

# MANUAL TÉCNICO

---

Grupo#1

Guatemala, 8 de noviembre de 2020

## Manual Técnico

Este proyecto consta de cinco fases para su respectiva implementación.

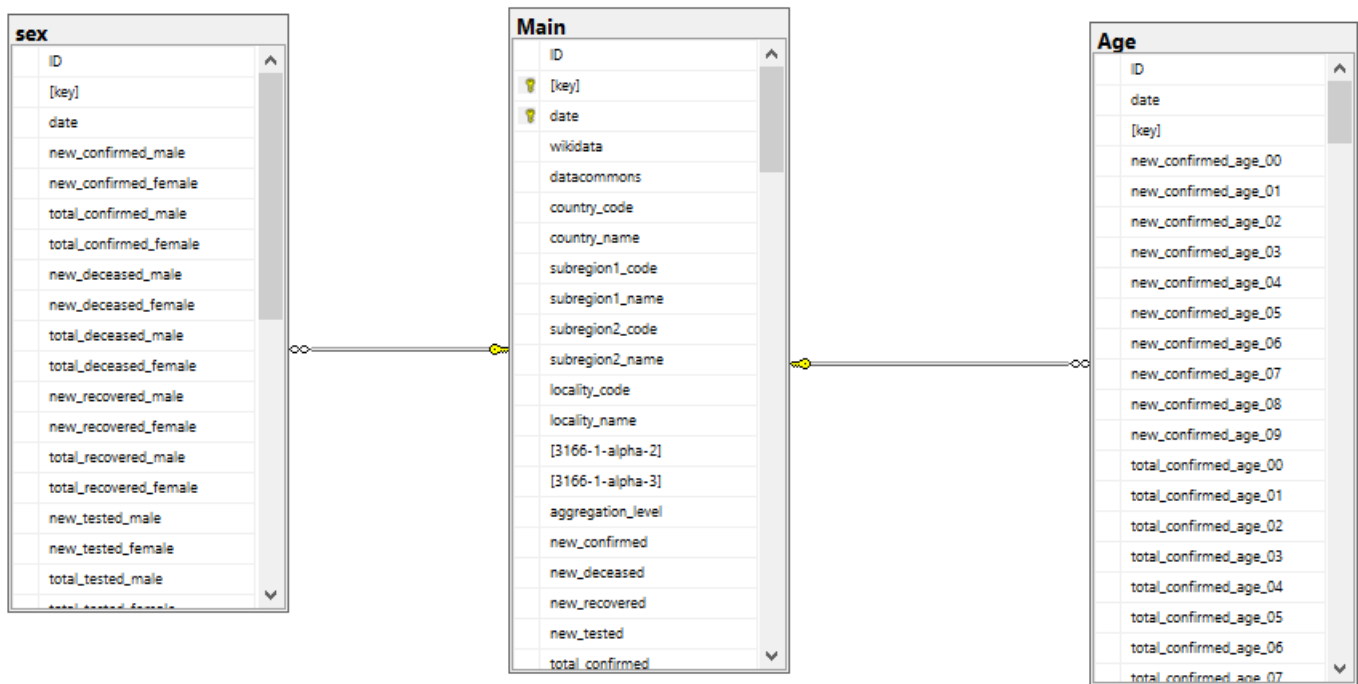
### Fase 1

Creación de base de datos relacional llamada COVID\_GRUPO1. Esta base de datos cuenta con tres tablas, teniendo como tabla principal la tabla “Main” que tiene una llave primaria compuesta conformada por [date] y [key]. Esta tabla es el dataset principal de la base de datos que contiene todos los datos con su respectivo detalle de los casos más recientes del COVID19.

Luego está la tabla “sex” y la tabla “Age” que tienen una llave foránea compuesta conformada por [date] y [key].

La tabla sex y Age son dos visualizaciones diferentes de ver el Main. El sex es la data dispersa por genero mientras que la Age es la data dispersa por la edad.

Diagrama Relacional de la base de datos COVID\_GRUPO1.



## Fase 2

Creación de la base de datos COVID\_GRUPO1\_DWH. Esta base de datos multidimensional fue creada con el modelo Estrella donde todas las dimensiones se van conectadas a una o varias tablas de hechos.

Esta base de datos está compuesta por tablas de Dimensiones.

Hay un total de cinco dimensiones: Demographics, Economy, Geography, Health y Fecha.

La dimensión Demographics contiene información en el ámbito demográfico del dataset de COVID.

La dimensión Economy contiene información en el ámbito económico del dataset de COVID.

La dimensión Geography contiene información en el ámbito geográfico del dataset de COVID.

La dimensión de Health contiene información en el ámbito de la salud del dataset de COVID.

La dimensión de Fecha fue una dimensión creada por nosotros para que el cubo multidimensional tenga una referencia respecto al tiempo.

Además consta de tablas de hechos (Fact) que representa las actividades de análisis.

Hay un total de tres tablas de hechos: Epidemiology, Sex y Age.

La tabla de hechos de Epidemiology es una unión de la tabla Epidemiology con la tabla Hospitalizations, tiene como llave primaria una llave compuesta conformada de [key] y [date].

La tabla de hechos de Sex, es una tabla de hechos que viene del dataset Age de la base de datos relacional, igual es el caso con la tabla Age.

Las tres tablas Fact están relacionadas con las cinco dimensiones previamente explicadas por medio de la llave [key].

Esta base de datos también consta de tablas Staging para cada tabla de hechos, el cual nos ayuda a insertar al Fact ya que se almacenan los nuevos registros a insertar en la tabla de hechos.

