

Contrato Interadministrativo Nro. 3021063 de 2021
Suscrito entre la Empresa de Telecomunicaciones de Bogotá y el Fondo Financiero
Distrital de Salud - Secretaría Distrital de Salud

Fase 3 del Proyecto
Ejecución de la Migración Fase 3

Código:
SISEM-3-E18-1

Bogotá, 25 de Agosto de 2025

FICHA DE IDENTIFICACIÓN DEL DOCUMENTO

Título del Documento	“Ejecución de la Migración Fase 3”	
Resumen	Este documento aborda la ejecución de la fase 3 de la migración, detallando la continuidad del proceso que se inició en las etapas anteriores. Se describen los pasos a seguir para la creación de los backups finales, así como la planificación para su posterior utilización. Además, se proporcionará una explicación detallada del proceso llevado a cabo para formalizar la migración de datos y la generación de informes correspondientes.	
Código	SISEM-3-E18-1	
Autor	ETB	
Estado	V2	25/08/2025
Revisado por		
Aprobado por		

CONTROL DE CAMBIOS

Versión	Causa del Cambio	Responsable del Cambio	Fecha del Cambio
V1	Versión inicial	Cesar Daniel Urrea López	14/11/2023
V2	Atención ajustes SDS	Cesar Daniel Urrea López	25/08/2025
V3	Versión final	Cesar Daniel Urrea López	25/09/2025

DERECHOS DE AUTOR

Derechos de autor
ETB. Todo el material de este entregable puede ser utilizado libremente por la SDS y las personas a quienes ésta asigne en el marco del convenio citado.

CRÉDITOS

Créditos
Grupo de trabajo de ETB

Tabla de Contenido

1. Introducción.....	7
2. Audiencia.....	8
3. Terminología	9
3.1. Siglas y Acrónimos	9
3.2. Definiciones y conceptos	9
4. Perfilamiento de las Fuentes de Información (SIDCRUE).....	10
4.1. Identificación de la Estructura de Datos.....	10
4.2. Auditoría y Análisis de Calidad	33
4.2.1. Profiling Referencia	33
4.2.2. Profiling Intoxicados	37
4.2.3. Profiling inc_smental	40
4.2.4. Profiling traslados.....	42
4.2.5. Profiling Ambulancias.....	46
4.2.6. Profiling usuarios_capacitados_final.....	49
4.3. Alistamiento de Infraestructura de Migración.....	52
4.4. Preparación de los Repositorios Finales e Intermedios.....	53
4.5. Construcción de scripts de migración	53
5. Copias de Respaldo de la Información.....	57
5.1. Copia del SIDCRUE en BD.....	57
5.2. Actualización de la Copia de Respaldo.....	57
5.3. Backup de Información Histórica	57
6. Evaluación de Calidad de los Datos Durante el Proceso de Migración y Etapas de Migración	58
6.1. Proceso de Migración y Etapas	58
6.2. Reglas de Validación de la Migración.....	59
6.2.1. Completitud	59
6.2.2. Concordancia	66
6.2.3. Integridad	67
6.2.3.1. Automatización del proceso.....	67

Tabla de Ilustraciones

Ilustración 1 Base restaurada en el servidor de pruebas.....	18
Ilustración 2 BUCKET habilitado en GCP para el restaure de BACKUP	18
Ilustración 3 Base restaurada en el servidor de pruebas 2024.....	19
Ilustración 4 Overview referencias 2023.....	34
Ilustración 5 Overview referencias 2024.....	34
Ilustración 6 Vista parcial de los campos tabla referencia	35
Ilustración 7 Campo nombre_hospital 2023.....	35
Ilustración 8 Campo nombre_hospital 2024.....	35
Ilustración 9 campo tipo_evento 2023	36
Ilustración 10 campo tipo_evento 2024	36
Ilustración 11 campo tipodoc 2023.....	36
Ilustración 12 campo tipodoc 2024.....	36
Ilustración 13 Campo edad 2023	37
Ilustración 14 Campo edad 2024	37
Ilustración 15 campo servicio_pte 2023	37
Ilustración 16 campo servicio_pte 2023	37
Ilustración 17 Overview intoxicados 2023	38
Ilustración 18 Overview intoxicados 2024	38
Ilustración 19 Vista parcial de los campos tabla intoxicados 2023	38
Ilustración 20 Vista parcial de los campos tabla intoxicados 2024	39
Ilustración 21 Campo edad 2023.....	39
Ilustración 22 Campo edad 2024	39
Ilustración 23 campo incidente 2023.....	40
Ilustración 24 campo incidente 2024	40
Ilustración 25 Overview inc_smental 2023.....	40
Ilustración 26 Overview inc_smental 2024.....	41
Ilustración 27 Campo prioridad_inc_nuse 2023	41
Ilustración 28 Campo prioridad_inc_nuse 2024	41
Ilustración 29 Campo codigo_tip_inc_nuse 2023	41
Ilustración 30 Campo codigo_tip_inc_nuse 2024	42
Ilustración 31 Oveview traslados 2023	42
Ilustración 32 Oveview traslados 2024	42
Ilustración 33 Campo tipo_traslado 2023.....	43
Ilustración 34 Campo tipo_traslado 2024	44
Ilustración 35 campo paciente 2023.....	45
Ilustración 36 campo paciente 2024.....	45

Ilustración 37 Overview ambulancias 2023.....	46
Ilustración 38 Overview ambulancias 2024.....	46
Ilustración 39 Alerta campos horas_dia, cilindraje, modelo,valor_hora, potencia, id_base 2023.....	46
Ilustración 40 Alerta campos horas_dia, cilindraje, modelo,valor_hora, potencia, id_base 2024.....	47
Ilustración 41 Campo marca 2023.....	47
Ilustración 42 Campo marca 2024.....	48
Ilustración 43 Campo cilindraje 2023.....	48
Ilustración 44 Campo cilindraje 2024.....	49
Ilustración 45 Overview usuarios_capacitacitados_final 2023.....	49
Ilustración 46 Overview usuarios_capacitacitados_final 2024.....	50
Ilustración 47 Campos num_via_prin, num_via_gen, num_placa 2023.....	50
Ilustración 48 Campos num_via_prin, num_via_gen, num_placa 2024.....	50
Ilustración 49 Campo tipo_id 2023.....	51
Ilustración 50 Campo tipo_id 2024.....	51
Ilustración 51 Servidor MySql.....	52
Ilustración 52 Bucket backupsidcrue.....	52
Ilustración 53 -1 Diagrama automatización.....	53
Ilustración 54 distribución de servidores.....	53
Ilustración 55 Detalle Bucket data cargada.....	54
Ilustración 56 Proceso de ejecución unificación tablas.....	55
Ilustración 57 Vista migration_pubsub_message_list_table_joined_sidcrue_v1.....	55
Ilustración 58 Vista migration-transform-data-tables-joined-2-gen-v2.....	56
Ilustración 59 Vista migration_transform_data_tables_not_joined.....	56
Ilustración 60 Vista data_collection_to_bigquery_migration_v1.....	57
Ilustración 61 Servidores bases de datos Google Cloud.....	57
Ilustración 62 Ejemplo copias de seguridad.....	58
Ilustración 63 Diagrama arquitectura de datos.....	59

Tabla de Tablas

Tabla 1 Tablas en BigQuery Fase 2	17
Tabla 2 Tablas no unificadas.....	26
Tabla 3 Tablas unificadas.....	33
Tabla 4 Resultado tablas no unificadas.....	66

1. Introducción

El propósito de este entregable es extender la ejecución del plan de migración, para esto se hará perfilamiento de fuentes de información realizando: Identificación de la estructura de datos, auditoría y análisis de calidad, alistamiento de infraestructura de migración, preparación de los repositorios finales e intermedios, construcción y adaptación de los scripts de migración para lograrlo se incluyen los análisis realizados durante la ejecución que permitieron realizar la migración en esta fase del proyecto entendiendo que se ha planteado una migración por fases de acuerdo con las necesidades de operación del sistema de emergencias médicas y tomando como única fuente de migración el sistema SIDCRUE.

Este documento presenta el alcance, los entregables asociados y los criterios de aceptación de cada uno en el marco del contrato interadministrativo # 3021063 suscrito entre la ETB y el Fondo Financiero Distrital de Salud. Para que sirvan como referencia a la Dirección TIC (Tecnologías de la Información y Comunicaciones) y la Dirección de Urgencias y Emergencias en Salud, de la Secretaría Distrital de Salud de Bogotá.

2. Audiencia

Este documento está dirigido a los supervisores y los líderes del proyecto, más específicamente, a aquellos que se desempeñan en la Dirección de Urgencias y Emergencias en Salud - DUES - y en la Dirección de Tecnologías de la Información y la Comunicación - TIC -, en su calidad de solicitantes y verificadores del cumplimiento de los servicios asociados a este contrato.

3. Terminología

Se presentan las siglas (palabras que se forman a partir de las letras iniciales de un término); los acrónimos (siglas que se forman por la fusión de elementos de dos o más términos, pero que se leen en forma de palabra); las definiciones y los conceptos utilizados en el entregable.

Para facilitar su lectura, se han agrupado así, por orden alfabético:

3.1. Siglas y Acrónimos

APH	Atención Pre Hospitalaria
BD	Base de Datos
CRUE	Centro Regulador de Urgencias y Emergencias
DUES	Dirección de Urgencias y Emergencias en Salud
ETB	Empresa de Telecomunicaciones de Bogotá
IPS	Institución Prestadora de Servicios de Salud
SIDCRUE	Sistema de Información del CRUE
SISEM	Sistema de Información del Sistema de Emergencias Médicas
TARM	Técnico Auxiliar de Regulación Médica
TIC	Tecnologías de la Información y las Comunicaciones

3.2. Definiciones y conceptos

Base de Datos de Origen	Es una base de datos que contiene datos para migrar a una o más bases de datos de destino.
Base de Datos de Destino	Es una base de datos que recibe datos migrados desde una o más bases de datos de origen.
Premier One	Sistema de despacho automático del Centro de Comando, Control, Comunicaciones y Cómputo de la Secretaría Distrital de Seguridad, Convivencia y Justicia y utilizado por el equipo del CRUE.

Recurso APH	Recurso de transporte utilizado en la atención prehospitalaria, usualmente asociado a las ambulancias.
--------------------	--

4. Perfilamiento de las Fuentes de Información (SIDCRUE)

4.1. Identificación de la Estructura de Datos

Como la estructura de datos fue documentada en las fases 1 y 2; en este documento se adjunta el listado de las tablas que la secretaria de salud identificó en la última validación que se realizó de la base de datos, como parte de este documento en los adjuntos se encuentra el diccionario de datos del SIDCRUE.

En la primera Fase de la migración se identificaron 534 tablas, esto se logró realizando un proceso de validación de nombres de columnas y tipo de información almacenada, se agruparon en un total de 154 grupos; de este gran total para la fase 2 se cargaron en BigQuery un total de 40 grupos unificados y 72 tablas sin unificar.

Catalogo/ Schema BigQuery	Tipo	Nombre Tabla
sisem-skg / intermediate_tables_migration_quality	Unificada	amedAdvisory_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	retentionStretchers_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	stretchers_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	assessment_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	userRoles_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	refandcont_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	cie10_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	providersBases_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	payers_table_list

sisem-skg / intermediate_tables_migration_quality	Unificada	ambulance_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	mentalHealthIntervention_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	pendingReview_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	stretcherCar_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	news_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	covidAdvisory_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	liberations_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	country_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	incidents_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	inscribed_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	ambformato_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	amedRequest_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	departmentsColombia_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	availability_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	weights_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	reference_table_list

sisem-skg / intermediate_tables_migration_quality	Unificada	reviewResources_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	burned_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	spendingAmount_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	outOfServiceReview_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	idProcad_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	comments_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	inventories_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	complaints_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	indicatorsIpsAmedChanges_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	hospitals_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	surveys_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	amedTemporal_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	registration_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	data_table_list
sisem-skg / intermediate_tables_migration_quality	Unificada	enterprisesAph_table_list
sisem-skg / intermediate_tables_migration_quality	Sin unificar	lugarinc

sisem-skg / intermediate_tables_migration_quality	Sin unificar	electivas
sisem-skg / intermediate_tables_migration_quality	Sin unificar	rpc_adultos
sisem-skg / intermediate_tables_migration_quality	Sin unificar	horas
sisem-skg / intermediate_tables_migration_quality	Sin unificar	bd_supersalud
sisem-skg / intermediate_tables_migration_quality	Sin unificar	ips_amed
sisem-skg / intermediate_tables_migration_quality	Sin unificar	amed_paciente_reg
sisem-skg / intermediate_tables_migration_quality	Sin unificar	causalidad
sisem-skg / intermediate_tables_migration_quality	Sin unificar	fservicio
sisem-skg / intermediate_tables_migration_quality	Sin unificar	historico_avl
sisem-skg / intermediate_tables_migration_quality	Sin unificar	estados_vehiculos_covid
sisem-skg / intermediate_tables_migration_quality	Sin unificar	circular012
sisem-skg / intermediate_tables_migration_quality	Sin unificar	productividad_amed
sisem-skg / intermediate_tables_migration_quality	Sin unificar	pacformimpactosocial
sisem-skg / intermediate_tables_migration_quality	Sin unificar	desplazamientos
sisem-skg / intermediate_tables_migration_quality	Sin unificar	institucion

sisem-skg / intermediate_tables_migration_quality	Sin unificar	localidad
sisem-skg / intermediate_tables_migration_quality	Sin unificar	dia_cierre
sisem-skg / intermediate_tables_migration_quality	Sin unificar	disponibles
sisem-skg / intermediate_tables_migration_quality	Sin unificar	empresas_publicas
sisem-skg / intermediate_tables_migration_quality	Sin unificar	candidatos
sisem-skg / intermediate_tables_migration_quality	Sin unificar	directorio
sisem-skg / intermediate_tables_migration_quality	Sin unificar	pdf_uci
sisem-skg / intermediate_tables_migration_quality	Sin unificar	ciudades_uci
sisem-skg / intermediate_tables_migration_quality	Sin unificar	asignaciones
sisem-skg / intermediate_tables_migration_quality	Sin unificar	cups
sisem-skg / intermediate_tables_migration_quality	Sin unificar	clasific_novedad
sisem-skg / intermediate_tables_migration_quality	Sin unificar	aglomeraciones
sisem-skg / intermediate_tables_migration_quality	Sin unificar	fechas_ejecucion_scripts
sisem-skg / intermediate_tables_migration_quality	Sin unificar	indicadores_serv_urgencias
sisem-skg / intermediate_tables_migration_quality	Sin unificar	fechas_ambulancias
sisem-skg / intermediate_tables_migration_quality	Sin unificar	llamada_quemados

sisem-skg / intermediate_tables_migration_quality	Sin unificar	hojas_de_vida
sisem-skg / intermediate_tables_migration_quality	Sin unificar	medicoregulator
sisem-skg / intermediate_tables_migration_quality	Sin unificar	reclamos
sisem-skg / intermediate_tables_migration_quality	Sin unificar	historico_radio
sisem-skg / intermediate_tables_migration_quality	Sin unificar	indicadores_emergencias
sisem-skg / intermediate_tables_migration_quality	Sin unificar	notas_inc
sisem-skg / intermediate_tables_migration_quality	Sin unificar	enfermero
sisem-skg / intermediate_tables_migration_quality	Sin unificar	nedocs
sisem-skg / intermediate_tables_migration_quality	Sin unificar	llamantes_sem
sisem-skg / intermediate_tables_migration_quality	Sin unificar	entidad_expide_soat
sisem-skg / intermediate_tables_migration_quality	Sin unificar	informacion_tripulantes_vacuna_covid
sisem-skg / intermediate_tables_migration_quality	Sin unificar	correo_activo
sisem-skg / intermediate_tables_migration_quality	Sin unificar	gestion_liberacion_camillas_eps
sisem-skg / intermediate_tables_migration_quality	Sin unificar	emergencia_sanitaria
sisem-skg / intermediate_tables_migration_quality	Sin unificar	pdf_quejas
sisem-skg / intermediate_tables_migration_quality	Sin unificar	ordendetrabajo

sisem-skg / intermediate_tables_migration_quality	Sin unificar	coordiadores_aph
sisem-skg / intermediate_tables_migration_quality	Sin unificar	pdf_referencia
sisem-skg / intermediate_tables_migration_quality	Sin unificar	pac_smental
sisem-skg / intermediate_tables_migration_quality	Sin unificar	psicologia
sisem-skg / intermediate_tables_migration_quality	Sin unificar	personal_veh_emerg
sisem-skg / intermediate_tables_migration_quality	Sin unificar	entrega_psicologo
sisem-skg / intermediate_tables_migration_quality	Sin unificar	estamentos
sisem-skg / intermediate_tables_migration_quality	Sin unificar	inc_smental
sisem-skg / intermediate_tables_migration_quality	Sin unificar	asistencia_clinica
sisem-skg / intermediate_tables_migration_quality	Sin unificar	ingresonovedades
sisem-skg / intermediate_tables_migration_quality	Sin unificar	apertura_sistema
sisem-skg / intermediate_tables_migration_quality	Sin unificar	camasred
sisem-skg / intermediate_tables_migration_quality	Sin unificar	convocatorias
sisem-skg / intermediate_tables_migration_quality	Sin unificar	nombres_indicadores
sisem-skg / intermediate_tables_migration_quality	Sin unificar	reg_asegurador
sisem-skg / intermediate_tables_migration_quality	Sin unificar	rpc_ninos

sisem-skg / intermediate_tables_migration_quality	Sin unificar	resumen_conglomerados
sisem-skg / intermediate_tables_migration_quality	Sin unificar	representantes
sisem-skg / intermediate_tables_migration_quality	Sin unificar	nov_thumano_covid19
sisem-skg / intermediate_tables_migration_quality	Sin unificar	estados_recursos_bk
sisem-skg / intermediate_tables_migration_quality	Sin unificar	guiones_emergencias
sisem-skg / intermediate_tables_migration_quality	Sin unificar	regimen
sisem-skg / intermediate_tables_migration_quality	Sin unificar	cursos_pers_veh
sisem-skg / intermediate_tables_migration_quality	Sin unificar	repcion

Tabla 1 Tablas en BigQuery Fase 2

Para la fase 3 se realizó el cargue del backup enviado por la Secretaria de Salud el cual fue entregado en un servidor FTP habilitado para este fin en Agosto de 2023. Se realizó la restauración, la cual tardó aproximadamente 3 horas.

