Set de datos:

Operacion

Recibo

Feccobro

Horcobro

Agencia

Nombre

Tipo

Departamento

Municipio

Producto

Operador

Valorefectivo

Valorcheque

Valorotrosmedios

Valor total

identificar las entidades principales y sus relaciones: En este caso, las entidades principales son:

Operación

Recibo

Agencia

Producto

Cajero

Las relaciones entre estas entidades se pueden establecer de la siguiente manera:

Cada operación se relaciona con un recibo y una agencia.

Cada recibo se relaciona con un producto.

Cada agencia se relaciona con un cajero.

Modelo dimensional identificar los hechos y las dimensiones

Hechos: eventos medibles

idfecha, idhora, idagencia, recibo,operación, idproducto, valorefectivo,valorcheque,valorotrosmedio,valor total

Dimensiones: atributos que describen los hechos

* Fecha : fecha de cobro
* Hora: hora de cobro
* Agencia: código de agencia, nombre, tipo, departamento y municipio, código del cajero que recibió el pago.
* Producto: descripción del producto del que se recibió el pago.

Tabla fecha:

|  |  |
| --- | --- |
| Key\_Fecha [PK] | Clave primaria de la dimensión |
| Fecha | Fecha en la que se efectuó el cobro. |
| Dia | Dia de la fecha. |
| Mes | Mes de la fecha. |
| Año | Año de la fecha. |
| EstacióndelAño | Estación del año correspondiente a una fecha. |

Tabla Hora:

|  |  |
| --- | --- |
| Key\_hora [PK] | Clave primaria de la dimensión. |
| Hora | La hora de cobro |
| Horas | Horas del cobro. |
| Minutos | Minutos del cobro. |
| Segundos | Segundos del cobro. |
| Periodo [madrugada-mañana-tarde-noche] | El periodo de atención para la hora cobro. |
| HoraIngreso | Si contiene hora de ingreso S o N. |
| HoraCierreAgencia | Si contiene hora de cierre de la agencia S o N. |
| Horaalmuerzo | Si contiene hora de almuerzo S o N. |
| Horasalida | Si contiene hora de salida S o N. |
| Horalaborable | Si contiene hora laborable S o N. |

Tabla agencia:

|  |  |
| --- | --- |
| Key\_Agencia [PK] | Clave primaria de la dimensión. |
| Agencia\_Nombre | Nombre de la agencia. |
| Agencia\_Tipo | El tipo de agencia si es normal, caseta, premium. |
| Departamento | El departamento donde se ubica la agencia. |
| Municipio | Municipio donde se ubica la agencia. |
| Código\_operador | Código del cajero que recibió el pago. En conjunto el código de agencia identifica de manera única al cajero. |

Tabla producto:

|  |  |
| --- | --- |
| Key\_Producto [PK] | Clave primaria de la dimensión |
| Producto\_Descripcion | Descripción del producto del que se recibió el pago. |

Tabla de hechos [Fact\_Pagos]:

|  |  |
| --- | --- |
| Key\_operacion [PK] | Clave primaria identificación única de la operación efectuada en una agencia, unión entre la columna de OPERACIÓN y RECIBO, RECIBO Identificador del pago relevante para el cliente de la transacción, junte ambas para sacar la llave primaria de mi tabla de hechos. |
| Key\_Agencia [FK] | Clave foránea de la tabla de agencia. |
| Key\_Producto [FK] | Clave foránea de la tabla de producto. |
| Key\_Fecha [FK] | Clave foránea de la tabla fecha. |
| Key\_Hora [FK] | Clave foránea de la tabla hora. |
| ValorEfectivo | Monto del pago recibido en efectivo. |
| ValorCheque | Monto del pago recibido en cheques del banco receptor. |
| ValorOtrosMedios | Monto del pago recibido en cheques de otros bancos que se reciben bajo la reserva usual. |
| ValorTotal | Monto total del pago recibido, esto sumara las tres columnas anteriores. |



Clasificando las columnas por campos agrupados son los group y los campos agregados los sum.

Creacion de la tabla de hechos FACT\_PAGOS.

Aqui se encuentra el dataset brindado por el enunciado del proyecto.

Se realizo la limpieza de cada columna del dataset tomando como filtro valores no NULL.

Aqui se crea la tabla SETDATOS.