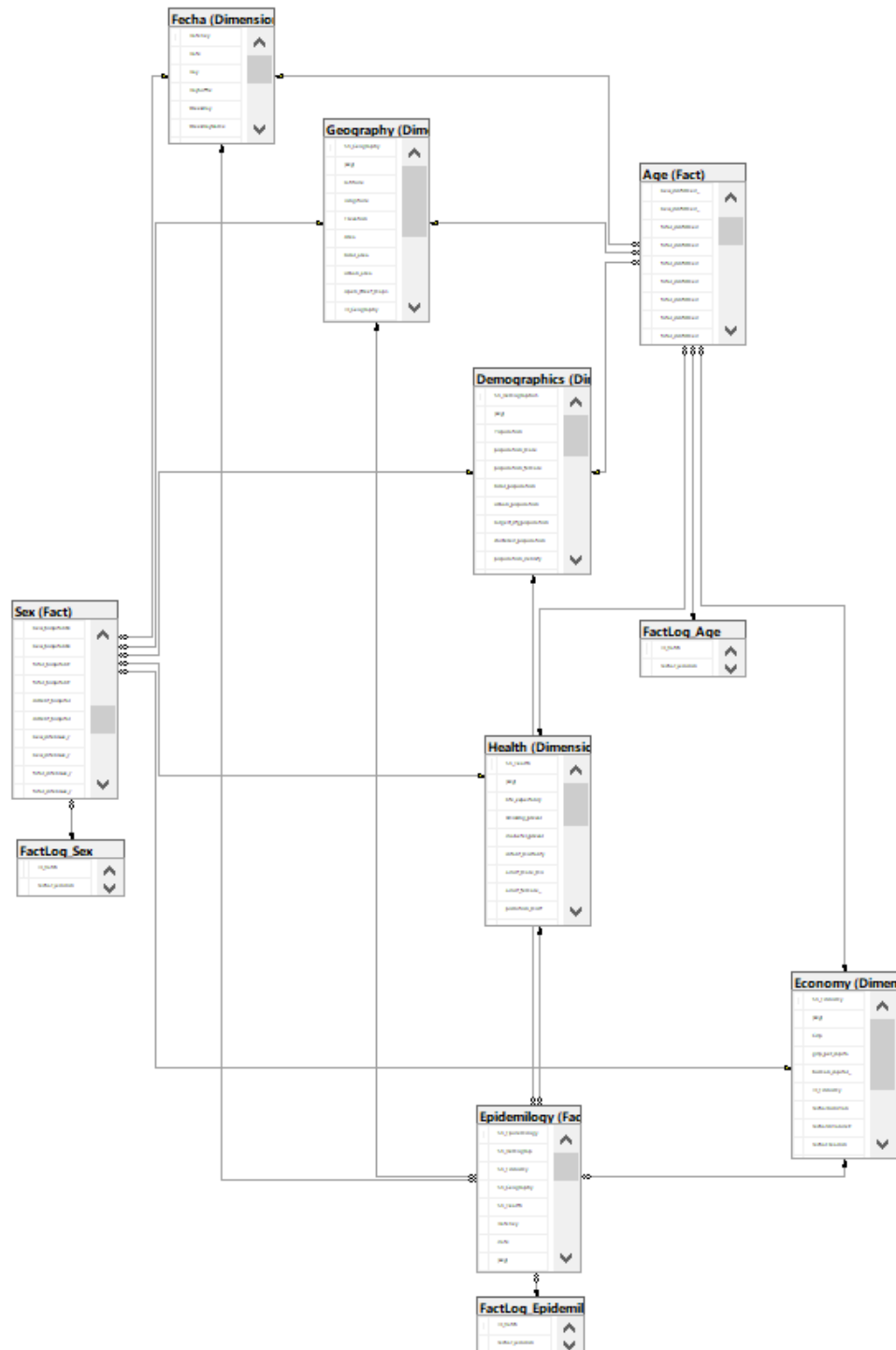
Sex (staging)	
date	↑
[key]	
new_confirmed_male	
new_confirmed_female	
total_confirmed_male	
total_confirmed_female	
new_deceased_male	
new_deceased_female	
total_deceased_male	
total_deceased_female	
new_recovered_male	
new_recovered_female	
total_recovered_male	
total_recovered_female	
new_tested_male	
new_tested_female	
total_tested_male	
total_tested_female	
new_hospitalized_male	
new_hospitalized_female	↓

Epidemiology (staging)	
date	
[key]	
new_confirmed	
new_deceased	
new_recovered	
new_tested	
total_confirmed	
total_deceased	
total_tested	
new_hospitalized	
new_intensive_care	
new_ventilator	
total_hospitalized	
total_intensive_care	
total_ventilator	
current_hospitalized	
current_intensive_care	
current_ventilator	

Age (staging)	
date	↑
[key]	
new_confirmed_age_00	
new_confirmed_age_01	
new_confirmed_age_02	
new_confirmed_age_03	
new_confirmed_age_04	
new_confirmed_age_05	
new_confirmed_age_06	
new_confirmed_age_07	
new_confirmed_age_08	
new_confirmed_age_09	
total_confirmed_age_00	
total_confirmed_age_01	
total_confirmed_age_02	
total_confirmed_age_03	
total_confirmed_age_04	
total_confirmed_age_05	
total_confirmed_age_06	
total_confirmed_age_07	↓

La base de datos cuenta también con tablas FactLog que nos permiten guardar el registro ya sea la fecha, ID de registros, cuando se hace una ejecución, entonces es cuando se hacen nuevos registros en la tabla de hechos se hace uso de esta tabla.

Diagrama Relacional de la base de datos COVID\_GRUPO1\_DWH:



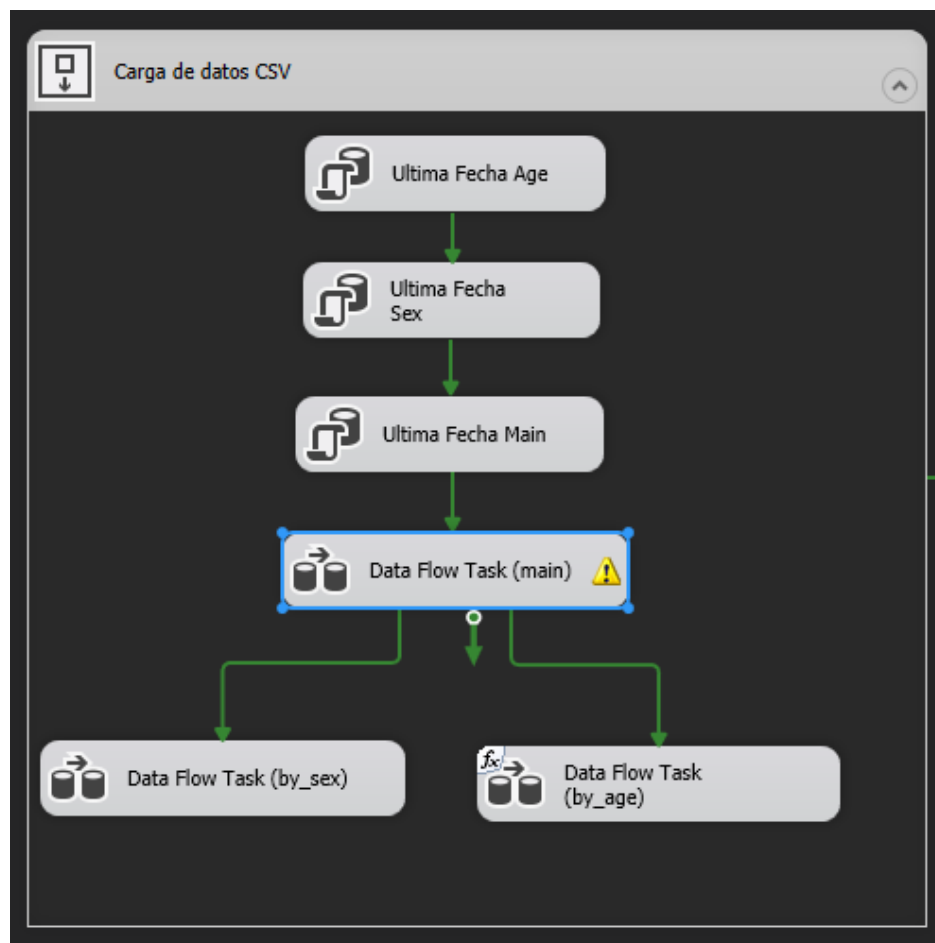
En la base de datos también cuenta con stored procedures que ayudan a realizar funciones específicas e importantes para el Data Warehouse.

