

# **UNIVERSIDAD DON BOSCO**



## **DIRECCIÓN DE EDUCACIÓN A DISTANCIA**

**Asignatura: Datawarehouse y Minería de Datos**

**Proyecto de Cátedra Fase 1**

**Presentado por:**

**Rigoberto Alcides Villalta - VV000329**

**El Salvador, 13 de noviembre de 2022**

## Enunciado de proyecto

El ministerio de obras publicas(MOP), ha recolectado los datos de las esquelas impuestas a los conductores de tránsito durante el año 2018 y no sabe como presentar el informe de esto al gobierno central para el cierre del año.

Por tal motivo a contratado un grupo de consultores para que analicen la información y proporcione un documento con el análisis que realizaron, solo tiene una restricción el análisis deben de ocupar dos estrategias por tema de minería de datos, puede ser cualquier de las descritas.

Se le pide elaborar dos estrategias de minería de datos para estos datos:

- Cubos OLAP
- Reporting Services

Y presentar la implementación de estas estrategias.

NOTA: Ing. No pude terminar ninguno de los dos puntos, estuve mucho tiempo primero peleando para hacer un script para normalizar los datos a una base de datos, probablemente lo hubiera hecho mejor con Calc o Excel, pero quería hacerlo con T-SQL. Le dejo muchos pantallazos de evidencia.

Luego no se porque el Cubo no se conectaba a la base. Pero adjunto los proyectos y lo hecho, espero mejorar drásticamente para la segunda fase.

## Desarrollo

Dado que tenemos un solo archivo separado por comas, utilizaremos el programa de hoja de cálculo LibreOffice Calc para hacer alguna limpieza de los datos antes de utilizarlo.

Primero lo importamos, aprovechando el asistente de importación de LibreOffice Calc eliminamos espacios innecesarios, asignamos tipos de datos y la codificación de caracteres:

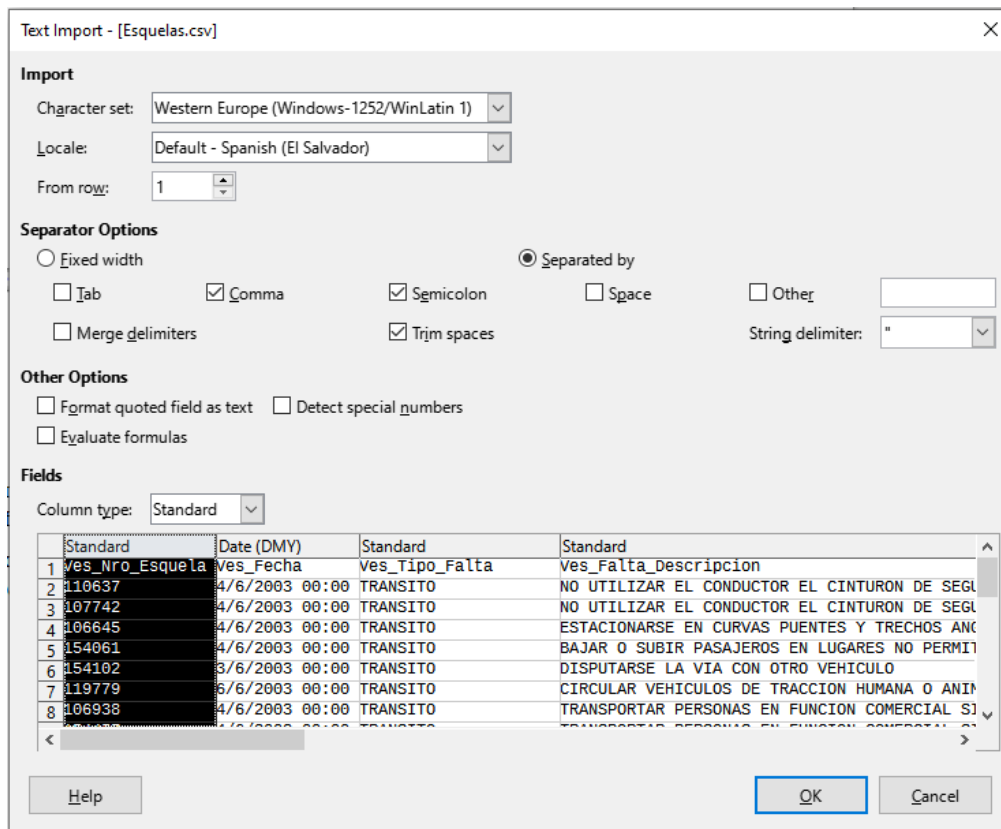


Imagen 1: Importación de los datos en LibreOffice Calc

Ahora utilizaremos la interfaz del programa para las siguientes normalizaciones:

1. Eliminaremos la columna “Estado”, ya que la columna con descripción nos da una mejor perspectiva de la información.
2. Cambiamos los encabezados para remover el prefijo: “Ves\_”
3. Eliminamos la parte que corresponde a la hora en las fechas, ya que esta no aporta información y no se encuentra en todas las filas
4. Filtramos los datos para evitar filas vacías en las columnas: Departamento, Estado\_Descripción y Valor, ya que estos generan inconsistencias y no hay manera de recuperar esta información.
5. Filtramos otros datos inconsistentes que se pueden observar.

