

Inteligencia de Negocios

EsSalud



EsSalud



Github: <https://github.com/DavidI7I2uni/Contralor-a-SIN>

Preguntas de negocio

Área	Rol del Usuario	Pregunta de Negocio	Nivel de Prioridad	Fuente de Datos Actual
Dirección Estratégica	Alta Dirección de EsSalud	¿Cuál es la tasa de diagnósticos de enfermedades priorizadas (diabetes, hipertensión, obesidad) por cada 1000 habitantes?	Alta	Datos Abiertos EsSalud (consultas externas), CIE10
Dirección Estratégica	Presidencia Ejecutiva	¿Cómo está evolucionando la tendencia de diagnósticos en el tiempo y en qué zonas se proyecta mayor riesgo?	Alta	Datos Abiertos EsSalud, INEI
Gestión Táctica	Gerencias Regionales	¿Qué regiones presentan mayor concentración de diagnósticos y cómo se distribuyen los casos por sexo y edad?	Alta	Datos Abiertos EsSalud, Ubigeo INEI

Preguntas de negocio

Área	Rol del Usuario	Pregunta de Negocio	Nivel de Prioridad	Fuente de Datos Actual
Gestión Táctica	Dirección de Prevención y Promoción	¿Cuál es la edad promedio al momento del diagnóstico y cómo puede orientar las campañas preventivas?	Media	Datos Abiertos EsSalud
Gestión Táctica	Coordinación de Control Médico	¿Cuánto tiempo pasa, en promedio, entre el diagnóstico inicial y el primer control posterior?	Alta	Registros hospitalarios, Datos Abiertos EsSalud
Gestión Operativa	Jefatura Hospitalaria	¿Cuál es la cobertura de diagnósticos lograda por cada red hospitalaria respecto a su población asegurada?	Alta	Datos Abiertos EsSalud, RENIPRESS
Gestión Operativa	Unidad de Recursos Humanos	¿Qué nivel de disponibilidad de especialistas se tiene en relación con los pacientes diagnosticados?	Media	RRHH EsSalud, RENIPRESS