Table Name	Engine	Increment	Data Length	Description	Row Count	Partitioned
a_temppersonas	InnoDB	16.872.084	1,1G		16.510.120	[]
activacion_ambulancia	InnoDB	24	16K		17	[]
actualizacion_dea	InnoDB	38	16K		37	[]
actualizacion_personal	InnoDB	4	16K		3	[]
adherencia	InnoDB	388	192K		367	[]
aglomeraciones	InnoDB	5.937	2,5M		4.625	[]
alertas_ips	InnoDB	234	96K		233	[]
amb_ajust_sql	InnoDB	0	16K		95	[]
amb_rep_montar	InnoDB	1	16K		0	[]
ambformatoemerg	InnoDB	22.368	1,5M		14.942	[]
ambformatoeventos	InnoDB	1.639	224K		1.526	[]
ambformatoimpactosocial	InnoDB	18.005	1,5M		18.083	[]
ambulancias	InnoDB	0	1,5M		1.118	[]
ambulancias1	InnoDB	0	144K		382	[]
ambulancias13_junio	InnoDB	0	240K		536	[]
ambulancias2010	InnoDB	0	48K		172	[]
ambulancias2011	InnoDB	0	48K		198	[]
ambulancias_09-05-2018	InnoDB	0	304K		564	[]
ambulancias_1	InnoDB	0	240K		536	[]
ambulancias_2015	InnoDB	0	208K		511	[]
ambulancias_2018	InnoDB	0	336K		567	[]
ambulancias_6sep	InnoDB	0	256K		530	[]
ambulancias_agosto	InnoDB	0	256K		530	[]
ambulancias_copias	InnoDB	0	1,5M		2.065	[]
ambulancias_corferias	InnoDB	0	80K		543	[]
ambulancias_historico	InnoDB	13.338	5,5M		12.945	[]
ambulancias_nuevas1	InnoDB	0	256K		530	[]
ambulancias_reps	InnoDB	9.455	1,5M		1.840	[]

Ilustración 1 Base restaurada en el servidor de pruebas

En esta base de datos restaurada en los servidores de GCP se identifican 513 tablas. Al validar esta información contra el archivo de unificación generado por la Secretaria de Salud se identifica una diferencia ya que en este se lista un total de 572 tablas. De estas, en la nueva propuesta realizada por el grupo de la SDS, 290 no están agrupadas y las 282 tablas restantes se agrupan en 41 grupos.

En Marzo de 2024 se recibe desde OneDrive de la Secretaria de Salud un nuevo backup el cual es subido a un BUCKET en GCP que queda habilitado para poder colocar los nuevos backup que se generen. Para esta tarea se automatiza un proceso que toma el backup de la ruta del bucket y lo restaura en la base de datos MySQL dispuesta para tal fin la cual siempre se llamará **quejas**; en este caso el total de tablas cargadas fue de 586.

Name	Size	Type	Created	Storage class
quejas.sql	29 GB	application/octet-stream	May 8, 2024, 9:22:00 PM	Standard

Ilustración 2 BUCKET habilitado en GCP para el restaure de BACKUP

Table Name	Engine	InnoDB	Data Length	Description	Row Count	Partitioned
a_tempnumpersonas	InnoDB	1	1G		17,514,805	0
activacion_ambulancia	InnoDB	1	24	16K	17	0
actualizacion_dea	InnoDB	1	76	16K	75	0
actualizacion_personal	InnoDB	1	50	16K	49	0
adherencia	InnoDB	1	1,387	1,5M	1,328	0
aglomeraciones	InnoDB	1	5,990	2,5M	4,579	0
alertas_ips	InnoDB	1	667	288K	666	0
amb_ajust_sql	InnoDB	1	0	16K	95	0
amb_reps_montar	InnoDB	1	1	16K	0	0
ambformatoemerg	InnoDB	1	22,531	1,5M	16,180	0
ambformeventos	InnoDB	1	1,640	224K	1,527	0
ambformimpactosocial	InnoDB	1	18,180	1,5M	18,297	0
ambulancias	InnoDB	1	0	1,5M	1,091	0
ambulancias1	InnoDB	1	0	144K	382	0
ambulancias13_junio	InnoDB	1	0	240K	536	0
ambulancias2010	InnoDB	1	0	48K	172	0
ambulancias2011	InnoDB	1	0	48K	198	0
ambulancias_09-05-2018	InnoDB	1	0	304K	564	0
ambulancias_1	InnoDB	1	0	240K	536	0
ambulancias_2015	InnoDB	1	0	208K	511	0
ambulancias_2018	InnoDB	1	0	336K	567	0
ambulancias_6sep	InnoDB	1	0	256K	530	0
ambulancias_agosto	InnoDB	1	0	256K	530	0
ambulancias_copias	InnoDB	1	0	1,5M	1,996	0
ambulancias_corferias	InnoDB	1	0	80K	543	0
ambulancias_historico	InnoDB	1	14,184	5,5M	13,862	0
ambulancias_nuevas1	InnoDB	1	0	256K	530	0
ambulancias_reps	InnoDB	1	10,206	384K	890	0
ambulanciasfs1	InnoDB	1	324,890	192K	833	0
ambulanciasfs2010	InnoDB	1	51,069	7,5M	51,088	0
ambulanciasfs2011	InnoDB	1	72,113	10M	70,307	0
ambulanciasfs2012	InnoDB	1	40,205	6,5M	39,440	0
ambulanciasfs2013	InnoDB	1	242,464	10M	58,030	0
ambulanciasfs2014	InnoDB	1	323,594	15M	79,693	0
ambulanciasfs2015	InnoDB	1	425,658	18M	100,977	0
ambulanciasfs_2016	InnoDB	1	498,728	15M	73,188	0
ambulanciasfs_agosto	InnoDB	1	5	16K	3	0
ambulanciasfs_nuevo	InnoDB	1	514,923	105M	510,417	0
ambulanciasfs_nuevo_2016	InnoDB	1	622,386	10M	44,247	0

Ilustración 3 Base restaurada en el servidor de pruebas 2024

Las tablas no unificadas de acuerdo con la propuesta realizada por la secretaria de salud son:

Tablas no Unificadas
a_tempnumpersonas
activacion_ambulancia
actualizacion_dea
actualizacion_personal
adherencia
aglomeraciones
alertas_ips
amb_ajust_sql
amb_reps_montar
ambformatoemerg
ambformeventos
ambformimpactosocial
ambulancias_reps
amed_eps
amed_paciente
amed_paciente_reg

asesorias_covid_old
asignaciones
asistencia_clinica
barrios
bases
bd_supersalud
bk_archivos_avl
bk_archivos_ficheros
bk_archivos_quejas
bk_archivos_radio
bk_novedades
bk_turnos_centro_operativo
camasred
camillas
camillas_fs
candidatos
cantidad_sgtos_covid
cantidad_sgtos_referencia
capacitaciones
capacitaciones_fecha
carrocamillero
carrocamillero2012
causalidad
cda_ministerio
circular012
ciudades_uci
clase_evento
clasific_novedad
cod_incidientes_p1
comentarios_referencia_uci
conglomerados_amed
conteo de usuarios por empresa
contratos_aph
convocatorias
coordiadores_aph
correo_activo
cups
cursos_pers_veh
datos
datos_dias_uci
datos_dias_uci_covid

datos_hospitales
datos_qr
datos_uci
datos_uci_cerrados
departamentos
despachos_amed_cargue
despachos_amed_encuestas
despachos_amed_no_contacto
despachos_cambios_amed
desplazamientos
dia_cierre
directorio
disponibilidad
disponibilidad_uci
disponibles
dpais
dptos_colombia
eapb
electivas
emergencia_sanitaria
empre_pub_subentidad
empresas_ctos
empresas_publicas
empresasaph
empresasaph_10_01_2017
empresasaphhistorico
encuentas
encuentas_linea_123
encuestas
encuestas_covid19
encuestas_inicial
encuestas_linea_123
encuestas_sem
enfermero
entidad_expide_soat
entrega_gestor_inf
entrega_psicologo
eps_traslados
escenario_evento
estados_moviles_gestores
estados_recursos_bk

estados_vehiculos_covid
estamentos
evaluacion_dea
evaluacion_despacho
evaluacion_enfermeros
evaluacion_medico_regulador
evaluacion_psicologia
evaluacion_referencia
evaluacion_tarm
evento_aph
fechas_ejecucion_scripts
formatoemerg
formeventos
formimpactosocial
fs_descuentos
fservicio
fuera_servicio_totales1
gestion_conglomerados_amed
gestion_liberacion_camillas_eps
guiones_emergencias
historico_avl
historico_radio
hojas_de_vida
horas
inc_smental
incidentes_sem
incidentes_sm
incidentes_uci
indicadores_amed
indicadores_emergencias
indicadores_ips_amed
indicadores_ips_amed_cambios
indicadores_serv_urgencias
informacion_tripulantes_vacuna_covid
infraccion_medica
ingresonovedades
inscritos_curso_sem
institucion
interv_smental
intervenciones_sm
intoxicados

inventario_dea
ips_amed
ips_radios
ips_traslados
liberacion_gestor
liberacion_uci
liberaciones
llamada_quemados
llamantes_sem
localidad
lugarinc
marca_veh_min
medicoreg
medicoregulador
mes_facturar
mesa_ayuda_sem
municipios
nedocs
nombres_indicadores
notas_inc
nov_thumano_covid19
noved_turno_gestor_inf
novedad_aph
novedades
novedades_despachos_corferias
novedades_despachos_referencia_uci
novedades_disponibilidad_uci
novedades_enfermeros
novedades_gestor
novedades_instituciones
novedades_noconformes
novedades_psicologia
novedades_psicologo
novedades_referencia
novedades_reg
novedades_retencion_sem
novedades_seguimientos
novedades_th
novedades_turno
novedades_turno_psiq
novedades_turnos_medicos

novedadesamb
ordendetrabajo
pac_interv_smental
pac_smental
pacformatoemerg
pacformeventos
pacformimpactosocial
pacientes_despachos2015
pacientes_equipo_comando
pacientes_seguimientos
pacientes_sem
pacientes_sm
pacientes_uci
pacientes_urg
pagad_reg_aseg
países
pdf_dea
pdf_quejas
pdf_referencia
pdf_uci
personal_dea
personal_fs
personal_veh_emerg
ponderaciones_adherencia
ponderaciones_encuestas_despachos
ponderaciones_encuestas_enfermeros
ponderaciones_encuestas_medicos
ponderaciones_encuestas_piscologia
ponderaciones_encuestas_referencia
ponderaciones_encuestas_tarm
prestadores_seg_pac
productividad_amed
profesiones_capacitacion
psicologia
ptes_amed
ptes_intervenciones_sm
ptes_llamadas
repcion
reclamos
referencia
referencia_covid

referencia_uci
referencia_uci_reabiertos
reg_asegurador
regimen
reporte_incidentes
representantes
respuestas_entrevistas
resumen_conglomerados
revision_ambulancias
rpc_adultos
rpc_ninos
sci201a
sci201resumenincidente
sci206ambincidente
sci206ambrequeridas
sci206anexo
sci206estacionesasimedica
sci206final
sci206hospitales
sci211solicitudrecurso
seguimiento_bases
seguimientos_crue
seguimientos_pacientes_sem
seguimientos_referencia_covid
seguimientos_sm
servicio_ocurrencia
servicios
servicios_urgencias
sgtos_ips
sgtos_revision_ambulancias
simulacros_dea
siras_tabla
solicitud_emblema_md
solicitud_tm
sucesos_seg_paciente
supersalud
tablamedicos
talento
tbl_entrega_turnos
tbl_novedades_turnos
tbl_turnos

tbl_turnos_medicos
tecnicos
temporal_bodega
temporal_quejas
tiempos_opetativos_liq
tipo_ent_resp
tipo_novedad
tipo_resp_pago
tipoinc
tras_sec_referencia_covid
traslados_comando
traslados_secundarios
turno_conductores
turno_despachador
turno_gestor
turno_medico_reg
turno_referencia
turno_tarm
turnos_centro_operativo
uso_dea
usuarios_dea
vehiculos_amed
vehiculos_amed_fs
vehiculos_salud_mental
vehiculos_sm_fs
visita_ips
visita_servicios_ips
visitas_gestores_uci

Tabla 2 Tablas no unificadas

Las tablas unificadas de acuerdo con la propuesta realizada por la secretaria de salud son:

Unificación	Tablas
Ambulancias	ambulancias
	ambulancias_09-05-2018
	ambulancias_1
	ambulancias_2015
	ambulancias_2018
	ambulancias_6sep
	ambulancias_agosto
	ambulancias_copias
	ambulancias_corferias

	ambulancias_nuevas1
	ambulancias1
	ambulancias13_junio
	ambulancias2010
	ambulancias2011
	bk_ambulancias
Ambulancias FS	ambulanciasfs_2016
	ambulanciasfs_agosto
	ambulanciasfs_nuevo
	ambulanciasfs_nuevo_2016
	ambulanciasfs_nuevo_2018
	ambulanciasfs_nuevo_ajustado
	ambulanciasfs1
	ambulanciasfs2010
	ambulanciasfs2011
	ambulanciasfs2012
	ambulanciasfs2013
	ambulanciasfs2014
	ambulanciasfs2015
	bk_ambulanciasfs_nuevo
Ambulancias historico	ambulancias_historico
	ambulanciashistorico
amed_atenciones	amed_atenc_primer_sem
	amed_atenc_primer_sem_2020
	amed_atenc_segund_sem
	amed_atenc_segund_sem_2020
	amed_atenc_tercer_trim
	amed_atenciones
	amed_atenciones_2020
	amed_atenciones_cambios
	amed_atenciones_carga_control
	amed_atenciones_carga_datos
	amed_atenciones_carga_datos_prueba
	amed_atenciones_carga_datos_prueba_1
	amed_atenciones_carga_datos1
	amed_atenciones_carga_error
	amed_atenciones_carga_error_prueba
	amed_atenciones_historial
	amed_atenciones_prueba
amed_solicitud	amed_solicitud
	amed_solicitud_2020

	amed_solicitud_cambios
	amed_solicitud_historico
amed_temporal	amed_temporal
	amed_temporal_dos
archivos	apertura_sistema
	archivos_ficheros
	archivos_mantenimiento
	archivos_mesa_ayuda_sem
	archivos_novedades
	archivos_novedades_retencion
	archivos_quejas
	archivos_referencia
	archivos_smartdata
	archivos_ucis
asesorias_amed	asesorias_amed
	asesorias_amed_2021
asesorias_covid	asesorias_covid
	asesorias_covid_fase_ii
bases_prestadores	bases_prestadores
	bases_prestadores_historico
cie10	cie10
	cie10_psicologos
	cie10_resumido
	cie10_resumido_anterior
	cie10_resumido_principal
	cie10_resumido_samuel
	cie10_traslados
	cie101
comentarios_ref	comentarios_ref
	comentarios_ref_2012-2015
	comentarios_ref_2016
despachos	despachos
	despachos_moviles
	despachos_moviles_abri15
	despachos_moviles_dic
	despachos_moviles_enero2015
	despachos_moviles_junio2015
	despachos_moviles_mar2015
	despachos_moviles_mayo2015
	despachos_moviles_nov
	despachos_moviles_oct2014

