optimización de procesos no se limita únicamente a modelos matemáticos de maximización o minimización, sino también a la mejora de

la eficiencia operativa en el uso de recursos públicos para el cumplimiento de objetivos institucionales.

La automatización de procesos permite reducir tiempos operativos, disminuir actividades manuales repetitivas y fortalecer la disponibilidad de información confiable para la toma de decisiones.

Desde una perspectiva de gestión orientada a resultados, la mejora de procesos busca incrementar la capacidad institucional utilizando de manera más eficiente recursos humanos, tecnológicos y presupuestarios limitados (García López & García Moreno, 2010).

En el presente trabajo, la implementación automatizada de reglas de validación permitió reducir significativamente los tiempos asociados al proceso manual de revisión y fortalecer la generación oportuna de información confiable para las actividades de supervisión institucional.

CAPITULO IV: RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS OBTENIDOS

A continuación, se expondrá el caso de la Supervisión a Centros de Acogida Residencial (CAR), donde se logró implementar el Reporte de Errores automatizado.

4.1 Presentación del caso Supervisión Nacional a Centros de Acogida Residencial (CAR).

La Supervisión Nacional a Centros de Acogida Residencial, consistía en la aplicación de encuestas a los directores de los 223 CAR a nivel nacional. La encuesta, la cual se encuentra en un aplicativo web en línea (encuesta virtual), contaba con una sección de información general y otras 10 secciones de diferentes temáticas.

Figura 2. SECCIONES DE LA ENCUESTA PARA LA SUPERVISIÓN NACIONAL A CENTROS DE ACOGIDA RESIDENCIAL (CAR)



Nota. La encuesta virtual fue construida por desarrolladores de la Oficina de Ingeniería de Software, que recibieron la encuesta física elaborada y ya validada por la Adjuntía. Fuente: Defensoría del Pueblo

Dado el nivel de detalle de cada sección y la cantidad de preguntas involucradas, el desarrollo de la estructura de la encuesta se presenta entre los **Anexos 1 y Anexo 11**, con la finalidad de no sobrecargar el cuerpo principal del trabajo y mantener el enfoque en el análisis del proceso de validación de datos.

Las respuestas recogidas en las encuestas virtuales se almacenan y distribuyen en tiempo real en un conjunto de 21 tablas dentro de la base de datos Oracle institucional, cada una asociada a distintas secciones y niveles de desagregación de la información. Estas tablas son las siguientes:

- CAR_PADRON
- CAR_FICHA
- CAR_FICHA_S1
- CAR_FICHA_S1_P1_2
- CAR_FICHA_S1_P1_3
- CAR_FICHA_S1_P1_6
- CAR_FICHA_S2
- CAR_FICHA_S3
- CAR_FICHA_S3_P3_2
- CAR_FICHA_S3_P3_3
- CAR_FICHA_S4
- CAR_FICHA_S4_P4_2
- CAR_FICHA_S4_P4_4
- CAR_FICHA_S5
- CAR_FICHA_S6
- CAR_FICHA_S6_P6_1

- CAR_FICHA_S7
- CAR_FICHA_S8
- CAR_FICHA_S9
- CAR_FICHA_S9_P9_1
- CAR_FICHA_S10

Cada una de estas tablas se actualiza en línea de manera continua conforme se registran modificaciones en el aplicativo web de la encuesta, lo que permite contar con información actualizada en tiempo real para su procesamiento y análisis.

Los analistas empiezan a procesar la información conectándose a estas tablas y mientras lo hacen van descubriendo errores en la data, los cuales reportan, esperando su validación o corrección y, posteriormente, verificar nuevamente la información, generándose ciclos repetitivos de coordinación. Este procedimiento requería, una comunicación constante entre los analistas, los responsables de las Adjuntías y los equipos de campo, la cual se caracterizaba por ser reiterativa y poco eficiente.

El tiempo excesivo invertido en las actividades de detección, seguimiento, alerta y subsanación de errores prolongaba la persistencia de inconsistencias dentro de las bases de datos, retrasaba la disponibilidad de información confiable para las actividades de supervisión institucional y elevaba los costos asociados al uso intensivo de recursos humanos en el proceso de validación.

4.2. Participación en la solución del problema: aportes y resultados (evidencias).

Ante la problemática expuesta, desarrollé un reporte automatizado de detección continua de errores en los datos, denominado “Reporte de Errores”, generado mediante el lenguaje Python. Para su implementación, fue necesario realizar previamente una exploración de las preguntas con errores en el programa Oracle SQL Developer, a fin de aplicar sobre ellas reglas de validación que permitieran realizar el seguimiento de su subsanación.

El Reporte de Errores era exportado automáticamente en formato Excel y remitido diariamente, en horas de la mañana, a la Adjuntía correspondiente, con el objetivo de que se procediera a la subsanación de las preguntas que presentaban respuestas vacías o inconsistentes directamente en la encuesta virtual.

Como resultado de este proceso, se generaba de manera continua una versión actualizada del Reporte de Errores, que reflejaba el estado real de la depuración de la información.

Figura 3. FLUJO DE TRABAJO PARA LA IMPLEMENTACIÓN DEL REPORTE DE ERRORES EN LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR)

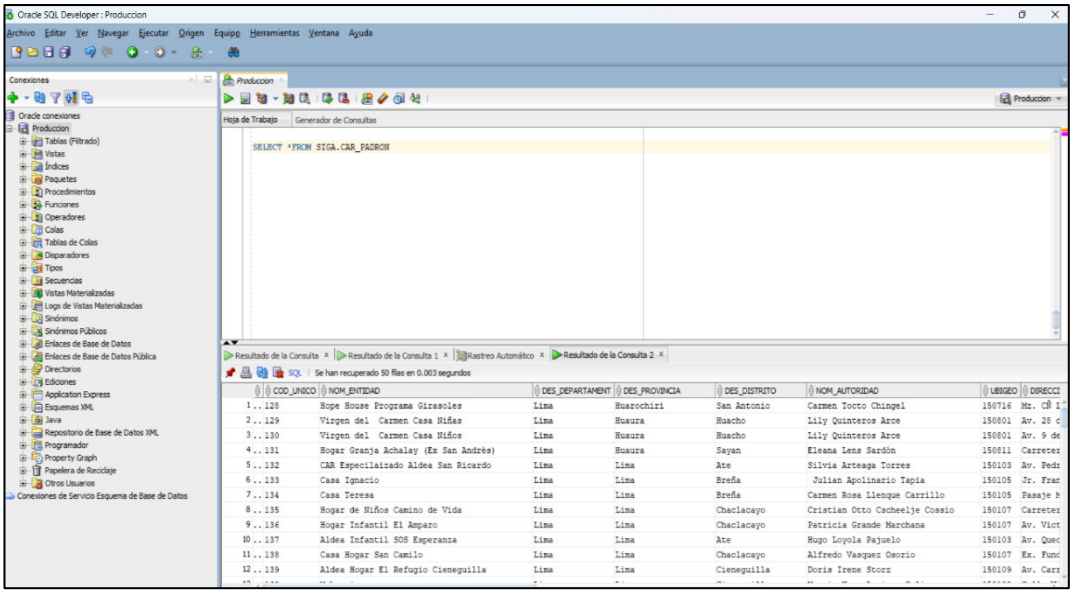


Nota. La información de campo ingresa a través de la encuesta virtual a la base de datos SQL de un servidor institucional, se visualizan las tablas en el programa Oracle SQL Developer, pero son consumidas en Python para la generación del Reporte de Errores en formato Excel, el cual se entrega diariamente a la Adjuntía.
Fuente: Defensoría del Pueblo.

4.2.1 Procedimiento para la implementación del Reporte de Errores

El procedimiento de validación se inicia con una revisión exploratoria de las tablas en el programa Oracle SQL Developer, donde se identifican las preguntas que presentan valores faltantes o inconsistencias en cada una de las tablas. Esta etapa permite reconocer patrones de error y definir las reglas de validación necesarias.

Figura 4. VISUALIZACIÓN DE TABLAS DE LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR) EN EL PROGRAMA ORACLE SQL DEVELOPER

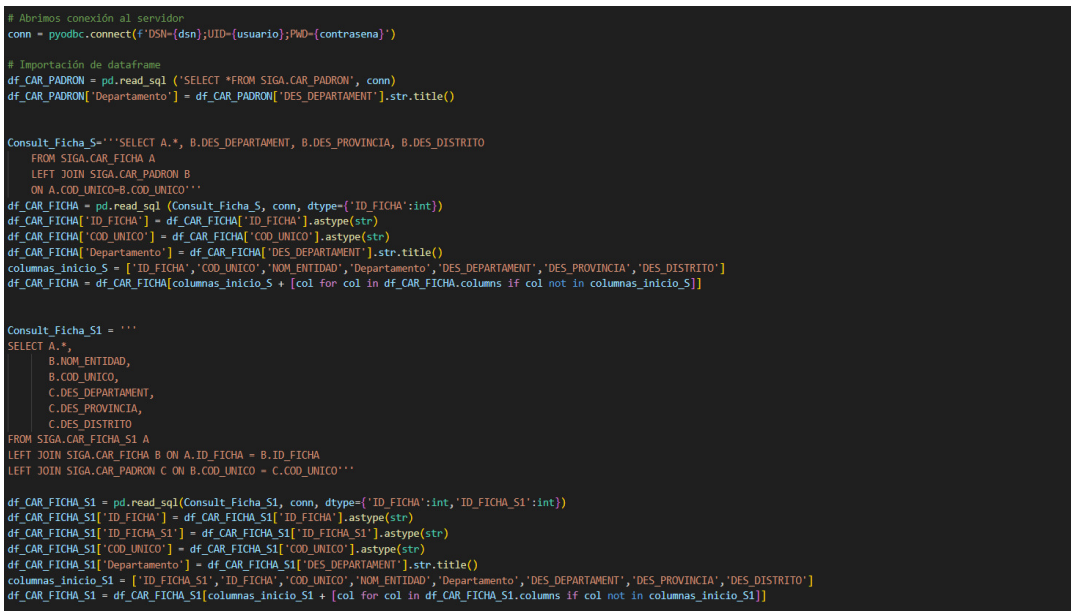


Nota. Para la visualización de las tablas, los analistas utilizan el programa Oracle SQL Developer, mediante el cual ejecutan consultas SQL (por ejemplo, *SELECT * FROM SIGA.CAR_PADRON*), lo que permite visualizar el contenido de la tabla en el panel de resultados. Fuente: Defensoría del Pueblo.

