UNIVERSIDAD DON BOSCO INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN DATAWAREHOUSE Y MINERÍA DE DATOS CICLO 02-2020



Entrega del segundo desafío practico

PRESENTADO POR:

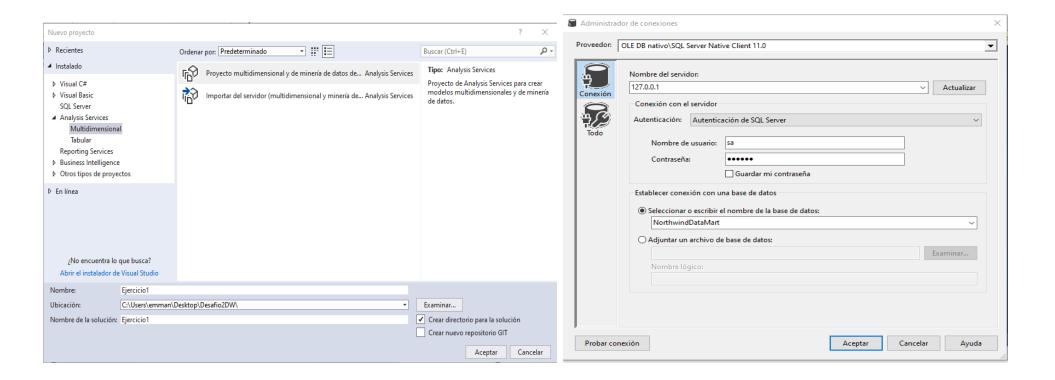
GIRÓN CORNEJO, EMMANUEL ANTONIO GC180313
ZEPEDA LÓPEZ, EDGARDO ANIBAL L180073

DOCENTE: Alexander Sigüenza

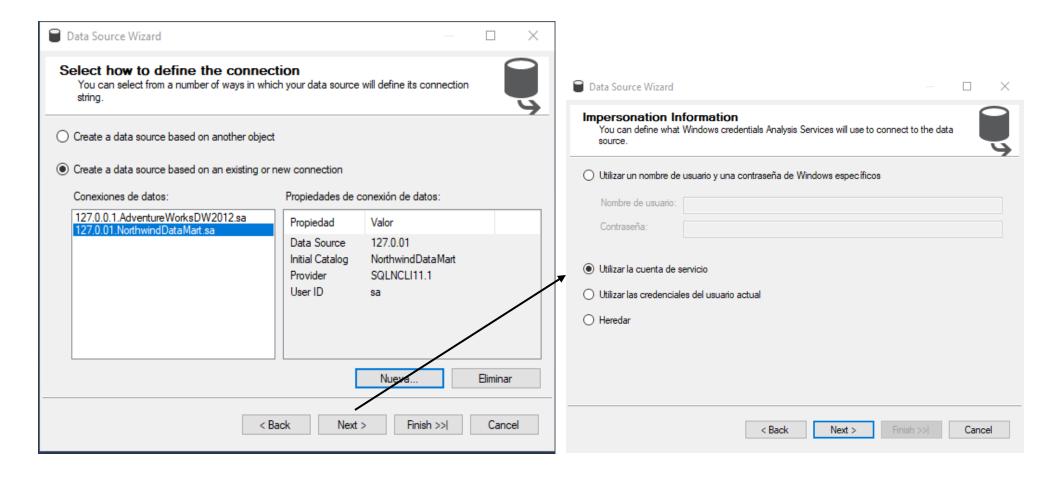
Grupo: 01L

EJERCICIO 1:

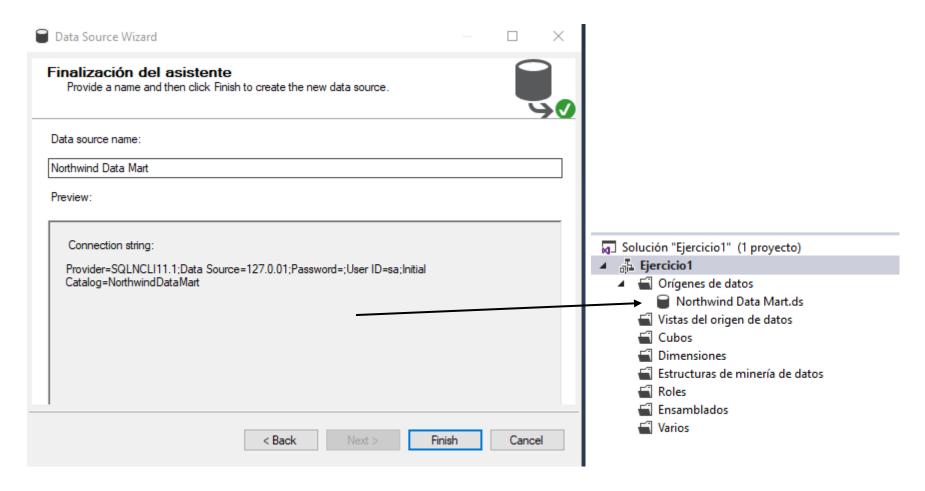
■ Para el ejericio 1 se requiere crear un proyecto en Analysis Service, Y en dicho proyecto re quiere un origen de datos para poder crear la partida de una base de datos hacia el Analysis service, luego especificamos en el apartado de origen de datos a que base de datos vamos a llamar en este caso requiere usar la base de datos NorthwindDataMart.



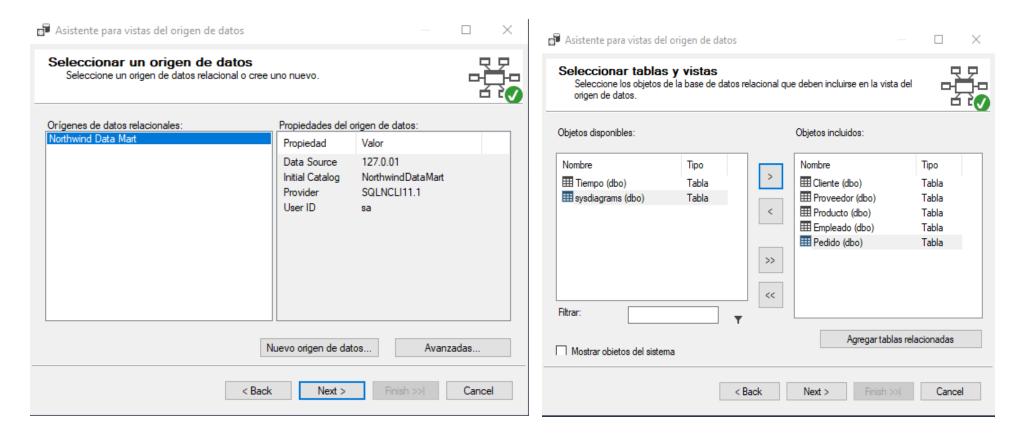
• Ya cuando ya tengamos el nuevo origen de datos nos aparecera asi, que podemos continuar en la selección de conexión y luego debe marcar en el apartado de información el utilizar cuenta de servicio para el origen de datos en la conexión a la base.