	despachos_moviles_sept_2014
	despachos_moviles_sidcrue
despachos_amed	despachos_amed
	despachos_amed_2021
despachos_amed_temporal	despachos_amed_temporal
	despachos_amed_temporal_2021
despachos_procad	despachos_procad
	despachos_procad_2016
	despachos_procad_219
	despachos_procad_abril
	despachos_procad_mayo
encuestas_efectivas_sm	encuestas_efectivas_sm
	encuestas_efectivas_sm1
hospitales	hospitales
	hospitales_bk
	hospitales_colombia
	hospitales_uci
	hospitales_upz
	hospitales1
	hospitalesdds
id_procad	id_procad
	id_procad_sep
incidentes	incidentes
	incidentes_2015
	incidentes_anterior
	incidentes_bk
	incidentes_bk_2017
	incidentes_bk_2018
	incidentes_bk_26-04-17
	incidentes_cop_01_sep
incidentes_tarm	incidentes_tarm
	incidentes_tarm_agosto17
	incidentes_tarm_oct_1
	incidentes_tarm_oct23
	incidentes_tarm_sept2017
inscritos_curso	inscritos_curso
	inscritos_curso_1
	inscritos_curso_12-13-2019
	inscritos_curso_bk20220915
	inscritos_curso-15-05-18
inventarioequipos	inventarioequipos

	inventarioequiposdebaja
pacientes	pacientes
	pacientes_antes
	pacientes_bk
	pacientes_bk_06
	pacientes_bk_2017
	pacientes_bk_2018
	pacientes_bk_26-04-17
	pacientes_bk_estructura
pagadores	pagadores
	pagadores_1
	pagadores_ant
	pagadores_bk_01_09_2017
	pagadores_desp
	pagadores_desp_jaime
	pagadores_desp3
prestadores_reps	prestadores_reps
	prestadores_reps_historico
quejas	quejas
	quejas_2018
	quejas_bk_2018
quemados	quemados
	quemados_2014
	quemados_2015
	quemados_2016
	quemados_2017
	quemados2012
	quemados2013
refandcont	refandcont
	refandcont_2012-2015
	refandcont_2016
registro_usuarios	registro_usuarios
	registro_usuarios_2015
	registro_usuarios_2016
	registro_usuarios2014
retcamillas	retcamillas
	retcamillas_2014-2015
	retcamillas_nuevo
retcamillas_nuevo_bk	retcamillas_nuevo_bk_2017
	retcamillas_nuevo_bk_2018
	retcamillas_nuevo_bk06

	retcamillas_nuevo_sem
revision_fuera_servicio	revision_fuera_servicio
	revision_fuera_servicio_bk
revision_pendientes	revision_pendientes
	revision_pendientes_bk
revision_recursos	revision_recursos
	revision_recursos_bk
roles_usuario	roles_usuario
	roles_usuario_10_01_2017
seguimientos_covid	seguimientos_asesorias_covid
	seguimientos_asesorias_covid_fase_ii
traslados	bk trasladosp
	bk trasladosp ene
	bk trasladosp feb
	bk trasladosp mrz
	bk traslad_eliminados
	bk trasladosp abr
	bk trasladosp agos
	bk trasladosp dic
	bk trasladosp julio
	bk trasladosp junio
	bk trasladosp may
	bk trasladosp nov
	bk trasladosp oct
	bk trasladosp sept
	consolidado_traslados
	traslado
	traslados
	traslados_febrero2016
	trasladosp
	trasladosp_abril2014
	trasladosp_abril2016
	trasladosp_abril2015
	trasladosp_agosto_2014
	trasladosp_agosto2015
	trasladosp_agosto2016_final
	trasladosp_copia
	trasladosp_dic_2014
	trasladosp_diciembre2015
	trasladosp_ene2014
	trasladosp_ene2015

	trasladospp_enero2016
	trasladospp_feb_2014
	trasladospp_febrero2015
	trasladospp_jul2015
	trasladospp_julio2014
	trasladospp_junio_2014
	trasladospp_junio2015
	trasladospp_junio2016
	trasladospp_marzo2014
	trasladospp_marzo2015
	trasladospp_mayo2014
	trasladospp_mayo2015
	trasladospp_mayo2016
	trasladospp_nov2015
	trasladospp_noviembre2014
	trasladospp_oct_2015
	trasladospp_octu2014
	trasladospp_sep_2014
	trasladospp_sept2015
	trasladospp_septiembre2016_final
	trasladospp1
	trasladospp2010
	trasladospp2011
	trasladosppcopia1
traslados_financieros	traslados_financieros
	traslados_financieros_piloto
usuarios	usuarios
	usuarios_10_01_2017
	usuarios_20dic16
	usuarios_ant
	usuarios_fapp
	usuarios_historico
	usuarios_idiger
	usuarios19feb
	usuarios2
	usuarioscapacitaciones123
	usuarioemerg
	usuariosprocad
	usuariosrec
	usuariosds
usuarios_capacitaciones	usuarios_capacitaciones

	usuarios_capacitaciones_historico
	usuarios_capacitacitados_final
valoracion_tm	valoracion_tm
	valoracion_tm_1
	valoracion_tm_sds
	valoracion_tm1
vdcamillas	vdcamillas
	vdcamillas_abril2014
	vdcamillas_agosto_2014
	vdcamillas_enero2014
	vdcamillas_febrero2014
	vdcamillas_julio2014
	vdcamillas_junio2014
	vdcamillas_marzo2014
	vdcamillas_mayo2014
	vdcamillas_noviembre2014
	vdcamillas_octubre2014
	vdcamillas_septiembre2014
	vdcamillas1
vehiculofs	vehiculofs
	vehiculofs2014

Tabla 3 Tablas unificadas

4.2. Auditoría y Análisis de Calidad

Para esta fase se utilizan las tecnologías propuestas de Python y Pandas, se despliega Pandas-Profiling herramienta que permite determinar el rango de cada una de las características, identificar los tipos de datos y calcular el porcentaje de valores nulos, en cada una de esas características.

La auditoría es el diagnóstico que permite garantizar que toda la información que ingresa a nuestra base de datos cumpla las condiciones necesarias para el correcto funcionamiento de las herramientas que los utilizan.

Una vez realizada la identificación de la estructura de datos, el siguiente paso es elegir las tablas que se consideren de uso más frecuente, estas son la muestra a la cual se les aplica el data profiling.

Las tablas seleccionadas son: Referencia, intoxicados, quemados, inc_smental, traslados, ambulancias, usuarios_capacitacitados_final.

4.2.1. Profiling Referencia

En la tabla referencia se registra la información de la gestión realizada a los pacientes del proceso de Referencia y Contra Referencia. Seguimiento al procedimiento solicitudes hechas al CRUE, para pacientes en otros municipios.

En el overview (vista general) del 2023 se identifican 60 campos y 2229 filas, con 3 campos de tipo numérico, 8 campos tipo fecha; de los cuales no son soportados 5 por el tipo de formato y 49 campos tipo carácter, la diferencia con la información del 2024 es que se agregaron dos campos naturaleza y entidad_territorial de tipo varchar(50).

Overview	Alerts 94	Reproduction
Dataset statistics		Variable types
Number of variables	60	Numeric 3
Number of observations	2229	DateTime 3
Missing cells	0	Categorical 49
Missing cells (%)	0.0%	Unsupported 5
Duplicate rows	0	
Duplicate rows (%)	0.0%	
Total size in memory	1.0 MiB	
Average record size in memory	480.1 B	

Ilustración 4 Overview referencias 2023

Overview		
Overview	Alerts 33	Reproduction
Dataset statistics		Variable types
Number of variables	62	Numeric 3
Number of observations	3167	DateTime 2
Missing cells	0	Text 29
Missing cells (%)	0.0%	Categorical 22
Duplicate rows	0	Unsupported 6
Duplicate rows (%)	0.0%	
Total size in memory	1.5 MiB	
Average record size in memory	496.0 B	

Ilustración 5 Overview referencias 2024

Ilustración 6 Vista parcial de los campos tabla referencia

nombre_hospital	Distinct 317	E.S.E. HOSPITAL DEPARTAMENTAL S... 128
Categorical	Distinct (%) 14.2%	HOSPITAL REGIONAL DE LA ORINOQ... 120
HIGH CARDINALITY	Missing 0	ESE HOSPITAL SAN VICENTE DE AR... 88
	Missing (%) 0.0%	HOSPITAL M.E. PATARROYO IPS S.A.S. 81
	Memory size 17.5 KiB	HOSPITAL JOSE CAYETANO VASQUEZ 75
		Other values (312) 1737

Ilustración 7 Campo nombre hospital 2023

Ilustración 8 Campo nombre hospital 2024

Los campos tipo_evento, tipodoc, edad, servicio_pte presentan la característica de alta correlación, lo que implica la posible existencia de una relación entre los valores que pueden tomar estos campos con relación a otras características de la fuente de datos.

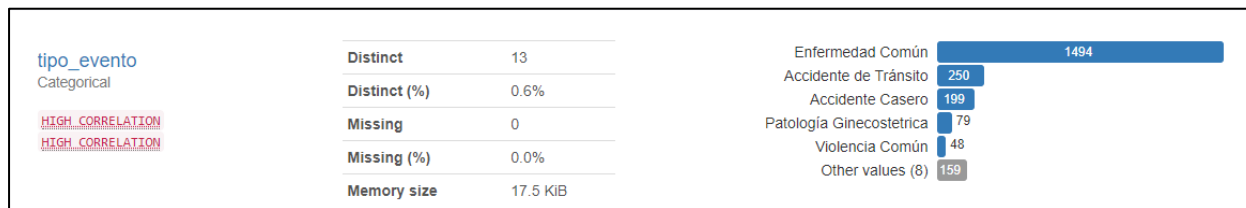


Ilustración 9 campo tipo_evento 2023

En el 2024 el dato se identifica un IMBALANCE(desequilibrado) que indica un dato con muchos mas ejemplos que los demás, esto pasa con el valor Enfermedad común



Ilustración 10 campo tipo_evento 2024

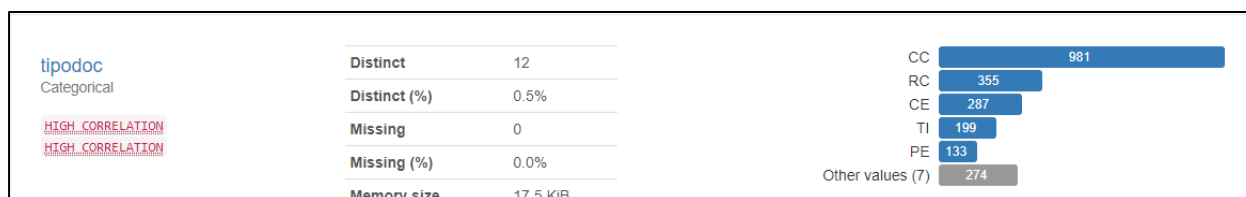


Ilustración 11 campo tipodoc 2023

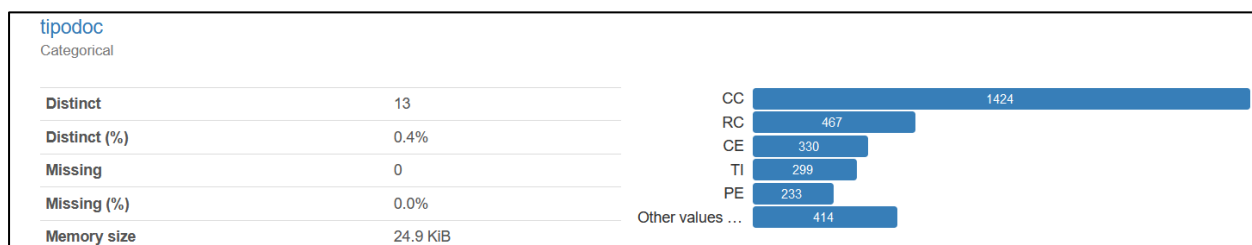


Ilustración 12 campo tipodoc 2024

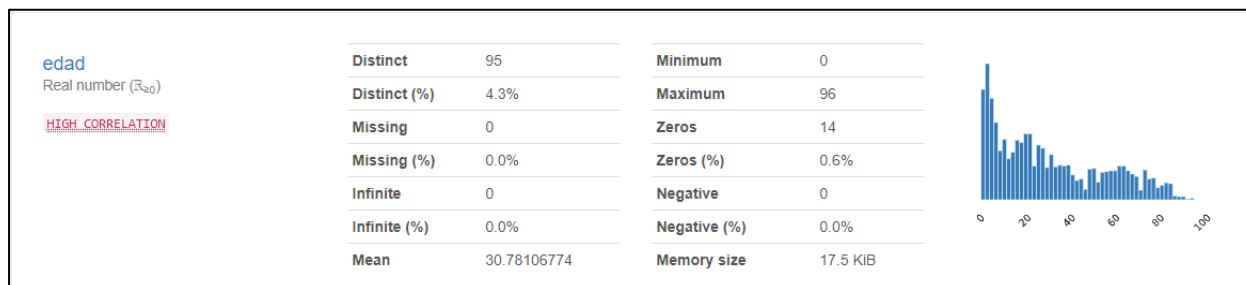


Ilustración 14 Campo edad 2024

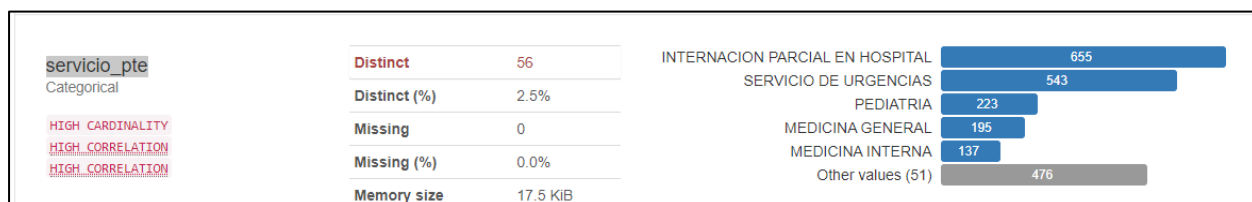


Ilustración 16 campo servicio pte 2023



Tabla donde se registran los datos de las personas que sufren envenenamiento por bebidas alcohólicas adulteradas u otras sustancias.

En el overview (vista general) de la base de 2023 se identifican 16 campos y 31 filas, con 3 campos de tipo numérico, 1 campos tipo fecha y 12 campos tipo carácter. En el 2024 se encuentran 23 campos, la diferencia está en los campos correo_ele, estado, usuario, usuario cierre, fecha cierre, localidad y fecha captura. A su vez los registros aumentaron a 56.

Overview

Alerts 38

Reproduction

Dataset statistics

Number of variables	16
Number of observations	31
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	4.0 KiB
Average record size in memory	132.1 B

Variable types

Numeric	3
Date Time	1
Categorical	12

Ilustración 17 Overview intoxicados 2023

Overview

Alerts 25

Reproduction

Dataset statistics

Number of variables	23
Number of observations	56
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	10.2 KiB
Average record size in memory	186.3 B

Variable types

Numeric	3
DateTime	1
Text	4
Categorical	13
Unsupported	2

Ilustración 18 Overview intoxicados 2024

Columns	Column Name	#	Data Type	Not Null
Constraints	numincidente	1	int(10) unsigned	[v]
Foreign Keys	fechainc	2	datetime	[v]
References	nompaciente	3	varchar(100)	[v]
Triggers	edad	4	int(10) unsigned	[v]
Indexes	unidad	5	varchar(45)	[v]
Partitions	genero	6	varchar(45)	[v]
Statistics	direccion	7	varchar(200)	[v]
DDL	telefono	8	int(10) unsigned	[v]
Virtual	incidente	9	varchar(80)	[v]
	causainc	10	varchar(45)	[v]
	observaciones	11	varchar(500)	[v]
	segsocial	12	varchar(50)	[v]
	pagador	13	varchar(100)	[v]
	traslado_pte	14	varchar(2)	[v]
	ips_destino	15	varchar(100)	[v]
	salud_publica	16	varchar(100)	[v]

Ilustración 19 Vista parcial de los campos tabla intoxicados 2023

Columns	Column Name	#	Data Type	Not Null
	numincidente	1	int(10) unsigned	[v]
	fechainc	2	datetime	[v]
	nompaciente	3	varchar(100)	[v]
	edad	4	int(10) unsigned	[v]
	unidad	5	varchar(45)	[v]
	genero	6	varchar(45)	[v]
	direccion	7	varchar(200)	[v]
	telefono	8	int(10) unsigned	[v]
	incidente	9	varchar(80)	[v]
	causanc	10	varchar(45)	[v]
	observaciones	11	varchar(500)	[v]
	epigrafiac	12	varchar(50)	[v]
	pagador	13	varchar(100)	[v]
	traslado_pte	14	varchar(2)	[v]
	ips_destino	15	varchar(100)	[v]
	salut_publica	16	varchar(100)	[v]
	cometo_ele	17	varchar(2)	[v]
	estado	18	varchar(30)	[v]
	usuario	19	varchar(30)	[v]
	usuario_cierre	20	varchar(30)	[v]
	fecha_cierre	21	date	[v]
	localidad	22	varchar(30)	[v]
	fecha_captura	23	datetime	[v]

Ilustración 20 Vista parcial de los campos tabla intoxicados 2024

El campo edad tiene una alta correlación con otros campos como nombre, incidente o pagador, pero también tiene registros donde su valor es cero. Esto se mantiene para las dos bases de datos.

