

PROYECTO 1

CUBOS DE

INFORMACIÓN

Análisis de Datos



Guatemala 21 de Marzo

JULIO ANTHONY ENGELS RUIZ COTO -
1284719

DISEÑO DIMENSIONAL

Modelo Dimensional identificar los hechos y las dimensiones:

Hechos: son eventos medibles.
Key_operacion, Key_Agencia ,Key_Producto, Key_Fecha, Key_Hora, ValorEfectivo, ValorCheque, ValorOtrosMedios, ValorTotal.

- 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.**

Para la tabla de fecha se agrego una columna de Estaciondel año para saber en que estacion del año se realizan mas cobros.

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.

Para la tabla de hora se agrego una columna de Periodo para saber en que periodo de atencion se realizan los cobros.

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.

El campo de Key_Operacion asumi en juntar la columna OPERACION con la columna de RECIBO para hacer mi llave primaria de la tabla de hechos.

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.

Diseño Dimensional:



DISEÑO ENTIDAD RELACIÓN

Las presentes imagenes muestran el diseño a nivel de la base de datos indicando los campos y llaves creadas.

```
DMLPROYECTOsql.sql...ANTOS\julio (52)) Desktop_Antos.AN...DATOS - Dia
|
CREATE TABLE FECHA (
  Key_Fecha int,
  Fecha DATE,
  Dia int not null,
  Mes int not null,
  Año int not null,
  Estaciondel año varchar(10)

  constraint PK_FECHA primary key (Key_Fecha)
)
DROP TABLE FECHA

CREATE TABLE HORA (
  Key_hora int,
  Hora TIME,
  Horas int not null,
  Minutos int not null,
  Segundos int not null,
  Periodo varchar(50),
  HoraIngreso char(1),
  Horacierreagencia char(1),
  Horaalmuerzo char(1),
  Horasalida char(1),
  Horalaborable char(1)

  constraint PK_HORA primary key (Key_hora)
)
DROP TABLE HORA

CREATE TABLE AGENCIA (
  Key_Agencia int,
  Agencia_nombre varchar(255),
  Agencia_tipo varchar(50),
  Departamento varchar(50),
  Municipio varchar(50),
  Codigo_operador int

  constraint PK_AGENCIA primary key (Key_Agencia)
)
DROP TABLE AGENCIA

CREATE TABLE PRODUCTO(
  Key_Producto int,
  Producto_descripcion varchar(255)

  constraint PK_Producto primary key (Key_Producto)
)
DROP TABLE PRODUCTO
```



```
CREATE TABLE TABLAHECHOS(  
Key_operacion int,  
Key_Agencia int,  
Key_Producto int,  
Key_Fecha int,  
Key_hora int,  
ValorEfectivo decimal(10,2),  
ValorCheque decimal(10,2),  
ValorOtrosMedios decimal(10,2),  
ValorTotal decimal(10,2)  
  
constraint Key_operacion primary key (Key_operacion),  
constraint FK_Agencias foreign key (Key_Agencia) references AGENCIA(Key_Agencia),  
constraint FK_Producto foreign key (Key_Producto) references PRODUCTO(Key_Producto),  
constraint FK_Fecha foreign key (Key_Fecha) references FECHA(Key_Fecha),  
constraint FK_HORA foreign key (Key_hora) references HORA(Key_hora)  
  
)  
DROP TABLE TABLAHECHOS
```

En la presente imagen se observa el DML de la creación de las tablas, TABLAHECHOS corresponde a FACT_PAGOS.

