UDB VIRTUAL

Universidad Don Bosco, El Salvador

Datawarehouse y Minería de Datos

DMD941 G01T

“DESAFÍO 1”

Instructora

Karens Lorena Medrano Mejía

Estudiantes

Nombre	Carné
Oliver Alejandro Erazo Reyes	ER231663
Fernando Alonso Cortez Rivas	RM161936
Kevin Oswaldo Sintigo Merino	SM172332

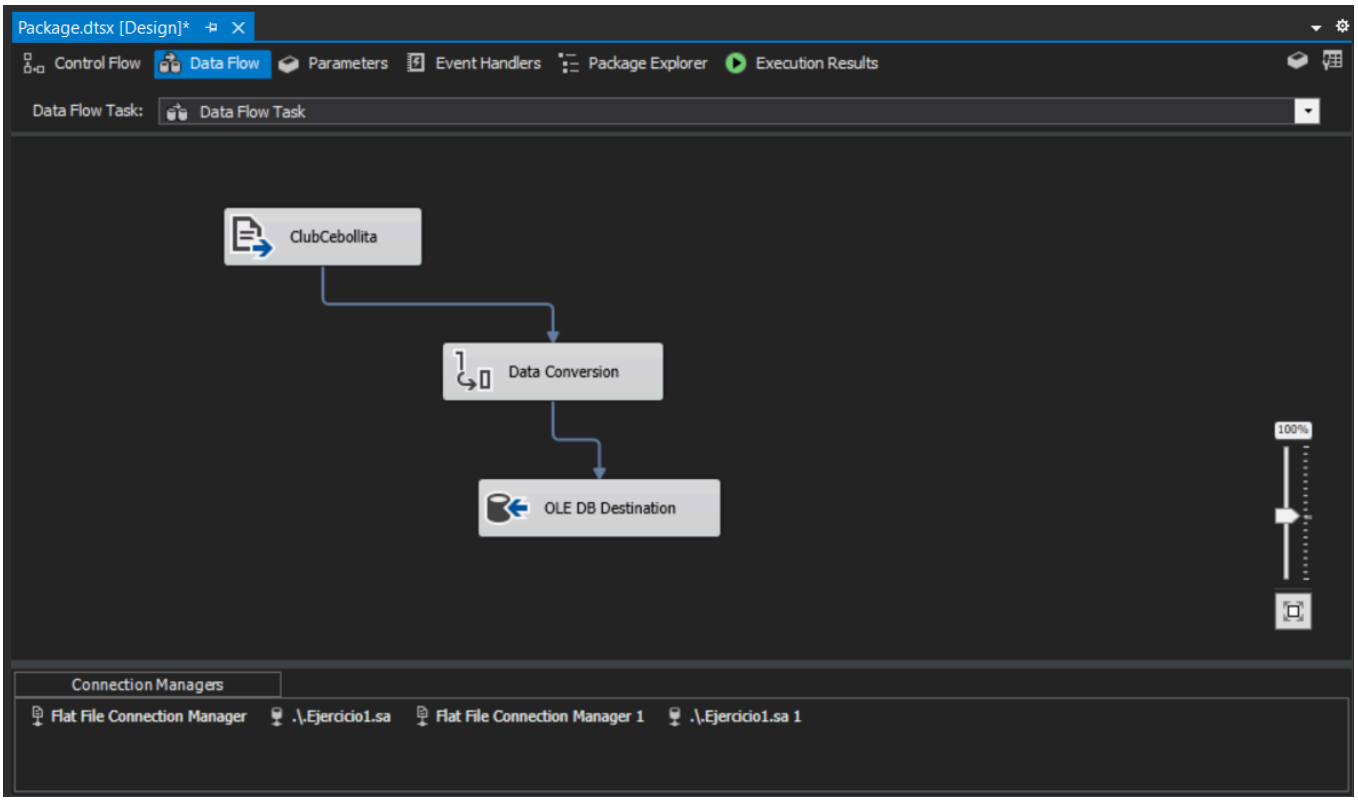
Enlace al vídeo

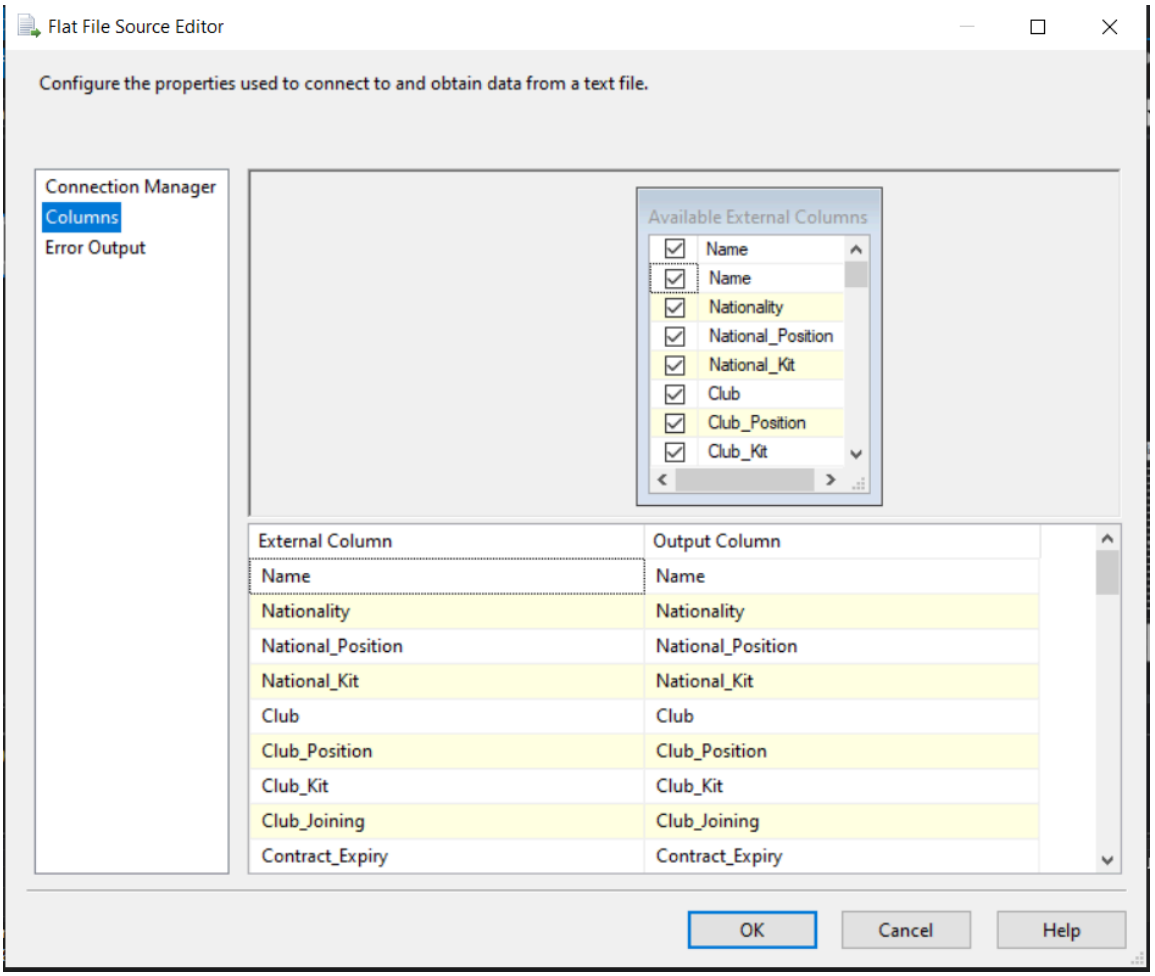
DMD - Desafio 1

Sábado 14 de septiembre de 2024

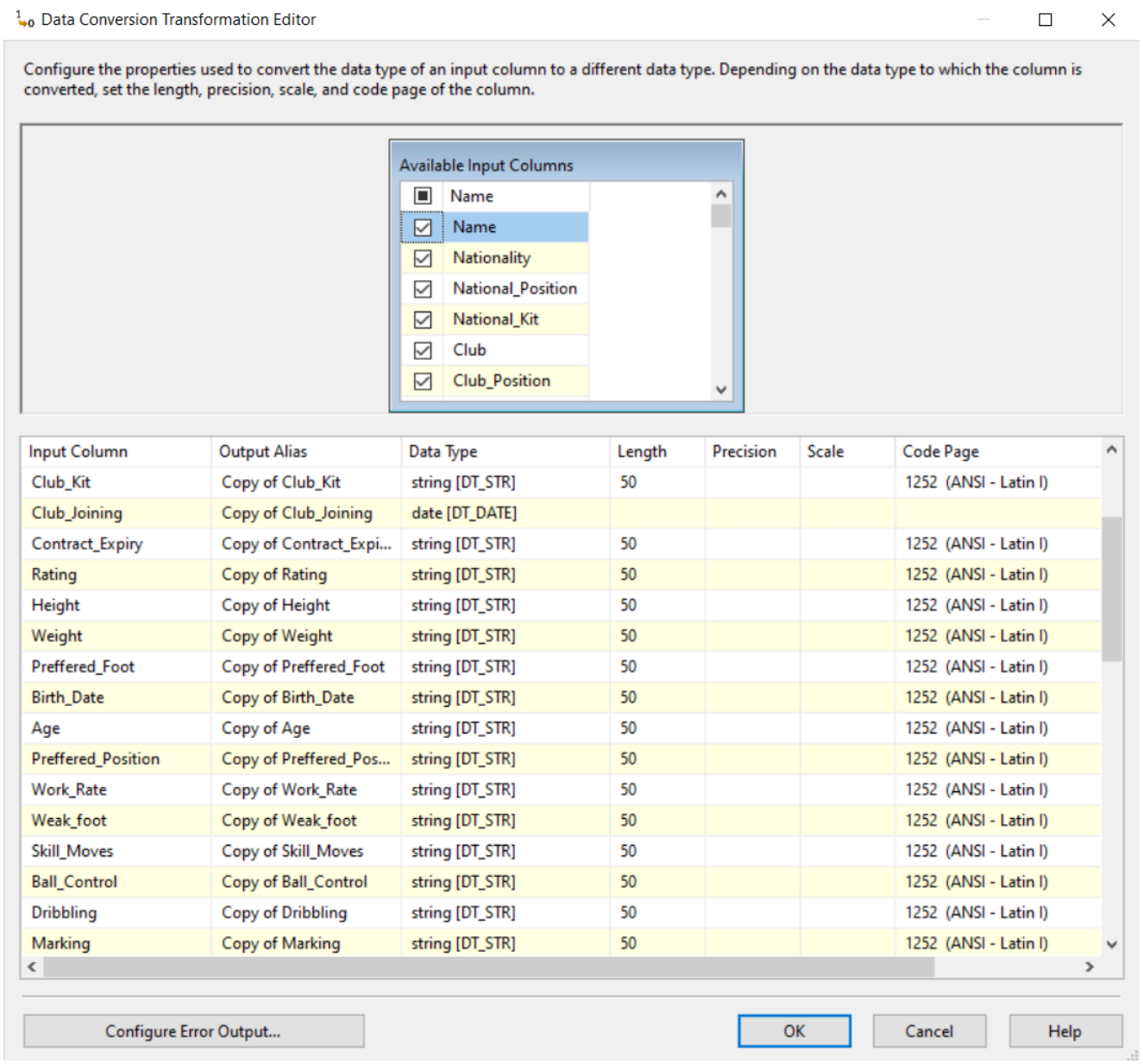
Ejercicio 1

- DATA FLOW.

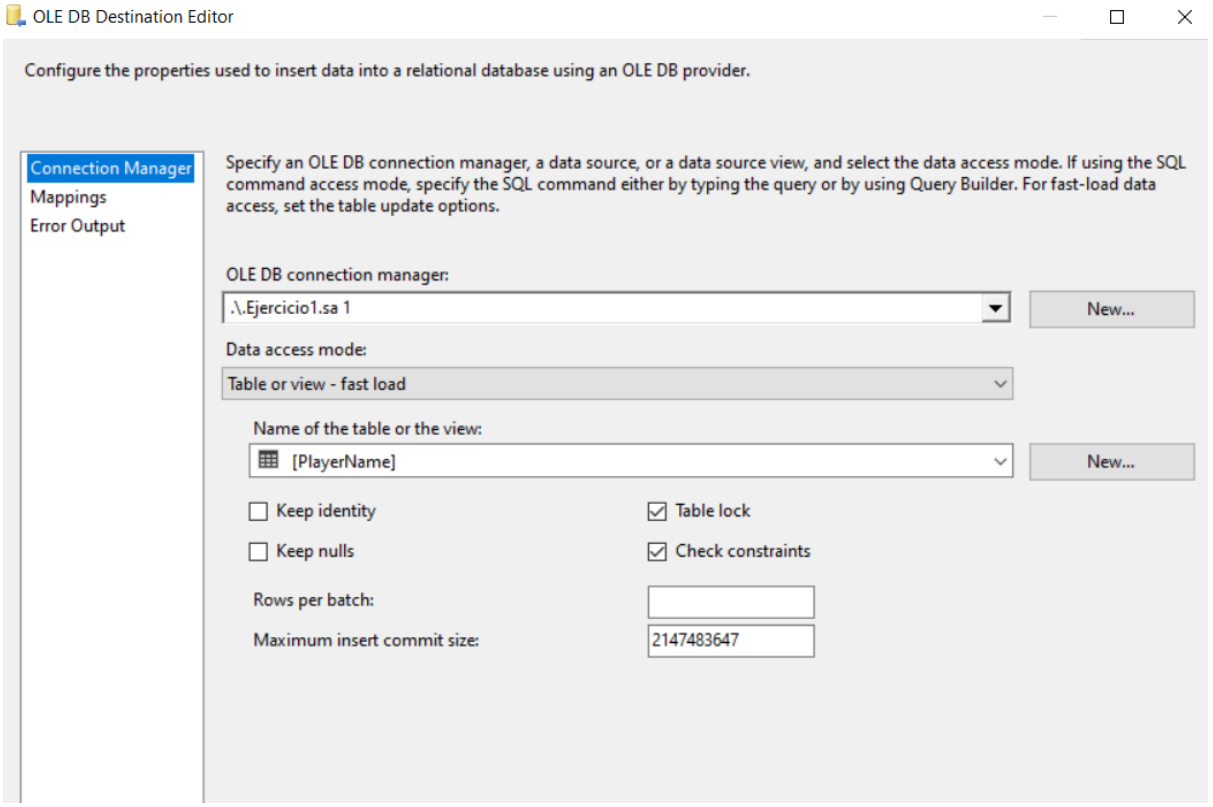


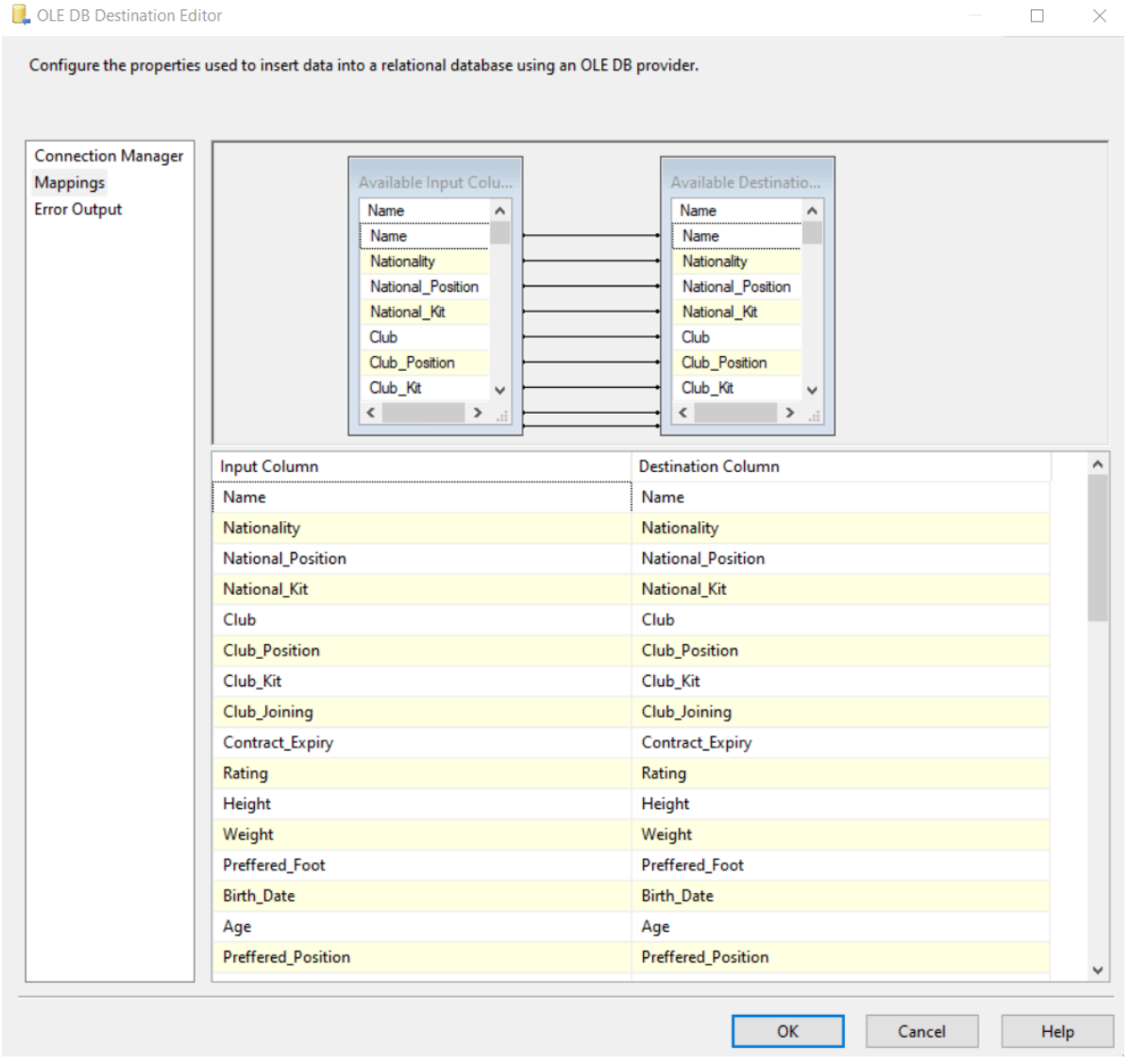


- DATA CONVERSION



- OLE DB DESTINATION





- Calcular Correlacion Rating, Edad .