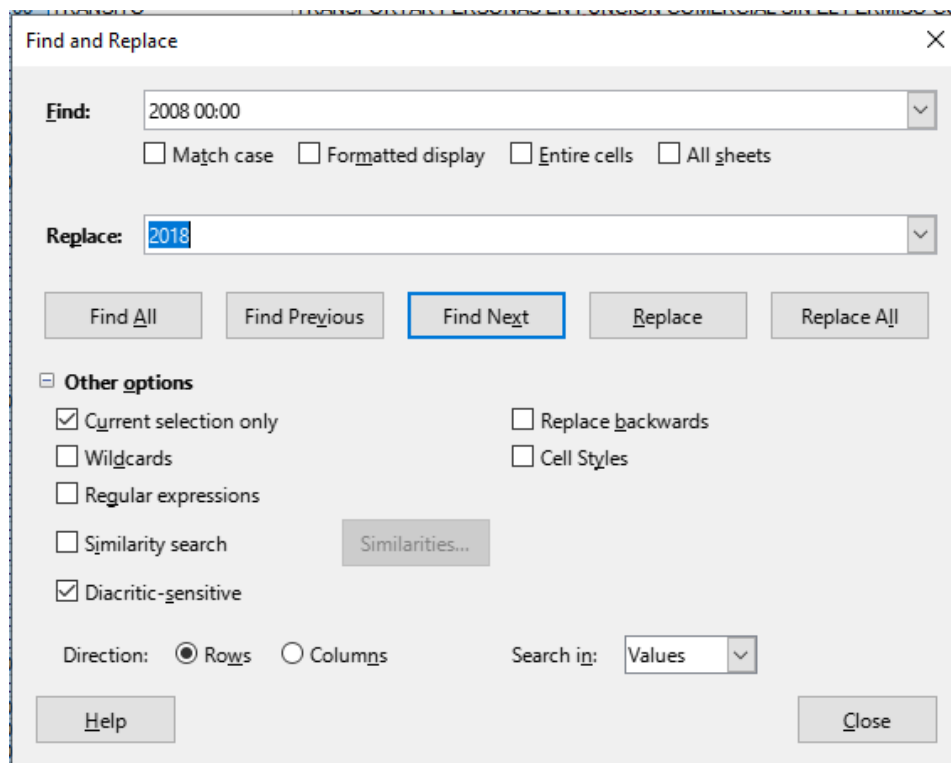


Imagen 2: Remplazo de datos

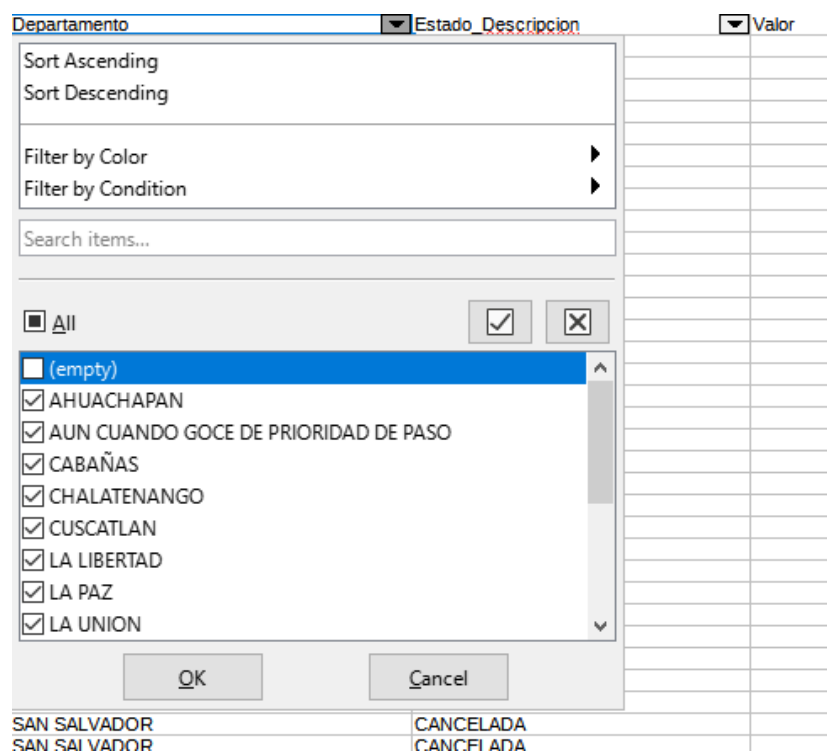


Imagen 3: Filtrado de Datos

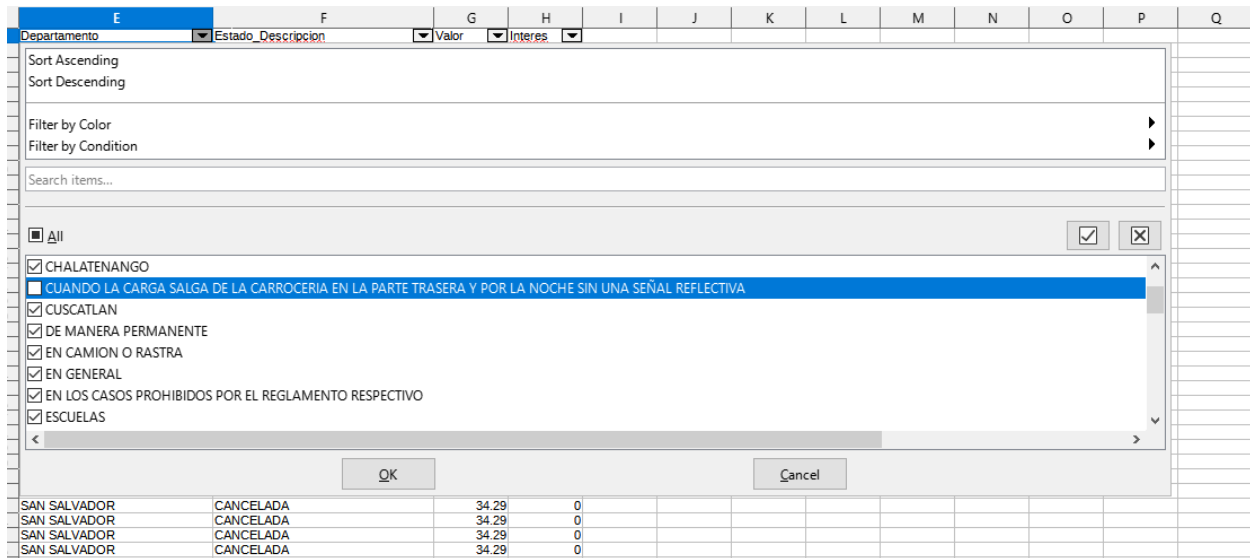


Imagen 4: Filtrado de datos inconsistentes en columna "Departamento"

Name	Date modified	Type	Size
Esquelas.csv	13/11/2022 17:57	CSV File	130,495 KB
Esquelas_normalizado.csv	13/11/2022 18:09	CSV File	104,883 KB