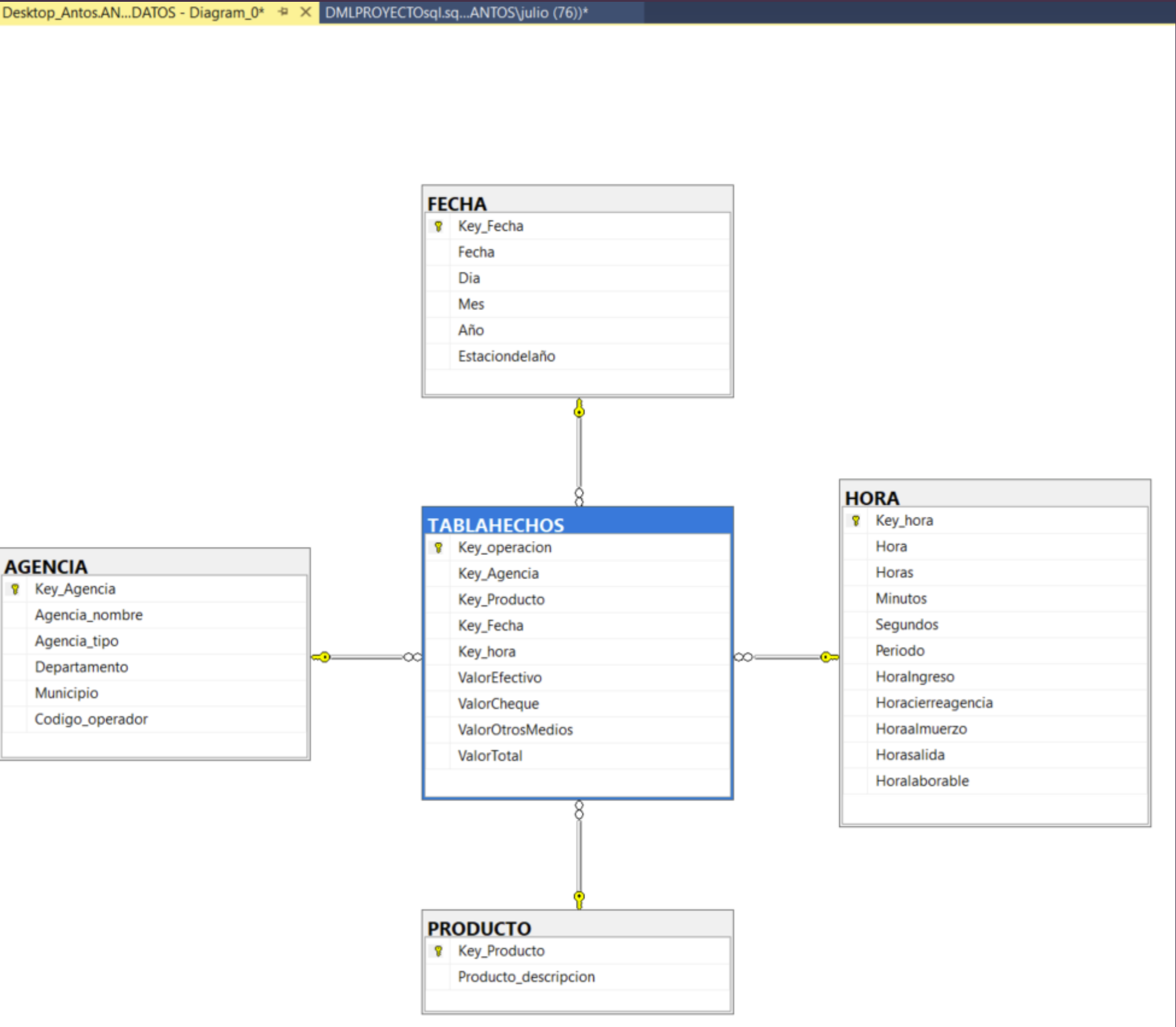
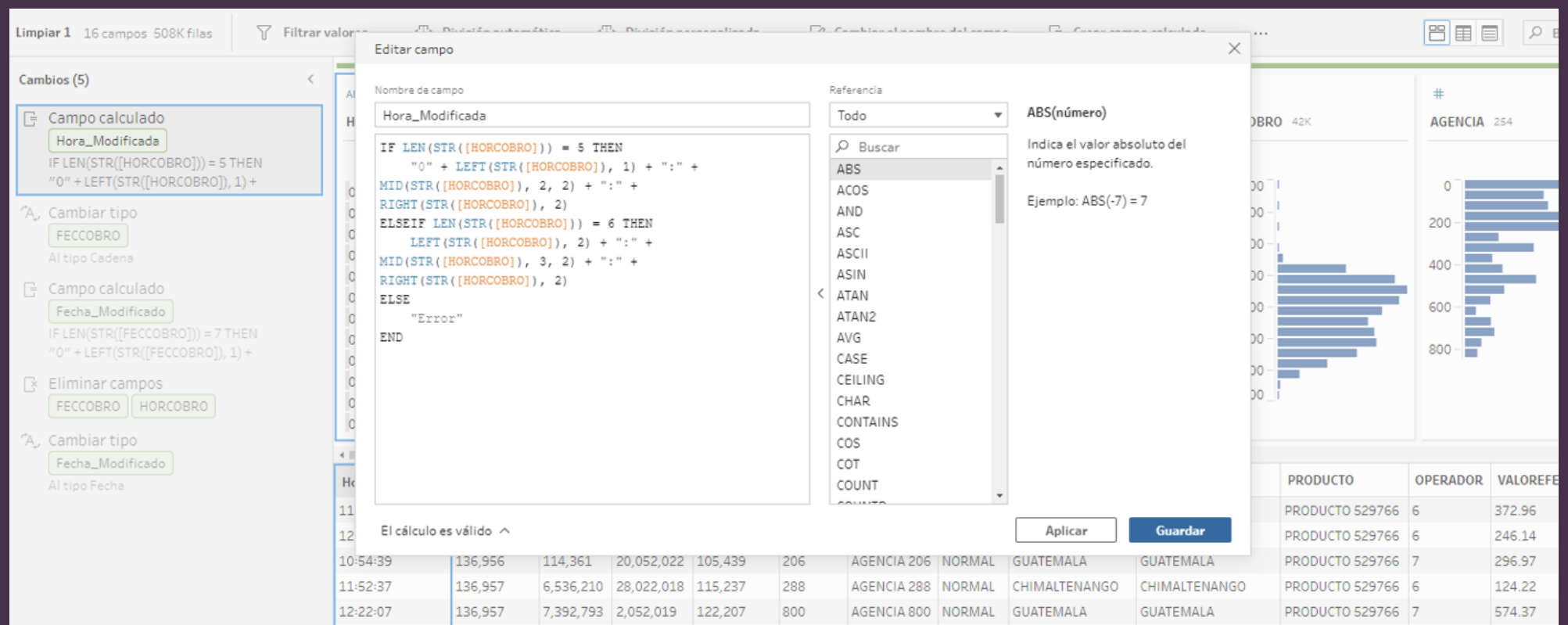
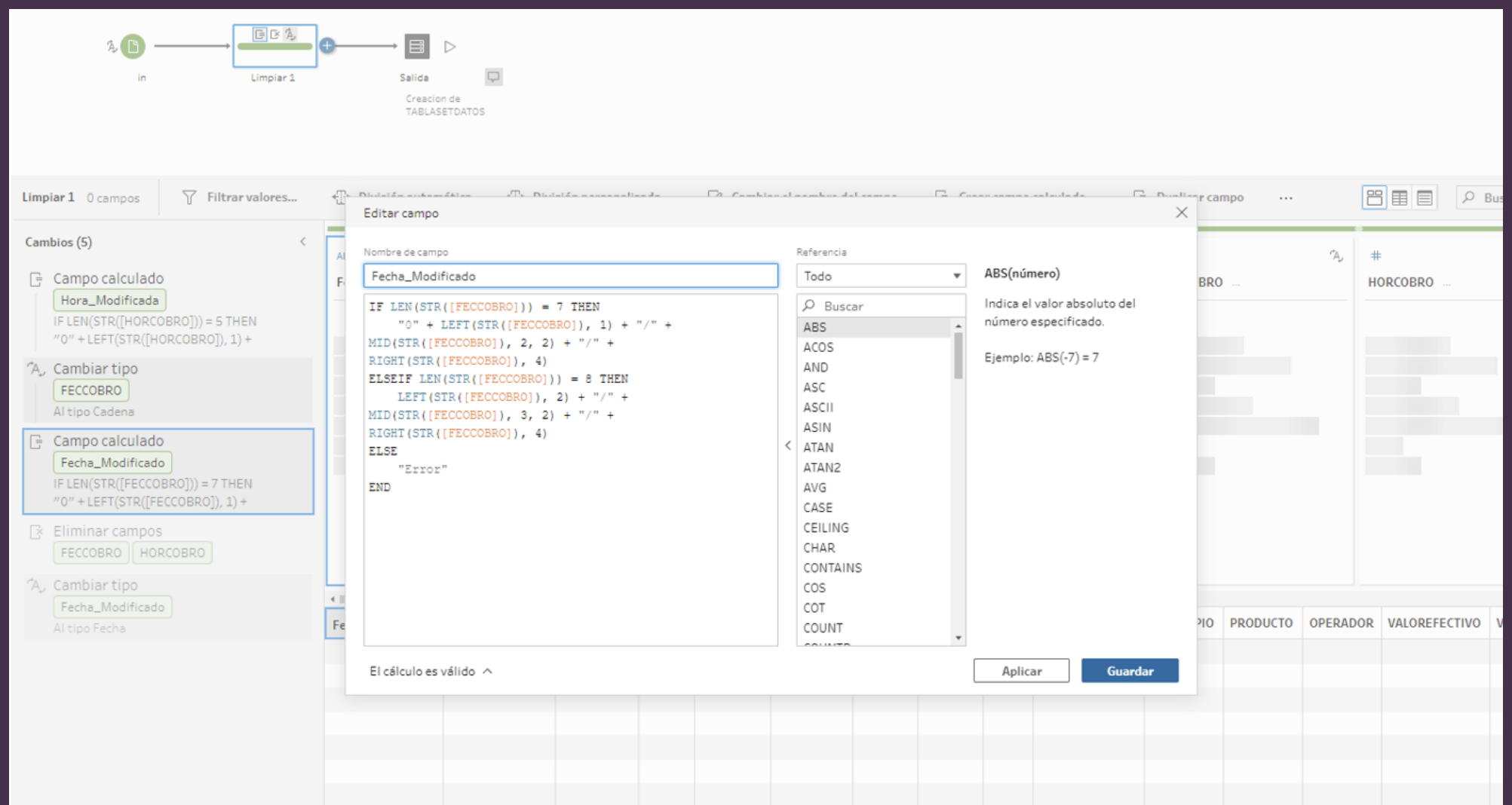