Posteriormente, mediante el lenguaje Python, se cargan las 21 tablas y se procede a aplicar las reglas de validación sobre las preguntas previamente identificadas. Para ello, se crean columnas de observación asociadas a cada pregunta evaluada, siguiendo una nomenclatura secuencial (Obs1, Obs2, Obs3, etc.).

Estas columnas permiten hacer seguimiento continuo a la presencia de errores, mediante la asignación de mensajes específicos en los casos en que se detectan valores vacíos o inconsistencias, y un valor neutro cuando no se registran observaciones.

Figura 5. CONEXIÓN PYTHON A LAS TABLAS DE LA BASE DE DATOS ORACLE DE LA SUPERVISIÓN NACIONAL A LOS CENTROS DE ACOGIDA RESIDENCIAL (CAR)



```
# Abrimos conexión al servidor
conn = pyodbc.connect(f'DSN={dsn};UID={usuario};PWD={contrasena}')

# Importación de dataframe
df_CAR_PADRON = pd.read_sql ('SELECT *FROM SIGA.CAR_PADRON', conn)
df_CAR_PADRON['Departamento'] = df_CAR_PADRON['DES_DEPARTAMENT'].str.title()


Consult_Ficha_S='''SELECT A.*, B.DES_DEPARTAMENT, B.DES_PROVINCIA, B.DES_DISTRITO
    FROM SIGA.CAR_FICHA A
    LEFT JOIN SIGA.CAR_PADRON B
    ON A.COD_UNICO=B.COD_UNICO'''
df_CAR_FICHA = pd.read_sql (Consult_Ficha_S, conn, dtype={'ID_FICHA':int})
df_CAR_FICHA['ID_FICHA'] = df_CAR_FICHA['ID_FICHA'].astype(str)
df_CAR_FICHA['COD_UNICO'] = df_CAR_FICHA['COD_UNICO'].astype(str)
df_CAR_FICHA['Departamento'] = df_CAR_FICHA['DES_DEPARTAMENT'].str.title()
columnas_inicio_S = ['ID_FICHA','COD_UNICO','NOM_ENTIDAD','Departamento','DES_DEPARTAMENT','DES_PROVINCIA','DES_DISTRITO']
df_CAR_FICHA = df_CAR_FICHA[columnas_inicio_S + [col for col in df_CAR_FICHA.columns if col not in columnas_inicio_S]]


Consult_Ficha_S1 = '''
SELECT A.*,
       B.NOM_ENTIDAD,
       B.COD_UNICO,
       C.DES_DEPARTAMENT,
       C.DES_PROVINCIA,
       C.DES_DISTRITO
FROM SIGA.CAR_FICHA_S1 A
LEFT JOIN SIGA.CAR_FICHA B ON A.ID_FICHA = B.ID_FICHA
LEFT JOIN SIGA.CAR_PADRON C ON B.COD_UNICO = C.COD_UNICO'''

df_CAR_FICHA_S1 = pd.read_sql(Consult_Ficha_S1, conn, dtype={'ID_FICHA':int,'ID_FICHA_S1':int})
df_CAR_FICHA_S1['ID_FICHA'] = df_CAR_FICHA_S1['ID_FICHA'].astype(str)
df_CAR_FICHA_S1['ID_FICHA_S1'] = df_CAR_FICHA_S1['ID_FICHA_S1'].astype(str)
df_CAR_FICHA_S1['COD_UNICO'] = df_CAR_FICHA_S1['COD_UNICO'].astype(str)
df_CAR_FICHA_S1['Departamento'] = df_CAR_FICHA_S1['DES_DEPARTAMENT'].str.title()
columnas_inicio_S1 = ['ID_FICHA_S1','ID_FICHA','COD_UNICO','NOM_ENTIDAD','Departamento','DES_DEPARTAMENT','DES_PROVINCIA','DES_DISTRITO']
df_CAR_FICHA_S1 = df_CAR_FICHA_S1[columnas_inicio_S1 + [col for col in df_CAR_FICHA_S1.columns if col not in columnas_inicio_S1]]
```

Nota. Por ejemplo, la tabla **CAR_PADRON** es cargada en Python mediante la consulta "SELECT * FROM SIGA.CAR_PADRON", siendo almacenada en el DataFrame denominado *df_CAR_PADRON*. De manera análoga, la tabla **CAR_FICHA** se carga en el DataFrame *df_CAR_FICHA*, y este procedimiento se replica para el resto de las tablas, garantizando la disponibilidad de la información para su posterior procesamiento y validación. Fuente: Defensoría del Pueblo.

A continuación, se presentan tres casos representativos de aplicación de reglas de validación de Python, los cuales ilustran el funcionamiento del sistema automatizado para la detección continuo de errores en las encuestas.

4.2.1.1 Reglas de validación

a) Pregunta 1 con Error ("Obs1").

Se inicia con la pregunta "Tipo de CAR", en la cual, si se selecciona la opción "CAR Especializado", se debe especificar el tipo correspondiente. Sin embargo, se identificaron encuestas en las que esta información no fue completada, generando vacíos en los datos recolectados.

Figura 6. PREGUNTA 1 CON ERROR EN LA ENCUESTA VIRTUAL



Nota. La figura muestra un caso de error en la encuesta virtual, donde se selecciona la opción "CAR Especializado" sin especificar el perfil correspondiente. Para la validación, ambas variables se encuentran en la tabla *df_CAR_FICHA*, específicamente en las columnas "TIPO_CAR" y "PERFIL_CAR", lo que permite identificar registros con valores faltantes. Fuente: Defensoría del Pueblo.

Para la detección de este tipo de error, se implementó una regla de validación orientada a identificar los casos en los que se selecciona la opción "CAR Especializado" sin especificar el perfil correspondiente.

Para ello, se crea la columna "Obs1", en la cual se asigna un mensaje de error a los registros que presentan estos vacíos ("En P0 de la Sección 0. INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL, no especifica el perfil de CAR"), mientras que aquellos que no presentan observaciones mantienen un valor neutro ("-").

De esta manera, mientras se subsanan los valores faltantes en campo a través de la encuesta virtual, es posible realizar el seguimiento de su corrección en la base de datos.

Figura 7. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 1 CON ERROR (OBS1)

```
###Tabla CAR_FICHA

df_CAR_FICHA['Obs1'] = np.where(
    (df_CAR_FICHA['TIPO_CAR'] == "C") &
    ((df_CAR_FICHA['PERFIL_CAR'] == " ") | (df_CAR_FICHA['PERFIL_CAR'].isna())),
    'En P0 de la Sección 0.INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL, No especifica perfil de CAR','-')
```

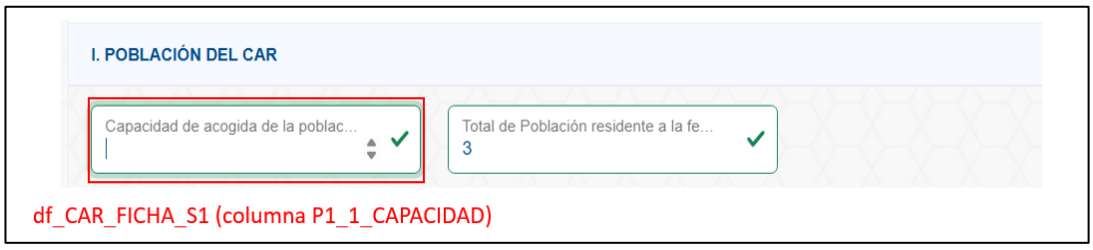
Nota. La figura muestra la implementación de la regla de validación en Python mediante la generación de la columna *Obs1* en la tabla *df_CAR_FICHA*. La condición evalúa los registros en los que la variable "TIPO_CAR" toma el valor "C" (CAR Especializado) y la variable "PERFIL_CAR" se encuentra vacía o nula, asignando un mensaje de error "En P0 de la Sección 0. INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL, no especifica perfil de CAR"; en caso contrario, se asigna el valor "-". Fuente: Defensoría del Pueblo.

b) Pregunta 2 con Error ("Obs2").

En la pregunta sobre la capacidad de acogida de la población, ubicada en la Sección I, se identificaron registros con valores vacíos en la información reportada.

Para la detección de este tipo de error, se implementó una regla de validación que permite identificar los casos en los que no se ha especificado la capacidad de acogida del centro.

Figura 8. PREGUNTA 2 CON ERROR EN LA ENCUESTA VIRTUAL



Nota. La figura muestra un caso de error en la encuesta virtual, donde no se ha completado la información correspondiente a la capacidad de acogida de la población en la Sección I. Esta variable se registra en la columna "P1_1_CAPACIDAD" de la tabla *df_CAR_FICHA_S1*. Fuente: Defensoría del Pueblo.

Para ello, se crea la columna Obs2, en la cual se asigna un mensaje de error a los registros que presentan estos vacíos ("En P1.1 de la Sección I. POBLACIÓN DEL CAR, no especifica la capacidad de acogida de la población"), mientras que aquellos que no presentan observaciones mantienen un valor neutro ("-").