Se hace join con la DIM\_FECHA(1) que contiene feccobro y día, mes, año

Se agregan los campos de fecha, día, mes, año

Se agrega la llave primaria

Se crea la tabla DIM\_FECHA

Se hace join con la DIM\_HORA(1) que contiene horacobro, hora,minuto,segundos

Se agregan los campos de horas, minutos,segundos,periodo

Se agrega la llave primaria

Se crea la tabla DIM\_HORA

Se agrega agencia, nombre,tipo,departamento,municipio

Al campo agencia se cambia por Key\_Agencia

Se crea la tabla DIM\_AGENCIA

Se agrega producto

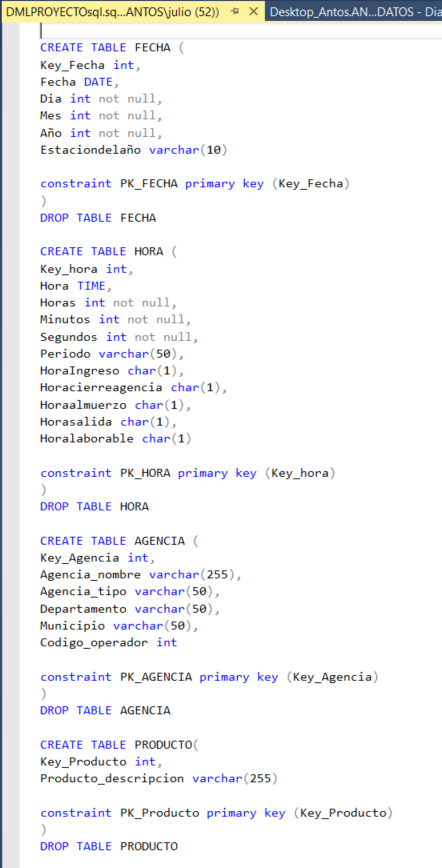
Se agrega la llave primaria

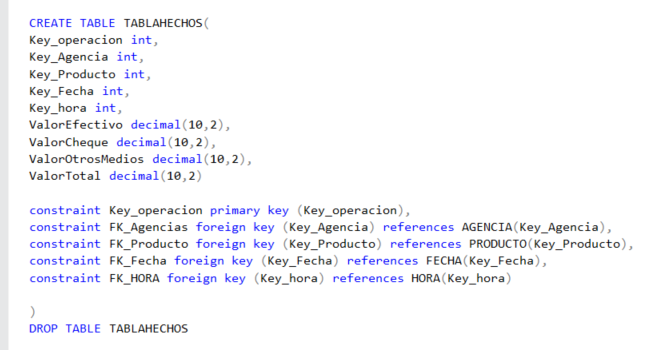
Se crea la tabla DIM\_PRODUCTO

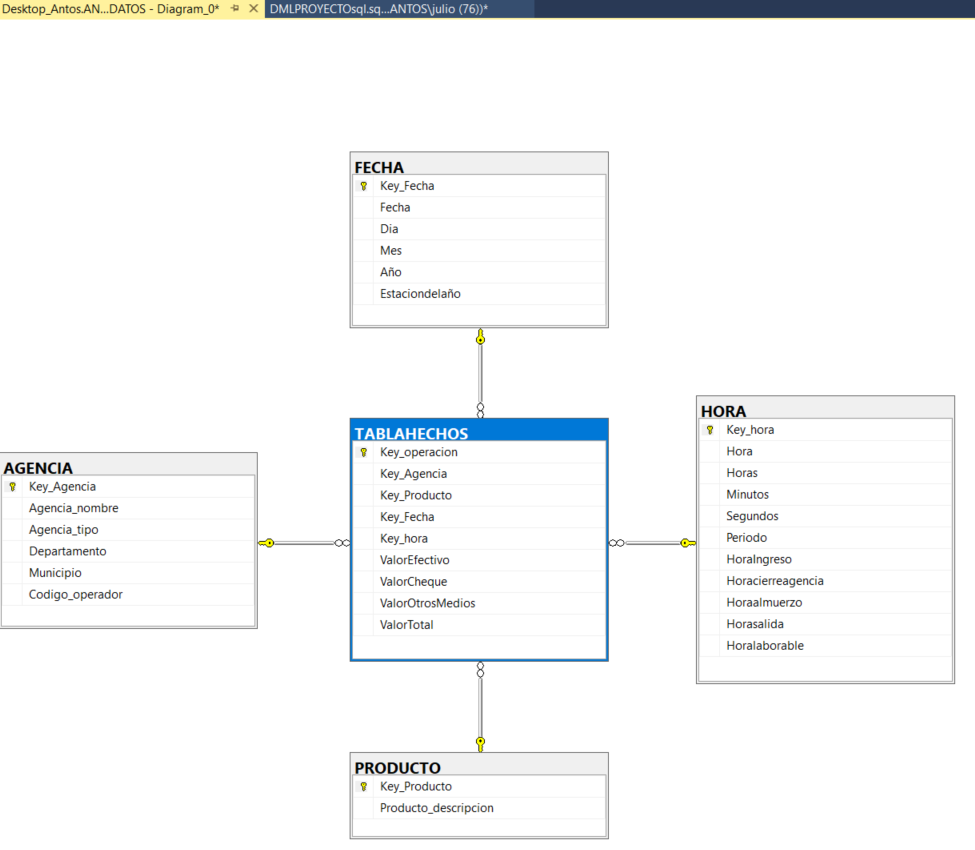
Se agrega operador

Se agrega la llave primaria

Se crea la tabla DIM\_OPERADOR







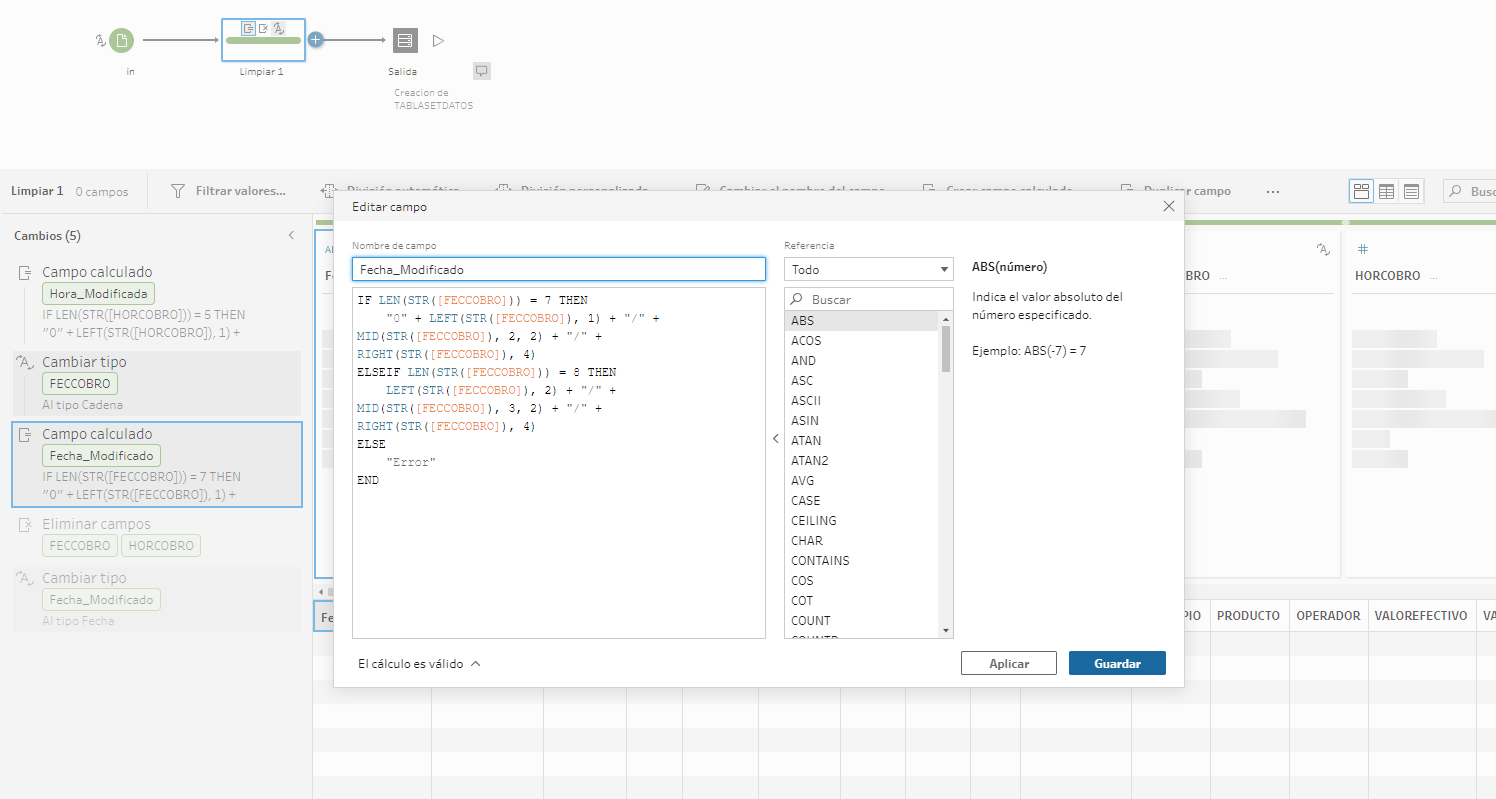
¿Cuáles son los productos más populares entre los clientes de los bancos?