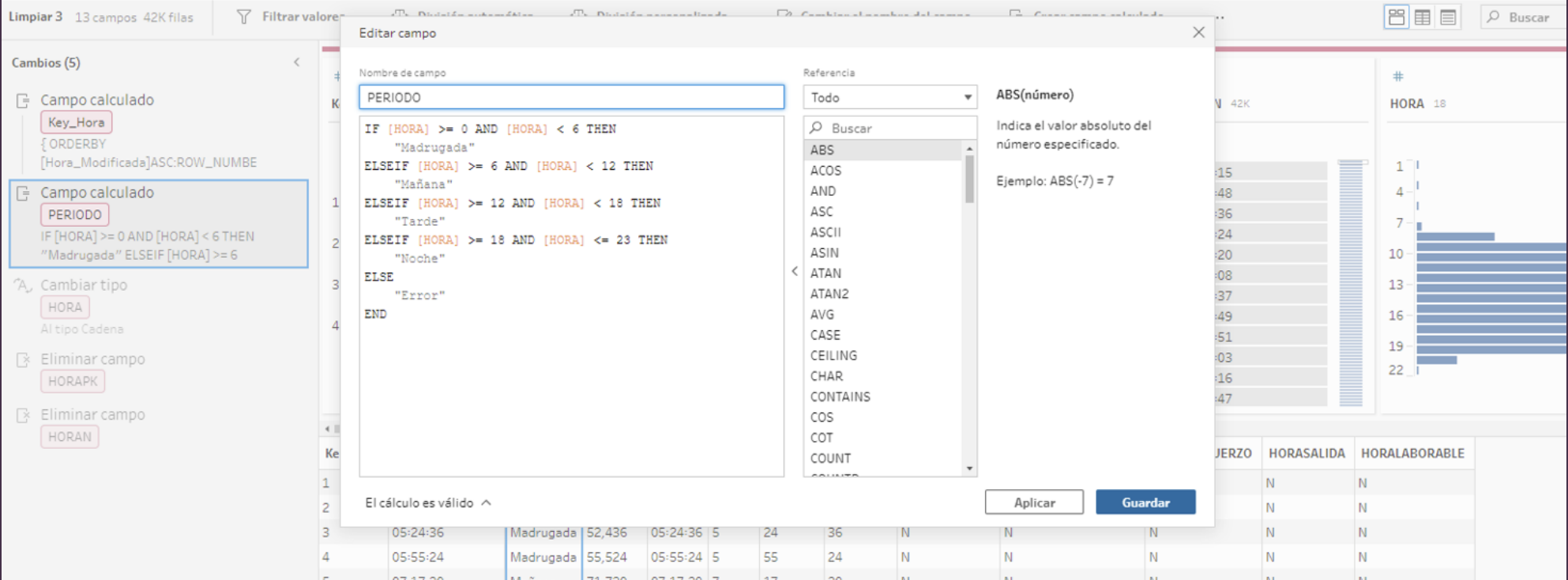
TABLEAU PREP

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 colocó de nombre Fecha_Modificada y se procedió a crear la expresión, esta evalúa la longitud del número en el campo de FECCOBRO y dependiendo de si tiene 7 u 8 dígitos, se le agregará 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 inglesa (MM/DD/YYYY).

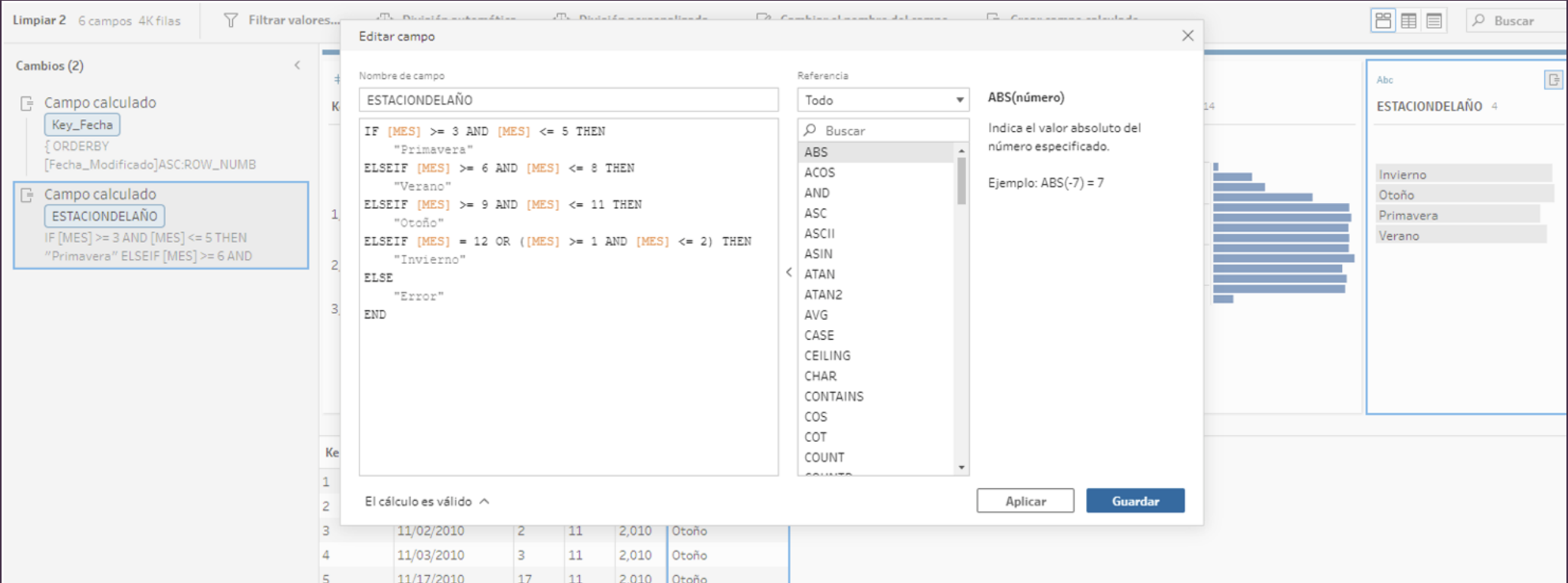


Luego de visualizar el dataset se percata que la columna HORCOBRO también hay que convertirla en formato de horas “00:00:00” en este caso, también se realiza un función la cual convierta los números de 5 dígitos o 6 dígitos, la expresión evalúa la longitud del campo HORCOBRO y dependiendo de si tiene 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 número en horas, minutos y segundos.

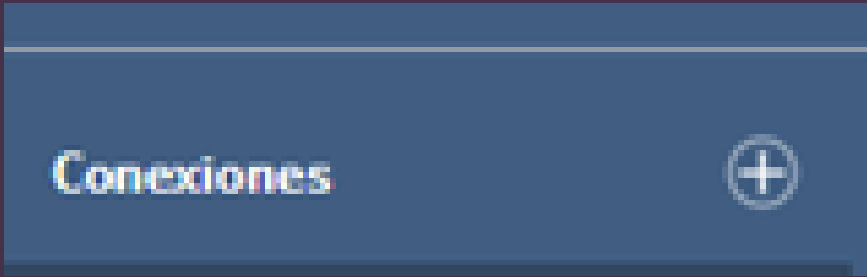
Uno de los atributos para la DIM_HORA es dada una hora identificar en qué periodo del día si hizo dicho pago por lo cual se realizo un campo calculado para realizar dicha función, la función clasifica el numero entero de la columna HORA en función de los parámetros que asumí: madrugada de 12am a 5am, 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.

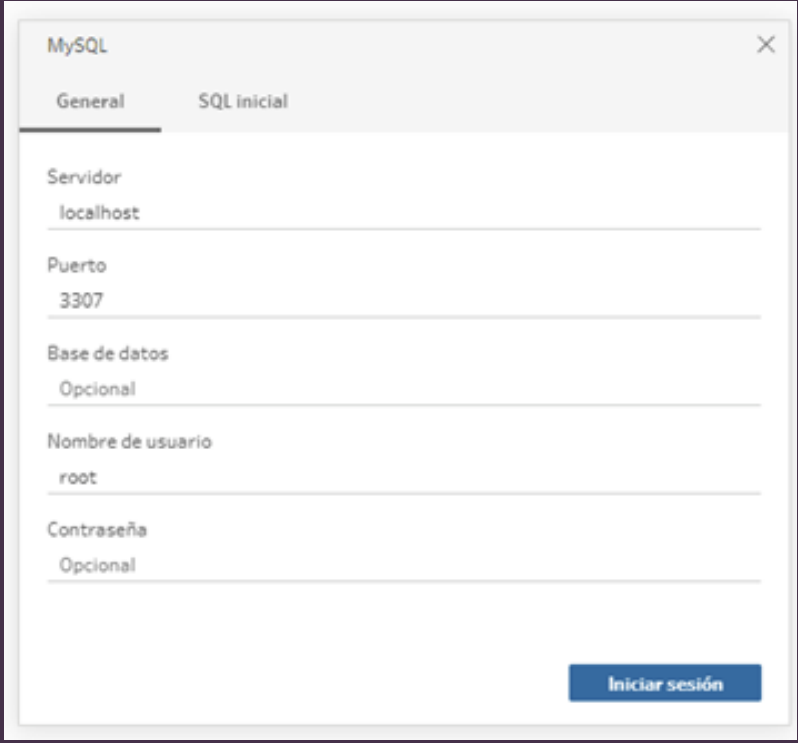
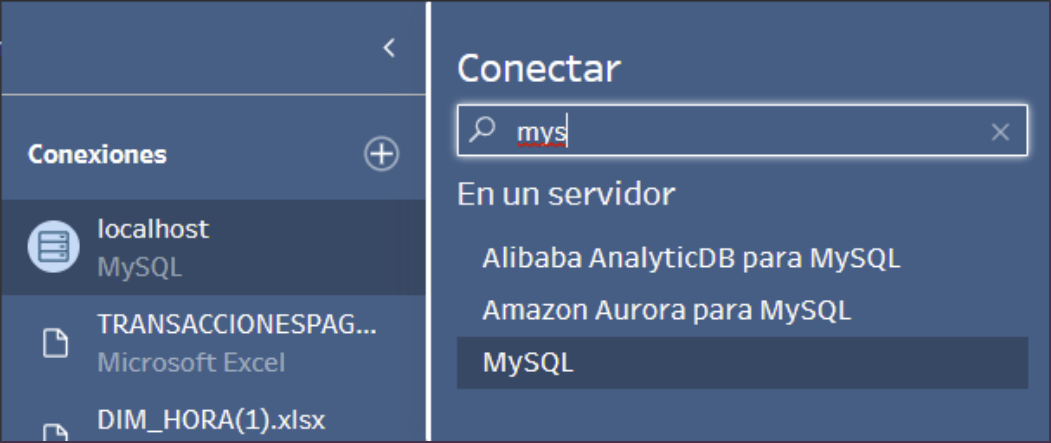


