

**UNIVERSIDAD DON BOSCO**  
**DIRECCIÓN DE EDUCACIÓN A DISTANCIA - UDB VIRTUAL**



**DATAWAREHOUSE Y MINERÍA DE DATOS DMD941 G01T**

**ACTIVIDAD**

PROYECTO FASE 1

**CATEDRATICO**

ING. KARENS MEDRANO

**PRESENTADO POR**

GARCÍA JUÁREZ, KELVIN VLADIMIR

GJ111587

**SOYAPANGO, EL SALVADOR, NOVIEMBRE 2022**

**Enlace GitHub: [https://github.com/kelvinjuarez/DMD941---](https://github.com/kelvinjuarez/DMD941---Proyecto-Fase-1)**  
**[Proyecto-Fase-1](https://github.com/kelvinjuarez/DMD941---Proyecto-Fase-1)**

## **Objetivo**

Aplicar las técnicas necesarias de minería de datos, con la finalidad de analizar la información para la ayuda de toma de decisiones estratégicas de una organización.

## **Marco teórico**

En la actualidad, la minería de datos se ubica en casi cualquier sector donde se pueda obtener información relevante de un volumen importante de datos. Las empresas son auxiliadas por esta actividad, pues siempre hay procesos o temas que se deben mejorar o solucionarse.

Permite además el mejor entendimiento de un negocio y la ejecución de acciones relevantes para guiar a una organización hacia el éxito. Las áreas de negocio más comunes en que se aplica la minería de datos son la segmentación de clientes, el marketing, etcétera.

## **Antecedentes**

El ministerio de obras públicas (MOP), ha recolectado mucha información sobre dos temas de suma importancia, pero no sabe cómo presentar los informes al gobierno central para el cierre del año 2018.

- Esquelas de infracción de tránsito
- Parque vehicular

Por tal motivo se ha 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.

## Situación actual

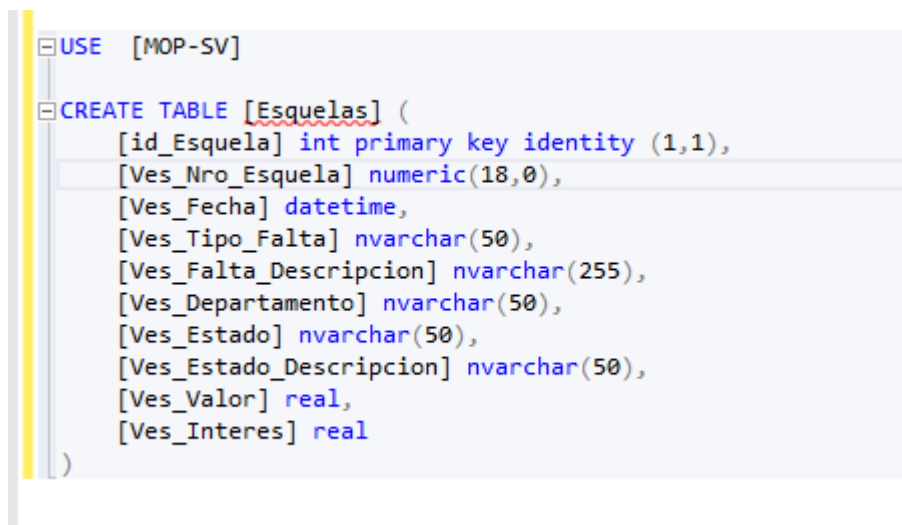
El Ministerio de Obras Publicas (MOP) actualmente la información la consolidad por medio de archivo en Excel a través de una herramienta interna donde no genera reportes automáticos. Nosotros analizaremos la información actual para ofrecer herramientas necesarias para analizar la información de acuerdo con la necesidad que poseen.

## ANÁLISIS ESQUELAS

**Estrategia usada:** Power BI

**Proceso:** Primeramente, implementamos el proceso ETL para transformar la información y almacenarla a nuestro servidor SQL.