El primero SP solamente llena la dimensión Fecha, luego hay tres SP, uno para cada tabla de hechos el cual se hace un MERGE de la tabla de Staging que contiene los nuevos registros a insertar a la tabla de hechos.

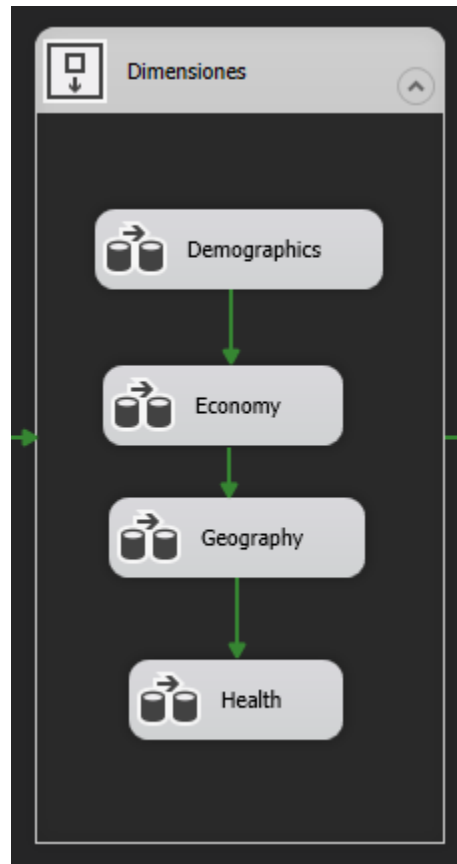
### Fase 3:

Esta fase se trata en la creación del paquete de Integration Services (SSIS).

En esta fase la cual fue la más compleja y es donde se ve el ciclo de vida de nuestra base de datos multidimensional se compone de varias sub fases de los cuales se agruparon por Sequence Containers, el primer Sequence Container es la carga de datos CSV, este contenedor esta compuesto por varios Execute SQL Task el cual primero lo que hace es que trae la última fecha de ejecución de las tres tablas de FactLog ya que esto sirve para solo insertar datos que no han sido insertados a nuestras tablas de la base de datos COVID\_GRUPO1, ya que en cada Data Flow Task tiene un CSV Source que inserta data de un CSV de internet, luego con una División Condisional se hace la condición que solo inserte cuando sea mayor a la ultima fecha traída en el SQL Task anteriormente, luego Ordenar lo que hace empieza a validar si la [key] y [date] de ese registro esta en la tabla destino, no lo inserta de lo contrario lo inserta y toda la data que pasa por todos esos filtros se insertan a un OLE DB Destino que en este caso serian nuestras tablas de Main, Sex y Age.

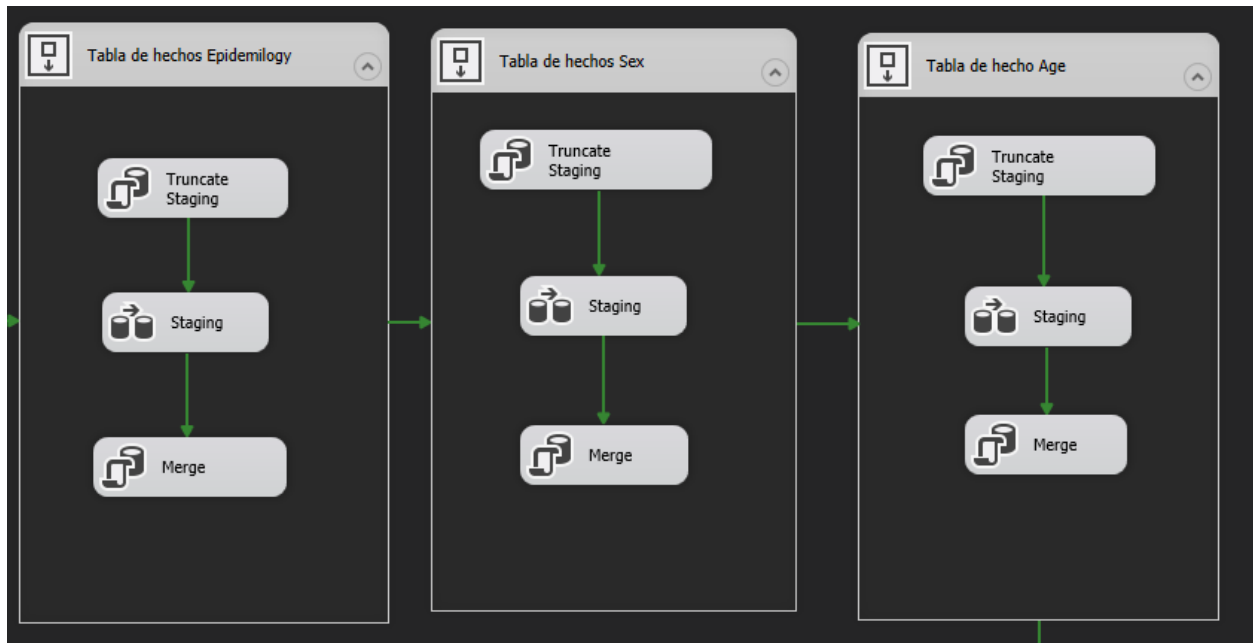


La otra sub fase es el Sequence Container de las dimensiones, cada dimensión es trabajada con un Data Flow Task, el cual tiene un OLE DB Origen que tiene un SQL Statement que contiene un Select para cada dimensión que también tiene la condición que solo inserte cuando la fecha actual sea mayor a la última fecha de ejecución que tiene registrada el Factlog , luego hay un convertidor de datos el cual convierte toda la data a la data requerida y manejada por nuestro base de datos DWH, ya con estas “nuevas columnas” se manejan como copias que se utilizaran en una Dimensión de Variación Lenta, ahí le hacemos saber que columnas de dicha Dimensión será de SCD de tipo 1 o de tipo 2 y finalmente tenemos un Insert Destination donde ya se inserta a cada respectiva Dimensión.



Las otras sub-fases son Sequence Container de Tabla de Hechos para cada tabla de hechos, en ese Sequence Container lo que hace primero es que se ejecuta un SQL Task que hace un Truncate a la tabla Staging respectiva de su tabla de hechos luego en un Data Flow Task que lo que hace es llenar la tabla Staging y finalmente hacer otro SQL Task que lo que hace es ejecutar el SP de Merge de esa respectiva tabla de hechos.

Ese proceso se realiza uno por cada tabla de hechos.



Finalmente esta la sub-fase donde se encuentran un Sequence Container el cual su función son servicios analíticos de SSAS, el primero son las dimensiones el cual son de tipo Process Update, luego son de cada tabla de hechos que es de tipo Process Add, y se ingresa el Select del SP\_Merge.