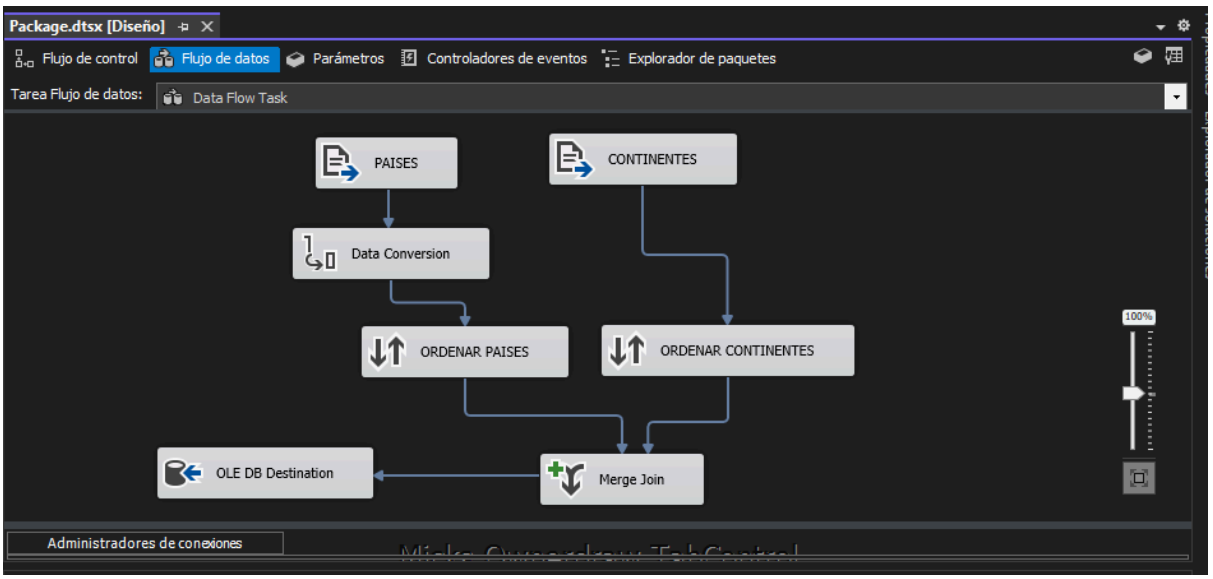
```sql
USE [Ejercicio1]
GO
/****** Object:  StoredProcedure [dbo].[CalcularCorrelacionRatingEdad]    Script Date: 9/14/2024 2:26:10 PM ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
ALTER PROCEDURE [dbo].[CalcularCorrelacionRatingEdad]
AS
BEGIN
    -- Obtener la media de rating y age
    WITH AvgValues AS (
        SELECT
            AVG(CAST(Rating AS FLOAT)) AS PromedioRating,
            AVG(CAST(Age AS FLOAT)) AS PromedioEdad
        FROM
            dbo.PlayerName
    ),
    -- Obtener la suma de las diferencias
    Covariance AS (
        SELECT
            SUM((CAST(Rating AS FLOAT) - AvgValues.PromedioRating) * (CAST(Age AS FLOAT) - AvgValues.PromedioEdad)) AS Covarianza,
            COUNT(*) AS NumeroFilas
        FROM
            dbo.PlayerName, AvgValues
    ),
    -- Obtener la desviación estándar
    StdDevs AS (
        SELECT
            STDEV(CAST(Rating AS FLOAT)) AS DesvEstRating,
            STDEV(CAST(Age AS FLOAT)) AS DesvEstEdad
        FROM
            dbo.PlayerName
    )
    -- Calcular la correlación
    SELECT
        ROUND((Covariance.Covarianza / (Covariance.NumeroFilas - 1)) /
        (StdDevs.DesvEstRating * StdDevs.DesvEstEdad), 4) AS Correlacion_Rating_Edad
    FROM
        Covariance, StdDevs;
END;
```

- Calcular Estadisticas Jugadores.

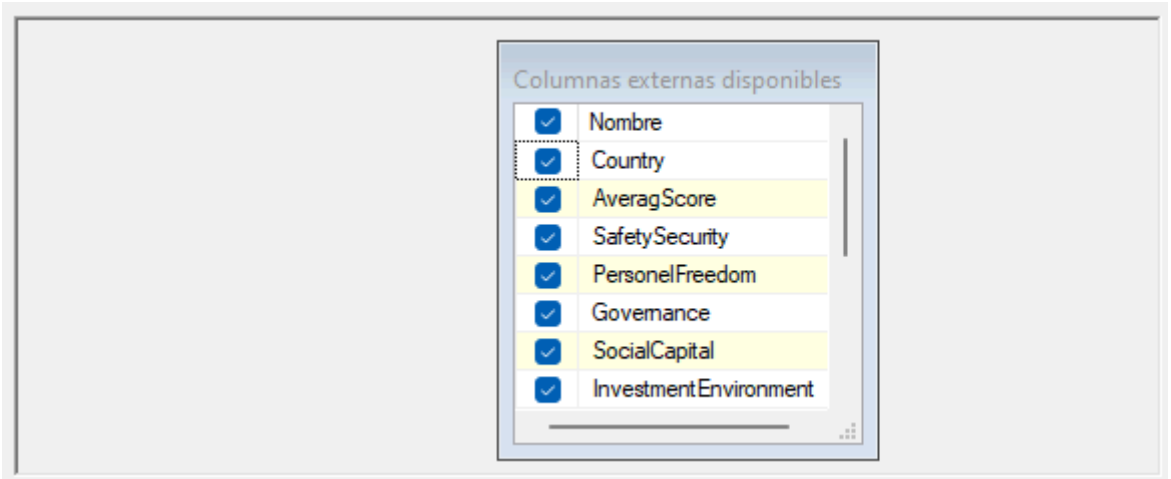
```sql
USE [Ejercicio1]
GO
/****** Object:  StoredProcedure [dbo].[CalcularEstadisticasJugadores]    Script Date: 9/14/2024 2:26:14 PM ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
ALTER PROCEDURE [dbo].[CalcularEstadisticasJugadores]
AS
BEGIN
    WITH CleanedData AS (
        SELECT
            Age AS Edad,
            -- Limpiar y convertir Height y Weight a FLOAT
            TRY_CAST(REPLACE(REPLACE(Height, ' cm', ''), ',', '') AS FLOAT) AS HeightCleaned,
            TRY_CAST(REPLACE(REPLACE(Weight, ' kg', ''), ',', '') AS FLOAT) AS WeightCleaned,
            TRY_CAST(Rating AS FLOAT) AS RatingCleaned
        FROM
            dbo.PlayerName
        WHERE
            -- Asegurarse de que después de eliminar las unidades, los valores pueden convertirse a números
            TRY_CAST(REPLACE(REPLACE(Height, ' cm', ''), ',', '') AS FLOAT) IS NOT NULL
            AND TRY_CAST(REPLACE(REPLACE(Weight, ' kg', ''), ',', '') AS FLOAT) IS NOT NULL
            AND TRY_CAST(Rating AS FLOAT) IS NOT NULL  -- Asegurarse de que Rating también sea numérico
    )
    -- Calcular las estadísticas
    SELECT
        Edad,
        ROUND(AVG(HeightCleaned), 2) AS Altura_Promedio_cm,
        ROUND(STDEV(HeightCleaned), 2) AS Desviacion_Estandar_Altura_cm,
        ROUND(AVG(WeightCleaned), 2) AS Peso_Promedio_kg,
        ROUND(STDEV(WeightCleaned), 2) AS Desviacion_Estandar_Peso_kg,
        ROUND(AVG(RatingCleaned), 2) AS Calificacion_Promedio,
        ROUND(STDEV(RatingCleaned), 2) AS Desviacion_Estandar_Calificacion
    FROM
        CleanedData
    GROUP BY
        Edad
    ORDER BY
        Edad;
END;
```

Ejercicio 2

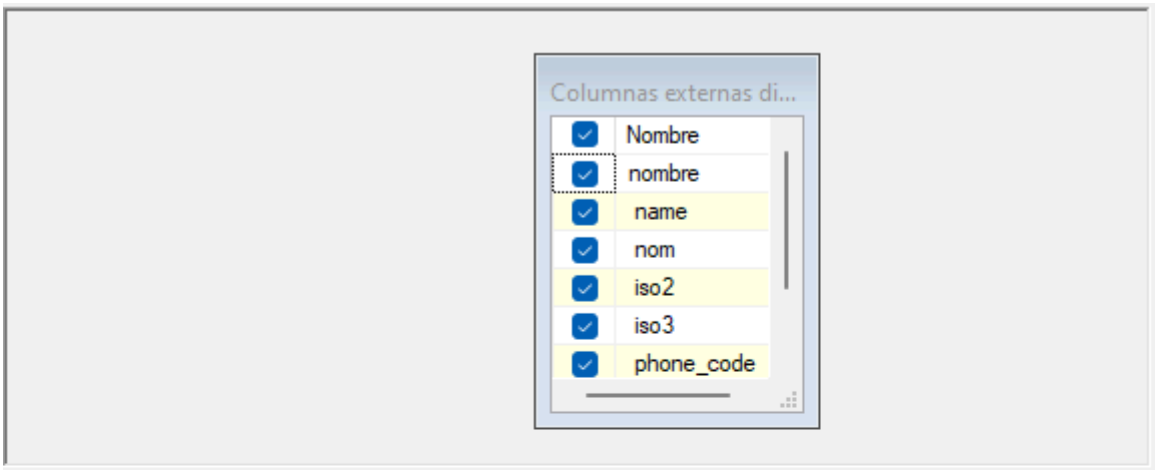
Creación de ETL



Orígenes de datos

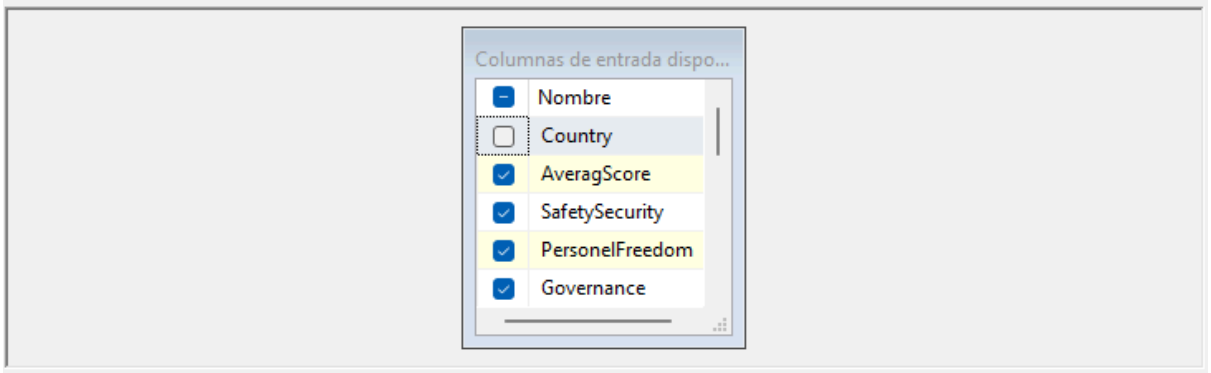


Columna externa	Columna de salida
Country	Country
AveragScore	AveragScore
SafetySecurity	SafetySecurity
PersonelFreedom	PersonelFreedom
Governance	Governance
SocialCapital	SocialCapital
InvestmentEnvironment	InvestmentEnvironment
EnterpriseConditions	EnterpriseConditions
MarketAccessInfrastructure	MarketAccessInfrastructure



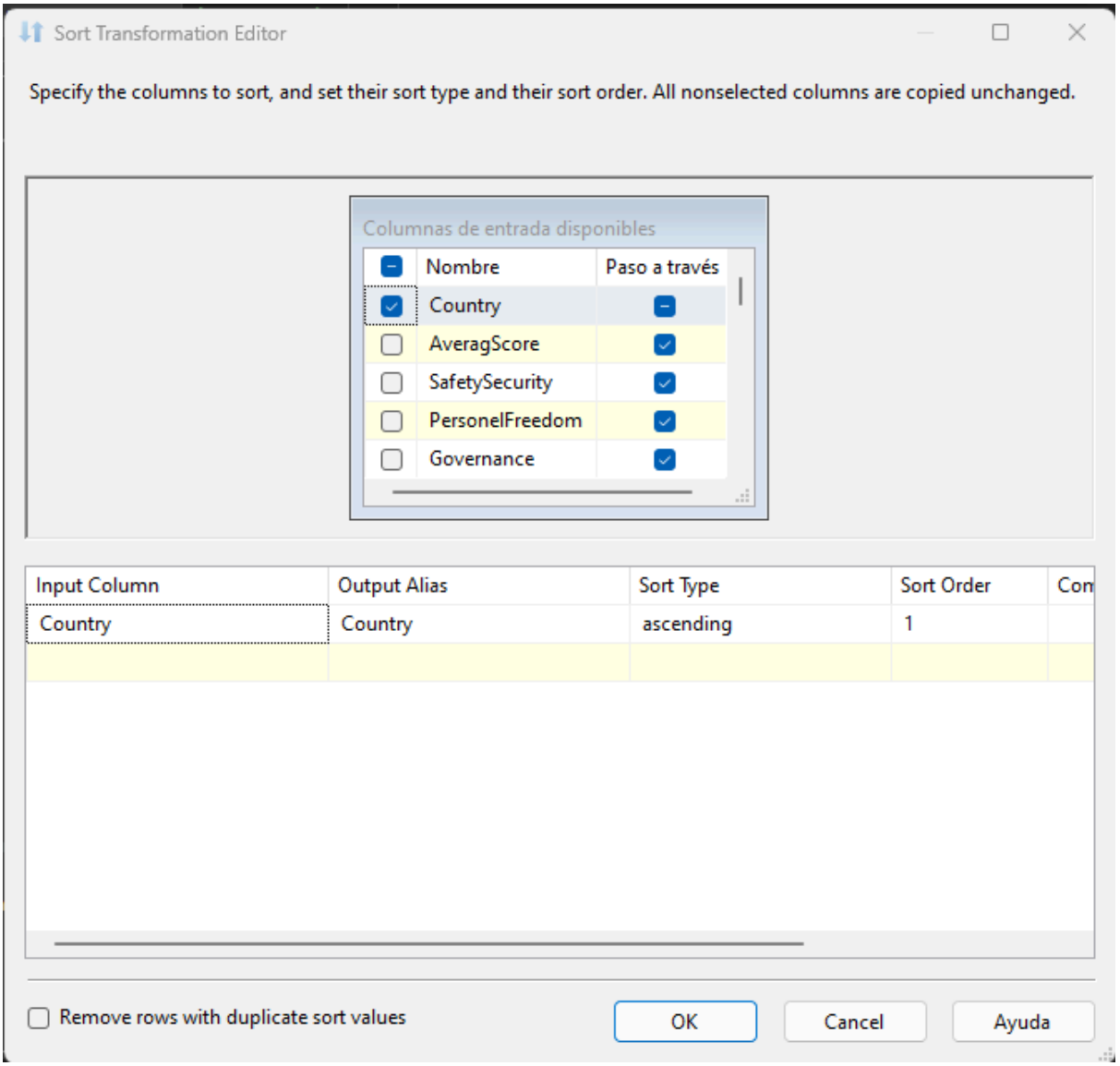
Columna externa	Columna de salida
nombre	nombre
name	name
nom	nom
iso2	iso2
iso3	iso3
phone_code	phone_code
continente	continente