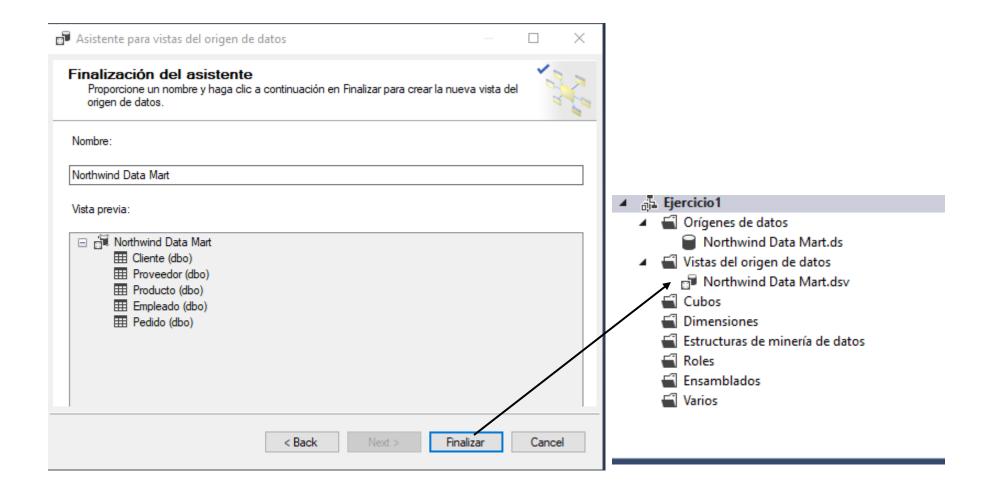
Luego ya llegamos al apartado de finalización donde ya podemos obtener nuestro origen de datos perteneciente a SQL server como se muestra en la imagen



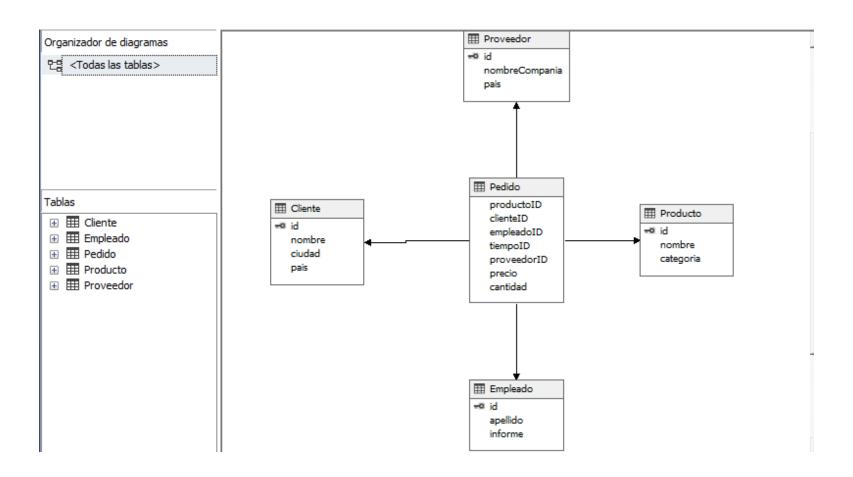
Luego pasamos al apartado de las vistas del origen de datos, ya al haber creado el origen de datos en la misma base de datos en uso, se debera elegir las tablas que seran para crear el diagrama dimensional y en este caso las tablas a elegir son: Cliente, Proveedor, Prodcuto, Empleado y pedido.



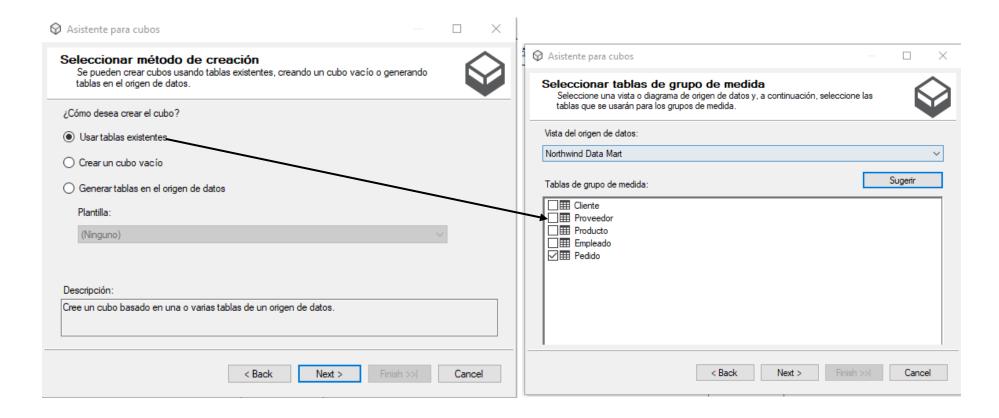
• Ya luego de haber seleccionado las tablas ahora podremos finalizar el proceso y ya en el aparatado de vistas del origen nos aparecera de esta manera.



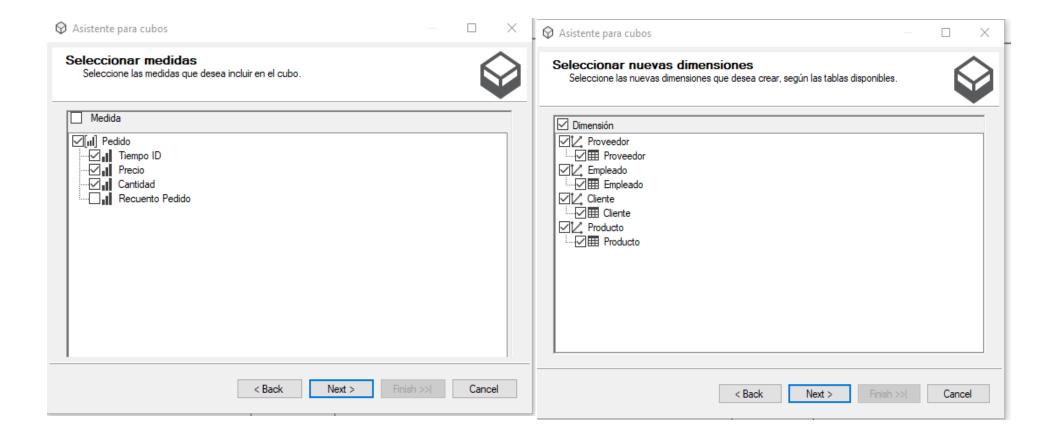
■ Ya cuando se muestre la vista del origen de datos abrimos el diagrama y nos encontramos el diagrama dimensional. De esta manera:



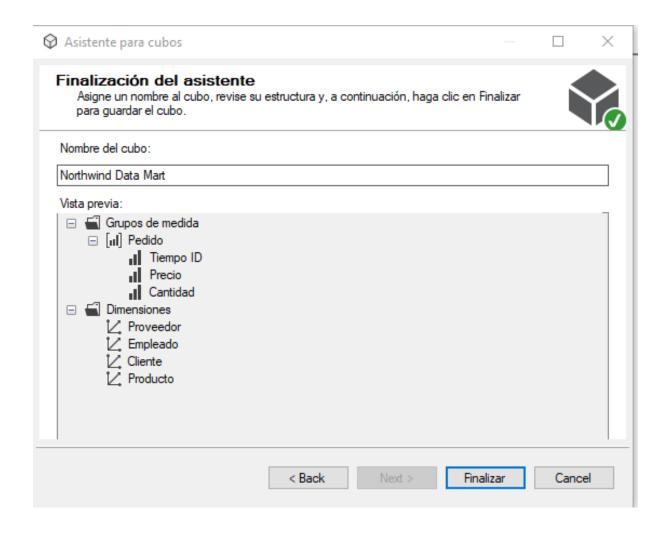
Ahora crearemos el cubo olap, ya con los datos existentes, seleccionamos la tabla hechos o podemos darle en sugerir para que sea mas comodo el notar la tabla hecho para asu vez siempre se puede ver en el diagrama dimensional. Siempre utilizando la base de datos Northwind Data Mart



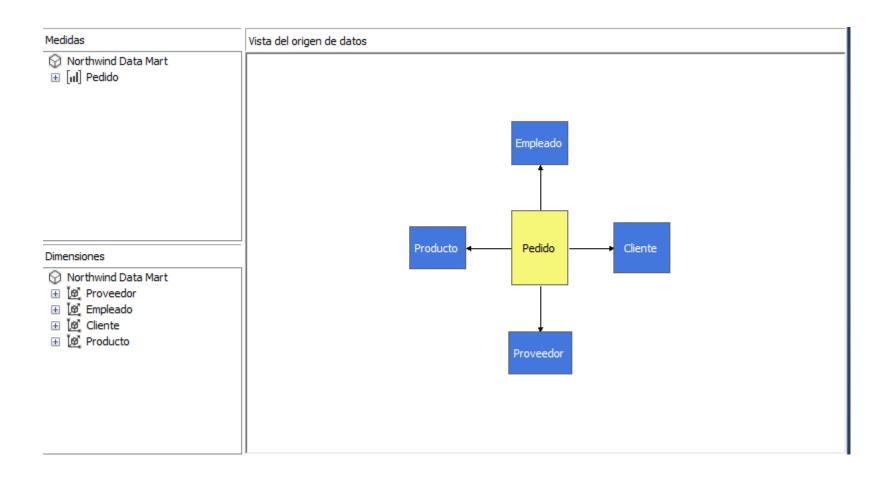
• Luego elegimos las medidas y las dimensiones, en las medidas tendremos que desmarcar el recuente de pedidos ya que no es necesario sabiendo que es la tabla hechos quien predomina solo seleccionaremos 3 campos que serian Tiempo ID, Precio y Cantidad. En el apartado de dimensiones todas deben estar enmarcadas ya que la tabla hechos no aparece ya que en este caso solo estan las que se relacionan de ella que son las dimensiones.



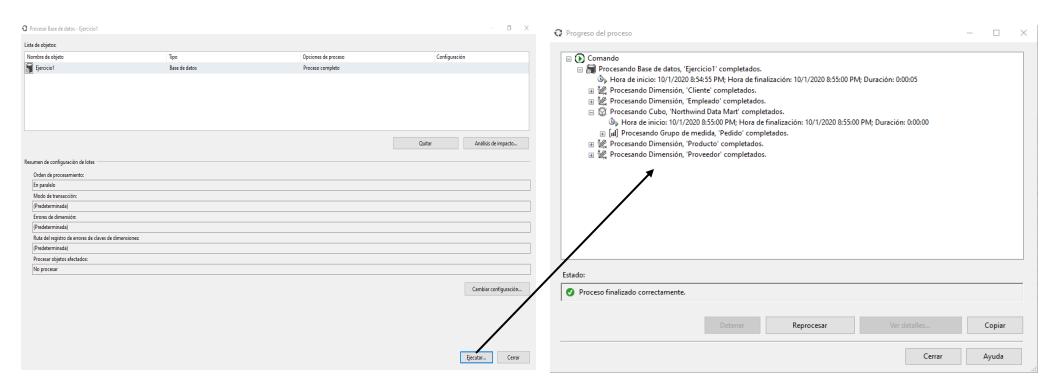
■ Y ya podremos finalizar el proceso de la creación del cubo olap ya con las medidas y dimensiones creadas.



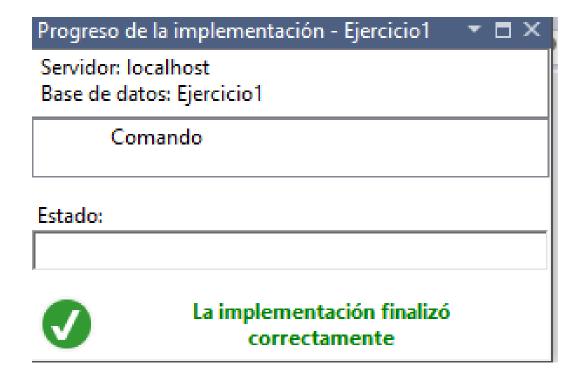
■ Es asi como nos quedaria el cubo olap, con sus dimensiones que seria: Empleado, Producto, Proveedor y Cliente. Y la tabla hechos seria Pedido.



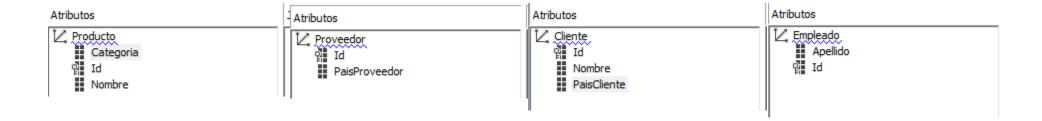
■ Luego de haber creado el cubo habra que procesarlo, y también ejercutar las acciones del cubo.