Una vez arreglado el dataset se procede a la herramienta de Tableau Prep:



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

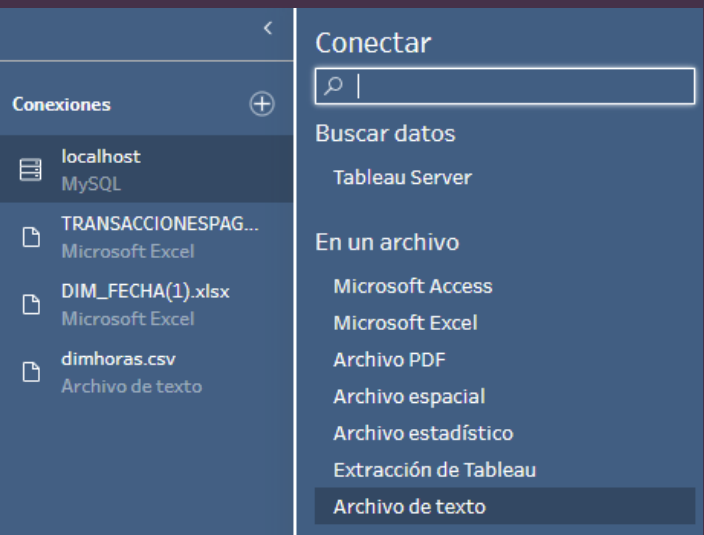
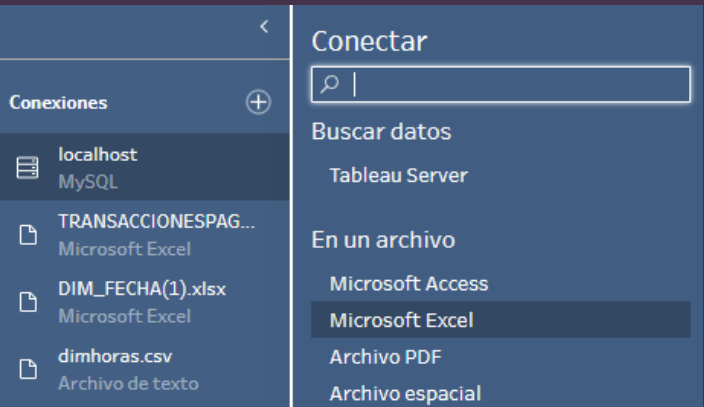
En el buscador colocamos mysql damos click en la opción de MySQL.



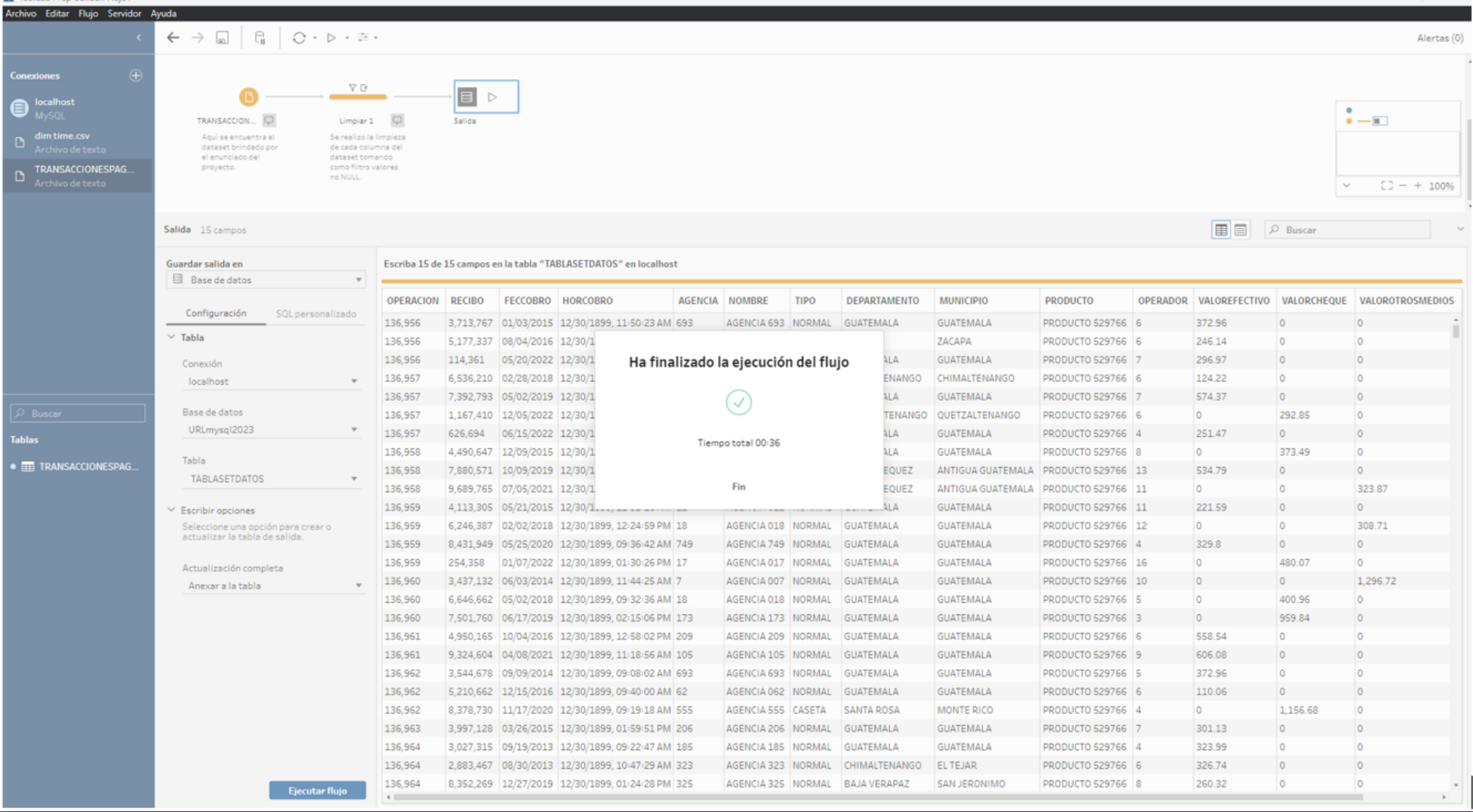
Luego se mostrará la siguiente ventana para colocar las credenciales de la instancia de la base de datos en este caso MySQL.