#### Fase 4:

Esta fase se trata de la creación del proyecto de SSAS (la creación del cubo multidimensional), en el cual primero se crea un Data Source el cual es la conexión de la base de datos donde esta la base de datos multidimensional (DWH), luego se crea un Data Source View el cual vamos a agregar las tres tablas Fact y las dimensiones. Ya con un Data Source View podemos crear nuestras dimensiones y Cubos, en este proyecto se crearon tres cubos que utilizan las mismas dimensiones. En cada dimensión trajimos solo las columnas que son relevantes para dar un punto de vista a nuestras tablas de hechos, en unas dimensiones creamos jerarquías que creemos que son relevantes. Para las tablas de hechos hicimos varios Calculations y KPI's, cada tabla de hechos tiene sus diferentes KPI'S y Calculations.

Calculations y KPI'S Cubo Epidemiology:

Script Organizer	
	Command
1	CALCULATE
2	[Cantidad Total Casos]
3	[AVG Casos]
4	[Cantidad Total Hospitalizaciones]
5	[AVG Hospitalizaciones]
6	[Cantidad Total Recuperados]
7	[AVG Recuperados]
8	[Cantidad Total Fallecidos]
9	[AVG Fallecidos]
10	[Cantidad Total Pruebas Confirmadas]
11	[AVG Pruebas Confirmadas]
12	[Cantidad Total Cuidados Intensivos]
13	[AVG Cuidados Intensivos]
14	[Cantidad Total Casos Ventilacion]
15	[AVG Casos Ventilador]

KPI Organizer	
	KPI Infectados
	KPI Porcentaje Fallecida
	KPI Recuperados





## Calculations y KPI'S Cubo Sex:

Script Organizer		
		Command
1		CALCULATE
2		[Cantidad Total Casos Hombres]
3		[Cantidad Total Casos Mujeres]
4		[Cantidad Total Fallecidos Hombres]
5		[Cantidad Total Fallecidos Mujeres]
6		[Cantidad Total Recuperados Hombres]
7		[Cantidad Total Recuperados Mujeres]
8		[Cantidad Total Pruebas Confirmadas Hombres]
9		[Cantidad Total Pruebas Confirmadas Mujeres]
10		[Cantidad Total Hospitalizados Hombres]
11		[Cantidad Total Hospitalizados Mujeres]
12		[Cantidad Total Cuidados Intensivos Hombres]
13		[Cantidad Total Cuidados Intensivos Mujeres]
14		[Cantidad Total Ventilados Hombres]
15		[Cantidad Total Ventilados Mujeres]
16		[AVG Casos Hombres]
17		[AVG Casos Mujeres]
18		[AVG Fallecidos Hombres]
19		[AVG Fallecidos Mujeres]
20		[AVG Recuperados Hombres]
21		[AVG Recuperados Mujeres]
22		[AVG Pruebas Confirmadas Hombres]
23		[AVG Pruebas Confirmadas Mujeres]
24		[AVG Hospitalizados Hombres]
25		[AVG Hospitalizados Mujeres]
26		[AVG Cuidados Intensivos Hombres]
27		[AVG Cuidados Intensivos Mujeres]
28		[AVG Ventilados Hombres]
29		[AVG Ventilados Mujeres]

KPI Organizer	
	KPI Porcentaje Fallecida Hombres
	KPI Recuperados Hombres
	KPI Infectados Hombres
	KPI Infectados Mujeres
	KPI Porcentaje Fallecida Mujeres
	KPI Recuperados Mujeres

Calculations y KPI'S Cubo Age:

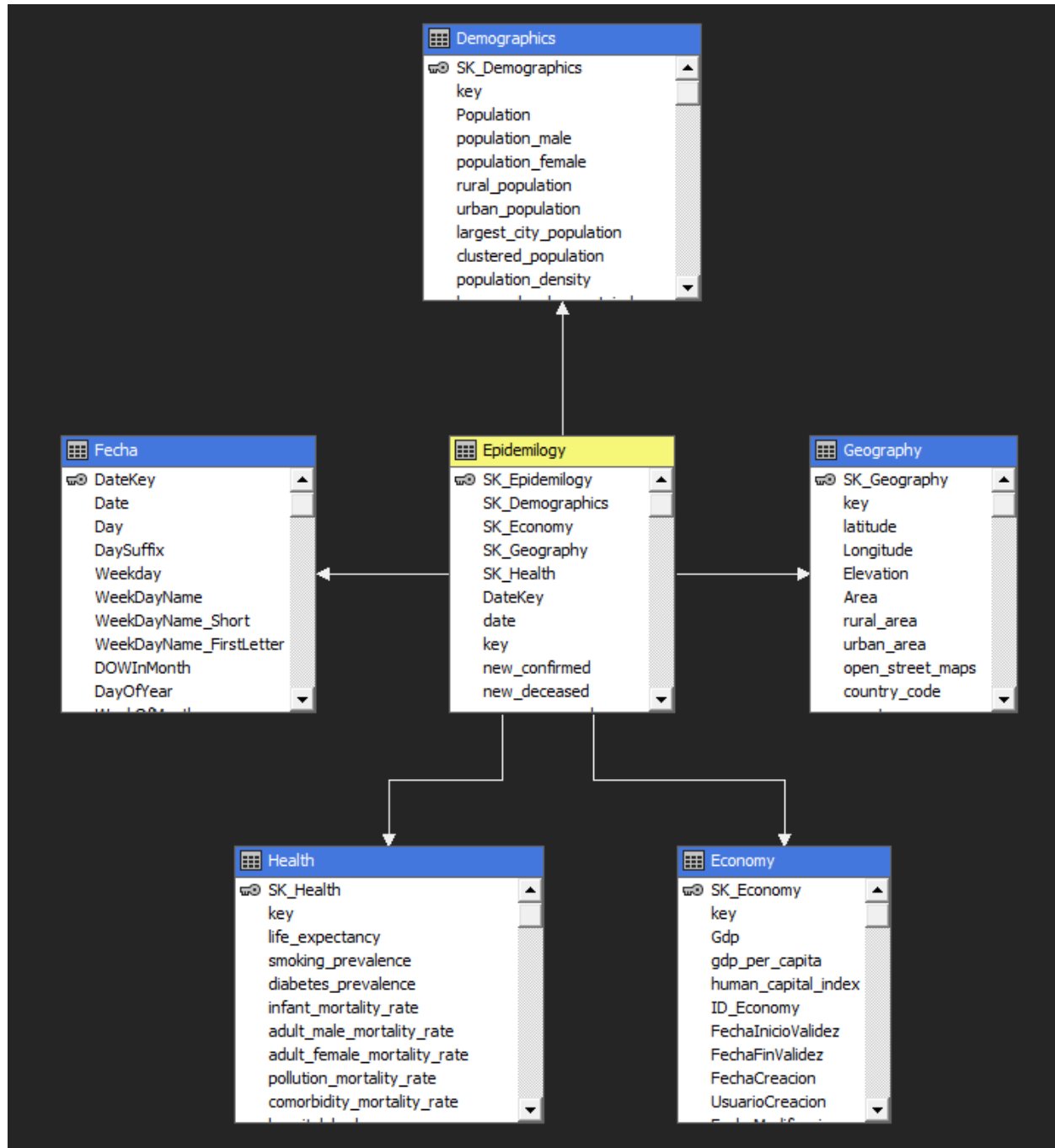
Script Organizer		
		Command
1		CALCULATE
2		[Cantidad Total Casos Niños]
3		[Cantidad Total Casos Adultos]
4		[Cantidad Total Casos Ancianos]
5		[AVG Casos Niños]
6		[AVG Casos Adultos]
7		[AVG Casos Ancianos]
8		[Cantidad Total Recuperados Niños]
9		[Cantidad Total Recuperados Adultos]
10		[Cantidad Total Recuperados Ancianos]
11		[AVG Recuperados Niños]
12		[AVG Recuperados Adultos]
13		[AVG Recuperados Ancianos]
14		[Cantidad Total Fallecidos Niños]
15		[Cantidad Total Fallecidos Adultos]
16		[Cantidad Total Fallecidos Ancianos]
17		[AVG Fallecidos Niños]
18		[AVG Fallecidos Adultos]
19		[AVG Fallecidos Ancianos]