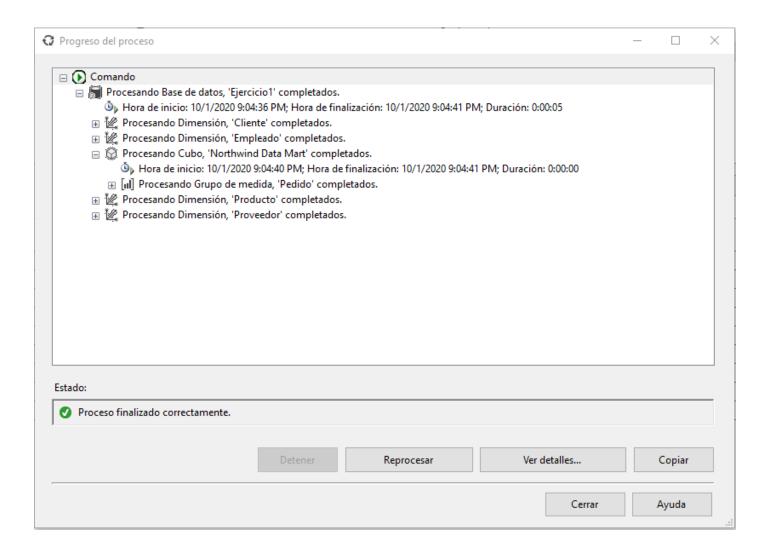
■ Y prodremos notar que ya el cubo se proceso correctamente.



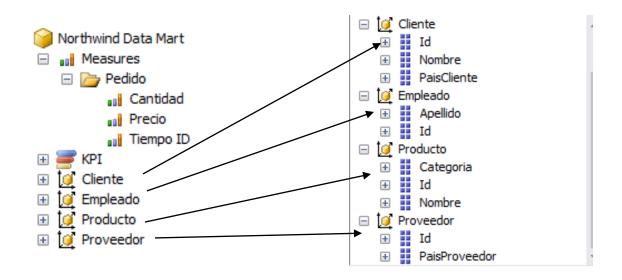
■ Luego de haber procesado el cubo, ya podremos examinarlo, pero habra un inconveniente que en cada campo solamente aparecera el ID, entonces trendemos que seleccionar en cada dimensión y agregar atributos a cada campo. En el caso que se nota un cambio de nombre en ciertos atributos es porque, para que no hubiese una redundancia en el nombre especificamos mejor el nombre del atributo para el campo.



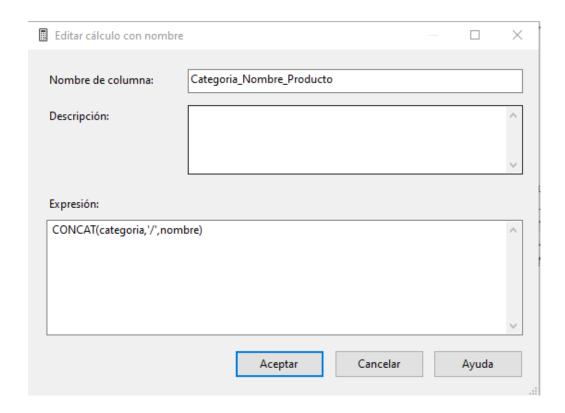
■ Luego procesamos el cubo otra vez al haber hecho dicho cambio y agregar mas atributos a los campos.



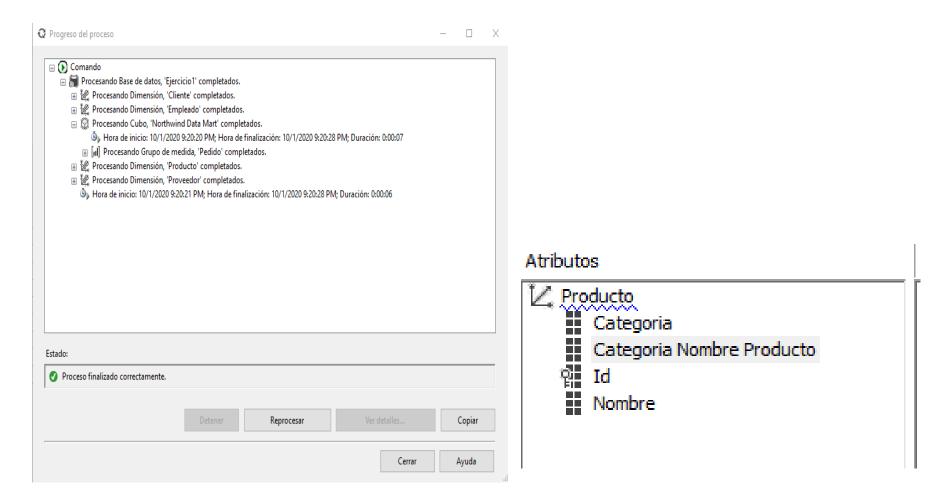
■ Volvemos a examinar el cubo y ya tendremos los atributos para cada campo en el cubo y ya se podra realizar consultas.



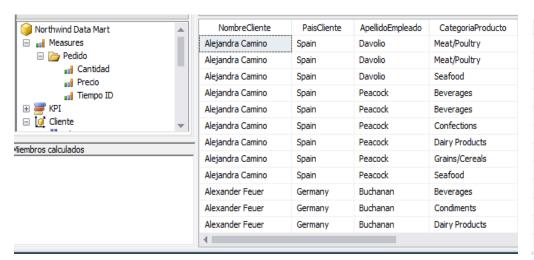
• Pero antes de poder realizar alguna consulta en el examinador se debe hacer un calculo, que pide el ejercicio es concatenar la categoria nombre y producto. En este caso a realizar el calculo, usamos la sintaxis CONCAT(categoria,'/',nombre) Esto permite unir estos campos para poder mostrar el nombre del producto en categoria.



• Volvemos a Procesar el cubo otra vez luego de haber añadido el calculo de nombre. Y podremos denotar que en el campo producto ya aparece el calculo nombre que creamos que por nombre lo definimos como: Categoria Nombre Producto.



• Ahora si se podra realizar las consultas. Como se puede observar se agrego también a la consulta. El calculo nombre que también muestra los datos del producto y del nombre divididos siempre por una pleca.