Imagen 5: Archivo normalizado reducido

Dado que SSIS se lleva mejor con los archivos de Excel, exportamos este archivo a ese formato:

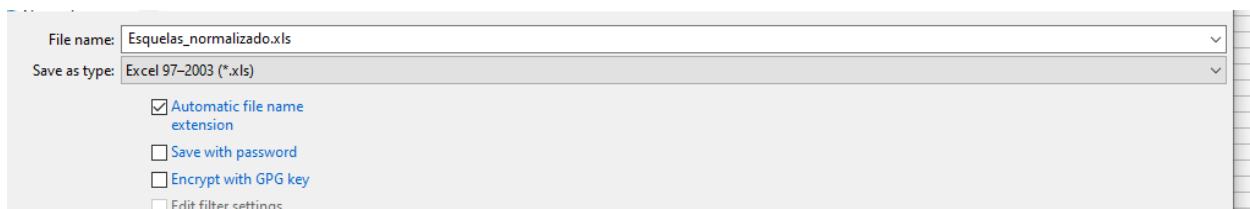


Imagen 6: Exportación de archivo a Excel

Ahora que tenemos el archivo normalizado podemos crear una base de datos normalizada.

## Creación de base de datos

Abrimos el SSMS y procedemos a crear una nueva base de datos:

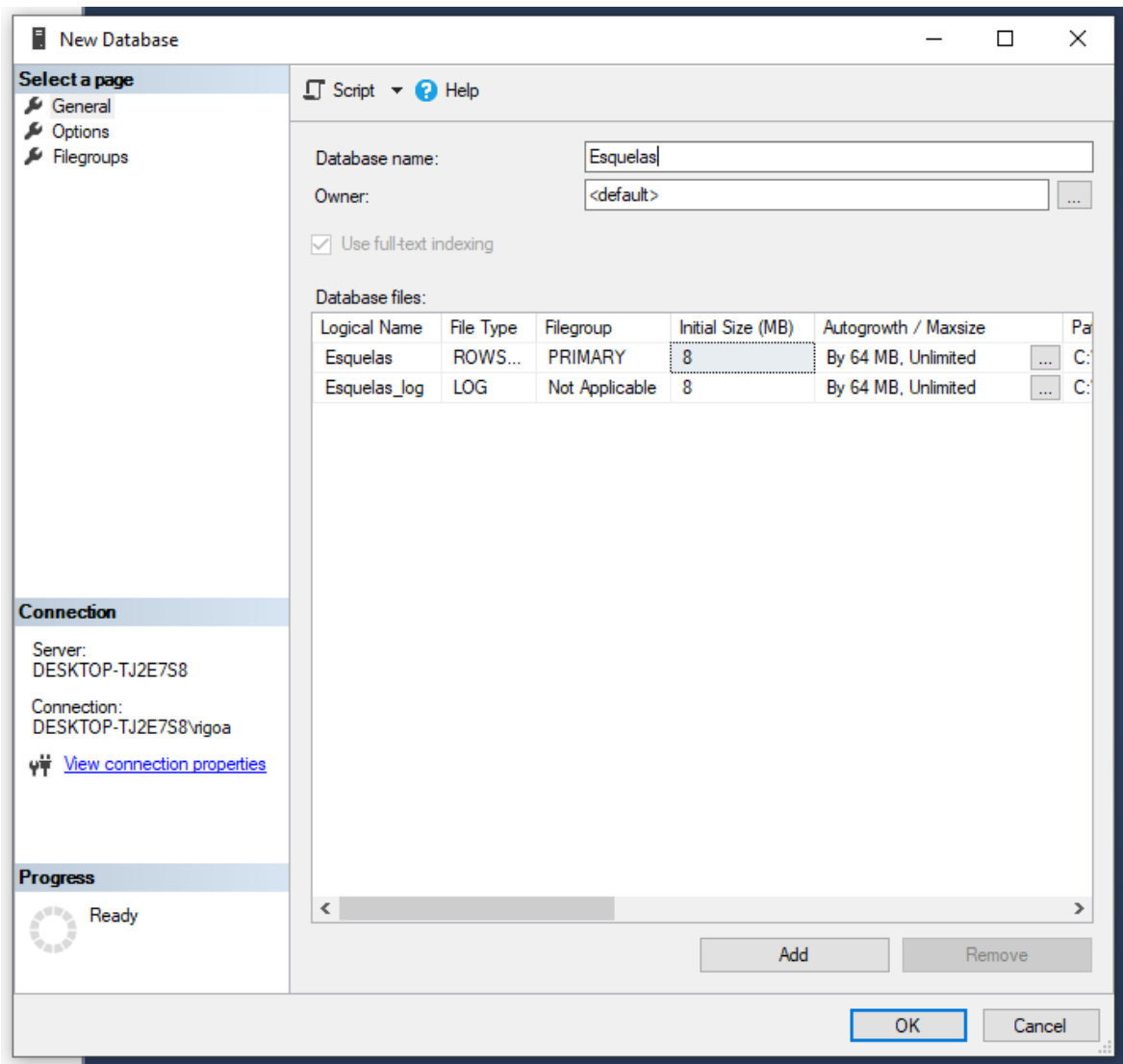


Imagen 7: Creación de la base de datos

Para poder tener normalizado nuestra base de datos necesitamos las siguientes tablas:

- Tipo de escuela: El tipo de escuela que es, la descripción de cada escuela y el monto que se impone por la infracción
- Departamento, con el departamento de El salvador
- Estado: El estado de trámite en que se encuentra la escuela
- Esquelas impuestas: el Número de escuela, la fecha, el tipo de escuela y su estado con llaves foráneas a cada tabla correspondiente.

Creamos las tablas con el siguiente comando:

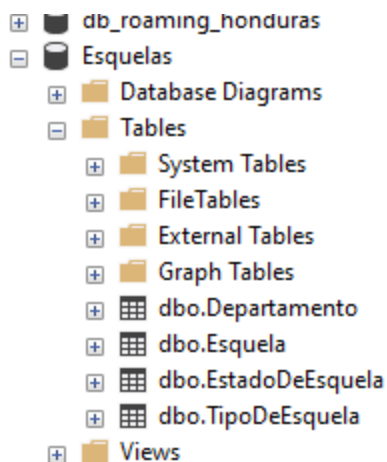
```
USE Esquelas
GO
CREATE TABLE Departamento
    (Id int NOT NULL PRIMARY KEY, Nombre nvarchar(25) NOT NULL);

CREATE TABLE TipoDeEsquela
    (Id int NOT NULL PRIMARY KEY, TipoDeFalta nvarchar(50) NOT NULL,
    Descripcion text NOT NULL, MontoInfraccion smallmoney NOT NULL);

CREATE TABLE EstadoDeEsquela
    ( Id int NOT NULL PRIMARY KEY, Estado nvarchar(50) NOT NULL);

CREATE TABLE Esquela
    (Id int NOT NULL PRIMARY KEY, TipoDeEsquela int NOT NULL,
    Departamento int NOT NULL, EstadoDeEsquela int NOT NULL,
    Interes smallmoney NOT NULL,
    FOREIGN KEY (TipoDeEsquela) REFERENCES TipoDeEsquela(id)
    ON DELETE CASCADE ON UPDATE CASCADE,
    FOREIGN KEY (Departamento) REFERENCES Departamento(id)
    ON DELETE CASCADE ON UPDATE CASCADE,
    FOREIGN KEY (EstadoDeEsquela) REFERENCES EstadoDeEsquela(id)
    ON DELETE CASCADE ON UPDATE CASCADE );
```

Y comprobamos que el Script haya corrido con éxito, las tablas se muestran:



*Imagen 8: Tablas creadas en la base de datos*

Me doy cuenta que voy a necesitar que los Id de cada tabla se incremente automaticamente, por lo que corremos el siguiente script para cada una:

```
ALTER TABLE Persons AUTO_INCREMENT=1
```

DESKTOP-TJ2E7S8.Es...asDB - dbo.Table_1* -> X			
	Column Name	Data Type	Allow Nulls
E7S8\rigor	NoEsquela	nchar(10)	<input type="checkbox"/>
	Fecha	date	<input type="checkbox"/>
	TipoFalta	nchar(35)	<input type="checkbox"/>
	Descripcion	text	<input type="checkbox"/>
	Departamento	nchar(20)	<input type="checkbox"/>
	Estado	nchar(20)	<input type="checkbox"/>
	Valor	smallmoney	<input type="checkbox"/>
	Interes	smallmoney	<input type="checkbox"/>
			<input type="checkbox"/>

- EsquelasDB
  - Database Diagrams
  - Tables
    - System Tables
    - FileTables
    - External Tables
    - Graph Tables
    - dbo.Esquelas
  - Views

SQL Server Import and Export Wizard

**Choose a Data Source**  
Select the source from which to copy data.

Data source: Microsoft Excel

Excel connection settings

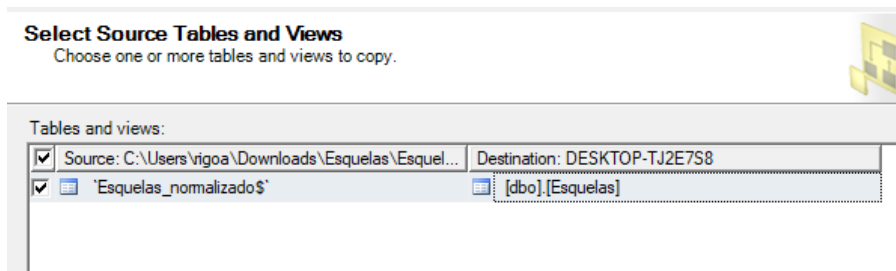
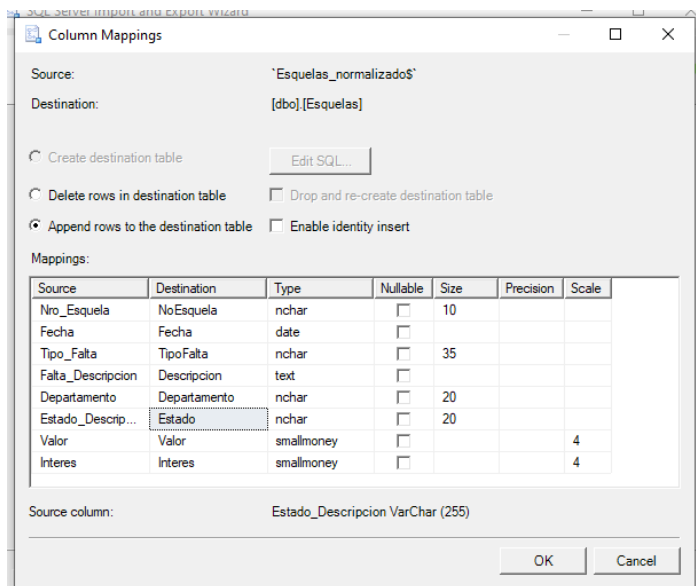
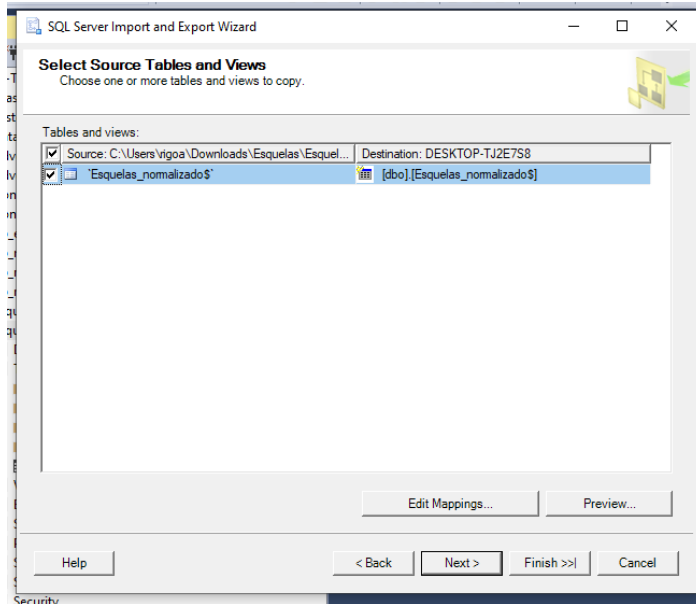
Excel file path: C:\Users\vigoa\Downloads\Esquelas\Esquelas\_normalizado.xls Browse...