Figura 9. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 2 CON ERROR (OBS2)

```
###Tablas CAR_FICHA_S1

df_CAR_FICHA_S1['Obs2'] = np.where(
    (df_CAR_FICHA_S1['P1_1_CAPACIDAD'] == " ") | (df_CAR_FICHA_S1['P1_1_CAPACIDAD'].isna()),
    'En P1.1 de la Sección I.POBLACIÓN DEL CAR, no especifica capacidad de acogida de la población', '-')
```

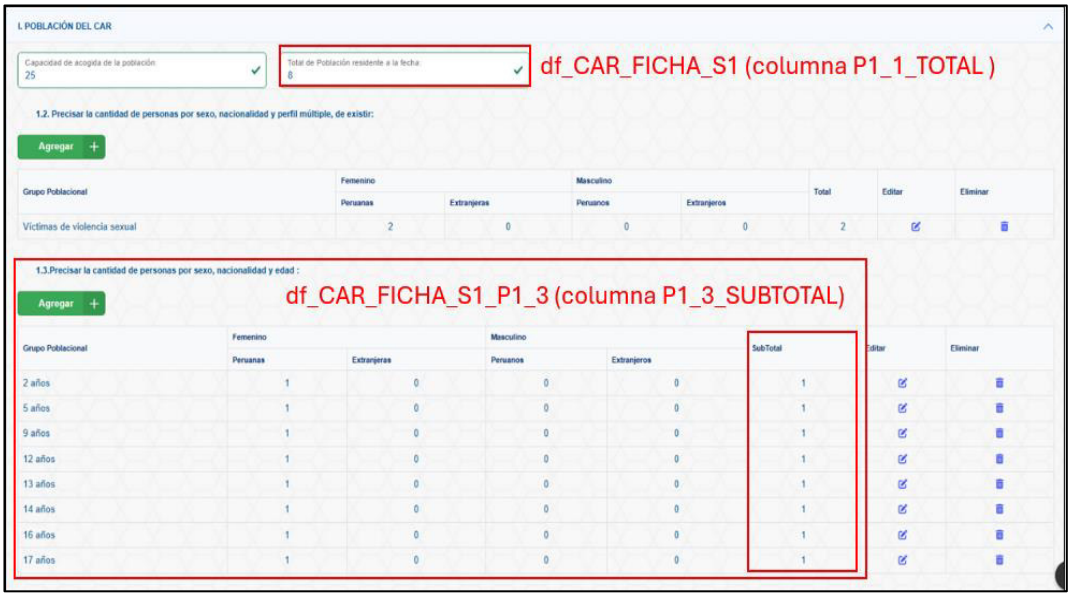
Nota. La figura muestra la implementación de la regla de validación en Python mediante la generación de la columna *Obs2* en la tabla *df_CAR_FICHA_S1*. La condición evalúa los registros en los que la variable "P1_1_CAPACIDAD" se encuentra vacía o nula, asignando un mensaje de error; en caso contrario, se asigna el valor "-". Fuente: Defensoría del Pueblo.

c) Pregunta 3 con Error ("Obs3").

Para la detección de este error, se implementó una regla de validación orientada a identificar inconsistencias en los totales de población. En la Sección I, el total de población residente declarado debe coincidir con la suma del detalle desagregado por sexo, nacionalidad y edad.

A partir de esta condición, la regla permite identificar dos situaciones: la existencia de valores vacíos en la información registrada (ya sea en el total declarado o en la suma del detalle) y la diferencia entre ambos totales.

Figura 10. PREGUNTA 3 CON ERROR EN LA ENCUESTA VIRTUAL



Nota. La figura muestra un caso de inconsistencia en la encuesta virtual, donde el total de población residente no coincide con la suma del detalle desagregado por sexo, nacionalidad y edad. Estas variables se registran en las columnas "P1_1_TOTAL" de la tabla *df_CAR_FICHA_S1* y "P1_3_SUBTOTAL" de la tabla *df_CAR_FICHA_S1_P1_3*. Fuente: Defensoría del Pueblo.

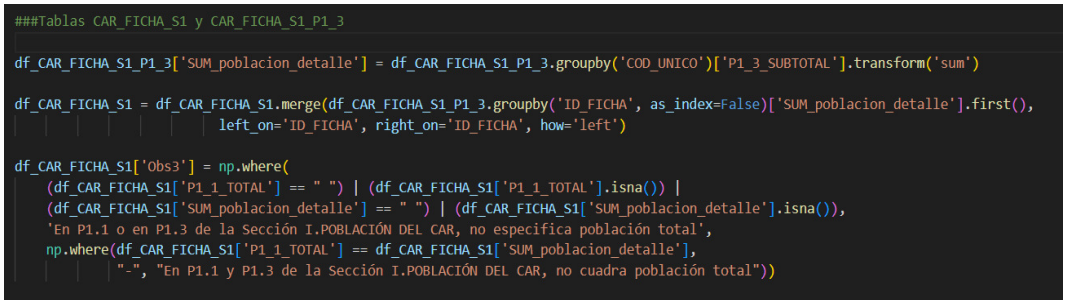
A los registros que no presentan estas inconsistencias se les asigna un valor neutro ("-"); mientras que a aquellos que sí las presentan se les asigna un mensaje de error, según corresponda.

Para los registros con valores vacíos, se asigna el mensaje: “En P1.1 o en P1.3 de la Sección I. Población del CAR, no especifica población total”.

En los casos donde los valores existen, pero no coinciden, se asigna el mensaje: “En P1.1 y P1.3 de la Sección I. Población del CAR, no cuadra población total”.

Finalmente, los mensajes se guardan en la columna *Obs3*.

Figura 11. IMPLEMENTACIÓN DE REGLA DE VALIDACIÓN EN PREGUNTA 3 CON ERROR (OBS3)



```
###Tablas CAR_FICHA_S1 y CAR_FICHA_S1_P1_3

df_CAR_FICHA_S1_P1_3['SUM_poblacion_detalle'] = df_CAR_FICHA_S1_P1_3.groupby('COD_UNICO')['P1_3_SUBTOTAL'].transform('sum')

df_CAR_FICHA_S1 = df_CAR_FICHA_S1.merge(df_CAR_FICHA_S1_P1_3.groupby('ID_FICHA', as_index=False)['SUM_poblacion_detalle'].first(),
                    left_on='ID_FICHA', right_on='ID_FICHA', how='left')

df_CAR_FICHA_S1['Obs3'] = np.where(
    (df_CAR_FICHA_S1['P1_1_TOTAL'] == " ") | (df_CAR_FICHA_S1['P1_1_TOTAL'].isna()) |
    (df_CAR_FICHA_S1['SUM_poblacion_detalle'] == " ") | (df_CAR_FICHA_S1['SUM_poblacion_detalle'].isna()),
    'En P1.1 o en P1.3 de la Sección I.POBLACIÓN DEL CAR, no especifica población total',
    np.where(df_CAR_FICHA_S1['P1_1_TOTAL'] == df_CAR_FICHA_S1['SUM_poblacion_detalle'],
             "-", "En P1.1 y P1.3 de la Sección I.POBLACIÓN DEL CAR, no cuadra población total"))
```

Nota. La figura muestra la implementación de la regla de validación en Python mediante la generación de la columna *Obs3*. Primero, se calcula la suma de la población desagregada a partir de la variable “P1_3_SUBTOTAL”, agrupada por identificador único, generando el campo “SUM_poblacion_detalle”. Posteriormente, este resultado se integra a la tabla *df_CAR_FICHA_S1*, permitiendo comparar el total declarado “P1_1_TOTAL” con el total calculado. La condición identifica registros con valores vacíos o inconsistencias entre ambos totales, asignando un mensaje de error; en caso contrario, se asigna el valor “-”. Fuente: Defensoría del Pueblo.

Este proceso de implementación de reglas de validación en Python se aplica a todas las preguntas con errores identificadas durante la fase de exploración inicial en el programa Oracle SQL Developer. Como resultado, se definieron 164 reglas de validación, generándose columnas de observación (Obs) para cada caso evaluado.

4.2.1.2 Generación del Reporte de Errores (Excel)

Finalmente, *en Python* se genera un*a tabla* denominad*a* df_CAR_FICHA_ERRORES, en *la* cual se consolidan los registros correspondientes a los 223 Centros de Acogida Residencial (CAR) *y las 164 columnas generadas con las reglas de validación.*

Figura 12. PROCESO DE EXPORTACIÓN DEL REPORTE DE ERRORES DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL (CAR)

```
ruta_excel_errores = r'D:\Defensoria Balery\Proyecto CAR\Reporte de Errores\Reporte_Errores_CAR_Funcionarios.xlsx'

columnas_base = [
    'COD_UNICO', 'NOM_ENTIDAD', 'DES_DEPARTAMENT', 'DES_PROVINCIA',
    'DES_DISTRITO', 'USU_REGISTRO', 'FCH_REGISTRO', 'USU_ACTUALIZA',
    'FCH_ACTUALIZA', 'CEL_COMISIONADO', 'CORREO_COMISIONADO'
]

observaciones = [f'Obs{i}' for i in range(1, 164)]

columnas_seleccionadas = columnas_base + observaciones

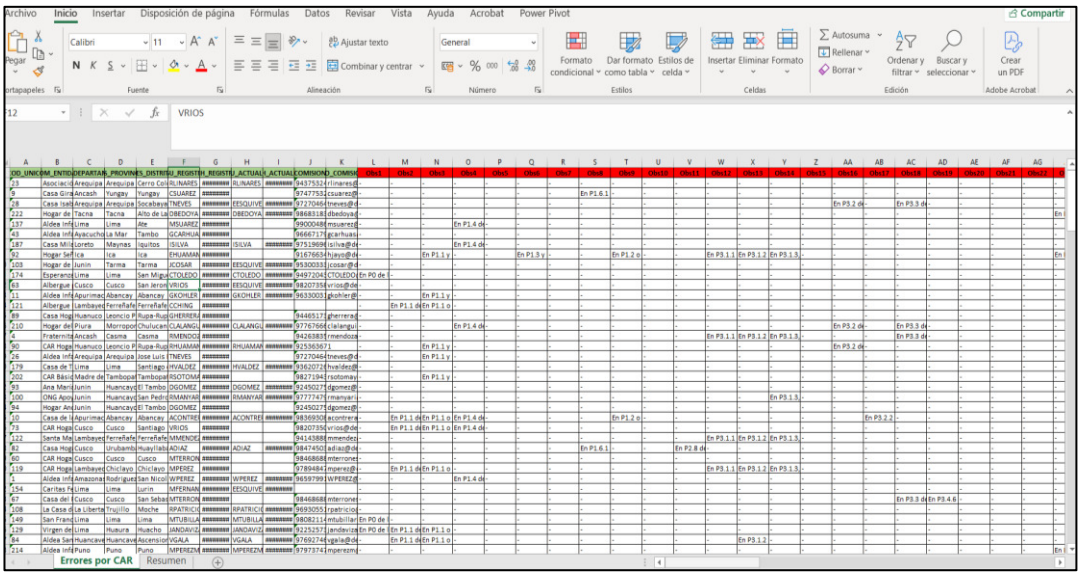
df_CAR_FICHA_ERRORES = df_CAR_FICHA[columnas_seleccionadas]

with pd.ExcelWriter(ruta_excel_errores, engine='openpyxl', mode='a', if_sheet_exists='replace') as writer:
    df_CAR_FICHA_ERRORES.to_excel(writer, sheet_name='Errores por CAR', index=False)
```

Nota. La figura muestra el proceso de exportación del reporte de errores en Python. Se construye el DataFrame *df_CAR_FICHA_ERRORES*, que contiene variables de identificación del CAR (como "COD_UNICO", nombre de la entidad, ubicación geográfica y datos de contacto del encuestador) y 164 columnas de observación (*Obs1* a *Obs164*). Posteriormente, este DataFrame es exportado a un archivo Excel denominado "Reporte_Errores_CAR_Funcionarios.xlsx", generando la hoja "Errores por CAR" mediante el uso de la librería *pandas* y el motor *openpyxl*. Fuente: Defensoría del Pueblo.