¿En qué días de la semana se realizan más pagos en efectivo?

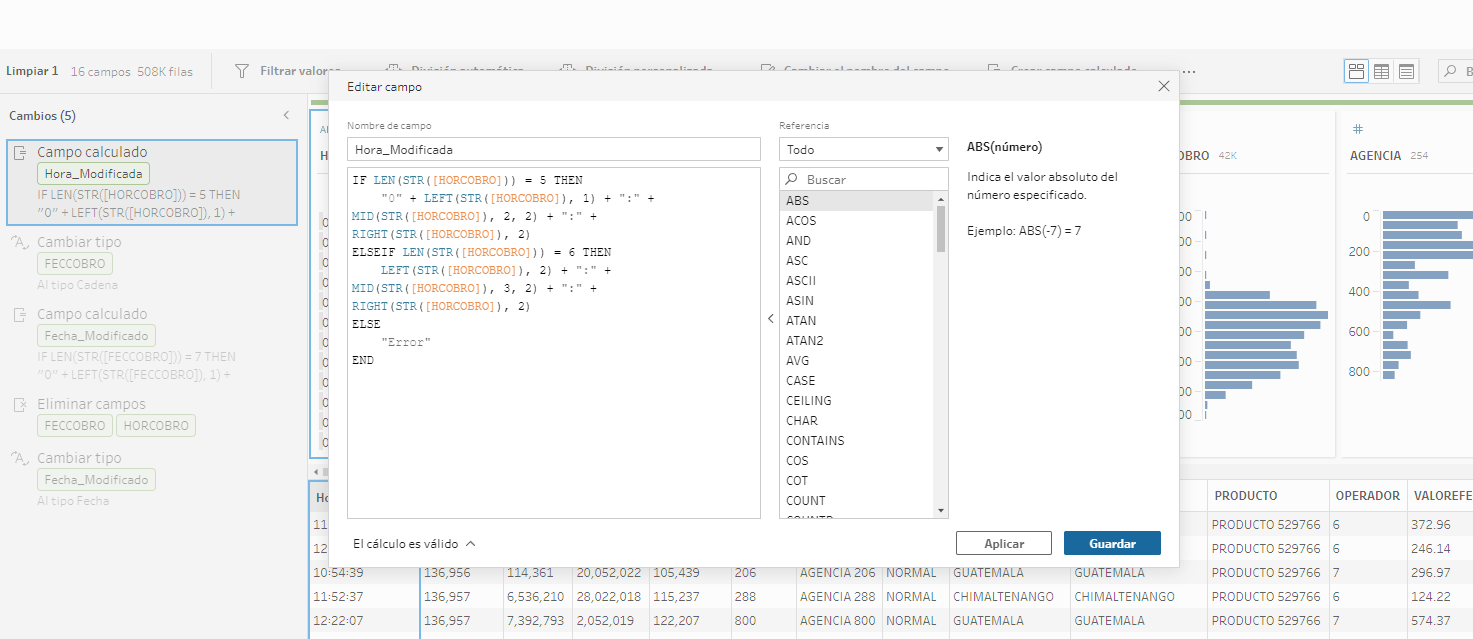
¿Cuáles son las agencias que reciben más pagos por departamento?

¿Cuáles son los municipios con mayor cantidad de pagos realizados?