KPI Organizer	
	KPI Porcentaje Infectados Niños
	KPI Porcentaje Infectados Adultos
	KPI Porcentaje Infectados Ancianos
	KPI Porcentaje Fallecida Niños
	KPI Porcentaje Fallecida Adultos
	KPI Porcentaje Fallecida Ancianos
	KPI Recuperados Niños
	KPI Recuperados Adultos
	KPI Recuperados Ancianos

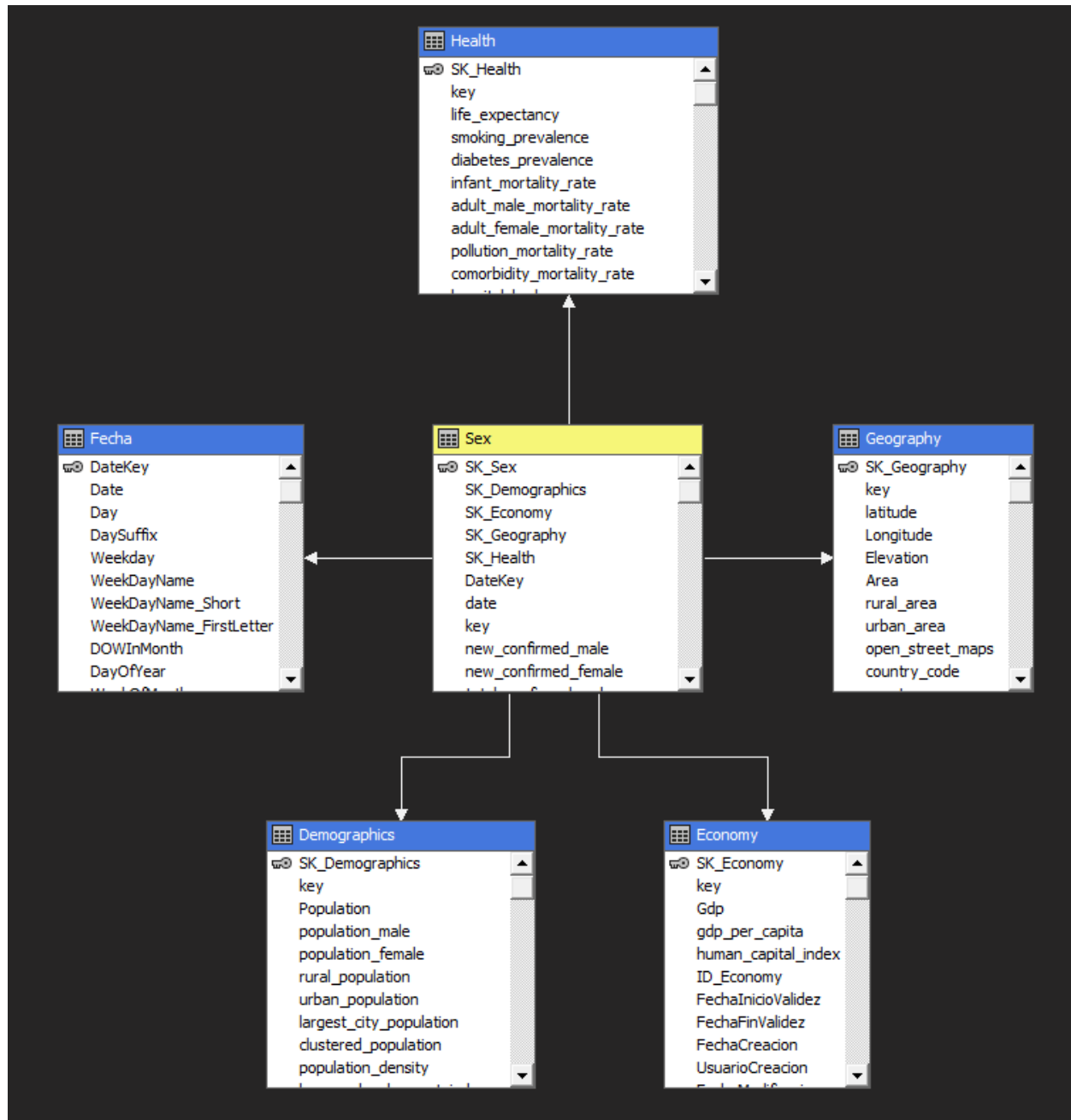


## Epidemiology Cube



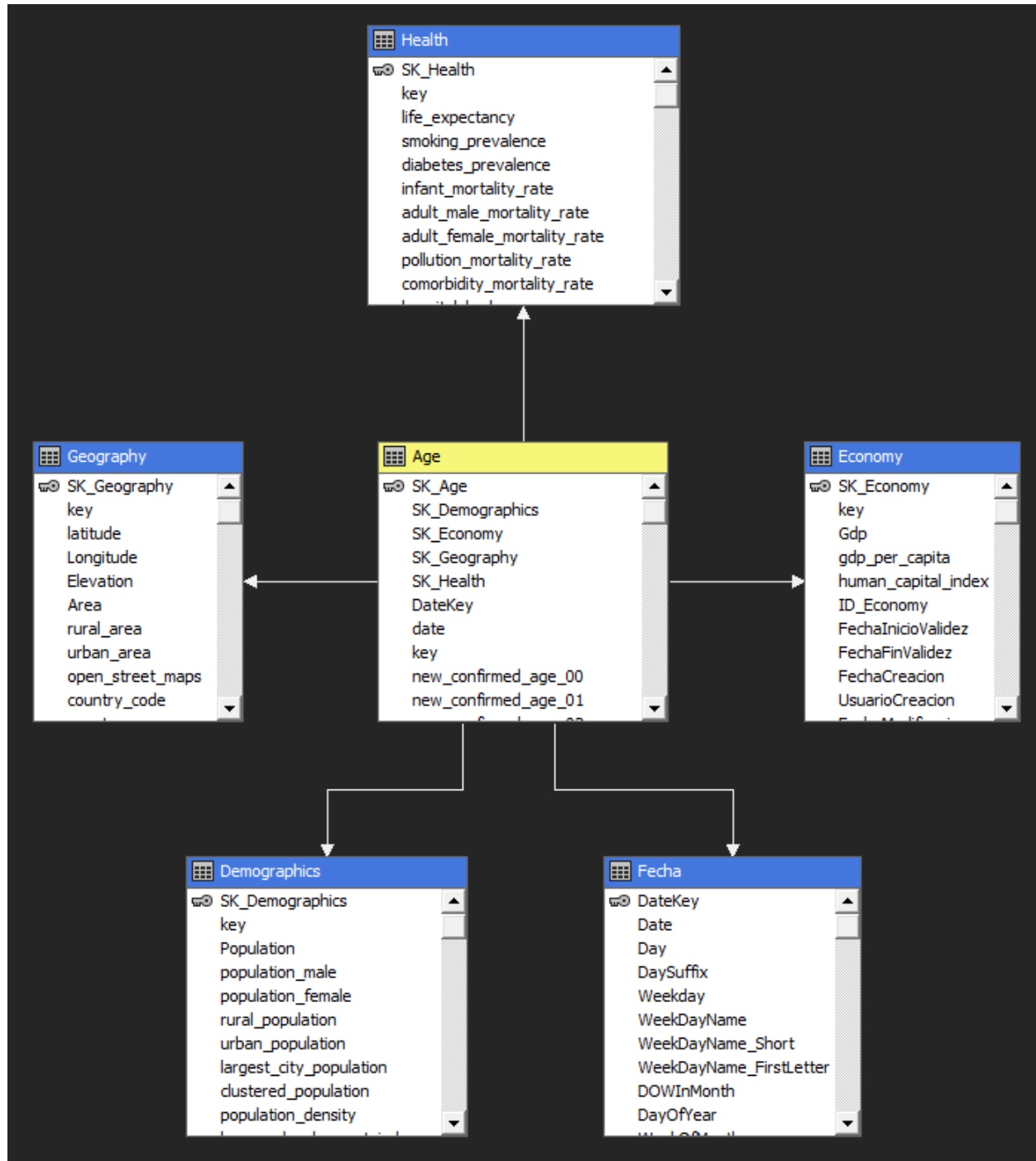


## Sex Cube





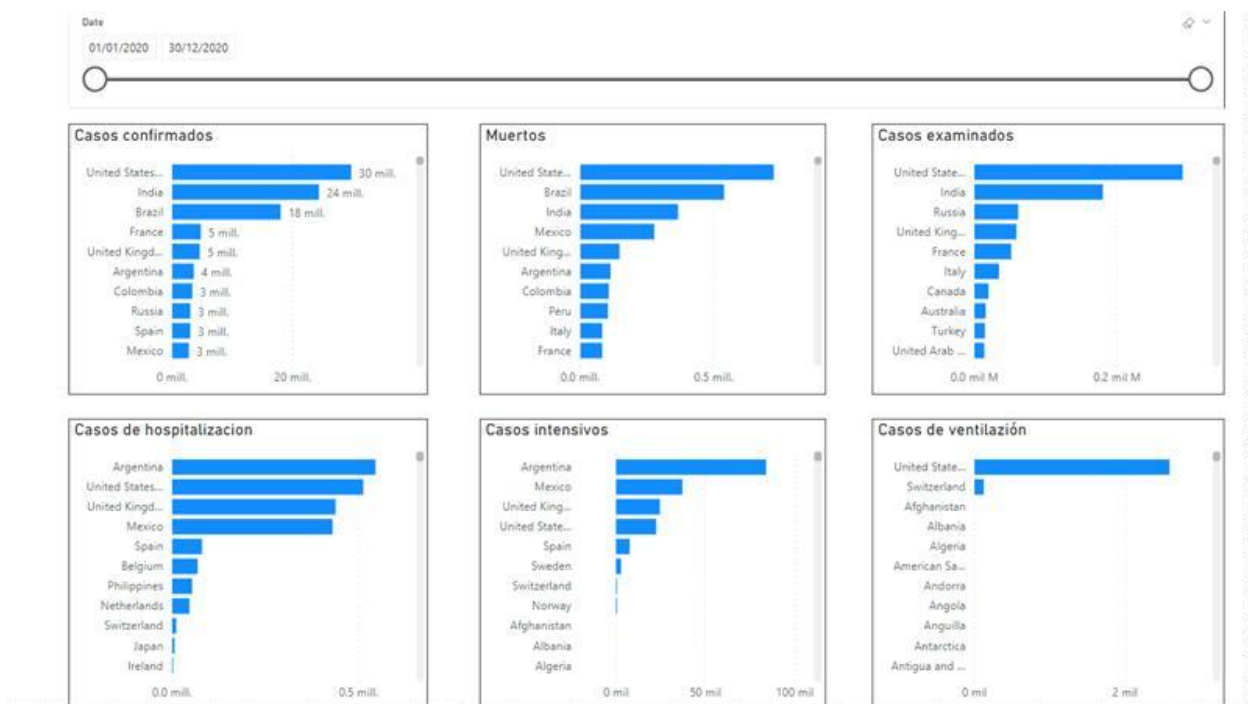
## Age Cube



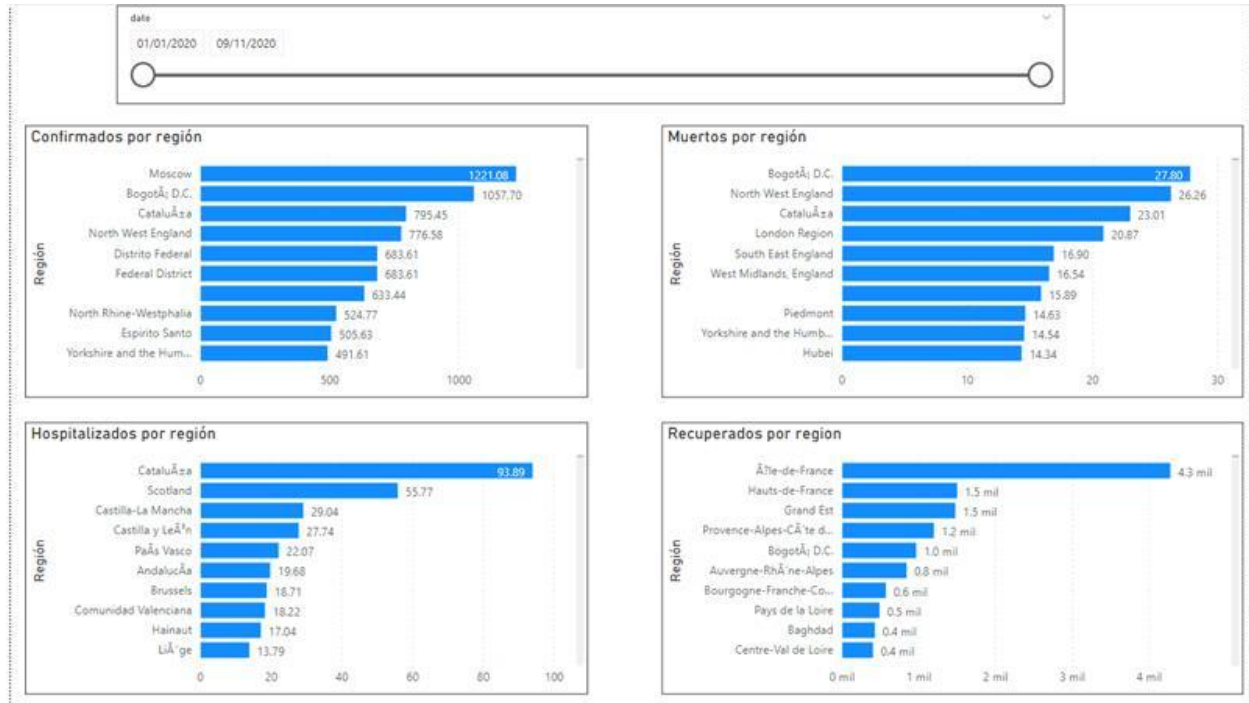
## Fase 5:

En esta fase del proyecto nos enfocamos en presentar los datos de forma amigable a las personas para que cualquier persona que desee observar las gráficas pueda entender que es lo que le quiere transmitir cada una.

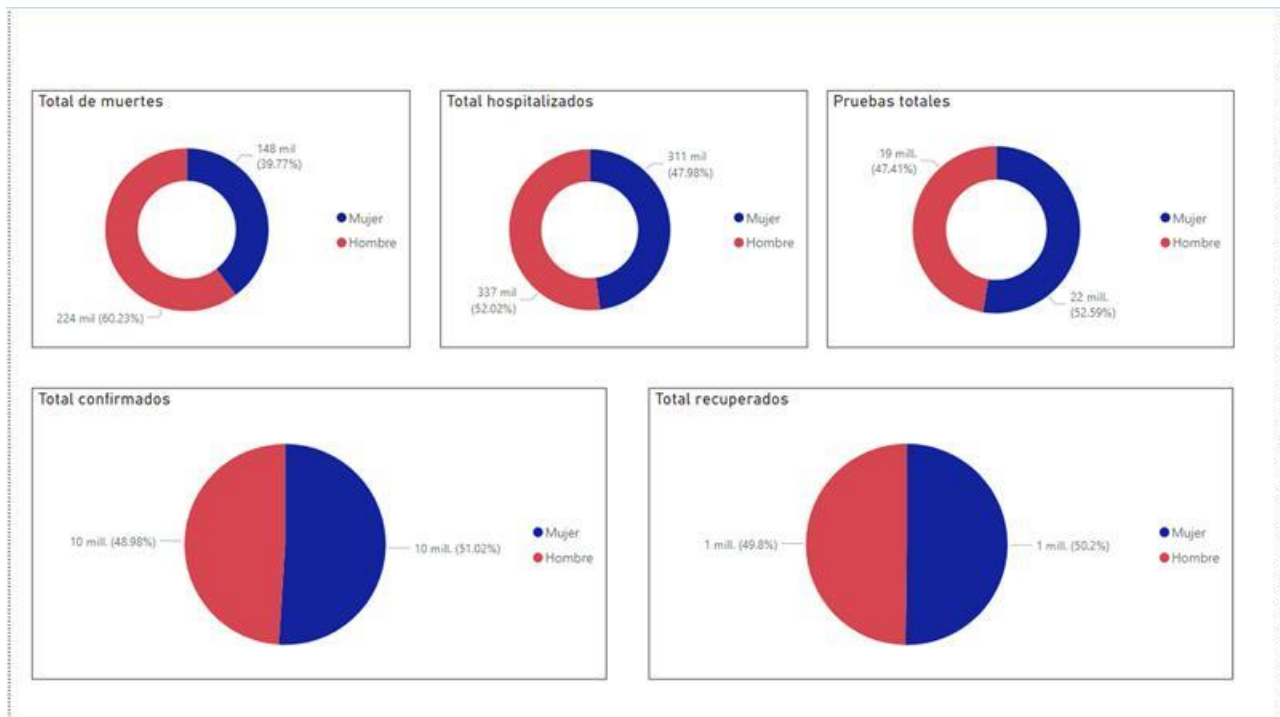
Como primera pestaña está el grupo de totales, que resume la información de diferente país en cada una de las categorías que se presentan en la misma. Para realizar las diferentes vistas se tomaron diferentes campos, pero todos con el denominador común de new, al cual se le hace una suma sobre todo el campo y se agrupan todos por el nombre del país. Podemos observar vistas como total de casos confirmados, muertes, casos examinados, casos hospitalizados, casos intensivos, y por último casos que requieren ventilación. Además, esta pestaña cuenta con un filtrado por fechas, en el cual el usuario puede escoger dos fechas y se le mostrara los datos totales de los diferentes campos a través del tiempo.



En la segunda pestaña “Promedios” encontramos el promedio de las siguientes cuatro categorías: confirmados, muertes, hospitalización y recuperados agrupados por región. Además de tener un filtro en el cual el usuario que desee observar el promedio de los diferentes casos pueda realizarlo.



En la tercera pestaña “Genero” es una pestaña general que da al usuario una vista de los siguientes casos totales: muertes, hospitalizaciones, pruebas, confirmaciones y recuperados de los dos géneros a nivel superficial.

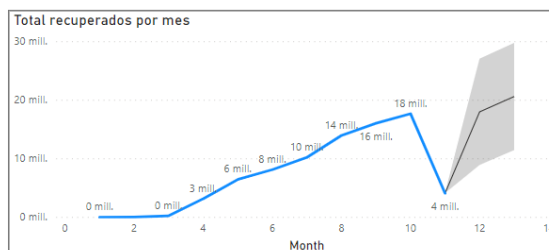
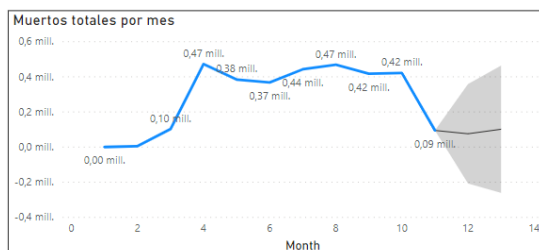
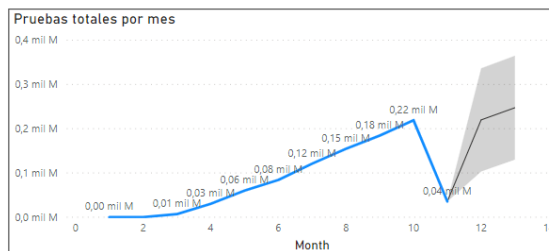
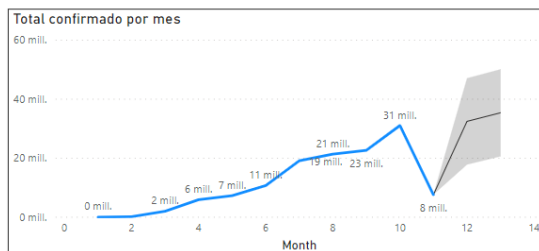


En la cuarta pestaña “Edad” se observa la sumatoria de los diferentes rangos de edad en cada uno de los casos que son afectados por la enfermedad. Entre los campos podemos ver los totales de: confirmados, recuperados, muertes y por último pruebas.