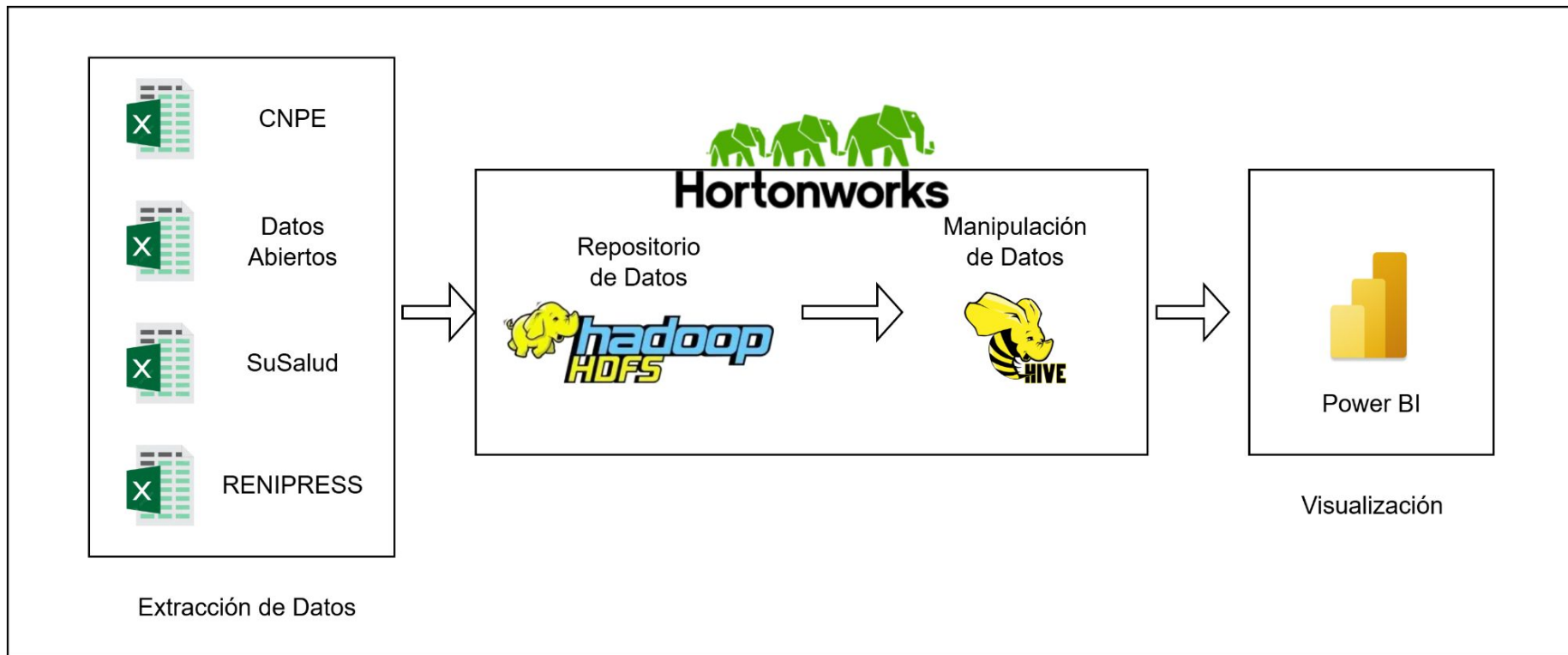
KPI's Definidos

Nombre del KPI	Descripción	Formula	Unidad	Frecuencia
Tasa de diagnósticos por 1000 habitantes	Mide la frecuencia de diagnósticos de estas patologías en la población.	$(N^{\circ} \text{ diagnósticos} \div \text{Población total}) \times 1000$	Casos por 1000 hab.	Mensual
Tendencia de crecimiento de casos	Mide la variación porcentual de diagnósticos en el tiempo para anticipar zonas de riesgo.	$((\text{Casos periodo actual} - \text{Casos periodo anterior}) \div \text{Casos periodo anterior}) \times 100$	%	Mensual
Concentración geográfica de diagnósticos	(Casos en región ÷ Total de casos) × 100	$(\text{Casos en región} \div \text{Total de casos}) \times 100$	%	Trimestral
Edad promedio de diagnóstico	Estima la edad promedio de los pacientes al momento del diagnóstico, útil para segmentar campañas.	$\Sigma (\text{Edad de diagnóstico}) \div N \text{ pacientes}$	Años	Trimestral
Tiempo promedio entre diagnóstico y control	Mide días transcurridos entre diagnóstico inicial y primer control posterior.	$\Sigma (\text{Fecha control} - \text{Fecha diagnóstico}) \div N \text{ pacientes}$	Días	Semestral
Distribución por sexo de diagnósticos	Diferencia en proporción de diagnósticos por sexo (Paciente)	$(\text{Casos sexo} \div \text{Total casos}) \times 100$	%	Trimestral

KPI's Definidos

Nombre del KPI	Descripción	Formula	Unidad de medida	Frecuencia
Índice de diagnósticos en población joven (<40 años)	Casos de hipertensión/diabetes/o besidad en menores de 40 años (Paciente + Diagnóstico).	$(\text{Casos } <40 \text{ años} \div \text{Total casos}) \times 100$	%	Trimestral
Disponibilidad de especialistas por paciente	Relación médicos especialistas/pacientes diagnosticados (Médico + Diagnóstico).	$N \text{ Médicos} \div N \text{ pacientes atendidos}$	Ratio	Trimestral
Cobertura de diagnósticos por red hospitalaria	Casos diagnosticados en cada red vs asegurados en esa red (IPRESS + Ubigeo).	$(\text{Pacientes diagnosticados} \div \text{Población asegurada en red}) \times 100$	%	Trimestral
Variabilidad regional de diagnósticos	Diferencia entre la región con mayor y menor tasa de diagnósticos (Ubigeo + Diagnóstico).	$\text{Máx}(\text{tasa diagnósticos}) - \text{Mín}(\text{tasa diagnósticos})$	%	Trimestral