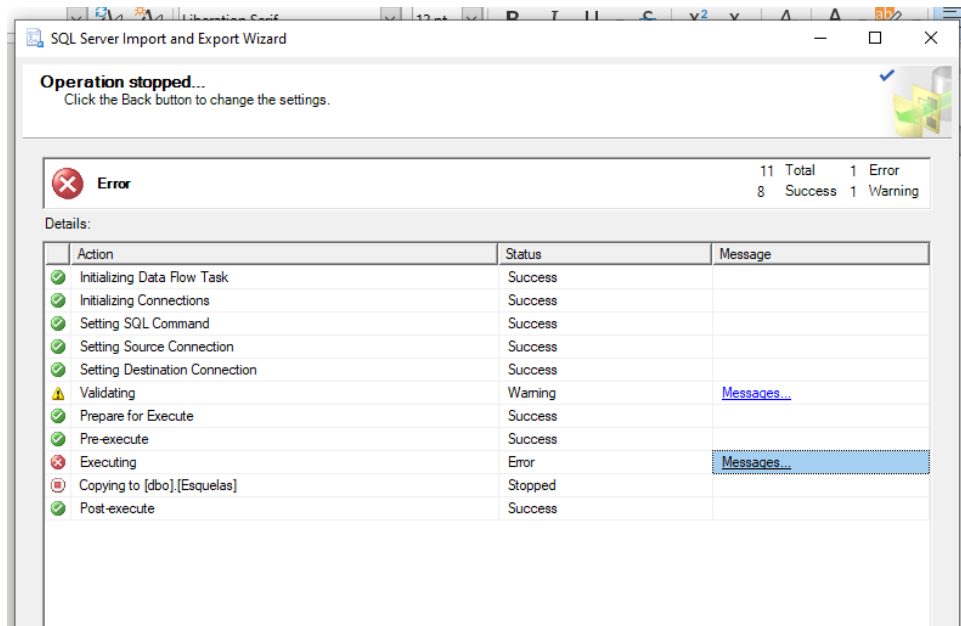
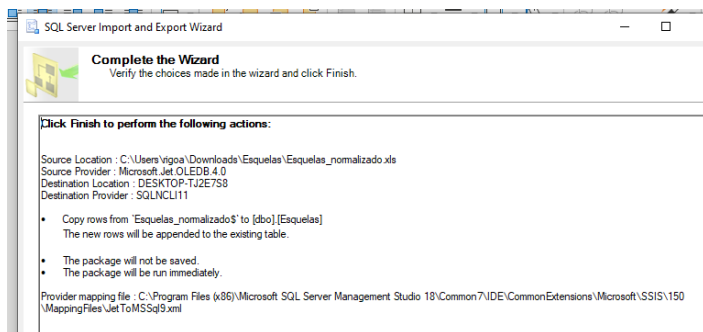
Excel version: Microsoft Excel 97-2003

☒ First row has column names

Help
< Back
Next >
Finish >>
Cancel







```

-- ***** SCRIPT FOR SELECT TOP N ROWS COMMAND *****
SELECT [NoEtiqueta]
      ,[Fecha]
      ,[TipoFalta]
      ,[Descripcion]
      ,[Departamento]
      ,[Estado]
      ,[Valor]
      ,[Interes]
FROM [EtiquetasDB].[dbo].[Etiquetas]

```

Results	Messages	NoEtiqueta	Fecha	TipoFalta	Descripcion	Departamento	Estado	Valor	Interes
3363		160296	2003-06-20	TRANSITO	NO RESPETAR SEÑALES DE PRECAUCION	SAN SALVADOR	CANCELADA	34.29	0.00
3364		160297	2003-06-22	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIMAL C...	SAN SALVADOR	CANCELADA	57.14	0.00
3365		160298	2003-06-22	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIMAL C...	SAN SALVADOR	CANCELADA	57.14	0.00
3366		160299	2003-06-23	TRANSITO	NO RESPETAR SEÑALES DE PRECAUCION	SAN SALVADOR	CANCELADA	34.29	0.00
3367		160300	2003-06-23	TRANSITO	NO RESPETAR SEÑALES DE PRECAUCION	SAN SALVADOR	CANCELADA	34.29	0.00
3368		160301	2003-06-11	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00
3369		160302	2003-06-18	TRANSITO	NEGARSE A PRESENTAR LOS DOCUMENTOS DE TRANSITO...	SAN SALVADOR	CANCELADA	34.29	0.00
3370		160376	2003-06-11	TRANSPORTE TERRESTRE	ESTACIONARSE MAS TIEMPO DE LO NECESARIO PARA SUBIR	SAN SALVADOR	CARGADA	11.43	7.31
3371		160377	2003-06-11	TRANSPORTE TERRESTRE	ESTACIONARSE MAS TIEMPO DE LO NECESARIO PARA SUBIR	SAN SALVADOR	CANCELADA	11.43	0.00
3372		160378	2003-06-11	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00
3373		160379	2003-06-11	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00
3374		160381	2003-06-13	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00
3375		160383	2003-06-13	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00
3376		160384	2003-06-13	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CARGADA	34.29	21.94
3377		160385	2003-06-13	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN E...	SAN SALVADOR	CANCELADA	34.29	0.00

## Cubo Olap

Iniciamos creando un proyecto de Business Intelligence, particularmente un Proyecto multidimensional y de minería de datos de Analysis Services.

