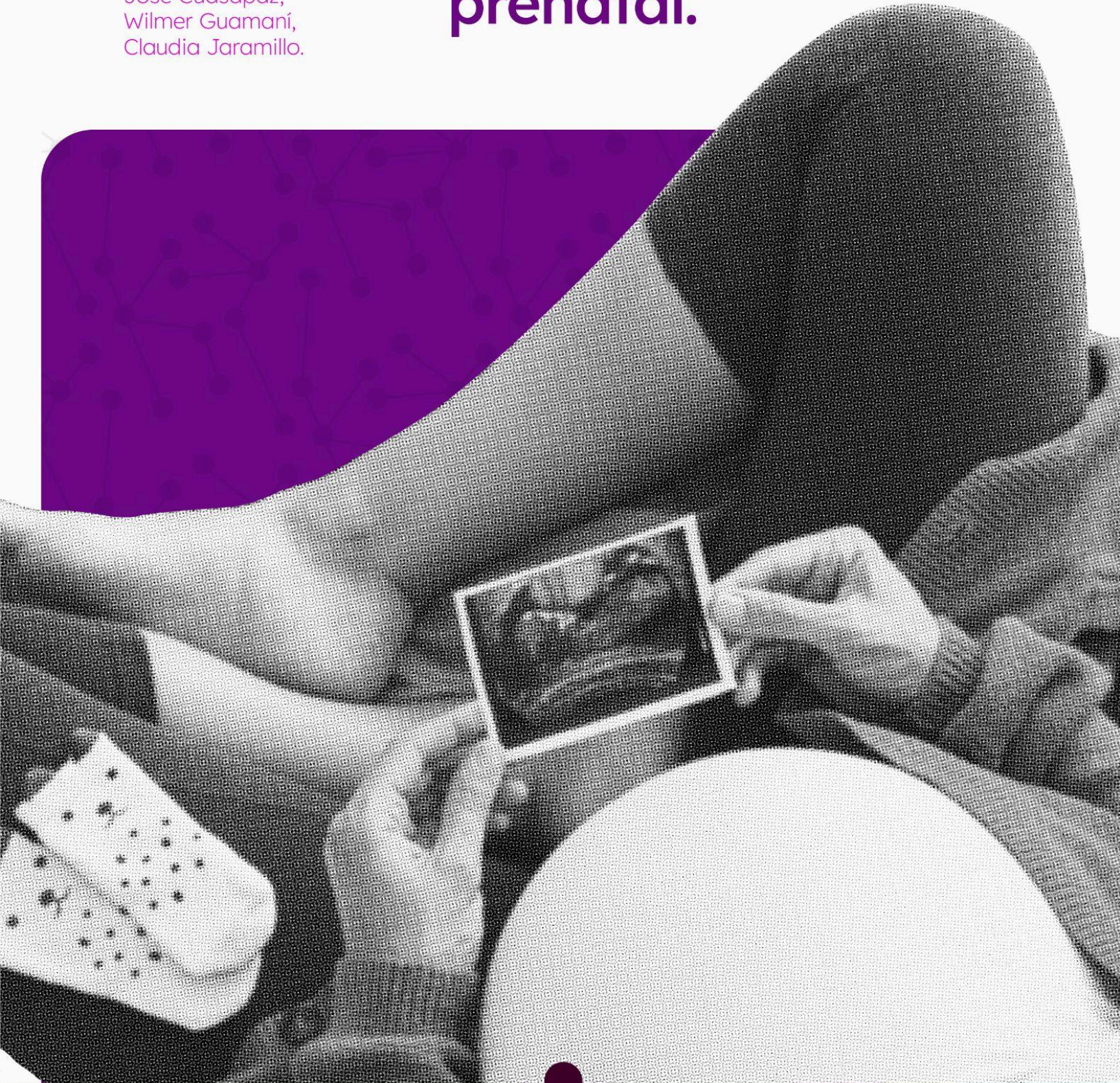


Autores:

Daniel Alvarez,
Jose Cuasapaz,
Wilmer Guamaní,
Claudia Jaramillo.

Guía metodológica para la creación de un dashboard del indicador de control prenatal.





Guía metodológica para la creación de un dashboard del indicador de control prenatal

Elaborado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo.

Diseño y Diagramación:

Fundación Datalat

Proyecto +Datos +Transparencia

Fundación Datalat

En el marco del **Convenio de Cooperación Interinstitucional entre el Ministerio de Telecomunicaciones y de la Sociedad de la Información y la Fundación Datalat No. 024-2022**, y los compromisos del **Primer Plan de Acción de Gobierno Abierto Ecuador 2019-2021** y el **Segundo Plan de Acción de Gobierno Abierto Ecuador 2022-2024** apoyados por la **Secretaría de Gobierno Abierto**.

Apoyado por:

National Endowment for Democracy

Ecuador, enero del 2024

La presente guía puede ser reproducida y difundida de forma libre siempre y cuando reconozca la autoría y se asegure un uso ético de la información.

Esta guía ha sido producida con el financiamiento de National Endowment for Democracy para equipos seleccionados como ganadores del evento participativo “Datajam 2da. edición: Mide, analiza, actúa” organizado por Fundación Datalat, el Ministerio del Ambiente, Agua y Transición Ecológica (MAATE), el Ministerio de Agricultura y Ganadería (MAG), el Ministerio de Salud Pública (MSP), el Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL), Subsecretaría de Gobierno Abierto de la Presidencia de la República y la Dirección de Innovación y Vinculación de la EPN, en el marco del proyecto “+Datos +Transparencia”, su contenido es responsabilidad exclusiva de sus autores, no necesariamente es un reflejo de los puntos de vista de Fundación Datalat, MAATE, MAG, MSP, MINTEL, Subsecretaría de Gobierno Abierto de la Presidencia de la República, o la Dirección de Innovación y Vinculación de la EPN.

CONTENIDO

1.	OBJETIVOS.....	5
1.1.	Objetivo General.....	5
1.2.	Objetivos Específicos.....	5
2.	HERRAMIENTAS DE TRABAJO.....	5
2.1.	Metodología Ralph Kimball.....	5
2.2.	Python.....	6
2.3.	Jupyter Notebook.....	6
2.4.	Pentaho Data Integration.....	6
2.5.	Postgresql.....	7
2.6.	DBeaver.....	7
2.7.	Microsoft PowerBI.....	7
3.	INSUMOS.....	7
4.	PROCEDIMIENTOS.....	8
4.1.	Definición de Requerimientos de Negocio.....	9
4.1.1.	Comprender el problema.....	9
4.1.2.	Diagnósticos Iniciales.....	11
4.1.3.	Depuración del dataset.....	16
4.2.	Diseño de Arquitectura y artefactos.....	21
4.3.	Diseño de modelos.....	22
4.3.1.	Definir los modelos dimensionales.....	22
4.3.2.	Crear el modelo de datos en PostgreSQL.....	23
4.4.	Diseño de ETL.....	27
4.5.	Especificación de Dashboard.....	32
4.6.	Desarrollo de Dashboard.....	33
4.6.1.	Conectar al modelo de datos.....	33

4.6.2.	Crear variables.....	36
4.6.3.	Crear las visualizaciones (gráficos e indicadores).....	36
4.7.	Publicar el Dashboard.....	39
5.	RESULTADOS.....	40
6.	RECOMENDACIONES TÉCNICAS.....	42
7.	ANEXOS.....	42
8.	BIBLIOGRAFÍA.....	43

ILUSTRACIONES

Ilustración 1.	Metodología Kimball	6
Ilustración 2.	Librerías empleadas en el análisis de datos	8
Ilustración 3.	Variables contenidas en el DataSet	10
Ilustración 4.	Dimensiones del DataSet	10
Ilustración 5.	Tipo de variables y tipo de datos en el DataSet	11
Ilustración 6.	Frecuencia de la variable CODIGO_CIE10_1	12
Ilustración 7.	Embarazo Confirmado según la Zona	13
Ilustración 8.	Embarazo Confirmado según la Provincia	14
Ilustración 9.	Embarazo Confirmado según la Edad	15
Ilustración 10.	Embarazo Confirmado según la Nacionalidad	15
Ilustración 11.	Embarazo Confirmado según la Etnia	16
Ilustración 12.	Importación de Librerías	17
Ilustración 13.	Creación de dataframe	17
Ilustración 14.	Creación de columnas jupyter	18
Ilustración 15.	Consulta de información Jupyter	18
Ilustración 16.	Ingreso de datos para filtros de información Jupyter	19
Ilustración 17.	Creación de banderas Jupyter	20
Ilustración 18.	Resultados de filtro de información Jupyter	20
Ilustración 19.	Exportación de información Jupyter	21
Ilustración 20.	Dataset resultante del proceso realizado con la herramienta Python	21

Ilustración 21. Arquitectura del proceso de creación de un dashboard de indicadores de salud pública	22
Ilustración 22. Creación de la Base de Datos “DB_MSP”	24
Ilustración 23. Creación del esquema consultaexterna2021.	24
Ilustración 24. Tabla Embarazadas PostgreSQL.	25
Ilustración 25. Creación de Tabla aten_pren_resultado postgresQL.	27
Ilustración 26. Base de datos DB_MSP.	27
Ilustración 27. Proceso ETL de datos sobre el indicador Control Prenatal del 2021.	28
Ilustración 28. Configuración parámetros lectura archivo ATEN_PREN_RESULTADO.csv.	28
Ilustración 29. Selección de los campos del archivo ATEN_PREN_RESULTADO.csv.	29
Ilustración 30. Mapeo de los campos de la tabla ATEN_PREN_RESULTADO.	30
Ilustración 31. Configuración a la base de datos DB_MSP.	32
Ilustración 32. Proceso ETL de datos sobre personas embarazadas por provincia.	32
Ilustración 33. Información de mujeres embarazadas por provincia.	33
Ilustración 34. Pasos para Conectar al Modelo de Datos	34
Ilustración 35. Ingreso de Datos para la Conexión a Base de Datos	35
Ilustración 36. Tablas de la Base de Datos DB_MSP	36
Ilustración 37. Pasos para la creación de Variable	37
Ilustración 38. Pasos para crear visualizaciones en PowerBI	37
Ilustración 39. Visualizaciones en el Dashboard	38
Ilustración 40. Power BI Desktop – Publicar Reporte	40
Ilustración 41. Power BI Desktop – Obtener Link público del Reporte	40
Ilustración 42. Power BI Desktop – Obtener Link público del Reporte	41
Ilustración 43. Dashboard CCP-MSP	42

TABLAS

Tabla 1. Relación del Objetivos de la Guía Vs Fases de Kimball	8
Tabla 2. Tabla Embarazadas.	24
Tabla 3. Tabla aten_pren_resultado.	25

Introducción

La guía metodológica para la creación de un dashboard del indicador de control prenatal busca sugerir una arquitectura que permita visualizar de manera dinámica y efectiva los datos correspondientes al indicador de control prenatal del año 2021, suministrados por el Ministerio de Salud Pública.

El proceso se estructura en un análisis y procesamiento exhaustivo de los datos del indicador de control prenatal proporcionados por el Ministerio de Salud Pública. La guía propone diseñar un proceso automático que abarque desde la obtención hasta la carga de estos datos en un repositorio específico, haciendo uso de una herramienta E.T.L. y analítica de datos.

Entre las herramientas de trabajo destacadas se encuentra la Metodología Ralph Kimball, ampliamente reconocida en la construcción de Almacenes de Datos. Esta metodología, que se puede ejecutar de forma paralela o secuencial, se erige como un marco sólido para la integración de datos a lo largo del tiempo.

El conjunto de herramientas técnicas abarca desde el omnipresente lenguaje de programación Python hasta la versatilidad del Jupyter Notebook para la ejecución de código. Pentaho Data Integration destaca en la gestión avanzada de datos mediante técnicas ETL, y PostgreSQL, con su capacidad de funcionar como almacén de datos, se integra como sistema de gestión de bases de datos. El flujo metodológico también abraza herramientas especializadas como DBeaver para la administración de bases de datos y Microsoft PowerBI para la visualización de datos y Business Intelligence.

Los insumos utilizados en la implementación incluyen la Base de Datos de Consulta externa del MSP para el servicio ambulatorio, proyecciones poblacionales a nivel de provincia, y la aplicación de diversas librerías de Python como Pandas, Matplotlib y Seaborn durante el análisis exploratorio de datos.

1. OBJETIVOS

1.1. Objetivo General

Implementar una arquitectura y guía metodológica para visualización del indicador de control prenatal año 2021 - Ministerio de Salud Pública a través de un dashboard / tablero de información dinámico.

1.2. Objetivos Específicos

1. Realizar el análisis y procesamiento de los datos proporcionados por el Ministerio de Salud Pública sobre el indicador Control Prenatal - 2021.
2. Diseñar el proceso automático de obtención, transformación y carga de los datos de control prenatal del año 2021 al repositorio de datos con la utilización de una herramienta E.T.L. y analítica de datos.
3. Desarrollar un dashboard dinámico del indicador del Control Prenatal del 2021.

2. HERRAMIENTAS DE TRABAJO

A continuación, se detallan las metodologías, técnicas y herramientas utilizadas durante la implementación de la Arquitectura que se planteó como objetivo. El orden corresponde a las fases de la metodología adoptada.

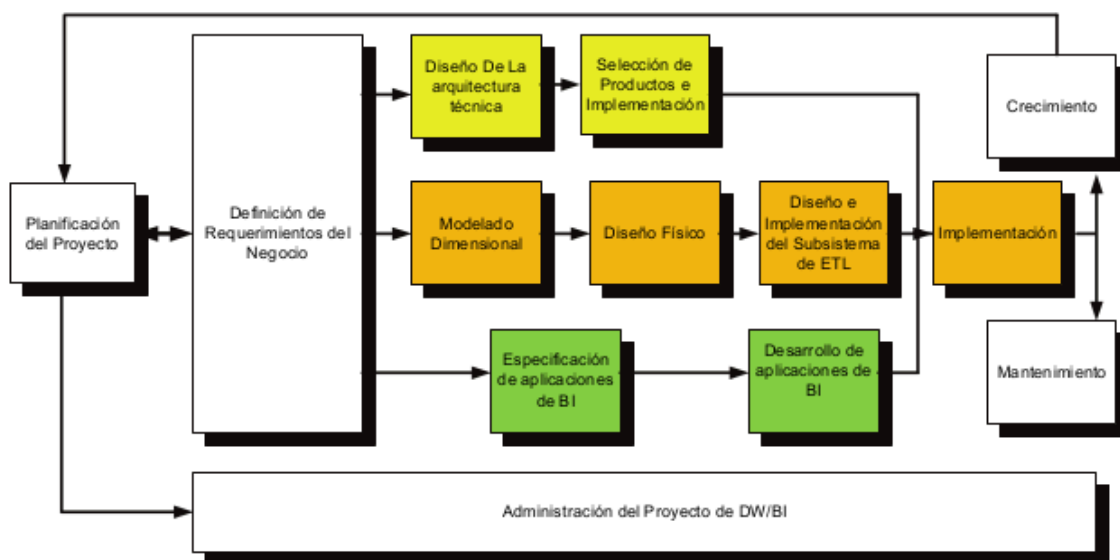
1.1. Metodología Ralph Kimball

La Metodología Kimball es ampliamente utilizada para la construcción de Almacenes de Datos, como almacén de datos nos referimos a un conjunto de datos

que sea integrado, persistente en el tiempo, dinámico en cuanto a su actualización y que ayude a la toma de decisiones para la entidad que lo implementa.