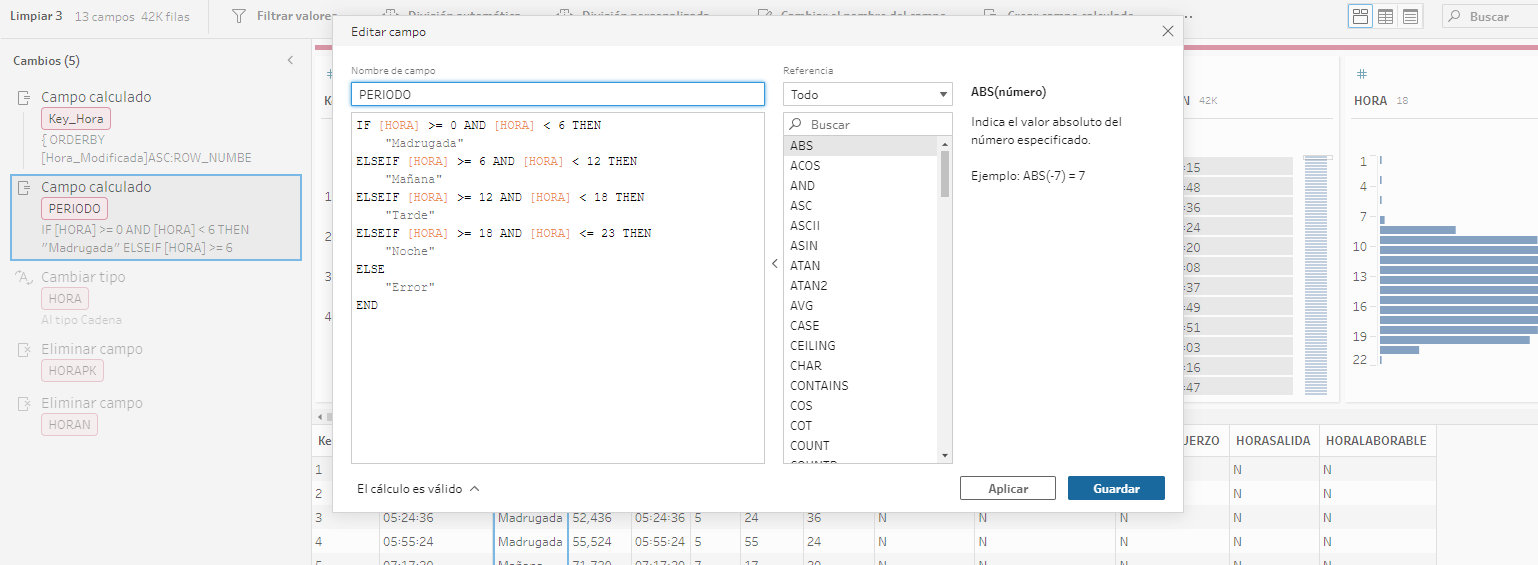
¿Cuáles son los operadores que reciben más pagos en efectivo?



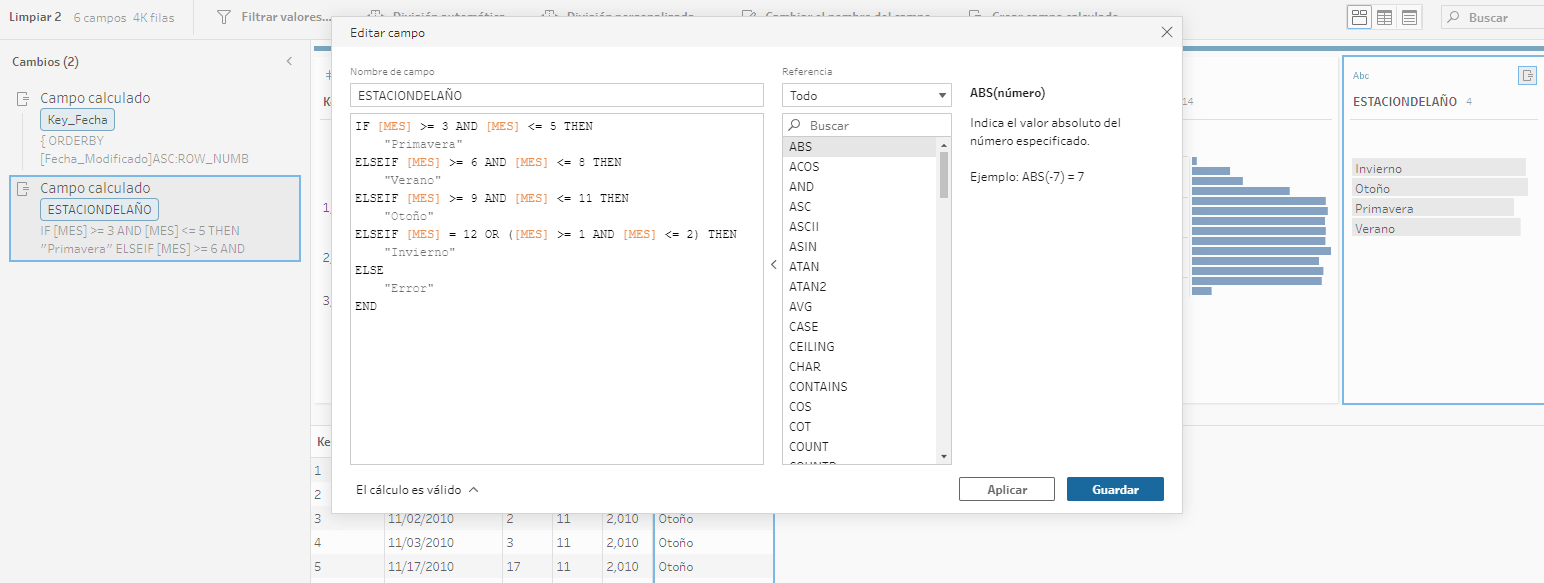
Una vez cargado nuestro dataset (xlsx) en tableau procedemos a hacer una función que convierta la columna de FECCOBRO a formato de fecha inglesa (MM/DD/YYYY), para ello se creó un campo calculado y se coloco de nombre Fecha\_Modificada y se procedió a crear la expresión, esta evalúa la longitud del numero en el campo de FECCOBRO y dependiendo de si tiene 7 u 8 dígitos, se le agregara un 0 al inicio si es necesario a los números de 7 dígitos, para los de 8 dígitos no es necesario agregar dicho 0 y luego separar los números en formato de fecha ingles (MM/DD/YYYY).