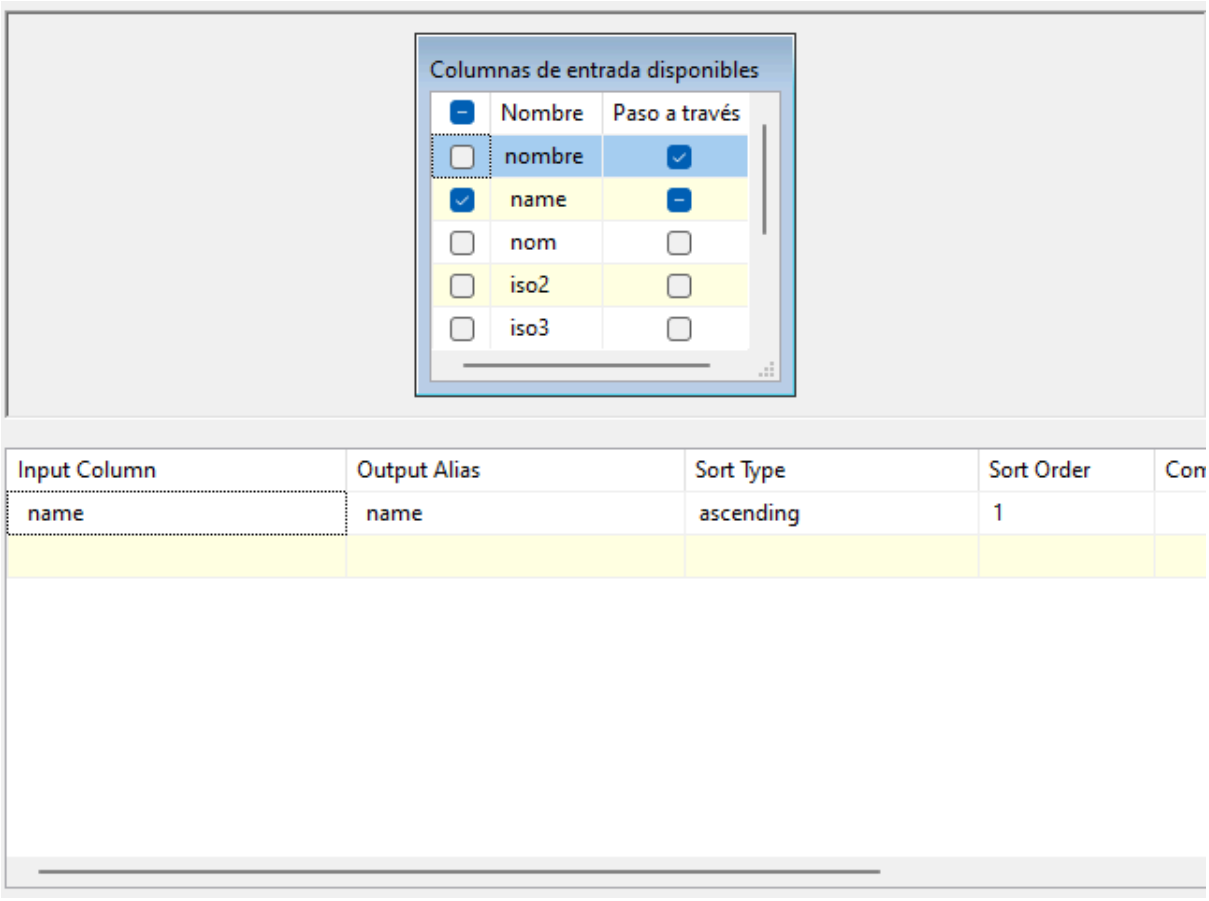
Conversión de datos



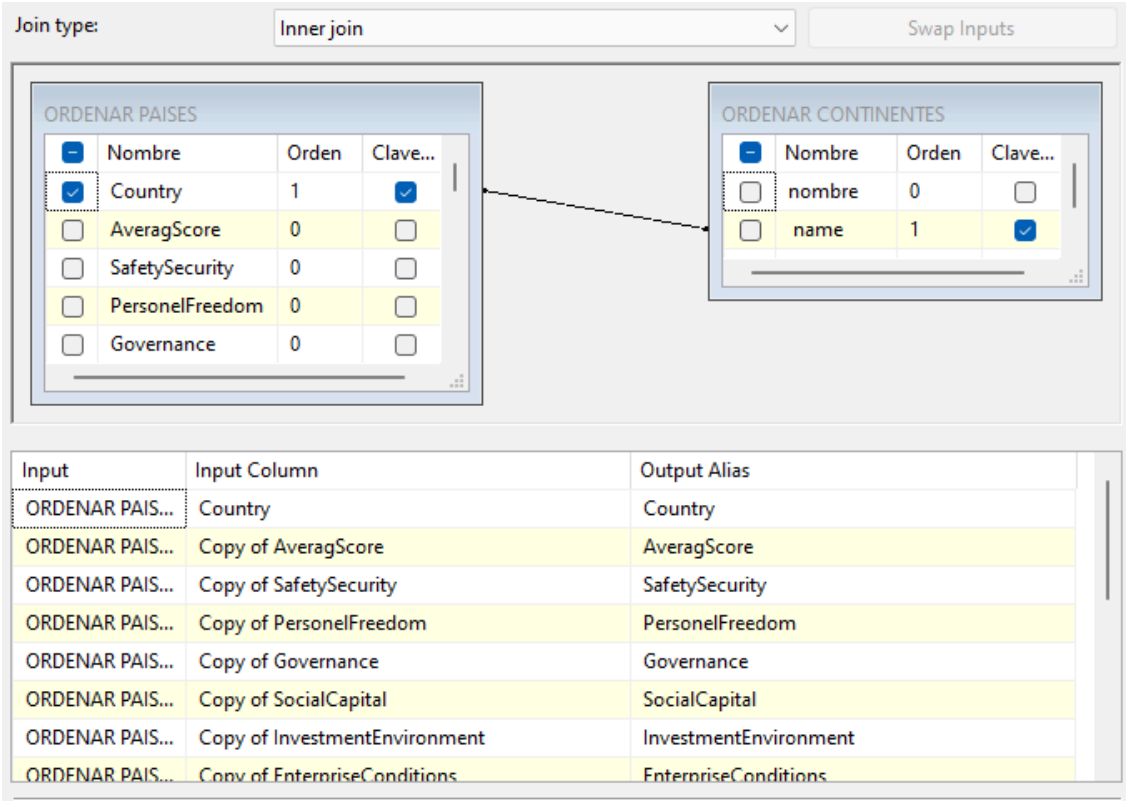
Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Pag
AveragScore	Copy of AveragScore	decimal [DT_DECIMAL]			0	
SafetySecurity	Copy of SafetySecurity	decimal [DT_DECIMAL]			0	
PersonelFreedom	Copy of PersonelFree...	decimal [DT_DECIMAL]			0	
Governance	Copy of Governance	decimal [DT_DECIMAL]			0	
SocialCapital	Copy of SocialCapital	decimal [DT_DECIMAL]			0	
InvestmentEnvironment	Copy of InvestmentEn...	decimal [DT_DECIMAL]			0	
EnterpriseConditions	Copy of EnterpriseCo...	decimal [DT_DECIMAL]			0	
MarketAccessInfrastru...	Copy of MarketAccess...	decimal [DT_DECIMAL]			0	
EconomicQuality	Copy of EconomicQu...	decimal [DT_DECIMAL]			0	