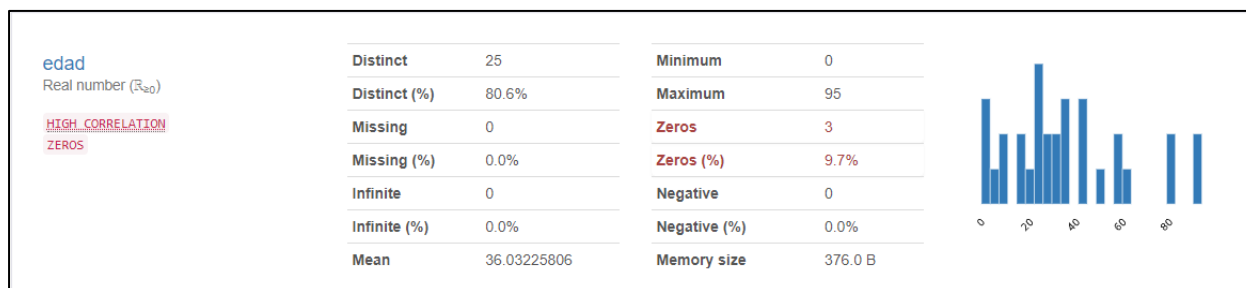


Ilustración 21 Campo edad 2023

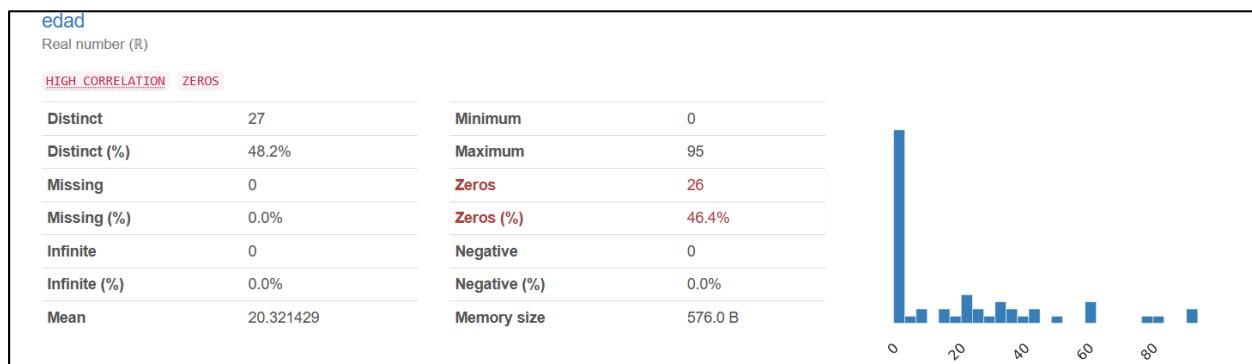


Ilustración 22 Campo edad 2024

El campo incidente también tiene una alta correlación. Así mismo se identifica que la mayoría de los registros no poseen información del incidente. Esto se mantiene para las dos bases de datos.

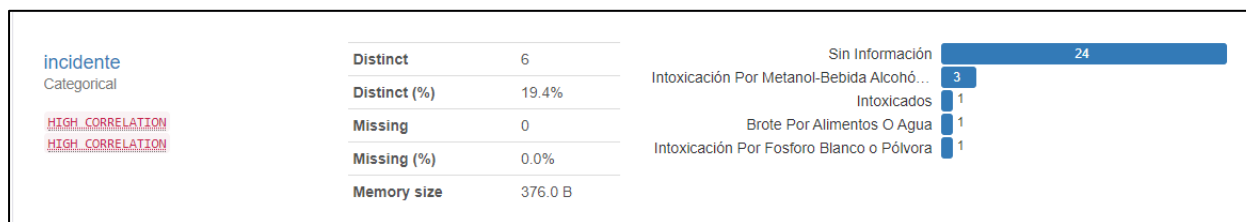


Ilustración 23 campo incidente 2023

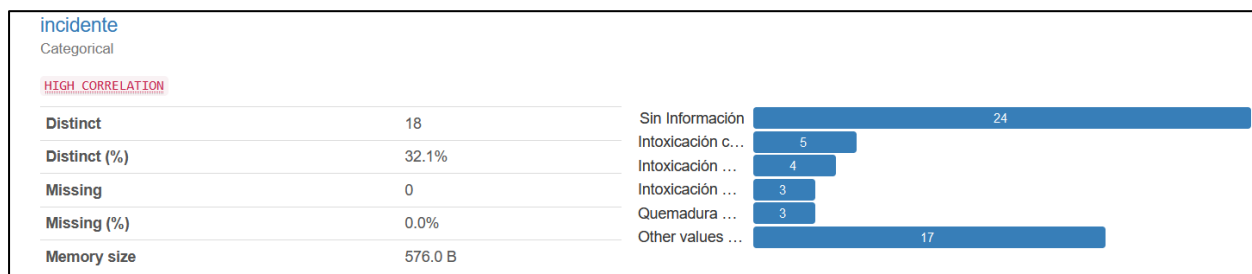


Ilustración 24 campo incidente 2024

4.2.3. Profiling inc_smental

La información Propia de Salud Mental se registra en la tabla *inc_smental* cuando se genera para Premier One.

En el overview (vista general) se identifican 33 campos y 37366 filas, con 1 campo de tipo numérico, 1 campo tipo fecha y 31 campos tipo carácter. No se modifica la tabla en la base de datos del 2024 pero si aumentan las filas a 52180.

Overview Alerts 50 Reproduction	
Dataset statistics	
Number of variables	33
Number of observations	37366
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	9.4 MiB
Average record size in memory	264.0 B
Variable types	
Numeric	1
Date Time	1
Categorical	31

Ilustración 25 Overview inc_smental 2023

Overview

Alerts16

Reproduction

Dataset statistics

Number of variables	33
Number of observations	52180
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	13.1 MiB
Average record size in memory	264.0 B

Variable types

Numeric	1
DateTime	1
Text	18
Categorical	13

Ilustración 26 Overview inc_smental 2024

Los campos prioridad_inc_nuse y código_tip_inc_nuse presentan una alta correlación entre sí, también son campos que comparados con el total de registros presentan un bajo porcentaje de valores distintos lo que implica que pueden ser paramétricos. Ambos campos se mantienen para las dos bases de datos.

prioridad_inc_nuse	Distinct	4	Media	26266
Categorical	Distinct (%)	< 0.1%	Alta	9143
HIGH CORRELATION	Missing	0	Baja	1364
HIGH CORRELATION	Missing (%)	0.0%	Crítica	593
	Memory size	292.0 KiB		

Ilustración 27 Campo prioridad_inc_nuse 2023

prioridad_inc_nuse	Distinct	4	Media	36715
Categorical	Distinct (%)	< 0.1%	Alta	13237
HIGH CORRELATION	Missing	0	Baja	1470
	Missing (%)	0.0%	Crítica	758
	Memory size	407.8 KiB		

Ilustración 28 Campo prioridad_inc_nuse 2024

código_tip_inc_nuse	Distinct	5	941	22240
Categorical	Distinct (%)	< 0.1%	611	5640
HIGH CORRELATION	Missing	0	609	3746
HIGH CORRELATION	Missing (%)	0.0%	918	3515
	Memory size	292.0 KiB	906	2225

Ilustración 29 Campo código_tip_inc_nuse 2023

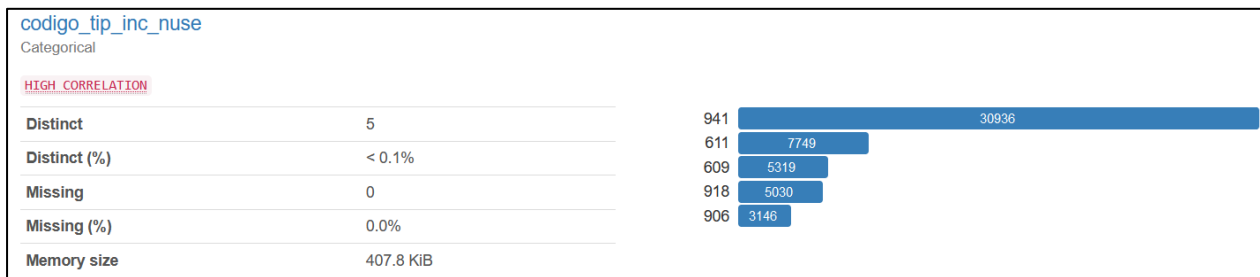


Ilustración 30 Campo codigo_tip_inc_nuse 2024

4.2.4. Profiling traslados

El seguimiento a la evolución clínica de pacientes trasladados de eventos de múltiples víctimas (emergencias y desastres) se registra en la tabla traslados.

En el overview (vista general) se identifican 15 campos y 258682 filas, con 1 campo de tipo numérico, 1 campo tipo fecha y 13 campos tipo carácter. Esto se mantiene para las dos bases de datos aunque en el 2024 aumenta a 259636 filas.

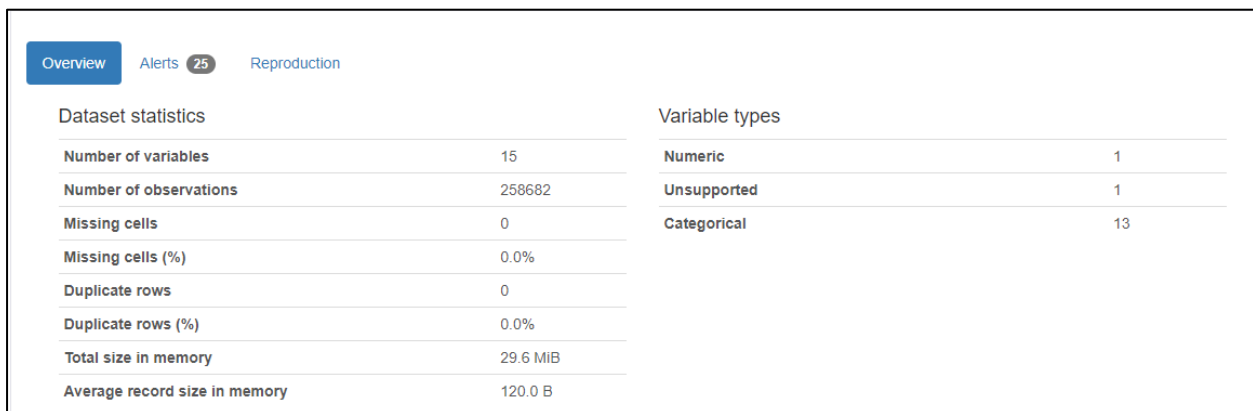


Ilustración 31 Oveview traslados 2023

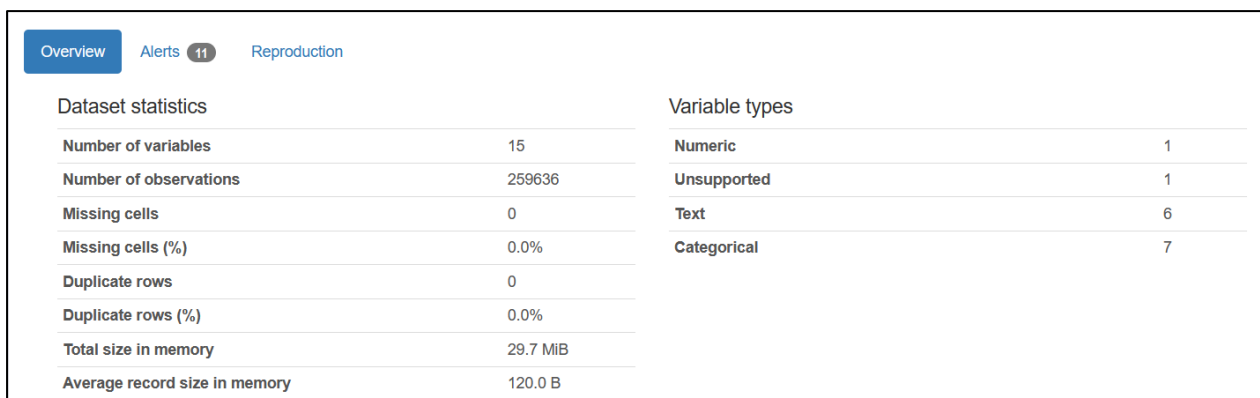


Ilustración 32 Oveview traslados 2024

El campo tipo_traslado a pesar de ser un campo que parece describir la condición del registro presenta datos donde este valor está vacío, lo que sucede también con el campo paciente que a pesar de tener una descripción como NN presenta registros en los que este valor no fue diligenciado. El campo se mantiene para las dos bases de datos.