NombreProducto	Categoria Nombre Producto	PaisProveedor	Cantid 4
Perth Pasties	Meat/Poultry/Perth Pasties	Australia	10
Tourtière	Meat/Poultry/Tourtière	Canada	5
Nord-Ost Matjeshering	Seafood/Nord-Ost Matjeshering	Germany	10
Guaraná Fantástica	Beverages/Guaraná Fantástica	Brazil	6
Steeleye Stout	Beverages/Steeleye Stout	USA	4
Teatime Chocolate Biscuits	Confections/Teatime Chocolate Biscuits	UK	1
Camembert Pierrot	Dairy Products/Camembert Pierrot	France	10
Ravioli Angelo	Grains/Cereals/Ravioli Angelo	Italy	2
Nord-Ost Matjeshering	Seafood/Nord-Ost Matjeshering	Germany	7
Lakkalikööri	Beverages/Lakkalikööri	Finland	10
Vegie-spread	Condiments/Vegie-spread	Australia	6
Mozzarella di Giovanni	Dairy Products/Mozzarella di Giovanni	Italy	30 🔻

Categoria Nombre Producto	PaisProveedor	Cantidad	Precio
Meat/Poultry/Perth Pasties	Australia	10	262
Meat/Poultry/Tourtière	Canada	5	29.5
Seafood/Nord-Ost Matjeshering	Germany	10	207
Beverages/Guaraná Fantástica	Brazil	6	21.6
Beverages/Steeleye Stout	USA	4	57.6
Confections/Teatime Chocolate Biscuits	UK	1	7.3
Dairy Products/Camembert Pierrot	France	10	340
Grains/Cereals/Ravioli Angelo	Italy	2	31.2
Seafood/Nord-Ost Matjeshering	Germany	7	150.09
Beverages/Lakkalikööri	Finland	10	180
Condiments/Vegie-spread	Australia	6	263.4
Dairy Products/Mozzarella di Giovanni	Italy	30	1044

Ejercicio 2 / desafío practico DMD

Se nos pide determinar como a cambiado la movilidad de las personas, en un país de Centroamérica, debido a la pandemia del COVID-19 durante los meses de Febrero a Septiembre del presente año, 2020.

El objetivo de realizar este análisis es encontrar una utilidad práctica, a través del tiempo, para así conocer si las medidas que se han tomado han sido las más acertadas para combatir la pandemia.

Se nos proporcionó una serie de archivos en formato .csv, pertenecientes a cada país de Centroamérica, con los datos del país en cuestión. Ejemplo: departamentos, fechas, cambio en la movilidad de comercio, lugares de trabajo, recreación, etc.

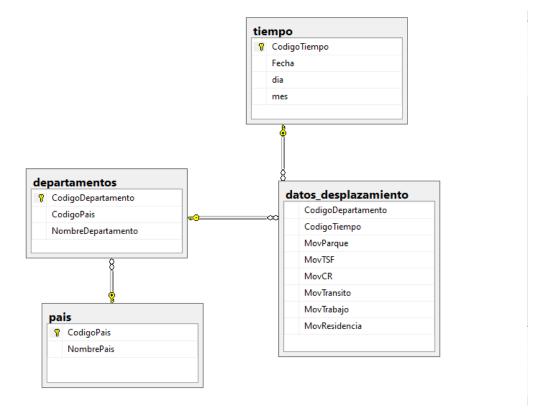
Como equipo decidimos analizar los datos de El Salvador, debido a que así el análisis tendrá un alcance local y será enfocado a la realidad de los salvadoreños.

El proceso a seguir fue el siguiente:

- 1.Previamente se convirtió el archivo .csv a xlxs. para una mayor comodidad en el proceso de ETL
- 2. Proseguimos a creación del datamart que posteriormente usamos para cargar la data.

```
create database desplazamiento_comunidad
use desplazamiento_comunidad
create table pais (
CodigoPais nchar(2) primary key,
NombrePais varchar(50) not null
create table departamentos
CodigoDepartamento nchar(5) primary key,
CodigoPais nchar(2),
NombreDepartamento varchar(50) not null
create table tiempo(
CodigoTiempo int identity(1,1) primary key,
Fecha date,
dia int,
mes int
create table datos desplazamiento(
CodigoDepartamento nchar(5) not null,
CodigoTiempo int identity(1,1),
MovParque float not null,
MovTSF float not null,
MovCR float not null,
MovTransito float not null,
MovTrabajo float not null,
MovResidencia float not null
```

3. Posteriormente relacionamos las tablas y obtuvimos el siguiente diagrama.



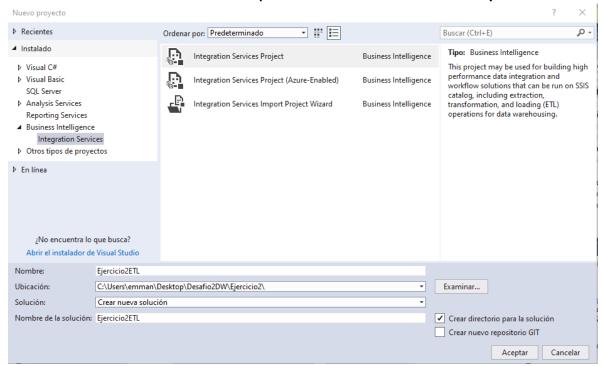
La tabla de pertenecientes a las dimensiones, que consideramos,son: la tabla país, departamentos y tiempo. Las cuales son las que proporcionaron, posteriormente, la información necesaria en el análisis del cubo OLAP.

La tabla de hechos la denominamos: datos_desplazamiento.

El significado de los campos es:

- MovParque: Cambio en Movilidad Parques y Espacios Públicos
- MovTSF: Cambio en Movilidad Tiendas, Supermercados y Farmacias
- MovCR: Cambio en Movilidad Comercio y Recreación
- MovTransito: Cambio en Movilidad Tránsito
- MovTrabajo: Cambio en Movilidad Lugares de Trabajo
- MovResidencia: Cambio en Movilidad Lugar de Residencia

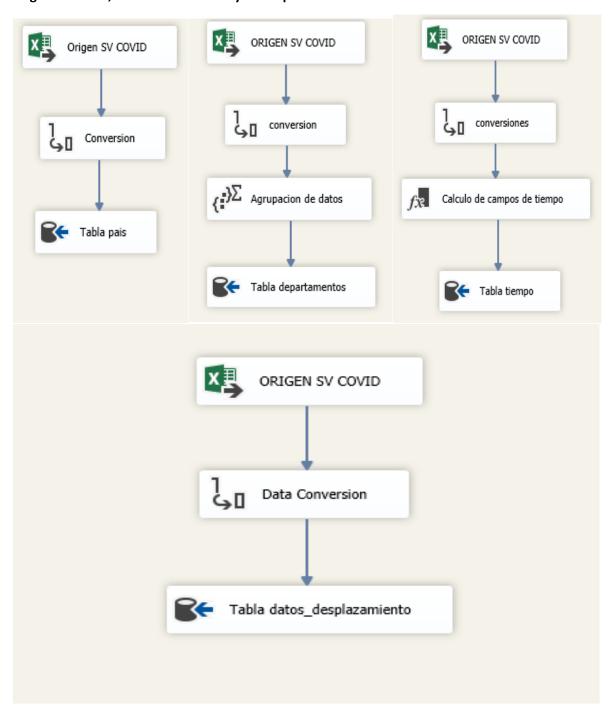
4. Continuamos con la creación de un ETL para llenar las tablas de dimensiones y hechos.