A continuación, en la ilustración 1, se muestran las fases de la Metodología Kimball, es importante recalcar que la metodología fue diseñada para que las fases puedan ser ejecutadas de forma paralela o secuencial (Kimball Group, s.f.).

Ilustración 1. Metodología Kimball



Fuente:

<http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html>

Creado por: WEB 2.0 - MediaWiki

1.2. Python

Es un lenguaje de programación que en la actualidad tiene un amplio uso entre los desarrolladores encargados de implementar soluciones de software, ciencia de datos y Machine Learning. Entre sus principales características se encuentran:

facilidad de aprendizaje, capacidad de ejecución en varias plataformas, ser un lenguaje interpretado y también ser de código abierto (AWS, 2023).

1.3. Jupyter Notebook

El Jupyter Notebook es una aplicación web de código abierto, desarrollada utilizando lenguaje HTML que permite crear, compartir y editar documentos en los que se puede ejecutar código python, hacer anotaciones, insertar ecuaciones, visualizar resultados y documentar funcionalidades (Martín, 2018).

1.4. Pentaho Data Integration

Pentaho Data Integration, PDI por sus siglas, es parte de la suite multiplataforma de inteligencia empresarial de código abierto PENTAHO. Su objetivo principal está orientado a la gestión avanzada de datos a través técnicas ETL (Extract, Transform y Load) para convertir datos brutos en datos netos; partiendo de descartar información poco útil y manteniendo aquella que puede agregar valor en un formato más limpio y estructurado. Esta información limpia nos permitirá generar informes, reportes, cuadros de mando, estadísticas y gráficas que sean más fáciles de usar e interpretar (Comunidad Empresas, 2023).

1.5. PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos relacionales (RDBMS) de código abierto, aunque fue diseñado para sistemas transaccionales, puede ajustarse y funcionar como almacén de datos. Las bases de datos transaccionales están pensadas para responder a muchas solicitudes de consulta en forma eficientemente; para esto, la mayoría de las bases de datos almacenan datos por filas; esto permite cargar filas completas de manera eficiente desde el disco.

PostgreSQL, aunque está orientado a filas, también puede funcionar fácilmente con consultas analíticas (Elsamadisi, 2021).

1.6. DBeaver

DBeaver es una aplicación de software cliente de SQL y una herramienta de administración de bases de datos. Para las bases de datos relacionales, utiliza la interfaz de programación de aplicaciones JDBC para interactuar con las bases de datos a través de un controlador JDBC (Jiménez, 2023).

1.7. Microsoft PowerBI

Microsoft PowerBI es una herramienta de Business Intelligence y visualización de datos, proporciona a los usuarios no especializados herramientas para agregar, analizar, visualizar y compartir datos. La interfaz de usuario de Microsoft Power BI es bastante intuitiva, dispone de una versión gratuita que está destinada a propietarios de pequeñas y medianas empresas, se puede conectar a una gran variedad de fuentes de datos, desde archivos Excel, bases de datos y aplicaciones en la nube (Microsoft, 2023).

3. INSUMOS

Los insumos usados en la implementación de la Arquitectura Técnica para la publicación del del dashboard del Indicador del Ministerio de Salud Pública (MSP) planteado:

1. Base de Datos de Consulta externa del MSP que corresponde al servicio ambulatorio para pacientes con una cita previa que acceden a atenciones médicas.

ANEXO 1

2. Proyecciones poblacionales a nivel de provincia. *ANEXO 2*
3. Python. Dentro del análisis exploratorio de los datos se emplearon las siguientes librerías:
 - Pandas: librería de Python especializada en el manejo y análisis de estructuras de datos.
 - Matplotlib: biblioteca para la generación de gráficos en dos dimensiones, a partir de datos contenidos en listas o arrays en el lenguaje de programación Python.
 - Seaborn: Proporciona varias funciones para personalizar los gráficos.

Ilustración 2. Librerías empleadas en el análisis de datos

```
In [73]: # Importo las librerías básicas:  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

4. PROCEDIMIENTOS

El procedimiento utilizado se basa en la metodología Ralph Kimball, para lo cual es necesario presentar la relación que tiene los objetivos específicos planteados y su correspondencia con los componentes de la metodología Ralph Kimball.

Tabla 1. Relación del Objetivos de la Guía Vs Fases de Kimball

Objetivo Específico	Fases de Metodología Kimball
1. Realizar el levantamiento y entendimiento de los requerimientos del negocio sobre el indicador de Control Prenatal del Ministerio de Salud Pública	1. Definición de Requerimientos de Negocio
	2. Diseño de Arquitectura y artefactos
2. Diseñar el proceso automático de obtención de los datos de control prenatal del año 2021, transformación y carga de datos al repositorio definido	3. Diseño de modelos
	4. Diseño de ETL
3. Desarrollar un dashboard dinámico, comprensible para visualizar el indicador Control Prenatal	5. Especificación de Dashboard
	6. Desarrollo de Dashboard
	7. Publicación de Dashboard

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.1. Definición de Requerimientos de Negocio

Los requerimientos del negocio tienen un rol central en el marco de trabajo ya que soportan de las tareas subsiguientes y a la vez definen la planificación y esfuerzo necesario para desarrollar el presente trabajo.

Como paso previo se realizó un entendimiento general del problema. Para lo cual el Ministerio de Salud Pública requiere comunicar en forma efectiva e intuitiva, desde la perspectiva de la ciudadanía, los indicadores de salud que actualmente se realiza a través de boletines; ya que su principal audiencia está conformada por la ciudadanía y las personas encargadas de la política pública de salud, cuya misión se enfoca en

garantizar el derecho a la salud de la población en el territorio ecuatoriano (Plan Estratégico Institucional 2021-2025).

Con base en el dataset de consulta externa del MSP, es necesario calcular el indicador de Control Prenatal correspondiente al año 2021, y los resultados obtenidos deben ser visualizados en un tablero mediante una herramienta de Business Intelligence.

Es importante realizar un estudio previo de los datos, es así que se procedió a realizarlo a través de Python, con el proceso detallado a continuación:

3.1.1. Comprender el problema

Contenido del conjunto de datos:

Para entender el contenido se tiene que definir el significado y la relevancia de cada variable con respecto al problema.

Los indicadores de salud están sujetos al análisis de las variables CODIGO_CIE10_1, CODIGO_CIE10_2, CODIGO_CIE10_3 (Código del diagnóstico CIE-10 el cual se obtiene de la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud).

Las demás variables se describen de la siguiente manera:

Ilustración 3. Variables contenidas en el Dataset

```
In [11]: # Revisaremos las columnas contenidas dentro del dataset
```

```
print('Información en el dataset:')  
print(data.keys())
```

```
Información en el dataset:  
Index(['ZONA', 'PROVINCIA', 'CANTON', 'CODIGO_ESTABLECIMIENTO',  
       'FECHA_ATENCION', 'FECHA_NACIMIENTO', 'SEXO', 'NACIONALIDAD',  
       'AUTOIDENTIFICACION_ETNICA', 'PROVINCIA_RESIDENCIA',  
       'CANTON_RESIDENCIA', 'GRUPO_PRIORITARIO_1', 'GRUPO_PRIORITARIO_2',  
       'GRUPO_PRIORITARIO_3', 'CODIGO_CIE10_1', 'DESCRIPCION_CIE10_1',  
       'TIPO_ATENCION1', 'CONDICION_DIAGNOSTICO1', 'CODIGO_CIE10_2',  
       'DESCRIPCION_CIE10_2', 'TIPO_ATENCION2', 'CONDICION_DIAGNOSTICO2',  
       'CODIGO_CIE10_3', 'DESCRIPCION_CIE10_3', 'TIPO_ATENCION3',  
       'CONDICION_DIAGNOSTICO3', 'EDAD_ANIOS'],  
      dtype='object')
```

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Cantidad de datos:

Para conocer la cantidad de datos que se tiene que analizar se muestra las dimensiones del Dataset, de la siguiente manera:

Ilustración 4. Dimensiones del Dataset

```
In [10]: #Muestra Las Dimensiones del DataSet  
print(data.shape)
```

```
(17493594, 27)
```

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Variables del dataset:

Para identificar qué variables componen el dataset se compone en su mayoría por variables categóricas que se detallan a continuación:

Ilustración 5. Tipo de variables y tipo de datos en el Dataset

```
In [20]: #Permite conocer el tipo de dato y de la clase de variable |
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17493594 entries, 0 to 17493593
Data columns (total 27 columns):
#   Column                                Dtype
---  -
0   ZONA                                    object
1   PROVINCIA                              object
2   CANTON                                  object
3   CODIGO_ESTABLECIMIENTO                 int64
4   FECHA_ATENCION                         object
5   FECHA_NACIMIENTO                      object
6   SEXO                                    object
7   NACIONALIDAD                          object
8   AUTOIDENTIFICACION_ETNICA            object
9   PROVINCIA_RESIDENCIA                  object
10  CANTON_RESIDENCIA                      object
11  GRUPO_PRIORITARIO_1                   object
12  GRUPO_PRIORITARIO_2                   object
13  GRUPO_PRIORITARIO_3                   object
14  CODIGO_CIE10_1                        object
15  DESCRIPCION_CIE10_1                   object
16  TIPO_ATENCION1                        object
17  CONDICION_DIAGNOSTICO1                object
18  CODIGO_CIE10_2                        object
19  DESCRIPCION_CIE10_2                   object
20  TIPO_ATENCION2                        object
21  CONDICION_DIAGNOSTICO2                object
22  CODIGO_CIE10_3                        object
23  DESCRIPCION_CIE10_3                   object
24  TIPO_ATENCION3                        object
25  CONDICION_DIAGNOSTICO3                object
26  EDAD_ANIOS                            int64
dtypes: int64(2), object(25)
memory usage: 3.5+ GB
```

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.1.2. Diagnósticos Iniciales

Para comprender qué códigos de tipo de diagnóstico prevalecen en el dataset se muestra una tabla de frecuencias.

Ilustración 6. Frecuencia de la variable CODIGO_CIE10_1

```
In [132]: # tabla de frecuencia de CODIGO_CIE10_1
pd.value_counts(data['CODIGO_CIE10_1'])
```

```
Out[132]: Z000      1302838
          Z012      1086553
          Z001      1083925
          K021       821967
          I10X       763564
          ...
          T794         1
          C104         1
          T817         1
          C101         1
          Z994         1
          Name: CODIGO_CIE10_1, Length: 9249, dtype: int64
```

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Las variables priorizadas en este ejemplo son:

- **CODIGO_CIE10_1**: Código del diagnóstico CIE-10 el cual se obtiene de la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud.
- **ZONA**: Zona a la que pertenece el establecimiento de salud.
- **PROVINCIA**: Provincia a la que pertenece el establecimiento de salud.
- **EDAD ANIOS**: Edad del paciente.
- **NACIONALIDAD**: Nacionalidad del paciente.
- **AUTOIDENTIFICACION ETNICA**: Auto identificación étnica del paciente.

Con base en las principales variables identificadas a continuación se interpreta las relaciones entre variables sugeridas:

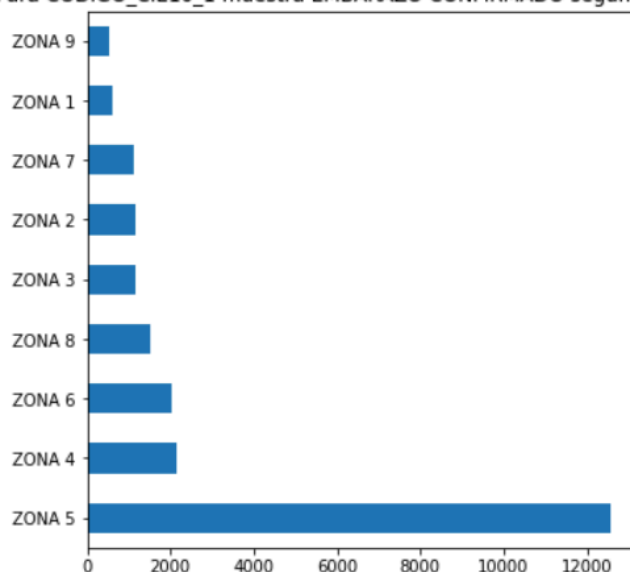
1. Para identificar embarazos confirmados por zona, se utilizan las variables CODIGO_CIE10_1=Z321 y ZONA. Se observa en la siguiente ilustración que una gran concentración de embarazos se da en la Zona 5.

Ilustración 7. Embarazo Confirmado según la Zona

In [33]:

```
# 1. Gráfico de EMBARAZO CONFIRMADO según Los ZONA
df = data[data['CODIGO_CIE10_1'] == "Z321"]
plot = df['ZONA'].value_counts().plot(kind='barh',
figsize=(6, 6),
title='Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los ZONA')
```

Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los ZONA



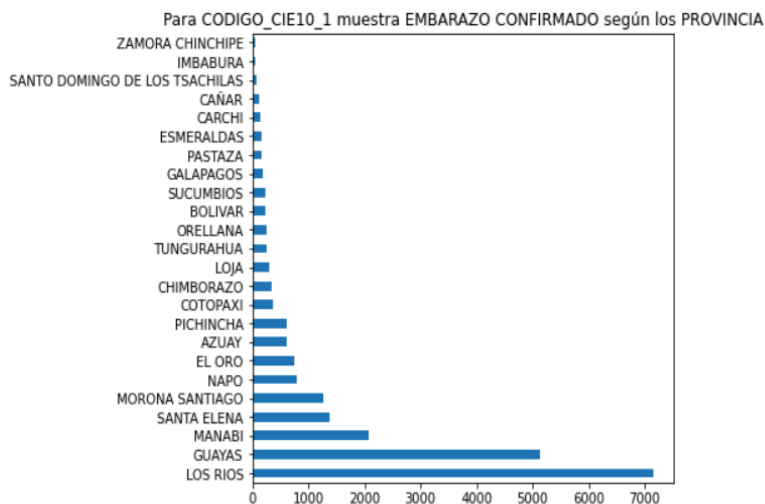
Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

2. Para identificar embarazos confirmados por provincia, se utilizan las variables CODIGO_CIE10_1=Z321 y PROVINCIA. Se observa en la siguiente ilustración las provincias con mayor concentración de embarazos son Los Ríos y Guayas.

Ilustración 8. Embarazo Confirmado según la Provincia

```
In [32]: # 2. Gráfico de EMBARAZO CONFIRMADO según Los PROVINCIA
df = data[data['CODIGO_CIE10_1'] == "Z321"]
plot = df['PROVINCIA'].value_counts().plot(kind='barh',
figsize=(6, 6),
title='Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los PROVINCIA')
```



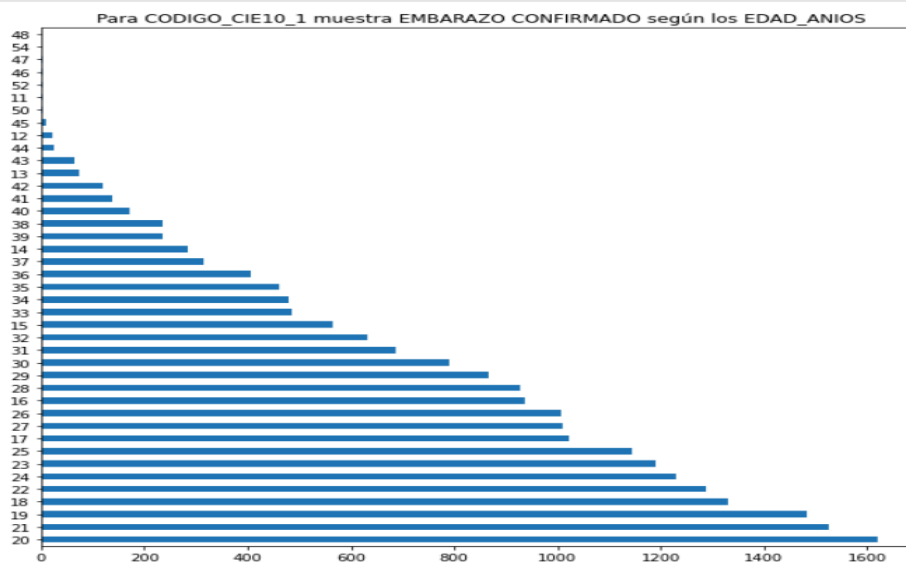
Fuente: Datos de MSP - Indicador CCP - 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

4. Para identificar embarazos confirmados por Edad, se utilizan las variables CODIGO_CIE10_1=Z321 y EDAD. Se observa en la siguiente ilustración que aproximadamente más del 50% de los embarazos se concentran entre las edades de 17 a 25 años.

Ilustración 9. Embarazo Confirmado según la Edad

```
In [41]: # 3. Gráfico de EMBARAZO CONFIRMADO según Los EDAD_ANIOS
df = data[data['CODIGO_CIE10_1'] == "Z321"]
plot = df['EDAD_ANIOS'].value_counts().plot(kind='barh',
figsize=(10, 10),
title='Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los EDAD_ANIOS')
```



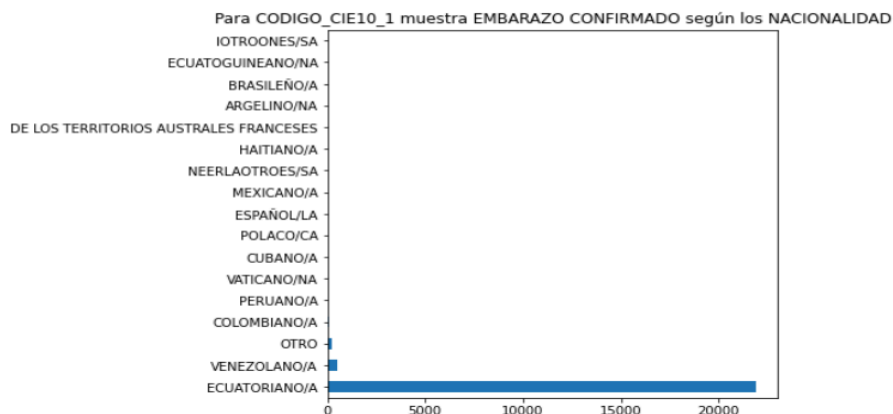
Fuente: Datos de MSP - Indicador CCP - 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Para identificar embarazos confirmados por Nacionalidad, se utilizan las variables CODIGO_CIE10_1=Z321 y NACIONALIDAD. Se observa en la siguiente ilustración que la segunda nacionalidad en registrar embarazo es la venezolana después de la ecuatoriana.

Ilustración 10. Embarazo Confirmado según la Nacionalidad

```
In [42]: # 4. Gráfico de EMBARAZO CONFIRMADO según Los NACIONALIDAD
df = data[data['CODIGO_CIE10_1'] == "Z321"]
plot = df['NACIONALIDAD'].value_counts().plot(kind='barh',
figsize=(6, 6),
title='Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los NACIONALIDAD')
```



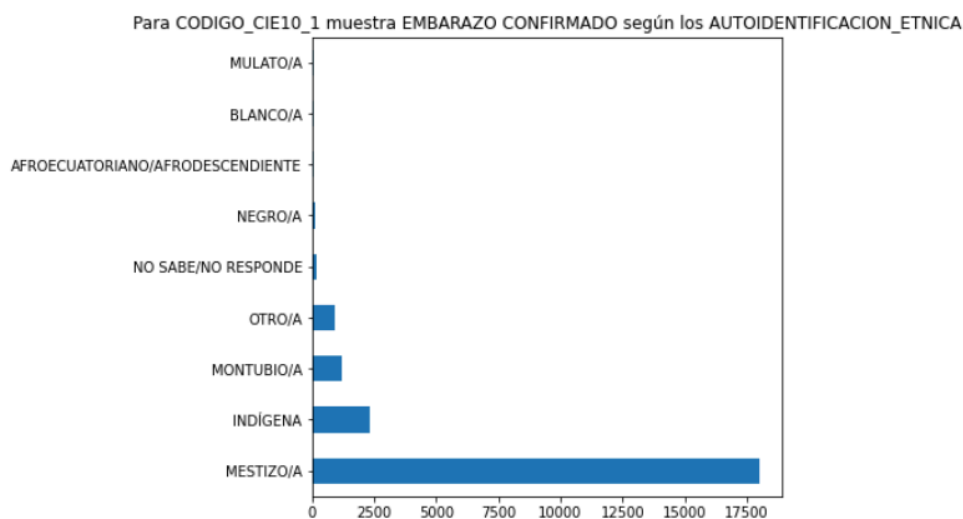
Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

5. Para identificar embarazos confirmados por Identidad Étnica, se utiliza las variables CODIGO_CIE10_1=Z321 y AUTOIDENTIFICACION_ETNICA. Se observa en la siguiente ilustración que en relación a la etnia la población que predomina en los embarazos es la mestiza, seguida de la indígena y la montubia.

Ilustración 11. Embarazo Confirmado según la Etnia

```
In [43]: # 5. Gráfico de EMBARAZO CONFIRMADO según los AUTOIDENTIFICACION_ETNICA
df = data[data['CODIGO_CIE10_1'] == "Z321"]
plot = df['AUTOIDENTIFICACION_ETNICA'].value_counts().plot(kind='barh',
figsize=(6, 6),
title='Para CODIGO_CIE10_1 muestra EMBARAZO CONFIRMADO según los AUTOIDENTIFICACION_ETNICA')
```



Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

En conclusión, después de realizar una revisión general de los datos y la interrelación de las variables se puede identificar claramente la temática sobre la cual debe presentar información del dashboard en una herramienta de visualización sin perder la visión de negocio.

3.1.3. Depuración del dataset

Una vez realizada la tarea de análisis exploratorio inicial de los datos, se procede a depurar y preparar el dataset para poder calcular el indicador que se ha planteado como alcance, es decir el Control Prenatal del 2021.

Dentro de la herramienta Jupyter fue creado un nuevo Notebook; a continuación, vamos a revisar el código que contiene el mencionado script.

1. Se realiza la importación de las librerías a utilizar:

Ilustración 12. Importación de librerías

```
In [1]: import pandas as pd
import datetime
from dateutil.relativedelta import relativedelta
import numpy as np
```

Fuente: Jupyter Noteboook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

2. Se crea un dataframe con los datos del archivo original del archivo: Consulta Externa 2021.csv.

Ilustración 13. Creación de dataframe


```

In [2]: df = pd.read_csv(r'D:\DESCARGAS\EDGE\Consulta Externa 2021\Consulta Externa 2021.csv', sep=';',
                        dtype={"DESCRIPCION_CIE10_2": "string", "DESCRIPCION_CIE10_3": "string"})

In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17493594 entries, 0 to 17493593
Data columns (total 27 columns):
#   Column                               Dtype
---  ---
0   ZONA                                  object
1   PROVINCIA                             object
2   CANTON                                 object
3   CODIGO_ESTABLECIMIENTO                int64
4   FECHA_ATENCION                        object
5   FECHA_NACIMIENTO                      object
6   SEXO                                   object
7   NACIONALIDAD                          object
8   AUTOIDENTIFICACION_ETNICA            object
9   PROVINCIA_RESIDENCIA                  object
10  CANTON_RESIDENCIA                      object
11  GRUPO_PRIORITARIO_1                   object
12  GRUPO_PRIORITARIO_2                   object
13  GRUPO_PRIORITARIO_3                   object
14  CODIGO_CIE10_1                        object
15  DESCRIPCION_CIE10_1                   object
16  TIPO_ATENCION1                        object
17  CONDICION_DIAGNOSTICO1                object
18  CODIGO_CIE10_2                        object
19  DESCRIPCION_CIE10_2                   string
20  TIPO_ATENCION2                        object
21  CONDICION_DIAGNOSTICO2                object
22  CODIGO_CIE10_3                        object
23  DESCRIPCION_CIE10_3                   string
24  TIPO_ATENCION3                        object
25  CONDICION_DIAGNOSTICO3                object
26  EDAD_ANIOS                            int64
dtypes: int64(2), object(23), string(2)
memory usage: 3.5+ GB

```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamani, Claudia Jaramillo

3. Con los datos cargados se crea una columna como identificador único de cada registro, se crea un nuevo data frame filtrado solo con los registros “MUJER”. Y se visualiza la información de este nuevo data frame creado.

Ilustración 14. Creación de columnas jupyter

Adicionamos una columna con un identificador único para cada registro

```
In [4]: df['id'] = df.index
```

Creamos un data set filtrado de mujeres para el indicador prenatal

```
In [5]: df_mujeres=df[df['SEXO']=="MUJER"].copy()
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

4. Para realizar las consultas se utiliza los siguientes comandos:

Ilustración 15. Consulta de información Jupyter

Consultamos la info general del dataframe creado

```
In [6]: df_mujeres.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11587025 entries, 2 to 17493593
Data columns (total 28 columns):
#   Column                                Dtype
---  ---
0   ZONA                                   object
1   PROVINCIA                             object
2   CANTON                                 object
3   CODIGO_ESTABLECIMIENTO                int64
4   FECHA_ATENCION                        object
5   FECHA_NACIMIENTO                     object
6   SEXO                                  object
7   NACIONALIDAD                          object
8   AUTOIDENTIFICACION_ETNICA            object
9   PROVINCIA_RESIDENCIA                  object
10  CANTON_RESIDENCIA                      object
11  GRUPO_PRIORITARIO_1                   object
12  GRUPO_PRIORITARIO_2                   object
13  GRUPO_PRIORITARIO_3                   object
14  CODIGO_CIE10_1                        object
15  DESCRIPCION_CIE10_1                   object
16  TIPO_ATENCION1                        object
17  CONDICION_DIAGNOSTICO1                 object
18  CODIGO_CIE10_2                        object
19  DESCRIPCION_CIE10_2                    string
20  TIPO_ATENCION2                        object
21  CONDICION_DIAGNOSTICO2                 object
22  CODIGO_CIE10_3                        object
23  DESCRIPCION_CIE10_3                    string
24  TIPO_ATENCION3                        object
25  CONDICION_DIAGNOSTICO3                 object
26  EDAD_ANIOS                            int64
27  id                                     int64
dtypes: int64(3), object(23), string(2)
memory usage: 2.5+ GB
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Se crean y se llenan tres campos nuevos para poder filtrar las atenciones que corresponden a control prenatal, en el segundo campo se registra si corresponde PRIMERA atención y en el tercer campo se registra si se cumplen o no ambas condiciones. Esto se aplica para los tres campos donde se registran las atenciones para cada registro.

Ilustración 16. Ingreso de datos para filtros de información Jupyter

```
In [7]: df_mujeres['AT_CIE1'] = df['CODIGO_CIE10_1'].apply(lambda x: 1 if x in
    ['Z340', 'Z348', 'Z349', 'Z350', 'Z351', 'Z352', 'Z353', 'Z354', 'Z355', 'Z356',
    'Z357', 'Z358', 'Z359', 'Z321'] else 0)
df_mujeres['AT_PRIM1'] = df['TIPO_ATENCION1'].apply(lambda x: 1 if x in ['PRIMERA'] else 0)
df_mujeres['AT_1'] = df_mujeres['AT_CIE1'] * df_mujeres['AT_PRIM1']

In [8]: df_mujeres['AT_CIE2'] = df['CODIGO_CIE10_2'].apply(lambda x: 1 if x in
    ['Z340', 'Z348', 'Z349', 'Z350', 'Z351', 'Z352', 'Z353', 'Z354', 'Z355', 'Z356',
    'Z357', 'Z358', 'Z359', 'Z321'] else 0)
df_mujeres['AT_PRIM2'] = df['TIPO_ATENCION2'].apply(lambda x: 1 if x in ['PRIMERA'] else 0)
df_mujeres['AT_2'] = df_mujeres['AT_CIE2'] * df_mujeres['AT_PRIM2']

In [9]: df_mujeres['AT_CIE3'] = df['CODIGO_CIE10_3'].apply(lambda x: 1 if x in
    ['Z340', 'Z348', 'Z349', 'Z350', 'Z351', 'Z352', 'Z353', 'Z354', 'Z355', 'Z356',
    'Z357', 'Z358', 'Z359', 'Z321'] else 0)
df_mujeres['AT_PRIM3'] = df['TIPO_ATENCION3'].apply(lambda x: 1 if x in ['PRIMERA'] else 0)
df_mujeres['AT_3'] = df_mujeres['AT_CIE3'] * df_mujeres['AT_PRIM3']
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Se aplica una fórmula de suma de las banderas de los tres campos de atención para identificar si cada registro cumple con al menos una atención primaria correspondiente a control prenatal.