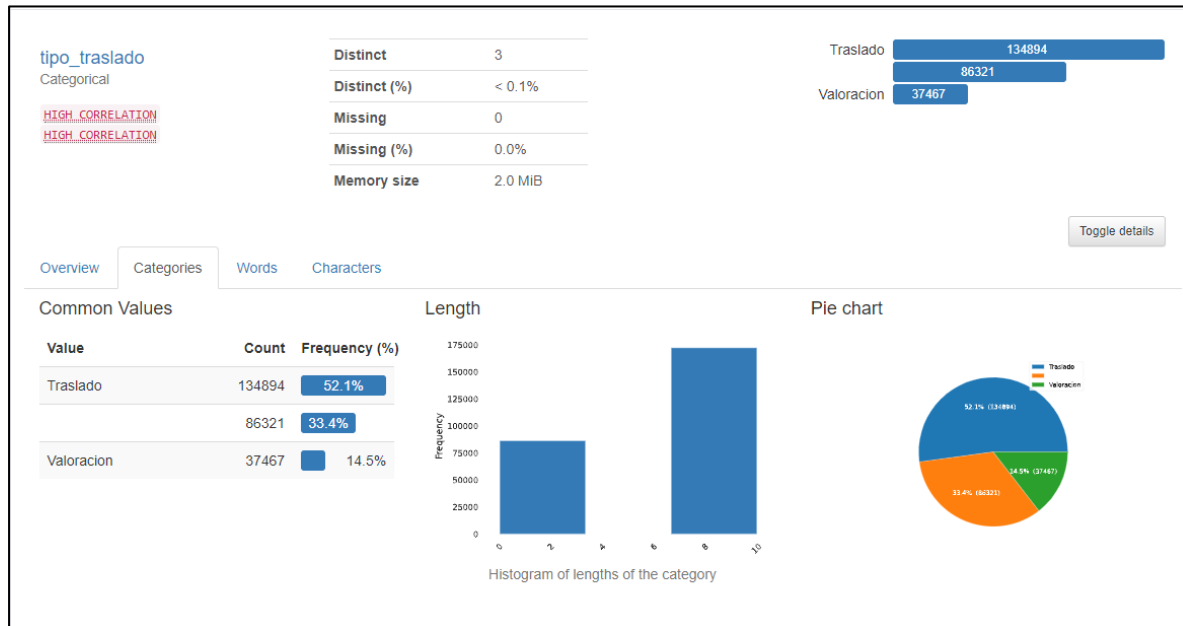


Ilustración 33 Campo tipo_traslado 2023

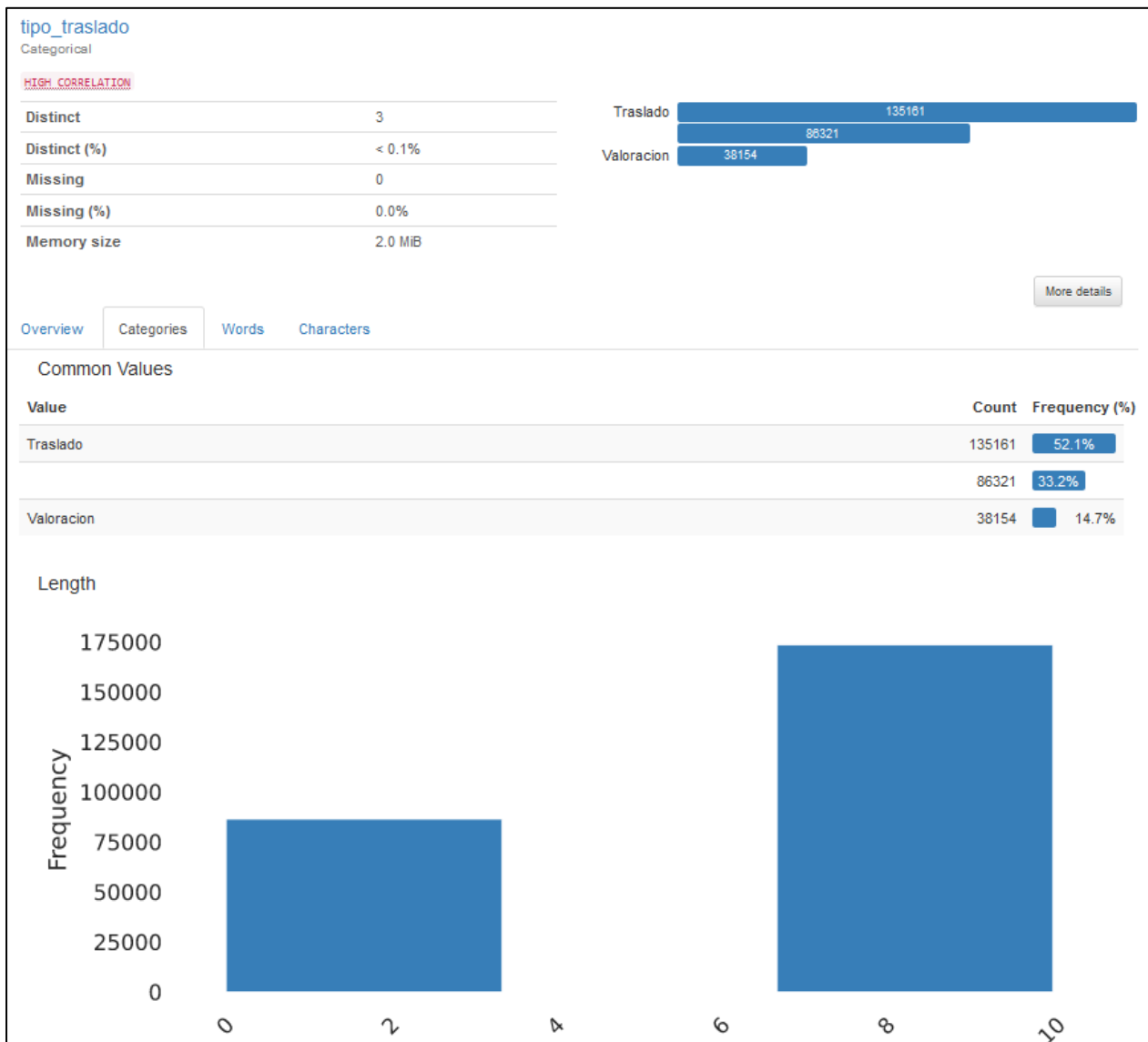


Ilustración 34 Campo tipo_traslado 2024

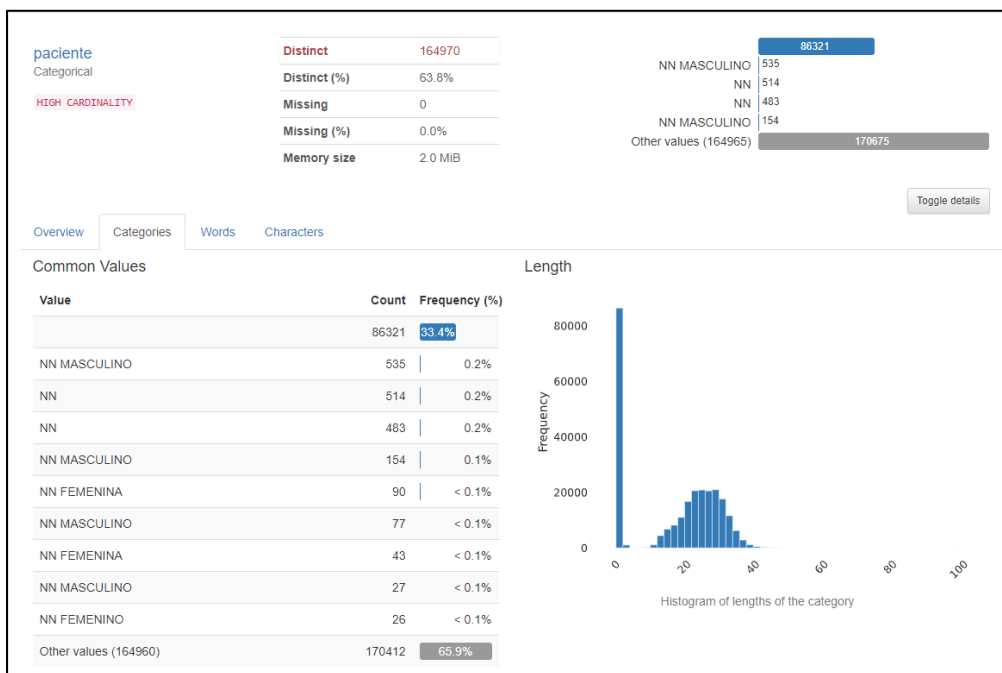


Ilustración 35 campo paciente 2023

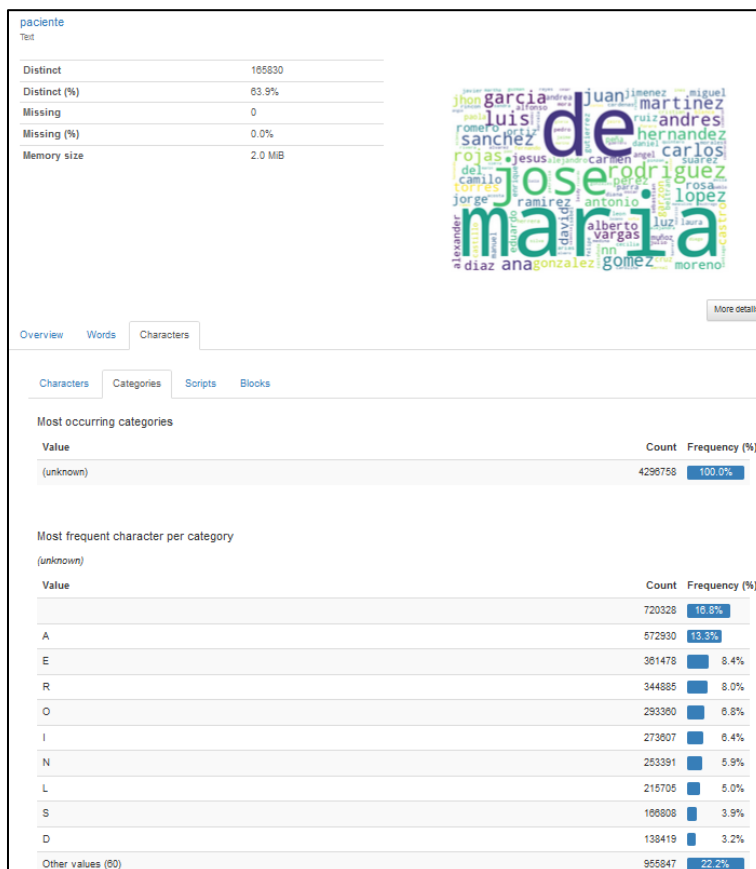


Ilustración 36 campo paciente 2024

4.2.5. Profiling Ambulancias

El seguimiento a la evolución clínica de pacientes trasladados de eventos de múltiples víctimas (emergencias y desastres) se registra en la tabla traslados y se cruza con la información de la tabla ambulancias para identificar los datos de quien realizó el traslado.

En el overview (vista general) se identifican 67 campos y 1151 filas, con 6 campos de tipo numérico, 12 campos tipo fecha los cuales no son soportados por el tipo de formato y 49 campos tipo carácter. En la base de datos de 2024 se identifica un campo nuevo llamado **aph** el número de filas disminuyo a 1124.

Overview

Alerts126

Reproduction

Dataset statistics

Number of variables	67
Number of observations	1151
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	602.6 KiB
Average record size in memory	536.1 B

Variable types

Categorical	49
Unsupported	12
Numeric	6

Ilustración 37 Overview ambulancias 2023

Overview

Alerts 79

Reproduction

Dataset statistics

Number of variables	68
Number of observations	1124
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	597.2 KiB
Average record size in memory	544.1 B

Variable types

Text	16
Unsupported	12
Numeric	6
Categorical	34

Ilustración 38 Overview ambulancias 2024

Es importante tener en cuenta que entre las alertas a validar están las de los campos horas_dia, cilindraje, modelo, valor_hora, potencia e id_base los cuales presentan registros con valores en cero. Los campos se mantienen para las dos bases de datos.

horas_dia	has 996 (86.5%) zeros	Zeros
cilindraje	has 245 (21.3%) zeros	Zeros
modelo	has 231 (20.1%) zeros	Zeros
valor_hora	has 1035 (89.9%) zeros	Zeros
potencia	has 526 (45.7%) zeros	Zeros
id_base	has 381 (33.1%) zeros	Zeros

Ilustración 39 Alerta campos horas_dia, cilindraje, modelo, valor_hora, potencia, id_base 2023

horas_dia	has 950 (84.5%) zeros	Zeros
cilindraje	has 245 (21.8%) zeros	Zeros
modelo	has 230 (20.5%) zeros	Zeros
valor_hora	has 1037 (92.3%) zeros	Zeros
potencia	has 502 (44.7%) zeros	Zeros
id_base	has 372 (33.1%) zeros	Zeros

Ilustración 40 Alerta campos horas_dia, cilindraje, modelo, valor_hora, potencia, id_base 2024

A pesar de que no presenta valores en cero, la tabla si presenta campos vacíos para campos que son parte de la descripción del registro como lo son marca y cilindraje. Los campos se mantienen para las dos bases de datos.

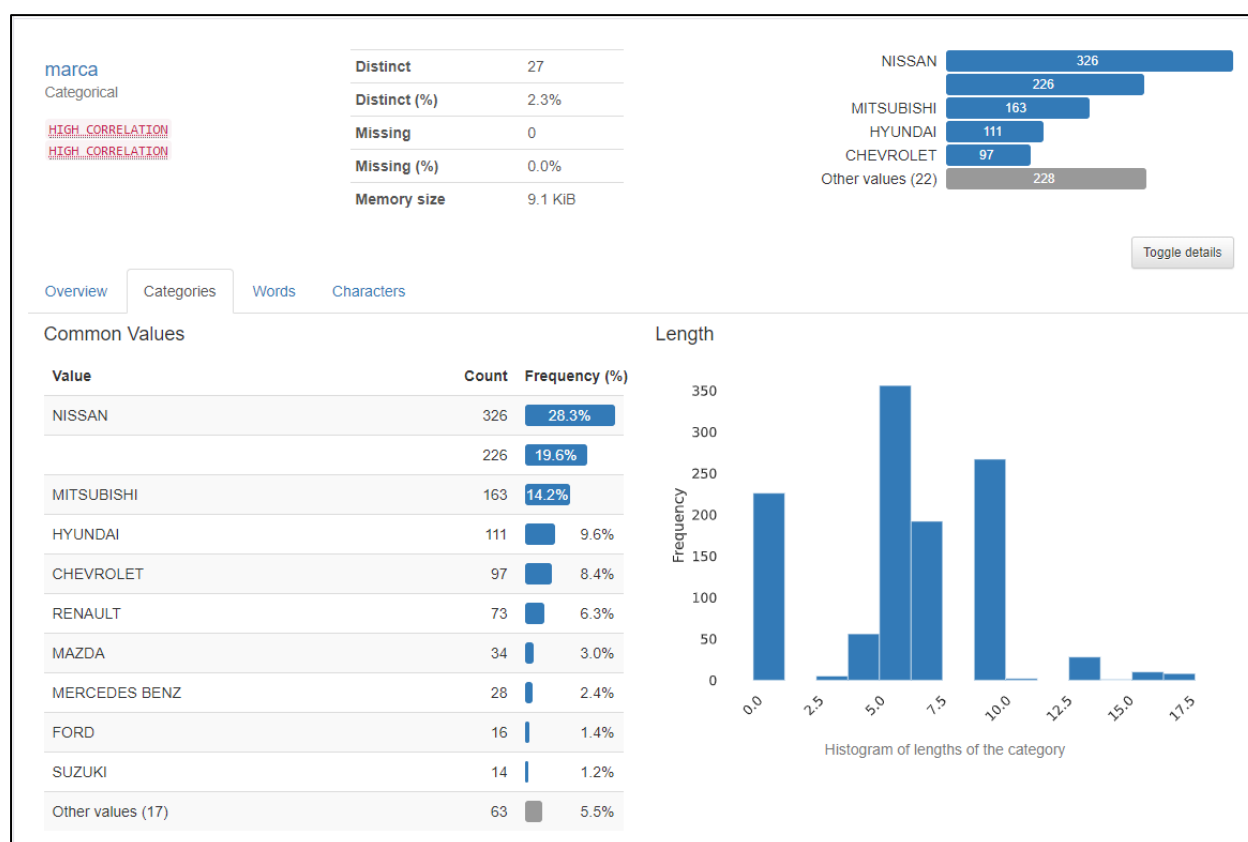


Ilustración 41 Campo marca 2023

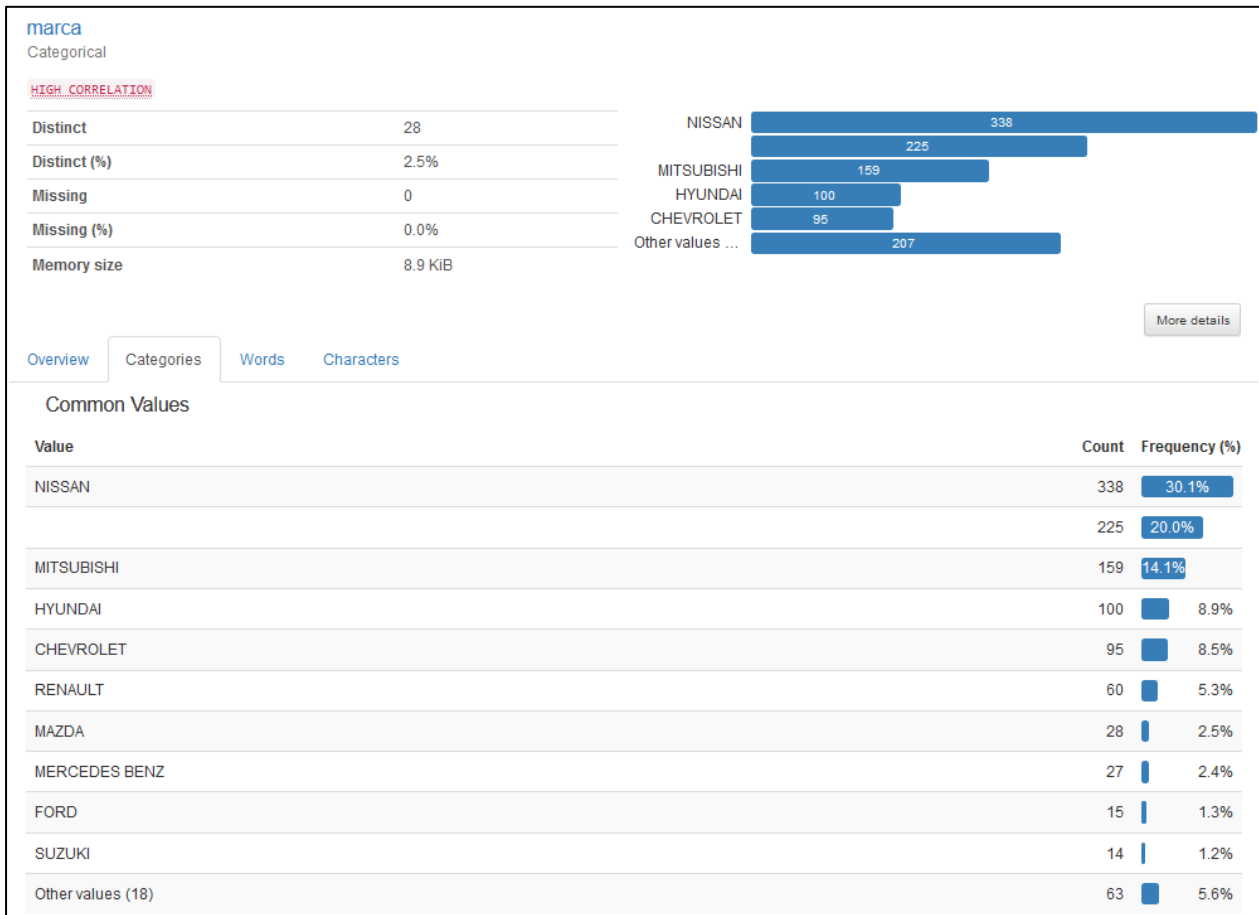


Ilustración 42 Campo marca 2024

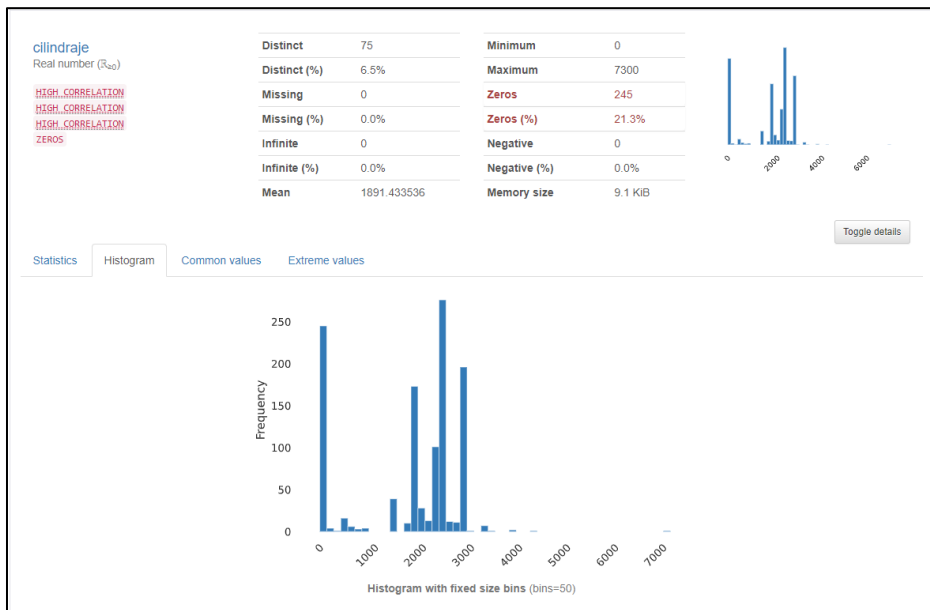


Ilustración 43 Campo cilindraje 2023



Ilustración 44 Campo cilindraje 2024

4.2.6. Profiling usuarios_capacitacitados_final

El registro de los usuarios interesados en participar en los cursos, jornadas, talleres que se dictan en forma presencial o virtual entre otros se realiza en la tabla usuarios_capacitacitados_final.