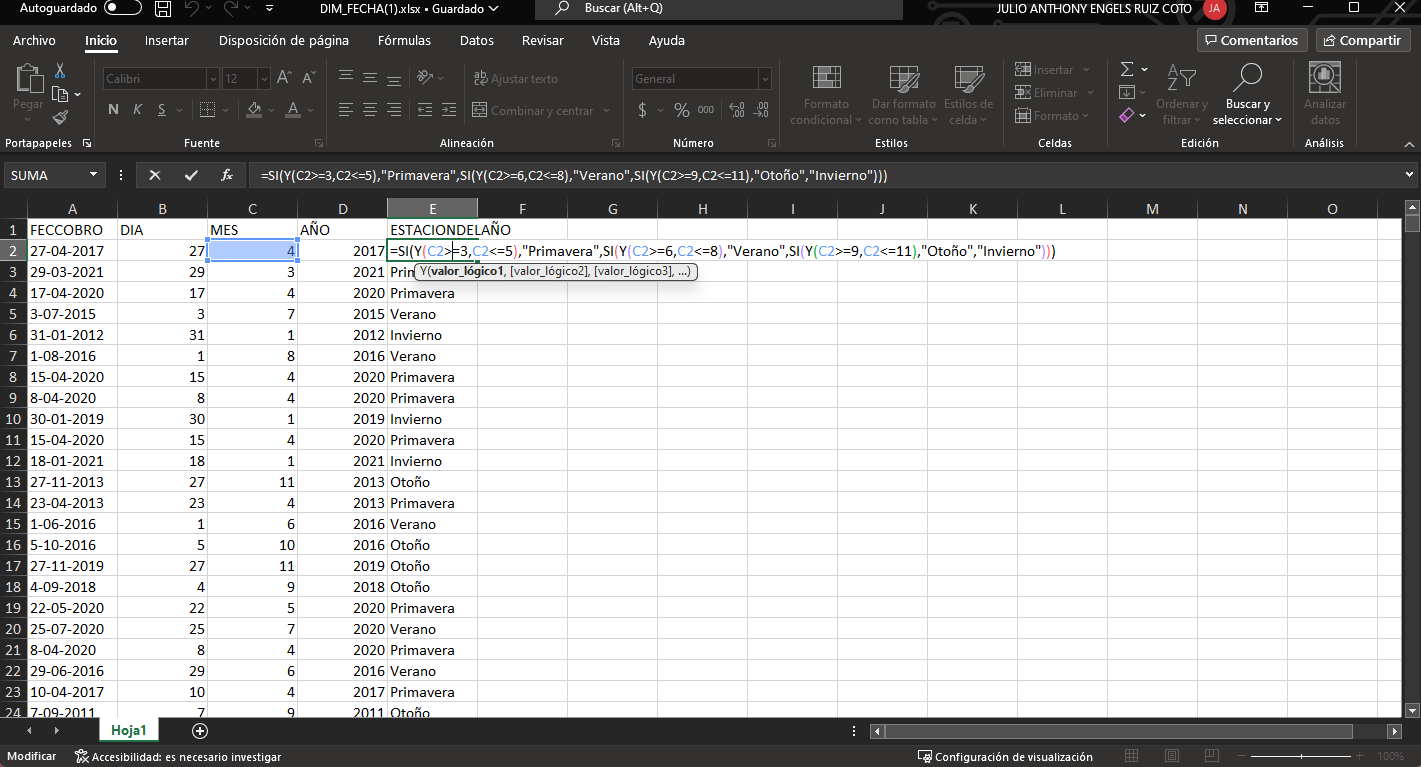
Luego de visualizar el dataset se percata que la columna HORCOBRO también hay que convertirla en formato de horas “00:00:00” en este caso para lo que también se realiza un función la cual lo convierta los números de 5 dígitos o 6 dígitos, la expresión evalúa la longitud del campo HORCOBRO y dependiendo de si tiene 5 o 6 dígitos agrega un 0 si es necesario a los de 5 dígitos caso que no pasa con los de 6 dígitos , luego separa los numero en horas, minutos y segundos.



Uno de los atributos para la DIM\_HORA es dada una hora identificar en qué periodo del día si hizo dicho pago por lo cual se realizo un campo calculado para realizar dicha función, la función clasifica el numero entero de la columna HORA en función de los parámetros que asumí: madrugada de 12am a 5am, mañana de 6am a 11am, tarde de 12pm a 17pm y noche de 18pm a 23pm, y si no cumple con estas condiciones devuelve un error.



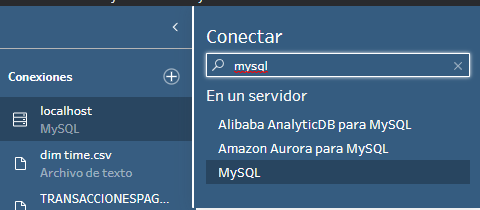
Uno de los atributos para la DIM\_FECHA es dada una fecha identificar en que estación del año se realizan más pagos por lo cual se realizó un campo calculado para realizar dicha función, la función clasifica el numero entero de la columna HORA en función de los parámetros basado en estaciones del hemisferio norte: entre los meses 3 a 5 marzo a mayo: primavera, meses 6 a 8 junio a agosto: verano, meses 9 a 11 septiembre a noviembre: otoño, y para los meses que quedan 12, 1, 2 diciembre, enero, febrero: invierno y si no cumple con estas condiciones devuelve error.



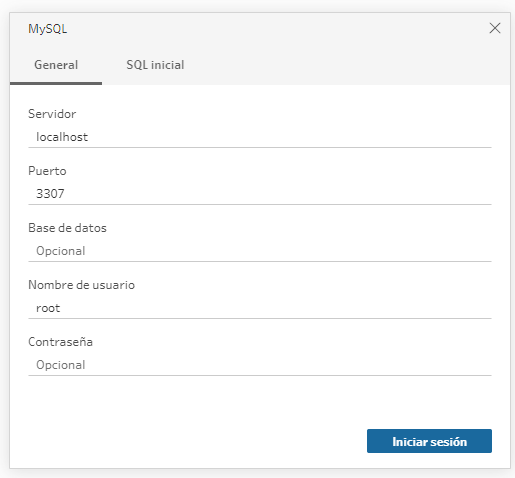
Para el CSV de la dimensión de fecha en la columna de ESTACIONDELAÑO se aplico la funcion SI, luego se aplica la funcion “y” es para controlar si esta entre esa estacion del año que son primavera, verano,otoño, si no cumple con ninguna de las 3 estaciones se valida que esta en invierno. Los rangos de los meses fueron asumidos no corresponden a a niungun tipo de hemisferios.



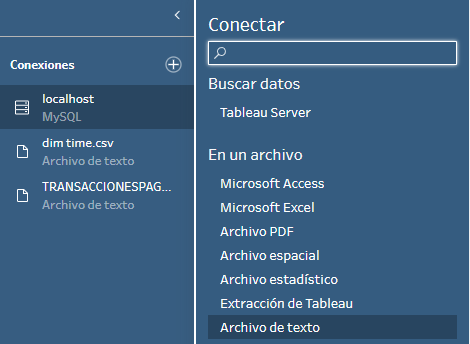
Una vez cargado Tableu Prep nos dirigimos con el cursor en el apartado de conexiones .