Ilustración 17. Creación de banderas Jupyter

```
In [10]: df_mujeres['AT_PREN_FINAL'] = df_mujeres['AT_1'] + df_mujeres['AT_2'] + df_mujeres['AT_3']

In [11]: df_mujeres_at_pren = df_mujeres[df_mujeres['AT_PREN_FINAL'] >= 1].copy()
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Se exporta el resultado con cada registro que cumple las condiciones para ser cargado en la BDD.

Ilustración 18. Resultados de filtro de información Jupyter

```
In [12]: df_mujeres_at_pren.to_csv("ATEN_PREN_RESULTADO.csv")

In [13]: df_mujeres_at_pren.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 269296 entries, 1538572 to 16932586
Data columns (total 38 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   ZONA                                   269296 non-null object
1   PROVINCIA                             269296 non-null object
2   CANTON                                  269296 non-null object
3   CODIGO_ESTABLECIMIENTO                 269296 non-null int64
4   FECHA_ATENCION                         269296 non-null object
5   FECHA_NACIMIENTO                       269296 non-null object
6   SEXO                                    269296 non-null object
7   NACIONALIDAD                           269296 non-null object
8   AUTOIDENTIFICACION_ETNICA             269296 non-null object
9   PROVINCIA_RESIDENCIA                  263804 non-null object
10  CANTON_RESIDENCIA                      263804 non-null object
11  GRUPO_PRIORITARIO_1                   269296 non-null object
12  GRUPO_PRIORITARIO_2                   269296 non-null object
13  GRUPO_PRIORITARIO_3                   269296 non-null object
14  CODIGO_CIE10_1                        269296 non-null object
15  DESCRIPCION_CIE10_1                   269296 non-null object
16  TIPO_ATENCION1                        269296 non-null object
17  CONDICION_DIAGNOSTICO1                 269296 non-null object
18  CODIGO_CIE10_2                        269296 non-null object
19  DESCRIPCION_CIE10_2                   164116 non-null string
20  TIPO_ATENCION2                        269296 non-null object
21  CONDICION_DIAGNOSTICO2                 269296 non-null object
22  CODIGO_CIE10_3                        269296 non-null object
23  DESCRIPCION_CIE10_3                   92706 non-null string
24  TIPO_ATENCION3                        269296 non-null object
25  CONDICION_DIAGNOSTICO3                 269296 non-null object
26  EDAD_ANIOS                            269296 non-null int64
27  id                                     269296 non-null int64
28  AT_CIE                                 269296 non-null int64
29  AT_PRIM                                 269296 non-null int64
30  AT_1                                    269296 non-null int64
31  AT_CIE2                                269296 non-null int64
32  AT_PRIM2                                269296 non-null int64
33  AT_2                                    269296 non-null int64
34  AT_CIE3                                269296 non-null int64
35  AT_PRIM3                                269296 non-null int64
36  AT_3                                    269296 non-null int64
37  AT_PREN_FINAL                          269296 non-null int64
dtypes: int64(13), object(23), string(2)
memory usage: 80.1+ MB
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Adicionalmente se exportan los resultados agregados por provincia para que sean cargados a la BDD.

Ilustración 19. Exportación de información Jupyter

```
In [14]: df_mujeres_at_pren_group=df_mujeres_at_pren.groupby(['PROVINCIA'])['AT_PREN_FINAL'].count()

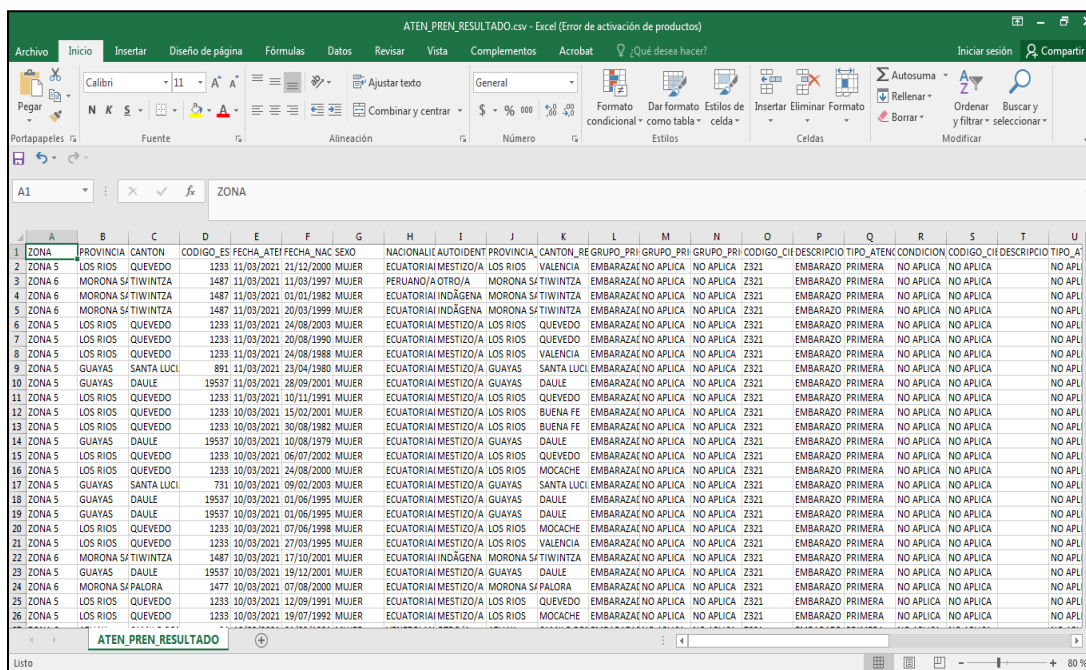
In [15]: df_mujeres_at_pren_group.to_csv("ATEN_PREN_RESULTADO_AGR.csv")
```

Fuente: Jupyter Notebook

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Finalmente se obtiene como salida el archivo .csv que contiene los registros (269.296 de un total de 17'493.594) y que van a ser tomados en cuenta para el cálculo del indicador.

Ilustración 20. Dataset resultante del proceso realizado con la herramienta Python



ZONA	PROVINCIA	CANTON	CODIGO_ES	FECHA_ATEI	FECHA	NAC	SEXO	NACIONALIDAD	AUTOIDENT	PROVINCIA	CANTON	RE	GRUPO_PRI	GRUPO_PRI	GRUPO_PRI	CODIGO_CHE	DESCRIPCIO	TIPO	ATEN	CONDICION	CODIGO_CHE	DESCRIPCIO	TIPO	ATEN
1	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	21/12/2000	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	VALENCIA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
2	ZONA 6	MORONA SÍ	TIWINTZA	1487	11/03/2021	11/03/1997	MUJER	PERUANO/A	OTRO/A	MORONA SÍ	TIWINTZA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
3	ZONA 6	MORONA SÍ	TIWINTZA	1487	11/03/2021	01/01/1982	MUJER	ECUATORIAI	INDÍGENA	MORONA SÍ	TIWINTZA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
4	ZONA 6	MORONA SÍ	TIWINTZA	1487	11/03/2021	20/03/1999	MUJER	ECUATORIAI	INDÍGENA	MORONA SÍ	TIWINTZA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
5	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	24/08/2003	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	QUEVEDO	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
6	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	20/08/1990	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	VALENCIA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
7	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	24/08/1988	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	VALENCIA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
8	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	23/04/1980	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	SANTA LUCI	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
9	ZONA 5	GUAYAS	SANTA LUCI	891	11/03/2021	28/09/2001	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
10	ZONA 5	GUAYAS	DAULE	19537	11/03/2021	10/11/1991	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	QUEVEDO	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
11	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	15/02/2001	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	BUENA FE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
12	ZONA 5	LOS RIOS	QUEVEDO	1233	11/03/2021	30/08/1982	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	BUENA FE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
13	ZONA 5	GUAYAS	DAULE	19537	10/03/2021	10/08/1979	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
14	ZONA 5	GUAYAS	DAULE	19537	10/03/2021	06/07/2002	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	QUEVEDO	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
15	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	24/08/2000	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	MOCACHE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
16	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	09/02/2003	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	SANTA LUCI	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
17	ZONA 5	GUAYAS	SANTA LUCI	731	10/03/2021	01/06/1995	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
18	ZONA 5	GUAYAS	DAULE	19537	10/03/2021	01/06/1995	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
19	ZONA 5	GUAYAS	DAULE	19537	10/03/2021	01/06/1995	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
20	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	07/06/1998	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	MOCACHE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
21	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	27/03/1995	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	VALENCIA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
22	ZONA 6	MORONA SÍ	TIWINTZA	1487	10/03/2021	17/10/2001	MUJER	ECUATORIAI	INDÍGENA	MORONA SÍ	TIWINTZA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
23	ZONA 5	GUAYAS	DAULE	19537	10/03/2021	19/12/2001	MUJER	ECUATORIAI	MESTIZO/A	GUAYAS	DAULE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
24	ZONA 6	MORONA SÍ	PALORA	1477	10/03/2021	07/08/2000	MUJER	ECUATORIAI	MESTIZO/A	MORONA SÍ	PALORA	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
25	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	12/09/1991	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	QUEVEDO	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA
26	ZONA 5	LOS RIOS	QUEVEDO	1233	10/03/2021	19/07/1992	MUJER	ECUATORIAI	MESTIZO/A	LOS RIOS	MOCACHE	EMBARAZAE	NO APLICA	NO APLICA	2321	EMBARAZO	PRIMERA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA	NO APLICA

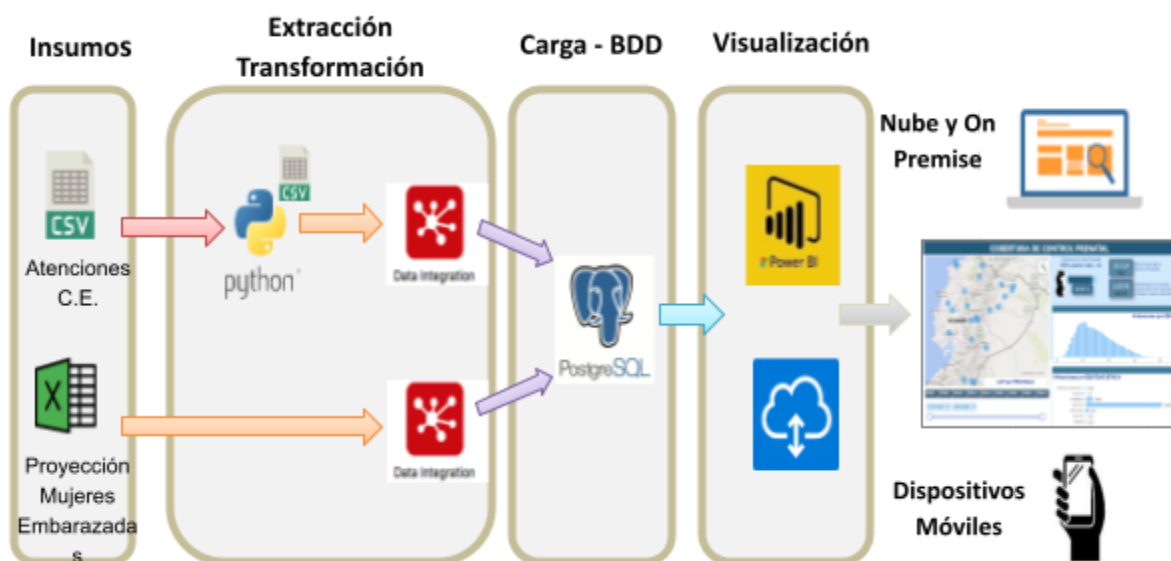
Fuente: Datos de MSP – Indicador CCP – 2021, MS Excel

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.2. Diseño de Arquitectura y artefactos

Y con el set de datos listos para ser tomados como insumo, se procede a definir la arquitectura y las diferentes herramientas para ejecutar el proceso de creación del dashboard del indicador de Control Prenatal del 2021.

Ilustración 21. Arquitectura del proceso de creación de un dashboard de indicadores de salud pública



Fuente: Realización Propia

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.3. Diseño de modelos

3.3.1. Definir los modelos dimensionales

El modelado dimensional es un método común para estructurar datos de BI y análisis en forma de un esquema en estrella. Un modelo dimensional permite

que los datos en un almacén de datos se representen de una manera diseñada para proporcionar un rendimiento rápido en consultas analíticas. Para lo cual es importante definir la tabla de hechos y tablas de dimensiones.

Medidas

Las medidas son valores cuantificables que se requiere medir o analizar.

- Indicador CCP: que mostrará el porcentaje (%) de cobertura por primera consulta de control prenatal por provincia en un tiempo determinado.
- Numerador NACCP: Número de primeras atenciones de control prenatal por provincia en un tiempo determinado.
- Denominador PME: Estimación de la población de mujeres embarazadas por provincia en un tiempo determinado.

Dimensiones

Las dimensiones son datos cualitativos de las medidas y define cómo se quiere medir

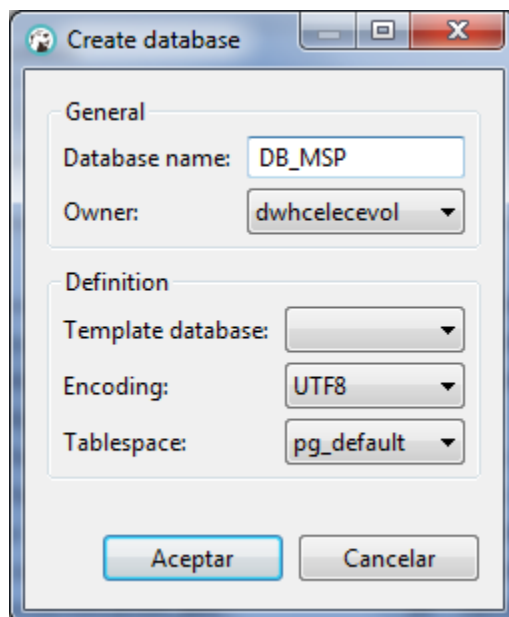
- Provincia: todas las provincias del Ecuador.
- Tiempo: días y meses del año de estudio 2021.
- Paciente: Datos básicos del paciente como; nombre, edad, identidad étnica, etc.