Finalmente, se exporta el archivo Excel denominado "Reporte_Errores_CAR_Funcionarios", el cual contiene la hoja "Errores por CAR". En esta hoja, las columnas con relleno verde corresponden a los identificadores del CAR, mientras que las columnas en rojo representan las preguntas con errores (Obs1–Obs164). En estas últimas se consigna el mensaje de error correspondiente a cada pregunta o, en su defecto, un valor neutro ("-") cuando no se registra error.

Figura 13. PRIMER REPORTE DE ERRORES DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA "ERRORES POR CAR"

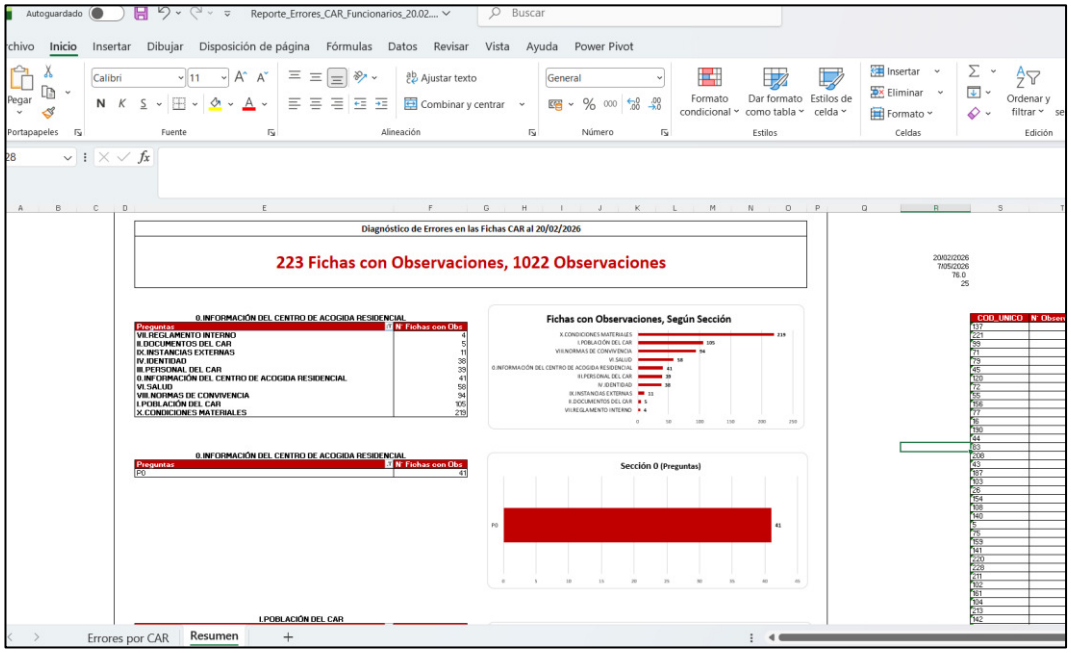


Nota. Archivo en formato Excel (hoja Errores por CAR) enviado a las Adjuntías el 20/02/2026 para subsanar los errores en la encuesta virtual.
Fuente: Defensoría del Pueblo.

Asimismo, el Excel incluye una hoja adicional llamada "Resumen" en la cual, a través de tablas dinámicas y gráficos, se resume la cantidad de errores y en qué secciones y preguntas se concentran, lo cual sirve para medir el avance en la subsanación.

Así, el primer Reporte de Errores fue generado y enviado a la Adjuntía el 20/02/2026 en el cual se registraron 223 Encuestas con errores y un total de 1022 errores, para que inicien la subsanación en la encuesta virtual.

Figura 14. PRIMER REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA "RESUMEN"

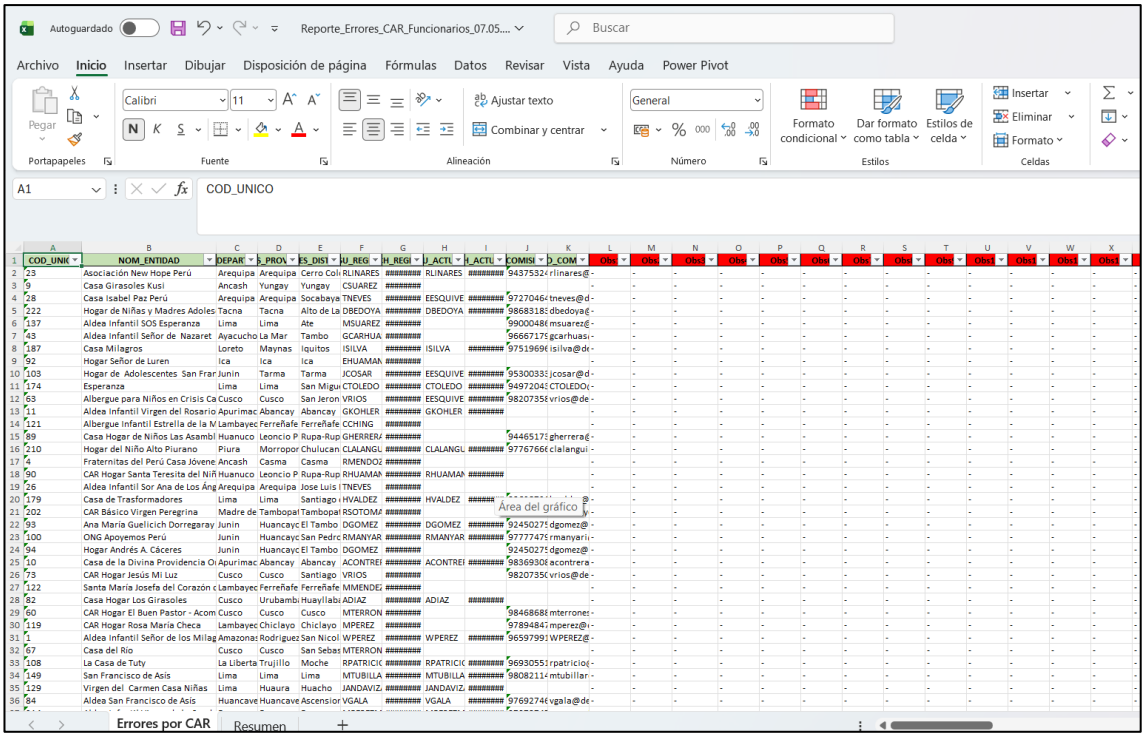


Nota. Archivo en formato Excel (hoja Resumen) enviado a las Adjuntías el 20/02/2026 para subsanar los errores en la encuesta virtual.
Fuente: Defensoría del Pueblo.

Durante un periodo aproximado de 3.5 meses, el archivo Python se ejecutó diariamente en horas de la mañana, generándose automáticamente el Reporte de Errores, el cual reflejaba el estado actualizado de los errores identificados. Posteriormente, dicho reporte era remitido a la Adjuntía correspondiente para la subsanación de errores directamente en la encuesta virtual.

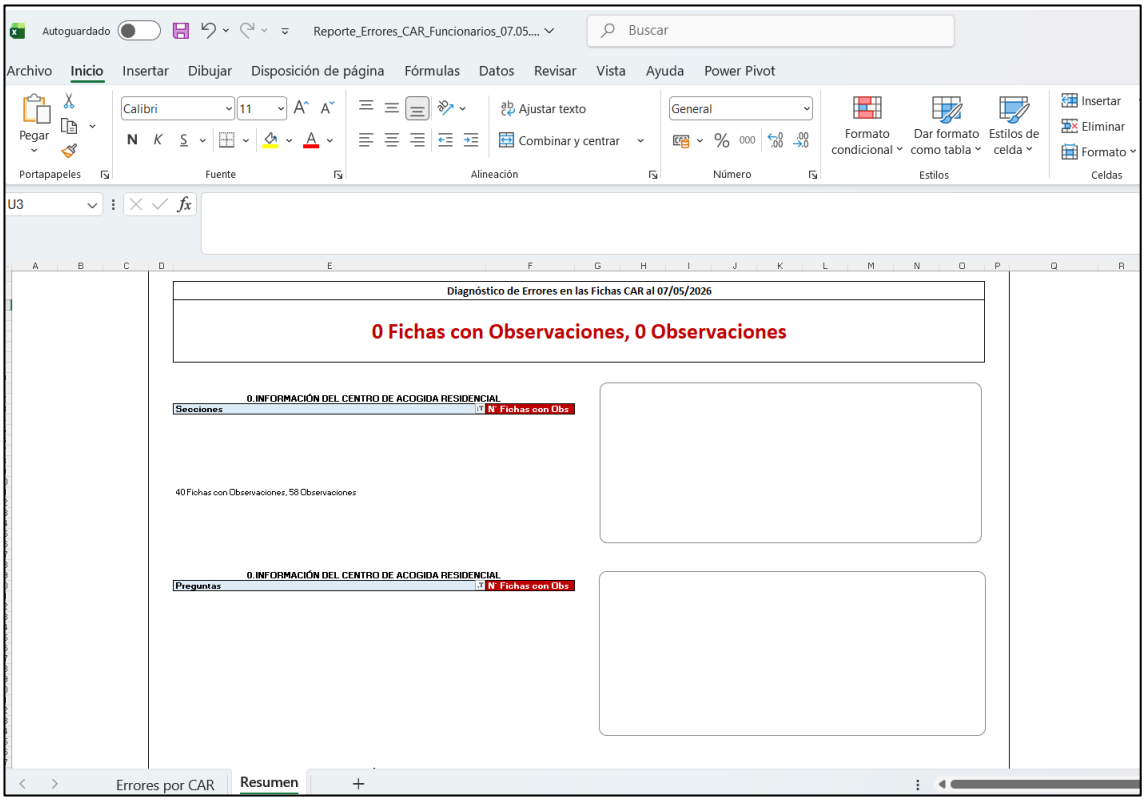
El último Reporte de Errores fue enviado el día 07/05/2026, en el cual se evidenció la subsanación de la totalidad de errores identificados, logrando obtener datos con mayores niveles de calidad, sin respuestas sin contestar ni inconsistencias. Durante todo ese lapso de tiempo, el analista se dedicó a realizar otras labores entre ellas avanzar con los indicadores para el informe.