En nuestro servidor se creó la siguiente tabla:

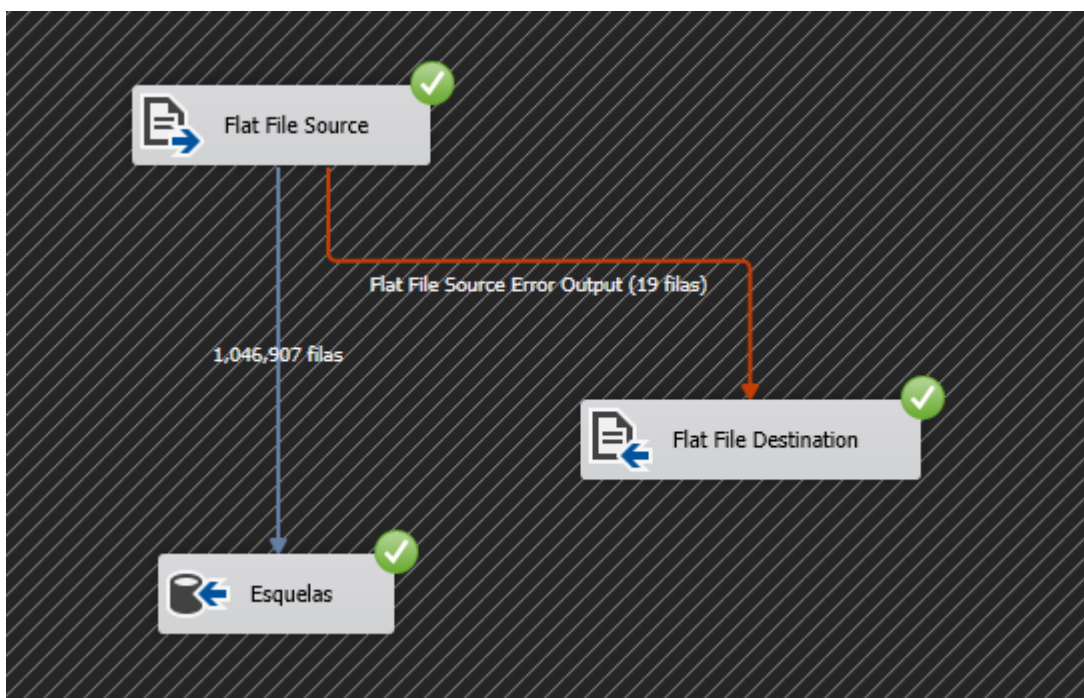


```
USE [MOP-SV]

CREATE TABLE [Esquelas] (
    [id_Esquila] int primary key identity (1,1),
    [Ves_Nro_Esquila] numeric(18,0),
    [Ves_Fecha] datetime,
    [Ves_Tipo_Falta] nvarchar(50),
    [Ves_Falta_Descripcion] nvarchar(255),
    [Ves_Departamento] nvarchar(50),
    [Ves_Estado] nvarchar(50),
    [Ves_Estado_Descripcion] nvarchar(50),
    [Ves_Valor] real,
    [Ves_Interes] real
)
```

Hacemos la transformación de los datos usando un proyecto SSIS con los siguientes componentes y manejamos los errores por medio de otro archivo, donde nuestro resultado fue el ingreso de 1,046,907 registros y 19 registro que no se incluyeron porque tenían conflictos con los tipos de datos proporcionados.

Estos datos deben ser validados para que posteriormente sean ingresados e incluirlo en nuestro análisis.



Luego en Power BI, establecemos la conexión y cargamos los datos para iniciar el análisis requerido.

**Navegador**

Opciones de presentación ▾

- ▲ KELVINPC [2]
- AdventureWorksDW2012
- ▲ MOP-SV [3]
  - ☐ Dim\_Departamento
  - ☐ Dim\_Falta
  - ☒ Esquelas

**Esquelas**

id_Esquila	Ves_Nro_Esquila	Ves_Fecha	Ves_Tipo_Falta	Ves
1	110637	4/6/2003 00:00:00	TRANSITO	
2	107742	4/6/2003 00:00:00	TRANSITO	
3	106645	4/6/2003 00:00:00	TRANSITO	
4	154061	4/6/2003 00:00:00	TRANSITO	
5	154102	3/6/2003 00:00:00	TRANSITO	
6	119779	6/6/2003 00:00:00	TRANSITO	
7	106938	4/6/2003 00:00:00	TRANSITO	
8	154457	4/6/2003 00:00:00	TRANSITO	
9	154951	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
10	154976	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
11	109210	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
12	109211	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
13	109207	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
14	105578	9/6/2003 00:00:00	TRANSITO	
15	109469	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
16	154826	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
17	154952	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
18	110968	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
19	106398	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	
20	106075	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	
21	107107	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	
22	154977	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	
23	148297	7/6/2003 00:00:00	TRANSITO	

Seleccionar tablas relacionadas

Cargar Transformar datos Cancelar

Teniendo en cuenta la información de la tabla esta presenta, datos como la fecha, tipo de esquila, departamento, estado, valor y la mora.