3.3.2. Crear el modelo de datos en PostgreSQL

Para crear la base de datos DB_MSP y las tablas Embarazadas y aten_pren_resultado se realiza los siguientes pasos.

1. Con la aplicación DBeaver 23.1.0 se crea la base de datos DB_MSP en PostgreSQL.

Ilustración 22. Creación de la Base de Datos “DB_MSP”

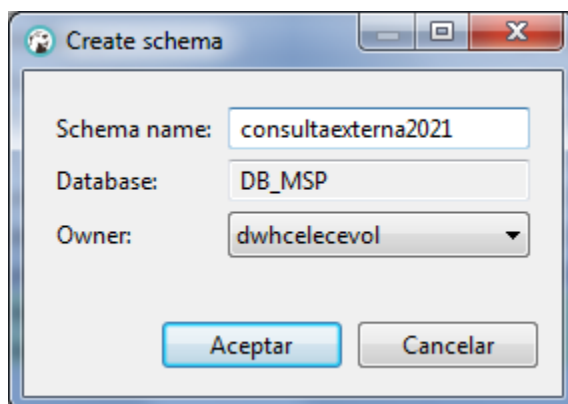


Fuente: DBeaver.

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

2. Crear el esquema consultaexterna2021 en la base de datos DB_MSP.

Ilustración 23. Creación del esquema consultaexterna2021.



Fuente: DBeaver.

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3. En el esquema consultaexterna2021 crear la tabla Embarazadas que contiene información de la cantidad de mujeres embarazadas por provincia con los siguientes campos:

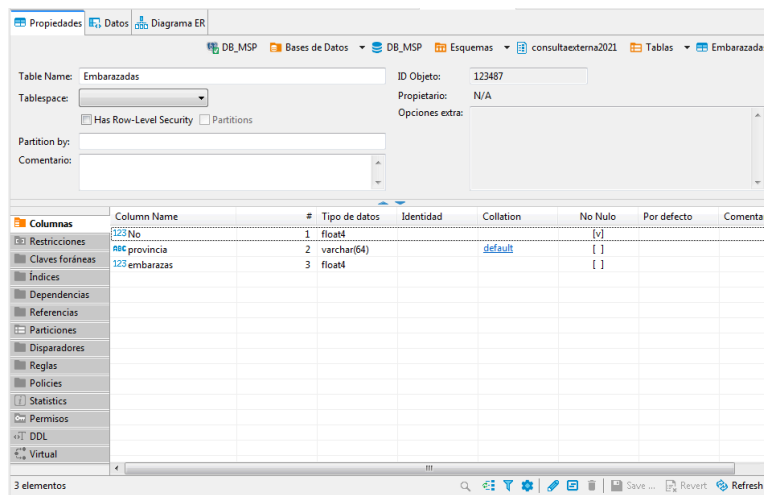
Tabla 2. Tabla Embarazadas.

Nombre Columna	Tipo de datos
No	float4
provincia	varchar(64)
embarazadas	float4

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Ilustración 24. Tabla Embarazadas PostgreSQL.



Fuente: DBeaver.

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

4. En el esquema consultaexterna2021 crear la tabla aten_pren_resultado con los siguientes campos:

Tabla 3. Tabla aten_pren_resultado.

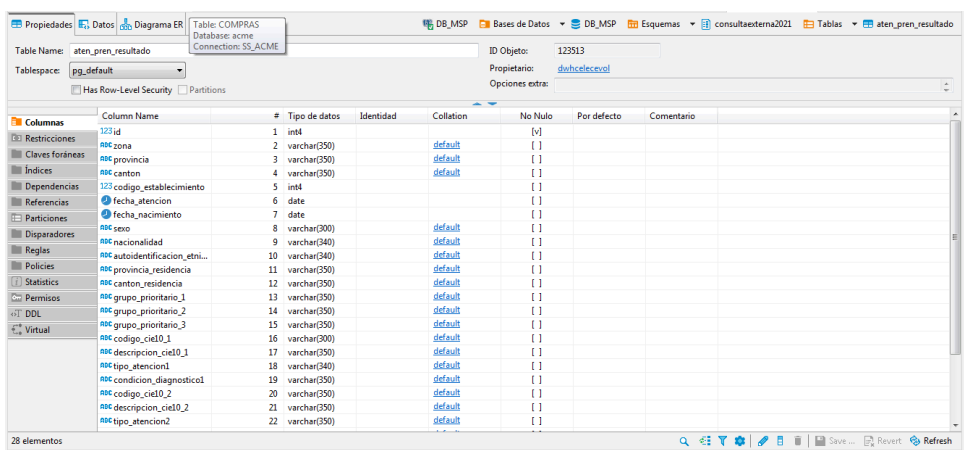
Nombre Columna	Tipo de datos
id	int4
zona	varchar(350)
provincia	varchar(350)
canton	varchar(350)
codigo_establecimiento	int4
fecha_atencion	date
fecha_nacimiento	date
sexo	varchar(300)
nacionalidad	varchar(340)
autoidentificacion_etnica	varchar(340)
provincia_residencia	varchar(350)
canton_residencia	varchar(340)
grupo_prioritario_1	varchar(350)
grupo_prioritario_2	varchar(350)
grupo_prioritario_3	varchar(350)
codigo_cie10_1	varchar(300)
descripcion_cie10_1	varchar(350)
tipo_atencion1	varchar(340)
condicion_diagnostico1	varchar(350)
codigo_cie10_2	varchar(350)
descripcion_cie10_2	varchar(350)
tipo_atencion2	varchar(350)
condicion_diagnostico2	varchar(350)
codigo_cie10_3	varchar(350)
descripcion_cie10_3	varchar(350)
tipo_atencion3	varchar(350)
condicion_diagnostico3	varchar(350)

edad_anios	int4
------------	------

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Ilustración 25. Creación de Tabla aten_pren_resultado postgreSQL.



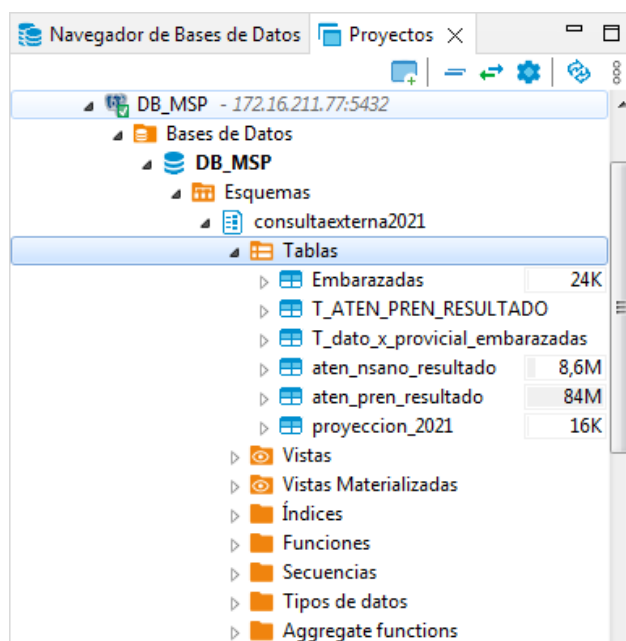
Columnas	Column Name	#	Tipo de datos	Identidad	Collation	No Nulo	Por defecto	Comentario
Restricciones	123 id	1	int4			[v]		
Claves foráneas	84 zona	2	varchar(350)		default	[]		
Indicadores	84 provincia	3	varchar(350)		default	[]		
Indicadores	84 canton	4	varchar(350)		default	[]		
Dependencias	123 codigo_establecimiento	5	int4			[]		
Referencias	4 fecha_atencion	6	date			[]		
Particiones	4 fecha_nacimiento	7	date			[]		
Disparadores	84 sexo	8	varchar(300)		default	[]		
Reglas	84 nacionalidad	9	varchar(340)		default	[]		
Políticas	84 autoidentificacion_etni...	10	varchar(340)		default	[]		
Statistics	84 provincia_residencia	11	varchar(350)		default	[]		
Permisos	84 canton_residencia	12	varchar(350)		default	[]		
DDL	84 grupo_prioritario_1	13	varchar(350)		default	[]		
Virtual	84 grupo_prioritario_2	14	varchar(350)		default	[]		
	84 grupo_prioritario_3	15	varchar(350)		default	[]		
	84 codigo_cie10_1	16	varchar(300)		default	[]		
	84 descripcion_cie10_1	17	varchar(350)		default	[]		
	84 tipo_atencion1	18	varchar(340)		default	[]		
	84 condicion_diagnostico1	19	varchar(350)		default	[]		
	84 codigo_cie10_2	20	varchar(350)		default	[]		
	84 descripcion_cie10_2	21	varchar(350)		default	[]		
	84 tipo_atencion2	22	varchar(350)		default	[]		

Fuente: DBeaver.

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Se puede visualizar la base de datos DB_MSP con las tablas embarazadas y aten_pren_resultado.

Ilustración 26. Base de datos DB_MSP.



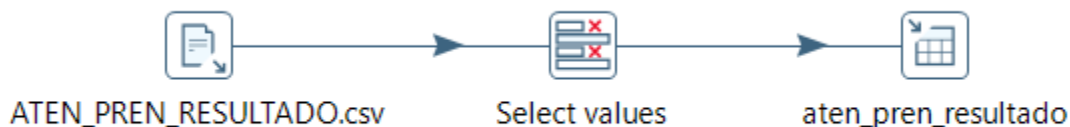
Fuente: DBeaver.

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.4. Diseño de ETL

Para cargar la información del dataset resultante del proceso realizado con la herramienta Python a la base de datos DB_MSP se realizó el proceso ETL con la herramienta Pentaho.

Ilustración 27. Proceso ETL de datos sobre el indicador Control Prenatal del 2021.



Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Para el proceso ETL de la información de atención prenatal se siguieron los siguientes pasos:

1. Configurar el step CSV file input con los siguientes parámetros con el fin de leer los datos del archivo “ATEN_PREN_RESULTADO.csv”, dataset resultante del paso 4.1.3. Depuración del dataset punto 8.

Ilustración 28. Configuración parámetros lectura archivo ATEN_PREN_RESULTADO.csv.

CSV file input

Step name: ATEN_PREN_RESULTADO.csv

Filename: C:\Users\wilmer.guamaní\Downloads\ATEN_PREN_RESULTADO.csv Browse...

Delimiter: ; Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional):

Running in parallel?

New line possible in fields?

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	ZONA	String		6		\$,	.	none
2	PROVINCIA	String		15		\$,	.	none
3	CANTON	String		21		\$,	.	none
4	CODIGO_ESTABLECIMIENTO	Integer	#	15	0	\$,	.	none
5	FECHA_ATENCION	Date	dd/MM/yyyy			\$,	.	none
6	FECHA_NACIMIENTO	Date	dd/MM/yyyy			\$,	.	none
7	SEXO	String		5		\$,	.	none
8	NACIONALIDAD	String		13		\$,	.	none
9	AUTOIDENTIFICACION_ETNICA	String		10		\$,	.	none
1..	PROVINCIA_RESIDENCIA	String		15		\$,	.	none
1..	CANTON_RESIDENCIA	String		21		\$,	.	none
1..	GRUPO_PRIORITARIO_1	String		11		\$,	.	none
1..	GRUPO_PRIORITARIO_2	String		9		\$,	.	none
1..	GRUPO_PRIORITARIO_3	String		9		\$,	.	none
1..	CODIGO_CIE10_1	String		4		\$,	.	none
1..	DESCRIPCION_CIE10_1	String		19		\$,	.	none
1..	TIPO_ATENCION1	String		7		\$,	.	none
1..	CONDICION_DIAGNOSTICO1	String		9		\$,	.	none
1..	CODIGO_CIE10_2	String		9		\$,	.	none
2..	DESCRIPCION_CIE10_2	Boolean				\$,	.	none
2..	TIPO_ATENCION2	String		9		\$,	.	none
2..	CONDICION_DIAGNOSTICO2	String		9		\$,	.	none
2..	CODIGO_CIE10_3	String		9		\$,	.	none

