



## **DATAWAREHOUSE Y MINERÍA DE DATOS**

**CATEDRATICO KAREN MEDRANO**

### **FASE 1**

<b>N°</b>	<b>ESTUDIANTES</b>	<b>CARNET</b>
1	MENA CORTEZ, BRYAN ANTONIO	MC211787
2	LOPEZ GUTIERREZ, CALEB VERENICE	LG211551

**13 DE NOVIEMBRE DEL AÑO 2022**

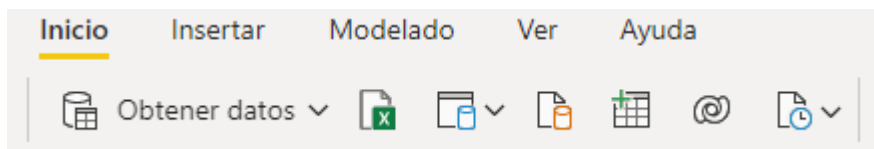
## I. OBJETIVOS

- Analizar la información del Ministerio de Obras Públicas (MOP) mediante las estrategias de minería de datos.
- Proporcionar la información y datos mediante K Means y Power BI.

## II. MARCO TEORICO

### TEMA: USO DE POWER BI

Para introducir el archivo CSV del cual queremos extraer los datos, debemos irnos a inicio y seleccionamos obtener datos, en ese apartado seleccionamos CSV.



Luego nos aparecerá una vista previa de nuestra tabla de datos, ahí podemos hacer cambios de formato o modificaciones, en este caso lo utilizaremos tal y como se nos entregaron los datos.

#### Esquelas.csv

Origen de archivo: 1252: Europeo occidental (Windows) | Delimitador: Punto y coma | Detección del tipo de datos: Basado en las primeras 200 filas

Ves_Nro_Esquila	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcion	Ves_Departamento
110637	04/06/2003 12:00:00 a. m.	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO
107742	04/06/2003 12:00:00 a. m.	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO
106645	04/06/2003 12:00:00 a. m.	TRANSITO	ESTACIONARSE EN CURVAS PUENTES Y TRECHOS ANGOSTOS	SAN SALVADOR
154061	04/06/2003 12:00:00 a. m.	TRANSITO	BAJAR O SUBIR PASAJEROS EN LUGARES NO PERMITIDOS	SAN SALVADOR
154102	03/06/2003 12:00:00 a. m.	TRANSITO	DISPUTARSE LA VIA CON OTRO VEHICULO	SAN SALVADOR
119779	06/06/2003 12:00:00 a. m.	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIMADA	SAN SALVADOR
106938	04/06/2003 12:00:00 a. m.	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
154457	04/06/2003 12:00:00 a. m.	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
154951	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZADAS	SAN SALVADOR
154976	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZADAS	SAN SALVADOR
109210	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
109211	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
109207	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
105578	09/06/2003 12:00:00 a. m.	TRANSITO	CIRCULAR CON VEHICULOS NO MATRICULADOS	SAN SALVADOR
109469	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
154826	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
154952	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR
110968	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN PERMISO	SAN SALVADOR
106398	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR
106075	07/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR

Extraer tabla mediante ejemplos | Cargar | Transformar datos | Cancelar

Para las consultas de datos que deseamos realizar necesitamos ejecutar diversos scripts, luego de un análisis estos son los que nosotros hemos creado:

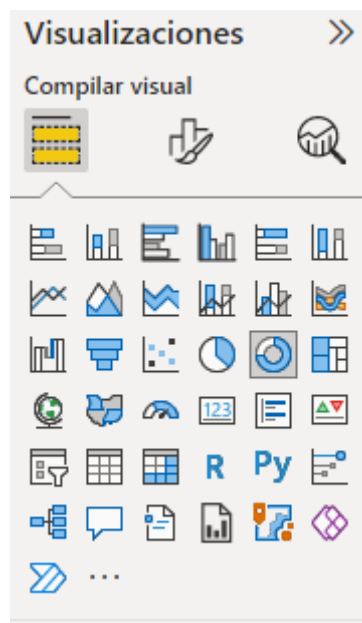
```
Estado = COUNT(Esqueles[Ves_Estado_Descripcion]) //Para obtener las cantidades de cada estado
```

```
Suma total recaudada = SUM(Esqueles[Ves_Valor]) // obtener el dinero total de todas las multas
```