Nuestro primer análisis será ver que departamentos son los que tiene más esquelas emitidas, ya que podemos dar mas prioridad a los departamentos que posean muchas y programar una campaña de concientización.

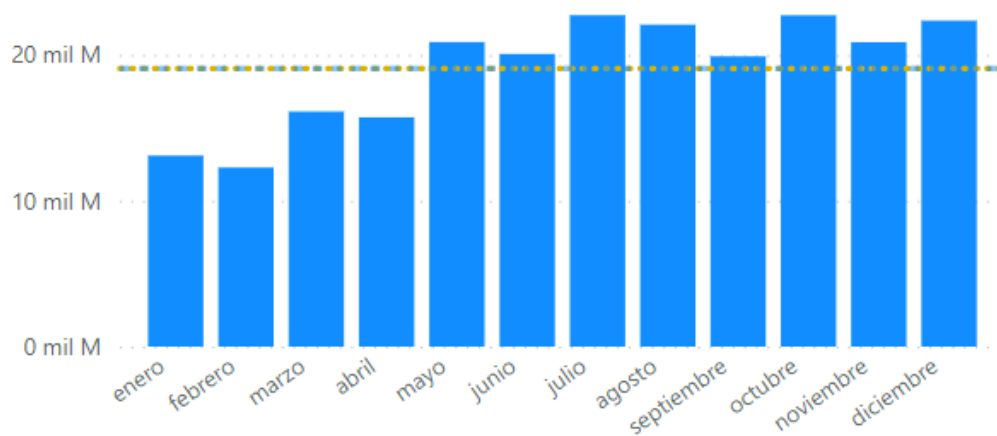
En nuestro análisis el departamento para el año 2007, San Salvador es el que tiene mas esquelas procesadas, seguido del departamento de La Paz y San Miguel.

### Esquelas por departamento

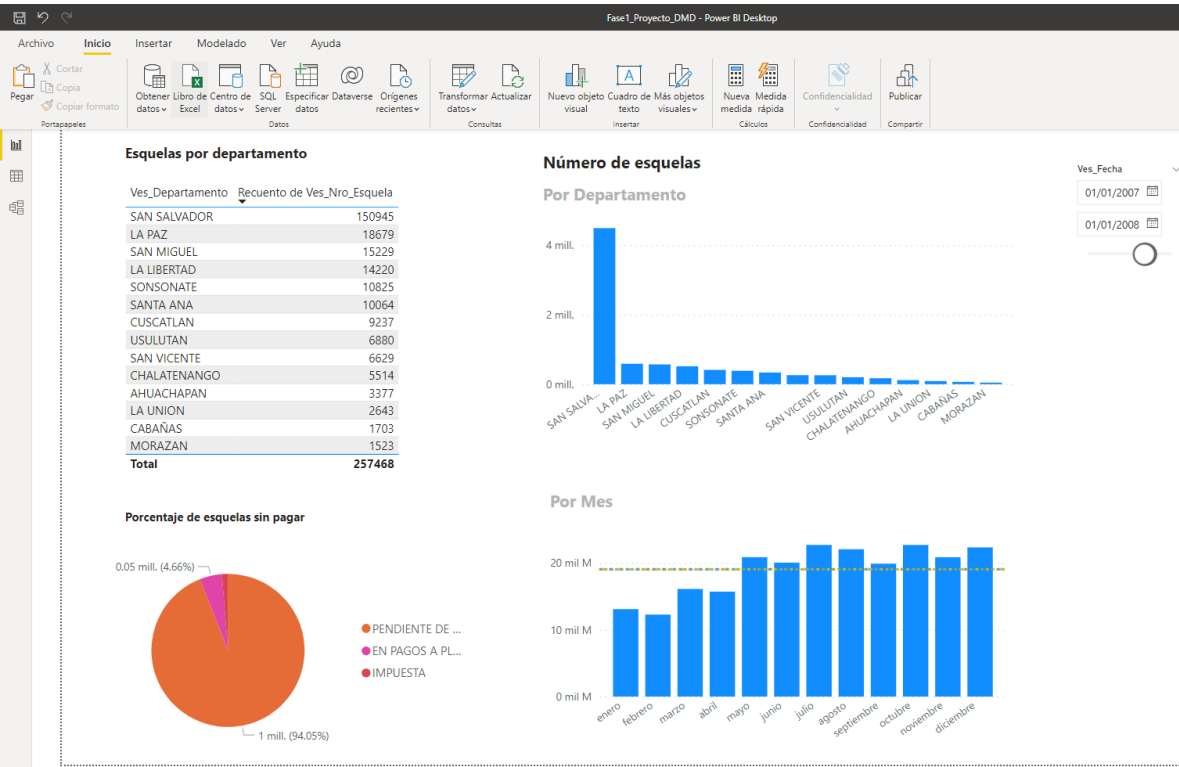
Ves_Departamento	Recuento de Ves_Nro_Esquila
SAN SALVADOR	150945
LA PAZ	18679
SAN MIGUEL	15229
LA LIBERTAD	14220
SONSONATE	10825
SANTA ANA	10064
CUSCATLAN	9237
USulután	6880
SAN VICENTE	6629
CHALATENANGO	5514
AHUACHAPAN	3377
LA UNION	2643
CABAÑAS	1703
MORAZAN	1523
<b>Total</b>	<b>257468</b>