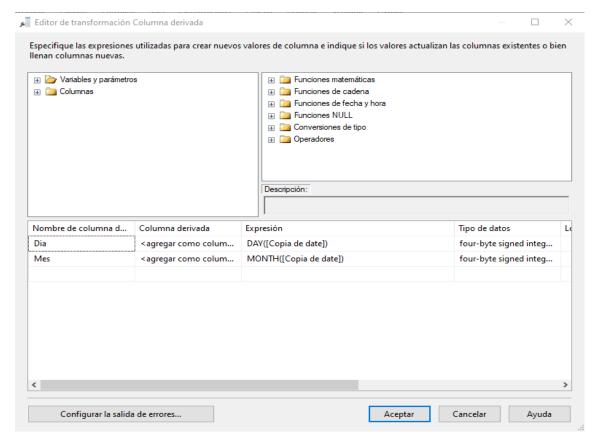
5.Se agregaron 4 dataflow correspondientes a las tablas a llenar.



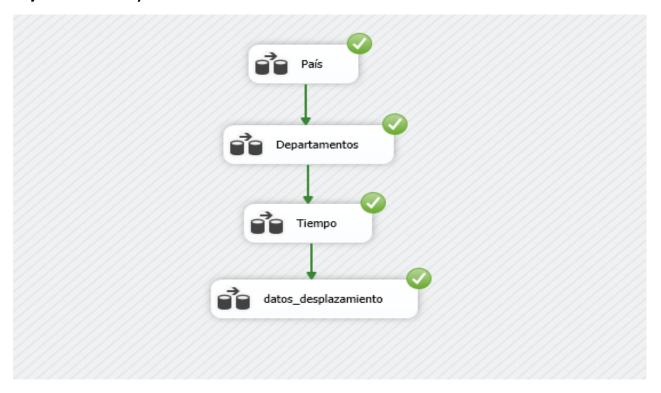
5. Se sigue la misma estructura para llenar las tablas de hechos y dimensiones. Configurando el origen de datos, el destino a la base y las respectivas conversiones de datos.