Figura 15. ÚLTIMO REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA: ERRORES POR CAR



Nota. Archivo en formato Excel (hoja Errores por CAR) enviado a las Adjuntías el 07/05/2026 para confirmar la subsanación total de los errores.
Fuente: Defensoría del Pueblo.

Figura 16. ÚLTIMO REPORTE DE ERRORES DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL (CAR): HOJA: RESUMEN



Nota. Archivo en formato Excel (hoja Resumen) enviado a las Adjuntías el 07/05/2026 para confirmar la subsanación total de los errores.
Fuente: Defensoría del Pueblo.

En su estado inicial, implicaba múltiples ciclos de revisión, comunicación constante con los equipos de campo y una elevada carga operativa, sin garantizar la corrección total de los errores. Con la automatización, se logró generar reportes diarios de errores en un tiempo aproximado de un minuto, permitiendo realizar el seguimiento continuo de los valores faltantes e inconsistencias en los datos y facilitando su subsanación oportuna.

4.2.2 Impacto de la implementación del Reporte de Errores

Antes de la implementación del sistema automatizado, el proceso de validación y subsanación de errores se realizaba de manera manual mediante coordinaciones entre un analista de datos responsable de la supervisión, con una remuneración mensual de S/ 6,000, y dos trabajadores de las Adjuntías, con remuneraciones mensuales de S/ 3,500 cada uno.

Considerando una jornada laboral de 160 horas mensuales, el costo por hora del analista ascendía aproximadamente a S/ 37.50, mientras que el costo por hora de cada trabajador de la Adjuntía ascendía aproximadamente a S/ 21.88.

Dado que el proceso manual de validación se extendía aproximadamente durante 7.5 meses, el analista destinaba 1,200 horas-hombre a dichas actividades, lo que representaba un costo de S/ 45,000. Por su parte, cada trabajador de la Adjuntía destinaba también 1,200 horas-hombre, equivalente a S/ 26,250 por persona; considerando la participación de dos personas, el costo total asociado a este personal ascendía a S/ 52,500.

En conjunto, el proceso manual demandaba aproximadamente 3,600 horas-hombre y representaba un costo organizacional estimado de S/ 97,500, asociado principalmente a actividades operativas de revisión, seguimiento y subsanación de errores. Esta situación evidenciaba elevados costos de transacción organizacionales, debido a la utilización intensiva de recursos humanos en actividades repetitivas de validación y depuración de información.

Tras la implementación del sistema automatizado desarrollado en Python, el tiempo del proceso de validación y subsanación se redujo a 3.5 meses. Bajo las mismas condiciones laborales, el analista destinó 560 horas-hombre al proceso de validación, representando un costo estimado de S/ 21,000. Por su parte, cada trabajador de la Adjuntía destinó también alrededor de 560 horas-hombre, equivalente a S/ 12,250 por persona; considerando la participación de dos personas, el costo total asociado a este personal ascendió a S/ 24,500.

En conjunto, el proceso automatizado demandó aproximadamente 1,680 horas-hombre y representó un costo operativo estimado de S/ 45,500.

En términos comparativos, la automatización permitió generar un ahorro operativo estimado de S/ 52,000, equivalente aproximadamente a una reducción del 53% en los costos de transacción asociados al proceso de validación. Del mismo modo, permitió liberar alrededor de 1,920 horas-hombre, representando una reducción aproximada del 53% en el tiempo destinado a actividades manuales de revisión, seguimiento y subsanación de errores.

Cabe precisar que, para efectos de la presente estimación, se consideraron únicamente los costos de transacción asociados al recurso humano involucrado en el proceso de

validación y subsanación de errores, debido a que estos representaban el principal componente de los costos de transacción organizacionales identificados. Por simplicidad metodológica, no se incorporaron costos indirectos como consumo de energía eléctrica, depreciación de equipos informáticos u otros costos administrativos asociados al desarrollo de las actividades.

Adicionalmente, el diseño e implementación de la solución automatizada tomó aproximadamente 2 semanas y fue desarrollado por un analista de datos con una remuneración mensual de S/ 6,000. Considerando una jornada laboral de 40 horas semanales, la implementación demandó aproximadamente 80 horas-hombre, representando un costo estimado de implementación de S/ 3,000.

En conclusión, la implementación del sistema automatizado incorporó un costo inicial aproximado de S/ 3,000; sin embargo, permitió generar un ahorro operativo estimado de S/ 52,000 en costos de transacción organizacionales asociados al proceso manual de validación y subsanación. Lo anterior evidencia una relación beneficio/costo altamente favorable, debido a que por cada S/ 1 invertido en la implementación del sistema automatizado se generó aproximadamente S/ 17.33 de ahorro operativo.

Asimismo, se logró subsanar la totalidad de los errores identificados en términos de completitud y consistencia, alcanzando una reducción del 100% en los errores y una disminución del 53% en el tiempo del proceso de validación (7.5 a 3.5 meses). Ello permitió generar información de mayor valor económico al fortalecer su calidad y garantizar su disponibilidad oportuna para la elaboración de Informes Defensoriales y la formulación de políticas públicas.

Tabla 1. IMPACTO ECONÓMICO Y OPERATIVO DE LA IMPLEMENTACIÓN DEL REPORTE DE ERRORES

Indicador	Antes de la automatización	Después de la automatización	Variación
Tiempo de proceso de validación	7.5 meses	3.5 meses	-4 meses (-53%)
Horas-hombre destinadas a validación	3,600 HH	1,680 HH	-1,920 HH (-53%)
Costos de transacción	S/97,500	S/45,500	-S/52,000 (-53%)
Costo de implementación	No existía	S/ 3,000	-
Detección/subsanación de errores	Parcial	Completa	100% de errores detectados y subsanados
Disponibilidad de la información	Tardía	Oportuna	Mejora significativa
Valor económico de la información	Bajo	Alto	Mejora significativa

Nota. La implementación del sistema automatizado permitió reducir aproximadamente en 53% el tiempo, las horas-hombre y los costos organizacionales asociados al proceso de validación y subsanación de errores. Asimismo, permitió detectar y subsanar todos los errores en menos tiempo fortaleciendo la disponibilidad oportuna de información para la elaboración de Informes Defensoriales y formulación de políticas públicas para la población vulnerables. Fuente: Defensoría del Pueblo.

4.3 Conclusiones

a) La implementación del sistema automatizado de validación permitió optimizar significativamente el proceso de detección continua de errores en las encuestas de supervisión de la Defensoría del Pueblo, mediante la automatización de reglas de validación orientadas a identificar valores faltantes e inconsistencias en los datos recolectados.

b) La automatización del proceso permitió fortalecer las dimensiones de completitud y consistencia de la información, logrando detectar y subsanar la totalidad de los errores identificados en las bases de datos utilizadas para la elaboración de Informes Defensoriales y generación de indicadores institucionales.

c) La implementación del Reporte de Errores permitió reducir el tiempo del proceso de validación de aproximadamente 7.5 meses a 3.5 meses,

representando una disminución aproximada del 53% en los tiempos operativos asociados a las actividades manuales de revisión, seguimiento y subsanación de errores.

d) La automatización permitió reducir significativamente los costos de transacción organizacionales asociados al proceso manual de validación, liberando aproximadamente 1,920 horas-hombre y generando un ahorro operativo estimado de S/ 52,000 en costos laborales asociados a actividades repetitivas de validación y depuración de información.

e) El proceso de automatización de detección continua de errores permitió obtener información de alto valor económico a bajo costo, fortaleciendo la capacidad institucional para una formulación más rápida y eficiente de políticas públicas orientadas a poblaciones vulnerables.

4.4 Recomendaciones

a) Incorporar herramientas complementarias de analítica y detección automatizada de anomalías que permitan identificar patrones de error no previstos inicialmente en las reglas de validación, fortaleciendo así la capacidad institucional de control y monitoreo de la calidad de los datos.

b) Integrar las reglas automatizadas de validación directamente en los aplicativos de captura de encuestas, a fin de prevenir errores desde el momento del registro de la información y disminuir los costos operativos asociados a procesos posteriores de depuración y corrección.

c) Fortalecer la capacitación del personal de campo en el registro y validación de información, especialmente en la correcta interpretación de preguntas y criterios de consistencia, con el propósito de reducir errores de origen y mejorar la calidad de los datos recolectados.

d) Extender la implementación del sistema automatizado a otras áreas de la Defensoría del Pueblo, con el fin de estandarizar los procesos de validación de datos a nivel institucional.

4.5 Referencias

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. The American Economic Review, 53(5), 941–973.

Azeroual, O., & Abuosba, M. (2019). Improving the data quality in information systems by using data mining techniques. International Journal of Advanced Computer Science and Applications, 10(1), 61–67.