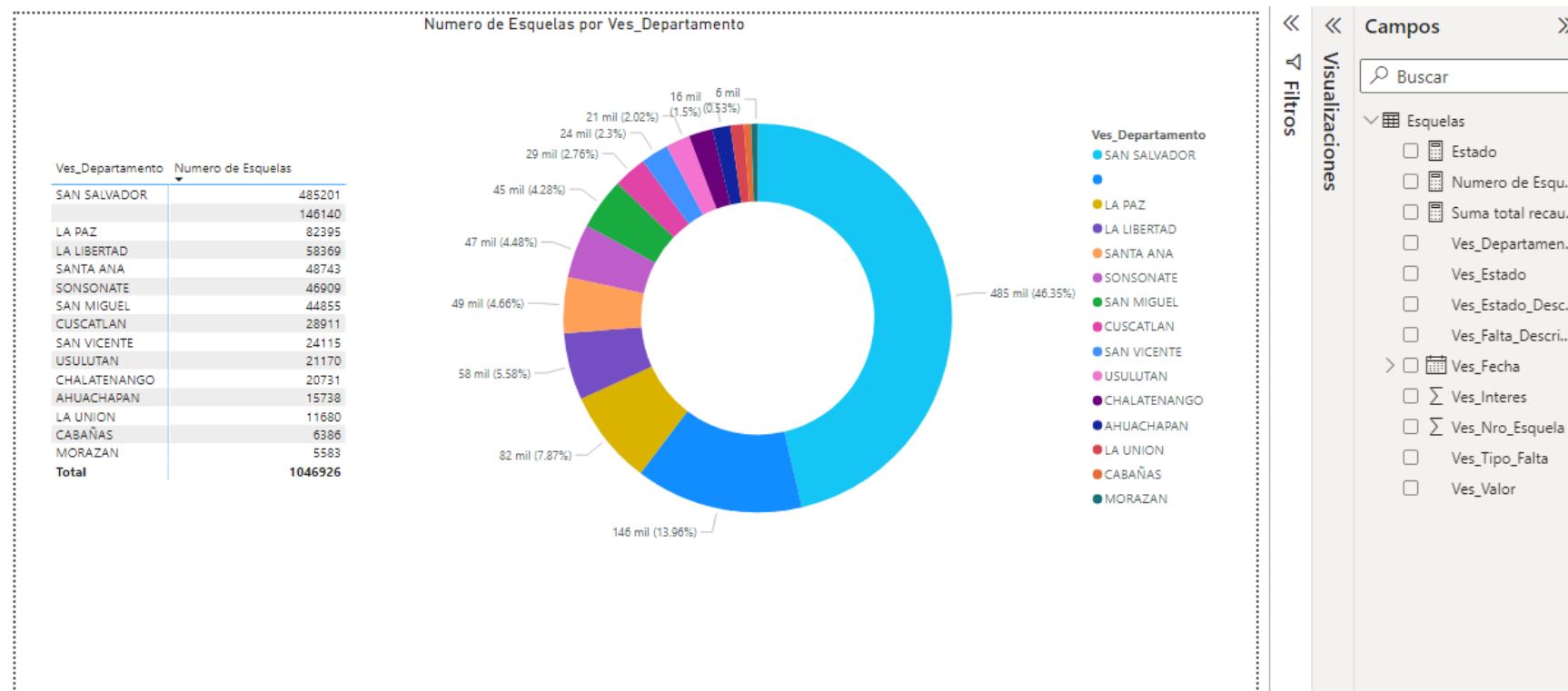
```
Numero de Esqueles = COUNT(Esqueles[Ves_Departamento]) //El numero total de multas impuestas
```

Estos scripts los hemos implementado ya que, si bien Power BI nos implementa ciertas consultas de datos de manera automática, siempre existirán diversas consultas para obtener un “Query” muy específico de algo que deseamos saber.