En el overview (vista general) se identifican 71 campos y 19041 filas, con 6 campos de tipo numérico, 5 campos tipo fecha los cuales no son soportados por el tipo de formato y 60 campos tipo carácter. La información se mantiene similar en las dos bases de datos.

Overview Alerts 121 Reproduction	
Dataset statistics	
Number of variables	71
Number of observations	19041
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	10.3 MiB
Average record size in memory	568.0 B
Variable types	
Numeric	6
Unsupported	5
Categorical	60

Ilustración 45 Overview usuarios_capacitacitados_final 2023

Overview

Alerts 70

Reproduction

Dataset statistics

Number of variables	71
Number of observations	19041
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	10.3 MiB
Average record size in memory	568.0 B

Variable types

Numeric	6
Unsupported	5
Text	23
Categorical	37

Ilustración 46 Overview usuarios_capacitados_final 2024

Aunque no representa campos descriptivos de los registros cabe resaltar que los campos num_via_prin, num_via_gen y num_placa presentan valores en cero. Los campos se mantienen para las dos bases de datos.

num_via_prin has 1850 (9.7%) zeros	Zeros
num_via_gen has 1744 (9.2%) zeros	Zeros
num_placa has 1786 (9.4%) zeros	Zeros

Ilustración 47 Campos num_via_prin, num_via_gen, num_placa 2023

num_via_prin has 1850 (9.7%) zeros	Zeros
num_via_gen has 1744 (9.2%) zeros	Zeros
num_placa has 1786 (9.4%) zeros	Zeros

Ilustración 48 Campos num_via_prin, num_via_gen, num_placa 2024

Si se debe validar y tener en cuenta los valores vacíos presenten un campo de identificación como tipo_id el cual si hace parte de la descripción del registro. El campo se mantiene para las dos bases de datos.

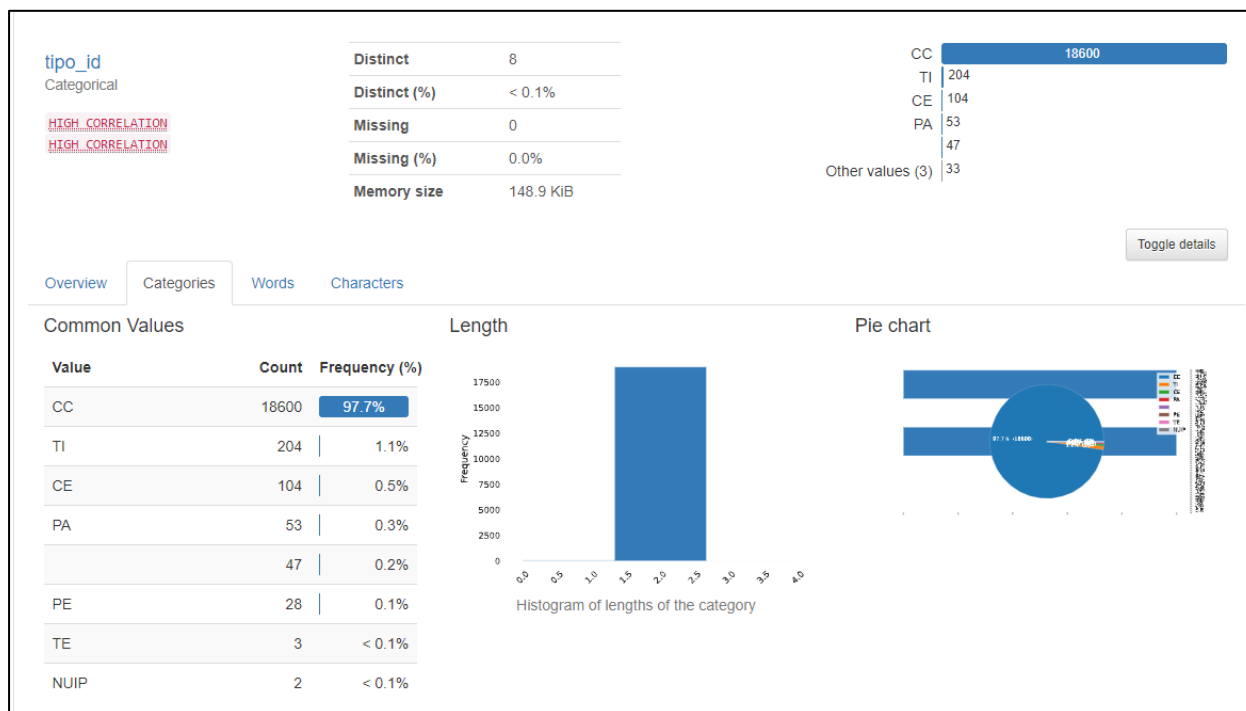


Ilustración 49 Campo tipo_id 2023

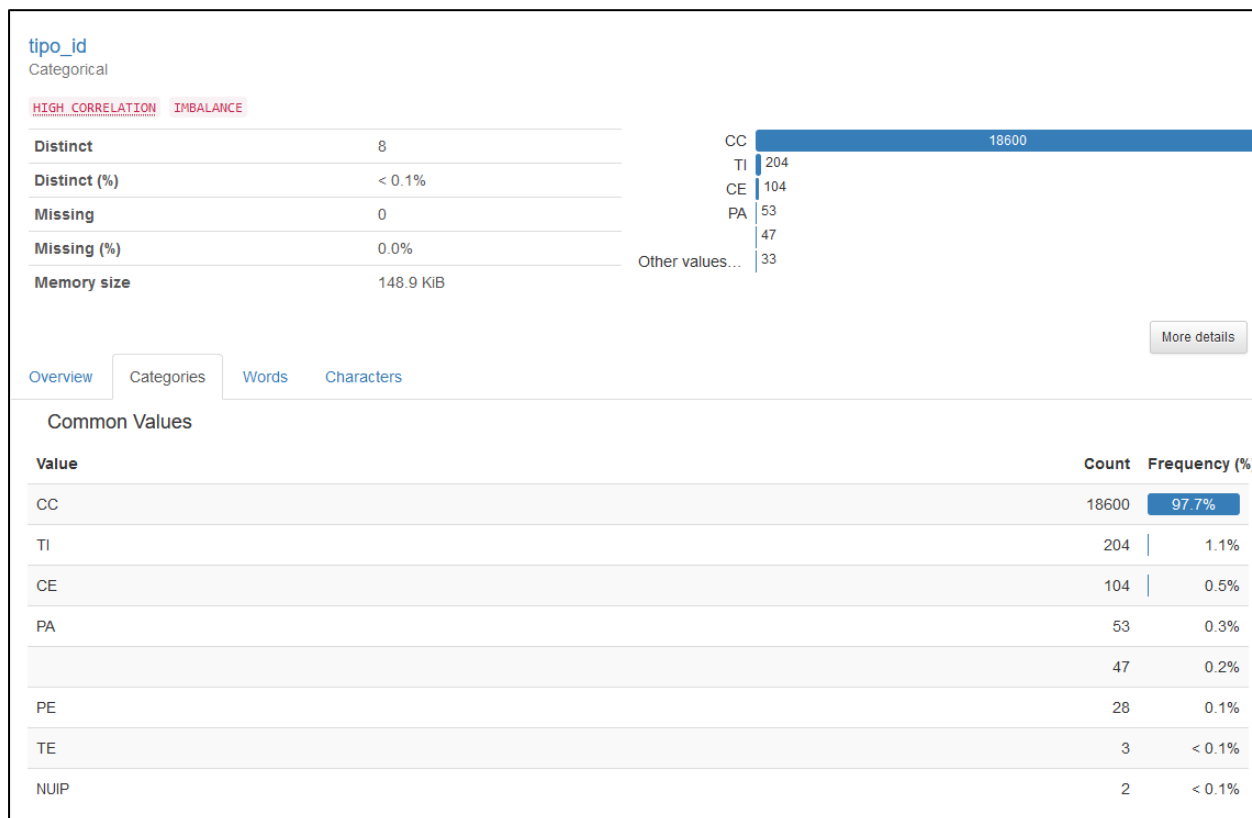


Ilustración 50 Campo tipo_id 2024

4.3. Alistamiento de Infraestructura de Migración

Se realizó la copia de la información desplegada por la Secretaría de Salud en el servidor FTP asignado para esta tarea, el cual tenía un tamaño en disco de 24,3 GB (26.134.806.528 bytes), a una ubicación local destinada para poder realizar el restore en la base de MySQL creada en Google Cloud. Este proceso se completó en aproximadamente 3 horas.

El archivo copiado se toma desde la herramienta DBeaver, para realizar el restore en el servidor de Google Cloud. Para esta tarea se tiene configurado la máquina con IP 104.198.67.167 puerto 3306 y se crea la base de datos quejas2024.

Se selecciona la opción de Database Import la cual nos permite seleccionar el archivo **.sql**, que contiene el restore de la base de datos de la Secretaría de Salud. Este proceso de restore tomó un tiempo aproximado de 12 horas, al terminar el restore se encontraron 513 tablas, una diferencia de 34 comparada con el restore anterior (547 quejas).

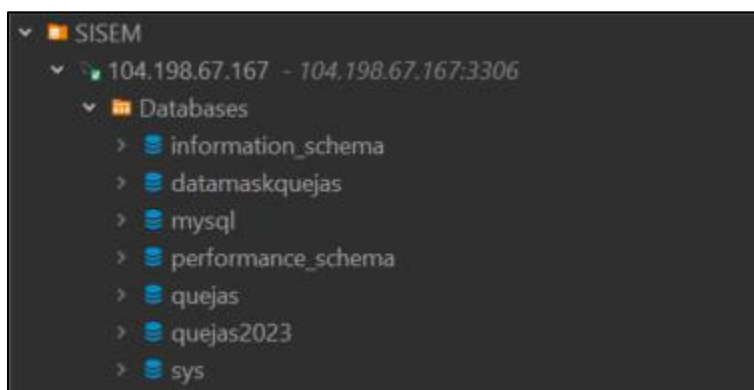


Ilustración 51 Servidor MySQL

En el caso de la información de marzo del 2024 se recibe desde OneDrive de la Secretaría de Salud el cual tenía un tamaño en disco de 29.0 GB (31,148,400,640 bytes) el cual es subido a un BUCKET nombrado **backupsidcrue** en GCP. Este bucket queda habilitado para poder colocar los nuevos backup que se generen. Para esta tarea se automatiza un proceso que toma el backup de la ruta del bucket y lo restaura en la base de datos MySQL dispuesta para tal fin la cual siempre se llamará **quejas**. El restore tomó aproximadamente 2 horas.

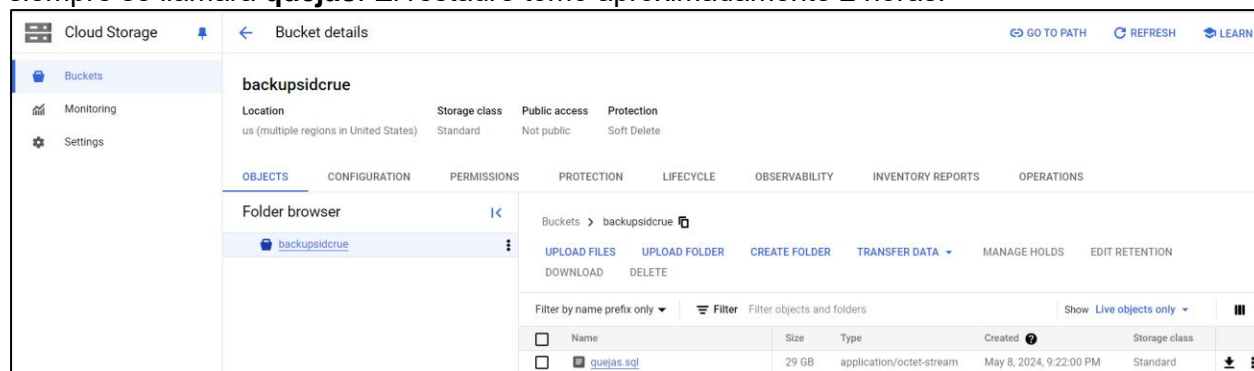


Ilustración 52 Bucket backupsidcrue

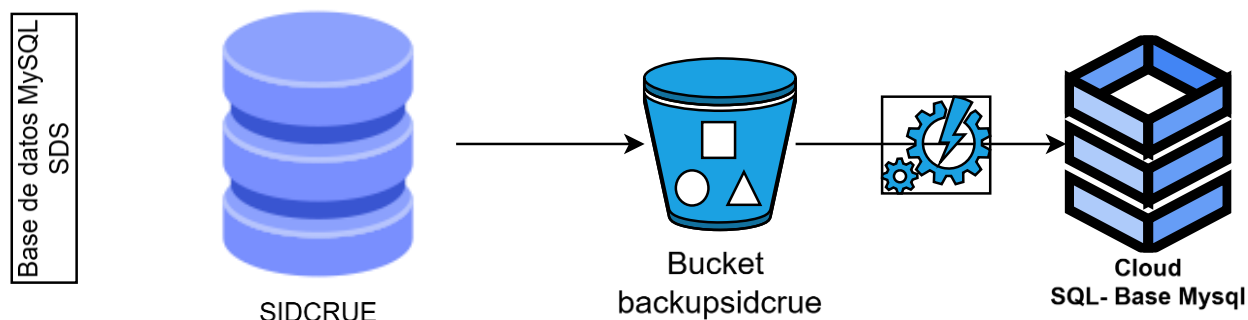


Ilustración 53 -1 Diagrama automatización

4.4. Preparación de los Repositorios Finales e Intermedios

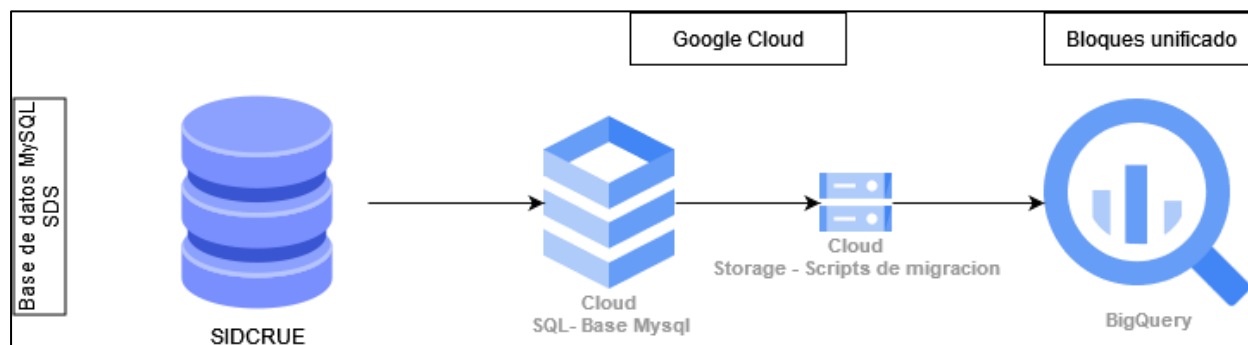


Ilustración 54 distribución de servidores

Réplica SIDCRUE: Para realizar el proceso de migración se realiza una réplica de la base de datos de SIDCRUE. Este proceso se realiza diariamente con un corte nocturno a partir de la información que se cargue en la base de datos actual en MySql.

Base de Datos Limpia Postgres SQL: En esta base de datos se almacena la información una vez se realizan todas las tareas de limpieza y agrupamiento de la base de datos original de SIDCRUE. Adicionalmente a este agrupamiento se realiza la validación y cruce contra las tablas paramétricas que conforman el modelo relacional de esta base de datos organizada.

Se puede identificar los nombres con los que se agrupa en el bucket y las tablas del modelo del SIDCRUE al final de la tabla.

4.5. Construcción de scripts de migración

Para realizar el agrupamiento de la información se realizó un desarrollo en Python el cual toma la lista de tablas a unificar. Se generan archivos PARQUET con toda la data de las tablas y se almacena en un bucket o repositorio, en los servidores de Google Cloud, los cuales pueden ser procesados para extraer la información que se necesita. Las piezas de código se anexan como archivo adjunto.