Orden de orígenes de datos

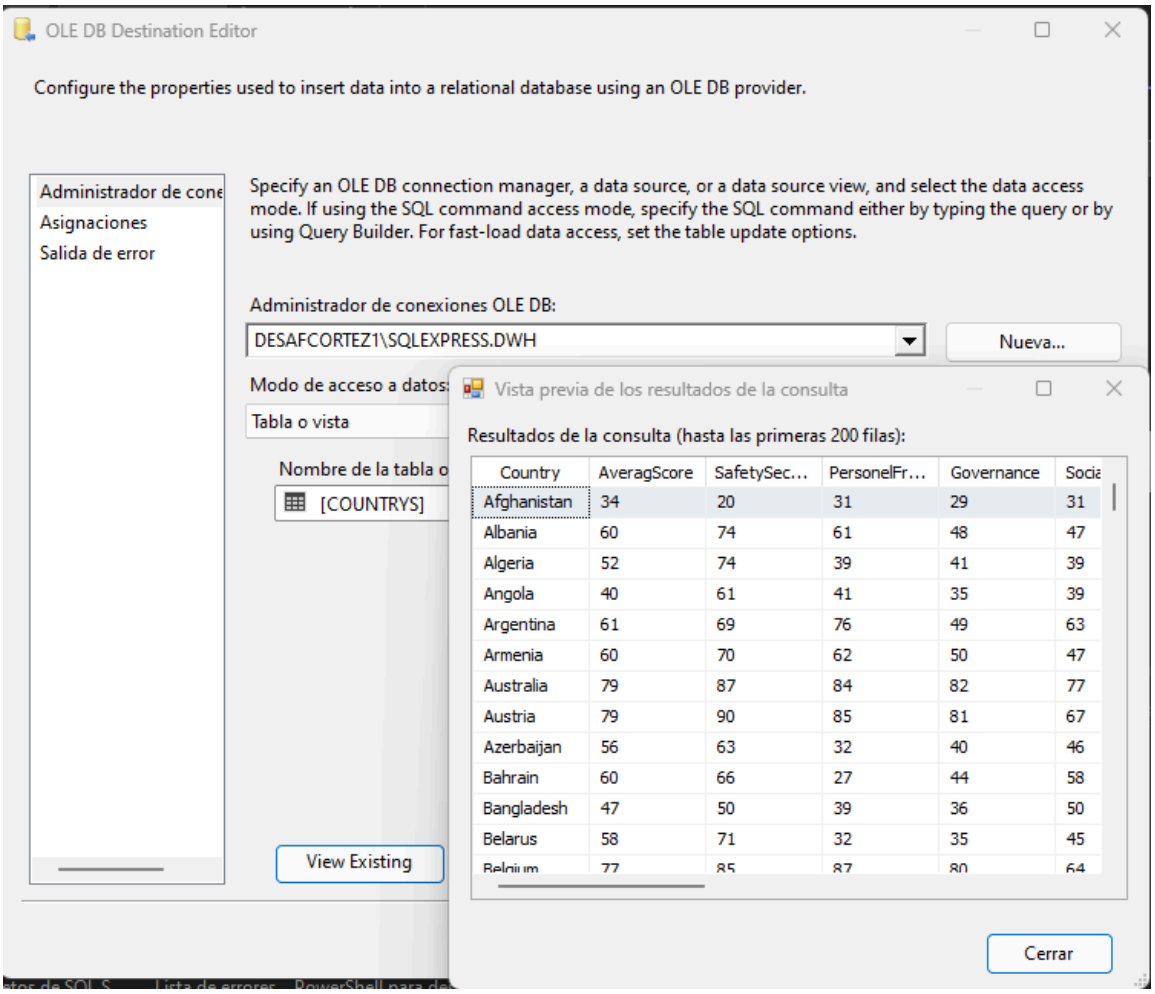




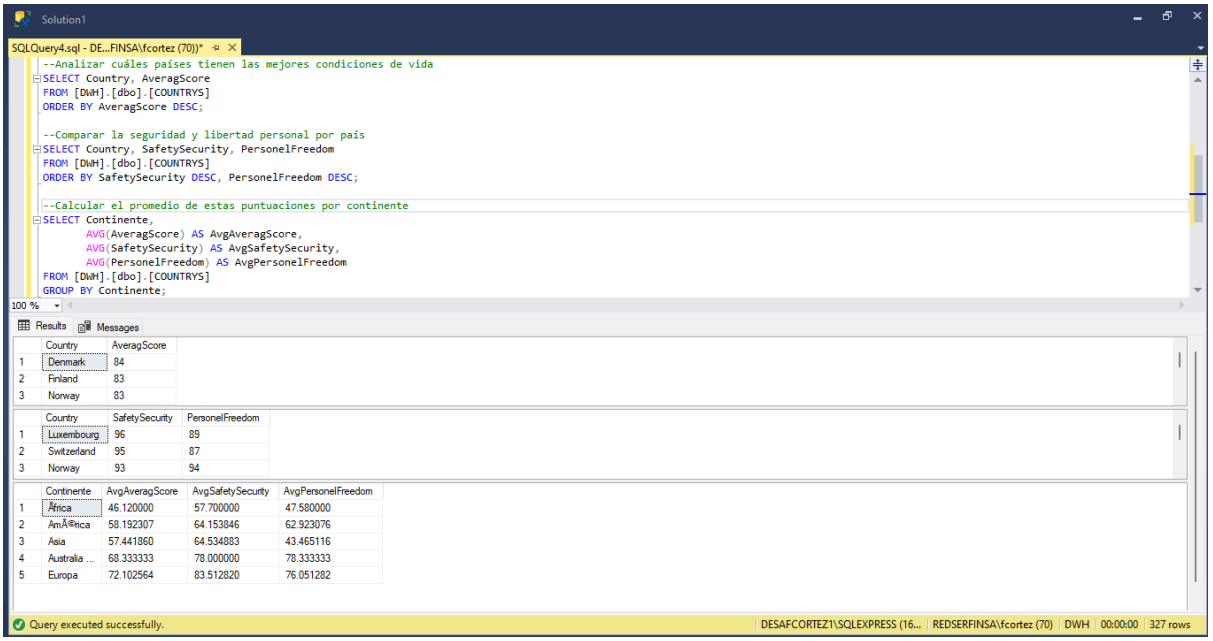
Unificación de datos



Destino de datos

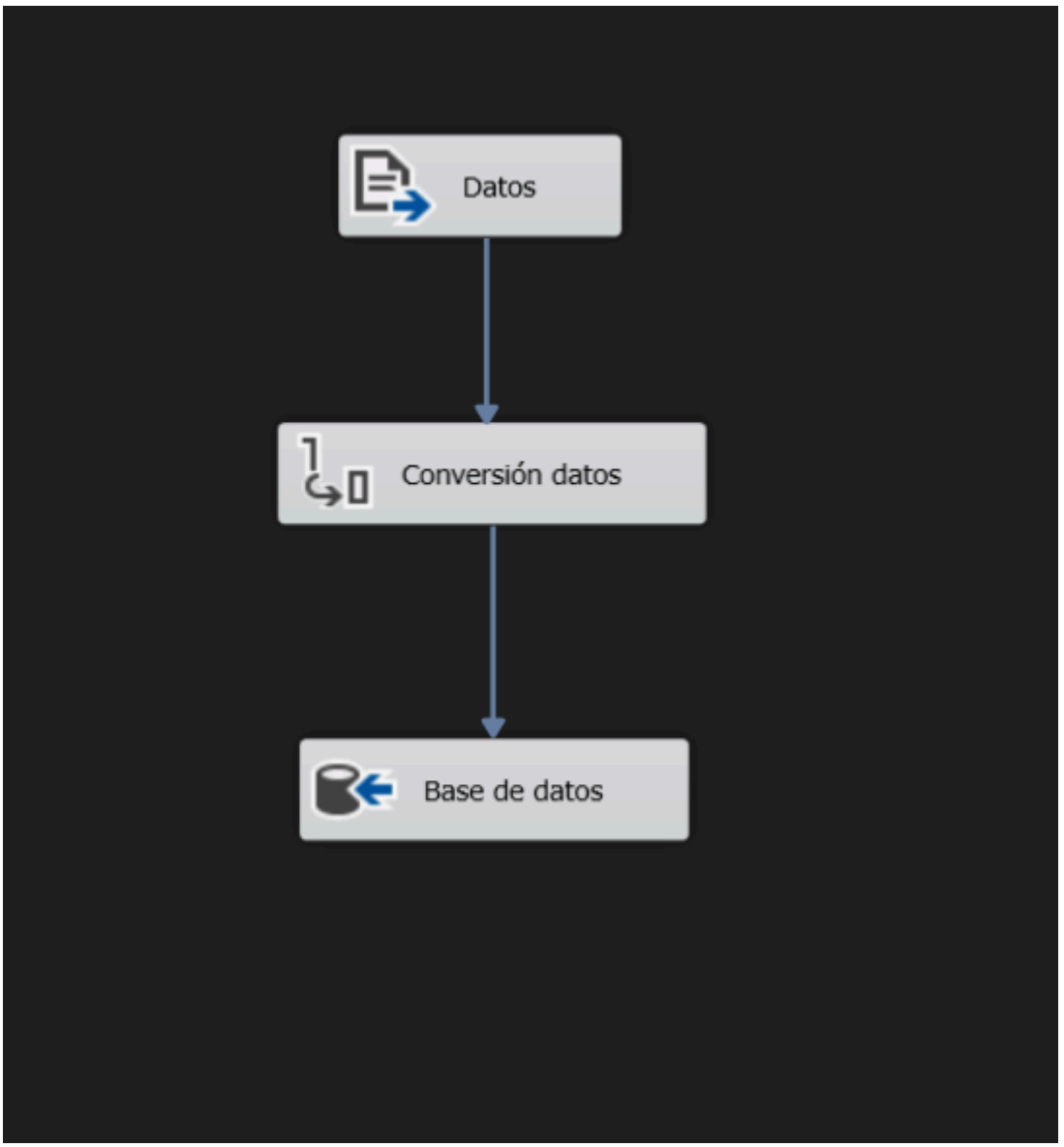


Consulta de estadísticas

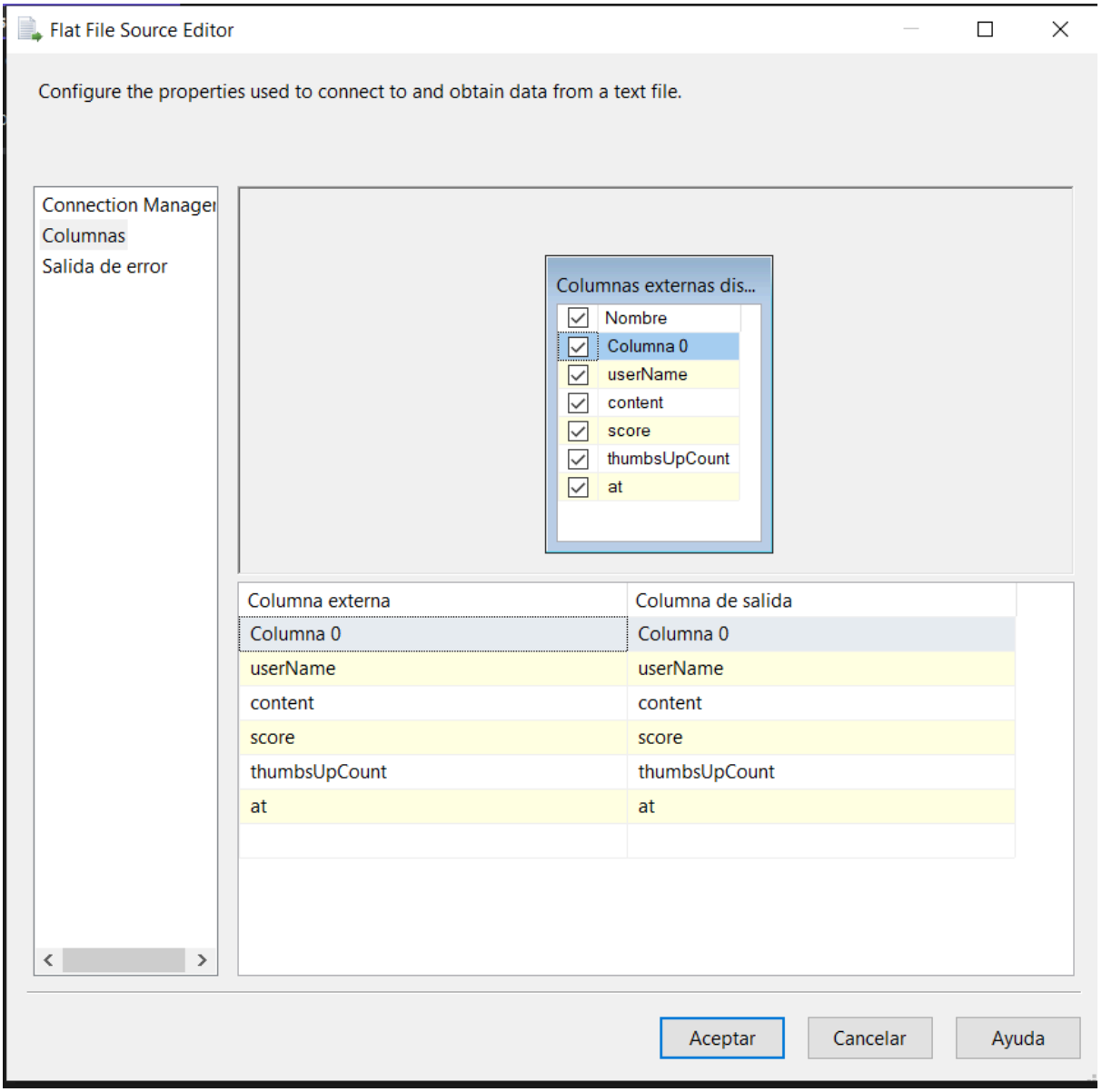


Ejercicio 3

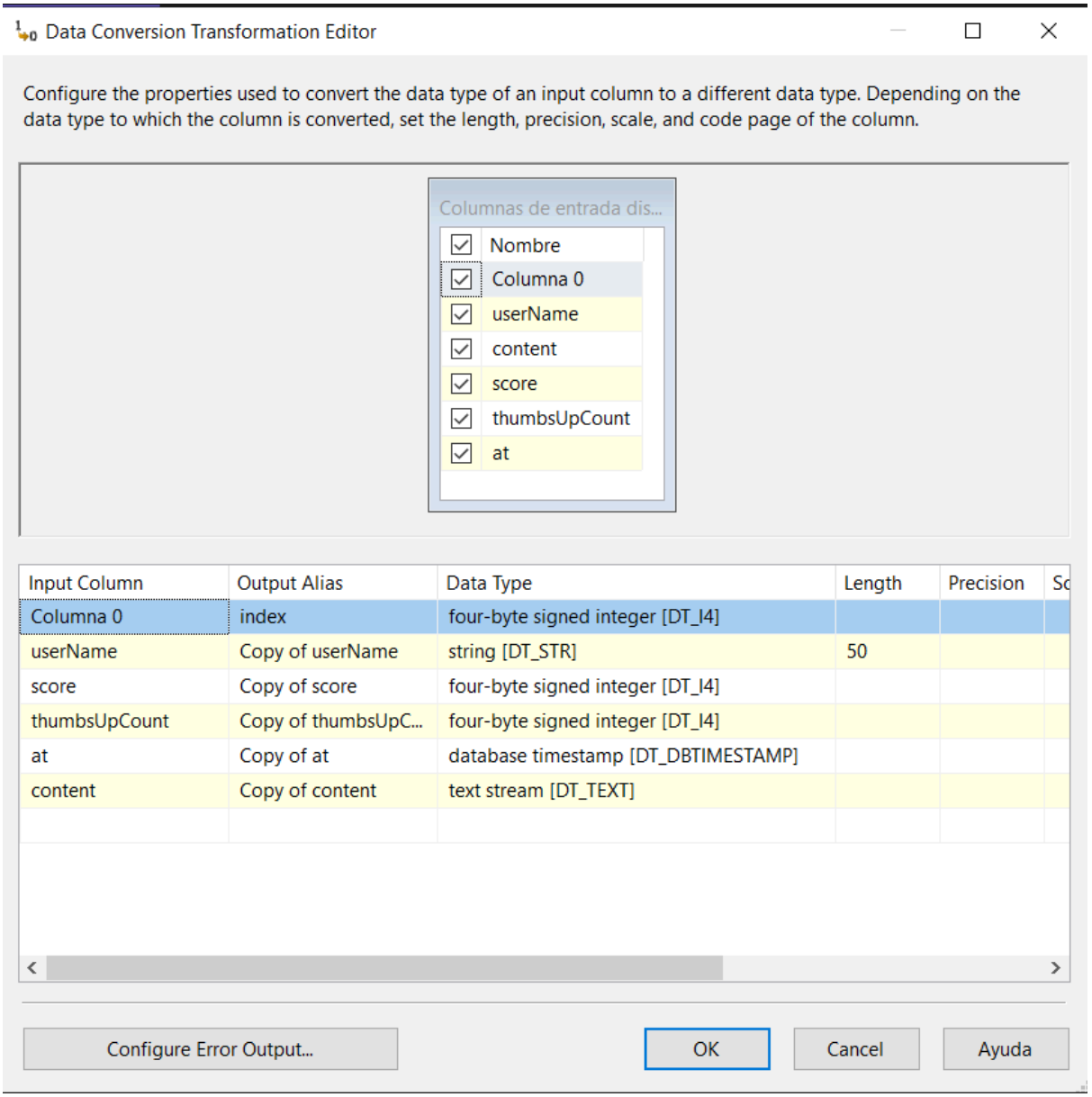
Data flow



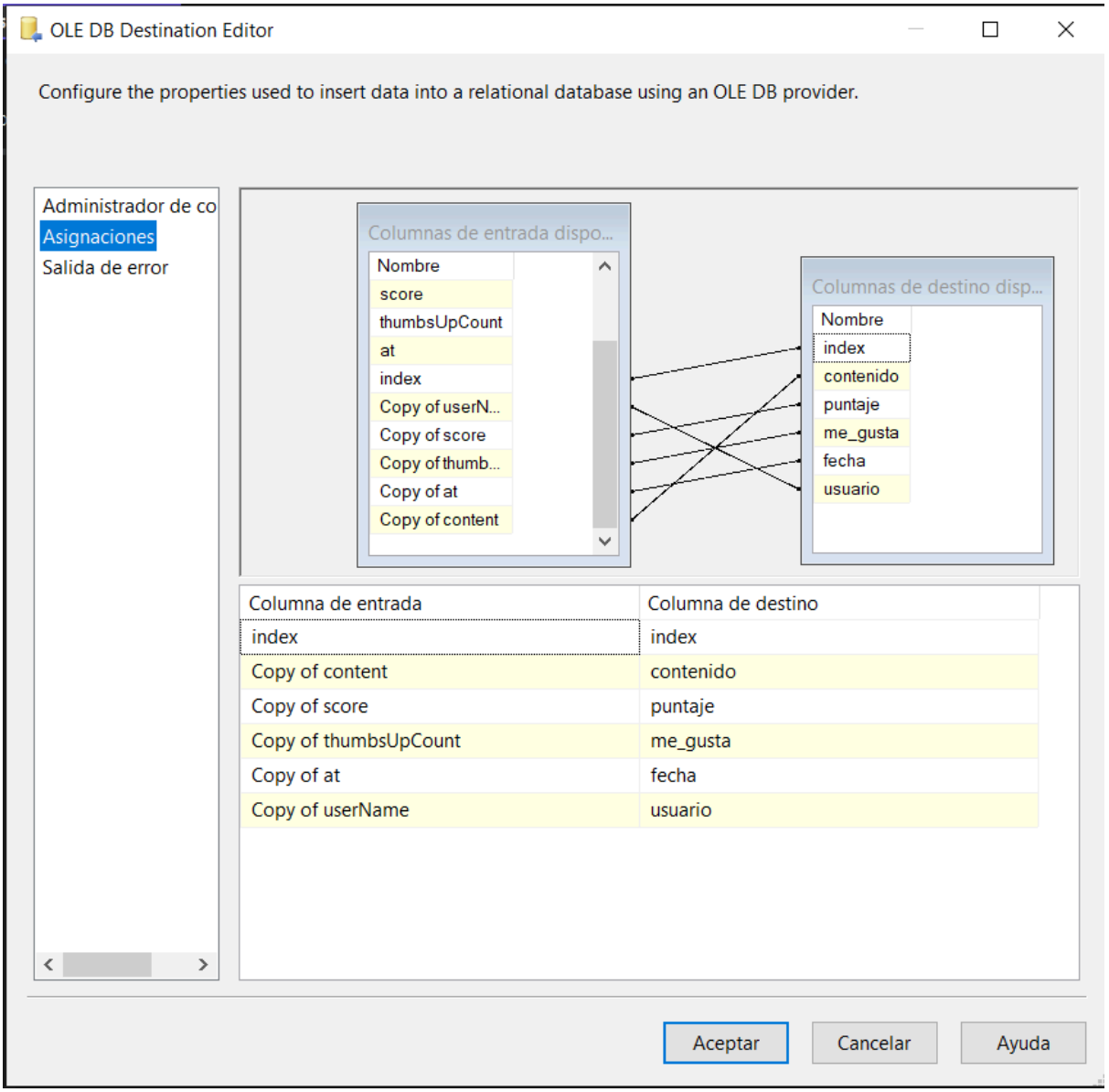
Carga de datos desde el csv



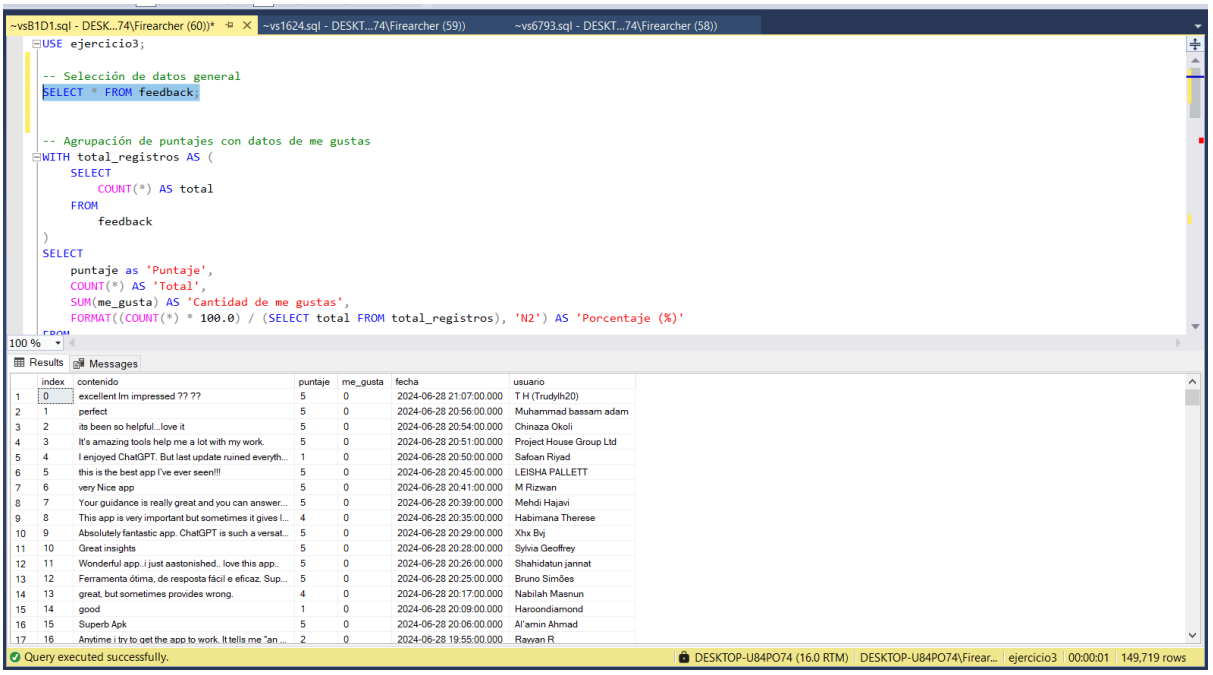
Conversión de datos



Carga de datos



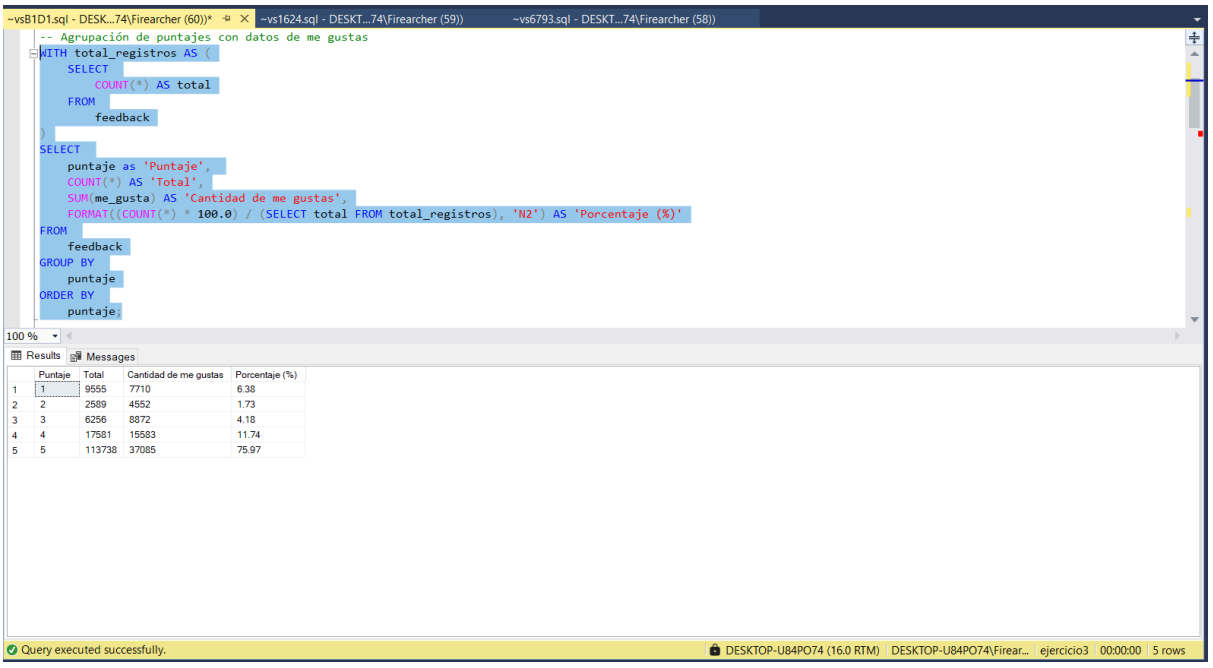
Selección de datos general



```
USE ejercicio3;

-- Selección de datos general
SELECT * FROM feedback;


-- Agrupación de puntajes con datos de me gustas
WITH total_registros AS (
    SELECT
        COUNT(*) AS total
    FROM
        feedback
)
SELECT
    puntaje as 'Puntaje',
    COUNT(*) AS 'Total',
    SUM(me_gusta) AS 'Cantidad de me gustas',
    FORMAT((COUNT(*) * 100.0) / (SELECT total FROM total_registros), 'N2') AS 'Porcentaje (%)'
FROM
```

	index	contenido	puntaje	me_gusta	fecha	usuario
1	0	excellent Im impressed ?? ??	5	0	2024-06-28 21:07:00.000	T H (Trudylh20)
2	1	perfect	5	0	2024-06-28 20:56:00.000	Muhammad bassam adam
3	2	its been so helpful...love it	5	0	2024-06-28 20:54:00.000	Chinaza Okoli
4	3	It's amazing tools help me a lot with my work.	5	0	2024-06-28 20:51:00.000	Project House Group Ltd
5	4	I enjoyed ChatGPT. But last update ruined everyth...	1	0	2024-06-28 20:50:00.000	Safoan Riyad
6	5	this is the best app I've ever seen!!!	5	0	2024-06-28 20:45:00.000	LEISHA PALLETT
7	6	very Nice app	5	0	2024-06-28 20:41:00.000	M Rizwan
8	7	Your guidance is really great and you can answer...	5	0	2024-06-28 20:39:00.000	Mehdi Hajavi
9	8	This app is very important but sometimes it gives l...	4	0	2024-06-28 20:35:00.000	Habimana Therese
10	9	Absolutely fantastic app. ChatGPT is such a versat...	5	0	2024-06-28 20:29:00.000	Xhx Bvj
11	10	Great insights	5	0	2024-06-28 20:28:00.000	Sylvia Geoffrey
12	11	Wonderful app..i just aastonished.. love this app..	5	0	2024-06-28 20:26:00.000	Shahidatun jannat
13	12	Ferramenta ótima, de resposta fácil e eficaz. Sup...	5	0	2024-06-28 20:25:00.000	Bruno Simões
14	13	great, but sometimes provides wrong.	4	0	2024-06-28 20:17:00.000	Nabilah Masnun
15	14	good	1	0	2024-06-28 20:09:00.000	Haroondiamond
16	15	Superb Apk	5	0	2024-06-28 20:06:00.000	Al'amin Ahmad
17	16	Anytime i try to get the app to work. It tells me "an ...	2	0	2024-06-28 19:55:00.000	Rayyan R