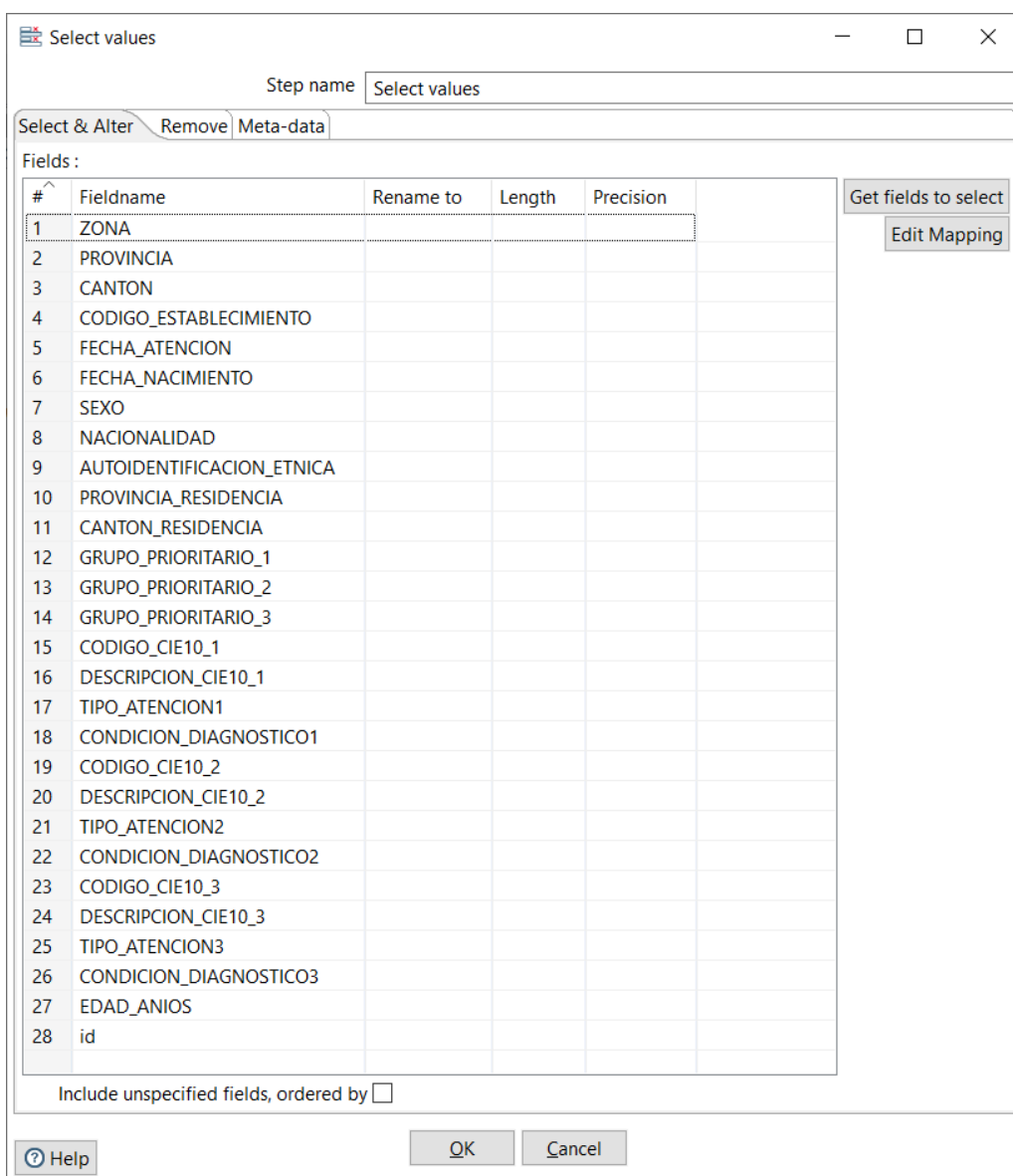
Buttons: Help, OK, Get Fields, Preview, Cancel

Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Step name: ATEN_PREN_RESULTADO.csv.
 - Filename: Seleccionamos el archivo ATEN_PREN_RESULTADO.csv.
 - Delimiter: Se coloca el carácter que separa las columnas (;).
 - Con Get Fields se obtienen todos los campos del archivo ATEN_PREN_RESULTADO.csv.
2. Con el Select values seleccionar las columnas que se van a cargar a la base de datos “DB_MSP”.

Ilustración 29. Selección de los campos del archivo ATEN_PREN_RESULTADO.csv.



Step name: Select values

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	ZONA			
2	PROVINCIA			
3	CANTON			
4	CODIGO_ESTABLECIMIENTO			
5	FECHA_ATENCION			
6	FECHA_NACIMIENTO			
7	SEXO			
8	NACIONALIDAD			
9	AUTOIDENTIFICACION_ETNICA			
10	PROVINCIA_RESIDENCIA			
11	CANTON_RESIDENCIA			
12	GRUPO_PRIORITARIO_1			
13	GRUPO_PRIORITARIO_2			
14	GRUPO_PRIORITARIO_3			
15	CODIGO_CIE10_1			
16	DESCRIPCION_CIE10_1			
17	TIPO_ATENCION1			
18	CONDICION_DIAGNOSTICO1			
19	CODIGO_CIE10_2			
20	DESCRIPCION_CIE10_2			
21	TIPO_ATENCION2			
22	CONDICION_DIAGNOSTICO2			
23	CODIGO_CIE10_3			
24	DESCRIPCION_CIE10_3			
25	TIPO_ATENCION3			
26	CONDICION_DIAGNOSTICO3			
27	EDAD_ANIOS			
28	id			

Include unspecified fields, ordered by

Buttons: Get fields to select, Edit Mapping, Help, OK, Cancel

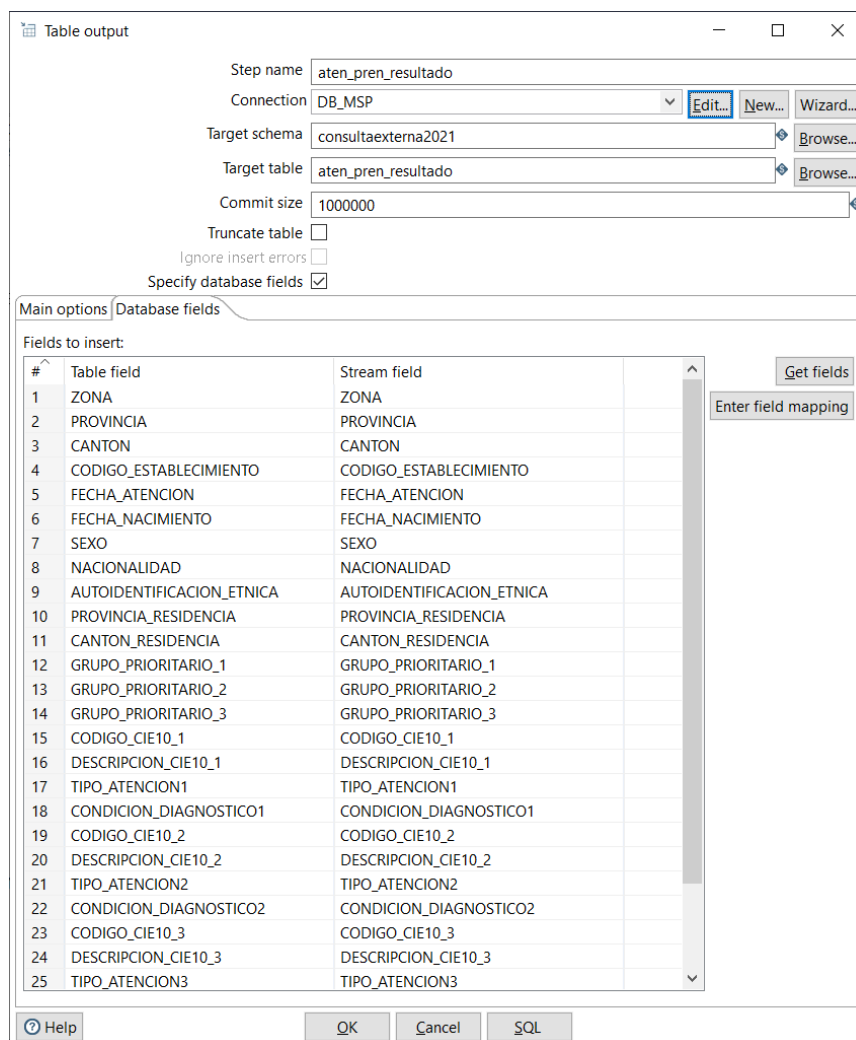
Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Con Get Fields to select seleccionamos los campos que se requieren para cargar a la base de datos.

- En el step table output se realiza el mapeo de las columnas y además se realiza la conexión a la base de datos “DB_MSP” esquema destino consultaexterna2021 y tabla “aten_pren_resultado”.

Ilustración 30. Mapeo de los campos de la tabla ATEN_PREN_RESULTADO.

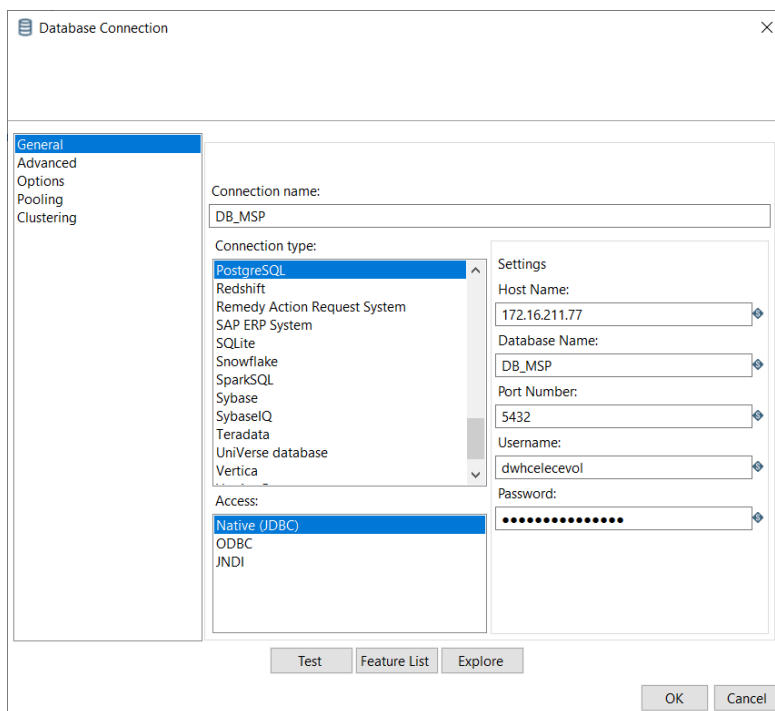


Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Step name: aten_pren_resultado.
 - Connection: Con la opción new realizamos la conexión a la base de datos DB_MSP.
 - Connection: Después de realizar la configuración de la conexión a la base de datos DB_MSP seleccionamos la conexión DB_MSP.
 - Target schema: consultaexterna2021.
 - Target table: aten_pren_resultado.
 - Specify database fields: Seleccionar esta opción.
 - Con Get fields obtenemos los campos a cargar a la base de datos y realizamos el mapeo.
4. En el paso anterior en connection con la opción new realizamos la configuración a la base de datos DB_MSP.

Ilustración 31. Configuración a la base de datos DB_MSP.



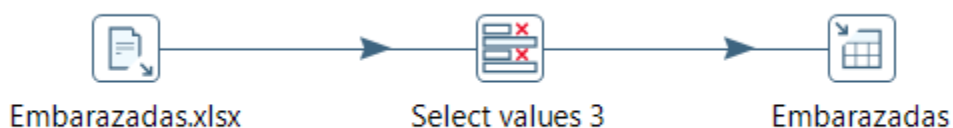
Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- En connection type seleccionamos PostgreSQL y en Access Native (JDBC).
- Connection name: DB_MSP.
- Host Name: Dirección IP del servidor de base de datos.
- Database Name: Nombre de la base de datos (DB_MSP).
- Port Number: 5432.
- Username: Nombre de usuario de acceso a la base de datos.
- Password: Password de acceso a la base de datos.

- Se realiza el proceso ETL de la información de las personas embarazadas por provincia, información proporcionada por el Ministerio de Salud Pública.

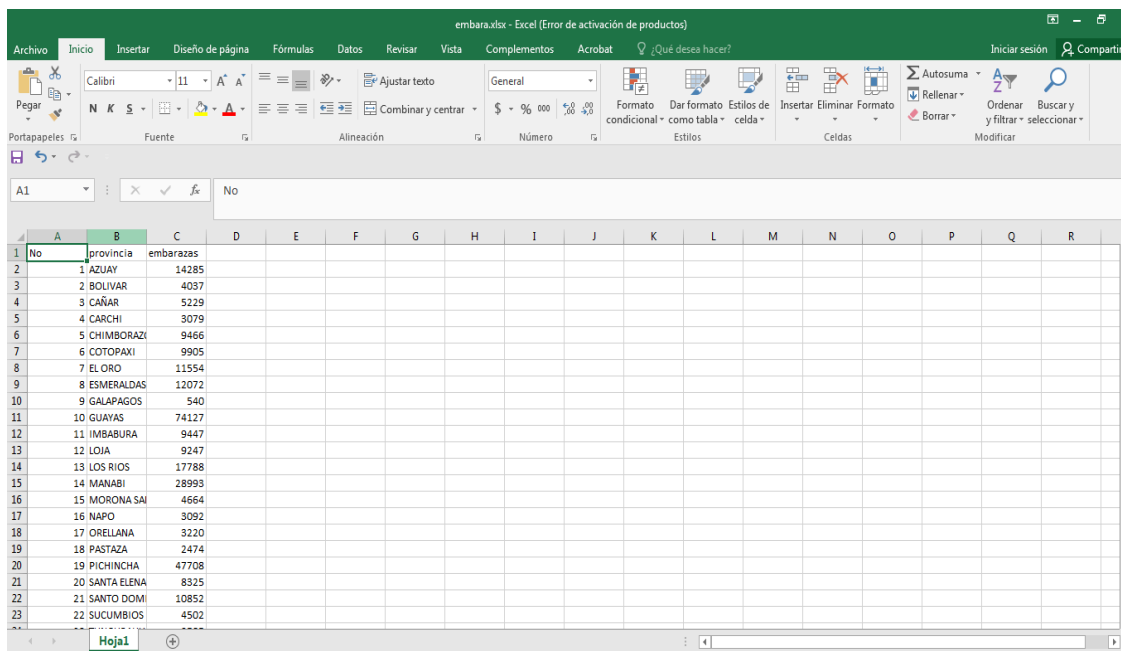
Ilustración 32. Proceso ETL de datos sobre personas embarazadas por provincia.



Fuente: Pentaho Data Integration (PDI).

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Ilustración 33. Información de mujeres embarazadas por provincia.



No	provincia	embarazadas
1	AZUAY	14285
2	BOLIVAR	4037
3	CAÑAR	5229
4	CARCHI	3079
5	CHIMBORAZO	9466
6	COTOPAXI	9905
7	EL ORO	11554
8	ESMERALDAS	12072
9	GALAPAGOS	540
10	GUAYAS	74127
11	IMBABURA	9447
12	LOJA	9247
13	LOS RIOS	17788
14	MANABI	28993
15	MORONA SAI	4664
16	NAPO	3092
17	ORELLANA	3220
18	PASTAZA	2474
19	PICHINCHA	47708
20	SANTA ELENA	8325
21	SANTO DOMI	10852
22	SUCUMBIOS	4502

Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Para el proceso ETL de la información de mujeres embarazadas por provincia es similar a los pasos que se realizaron en el proceso ETL de atención prenatal.

3.5. Especificación de Dashboard

Dentro de los insumos entregados, está la información detallada del indicador de cobertura prenatal, es decir, la fórmula y metodología de cálculo de este indicador. En base a lo proporcionado y con los datos disponible se define las especificaciones que tendrá el dashboard:

- Ser un dashboard comprensible y dinámico.
- Tener imágenes que hagan referencia a mujer embarazada.
- Visualizar el Indicador CCP en un mapa del Ecuador por provincia.
- Visualizar los datos del numerador y denominador del Indicador CCP.
- Poder realizar consultas dinámicas por provincia del Indicador CCP.
- Poder realizar consultas dinámicas por tiempo del Indicador CCP.
- Visualizar el número de atenciones por edad, identidad étnica, o por zona.