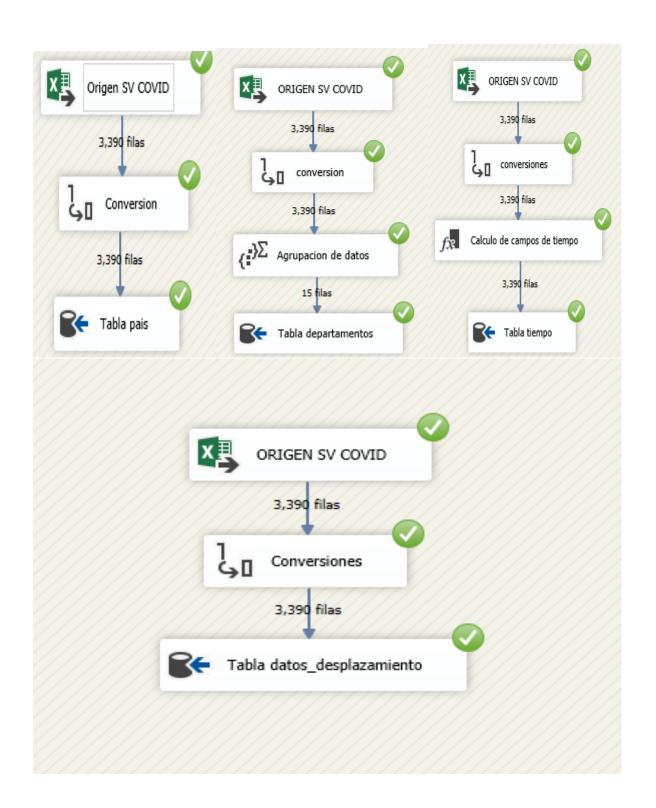


6. Para la tabla tiempo se derivaron las siguientes columnas:

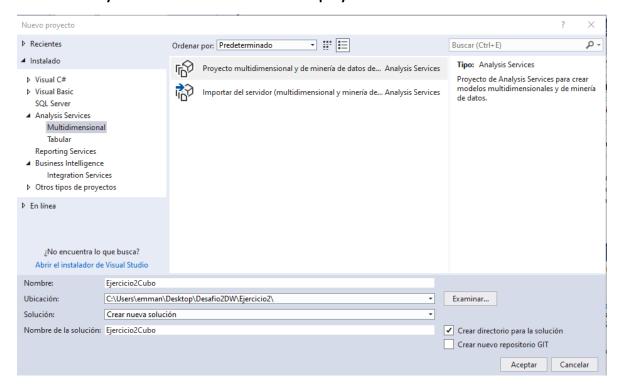


7. Ejecutamos el ETL y el resultado fue satisfactorio.

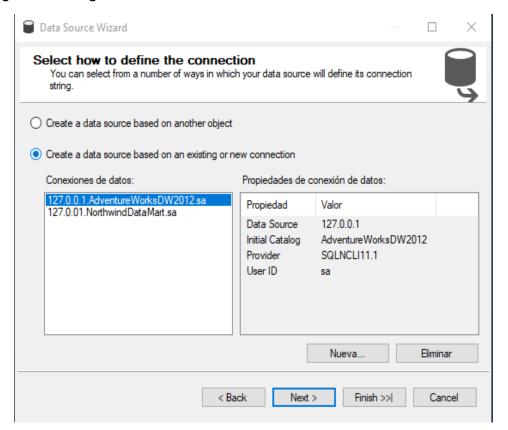


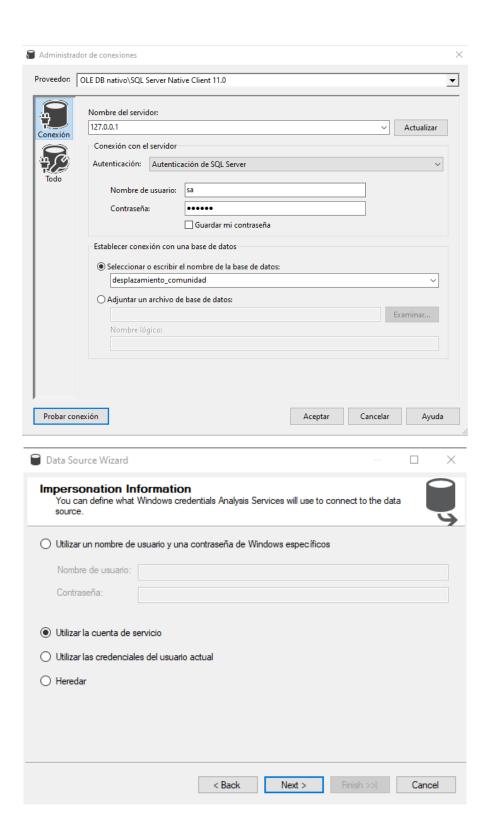


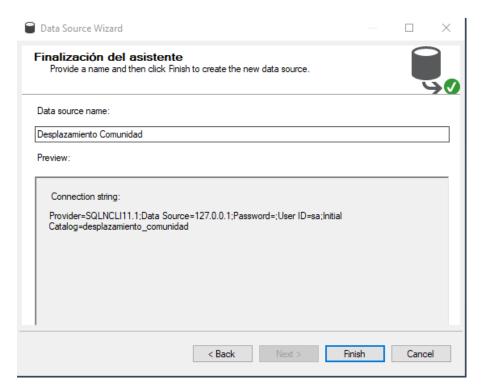
8.Con los datos cargados en el modelo de tipo copo de nieve, proseguimos a la elaboración de un cubo OLAP y así analizar los datos. Creamos el proyecto multidimensional.



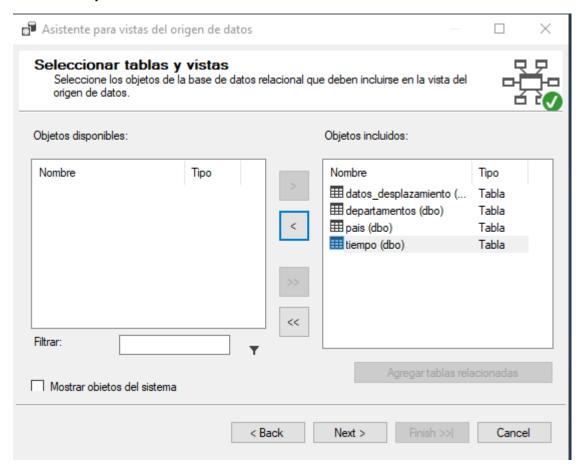
9. Configuramos el origen de datos.

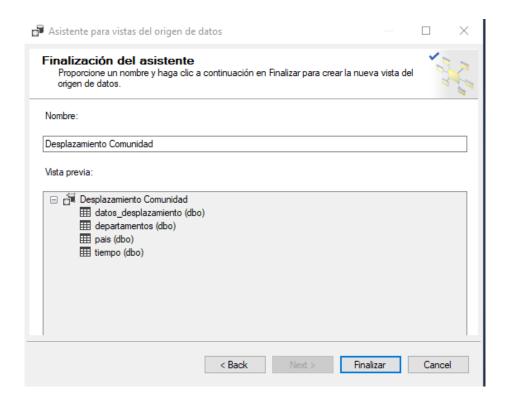




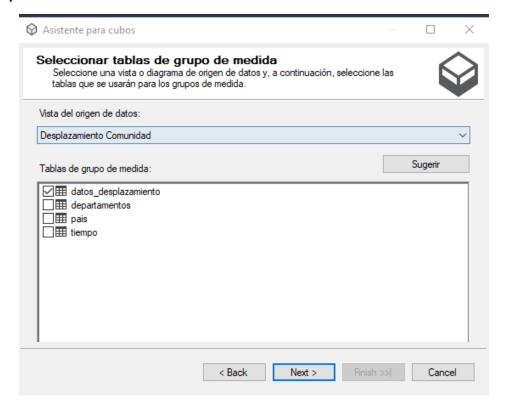


10. Como siguiente punto las vistas de origen de datos, destacamos la selección de las dimensiones y la tabla hechos.

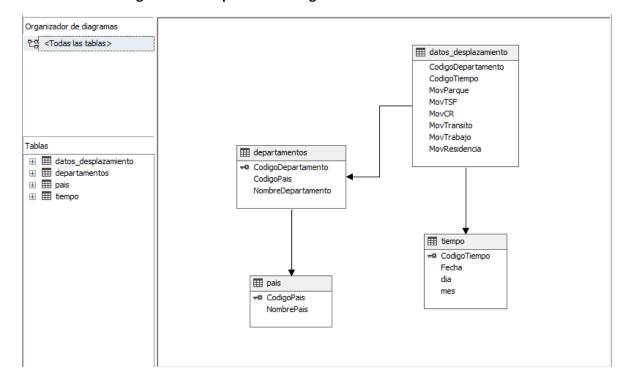




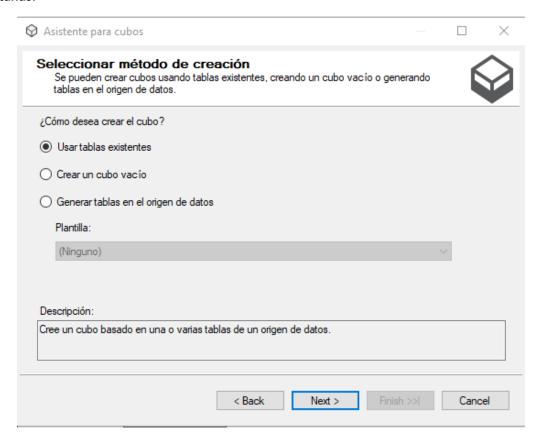
10. Es aquí donde seleccionamos la tabla de hechos en donde están nuestras medidas.

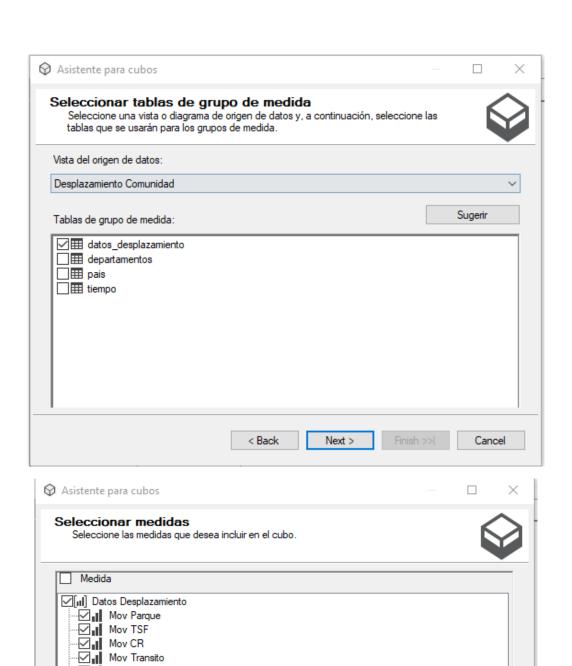


11.Las vistas de origen de datos quedan de la siguiente manera:



12. Proseguimos en la creación del cubo, aplicando la siguiente configuración a las diferentes ventanas:



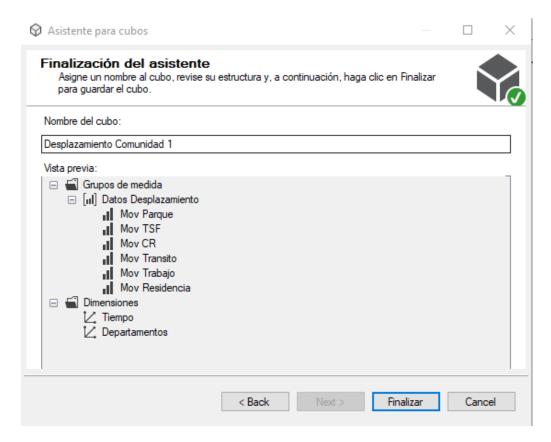


< Back

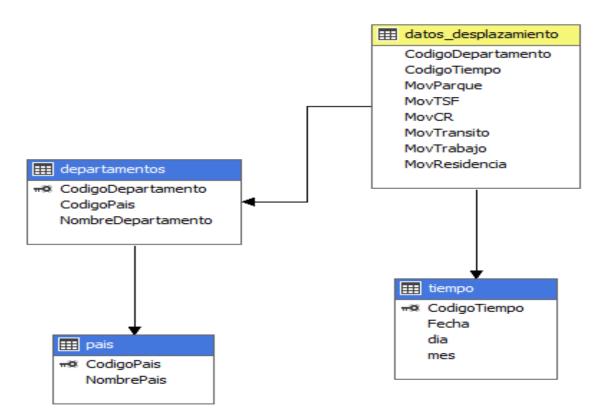
Next >

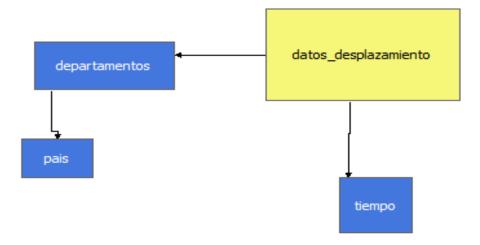
Cancel

Recuento Datos Desplazamiento



14. Obteniendo como resultado el cubo OLAP

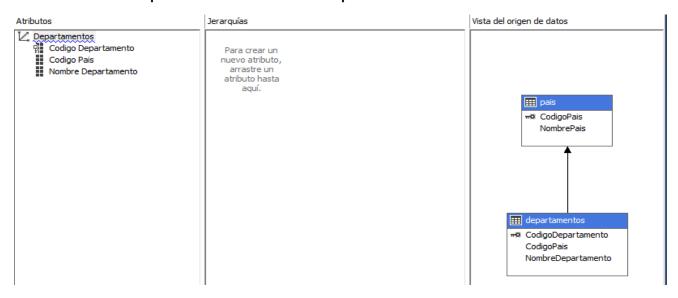




15. Agregmos los atributos a las dimensiones correspondientes.



16. Hacemos las distinción que para la dimensión de departamento se incluye el Código del país como atributo representando a la dimensión de país.



17. Proseguimos a realizar unas consultas para su análisis.

Nombre Departa	Mov Parque	Mov CR	Mov Residencia	Mov Trabajo	Mov Transito	Mov TSF
La Libertad Depa	-10762	-9614	4724	-8451	-7959	-6861
La Paz Department	-213	-50	0	90	-544	-122
San Miguel Depa	-7802	-6785	2925	-6138	-8428	-5036
San Salvador De	-10101	-10284	4464	-8530	-9924	-6302
Santa Ana Depa	-8520	-8020	3416	-6769	-9395	-5274
Sonsonate Depa	-6438	-4794	2566	-4991	-7071	-4109

Tomando de muestra los departamentos que presentaron una mayor calidad en los datos, como La Libertad, San Miguel, San Salvador, Santa Ana y Sonsonate. Es posible notar que para los diferentes tipos de movilidad en los diferentes rubros, como: cambio en movilidad parques y espacios públicos, cambio en movilidad tiendas, supermercados y farmacias, cambio en movilidad comercio y recreación, cambio en movilidad tránsito y por último, cambio en movilidad lugares de trabajo, presentan valores negativos lo que representa que por ejemplo: para la libertad hubo -10762 personas que se movilizaron en parques y espacios públicos. Los valores negativos indican el menor grado de afluencia de personas comparado a un valor base. Y es interesante ver que todos los rubros presentan valores negativos, excepto el de la movilidad en la residencia, presentando valores positivos indicando que aumento la afluencia de personas en casa, lo que es obvio debido a que una buena parte de la población acato las medidas de distanciamiento.

Mes	Mov Parque	Mov CR	Mov Residencia	Mov Trabajo	Mov Transito	Mov TSF
2	-2106	-2047	841	-1519	-1969	-1262
3	-6898	-6382	2800	-5445	-6695	-4125
4	-6427	-5849	2677	-5000	-5970	-4148
5	-7399	-6558	3047	-6053	-7127	-4827
6	-5712	- 4 901	2333	-4517	-6164	-3565
7	-5678	-5089	2413	-4669	-5615	-3506
8	-5957	-5255	2454	-4756	-5832	-3844
9	-3659	-3466	1530	-2830	-3949	-2427

Ahora Podemos verlo a través del paso de los meses, de febrero a septiembre, se puede ver en todos los rubros, menos en MovResidencia, que la movilidad va disminuyendo paulatinamente, hasta que alcanza un pico en mayo para aumentar posteriormente de la misma forma.

Los últimos datos que se tienen por ejemplo para la movilidad en lugares de trabajo, hubo 2830 personas menos movilizándose en ese rubro, en comparación al mes de febrero que hubo menos 1519 personas. Se comprueba que la situación se esta normalizando poco a poco.

Es evidente notar que la movilidad en las residencias aumento, pero en los último mes presento un descenso, lo que significa que ya menos personas se esta quedando en casa.

En conclusión del análisis, podemos decir que esta información puede ayudar a determinar si las medidas que se toman para combatir la pandemia son las más efectivas, se puede comparar las cifras de movilidad con los casos de COVID que se presentaron en un mes determinado. Por

ejemplo: en el mes de julio hubo menos 4669 personas en lugares de trabajos y 600 casos de COVID, y en el mes de agosto menos 4756 personas en el mismo rubro y 500 casos de COVID, se comprueba las medidas tomadas en la fecha, para su posterior aplicación en el futuro. Sirviendo de indicador para así mantener un equilibrio entre la economía y la salud de las personas. Y así juntos prepararnos para futuras emergencias.