The screenshot shows the Google Cloud Storage console. The left sidebar has 'Cloud Storage' selected. The main area is titled 'Detalles del bucket' and shows the path 'Depósitos > output_migracion > data_lake > raw'. Below the path are buttons for 'SUBIR ARCHIVOS', 'SUBIR CARPETA', 'CREAR CARPETA', 'TRANSFERIR LOS DATOS', 'ADMINISTRAR CONSERVACIONES', 'DESCARGAR', and 'BORRAR'. A table lists the objects in the bucket, including 'activacion_ambulancia.parquet', 'actualizacion_dea.parquet', 'actualizacion_personal.parquet', 'adherencia.parquet', 'aglomeraciones.parquet', 'alertas_ips.parquet', 'amb_ajust_sql.parquet', 'amb_reps_montar.parquet', and 'ambformatoemerg.parquet'.

Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacenamiento
activacion_ambulancia.parquet	3 KB	application/octet-stream	31 oct 2023 12:15:05	Standard
actualizacion_dea.parquet	7 KB	application/octet-stream	31 oct 2023 12:15:06	Standard
actualizacion_personal.parquet	2.9 KB	application/octet-stream	31 oct 2023 12:15:06	Standard
adherencia.parquet	34.3 KB	application/octet-stream	31 oct 2023 12:15:06	Standard
aglomeraciones.parquet	499 KB	application/octet-stream	31 oct 2023 12:15:07	Standard
alertas_ips.parquet	23 KB	application/octet-stream	31 oct 2023 12:15:08	Standard
amb_ajust_sql.parquet	17.5 KB	application/octet-stream	31 oct 2023 12:15:08	Standard
amb_reps_montar.parquet	5.8 KB	application/octet-stream	31 oct 2023 12:15:08	Standard
ambformatoemerg.parquet	458.1 KB	application/octet-stream	31 oct 2023 12:15:10	Standard

Ilustración 55 Detalle Bucket data cargada

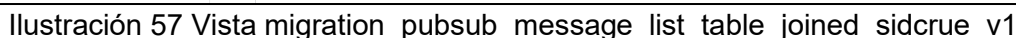
Se tiene disponible la base de datos con la réplica de SIDCRUE en MySQL. Se maneja de dos formas, una la unificación de tablas y la otra el que procesa las tablas que no tienen un agrupamiento. Para generar esta tarea se ejecuta un **Cloud scheduler** (programador de tareas de Google Cloud que permite automatizar la ejecución de procesos de acuerdo con la configuración del usuario), el cual se va a ejecutar cada vez que se requiera. Este activa un **PubSub** (el cual permite la comunicación asíncrona entre microservicios y se utiliza para la canalización de integración de datos, funciona como un middleware), enviando un mensaje para que la **Cloud function** (que es un entorno de ejecución permitiendo crear y conectar servicios en la nube) comience su ejecución.

Se activa e inicia el procesamiento de cada una de las listas de tablas, cada vez que termina el procesamiento de una lista envía un mensaje de inicio utilizando el **PubSub** a una **Cloud function**, la cual va ejecutando el proceso de unificación de la lista donde el primer ítem es el nombre de la tabla como se unificará. Los siguientes elementos de la lista son los nombres de las tablas que están en el grupo a unificar.

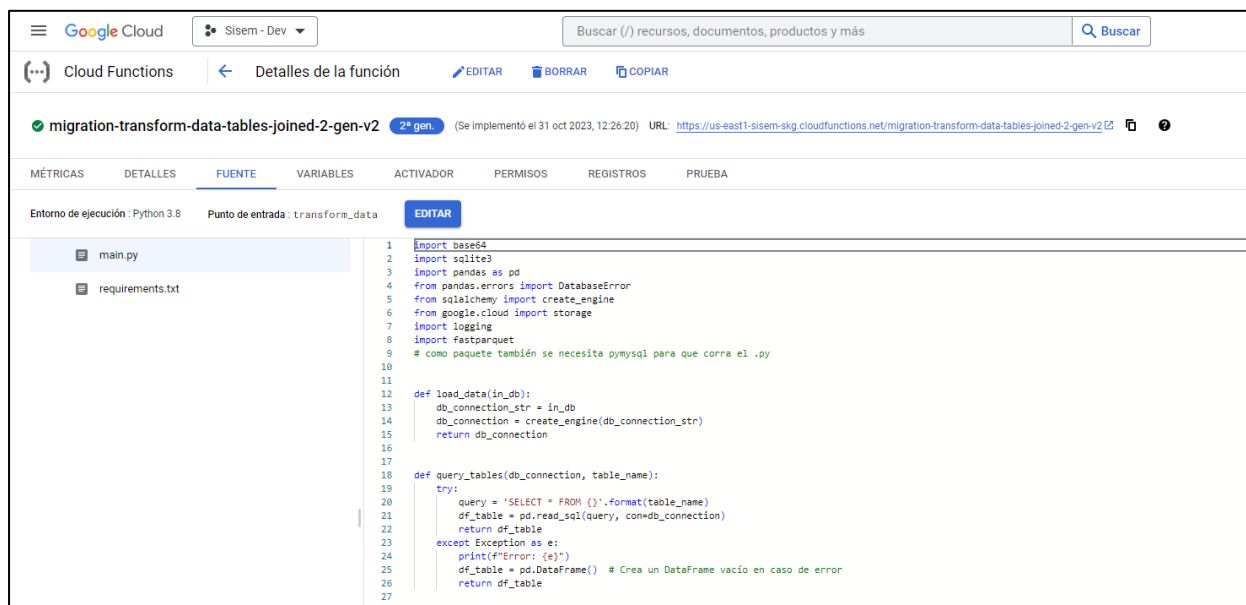
Finalmente, esto se almacena en el **Cloud Storage** que se encuentra en el Data Lake. En el caso de las tablas que no están agrupadas se ejecuta el otro camino el cual puede ir ejecutando cada una de las tablas y generar el almacenamiento en el Cloud Storage.



migration_pubsub_message_list_table_joined_sidcrue_v1 contiene todos los listados de las tablas a agrupar y es quien se encarga de enviar el mensaje al PubSub, para que inicie la ejecución. Este mensaje lo recibe la función **migration-transform-data-tables-joined-2-gen-v2**, el cual recibe la lista de tablas a unificar e inicia el proceso de unificación de las tablas al archivo csv generando un archivo log que permite identificar los errores.



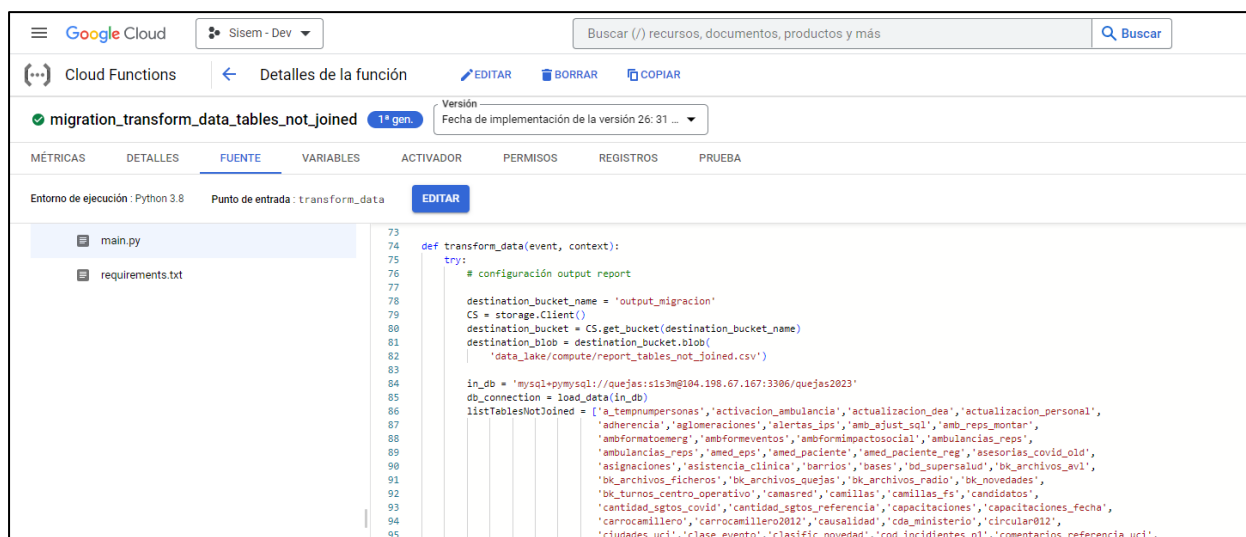
55



```
1 import base64
2 import sqlite3
3 import pandas as pd
4 from pandas.errors import DatabaseError
5 from sqlalchemy import create_engine
6 from google.cloud import storage
7 import logging
8 import fastparquet
9 # como paquete también se necesita pymysql para que corra el .py
10
11
12 def load_data(in_db):
13     db_connection_str = in_db
14     db_connection = create_engine(db_connection_str)
15     return db_connection
16
17
18 def query_tables(db_connection, table_name):
19     try:
20         query = 'SELECT * FROM {}'.format(table_name)
21         df_table = pd.read_sql(query, con=db_connection)
22         return df_table
23     except Exception as e:
24         print(f"Error: {e}")
25         df_table = pd.DataFrame() # Crea un DataFrame vacío en caso de error
26         return df_table
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

Ilustración 58 Vista migration-transform-data-tables-joined-2-gen-v2

migration_transform_data_tables_not_joined contiene todas las tablas que no van a ser unificadas, este es el segundo proceso en paralelo donde se envía el mensaje al PubSub y se ejecuta la tarea.



```
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

Ilustración 59 Vista migration_transform_data_tables_not_joined

data_collection_to_bigquery_migration_v1 finalmente se ejecuta este proceso el cual toma los archivos almacenados en el bucket de salida de la migración y los transforma en datasets de BigQuery

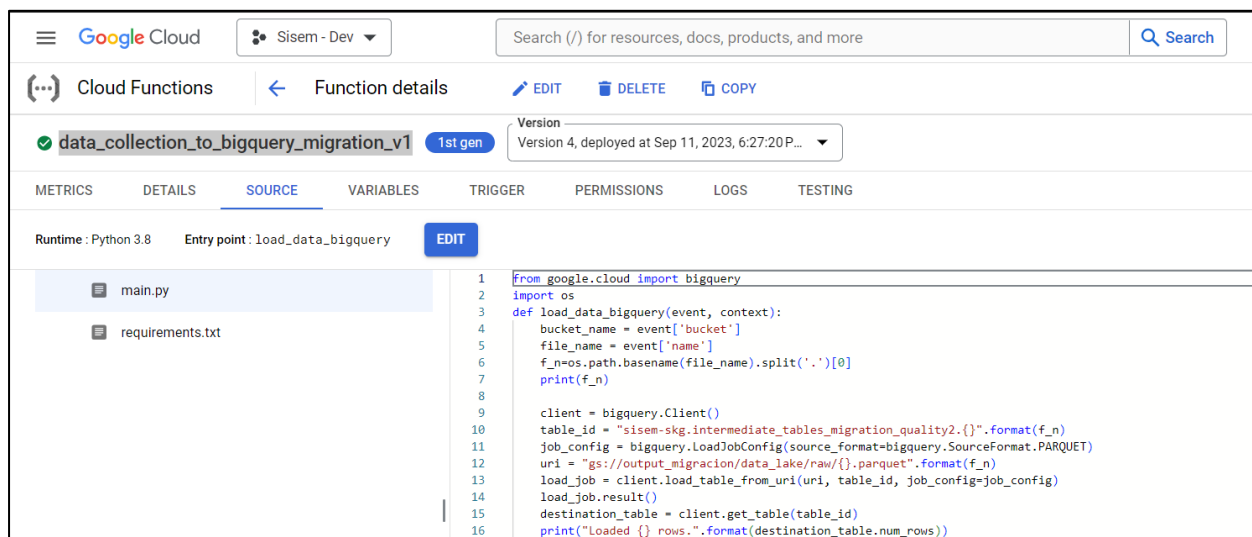


Ilustración 60 Vista data_collection_to_bigquery_migration_v1

5. Copias de Respaldo de la Información

5.1. Copia del SIDCRUE en BD

En la plataforma de Google Cloud se creó un servidor para restablecer la versión de SIDCRUE, este servidor tiene instalado la versión de MySQL 5.7 la dirección IP de acceso es la 104.198.67.167. Se realizó con base a la copia compartida por la secretaria de Salud la cual tenía un tamaño en disco de 29.0 GB (31,148,400,640 bytes). El proceso de restauración de este backup tardó aproximadamente 2 horas en completarse.

SQL

Instancias

CREAR INSTANCIA

MIGRAR DATOS

MOSTRAR PANEL DE INFORMACIÓN

Filtro

Ingresar el nombre o el valor de la propiedad

☐	ID de instancia	Tipo	Dirección IP pública	Dirección IP privada	Nombre de la conexión con la instancia	Alta disponibilidad	Ubicación	Acciones
☐	bd-sisem-01	PostgreSQL 13	34.148.74.49		sisem-skg:us-east1:bd-sisem-01	AGREGAR	us-east1-b	
☐	bd-sisem-01-prod	PostgreSQL 13	34.138.83.23		sisem-skg:us-east1:bd-sisem-01-prod	AGREGAR	us-east1-b	
☐	quejas	MySQL 5.7	104.198.67.167		sisem-skg:us-central1:quejas	AGREGAR	us-central1-f	

Ilustración 61 Servidores bases de datos Google Cloud

5.2. Actualización de la Copia de Respaldo

Para actualizar la copia del respaldo se solicitará un restaure semanal de la base de datos original del SIDCRUE, la cual se restablecerá en el servidor en Google Cloud que se destinó para tal finalidad.

5.3. Backup de Información Histórica

De acuerdo con la política de copias de seguridad se realiza semanalmente una vez se realice el restaure de la base de datos original.

El respaldo se realiza en frío (cold backup) o una copia de seguridad fuera de línea (offline backup). Es un proceso que se realiza mientras la base de datos esta fuera de línea y no se

puede acceder a la misma. De todas formas, como no se está realizando una actualización continua esto no afectaría en nada la copia que se está realizando.

Copias de seguridad

Recuperación de un momento determinado y copias de seguridad automáticas

Protege tus datos de pérdidas a un costo mínimo. [Más información](#)

☒ Copias de seguridad automáticas

Elige un período para que se cree una copia de seguridad de tus datos automáticamente. Aunque termine el período, el proceso continuará hasta que finalice. La hora está en tu zona horaria local (UTC-5).

12:00 – 16:00

✓ OPCIONES AVANZADAS

☐ Habilita la recuperación de un momento determinado

Te permite recuperar datos de un momento específico, con una precisión de fracciones de segundo, mediante el archivado de registros de escritura por adelantado. Asegúrate de tener suficiente almacenamiento o aumentos automáticos de almacenamiento para evitar quedarte sin espacio en el disco.

Mantenimiento

Las actualizaciones podrían realizarse en cualquier día de la semana. Cloud SQL define el momento en que se realiza el mantenimiento.

Resumen

Región	us-central1 (Iowa)
Versión de la base de datos	PostgreSQL 14
CPU virtuales	1 CPU virtual(es)
Memoria	3.75 GB
Almacenamiento	10 GB
Capacidad de procesamiento de la red (MB/s) ?	250 de 2,000
Capacidad de procesamiento del disco (MB/s) ?	Lectura: 4.8 de 240.0 Escritura: 4.8 de 72.0
IOPS ?	Lectura: 300 de 15,000 Escritura: 300 de 4,500
Conexiones	—
Copia de seguridad	Automatizada
Disponibilidad	Zona única
Recuperación de un momento determinado	Inhabilitada

Ilustración 62 Ejemplo copias de seguridad

6. Evaluación de Calidad de los Datos Durante el Proceso de Migración y Etapas de Migración

6.1. Proceso de Migración y Etapas

En el diagrama de arquitectura de datos se puede identificar la distribución de las bases de datos. Inicialmente tenemos la BD_SIDCRUE que representa la base de datos actual del SIDCRUE en MySQL; Base de datos no relacional y toda la información que contiene la plantilla 72 la descarga en archivos de Excel. que obtiene los datos del Premier ONE.

A partir del desarrollo e implementación de solución de arquitectura de datos en Google Cloud Platform, se gestiona un proceso ETL (Extracción, Transformación y Carga) con el objetivo de dimensionar la plantilla 72 junto con los atributos necesarios para la proyección de indicadores, gráficas y reportes.

En el **DataLake** se encuentra el bucket de **Cloud Storage** *datalake_r24_plantilla_72_input*, donde se almacenan los archivos Excel generados con la información de PREMIER ONE como primer componente en el procesamiento de información. La información contenida en estos archivos es procesada y transformada por la función *prueba_de_concepto_bi_ingest_data_r2_p72*. El proceso realizado por esta función reemplaza valores nulos, agrega las columnas necesarias según el archivo fuente y ejecuta las demás tareas de clasificación y limpieza de datos necesarias para poder generar los archivos intermedios que alimentarán los reportes. El resultado de esta primera etapa de procesamiento se almacena en el bucket *datalake_r24_plantilla_72_output*.