García López, R., & García Moreno, M. (2010). La gestión para resultados en el desarrollo: Avances y desafíos en América Latina y el Caribe. Banco Interamericano de Desarrollo.

International Organization for Standardization. (2008). ISO/IEC 25012:2008. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model. ISO.

Kimball, R., & Caserta, J. (2011). The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data (2nd ed.). Wiley.

Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). Wiley.

Niskanen, W. A. (1971). Bureaucracy and representative government. Aldine-Atherton.

Peralta Ascue, M. (2021). Modelo de evaluación de calidad de datos basado en la norma ISO/IEC 25012 aplicado a bases de datos institucionales [Tesis de maestría, Universidad Nacional Mayor de San Marcos].

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin, 23(4), 3–13.

Stiglitz, J. E. (2000). The contributions of the economics of information to twentieth century economics. The Quarterly Journal of Economics, 115(4), 1441–1478.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–33.

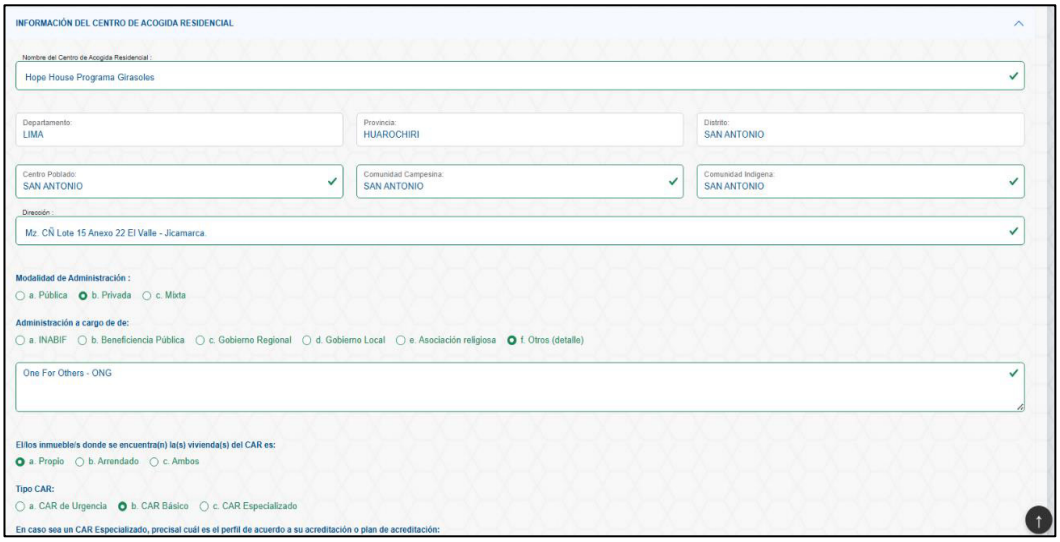
Williamson, O. E. (1985). The economic institutions of capitalism: Firms, markets, relational contracting. Free Press.

Yanes Pavón, R., Hernández García, Y., & Martínez Rodríguez, A. (2021). Evaluación de la calidad de datos: Una aproximación criptográfica. Revista Cubana de Ciencias Informáticas, 15(2), 45–58.

Zúñiga Segura, L., & Sánchez Godínez, M. del C. (2012). Evaluación de calidad de datos en sistemas de información institucionales. Tecnología en Marcha, 25(3), 54–63.

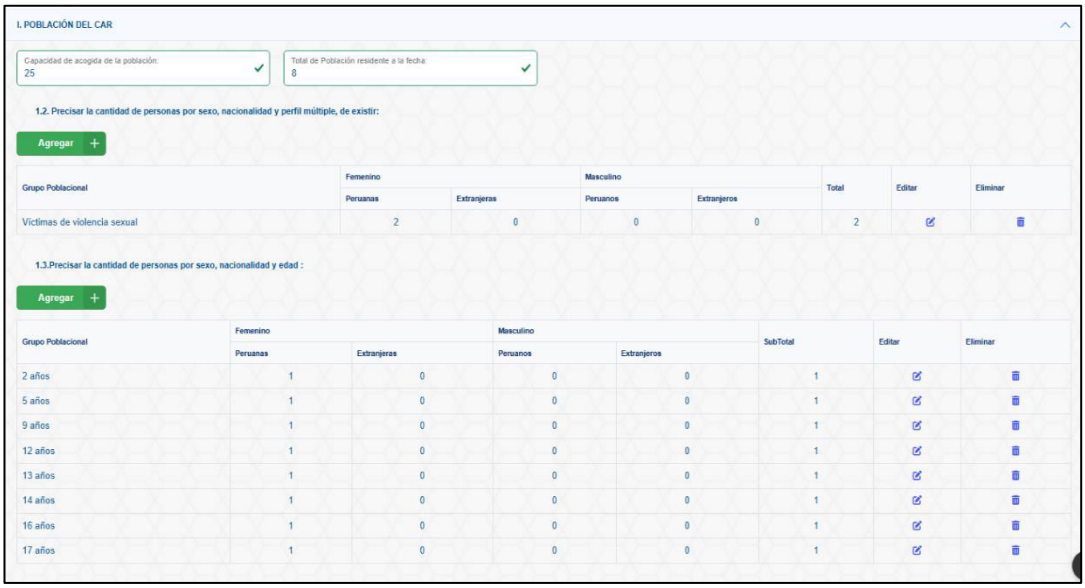
4.6 Anexos

Anexo 1. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN: INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL



Nota. En la sección INFORMACIÓN DEL CENTRO DE ACOGIDA RESIDENCIAL se identifican preguntas sobre la ubicación del CAR (departamento, provincia, distrito, dirección, nombre del CAR), modalidad (público, privado o mixto), entidad a cargo de la administración del CAR (INABIF, Beneficencia pública, Gobiernos Regional, Gobierno Local, Asociación Religiosa, otros). Fuente: Defensoría del Pueblo.

Anexo 2. ENCUESTA VIRTUAL DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN I: POBLACIÓN DEL CAR



I. POBLACIÓN DEL CAR

Capacidad de acogida de la población: 25

Total de Población residente a la fecha: 8

1.2. Precisar la cantidad de personas por sexo, nacionalidad y perfil múltiple, de existir:

Agregar +

Grupo Poblacional	Femenino		Masculino		Total	Editar	Eliminar
	Peruanas	Extranjeras	Peruanos	Extranjeros			
Víctimas de violencia sexual	2	0	0	0	2		

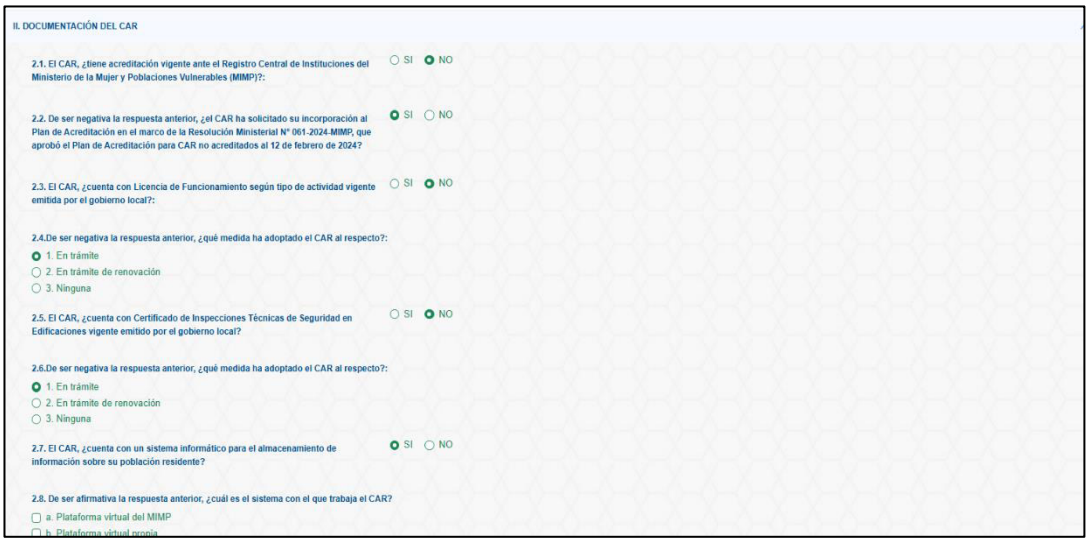
1.3.Precisar la cantidad de personas por sexo, nacionalidad y edad :

Agregar +

Grupo Poblacional	Femenino		Masculino		SubTotal	Editar	Eliminar
	Peruanas	Extranjeras	Peruanos	Extranjeros			
2 años	1	0	0	0	1		
5 años	1	0	0	0	1		
9 años	1	0	0	0	1		
12 años	1	0	0	0	1		
13 años	1	0	0	0	1		
14 años	1	0	0	0	1		
16 años	1	0	0	0	1		
17 años	1	0	0	0	1		

Nota. En la sección I- POBLACIÓN DEL CAR se identifican 6 preguntas entre ellas, declarativas del total de Niños, niñas y adolescentes (NNA) residentes, capacidad del CAR, y detalle del número de NNA por edad, sexo y nacionalidad. Una de las observaciones claves es que el número de NNA declarados debe coincidir con el total de detalle. También encontramos el detalle del número de NNA extranjeros por nacionalidad y sexo. Fuente: Defensoría del Pueblo.