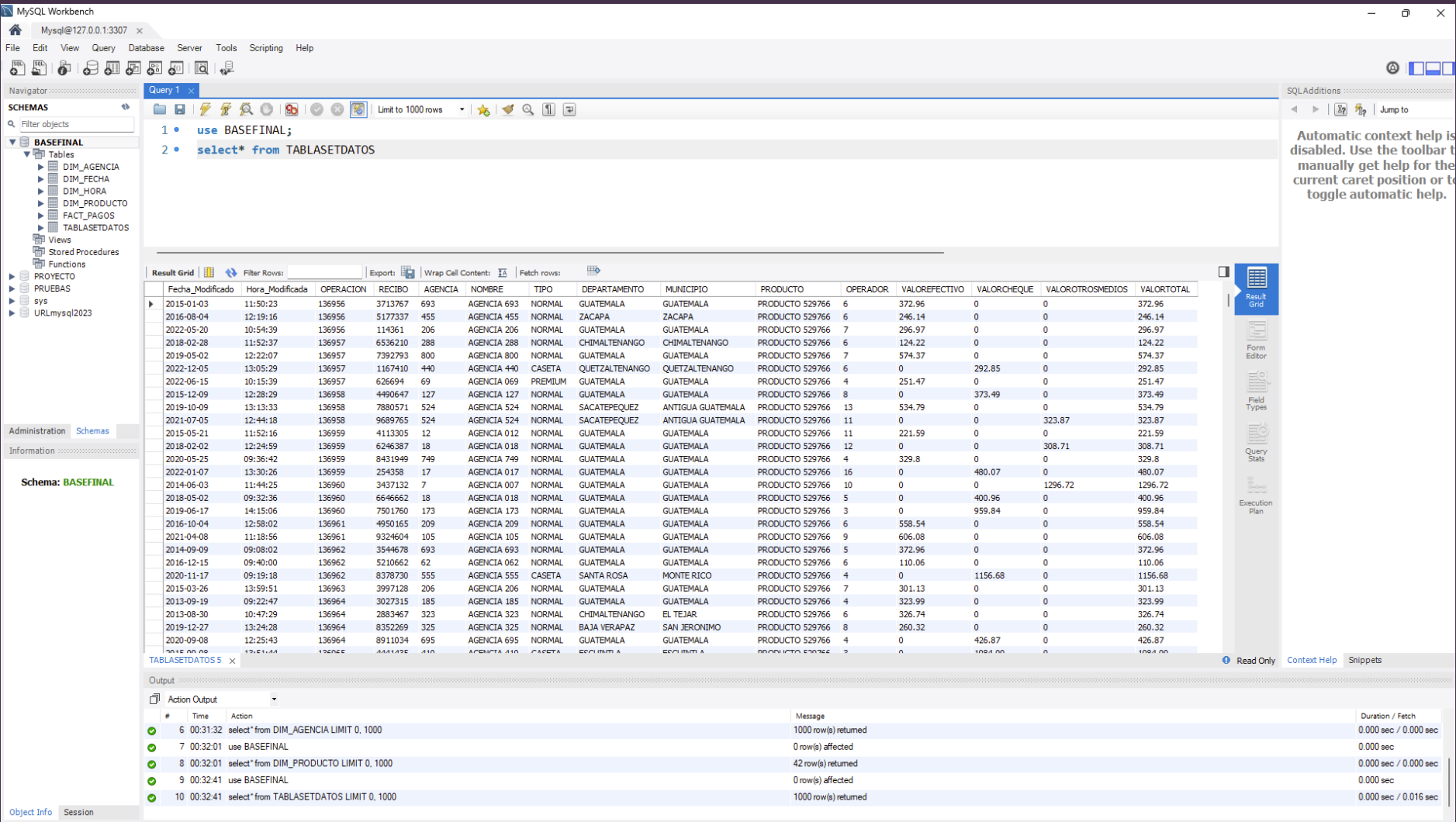
Una vez conectado nuestra instancia, agregamos otra conexión para agregar el archivo **TRANSACCIONESPAGOS.xlsx** del dataset brindado por la práctica. Se realiza el mismo proceso para agregar **DIM_FECHA(1).xlsx** y el **dimhoras.csv**



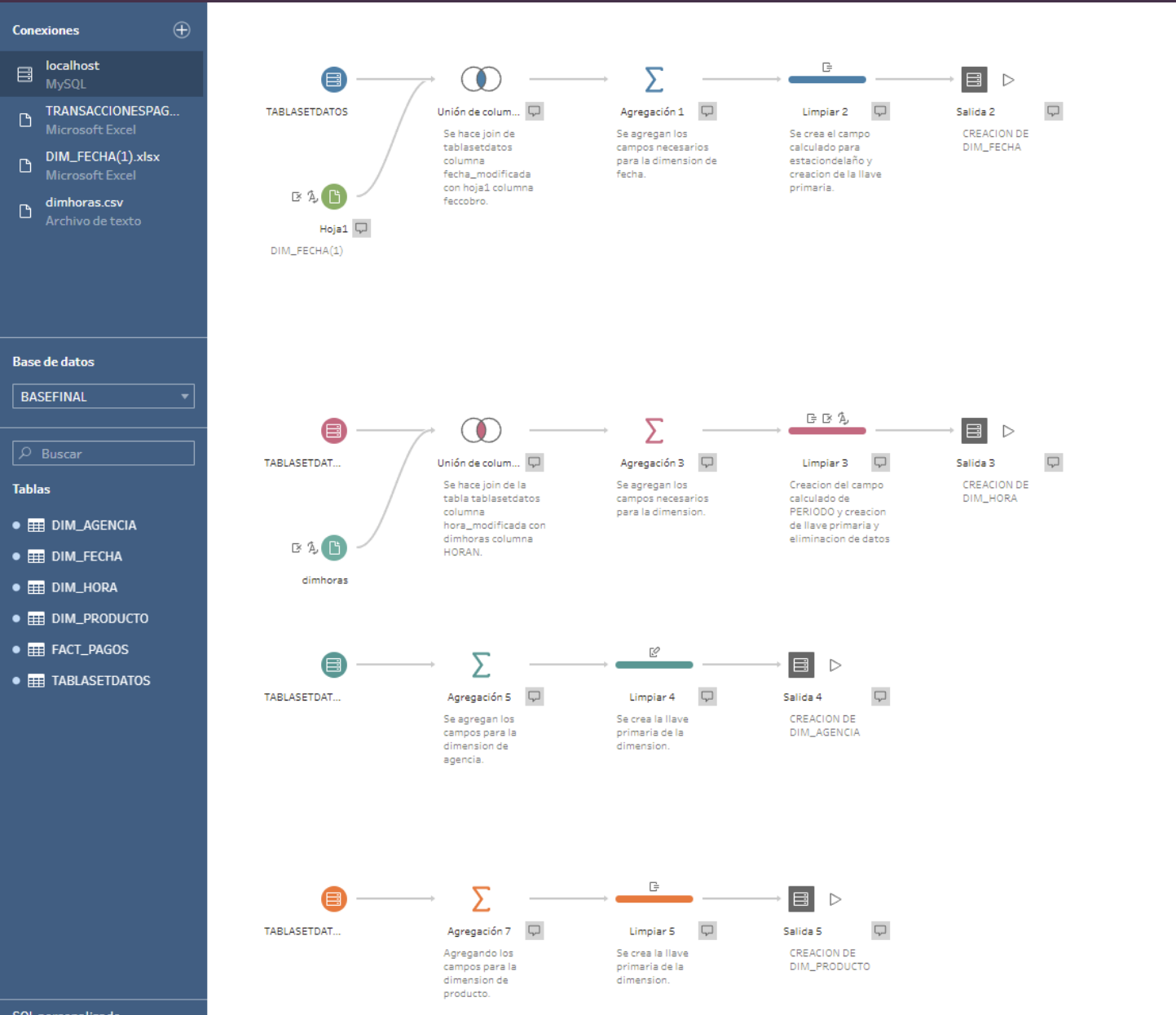
Luego arrastramos nuestro dataset a la hoja vacía del Tableau Prep para proceder con el análisis.