3.6. Desarrollo de Dashboard

Para la creación del Dashboard se utiliza la herramienta Microsoft Power BI Desktop Versión: 2.104.941.0 64-bit (abril de 2022) y los datos transformados sobre el Indicador de control prenatal:

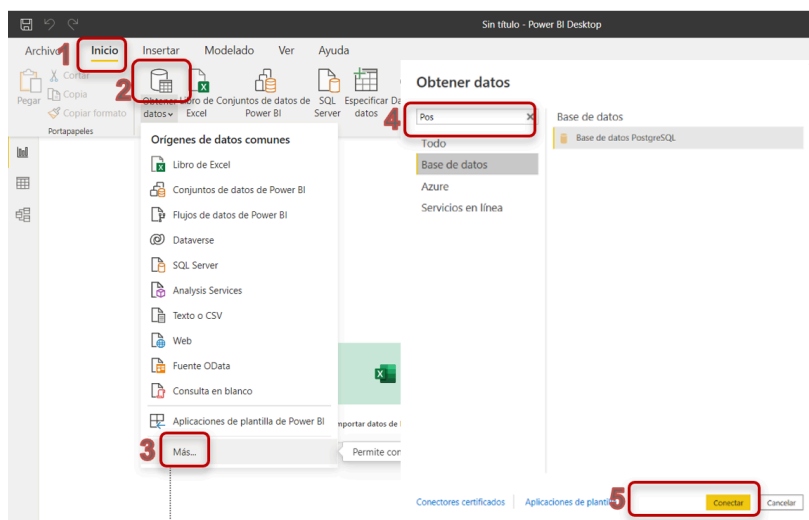
3.6.1. Conectar al modelo de datos

El modelo dimensional está creado en la base de datos PostgreSQL (DB_MSP)

1. Seleccionar el menú de “Inicio” del PowerBI.
2. Escoger la opción “Obtener datos”.
3. Escoger la opción “Más”.

4. En la opción de búsqueda se ingresa la palabra “PostgreSQL”.
5. Elegir Base de Datos PostgreSQL, Conectar.

Ilustración 34. Pasos para Conectar al Modelo de Datos



Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

6. Ingresar datos de conexión a la Base de Datos.
7. Escoger el Modo de Conectividad “Importar” (opción elegida para esta guía):
 - o **Importar:** Permite importar y consultar los datos desde el origen de datos (Postgres) y grabarlo en el archivo PowerBI (.pbix) el conjunto de datos.
 - o **DirectQuery:** Permite consultar los datos desde el origen de datos, para lo cual requiere necesariamente tener conexión a la base de datos.
8. Dar click en “Aceptar”.

Ilustración 35. Ingreso de Datos para la Conexión a Base de Datos

The image shows a dialog box titled "Base de datos PostgreSQL" with a close button (X) in the top right corner. It contains the following fields and options:

- Servidor:** A text input field containing "172.16.211.77".
- Base de datos:** A text input field containing "DB_MSP".
- Modo Conectividad de datos:** A section with a help icon (i) containing two radio button options: "Importar" (selected) and "DirectQuery".
- Opciones avanzadas:** A link with a right-pointing chevron (>).
- Buttons:** A yellow "Aceptar" button and a white "Cancelar" button.

Red boxes and numbers highlight specific elements: a red box labeled "6" encloses the "Servidor" and "Base de datos" fields; a red box labeled "7" encloses the "Modo Conectividad de datos" section; and a red box labeled "8" encloses the "Aceptar" and "Cancelar" buttons.

Fuente: Datos de MSP – Indicador CCP – 2021

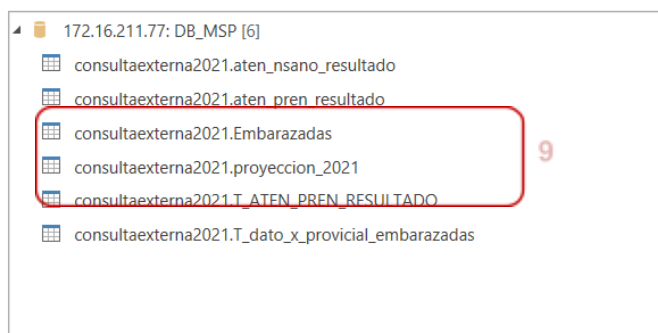
Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

9. Se elige las siguientes tablas:
- o **Aten_pren_resultado:** datos transformados que indican las atenciones por primera consulta preventiva de control prenatal a mujeres embarazadas (Código CIE_10) del año 2021.
 - o **Embarazadas:** tabla de datos entregados por el Ministerio de Salud que indican los estimados de población de mujeres embarazadas por provincia del año 2021.

Ilustración 36. Tablas de la Base de Datos DB_MSP

Navegación

Especifique un recurso al que se debe navegar.



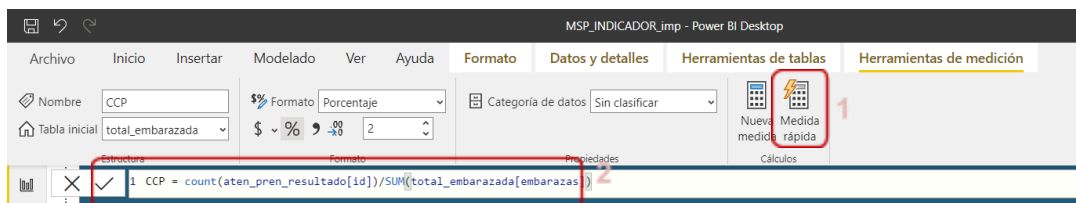
Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.6.2. Crear variables

1. Se crea la medida CCP (Cobertura de Control Prenatal).
2. Colocar la fórmula de cálculo, es decir, sumatoria de todas las atenciones por primera consulta preventiva de control prenatal a mujeres embarazadas sobre la estimación de la población de mujeres embarazadas.

Ilustración 37. Pasos para la creación de Variable



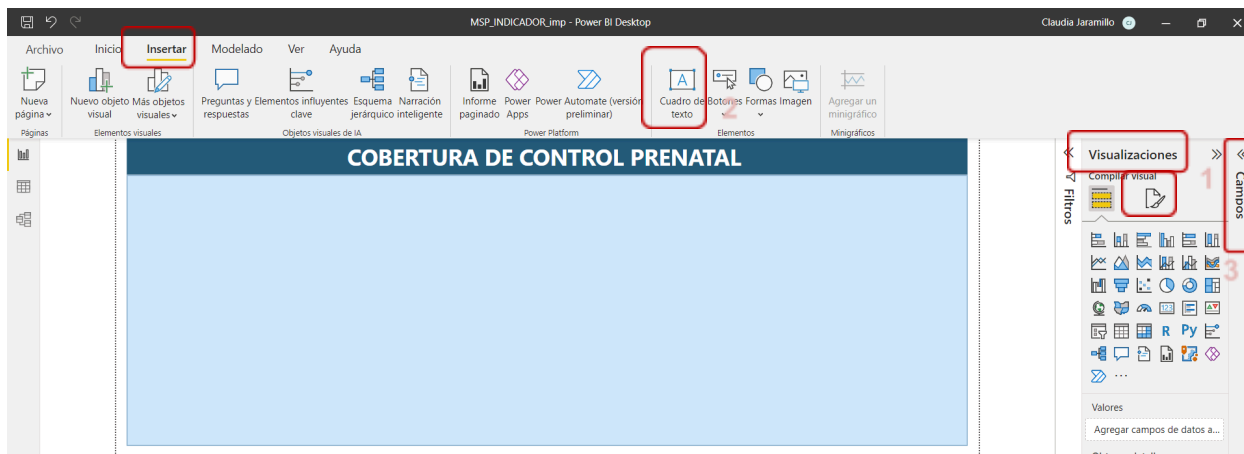
Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

3.6.3. Crear las visualizaciones (gráficos e indicadores)

1. Elegir los colores de fondo, en el panel de visualizaciones “dar formato de fondo”.
2. Ingresar el título del Indicador en el menú Insertar, opción “Cuadro de texto”.
3. **Insertar objetos al dashboard:**
 - o Para elegir los objetos y sus variables, elegir el panel de “visualizaciones”.
 - o Para elegir los campos de las tablas importadas en el punto 4.6.1 literal 9, elegir el panel “Campos.

Ilustración 38. Pasos para crear visualizaciones en PowerBI

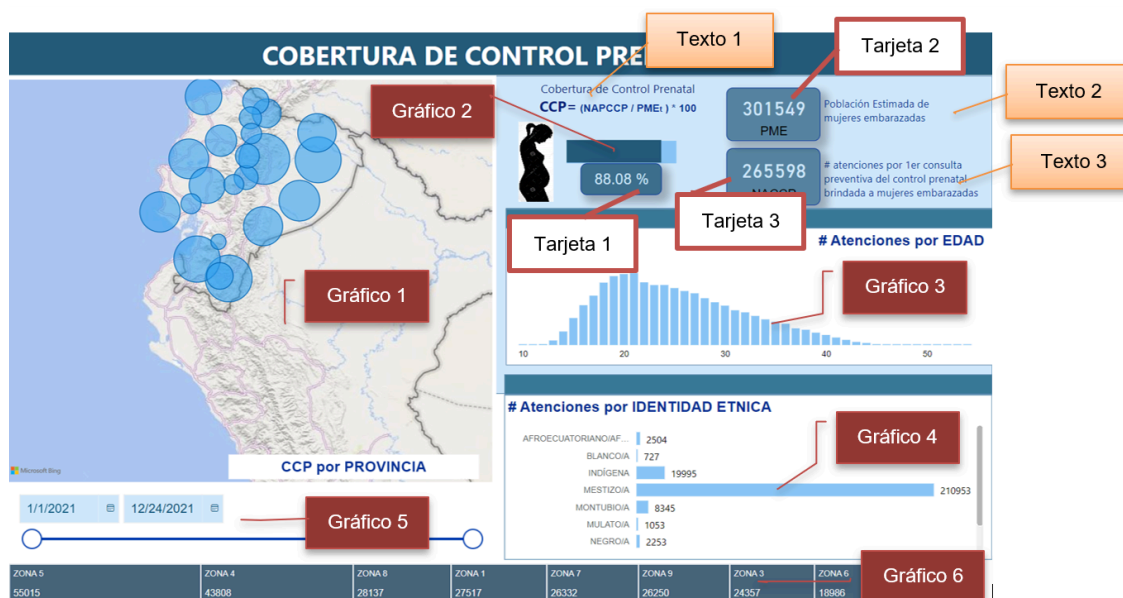


Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

En la siguiente ilustración se muestran todas las visualizaciones creadas en el dashboard del MSP.

Ilustración 39. Visualizaciones en el Dashboard



Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

- Gráfico 1:** Seleccionar el objeto “Mapa” y en las variables que lo componen ingresar:
 - “En Ubicación”: Colocar el campo “Provincia” de la tabla “aten_pren_resultado”.
 - “En Tamaño de Burbuja”: Colocar la medida CCP creada en el punto 4.6.2.
- Texto 1:** Seleccionar el objeto “Cuadro de texto” e ingresar el nombre del Indicador y la fórmula.
- Gráfico 2:** Seleccionar el objeto “Gráfico de barras agrupados”, y en las variables que lo componen ingresar:
 - En Eje X: Colocar la medida CCP creada en el punto 4.6.2.

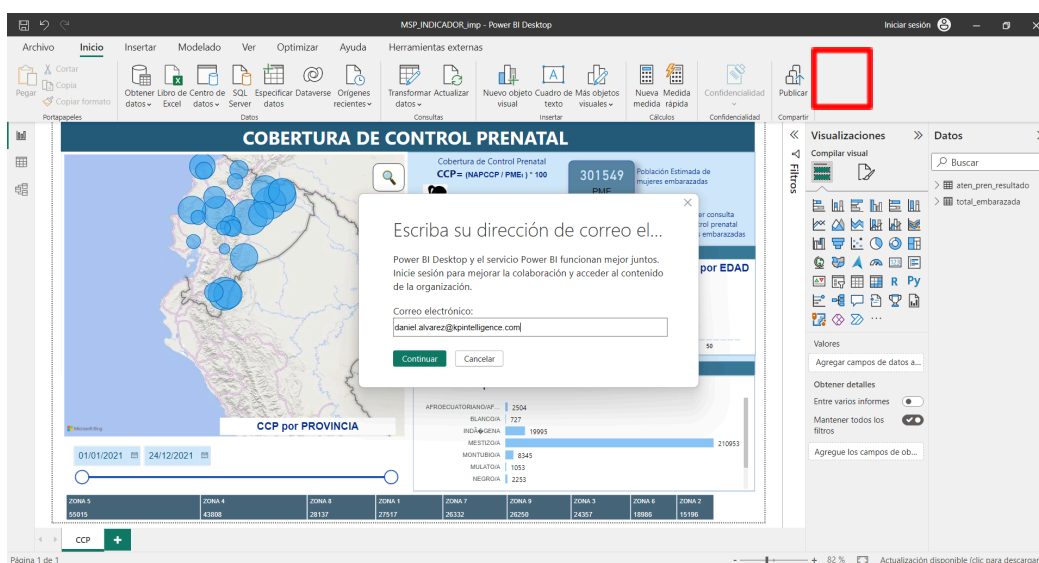
- d) **Tarjeta 1:** Seleccionar el objeto “Tarjeta”, y en las variables que lo componen ingresar:
- En Campos: Colocar la medida CCP creada en el punto 4.6.2.
- e) **Texto 2:** Seleccionar el objeto “Cuadro de texto” e Ingresar la descripción del denominador del Indicador CCP, es decir, (PME) Población Estimada de mujeres embarazadas
- f) **Tarjeta 2:** Seleccionar el objeto “Tarjeta”, y en las variables que lo componen ingresar:
- En Campos: Colocar el campo “embarazadas” de la tabla “total_embarazda”.
- g) **Texto 3:** Seleccionar el objeto “Cuadro de texto” e ingresar la descripción del numerador del Indicador CCP, es decir, (NACCP) número de atenciones por primera consulta preventiva del control prenatal brindada a mujeres embarazadas.
- h) **Tarjeta 3:** Seleccionar el objeto “Tarjeta”, y en las variables que lo componen ingresar:
- En Campos: Contabilizar el número de registros que existe en la tabla “aten_pren_resultado” (Recuento del “campo ID”).
- i) **Gráfico 3:** Seleccionar el objeto “Gráfico de columnas agrupados”, y en las variables que lo componen ingresar:
- En Eje X: Colocar el campo “edad” de la tabla “aten_pren_resultado”.
 - En Eje Y: Contabilizar el número de registros que existe en la tabla “aten_pren_resultado” (Recuento del “campo ID”).
- j) **Gráfico 4:** Seleccionar el objeto “Gráfico de barras agrupados”, y en las variables que lo componen:

- En Eje Y: Colocar el campo “identificación_etnica” de la tabla “aten_pren_resultado”.
 - En Eje X: Contabilizar el número de registros que existe en la tabla “aten_pren_resultado” (Recuento del “campo ID”).
- k) **Gráfico 5:** Seleccionar el objeto “Segmentación de dato”, y en las variables que lo componen ingresar:
- En Eje Y: Colocar el campo “fecha_atencion” de la tabla “aten_pren_resultado”.
- l) **Gráfico 6:** Seleccionar el objeto “Segmentación de dato”, y en las variables que lo componen ingresar:
- En Categoría: Colocar el campo “zona” de la tabla “aten_pren_resultado”.
 - En Valores: Contabilizar el número de registros que existe en la tabla “aten_pren_resultado” (Recuento del “campo ID”).