Anexo 3. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN II: DOCUMENTACIÓN DEL CAR



II. DOCUMENTACIÓN DEL CAR

2.1. El CAR, ¿tiene acreditación vigente ante el Registro Central de Instituciones del Ministerio de la Mujer y Poblaciones Vulnerables (MIMP)?: ○ SI ● NO

2.2. De ser negativa la respuesta anterior, ¿el CAR ha solicitado su incorporación al Plan de Acreditación en el marco de la Resolución Ministerial N° 061-2024-MIMP, que aprobó el Plan de Acreditación para CAR no acreditados al 12 de febrero de 2024? ● SI ○ NO

2.3. El CAR, ¿cuenta con Licencia de Funcionamiento según tipo de actividad vigente emitida por el gobierno local?: ○ SI ● NO

2.4.De ser negativa la respuesta anterior, ¿qué medida ha adoptado el CAR al respecto?:
- ● 1. En trámite
- ○ 2. En trámite de renovación
- ○ 3. Ninguna

2.5. El CAR, ¿cuenta con Certificado de Inspecciones Técnicas de Seguridad en Edificaciones vigente emitido por el gobierno local? ○ SI ● NO

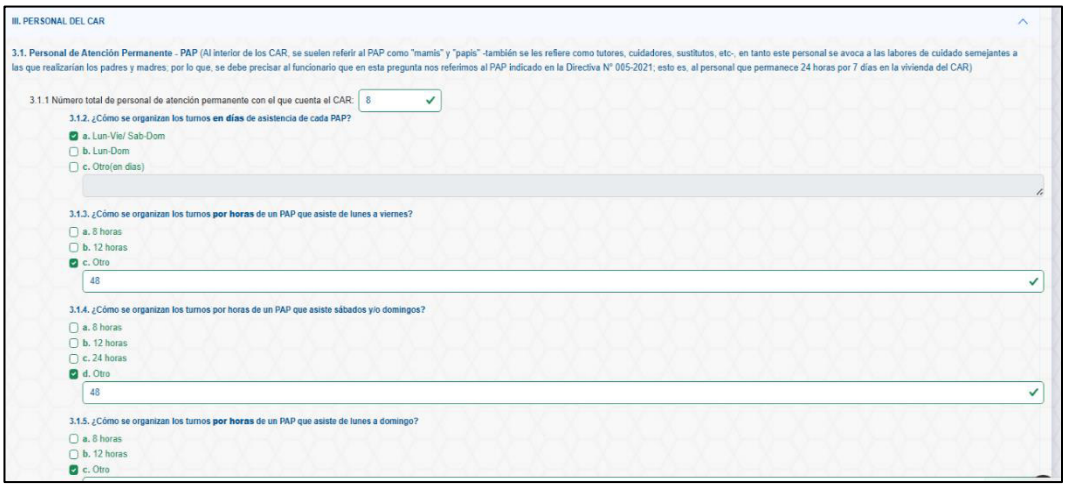
2.6.De ser negativa la respuesta anterior, ¿qué medida ha adoptado el CAR al respecto?:
- ● 1. En trámite
- ○ 2. En trámite de renovación
- ○ 3. Ninguna

2.7. El CAR, ¿cuenta con un sistema informático para el almacenamiento de información sobre su población residente? ● SI ○ NO

2.8. De ser afirmativa la respuesta anterior, ¿cuál es el sistema con el que trabaja el CAR?
- ☐ a. Plataforma virtual del MIMP
- ☐ b. Plataforma virtual propia

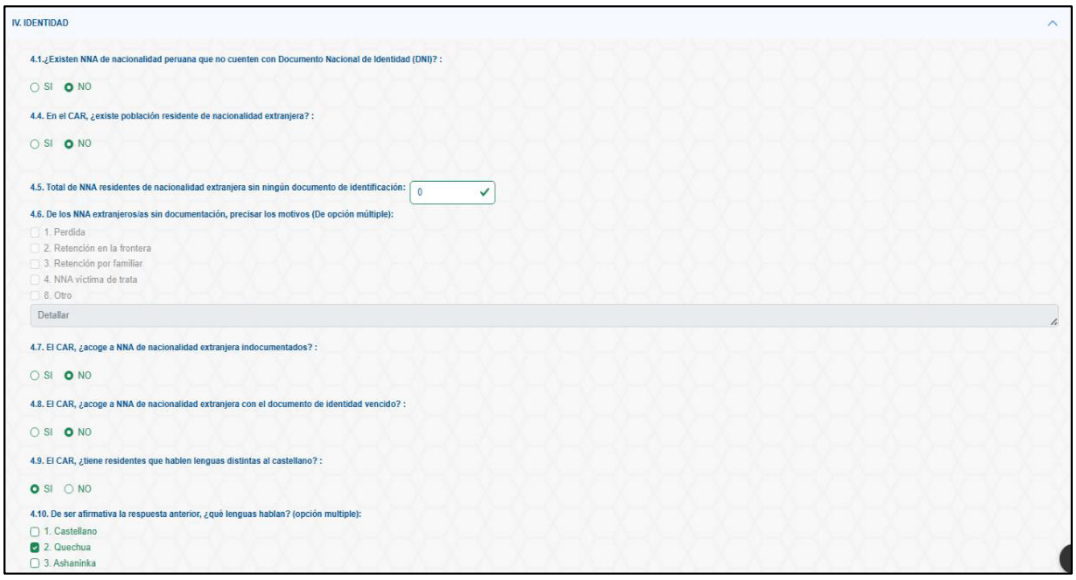
Nota. En la sección III- DOCUMENTACIÓN DEL CAR se identifican 8 preguntas entre ellas, si el CAR cuenta con Licencia de Funcionamiento, Certificado de Inspección Técnica de Seguridad en Edificaciones, Sistema Informático de Almacenamiento de Información e Población Residente. Fuente: Defensoría del Pueblo.

Anexo 4.ENCUESTA VIRTUAL DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN III: PERSONAL DEL CAR



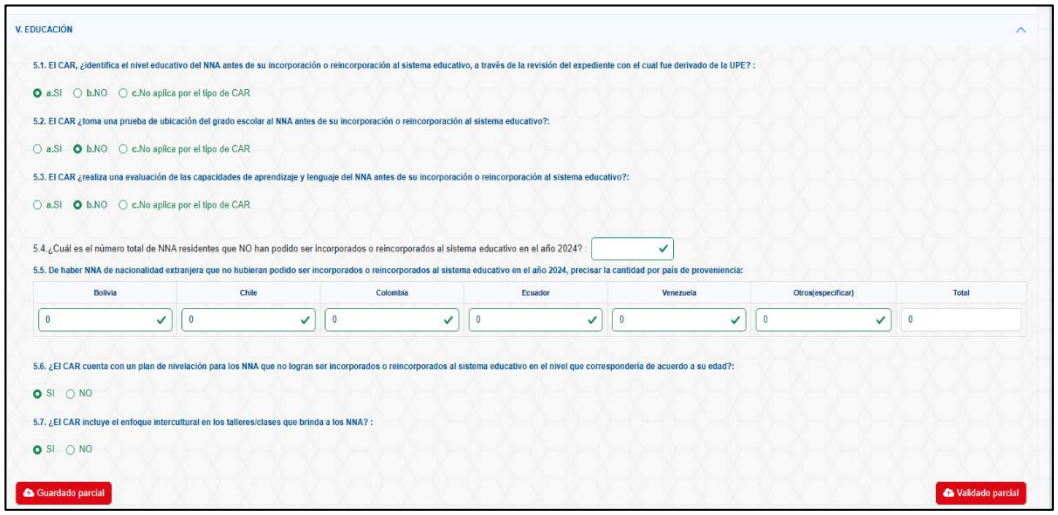
Nota. En la sección III- PERSONAL DEL CAR se identifican 19 preguntas entre ellas, las relacionadas a la jornada laboral del Personal de Atención Permanente (que son los tutores de los NNA), número de personal del equipo técnico (trabajadores sociales, psicólogos, educadores, etc.), equipo de apoyo (operadores de cocina, seguridad, etc). Fuente: Defensoría del Pueblo

Anexo 5. ENCUESTA DE LA SUPERVISIÓN DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN IV: IDENTIDAD



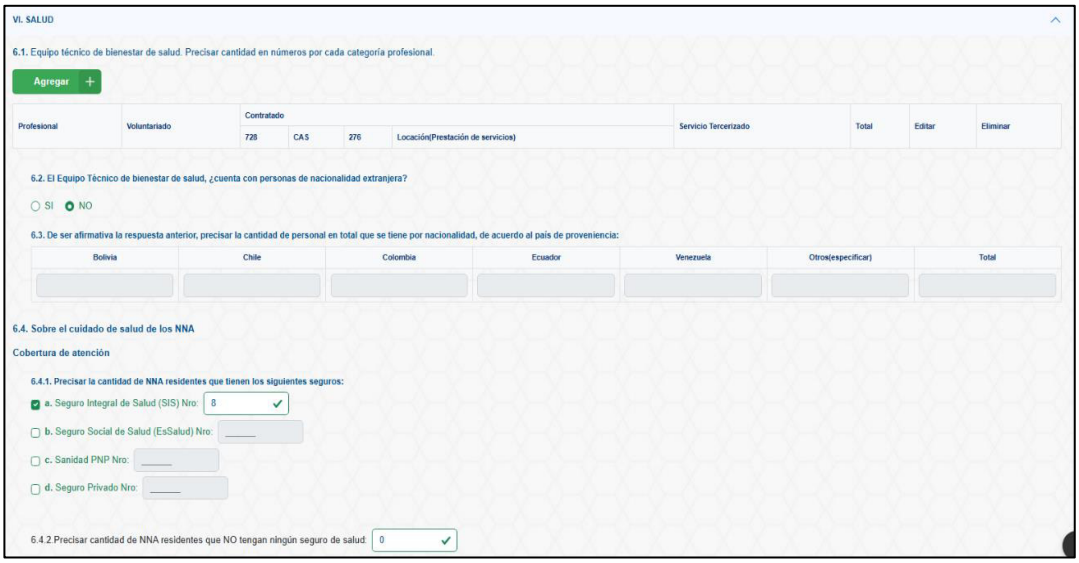
Nota. En la sección IV- IDENTIDAD se identifican 10 preguntas entre ellas, número de NNA sin documento de identificación, NNA que hablen lengua originaria. Fuente: Defensoría del Pueblo