Se agrega un paso de limpieza y 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.



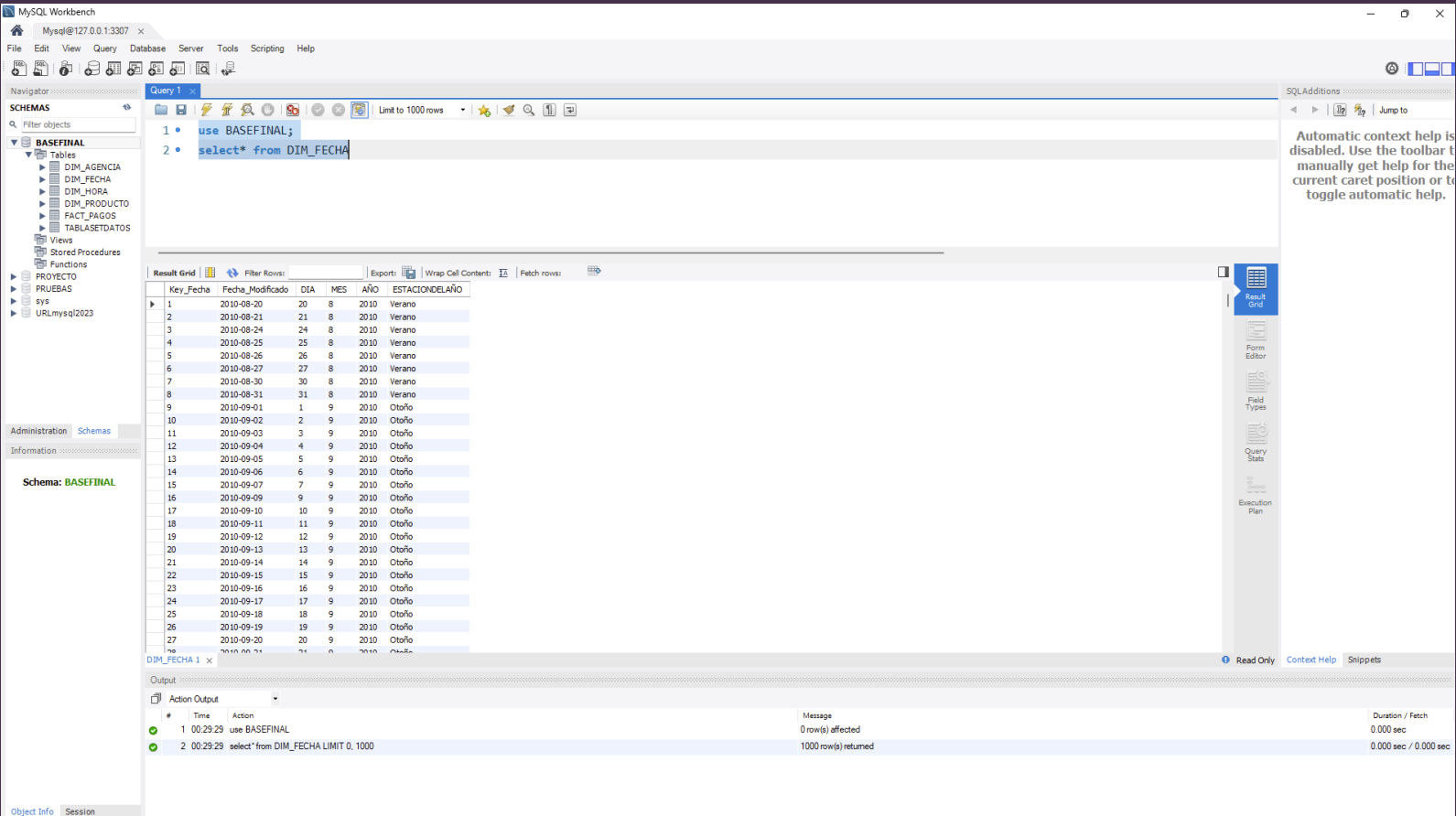
Una vez terminado la ejecución del flujo anterior abrimos MySQL Workbench y colocamos las presentes dos líneas para generar la consulta y comprobar que la tabla TABLASETDATOS se haya transformado correctamente.



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 borrarán datos de mi dataset.

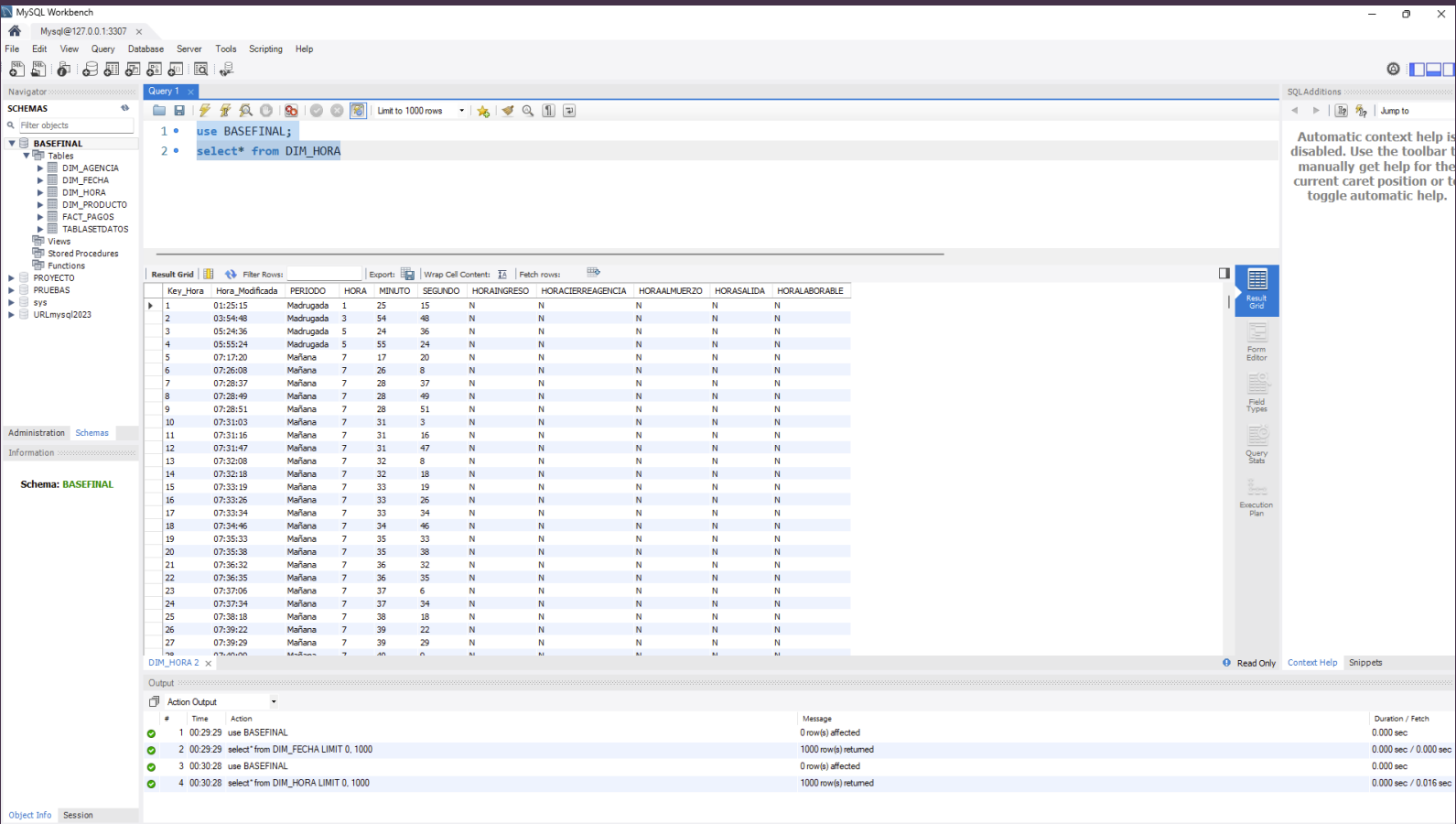
A nivel de base de datos a continuación se presenta la visualización de las dimensiones:

DIM_FECHA:

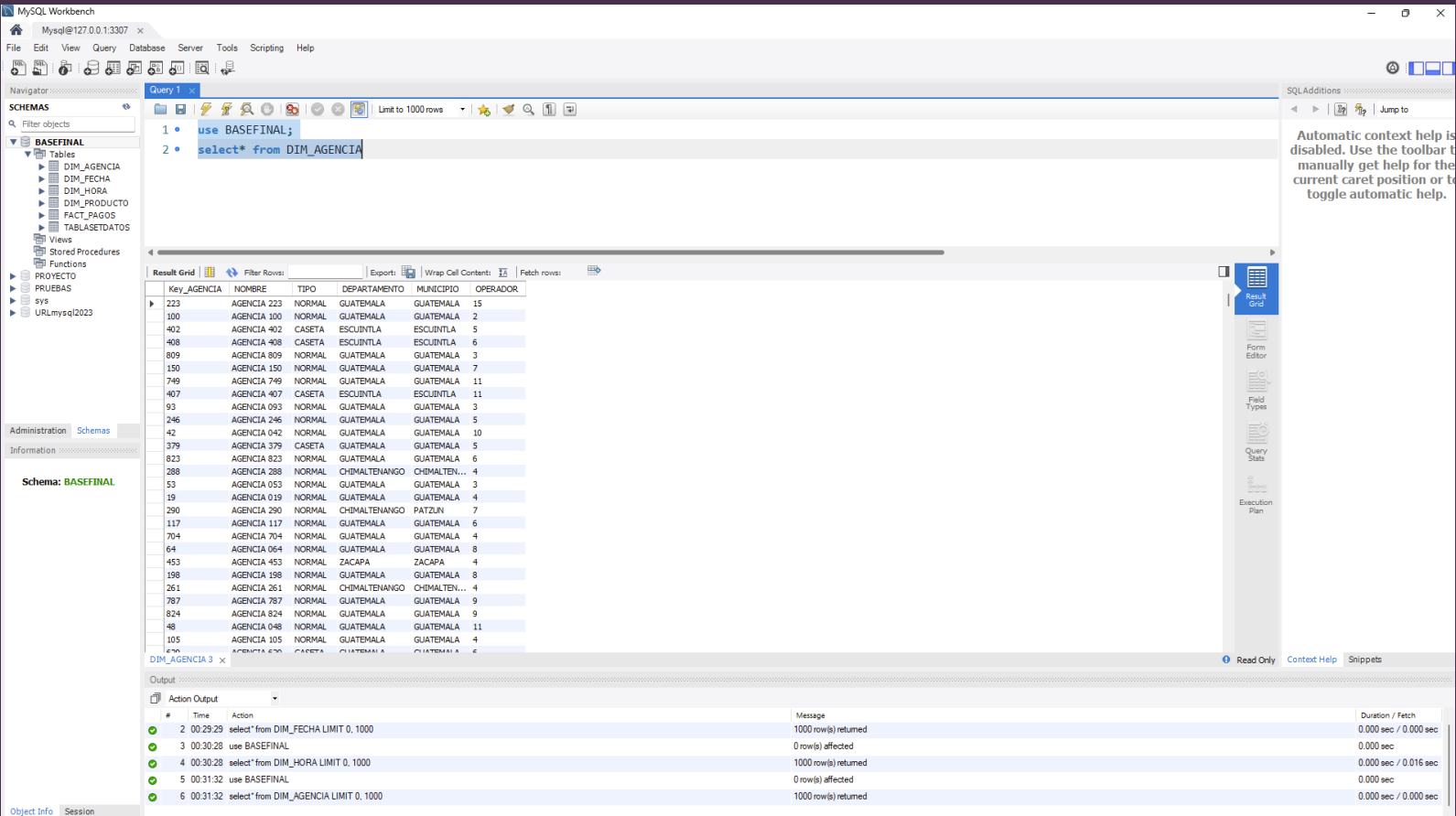


DIM_HORA:

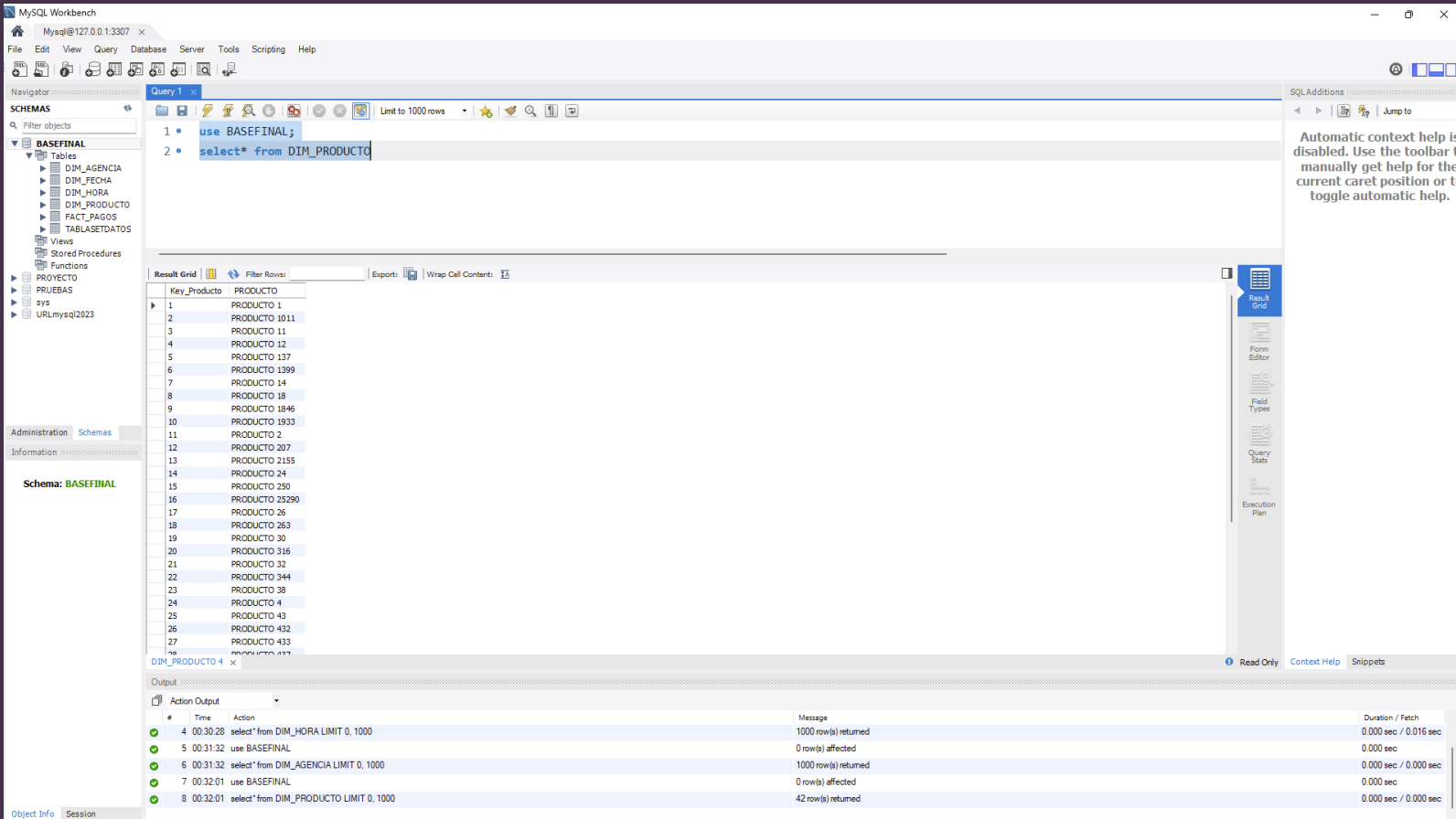
Para esta dimensión se utilizo la dimensión dim time.csv brindada por el catedrático y se aplico un join a la tabla TABLASETDATOS.