El siguiente análisis es revisar cual mes del año 2007 posee más esquelas, podemos observar que después del mes de abril las esquelas sobrepasan el promedio, aquí tenemos un factor preocupante ya que representa mas del 66% en todo el año.

### Por Mes



Finalmente este será nuestro dashboard del análisis.



**Estrategia usada:** Reporting Services

**Proceso:** Realizamos la conexión con el origen de datos.

The screenshot shows the 'Asistente para informes' (Report Wizard) window, specifically the 'Seleccionar el origen de datos' (Select data source) step. The window title is 'Asistente para informes'. The main heading is 'Seleccionar el origen de datos' with the instruction: 'Seleccione un origen de datos desde donde obtener datos para este informe o cree un nuevo origen de datos.' (Select a data source from where to obtain data for this report or create a new data source.)

There are two radio button options:

- ☐ Origen de datos compartido (Shared data source)
- ☒ Nuevo origen de datos (New data source)

Under the 'Nuevo origen de datos' option, there are several fields and buttons:

- Nombre:** (Name) - A text box containing 'Esquelas'.
- Tipo:** (Type) - A dropdown menu showing 'Microsoft SQL Server'.
- Cadena de conexión:** (Connection string) - A large text area for entering the connection string.
- Editar...** (Edit...) - A button to edit the connection string.
- Credenciales...** (Credentials...) - A button to manage credentials.

At the bottom of the window, there are navigation buttons: 'Help', '< Back', 'Next >' (highlighted with a blue border), 'Finish >>|', and 'Cancel'.

Debemos de agregar la consulta deseada para realizar el analisis, para ellos vamos a analizar la cantidad de esquelas por departamento, este será un reporte que vamos a generar a través de la red local del MOP y únicamente las personas autorizadas podrán ver la información.



Propiedades del conjunto de datos

Consulta

Campos

Opciones

Filtros

Parámetros

Elija un origen de datos y cree una consulta.

Nombre:

DataSet\_Esquelas

☐ Usar un conjunto de datos compartido.  
☒ Usar un conjunto de datos insertado en el informe.

Origen de datos:

ListadoPagosPendientes Nueva...