En la quinta pestaña “Tiempo” observaremos unos gráficos totales de los siguientes casos: confirmado, pruebas, muertos y recuperados. Todas estas graficas están agrupadas por un mes del año. Además, cuenta con una predicción de los siguientes dos meses con un nivel de confianza del 95%.



En la sexta pestaña “Mapas” se encuentran dos mapas que son el total de los siguientes casos: confirmados y muertes. En el primer mapa se encuentra un mapa de calor, donde se puede observar como las áreas con más casos confirmados se encuentran de un color más oscuro y a medida que este valor disminuye el color del mapa es directamente proporcional. Por otra parte, el segundo mapa cuenta con un mapa el cual se cuentan los casos y dependiendo de la cantidad de muertes este es más grande o más pequeño y es directamente proporcional a su color.



Para la realización de estos dos mapas se utilizó la longitud y latitud de los países para localizarlos en el mapa. Y adicionalmente se encuentra una tabla donde se menciona el nombre de los países, la cantidad total de confirmados y muertos con el detalle del país y sus respectivas subregiones en el cual si el usuario selecciona alguna fila esta información será filtrada en el mapa mostrando su ubicación, el número de casos confirmados totales y muertes.



country_name	subregion1_name	subregion2_name	Total confirmados	Total muertos
Afghanistan			42048	1556
Afghanistan	Badakhshan		454	7
Afghanistan	Badghis		685	12
Afghanistan	Baghlan		787	30
Afghanistan	Balkh		2474	100
Afghanistan	Bamyan		909	12
Afghanistan	Daykundi		468	3
Afghanistan	Farah		433	15
Afghanistan	Faryab		259	17
<b>Total</b>			<b>128153821</b>	<b>3180518</b>

## Fase 6:

### Deploy y job en servidor UrlServer1/Grupo 1

Se realizó deploy en el proyecto de SSIS, y en los procesos se realizó conexión al servidor y se especificó la ruta de acceso en la que se encontrar el paquete del proyecto.

Asistente para implementación de Integration Services

**Seleccionar destino**

Introducción  
Seleccionar origen  
Seleccionar destino de la implementación  
**Seleccionar destino**  
Revisar  
Resultados

Help

Escriba el nombre del servidor de destino y la ubicación del proyecto en el catálogo de Integration Services.

Nombre del servidor:  
40.117.179.53 Examinar...

Autenticación: Windows Authentication Conectar

Nombre de usuario:

Contraseña:

Ruta de acceso:  
Examinar...

Al finalizar el proceso de autenticación y de seleccionar el destino del paquete, se muestra una pantalla de resultados en la que si el proyecto no posee ningún error mostrara que se implementó de manera exitosa el paquete en el servidor.

Asistente para implementación de Integration Services

**Resultados**

Introducción  
Seleccionar origen  
Seleccionar destino de la implementación  
Seleccionar destino  
Revisar  
**Resultados**

Help

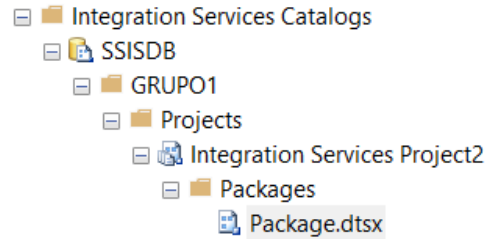
**Resultados**

Acción	Resultado
✓ Carga del proyecto	Correcto
✓ Conexión con el servidor de destino	Correcto
✓ Cambio del nivel de protección	Correcto
✓ Implementación del proyecto	Correcto

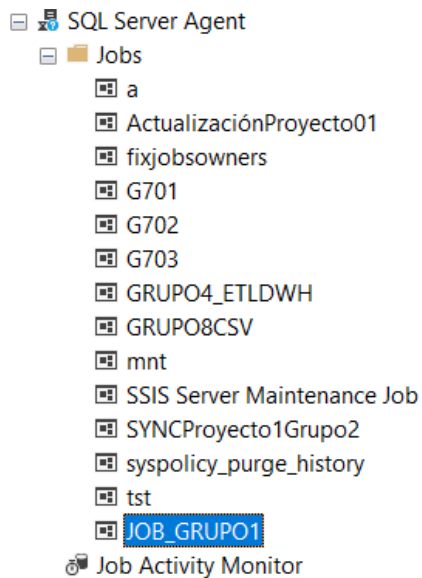
Guardar informe...

< Previous Next > Cerrar Cancelar

En el Server se encuentra la carpeta de integration services catalogs que posee todos los paquetes que desean implementar de SSIS.



Para que se pueda ejecutar el paquete de SSIS en el servidor de manera calendarizada, se crea un Job en el que se le asignara el nombre del usuario, una tarea y la fecha y hora en que se desea ejecutar.



Se le asigno la tarea a nuestro Job, especificando que es un servicio de integration services, seguido de especificar las credenciales para poder acceder al paquete de SSIS.



Job Step Properties - Correr Job

Select a page  
General  
Advanced

Script Help

Step name: Correr Job

Type: SQL Server Integration Services Package

Run as:

Package Configuration

Package source: SSIS Catalog

Server: 40.117.179.53

Log on to the server  
☒ Use Windows Authentication  
☐ Use SQL Server Authentication

User name: Password:

Package: \\SSISDB\GRUPO1\Integration Services Project2\Package.dtsx

Previous Next

OK Cancel

Connection

Server: URLServer1  
Connection: DESKTOP-T72OUIG\Ivan Arango  
[View connection properties](#)

Progress

Ready

Para nuestro Job se programó que se ejecute una vez por día, en la hora establecida de las 23:59:00 horas, comenzando por el 7 de noviembre de 2020 hasta el 22 de noviembre de 2020.

Job Schedule Properties - Calendarización Job

Name: Calendarización Job

Schedule type: Recurring ☒ Enabled

One-time occurrence

Date: 7/11/2020 Time: 20:52:17

Frequency

Occurs: Daily

Recurs every: 1 day(s)

Daily frequency

☒ Occurs once at: 23:59:00

☐ Occurs every: 1 hour(s)

Starting at: 23:59:00

Ending at: 23:59:59

Duration

Start date: 7/11/2020

☒ End date: 22/11/2020

☐ No end date:

Summary

Description: Occurs every day at 23:59:00. Schedule will be used between 7/11/2020 and 22/11/2020