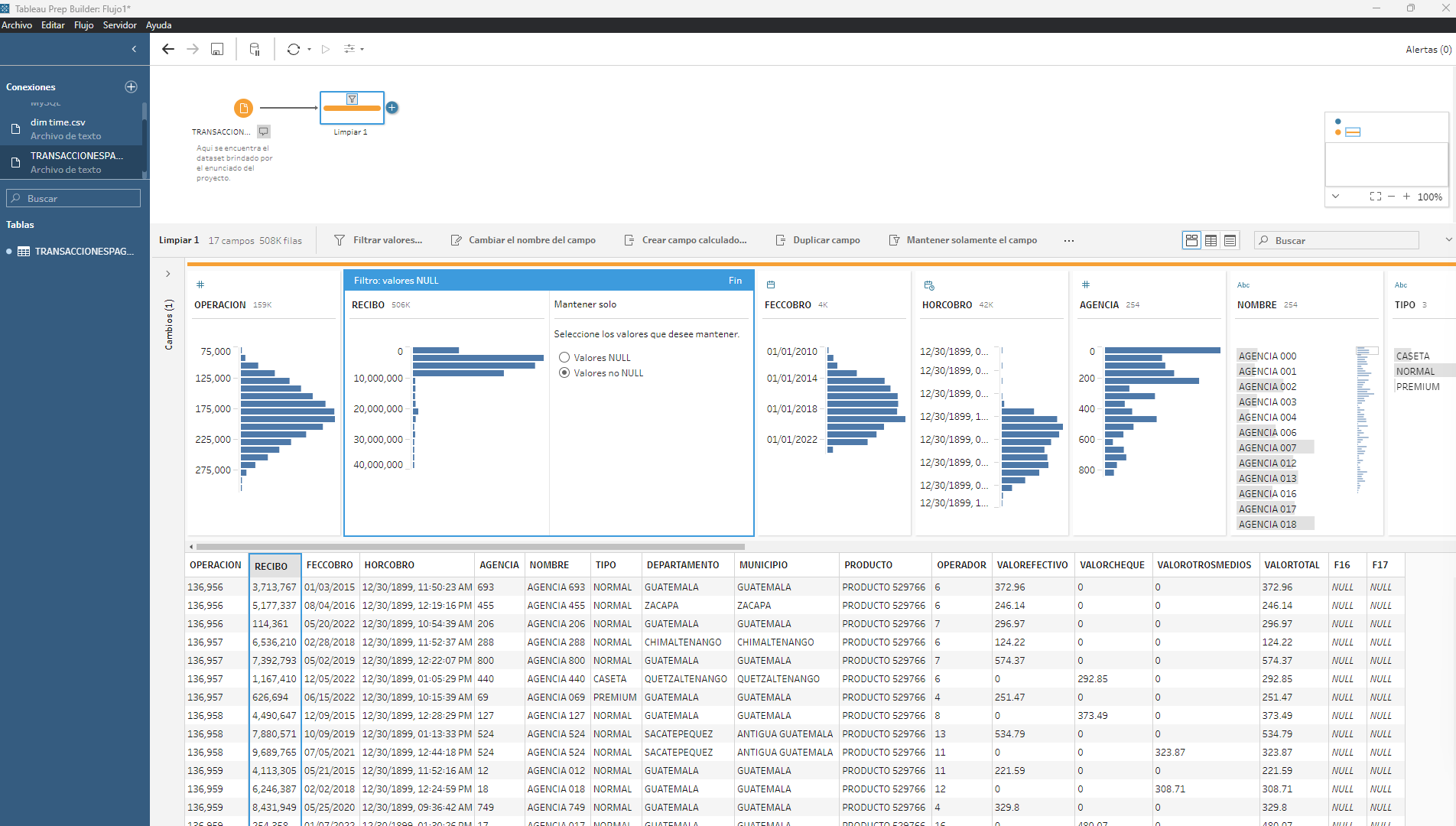
En el buscador colocamos mysql damos click en la opción de MySQL.



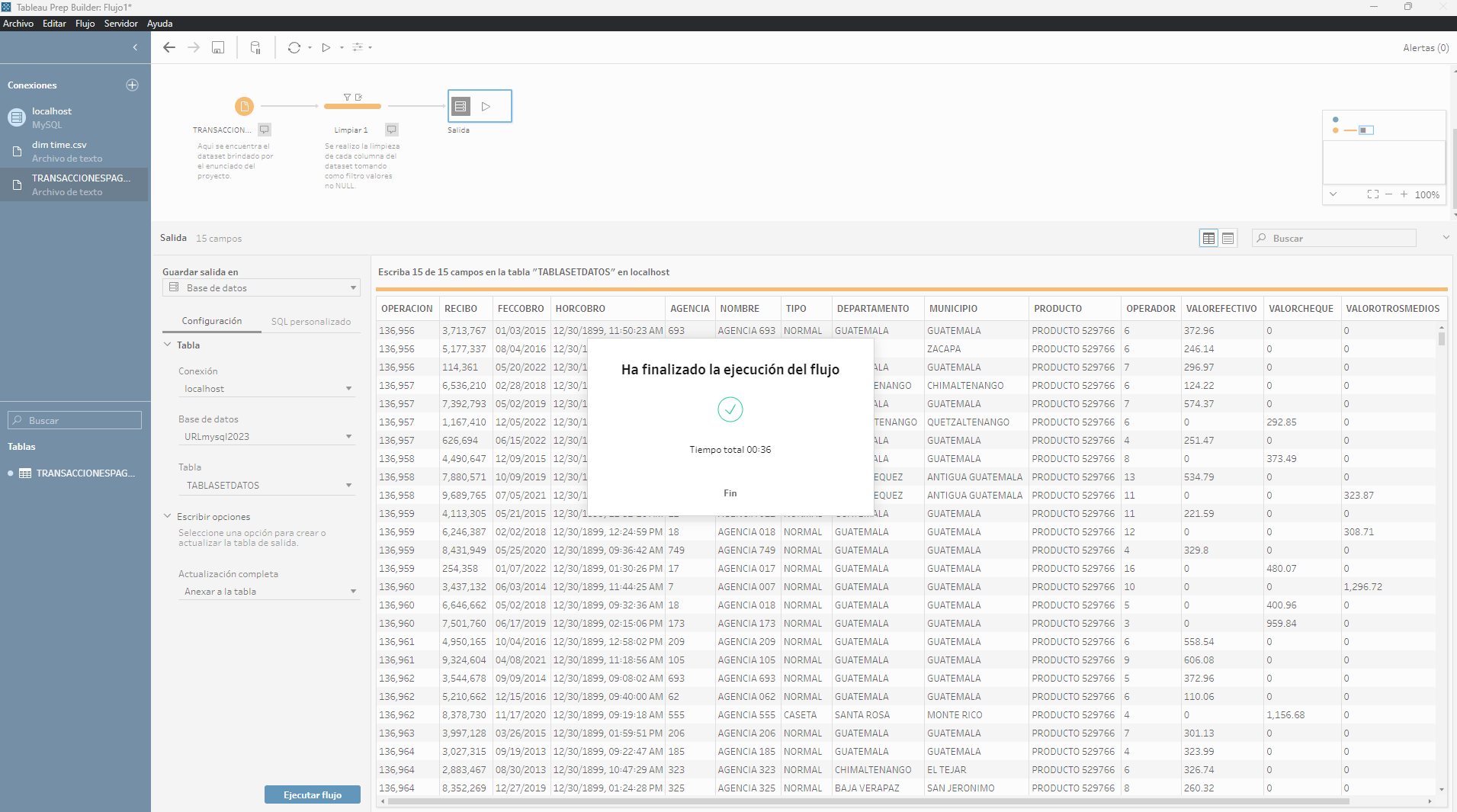
Luego se mostrará la siguiente ventana para colocar las credenciales de la instancia de la base de datos en este caso MySQL.