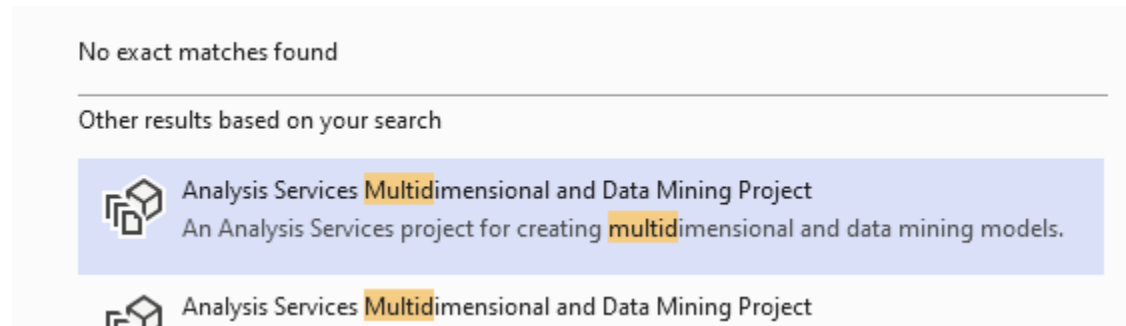


Imagen 9: Creación de proyecto Multidimensional

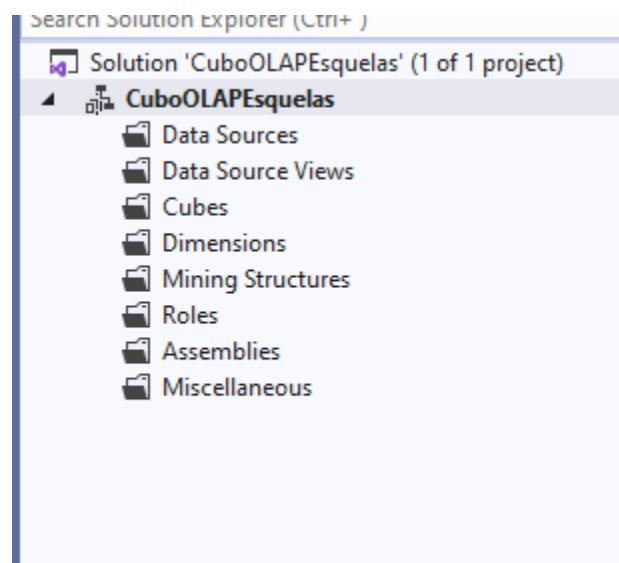


Imagen 10: Explorador de soluciones del proyecto Creado

Ahora agregamos una fuente de datos al proyecto:

Connection Manager

Provider: Native OLE DB\SQL Server Native Client 11.0

Server name: DESKTOP-TJ2E7S8 Refresh

Log on to the server

Authentication: Windows Authentication

User name: Password: Save my password

Connect to a database

Select or enter a database name: EsquelasDB

Attach a database file: Browse... Logical name:

Test Connection OK Cancel Help

Data Source Wizard

Select how to define the connection

You can select from a number of ways in which your data source will define its connection string.

Create a data source based on another object

Create a data source based on an existing or new connection

Data connections: DESKTOP-TJ2E7S8.EsquelasDB

Data connection properties:

Property	Value
Data Source	DESKTOP-TJ2E7S8
Initial Catalog	EsquelasDB
Integrated Se...	SSPI
Provider	SQLNCLI11.1

New... Delete

< Back Next > Finish >>| Cancel

Data Source Wizard

Completing the Wizard

Provide a name and then click Finish to create the new data source.

Data source name:

Esquelas DB

Preview:

Connection string:

Provider=SQLNCLI11.1;Data Source=DESKTOP-TJ2E7S8;Integrated Security=SSPI;Initial Catalog=EsquelasDB

< Back

Next >

Finish

Cancel

Cube Wizard

Select Creation Method

Cubes can be created by using existing tables, creating an empty cube, or generating tables in the data source.

How would you like to create the cube?

☐ Use existing tables

☒ Create an empty cube

☐ Generate tables in the data source

Template:

(None)

Description:

Create an empty cube. An empty cube is useful when users want to create everything manually, or when all dimensions are linked dimensions.

< Back

Next >

Finish >>|

Cancel

Cube Wizard

### Completing the Wizard

Name the cube, review its structure, and then click Finish to save the cube.

Cube name:

Preview:

< Back   Next >   **Finish**   Cancel


Cube Wizard

### Select Measure Group Tables

Select a data source view or diagram and then select the tables that will be used for measure groups.

Data source view:

Measure group tables:

☒  Esquelas

Suggest


< Back   **Next >**   Finish >> |   Cancel


Cube Wizard


### Select Measures


Select measures that you want to include in the cube.