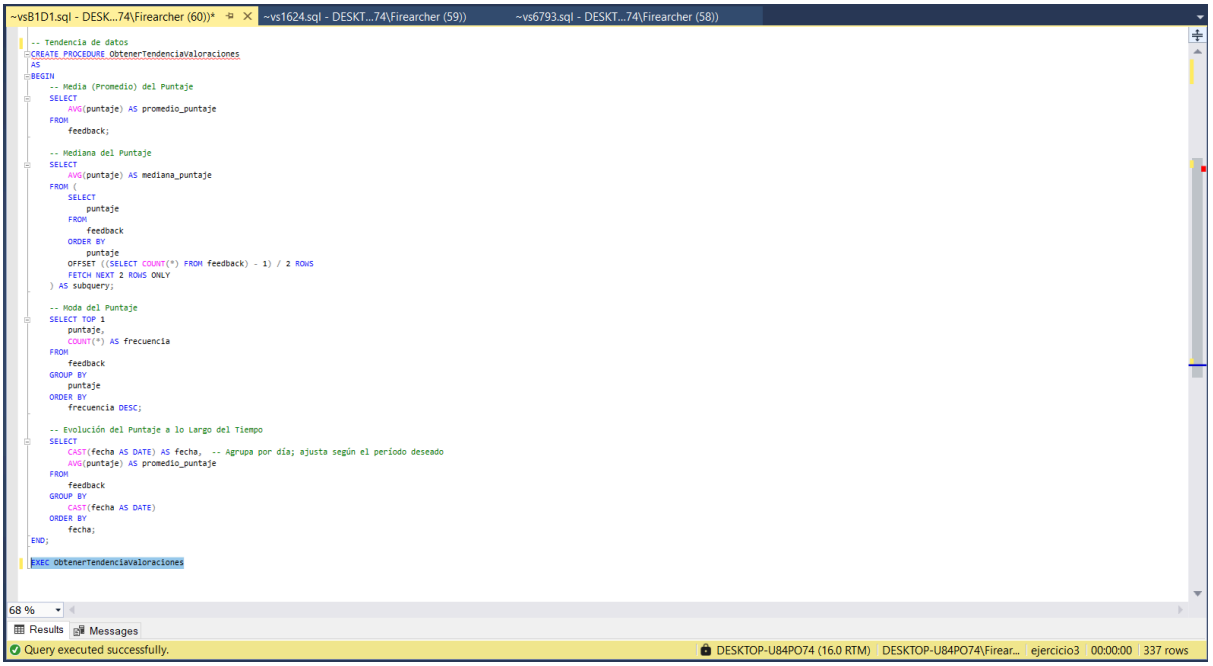
Selección de registros con suma de puntajes y porcentajes equivalentes



```
-- Agrupación de puntajes con datos de me gustas
WITH total_registros AS (
    SELECT
        COUNT(*) AS total
    FROM
        feedback
)
SELECT
    puntaje as 'Puntaje',
    COUNT(*) AS 'Total',
    SUM(me_gusta) AS 'Cantidad de me gustas',
    FORMAT((COUNT(*) * 100.0) / (SELECT total FROM total_registros), 'N2') AS 'Porcentaje (%)'
FROM
    feedback
GROUP BY
    puntaje
ORDER BY
    puntaje;
```

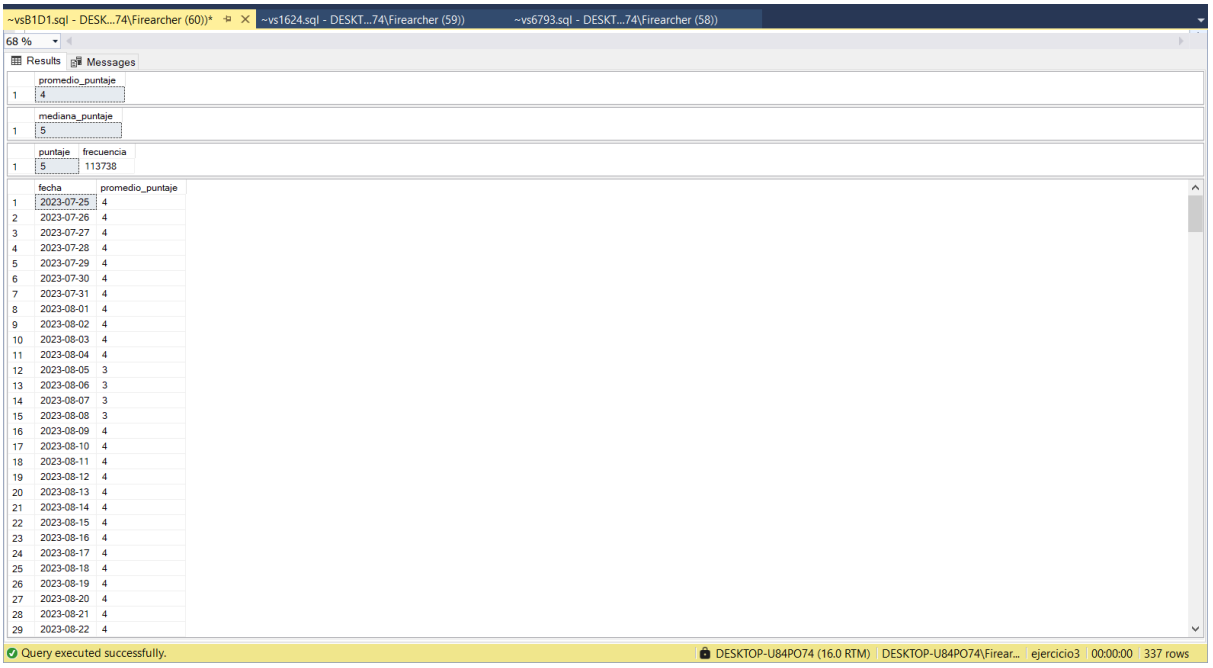
	Puntaje	Total	Cantidad de me gustas	Porcentaje (%)
1	1	9555	7710	6.38
2	2	2589	4552	1.73
3	3	6256	8872	4.18
4	4	17581	15583	11.74
5	5	113738	37085	75.97

Procedimiento con detalles estadísticos



```sql
-- Tendencia de datos
CREATE PROCEDURE ObtenerTendenciaValoraciones
AS
BEGIN
    -- Media (Promedio) del Puntaje
    SELECT
        AVG(puntaje) AS promedio_puntaje
    FROM
        feedback;

    -- Mediana del Puntaje
    SELECT
        AVG(puntaje) AS mediana_puntaje
    FROM (
        SELECT
            puntaje
        FROM
            feedback
        ORDER BY
            puntaje
        OFFSET ((SELECT COUNT(*) FROM feedback) - 1) / 2 ROWS
        FETCH NEXT 2 ROWS ONLY
    ) AS subquery;

    -- Moda del Puntaje
    SELECT TOP 1
        puntaje,
        COUNT(*) AS frecuencia
    FROM
        feedback
    GROUP BY
        puntaje
    ORDER BY
        frecuencia DESC;

    -- Evolución del Puntaje a lo Largo del Tiempo
    SELECT
        CAST(fecha AS DATE) AS fecha,  -- Agrupa por día; ajusta según el período deseado
        AVG(puntaje) AS promedio_puntaje
    FROM
        feedback
    GROUP BY
        CAST(fecha AS DATE)
    ORDER BY
        fecha;
END;

EXEC ObtenerTendenciaValoraciones
```

Resultado de ejecución



promedio_puntaje

1	4

mediana_puntaje

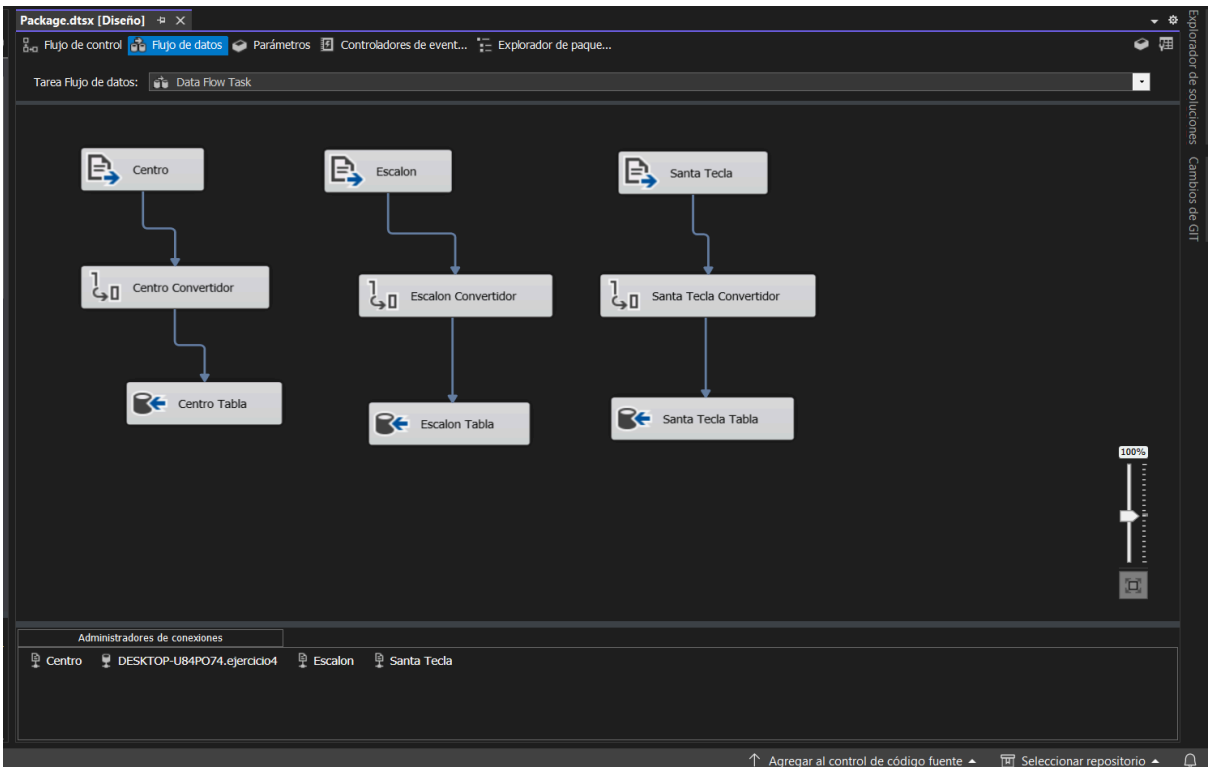
1	5

	puntaje	frecuencia
1	5	113738

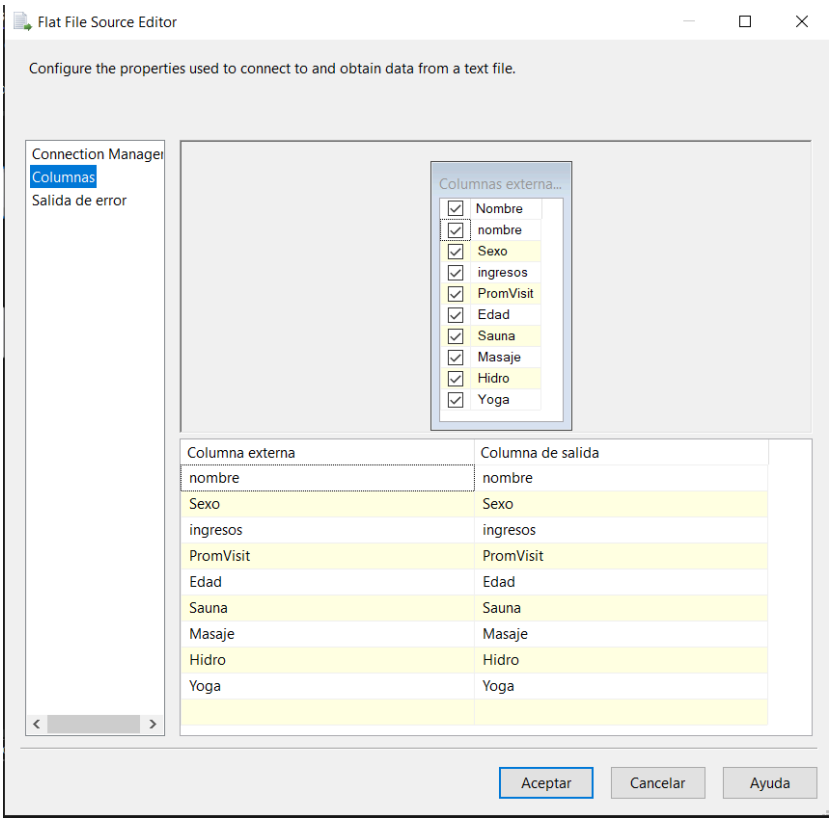
	fecha	promedio_puntaje
1	2023-07-25	4
2	2023-07-26	4
3	2023-07-27	4
4	2023-07-28	4
5	2023-07-29	4
6	2023-07-30	4
7	2023-07-31	4
8	2023-08-01	4
9	2023-08-02	4
10	2023-08-03	4
11	2023-08-04	4
12	2023-08-05	3
13	2023-08-06	3
14	2023-08-07	3
15	2023-08-08	3
16	2023-08-09	4
17	2023-08-10	4
18	2023-08-11	4
19	2023-08-12	4
20	2023-08-13	4
21	2023-08-14	4
22	2023-08-15	4
23	2023-08-16	4
24	2023-08-17	4
25	2023-08-18	4
26	2023-08-19	4
27	2023-08-20	4
28	2023-08-21	4
29	2023-08-22	4