Anexo 6. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN V: EDUCACIÓN



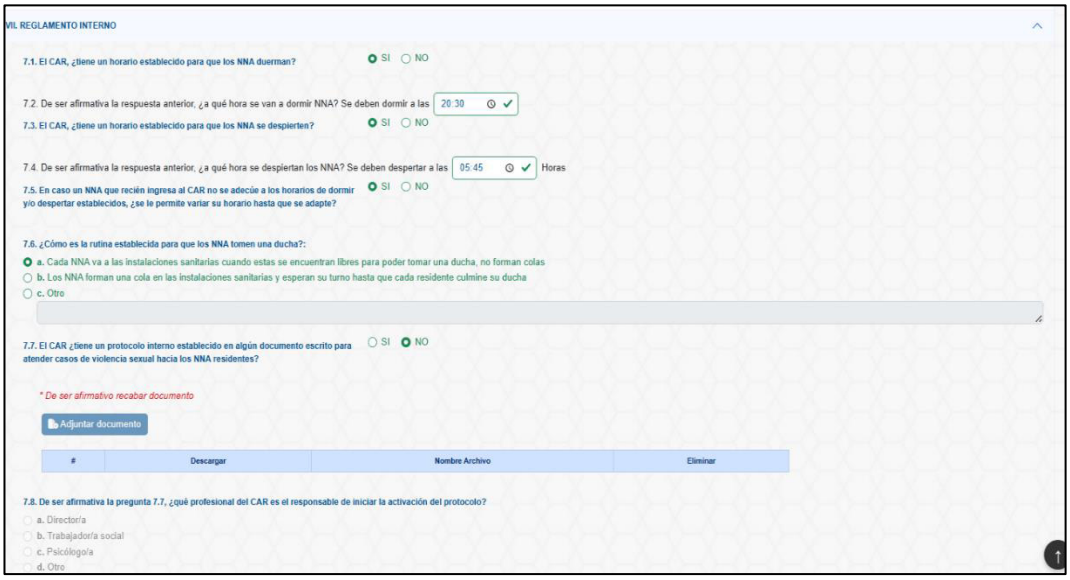
Nota. En la sección V- EDUCACIÓN se identifican 7 preguntas entre ellas, número si se identifica el nivel educativo de los NNA antes de su reincorporación al sistema educativo, si realiza evaluación de capacidades de aprendizaje, si el CAR cuenta con plan de nivelación de NNA, entre otras. Fuente: Defensoría del Pueblo.

Anexo 7. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VI: SALUD



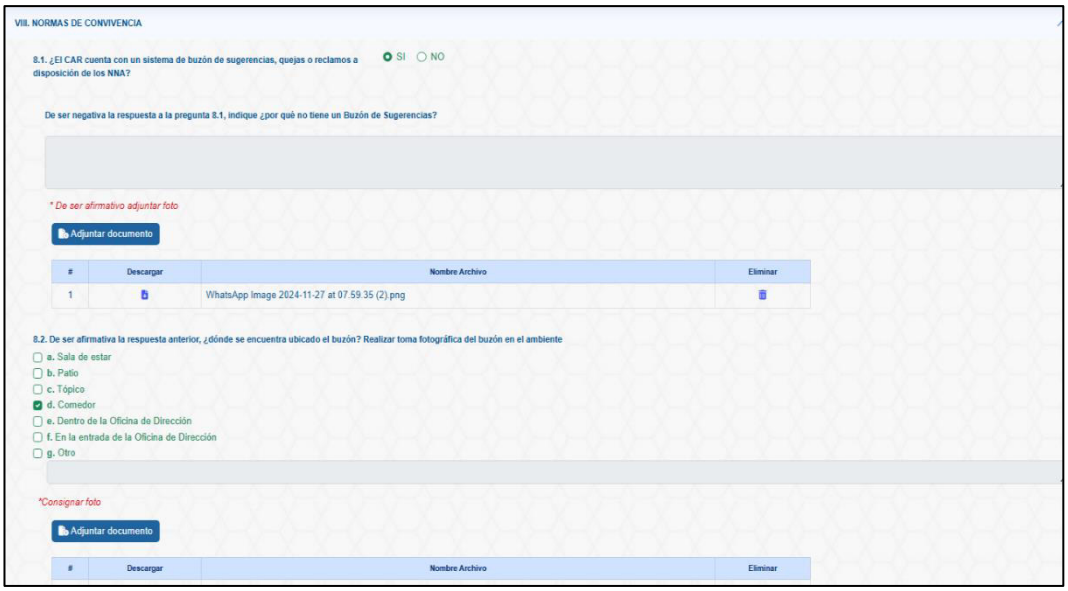
Nota. En la sección V- SALUD se identifican 25 preguntas entre ellas, el número de NNA con seguro, tipo de seguro y sin seguro, si el psicólogo del CAR brinda terapias individuales o grupales a todos los NNA, así como la frecuencia, si reciben controles médicos, entre otras preguntas. Fuente: Defensoría del Pueblo.

Anexo 8. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VII: REGLAMENTO INTERNO



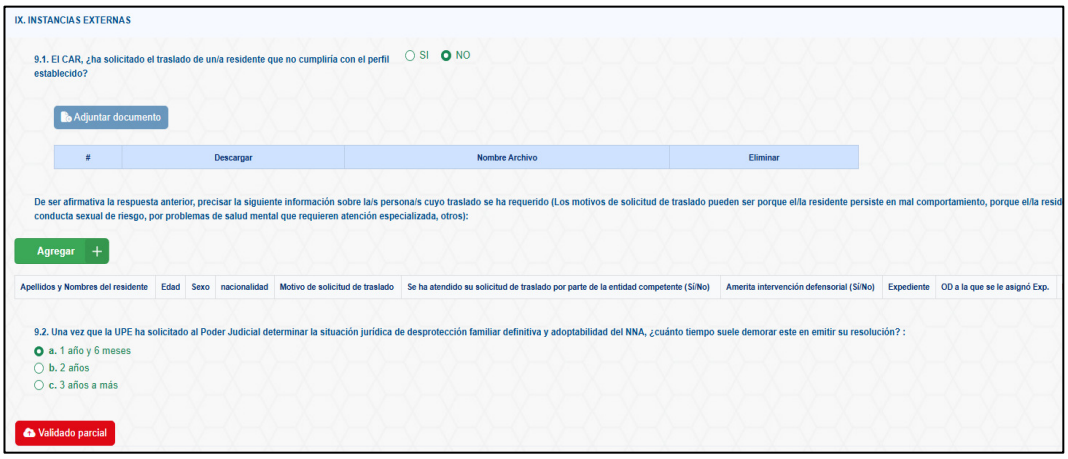
Nota. En la sección VI- REGLAMENTO INTERNO se identifican 10 preguntas entre ellas, si el CAR cuenta con un horario establecido para que los NNA duerman y se despierten, rutina para el uso de la ducha, protocolo para casos de violencia, entre otras. Fuente: Defensoría del Pueblo.

Anexo 9. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN VIII: NORMAS DE CONVIVENCIA



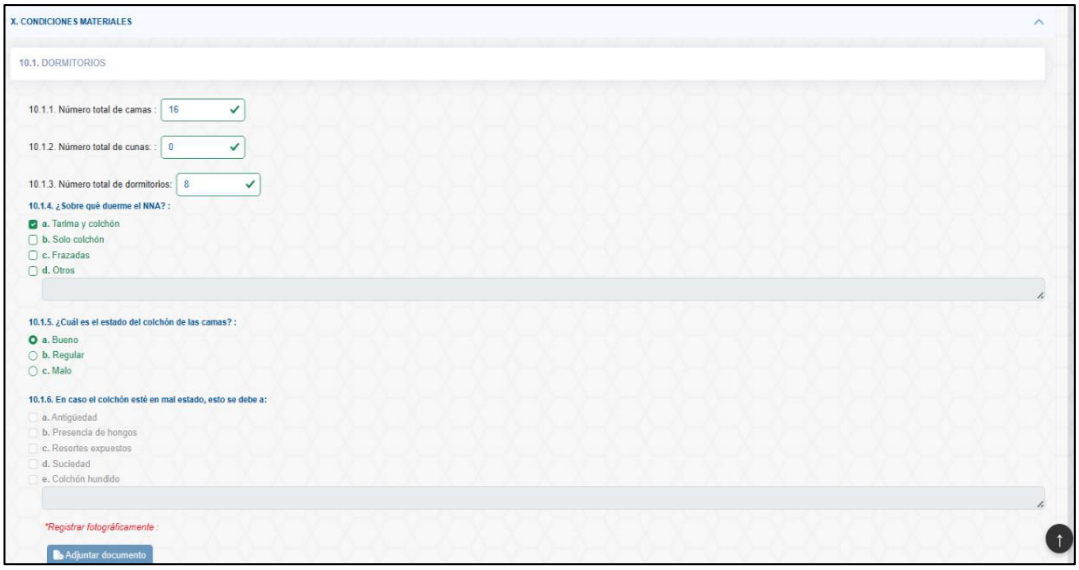
Nota. En la sección VI- NORMAS DE CONVIVENCIA se identifican 14 preguntas entre ellas, si el CAR cuenta con un sistema de buzón de sugerencia y su ubicación, si cuenta con registro de incidencia entre NNAs y personal, guía de metodología para crianza respetuosa, entre otras. Fuente: Defensoría del Pueblo.

Anexo 10. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN IX: INSTANCIAS EXTERNAS



Nota. En la sección IX- INSTANCIAS EXTERNAS se identifican 2 preguntas entre ellas, si el CAR ha solicitado traslado de algún NNA que no cumplía con el perfil del CAR y cuánto tiempo suele demorar emitir la resolución una vez que la UPE ha solicitado al Poder Judicial determinar la situación jurídica de desprotección familiar definitiva y adoptabilidad del NNA. Fuente: Defensoría del Pueblo.

Anexo 11. ENCUESTA VIRTUAL DE LA SUPERVISIÓN NACIONAL DE CENTROS DE ACOGIDA RESIDENCIAL, PREGUNTAS DE LA SECCIÓN X: CONDICIONES MATERIALES



Nota. En la sección IX- CONDICIONES MATERIALES se identifican 77 preguntas entre ellas, relacionadas a si el CAR cuenta con servicios básicos (agua, alcantarillado, energía eléctrica), seguridad (cámaras de videovigilancia, perímetro cercado), estado de los dormitorios y espacios lúdicos, si cuenta con rampas para NNA con discapacidad, entre otras. Fuente: Defensoría del Pueblo.