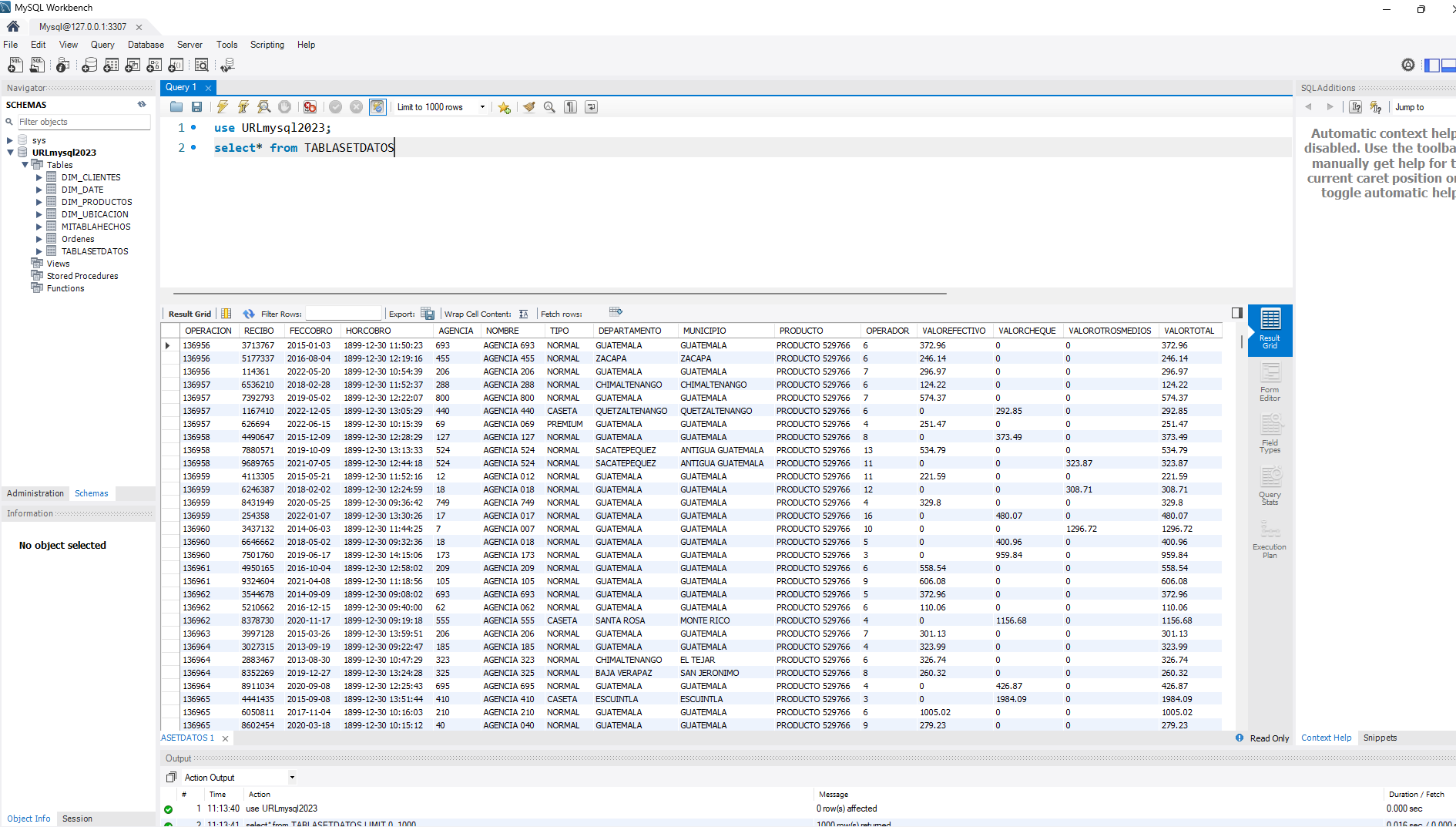
Una vez conectado nuestra instancia, agregamos otra conexión para agregar el CSV del dataset brindado por la práctica. Se realiza el mismo proceso para agregar la dim time.csv.