DIM_AGENCIA:



DIM_PRODUCTO:



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.



TABLASETDATOS

SQL personalizado



TABLA HECHOS FACT_PAGOS:

MySQL Workbench

Mysql@127.0.0.1:3307

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

BASEFINAL

Tables

DIM_AGENCIA

DIM_FECHA

DIM_HORA

DIM_PRODUCTO

FACT_PAGOS

TABLASETDATOS

Views

Stored Procedures

Functions

PROYECTO

PRUEBAS

sys

URLmysql2023

Administration

Schemas

Information

Schema: BASEFINAL

Object Info

Session

Query:1

Limit to 1000 rows

1 use BASEFINAL;

2 select* from FACT_PAGOS

3

4 Select count(*) from FACT_PAGOS

Result Grid

Filter Rows:

Exports

Wrap Cell Content:

Fetch rows:

PK_OPERACION	Key_AGENCIA	Key_Producto	Key_Fecha	Key_Hora	VALOREFECTIVO	VALORCHEQUE	VALOROTROSMEIOS	VALORTOTAL
326856	40	31	348	7667	0	0	289.92	289.92
432085	65	31	1720	9215	224.42	0	0	224.42
221891	127	31	3276	12114	277.37	0	0	277.37
486140	283	31	1222	4810	0	0	420.54	420.54
303177	103	31	1137	12024	167.37	0	0	167.37
267778	206	31	1994	30637	0	0	756.58	756.58
106591	182	33	3946	15117	1052.67	0	0	1052.67
344319	399	31	1320	5226	154.99	0	0	154.99
499932	405	36	2254	5479	950	0	0	950
35227	7	31	2443	34896	0	402.45	0	402.45
91531	410	31	3550	20659	0	0	132.01	132.01
441363	158	31	1254	19786	563.7	0	0	563.7
310076	230	31	2942	4678	242.14	0	0	242.14
112576	183	31	3178	29737	133.23	0	0	133.23
387850	409	31	3397	11082	369.79	0	0	369.79
188158	293	31	3368	33050	505.37	0	0	505.37
483629	655	31	1949	8340	493.87	0	0	493.87
288322	206	31	1322	32790	373.01	0	0	373.01
216891	117	31	3273	24613	347.91	0	0	347.91
488423	12	31	3143	13658	0	557.73	0	557.73
232772	535	31	3301	11823	153.04	0	0	153.04
107327	382	31	3409	16964	0	279.89	0	279.89
164068	293	31	3515	13695	278.04	0	0	278.04
165330	268	31	3899	13268	708.28	0	0	708.28
103256	17	16	3351	16307	0	0	2807.83	2807.83
167910	460	31	3143	39460	797.66	0	0	797.66
340900	270	31	3744	9628	0	561.83	561.83	561.83
320454	706	31	1107	11100	0	0	0	0

FACT_PAGOS 8

Output

Read Only

Context Help

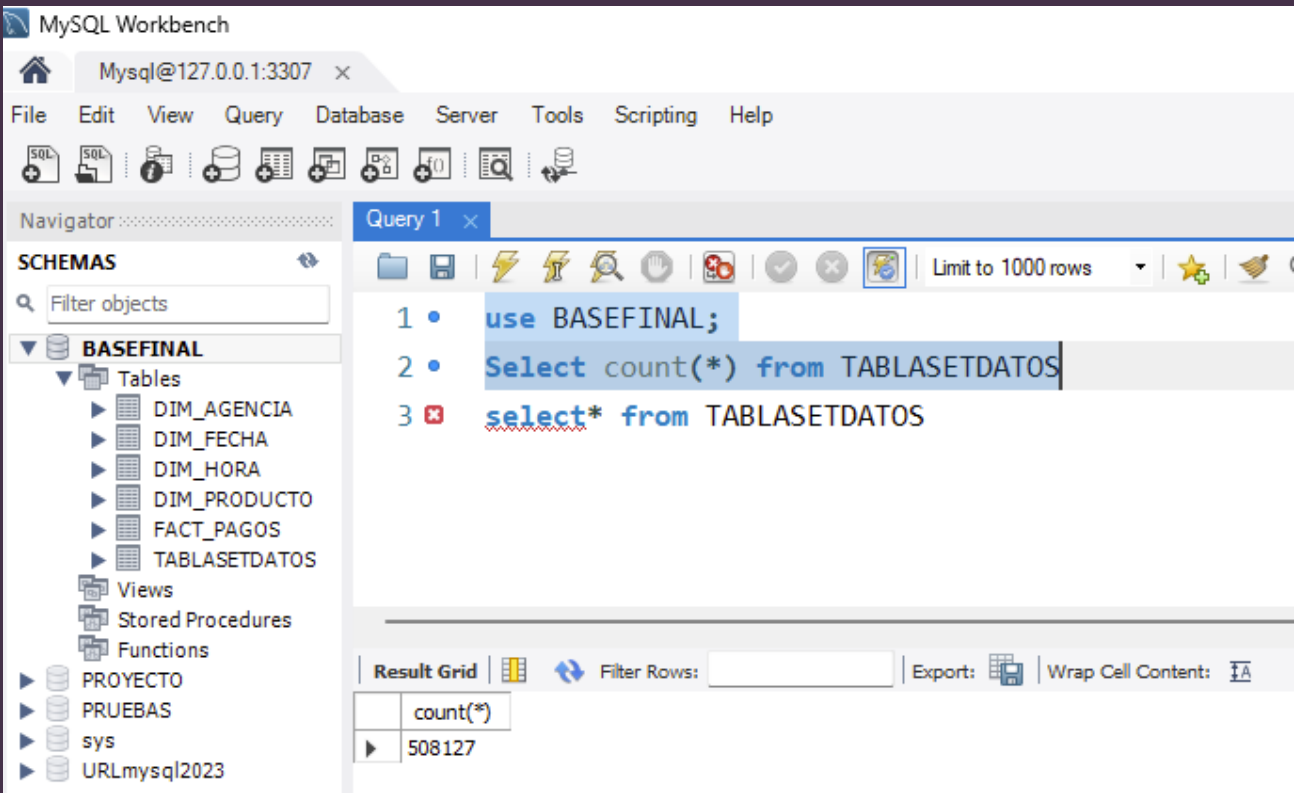
Snippets

SQLAdditions

Automatic context help disabled. Use the toolbar manually get help for the current caret position or toggle automatic help

#	Time	Action	Message	Duration / Fetch
12	00:34:02	Select count(*) from TABLASETDATOS LIMIT 0, 1000	1 row(s) returned	0.094 sec / 0.000 sec
13	00:34:45	use BASEFINAL	0 row(s) affected	0.000 sec
14	00:34:45	Select count(*) from FACT_PAGOS LIMIT 0, 1000	1 row(s) returned	0.047 sec / 0.000 sec
15	00:35:28	use BASEFINAL	0 row(s) affected	0.000 sec
16	00:35:28	select* from FACT_PAGOS LIMIT 0, 1000	1000 row(s) returned	0.015 sec / 0.000 sec

Aplicando un count al dataset para verificación de datos que tiene el dataset.



Aplicando un count al dataset para verificación de datos que tiene la tabla de hechos FACT_PAGOS.