Ejercicio 4

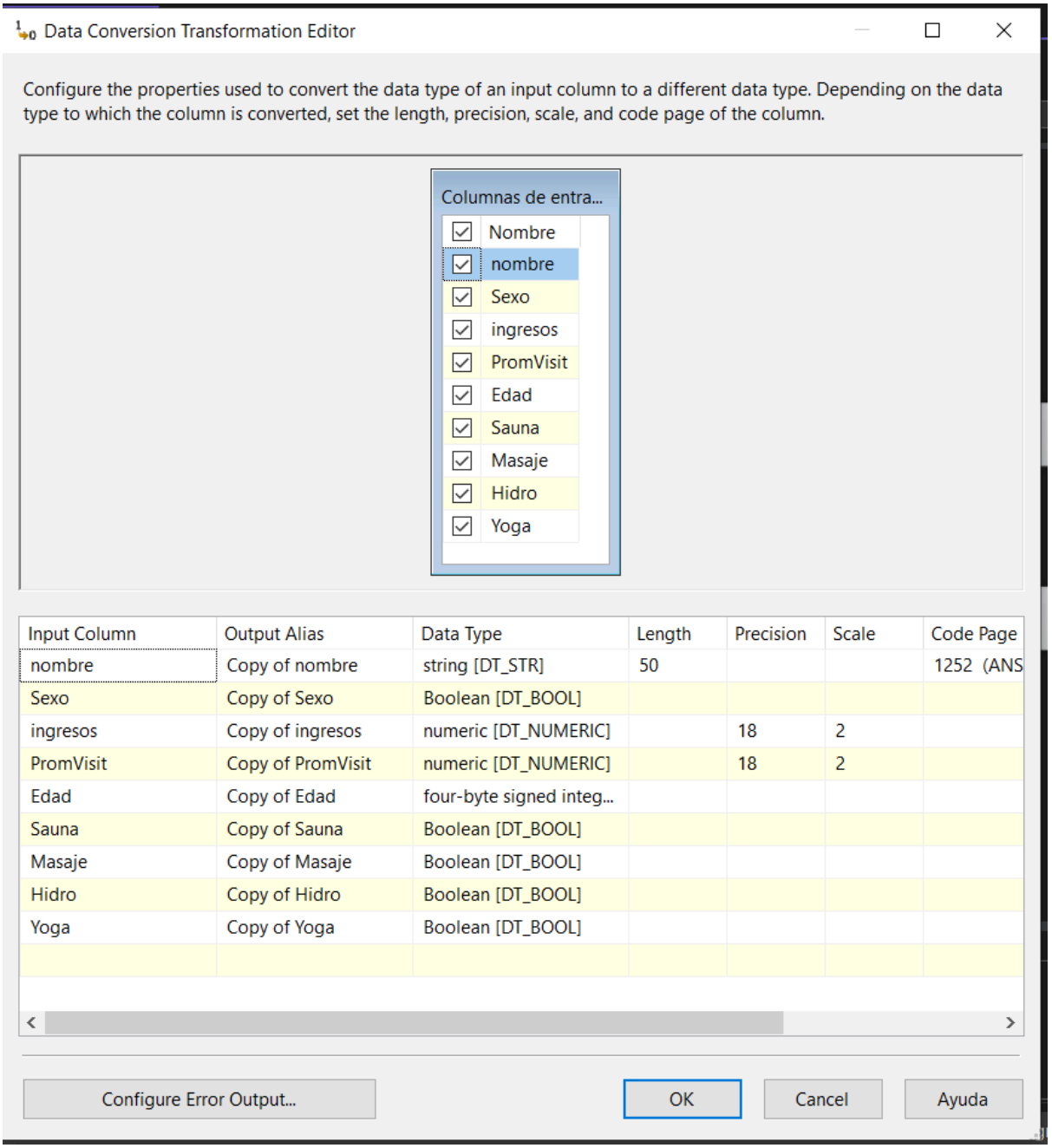
Data flow general



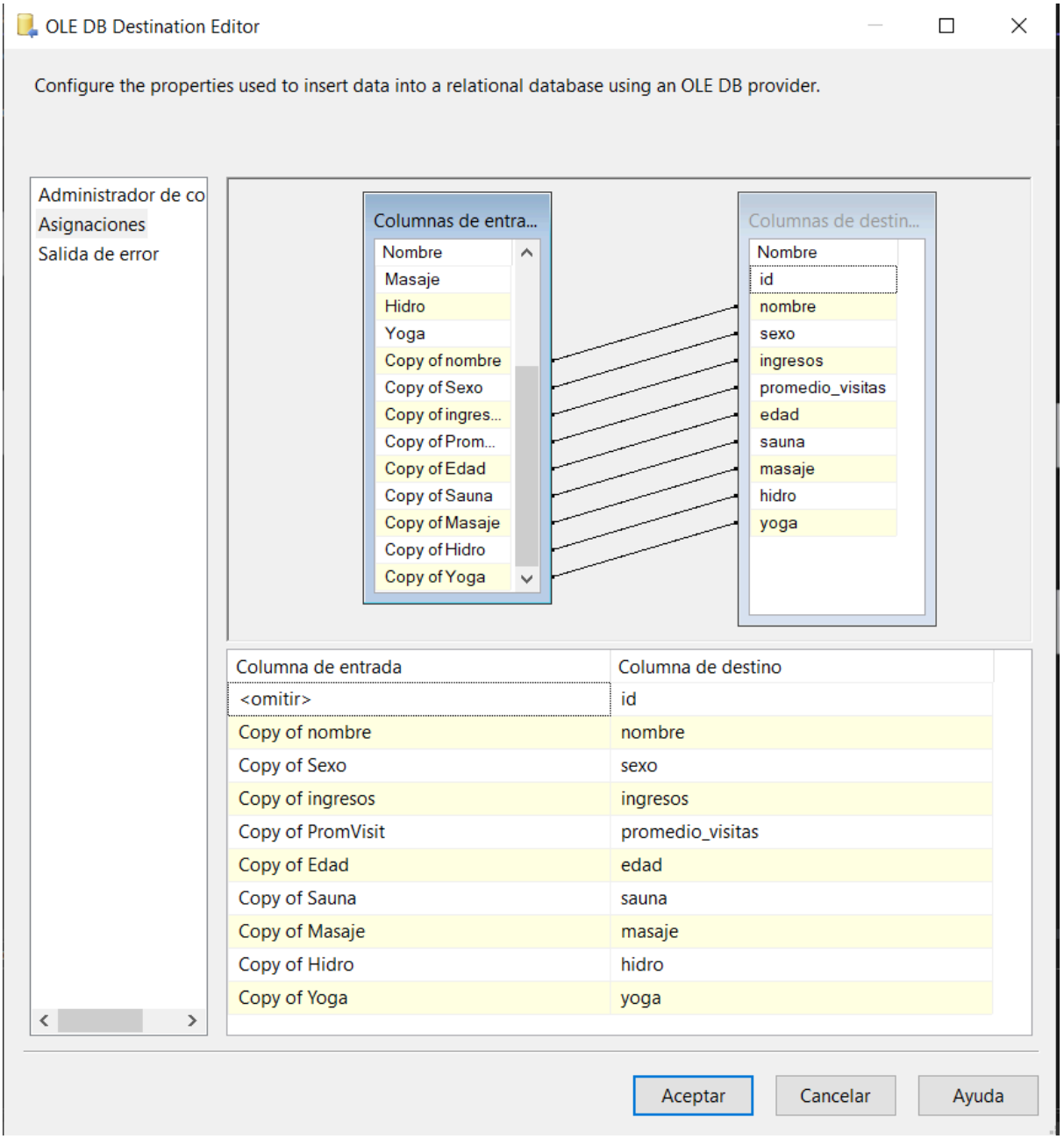
Obtención de datos (Igual para los tres casos)



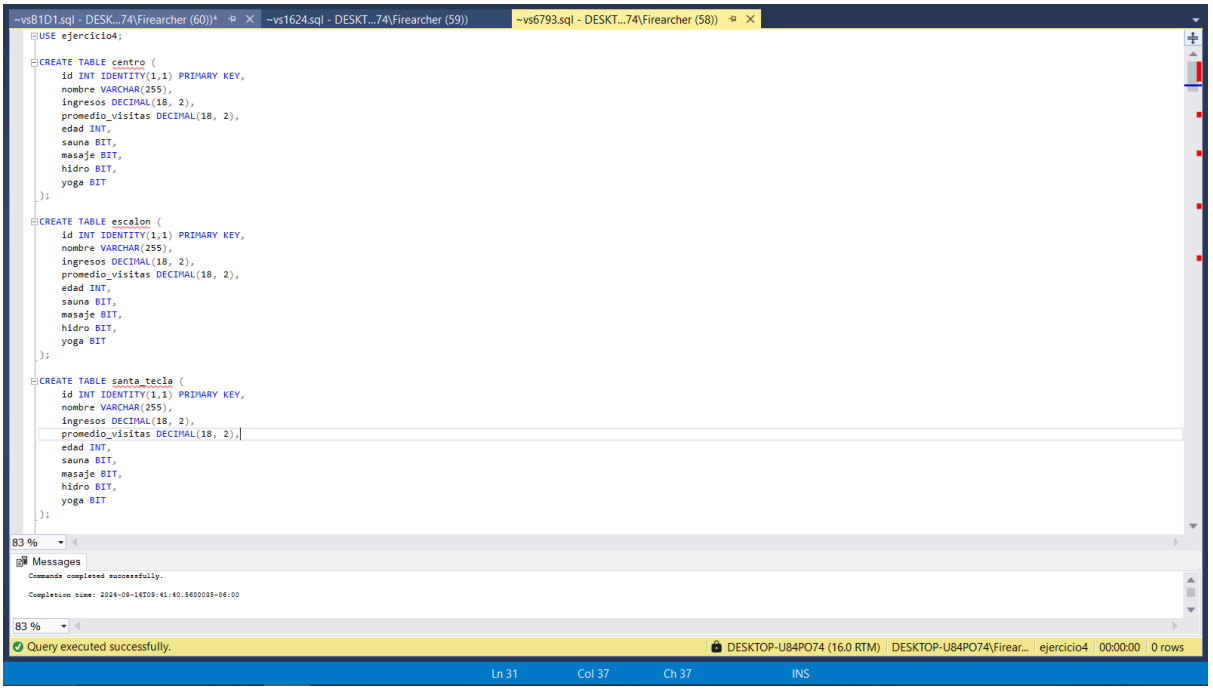
Conversión de datos (Igual en los tres casos)



Asignación de datos (Idéntico en los tres procesos)



Creación de tablas



```
USE ejercicio4;

CREATE TABLE centro (
    id INT IDENTITY(1,1) PRIMARY KEY,
    nombre VARCHAR(255),
    ingresos DECIMAL(18, 2),
    promedio_visitas DECIMAL(18, 2),
    edad INT,
    sauna BIT,
    masaje BIT,
    hidro BIT,
    yoga BIT
);

CREATE TABLE escalon (
    id INT IDENTITY(1,1) PRIMARY KEY,
    nombre VARCHAR(255),
    ingresos DECIMAL(18, 2),
    promedio_visitas DECIMAL(18, 2),
    edad INT,
    sauna BIT,
    masaje BIT,
    hidro BIT,
    yoga BIT
);

CREATE TABLE santa_tecla (
    id INT IDENTITY(1,1) PRIMARY KEY,
    nombre VARCHAR(255),
    ingresos DECIMAL(18, 2),
    promedio_visitas DECIMAL(18, 2),
    edad INT,
    sauna BIT,
    masaje BIT,
    hidro BIT,
    yoga BIT
);
```

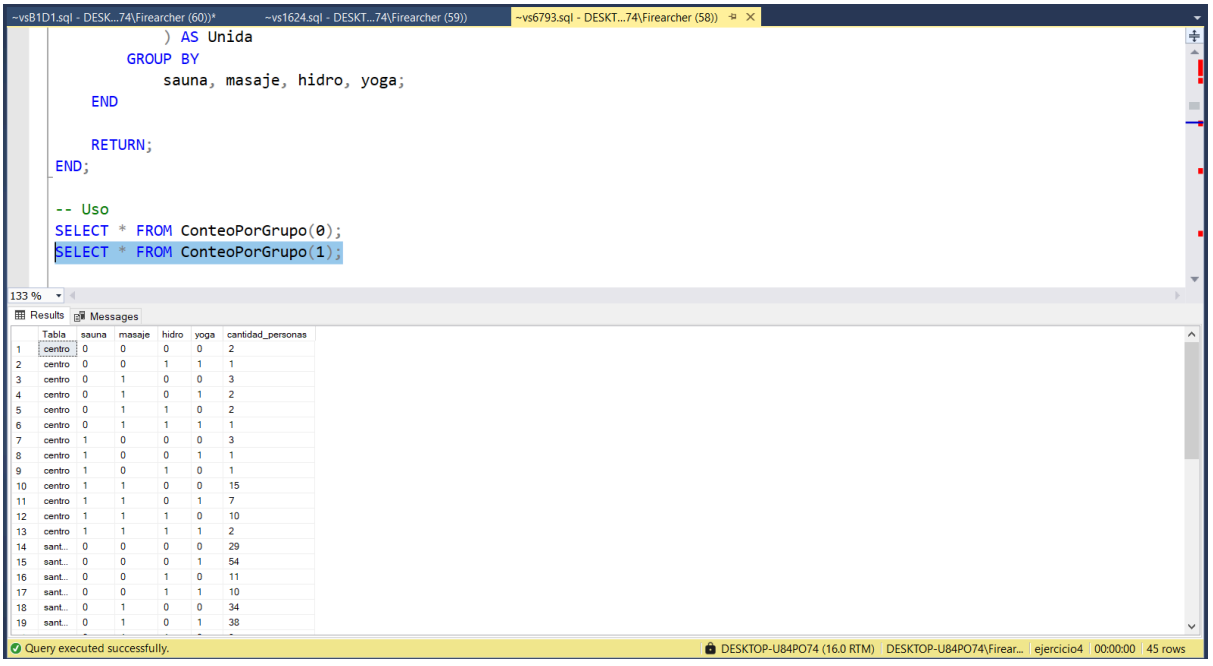
```
Commands completed successfully.

Completion time: 2024-09-14T09:41:40.5600035-06:00
```

Query executed successfully. DESKTOP-U84PO74 (16.0 RTM) DESKTOP-U84PO74\Firear... ejercicio4 00:00:00 0 rows

Ln 31 Col 37 Ch 37 INS

Ejecución de procedimientos

Conteo por grupo



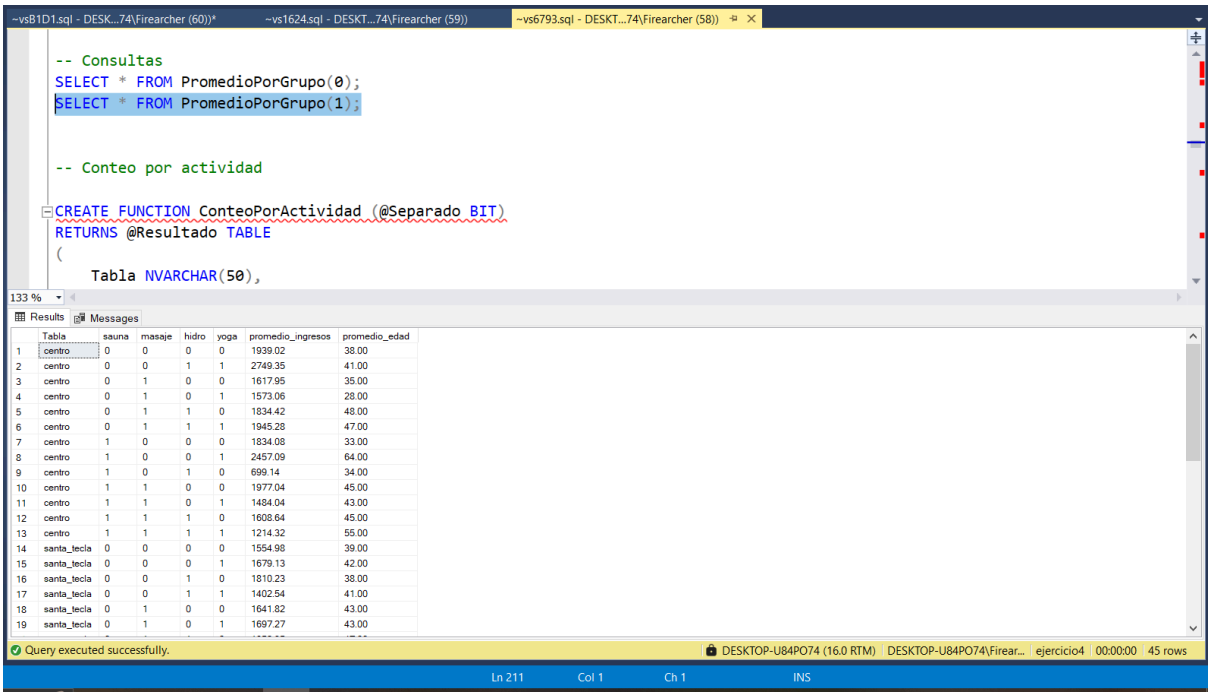
```
        ) AS Unida
    GROUP BY
        sauna, masaje, hidro, yoga;
END

RETURN;
END;

-- Uso
SELECT * FROM ConteoPorGrupo(0);
SELECT * FROM ConteoPorGrupo(1);
```

Results Messages

	Tabla	sauna	masaje	hidro	yoga	cantidad_personas
1	centro	0	0	0	0	2
2	centro	0	0	1	1	1
3	centro	0	1	0	0	3
4	centro	0	1	0	1	2
5	centro	0	1	1	0	2
6	centro	0	1	1	1	1
7	centro	1	0	0	0	3
8	centro	1	0	0	1	1
9	centro	1	0	1	0	1
10	centro	1	1	0	0	15
11	centro	1	1	0	1	7
12	centro	1	1	1	0	10
13	centro	1	1	1	1	2
14	sant...	0	0	0	0	29
15	sant...	0	0	0	1	54
16	sant...	0	0	1	0	11
17	sant...	0	0	1	1	10
18	sant...	0	1	0	0	34
19	sant...	0	1	0	1	38

Query executed successfully. DESKTOP-U84PO74 (16.0 RTM) DESKTOP-U84PO74\Firear... ejercicio4 00:00:00 45 rows

Promedio por grupos



```
-- Consultas
SELECT * FROM PromedioPorGrupo(0);
SELECT * FROM PromedioPorGrupo(1);


-- Conteo por actividad

CREATE FUNCTION ConteoPorActividad (@Separado BIT)
RETURNS @Resultado TABLE
(
    Tabla NVARCHAR(50),
```

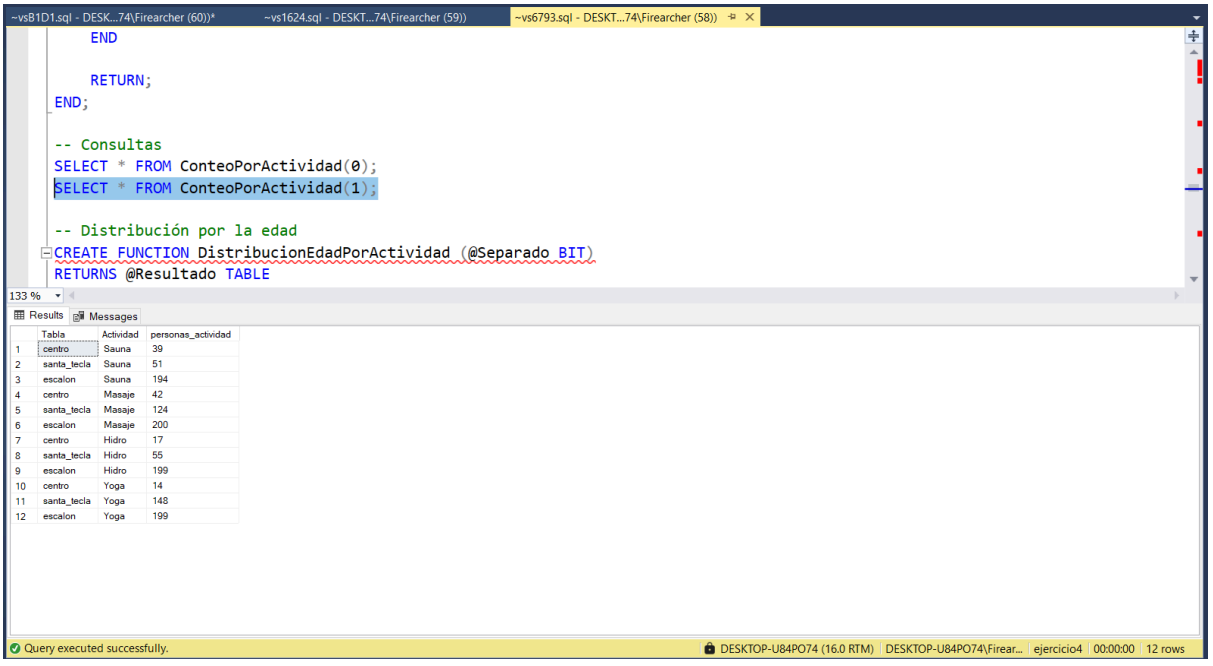
Results | Messages

	Tabla	sauna	masaje	hidro	yoga	promedio_ingresos	promedio_edad
1	centro	0	0	0	0	1939.02	38.00
2	centro	0	0	1	1	2749.35	41.00
3	centro	0	1	0	0	1617.95	35.00
4	centro	0	1	0	1	1573.06	28.00
5	centro	0	1	1	0	1834.42	48.00
6	centro	0	1	1	1	1945.28	47.00
7	centro	1	0	0	0	1834.08	33.00
8	centro	1	0	0	1	2457.09	64.00
9	centro	1	0	1	0	699.14	34.00
10	centro	1	1	0	0	1977.04	45.00
11	centro	1	1	0	1	1484.04	43.00
12	centro	1	1	1	0	1608.64	45.00
13	centro	1	1	1	1	1214.32	55.00
14	santa_tecla	0	0	0	0	1554.98	39.00
15	santa_tecla	0	0	0	1	1679.13	42.00
16	santa_tecla	0	0	1	0	1810.23	38.00
17	santa_tecla	0	0	1	1	1402.54	41.00
18	santa_tecla	0	1	0	0	1641.82	43.00
19	santa_tecla	0	1	0	1	1697.27	43.00