Tipo de consulta:

☒ Texto ☐ Tabla ☐ Procedimiento almacenado

Consulta:

```
SELECT id_Esquila, Ves_Nro_Esquila, Ves_Fecha, Ves_Tipo_Falta, Ves_Falta_Descripcion,
Ves_Departamento, Ves_Estado, Ves_Estado_Descripcion, Ves_Valor, Ves_Interes
FROM Esquelas
WHERE (Ves_Fecha BETWEEN '2007-01-01' AND '2007-12-31') AND
(Ves_Estado_Descripcion = 'PENDIENTE DE PAGO')
```

Diseñador de consultas... Importar... Actualizar campos

Esquelas - Visor de informes

No es seguro | kelvinpc/ReportServer/Pages/ReportViewer.aspx?%2fDMD\_Fase1\_GJ111587%2f

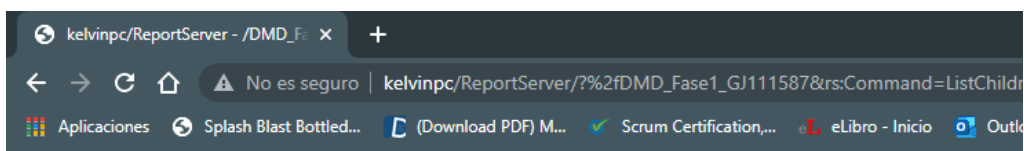
Aplicaciones Splash Blast Bottled... (Download PDF) M... Scrum Certification,... eBook - Inicio Outlook

Enter a user name and password to access the data source:

Nombre de inicio de sesión: Contraseña:

El personal tendrá visibilidad para dos reportes uno para el análisis de esquelas por departamento y otro con el listado de personas pendientes a pagar.

Esto a solicitud del mismo personal que necesita consultar por periodos y el detalle del valor pendiente de las esquelas, con la finalidad de dar un seguimiento oportuno al cobro de este.



kelvinpc/ReportServer - /DMD_Fase1_GJ111587		
<a href="#">[Al directorio principal]</a>		
lunes, noviembre 14, 2022 2:47 AM	52195	<a href="#">Esquelas</a>
lunes, noviembre 14, 2022 3:29 AM	52335	<a href="#">PagosPendientes</a>
lunes, noviembre 14, 2022 2:46 AM	13481	<a href="#">Report1</a>

Microsoft SQL Server Reporting Services Versión 15.0.1102.1002

Tendremos la visibilidad de esta manera:

Cambiar credenciales

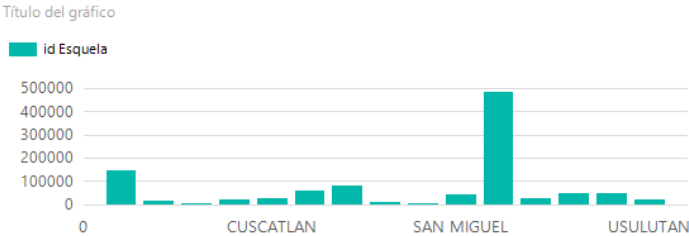


Ministerio de Obras Publicas

KELVINPC\Kelvi  
n

Reporte de esquelas

Detalle de Esquelas por departamento



	TRANSITO	TRANSPORTE CARGA	TRANSPORTE TERRESTRE
AHUACHAPAN	123594	705	21829
CABAÑAS	14450	49	1239
CHALATENAN GO	5946	21	419
CUSCATLAN	19410	184	1137
LA LIBERTAD	26507	125	2279
LA PAZ	46373	382	11614
LA UNION	77380	1472	3543
MORAZAN	9662	71	1945
SAN MIGUEL	4819	123	641
SAN SALVADOR	35033	674	9144
SAN VICENTE	365577	2107	117516
SANTA ANA	22209	371	1535
	43366	552	4825

PagosPendientes - Visor de info

No es seguro

kelvinpc/ReportServer/Pages/ReportViewer.aspx?%2fDMD\_Fase1\_GJ111587%2fPagosPendientes&rs:Co

Aplicaciones Splash Blast Bottled... (Download PDF) M... Scrum Certification,... eLibro - Inicio Outlook Aula Digital: Iniciar...

Cambiar credenciales

1 de 2 ?

100%

Buscar | Siguiend

MOP

Ministerio de Obras Publicas

Reporte de esquelas

Pagos Pendientes

01/01/2007 0:00:00

Ves Nro Esquela	Ves Valor
612843	57,14
654012	34,29
656236	57,14
672412	34,29
682172	44,42