Por otra parte, el bucket *datalake_plantilla_72_parametricas* almacena información extraída de los datos que maneja SIDCRUE tanto en su base de datos como en los archivos Excel y CSV provistos por la SDS para ser procesados en conjunto con los datos resultado de la limpieza y transformación inicial que se alojan en el bucket *datalake_r24_plantilla_72_output*. La función encargada de realizar este procesamiento conjunto es la **Cloud Function** *prueba_de_concepto_bi_build_p72* que a su vez genera un log del proceso ejecutado almacenando este log en el bucket *prueba_de_concepto_bi_log*. La información procesada que alimentará los reportes se almacena en *datalake_built_plantilla_72*. Esta data es reprocesada por la función *prueba_de_concepto_bi_mask_data* que ajusta el formato de algunas columnas, información que luego es almacenada en *plantilla_72_output* y cargada al conjunto de datos de **BigQuery** llamado *data_collection_to_bigquery_dataset_bi_p72*. De este último dataset es de donde PowerBI toma la información unificada de las bases de datos de SISEM para mostrar en los reportes y tableros.

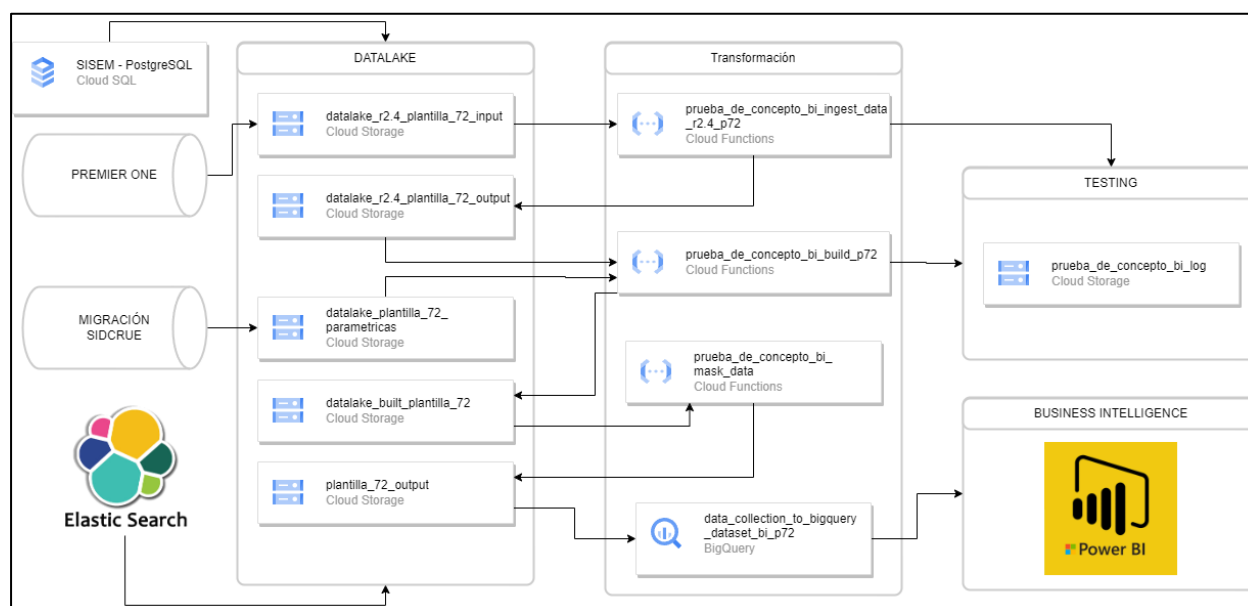


Ilustración 63 Diagrama arquitectura de datos

6.2. Reglas de Validación de la Migración

Para cada fase del proceso de migración se aplican reglas de validación que se pueden agrupar en tres: Completitud, concordancia e integridad.

6.2.1. Completitud

Para validar los datos que se migraron de la base de datos al bucket se generó un reporte, el cual se puede validar contra la información del Profiling o realizando un conteo de los registros en la base de datos original. Además, permite identificar las tablas que no contienen registros, ya que estas estructuras también serán migradas para conservar la información de todas las tablas.

name_table (Nombre de la tabla)	len_attributes(Cantidad de campos)	len_registry (Cantidad de registros)
---------------------------------	------------------------------------	--------------------------------------

ponderaciones_encuestas_d espachos	152	0
pdf_dea	6	0
pac_interv_smental	15	0
despachos_amed_cargue	42	0
nedocs	13	0
lugarinc	4	0
interv_smental	14	0
evaluacion_despacho	46	0
evaluacion_enfermeros	15	0
ponderaciones_encuestas_e nfermeros	152	0
historico_avl	15	0
electivas	37	0
historico_radio	16	0
indicadores_serv_urgencias	13	0
evaluacion_referencia	39	0
ponderaciones_encuestas_re ferencia	110	0
sci206ambincidente	8	0
sci206ambrequeridas	8	0
asistencia_clinica	9	0
sci206final	5	0
bk_archivos_radio	8	0
bk_archivos_avl	8	0
reporte_incidentes	38	0
servicios_urgencias	14	0
amb_reps_montar	37	0
sci211solicitudrecurso	25	1
datos_qr	3	1
sci206hospitales	7	1
dia_cierre	3	1
sci206estacionesasimedica	8	1
uso_dea	25	1
sci206anexo	10	1
novedades_th	5	1
sci201a	5	1
sci201resumenincidente	11	1
ordendetrabajo	17	2
datos_uci_cerrados	3	2
actualizacion_personal	8	3
rpc_ninos	67	3

correo_activo	8	3
encuentas	9	3
datos	3	4
ponderaciones_encuestas_tarm	152	4
personal_fs	4	4
novedades_retencion_sem	38	4
empresas_ctos	5	5
regimen	3	5
datos_uci	3	5
contratos_aph	8	5
evaluacion_tarm	33	6
seguimientos_referencia_covid	9	6
turno_conductores	69	7
reg_asegurador	3	7
servicio_ocurrencia	6	7
datos_hospitales	4	10
temporal_bodega	4	10
nombres_indicadores	2	10
novedades_despachos_corferias	4	10
tipo_ent_resp	3	11
ips_amed	3	11
ponderaciones_encuestas_piscologia	137	11
causalidad	2	12
evaluacion_dea	70	12
evento_aph	27	13
cursos_pers_veh	4	14
tecnicos	5	15
disponibles	4	15
fservicio	4	15
bk_novedades	30	16
activacion_ambulancia	6	17
infraccion_medica	62	18
cod_incidentes_p1	3	19
bk_archivos_ficheros	11	19
localidad	3	20
clase_evento	2	20
indicadores_ips_amed_cambios	47	20
escenario_evento	2	23

coordiadores_aph	7	24
sucesos_seg_paciente	24	24
simulacros_dea	27	24
tipoinc	2	26
comentarios_referencia_uci	4	28
intoxicados	16	31
dptos_colombia	3	33
departamentos	3	33
bk_archivos_quejas	8	34
encuestas_covid19	23	34
psicologia	50	37
actualizacion_dea	13	37
clasific_novedad	5	37
capacitaciones	3	40
mes_facturar	8	40
seguimientos_pacientes_se m	7	42
ips_radios	4	43
despachos_cambios_amed	8	44
emergencia_sanitaria	64	45
ponderaciones_encuestas_m edicos	152	46
pacientes_urg	17	47
empresasaphhistorico	7	50
camillas	4	51
reclamos	15	52
ponderaciones_adherencia	150	56
empresasaph_10_01_2017	5	56
tras_sec_referencia_covid	32	59
tipo_resp_pago	4	65
estamentos	5	66
encuestas_sem	27	67
empresas_publicas	8	70
turno_gestor	48	73
amb_ajust_sql	63	95
notas_inc	5	99
visita_servicios_ips	75	101
ptes_llamadas	21	107
tablamedicos	3	111
bases	22	115
cda_ministerio	14	116
ciudades_uci	2	128

novedad_aph	8	142
medicoregulador	175	145
recepcion	21	149
rpc_adultos	56	168
servicios	2	179
seguimiento_bases	14	186
novedades_psicologia	7	198
inscritos_curso_sem	12	204
eapb	4	205
formeventos	45	217
eps_traslados	5	220
inventario_dea	24	229
alertas_ips	9	233
dpais	4	240
guiones_emergencias	5	241
países	5	248
usuarios_dea	16	277
datos_dias_uci	6	297
personal_dea	14	363
adherencia	54	367
camillas_fs	10	368
mesa_ayuda_sem	11	394
visita_ips	40	442
capacitaciones_fecha	11	458
referencia_covid	49	483
profesiones_capacitacion	2	519
talento	37	620
hojas_de_vida	7	632
pacientes_equipo_comando	15	674
respuestas_entrevistas	47	720
datos_dias_uci_covid	14	745
traslados_comando	51	833
barrios	3	835
ptes_intervenciones_sm	15	843
indicadores_amed	45	877
encuentas_linea_123	41	909
marca_veh_min	3	928
tiempos_opetativos_liq	8	937
fuera_servicio_totales1	11	943
productividad_amed	11	992
carrocamillero	18	1001

convocatorias	8	1005
directorio	11	1018
liberaciones	7	1071
municipios	3	1123
asignaciones	8	1144
carrocamillero2012	18	1157
novedades_disponibilidad_u ci	5	1172
encuestas_linea_123	8	1200
pagad_reg_aseg	5	1219
resumen_conglomerados	6	1267
conglomerados_amed	35	1272
liberacion_gestor	30	1293
temporal_quejas	4	1310
horas	2	1440
ambformeventos	8	1526
pdf_quejas	4	1581
pacientes_seguimientos	20	1681
personal_veh_emerg	65	1697
gestion_conglomerados_am ed	11	1716
sgtos_ips	6	1861
pacformeventos	14	1877
ambulancias_reps	39	1896
ambulancias_reps	39	1896
solicitud_emblema_md	46	2074
sgtos_revision_ambulancias	43	2143
circular012	14	2224
bd_supersalud	23	2224
referencia	60	2229
visitas_gestores_uci	32	2230
novedades_turnos_medicos	5	2380
informacion_tripulantes_vac una_covid	21	2468
camasred	129	2573
cantidad_sgtos_referencia	4	2618
enfermero	45	2631
indicadores_emergencias	31	2654
tipo_novedad	25	3034
seguimientos_sm	23	3053
referencia_uci_reabierto	9	3231
nov_thumano_covid19	11	3293

amed_paciente_reg	63	3439
prestadores_seg_pac	22	3482
empre_pub_subentidad	4	3619
tbl_novedades_turnos	5	3724
fechas_ejecucion_scripts	3	3886
formatoemerg	48	3901
novedadesamb	15	4241
aglomeraciones	31	4631
tbl_turnos	53	4687
turno_referencia	21	5226
pdf_referencia	5	5961
tbl_turnos_medicos	24	6022
tbl_entrega_turnos	5	6044
supersalud	26	6411
ptes_amed	16	6528
ingresonovedades	105	6573
indicadores_ips_amed	48	6753
amed_paciente	61	6833
novedades_enfermeros	7	7253
cups	4	7394
pacientes_sem	60	7788
novedades_referencia	11	9216
llamada_quemados	18	10478
novedades	30	11537
formimpactosocial	45	11603
vehiculos_sm_fs	24	11837
despachos_amed_no_contacto	44	12133
novedades_turno	5	13508
encuestas	29	13661
incidentes_sem	17	13823
pacientes_sm	55	13918
turno_tarm	21	14169
pac_smental	62	14664
entrega_psicologo	71	14897
ambformatoemerg	8	15479
pacformimpactosocial	14	15961
ips_traslados	6	17218
ambformimpactosocial	8	18004
solicitud_tm	13	18817
novedades_seguimientos	6	19100

turno_despachador	25	21123
turno_medico_reg	17	21952
pacformatoemerg	15	24174
pacientes_despachos2015	31	26178
vehiculos_salud_mental	24	29952
medicoreg	19	34867
siras_tabla	11	35119
disponibilidad	21	35453
inc_smental	33	37366
novedades_noconformes	18	39267
incidentes_sm	31	39769
cantidad_sgtos_covid	4	47662
novedades_reg	10	51240
disponibilidad_uci	35	57648
gestion_liberacion_camillas_eps	14	58876
novedades_psicologo	36	61215
fs_descuentos	7	97941
novedades_instituciones	12	121474
bk_turnos_centro_operativo	5	129114
incidentes_uci	30	133023
turnos_centro_operativo	15	147874
referencia_uci	85	162317
seguimientos_crue	17	191329
despachos_amed_encuestas	36	209324
pdf_uci	5	229176
traslados_secundarios	21	320727
estados_recursos_bk	8	324751

Tabla 4 Resultado tablas no unificadas

6.2.2. Concordancia

En un primer ejercicio se trasladó la información directamente de la base de datos del SIDCRUE a la base de datos de SISEM.

- Para este ejercicio de las tablas, que cuya información se va a migrar que son; ambulancias, empresas APH y hospitales, se realizó el ejercicio con ambulancias.
- Se obtenía la información de la tabla ambulancias seleccionando los campos: Tipo, estado, modelo, avl y placa.
- Estos campos se asignaron a sus correspondientes tablas y atributos en SISEM, ambulanceType (resource_attribute_entity), Status (resource_entity), model (resource_attribute_entity), avlId (resource_attribute_entity), licensePlate (resource_attribute_entity).

Este proceso inicial se llevó a cabo una sola vez. Posteriormente, se decidió no continuar con este método y se optó por un enfoque más eficiente, utilizando una herramienta que carga directamente datos actualizados desde la página de REPS en SISEM. Esta nueva metodología evitó la necesidad de realizar el proceso de mapeo dos veces.

Se procesaron en total 48'906.414 campos para un gran total de 26.643'333.120 de registros cargados y validados en el proceso de cargue a BigQuery.

El proceso de mapeo de datos se ejecutará una sola vez, de manera integral. Esta ejecución final se llevará a cabo una vez que se cumplan dos condiciones clave:

1. **Consolidación de Datos:** Toda la información proveniente de las bases de datos de **SIDCRUE** esté completamente cargada en las tablas unificadas.
2. **Preparación de Reportes:** El proceso de generación de reportes en BigQuery, definido para la siguiente fase del proyecto, esté completamente configurado y listo para operar.

Al seguir este enfoque, nos aseguramos de que el proceso de concordancia sea preciso y eficiente, al tener la información completa y la infraestructura de reportes lista antes de la ejecución final.

6.2.3. Integridad

Para garantizar la **integridad de los datos**, se implementó un proceso de validación exhaustivo en cada fase del proyecto. Este proceso se centró en la unificación de tablas, lo que incluyó:

- **Identificación de Campos Comunes:** Se analizó cada tabla para identificar los campos con mayor cantidad de coincidencias, más allá del simple nombre de la tabla. Esto aseguró que la unificación fuera precisa y se basara en la lógica de los datos.
- **Homologación Automatizada:** Un script automatizado en **Python** se encarga de homologar y estandarizar los campos de las tablas unificadas. Este proceso garantiza que los datos de las distintas fuentes se consoliden de manera coherente antes de su carga final.

6.2.3.1. Automatización del proceso

El proceso de unificación de datos y carga final está completamente automatizado para garantizar su eficiencia, consistencia y trazabilidad. Una vez que la información de SIDCRUE ha sido extraída y transformada, una serie de scripts desarrollados en Python se encarga de gestionar el flujo de datos hasta su destino final **de acuerdo con lo** detallado en la sección de procesos de migración y carga.

Este pipeline de scripts realiza las siguientes operaciones clave:

- **Transformación y Homologación:** Como se mencionó en el apartado de integridad, el script **identifica y estandariza** automáticamente los campos de las tablas unificadas. Esto incluye la validación de tipos de datos, el formato de los campos y la aplicación de reglas de negocio para asegurar que los datos estén listos para su uso.
- **Carga y Persistencia:** La información ya homologada es cargada de forma persistente en un conjunto de datos en **BigQuery**. Esta base de datos actúa como el origen de datos

central para los componentes de inteligencia de negocios (BI), permitiendo que los equipos accedan a la información limpia y estructurada para sus análisis y reportes.

- **Auditoría y monitoreo de ejecución:** incluye logs y resultados de la ejecución de la migración tipificando los tipos de error que ocurren en caso de presentarse algún inconveniente en la extracción, transformación y carga de la información hacia los repositorios intermedios o en el repositorio final de la información.
- **Monitoreo de infraestructura:** al soportarse sobre herramientas de la consola de Google Cloud Platform se cuenta con el detalle de ejecución, orquestación, uso de recursos y cargas sobre componentes y diversos activadores del flujo de scripts durante toda la arquitectura de procesamiento de datos.

Este enfoque automatizado no solo reduce el riesgo de errores manuales, sino que también facilita la actualización de los datos en el futuro.

Fin del documento