Query executed successfully. DESKTOP-U84PO74 (16.0 RTM) DESKTOP-U84PO74\Firear... ejercicio4 00:00:00 45 rows

Ln 211 Col 1 Ch 1 INS

Conteo por actividad



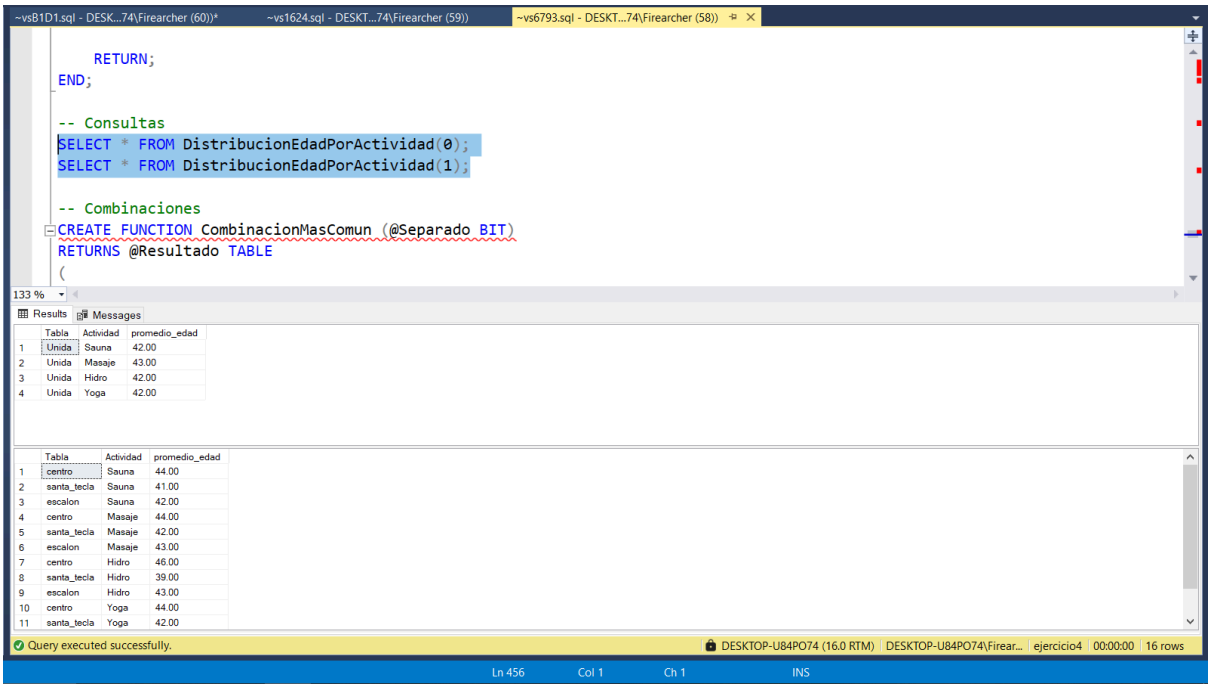
```
    END

    RETURN;
END;

-- Consultas
SELECT * FROM ConteoPorActividad(0);
SELECT * FROM ConteoPorActividad(1);

-- Distribución por la edad
CREATE FUNCTION DistribucionEdadPorActividad (@Separado BIT)
RETURNS @Resultado TABLE
```

Results | Messages

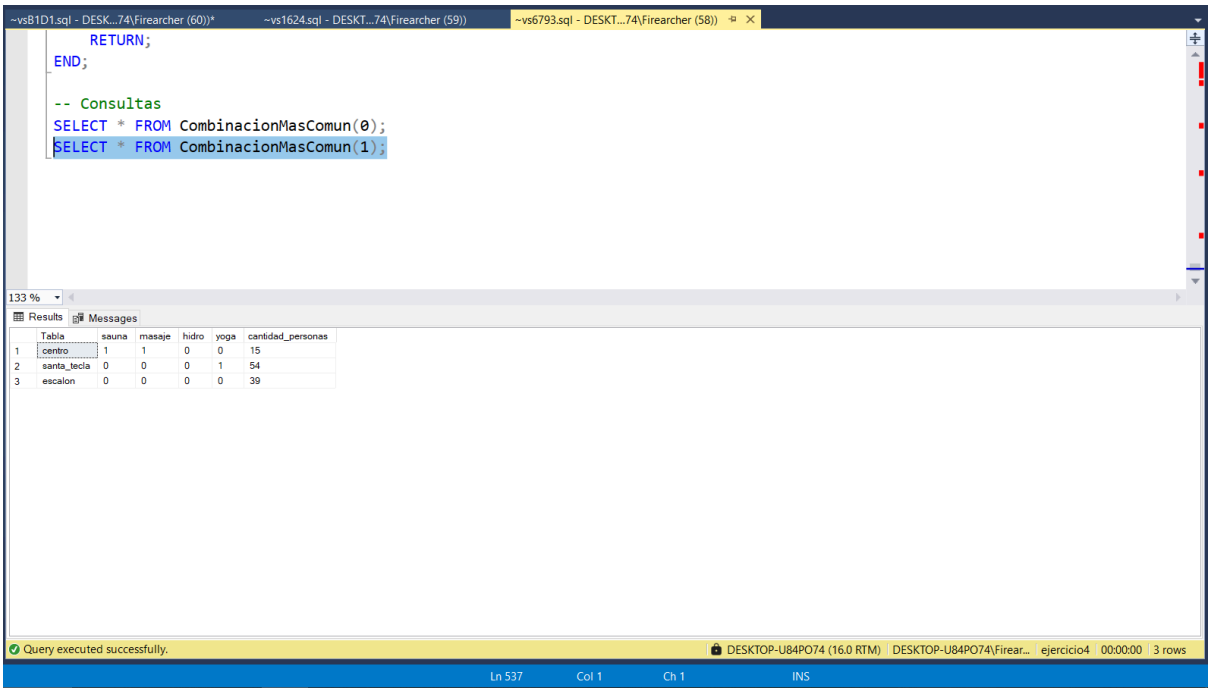
	Tabla	Actividad	personas_actividad
1	centro	Sauna	39
2	santa_tecla	Sauna	51
3	escalon	Sauna	194
4	centro	Masaje	42
5	santa_tecla	Masaje	124
6	escalon	Masaje	200
7	centro	Hidro	17
8	santa_tecla	Hidro	55
9	escalon	Hidro	199
10	centro	Yoga	14
11	santa_tecla	Yoga	148
12	escalon	Yoga	199

Query executed successfully. DESKTOP-U84PO74 (16.0 RTM) DESKTOP-U84PO74\Firear... ejercicio4 00:00:00 12 rows