3.7. Publicar el Dashboard

Para que el dashboard pueda ser accedido por un usuario a través del Internet se procede desde la herramienta Power BI Desktop a escoger la opción de publicar desde la pestaña Inicio. Una vez escogida la opción de Publicar la herramienta solicitará que se inicie sesión con una cuenta que haya sido activada dentro del servicio de Power BI en la nube.

Ilustración 40. Power BI Desktop – Publicar Reporte



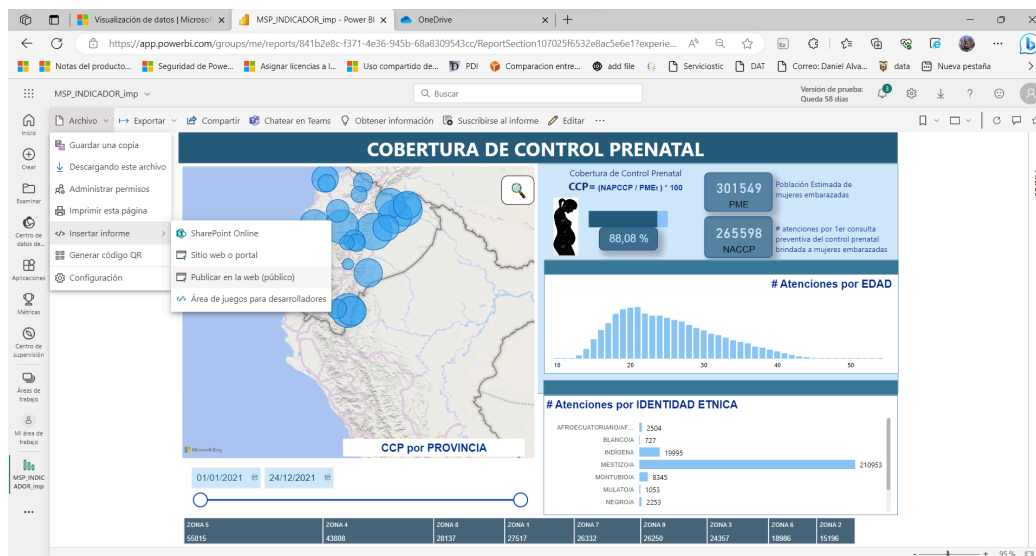
Fuente: Microsoft Power BI Desktop

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

De forma predeterminada podemos publicar el reporte en nuestra Área de Trabajo, o podemos publicarlo en otra área de trabajo que haya sido previamente creada.

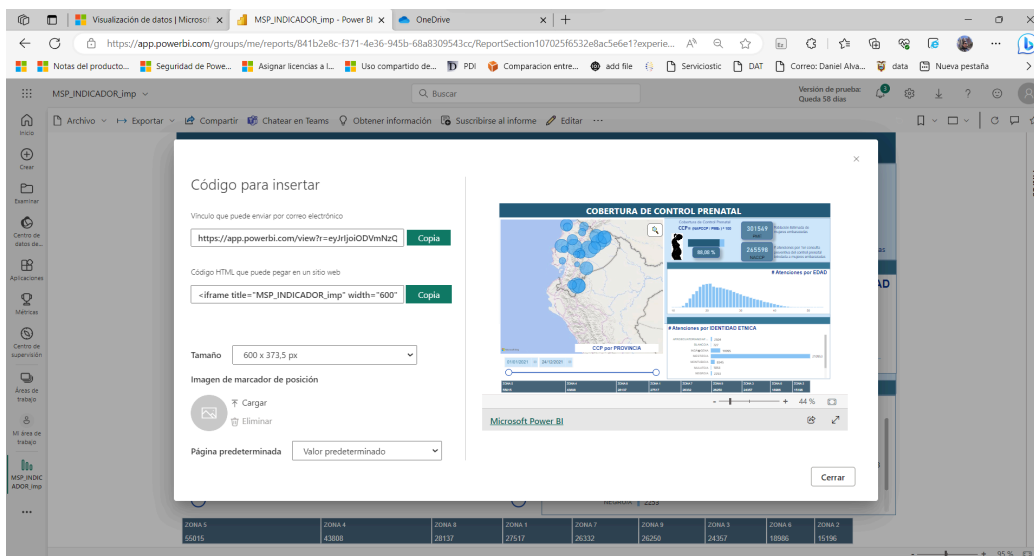
Finalmente, desde la página del servicio de Power BI en la nube podemos generar el link público que puede ser distribuido a los usuarios que vayan a ingresar y visualizar el dashboard desarrollado. Para esto se debe ingresar al servicio de Power BI con la misma cuenta con la que se publicó previamente el reporte, y una vez dentro del reporte dentro de la opción Archivo/Insertar informe se debe elegir “Publicar en la web (público)”.

Ilustración 41. Power BI Desktop - Obtener Link público del Reporte



Después de elegir la opción indicada se desplegará una pantalla que nos permitirá copiar al portapapeles el link para poder compartirlo directamente a los usuarios que van a ingresar a visualizar los datos pero también nos da la opción de obtener el código HTML que podemos insertar en una página web para que el reporte se inserte como un objeto web.

Ilustración 42. Power BI Desktop – Obtener Link público del Reporte



Fuente: Microsoft Power BI Desktop

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Link del dashboard desarrollado: [DASHBOARD INDICADOR CCP - MSP](#)

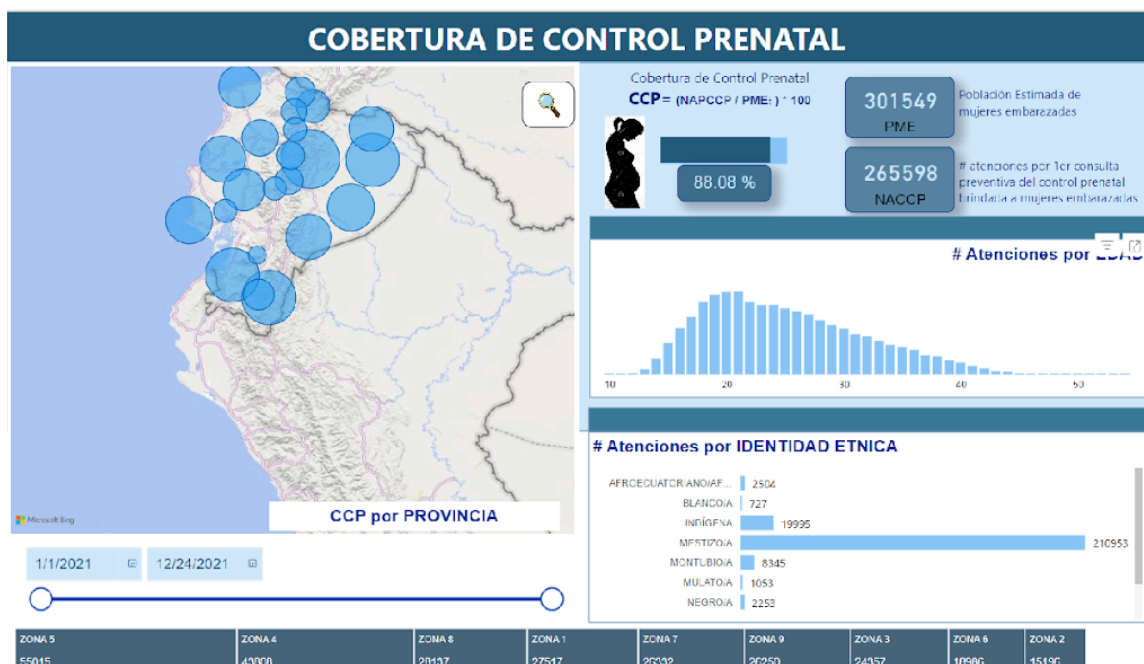
5. RESULTADOS

Se adjunta el resultado que se obtiene al aplicar los pasos/procesos de la sección 4.

- **Mapa:** Indica el porcentaje Cobertura de Control prenatal por provincia, para la cual, es decir, el tamaño de la burbuja representa el porcentaje de control prenatal.
- **Barra % CCP:** Indica el porcentaje Cobertura de Control prenatal total, o de acuerdo con el filtro aplicado (tiempo o provincia).

- **Gráfico # Atenciones por edad:** Indica el número de atenciones por primera consulta preventiva del control prenatal brindada a mujeres embarazadas por edades.
- **Gráfico # Atenciones por etnia:** Indica el número de atenciones por primera consulta preventiva del control prenatal brindada a mujeres embarazadas por identidad étnica.
- **Gráfico # Atenciones por Zona:** Indica el número de atenciones por primera consulta preventiva del control prenatal brindada a mujeres embarazadas por Zona.

Ilustración 43. Dashboard CCP-MSP



Fuente: Datos de MSP – Indicador CCP – 2021

Creado por: Daniel Alvarez, Jose Cuasapaz, Wilmer Guamaní, Claudia Jaramillo

Interpretación:

- **Mapa:** Se puede observar que la provincia de Napo, Galápagos, Zamora Chinchipe, el Oro, Orellana tiene el porcentaje de control prenatal más alto.
- **Gráfico # Atenciones por edad:** Se puede observar que las edades entre 17 - 23 años indican un mayor número de atenciones.
- **Gráfico # Atenciones por etnia:** Se puede observar que la etnia mestiza indica un mayor número de atenciones.
- **Gráfico # Atenciones por Zona:** Se puede observar que las Zonas 5 y 4 indican un mayor número de atenciones.

6. RECOMENDACIONES TÉCNICAS

- Como mejora técnica al proceso presentado se recomienda el uso de librerías que nos permita realizar la conexión desde Python a la BDD PostgreSQL, una opción sugerida en la librería Psycopg2, esto con el fin de poder realizar el registro del dataset que se requiera para cada cálculo de forma automática.
- Debido a que el objetivo final que se ha planteado corresponde a una herramienta de visualización (Power BI) orientada al uso de la ciudadanía, es recomendable que la experiencia del usuario satisfaga en la mayor medida posible las expectativas que estos tengan y logren una comprensión rápida de los datos desplegados. Para esto se recomienda contar con un modelo de datos optimizado y eficiente para consultas, así como no descuidar el aspecto visual en donde los objetos visuales como histogramas o gráficos de pastel, por ejemplo, predominen sobre tablas que contengan sólo números.

- Siempre será recomendable automatizar todo el proceso de consolidación, limpieza, procesamiento y despliegue de información; razón por la cual se recomienda también incursionar en el desarrollo de reportes de Power BI que estén conectados de forma directa al él o los repositorios corporativos. Para lograr este cometido es de gran ayuda la implementación del método de publicación Direct Query de Power BI junto con el uso del software de Microsoft conocido como On Premise Data Gateway de Microsoft.
- Finalmente, y a pesar de que existen herramientas que pueden realizar todo el proceso de consolidación, limpieza, procesamiento y despliegue de información como el mismo Microsoft POWER BI; recomendamos adoptar la buena práctica para que, en lo posible, se pueda utilizar en cada uno de los procesos mencionados herramientas especializadas y adaptadas a las necesidades específicas de cada organización.

7. ANEXOS

[Anexo 1. Base de datos](#)

[Anexo 2. Pronóstico por provincia](#)

[Anexo 3. PowerBI](#)

8. BIBLIOGRAFÍA

- AWS. (2023). ¿Qué es Python? - Explicación del lenguaje Python - AWS. Amazon Web Services, Inc.
<https://aws.amazon.com/es/what-is/python/#:~:text=Python%20es%20un%20lenguaje%20de,ejecutar%20en%20muchas%20plataformas%20diferentes.>
- Comunidad Empresas. (2023). Pentaho Data Integration: ¿Qué es y para qué sirve? Entel Comunidad Empresas.

<https://ce.entel.cl/articulos/pentaho-data-integration/#::-:text=Pentaho%20Data%20Integration%20es%20un,en%20un%20formato%20más%20limpio>.

- Elsamadisi, A. (2021, 10 de mayo). Using PostgreSQL as a Data Warehouse. Narrator Data Blog | Narrator.ai.
<https://www.narratordata.com/blog/using-postgresql-as-a-data-warehouse/>
- Kimball Group. (s.f.). Kimball DW/BI Lifecycle Methodology. Kimball Group.
<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/>
- Martín, G. (2018, 18 de enero). Primeros pasos con Jupyter Notebook - Adictos al trabajo. Adictos al trabajo.
<https://www.adictosaltrabajo.com/2018/01/18/primeros-pasos-con-jupyter-notebook/>
- Microsoft. (2023). ¿Qué es Power BI? Definición y características | Microsoft Power BI. Data Visualization | Microsoft Power BI.
<https://powerbi.microsoft.com/es-es/what-is-power-bi/>
- Jiménez, E. G., & Jinesta, R. A. (2023). Metodología de la programación: conceptos, lógica e implementación. Marcombo. ISBN: 978-958-778-842-6
- Plan Estratégico Institucional del MSP. (2023, 23 de mayo). Ministerio de Salud Pública.
<https://www.salud.gob.ec/wp-content/uploads/2023/06/PEI-MSP-MAYO-2023-USCRITO.pdf>