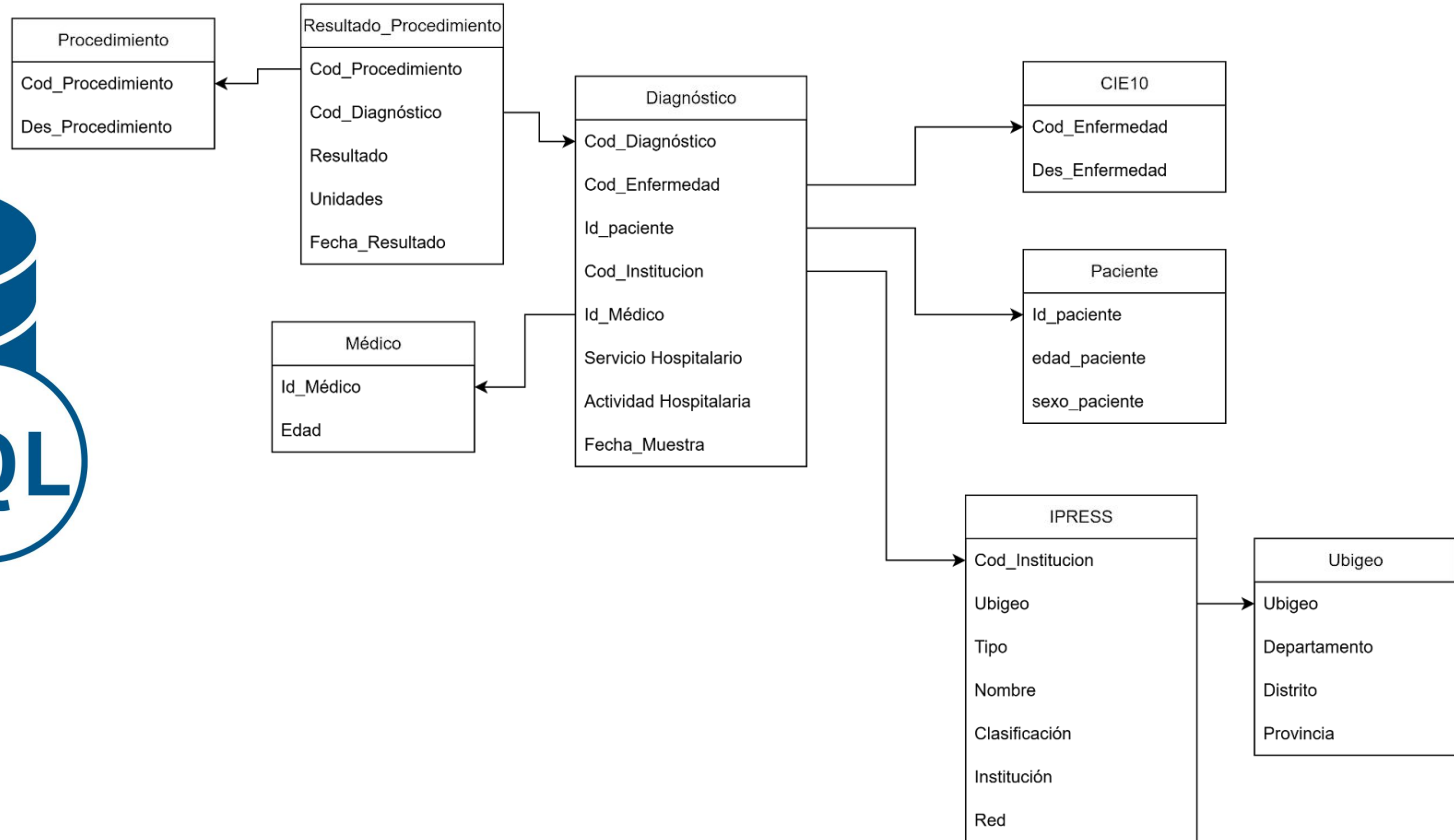
Arquitectura de datos



Fuente de datos

Repositorio	Archivo	Descripción
Centro Nacional de Planeamiento Estratégico	Datos-planeamiento-estrategico.xlsx	Incluye descripciones de cada ubigeo
Datos Abiertos Perú	DF_ExLab_CExt_Diabetes.csv	Contiene todas las consultas externas con diagnósticos de diabetes y sus detalles desde el 2020 hasta el 2024
	DF_ExLab_CExt_Hipertension.csv	Contiene todas las consultas externas con diagnósticos de hipertensión y sus detalles desde el 2020 hasta el 2024
	DF_ExLab_CExt_Obesidad.csv	Contiene todas las consultas externas con diagnósticos de obesidad y sus detalles desde el 2020 hasta el 2024
	Ubigeo.csv	Contiene el código de ubigeo, su distrito, provincia y departamento
SuSalud	CIEIO.csv	Contiene el código de todas la enfermedades y su descripción
RENIPRESS	USLRC20250923.xls	Contiene información de todos los hospitales del Perú