Distribución por actividad



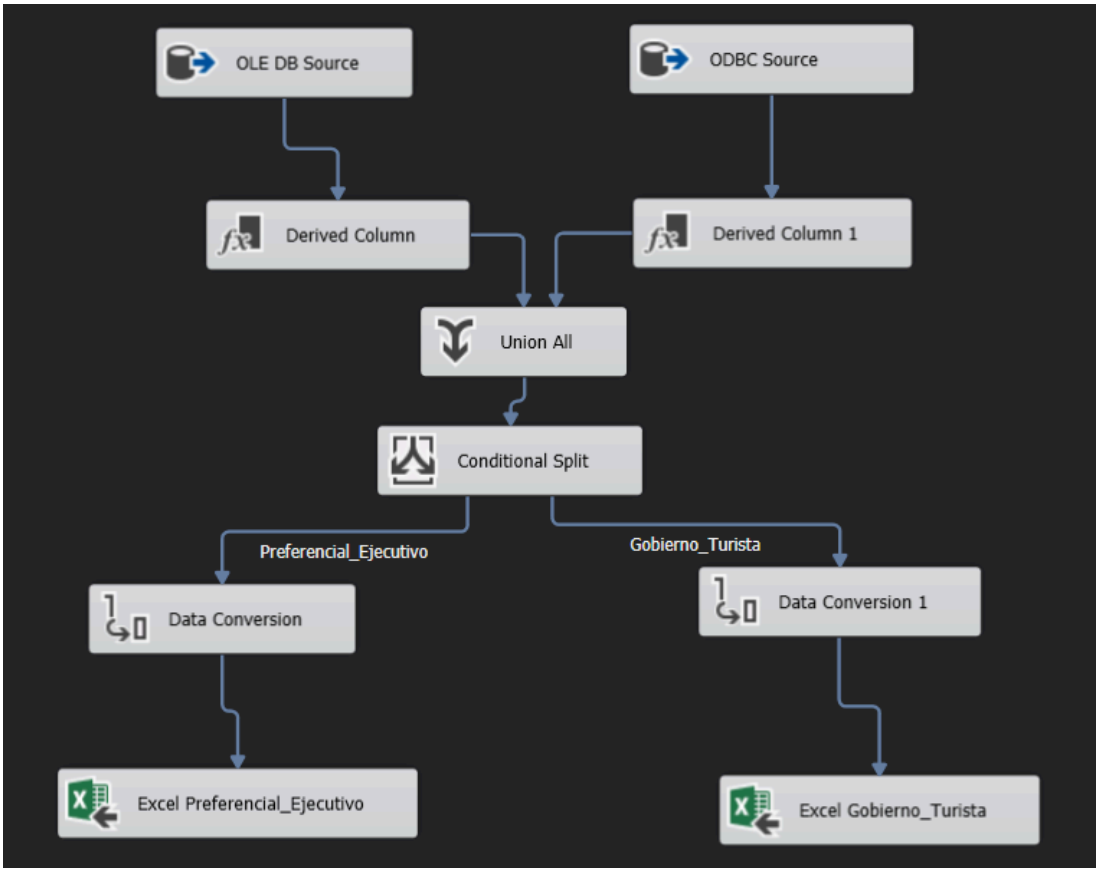
Por combinaciones



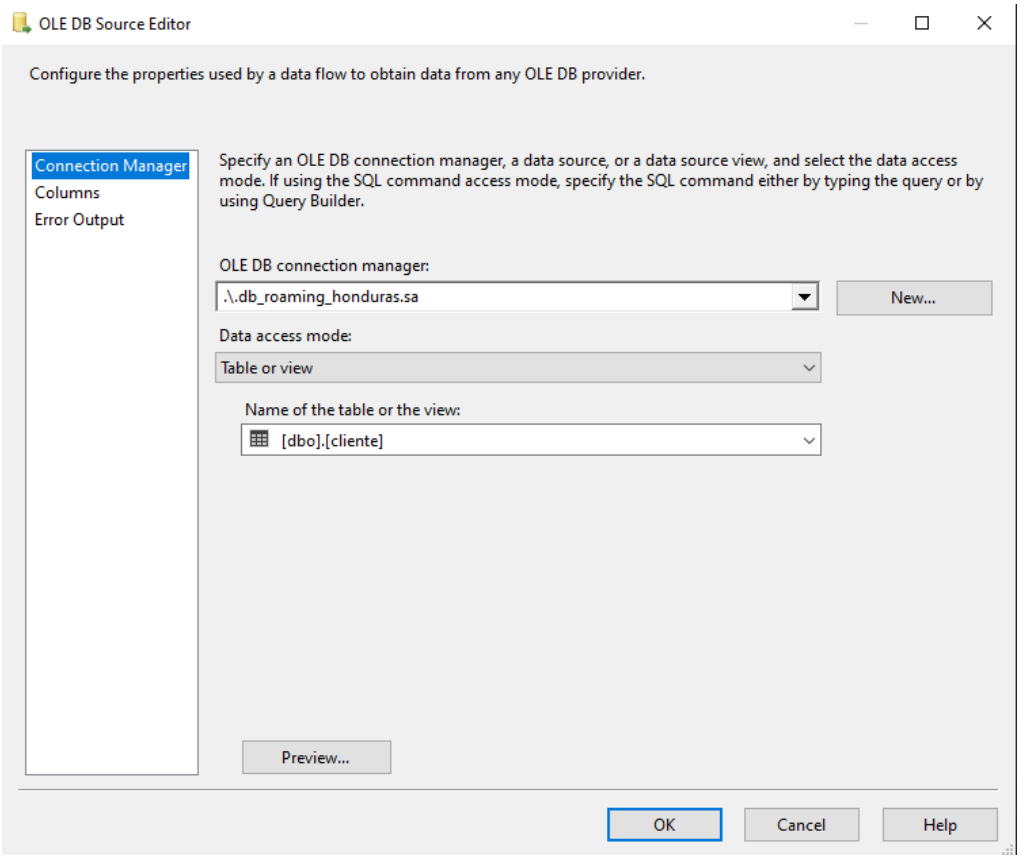
Ejercicio 5

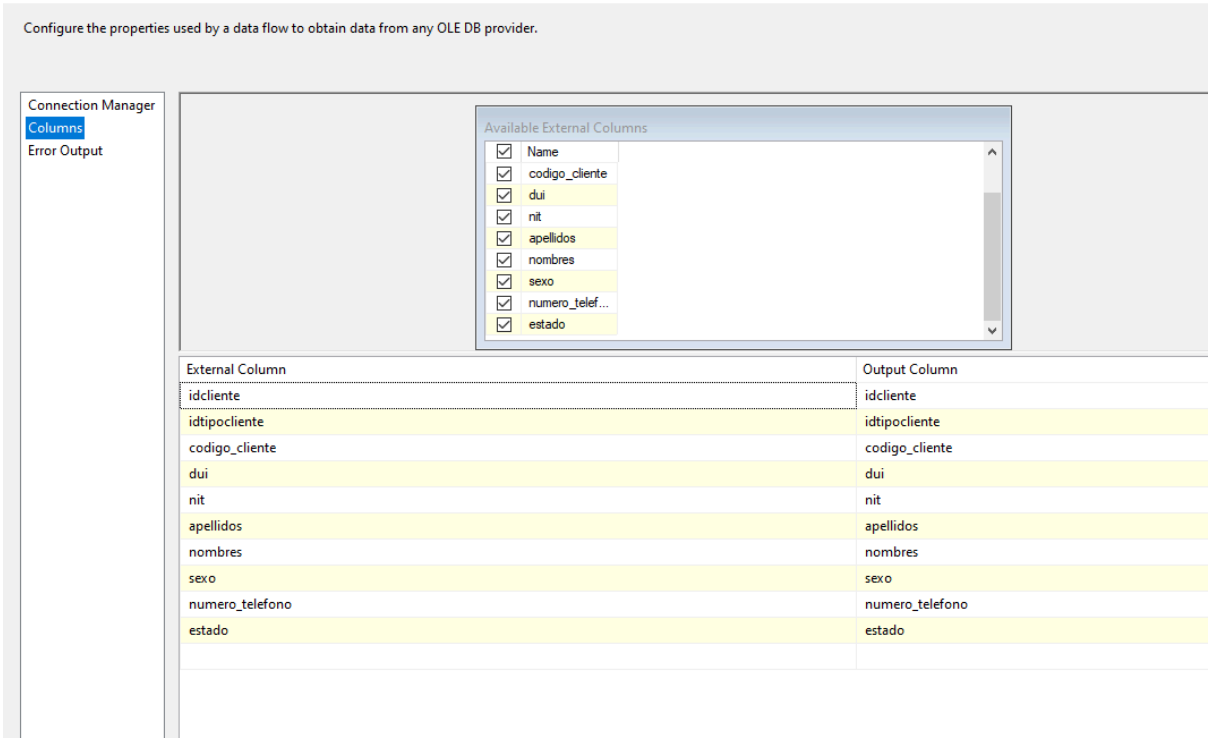
Ejercicio 6

- Data flow general

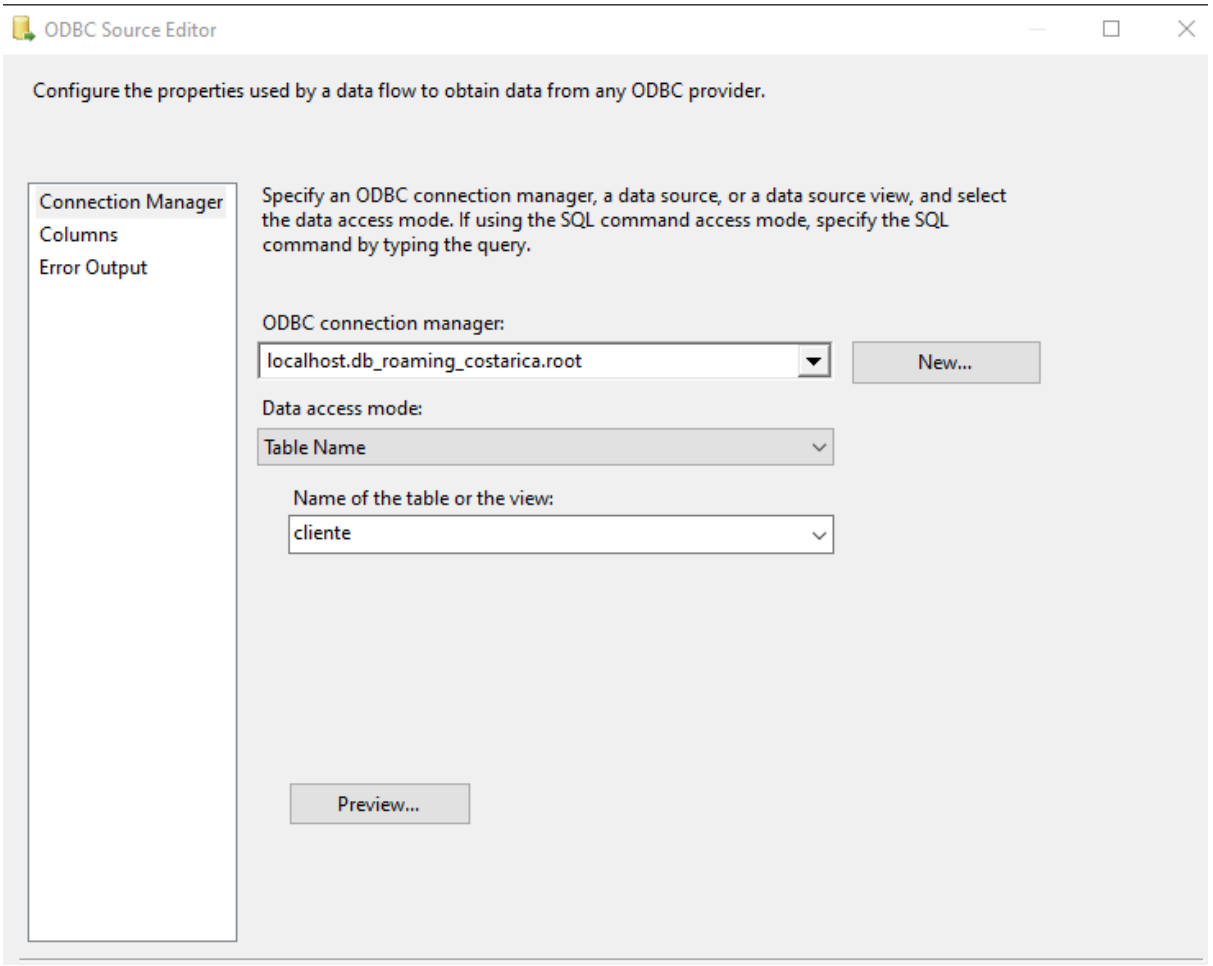


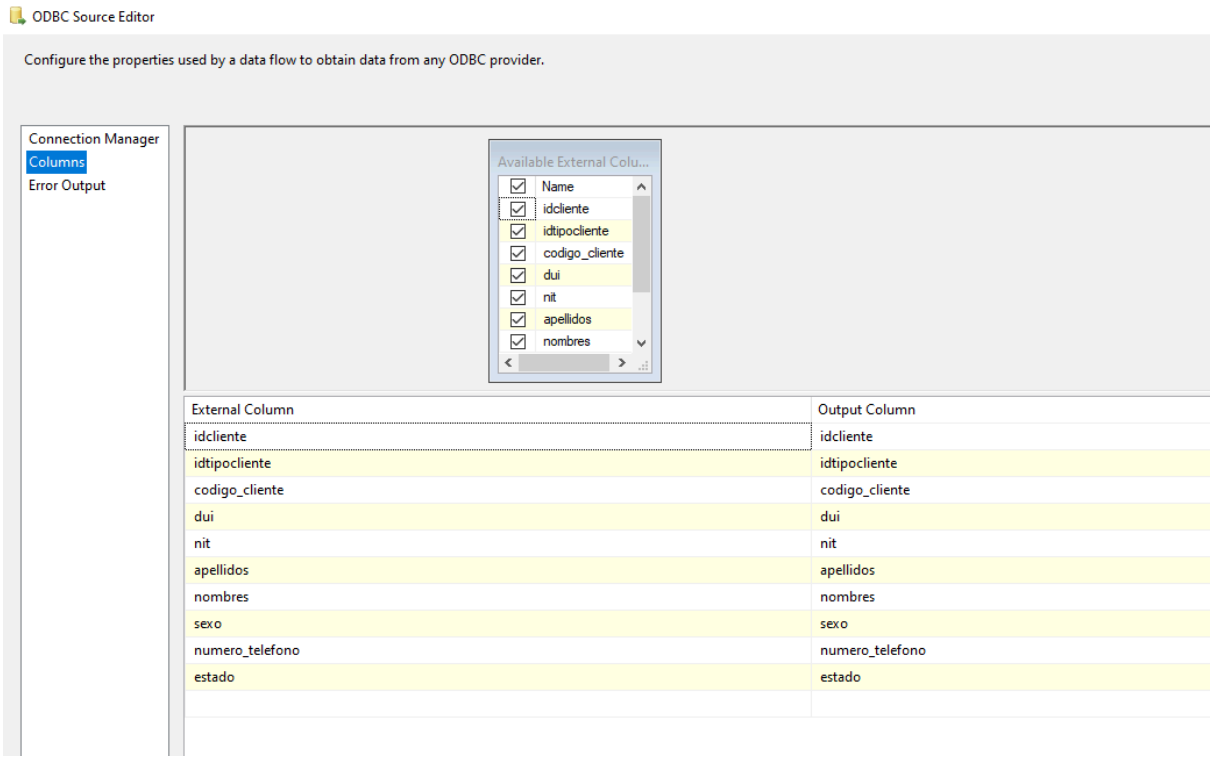
- SQL SERVER



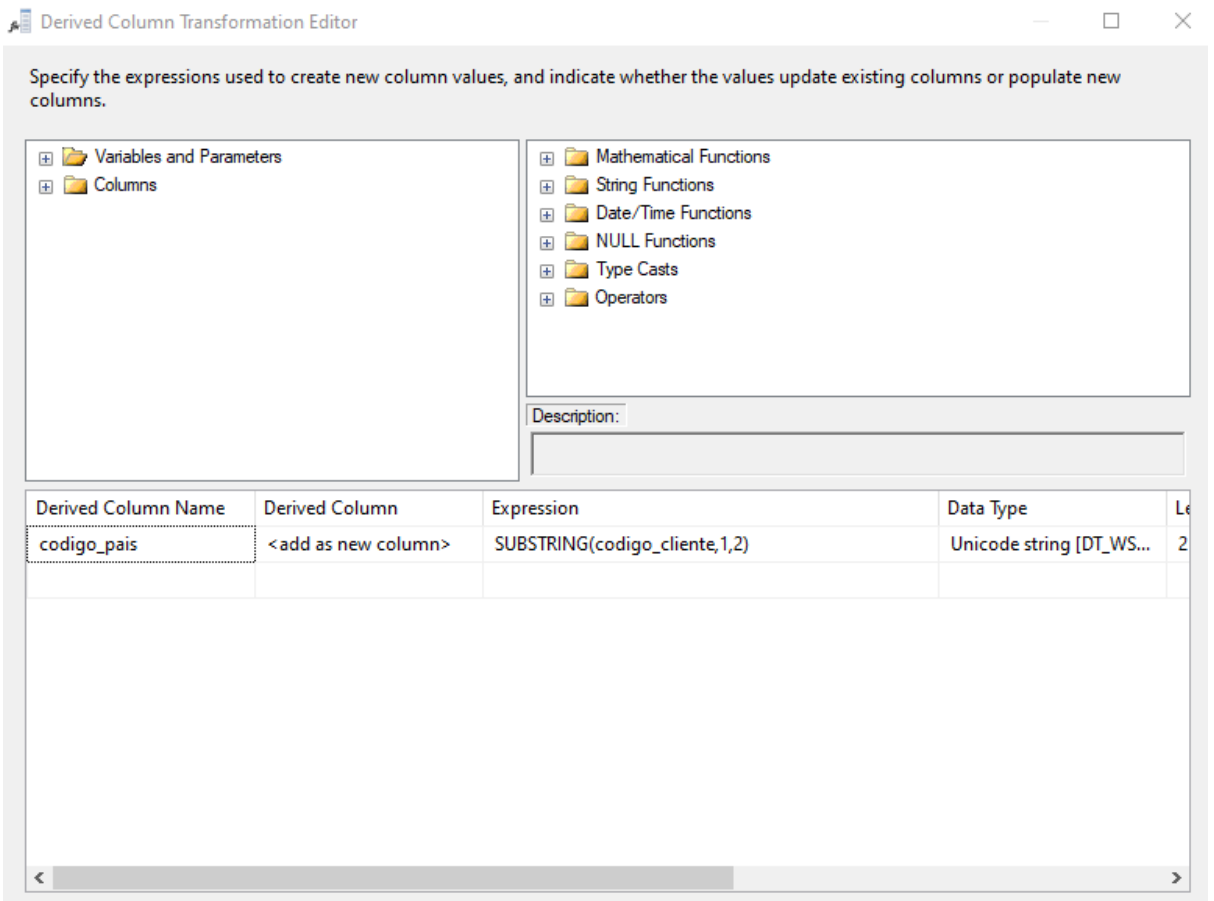


- MYSQL SERVER



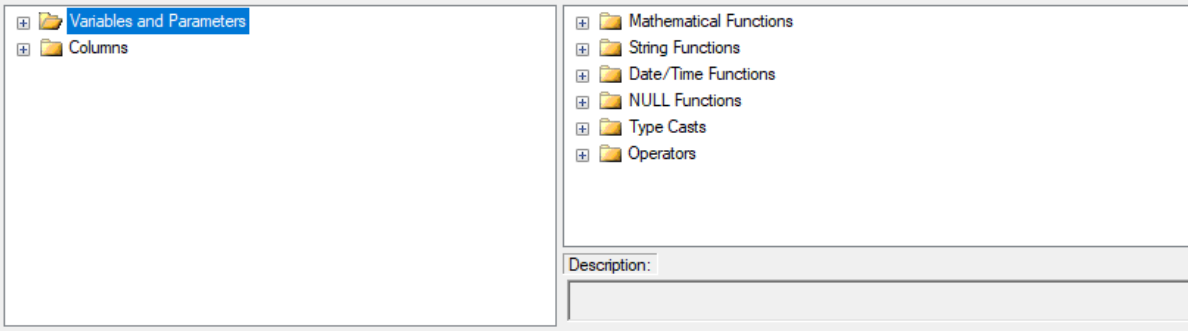


- Derived column Transformation SQL y MYSQL.



Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.



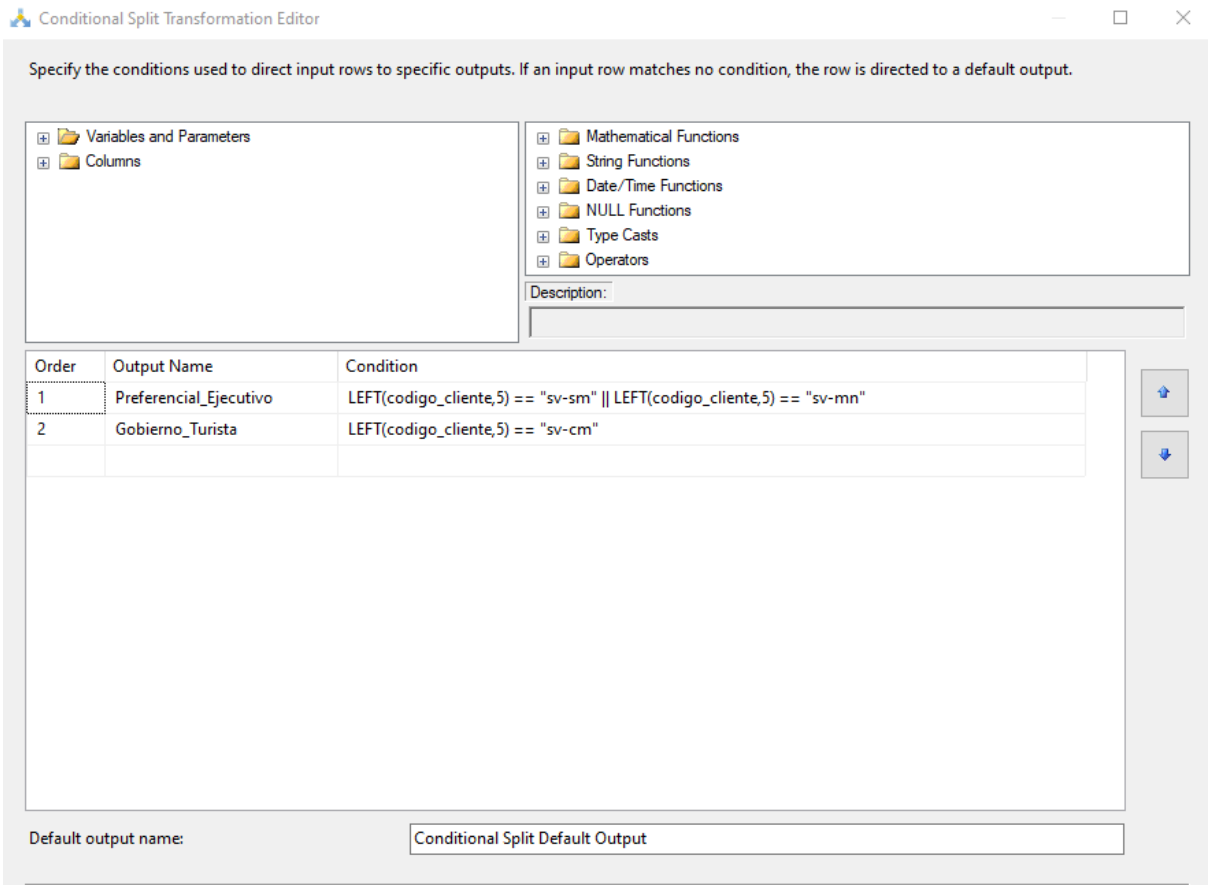
Derived Column Name	Derived Column	Expression	Data Type	Length
codigo_cliente	<add as new column>	SUBSTRING(codigo_cliente,1,2)	Unicode string [DT_WS...	2

Union All Transformation Editor

Configure the properties used to merge multiple inputs into one output by creating mappings between columns.

Output Column Name	Union All Input 1	Union All Input 2
idcliente	idcliente	idcliente
idtipocliente	idtipocliente	idtipocliente
codigo_cliente	codigo_cliente	ODBC Source.codigo_cliente
dui	dui	dui
nit	nit	nit
apellidos	apellidos	apellidos
nombres	nombres	nombres
sexo	sexo	sexo
numero_telefono	numero_telefono	numero_telefono
estado	estado	estado
codigo_pais	codigo_pais	Derived Column 1.codigo_cliente

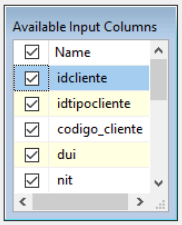
- Conditional Split



- Data Conversion.

Data Conversion Transformation Editor

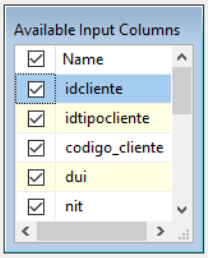
Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code pa



Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
idcliente	Copy of idcliente	Unicode string [DT_WSTR]	50			
idtipocliente	Copy of idtipocliente	Unicode string [DT_WSTR]	50			
codigo_cliente	Copy of codigo_cliente	Unicode string [DT_WSTR]	50			
dui	Copy of dui	Unicode string [DT_WSTR]	9			
nit	Copy of nit	Unicode string [DT_WSTR]	14			
apellidos	Copy of apellidos	Unicode string [DT_WSTR]	250			
nombres	Copy of nombres	Unicode string [DT_WSTR]	250			
sexo	Copy of sexo	Unicode string [DT_WSTR]	1			
numero_telefono	Copy of numero_telef...	Unicode string [DT_WSTR]	50			
estado	Copy of estado	Unicode string [DT_WSTR]	1			
codigo_pais	Copy of codigo_pais	Unicode string [DT_WSTR]	2			

Data Conversion Transformation Editor

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, p
the column.



Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
idcliente	Copy of idcliente	Unicode string [DT_WSTR]	50			
idtipocliente	Copy of idtipocliente	Unicode string [DT_WSTR]	50			
codigo_cliente	Copy of codigo_cliente	Unicode string [DT_WSTR]	15			
dui	Copy of dui	Unicode string [DT_WSTR]	9			
nit	Copy of nit	Unicode string [DT_WSTR]	14			
apellidos	Copy of apellidos	Unicode string [DT_WSTR]	250			
nombres	Copy of nombres	Unicode string [DT_WSTR]	250			
sexo	Copy of sexo	Unicode string [DT_WSTR]	1			
numero_telefono	Copy of numero_telef...	Unicode string [DT_WSTR]	8			
estado	Copy of estado	Unicode string [DT_WSTR]	1			
codigo_pais	Copy of codigo_pais	Unicode string [DT_WSTR]	2			

- Excel Destination.