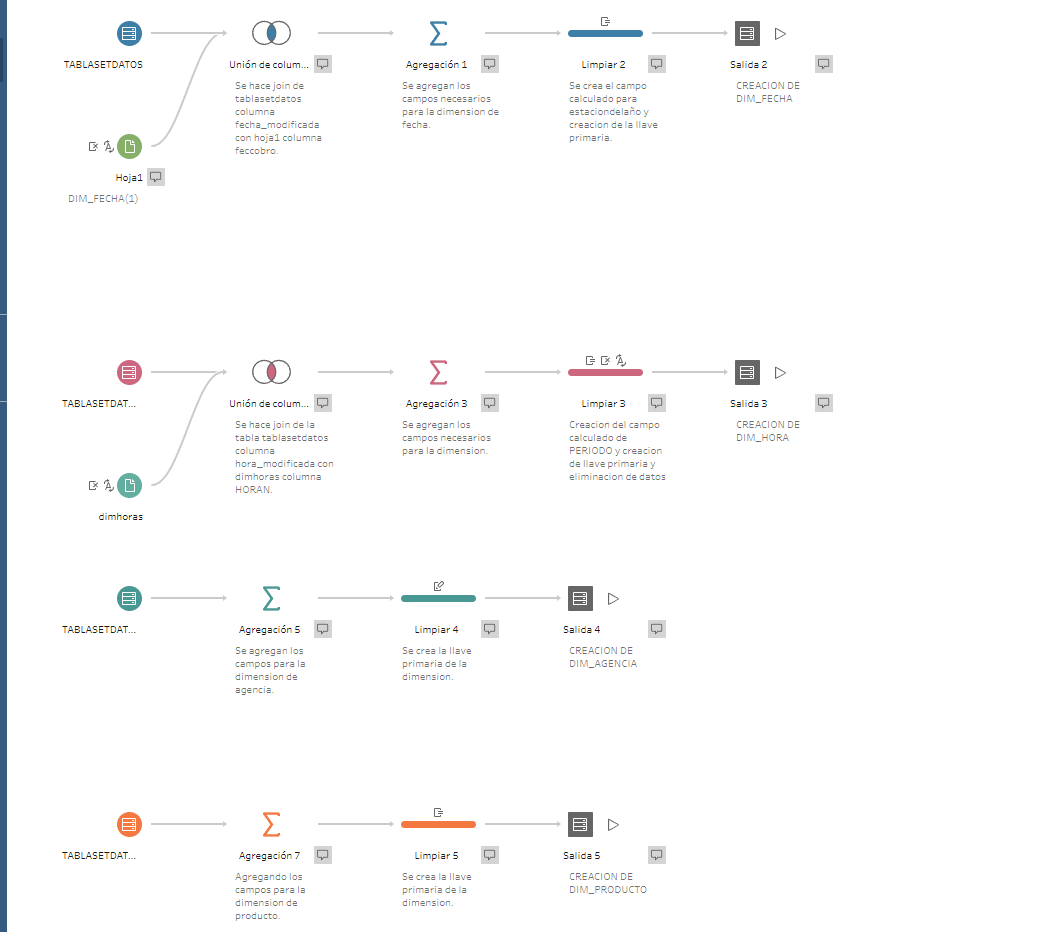
Luego arrastramos nuestro dataset a la hoja vacia del Tableau Prep para proceder con hacer la limpieza de datos se tomo encuenta el filtro de valores NO NULL para cada una de las columnas.



Se agrega una salida de datos en este caso se elige base de datos, hace referencia a la conexión de la instancia de base de datos, se selecciona la base de datos la cual se va a trabajar, luego se crea la tabla en este caso TABLASETDATOS, anexar a la tabla y ejecutar flujo se espera unos minutos hasta que haya finalizado la ejecución del flujo.



Una vez terminado la ejecución del flujo anterior abrimos MySQL Workbench y colocamos las presentes dos líneas para generar la consulta y comprobar que la tabla TABLASETDATOS se haya transformado correctamente.



En la siguiente imagen se muestra la creación de las dimensiones, para crear las dimensiones de fecha y de hora se aplicó joins, para el de fecha se hizo join con Hoja1 (DIM\_FECHA(1)) y TABLASETDATOS las columnas de FECHA\_MODIFICADA. Mismo proceso para la dimensión de hora se realizó join con dimhoras (CSV) brindado por el catedrático y TABLASETDATOS las columnas de HORA\_MOFICADA, para los dos casos se aplicaron agregaciones para ir identificando los atributos que tendrán las dimensiones así mismo se crearon campos calculados en para DIM\_FECHA ESTACIONESDELAÑO y para DIM\_HORA PERIODO, también la creación de las llaves primarias de las dimensiones, y sus respectivas salidas para crear la dimensión. Con respecto a las demás dimensiones, agencia y producto se aplicó una agregación en ambos casos para agregar los campos respectivos de dichas tablas con esto se aplico el paso de limpieza el cual también para crear las llaves primarias y con ello se procede a crear las tablas DIM\_AGENCIA y DIM\_PRODUCTO. En los paso de limpieza no aplique filtro de valores NO NULL ya que cuando termine de hacer mi tabla de hechos se borraban datos de mi dataset.

Para sacar la tabla de hechos la FACT\_PAGOS se realizaron 4 joins con la tabla de TABLASETDATOS, el primer join consiste en unir FECHA\_MODIFICADO con FECHA\_MODIFICADO de la dimensión de fecha, para el segundo join consiste en unir HORA\_MODIFICADA con HORA\_MODIFICADA de la dimensión de hora, para el tercer join se realizo el join de las columnas MUNICIPIO, DEPARTAMENTO, TIPO, NOMBRE, OPERADOR, AGENCIA con DEPARTAMENTO, TIPO, NOMBRE, OPERADOR, KEY\_AGENCIA de la dimensión de agencia, para el ultimo join consistió en unir PRODUCTO con PRODUCTO columna de la dimensión producto, después se agrego un paso de limpieza que consiste en eliminar columnas para solo dejar variables cuantitativas para la tabla de hechos, después se agrego un paso de agregación el cual consiste en agrupar los datos group y datos sum en los dos tipos de campos que hay, para este caso los datos group van en campos agrupados y los datos sum van en campos agregados. Por último, se agregó una salida y se creo la tabla de hechos llamada FACT\_PAGOS.