Modelo conceptual



Inventario de Fuentes de datos OLTP

Sistema	Área usuaria	Tipo	Tecnología	Frecuencia de utilización	Observaciones
Sistema de Planeamiento Estratégico – CEPLAN	Planeamiento Estratégico y Estadística	OLTP (maestro de referencia geográfica)	Archivo XLS	Mensual	Incluye descripciones y metadatos de cada ubigeo; sirve como insumo referencial para consolidar datos de salud y otras áreas del Estado.
Sistema de Consultas Externas – Datos Abiertos EsSalud	Red hospitalaria / Consulta externa	OLTP (registros médicos de atenciones y diagnósticos)	Archivos CSV	Diario	Incluye registros de exámenes de laboratorio en consultas externas con diagnóstico de diabetes, hipertensión y obesidad.
Sistema de Codificación Geográfica – Ubigeo (INEI)	Planeamiento / Estadística	OLTP (catálogo maestro geográfico)	Archivo CSV	Eventual / según necesidad	Contiene códigos y descripciones de distrito, provincia y departamento. Es la tabla de referencia estándar para enlazar información de salud con ubicación geográfica oficial.

Sistema	Área usuaria	Tipo	Tecnología	Frecuencia de utilización	Observaciones
Sistema de Clasificación Internacional de Enfermedades – CIE10	Áreas médicas y de estadística en EsSalud / SuSalud	OLTP (catálogo maestro de enfermedades)	Archivo CSV	Eventual / según necesidad	Contiene el código y descripción de enfermedades según la Clasificación Internacional de Enfermedades (CIE10).
RENIPRESS – Registro Nacional de Instituciones Prestadoras de Servicios de Salud	SuSalud / Gestión de infraestructura hospitalaria	OLTP (maestro de establecimientos de salud)	Archivo XLS	Eventual / según necesidad	Contiene información oficial de todos los hospitales, clínicas y centros de salud registrados en el Perú.

Carga de archivos

/ > Data

Total: 6 files or folders

+ Select All

New Folder

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
<div>↩</div>					
<div>📁 CIE10_2021</div>	--	2025-09-29 10:39	hive	hdfs	drwxr-xr-x
<div>📁 DF_ExLab_CExt_Diabetes</div>	--	2025-09-28 23:23	maria_dev	hdfs	drwxr-xr-x
<div>📁 DF_ExLab_CExt_Hipertension</div>	--	2025-09-28 23:23	maria_dev	hdfs	drwxr-xr-x
<div>📁 DF_ExLab_CExt_Obesidad</div>	--	2025-09-28 23:23	maria_dev	hdfs	drwxr-xr-x
<div>📁 Planeamiento_Estrategico_Ubigeo</div>	--	2025-09-29 10:39	hive	hdfs	drwxr-xr-x
<div>📁 geodir-ubigeo-inel</div>	--	2025-09-29 10:39	hive	hdfs	drwxr-xr-x

/ > Data > DF_ExLab_CExt_Dia...

Total: 2 files or folders

+ Select All

New Folder

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
<div>↩</div>					
<div>📄 DF_ExLab_CExt_Diabetes.csv</div>	124.8 MB	2025-09-28 22:27	maria_dev	hdfs	-rw-r--r--
<div>📄 DF_ExLab_CExt_Diabetes_2.csv</div>	104.6 MB	2025-09-28 22:28	maria_dev	hdfs	-rw-r--r--

Visualización de Archivos

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Data/
Found 6 items
drwxr-xr-x  - hive      hdfs      0 2025-09-29 15:39 /Data/CIE10_2021
drwxr-xr-x  - maria_dev hdfs      0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Diabetes
drwxr-xr-x  - maria_dev hdfs      0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Hipertension
drwxr-xr-x  - maria_dev hdfs      0 2025-09-29 04:23 /Data/DF_ExLab_CExt_Obesidad
drwxr-xr-x  - hive      hdfs      0 2025-09-29 15:39 /Data/Planeamiento_Estrategico_Ubigeo
drwxr-xr-x  - hive      hdfs      0 2025-09-29 15:39 /Data/geodir-ubigeo-inei
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /Data/CIE10_2021/
Found 1 items
-rw-r--r--  1 maria_dev hdfs    913690 2025-09-29 03:26 /Data/CIE10_2021/CIE10_2021.csv
```

Carga de datos



Query Editor

Worksheet ✕ Worksheet (1) ✕