☒ Measure

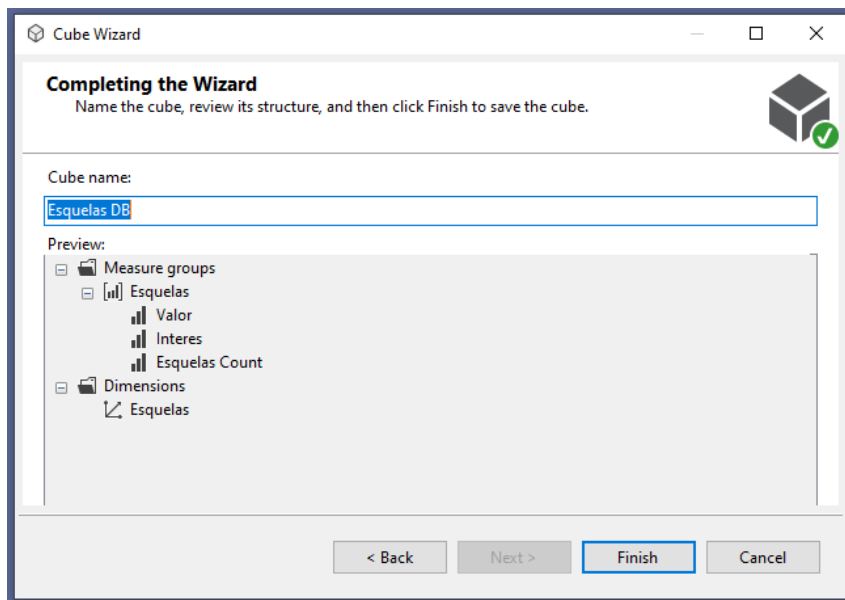
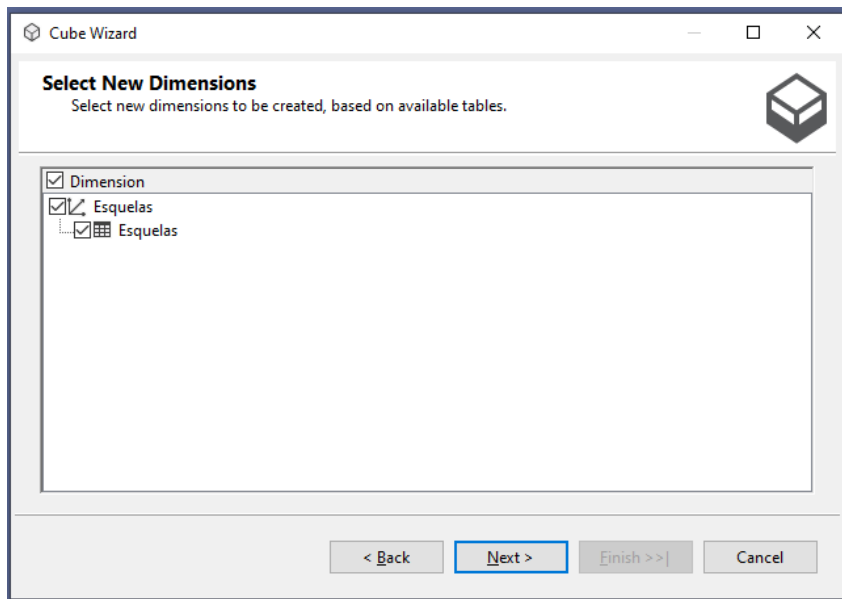
☒  Esquelas

☒  Valor

☒  Interes

☒  Esquelas Count

< Back   **Next >**   Finish >> |   Cancel



A connection cannot be made. Ensure that the server is running.



[Click here to see detailed error information](#)

[Click here to try loading the browser again](#)

# Reporting Services

## Create a new project

### Recent project templates


-  Analysis Services Multidimensional and Data Mining Project
-  Integration Services Project


Repor ✕ Clear all

All languages ▼ All platforms ▼ Stream Analytics ▼

No exact matches found


Other results based on your search

 **Report Server Project Wizard**  
Create a new **Report** Server project using **Report** Wizard.

 **Report Server Project**  
Create an empty **Report** Server project.

Not finding what you're looking for?  
[Install more tools and features](#)

Back Next

 Report Wizard

Select the Data Source

Select a data source from which to obtain data for this report or create a new data source.

☐ Shared data source

☒ New data source

Name:

DataSource1

Type:

Microsoft SQL Server

Connection string:

Edit...

Credentials...

Help

< Back

Next >

Finish >>|

Cancel



?

×

Connection Properties

Data source:  
Microsoft SQL Server (SqlClient)

Change...

Server name:  
DESKTOP-TJ2E7S8

Refresh

Log on to the server

Authentication: Windows Authentication

User name:Password:

Save my password

Connect to a database

Select or enter a database name:

EsqueiasDB

Attach a database file:

Browse...

Logical name:

Advanced...

Test Connection

OK

Cancel