```
1 --2. DF_ExLab_CExt_Obesidad
2 DROP TABLE IF EXISTS df_exlab_cext_obesidad;
3 CREATE EXTERNAL TABLE df_exlab_cext_obesidad (
4   fecha_corte STRING,
5   departamento STRING,
6   provincia STRING,
7   distrito STRING,
8   ubigeo STRING,
9   red STRING,
10  ipress STRING,
11  id_paciente STRING,
12  edad_paciente INT,
13  sexo_paciente STRING,
14  edad_medico INT,
15  id_medico STRING,
16  cod_diag STRING,
17  diagnostico STRING,
18  area_hospitalaria STRING,
19  servicio_hospitalario STRING,
20  actividad_hospitalaria STRING,
21  fecha_muestra STRING,
22  fec_resultado_1 STRING,
23  procedimiento_1 STRING,
24  resultado_1 STRING,
25  unidades_1 STRING,
26  fec_resultado_2 STRING,
27  procedimiento_2 STRING,
28  resultado_2 STRING,
29  unidades_2 STRING
30 )
31 ROW FORMAT DELIMITED
32 FIELDS TERMINATED BY '\073'
33 STORED AS TEXTFILE
34 LOCATION '/Data/DF_ExLab_CExt_Obesidad/'
35 TBLPROPERTIES ("skip.header.line.count"="1");
```

Worksheet ✕ Worksheet (1) ✕

```
1 --4. CIE10
2 DROP TABLE IF EXISTS cie10;
3 CREATE EXTERNAL TABLE cie10 (
4   codigo STRING,
5   descripcion STRING
6 )
7 ROW FORMAT DELIMITED
8 FIELDS TERMINATED BY ','
9 STORED AS TEXTFILE
10 LOCATION '/Data/CIE10_2021/'
11 TBLPROPERTIES ("skip.header.line.count"="1");
```

Query Editor

Worksheet ✕ Worksheet (1) ✕

```
1 --3. DF_ExLab_CExt_Diabetes
2 DROP TABLE IF EXISTS df_exlab_cext_diabetes;
3 CREATE EXTERNAL TABLE df_exlab_cext_diabetes (
4   fecha_corte STRING,
5   departamento STRING,
6   provincia STRING,
7   distrito STRING,
8   ubigeo STRING,
9   red STRING,
10  ipress STRING,
11  id_paciente STRING,
12  edad_paciente INT,
13  sexo_paciente STRING,
14  edad_medico INT,
15  id_medico STRING,
16  cod_diag STRING,
17  diagnostico STRING,
18  area_hospitalaria STRING,
19  servicio_hospitalario STRING,
20  actividad_hospitalaria STRING,
21  fecha_muestra STRING,
22  fec_resultado_1 STRING,
23  procedimiento_1 STRING,
24  resultado_1 STRING,
25  unidades_1 STRING,
26  fec_resultado_2 STRING,
27  procedimiento_2 STRING,
28  resultado_2 STRING,
29  unidades_2 STRING
30 )
31 ROW FORMAT DELIMITED
32 FIELDS TERMINATED BY '\073'
33 STORED AS TEXTFILE
34 LOCATION '/Data/DF_ExLab_CExt_Diabetes/'
35 TBLPROPERTIES ("skip.header.line.count"="1");
```

Execute

Explain

Upload

Save as...

Consultas



Worksheet

```
1 SELECT *
2 FROM df_exlab_cext_hipertension
3 LIMIT 10;
```

Query Process Results (Status: SUCCEEDED)

Save results...

Logs

Results

Filter columns...

previous

next

df_exlab_cext_hipertension.fecha_corte	df_exlab_cext_hipertension.departamento	df_exlab_cext_hipertension.provincia	df_exlab_cext_hipertension.districto	df_exlab_cext_hipertension.u
20240531	UCAYALI	CORONEL PORTILLO	MANANTAY	250107
20240531	UCAYALI	CORONEL PORTILLO	YARINACOCOA	250105
20240531	LIMA	LIMA	SURQUILLO	150141
20240531	LIMA	LIMA	SAN LUIS	150134
20240531	HUANUCO	HUANUCO	AMARILIS	100102
20240531	AREQUIPA	AREQUIPA	PAUCARPATA	40112
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201
20240531	ICA	CHINCHA	CHINCHA ALTA	110201

GRACIAS