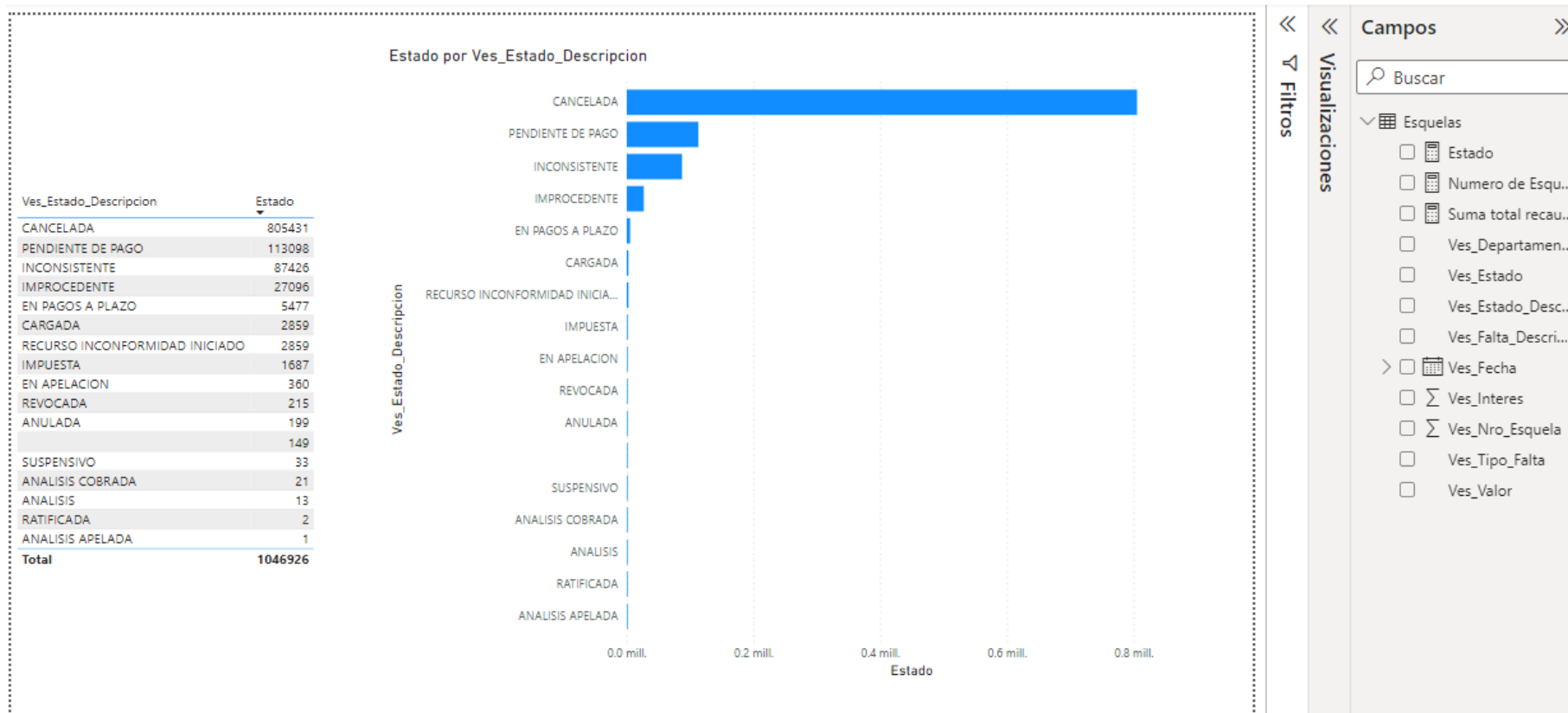
A la hora de mostrar datos, tenemos la pestaña de diseño, donde seleccionaremos el más adecuado según los datos que deseamos mostrar.



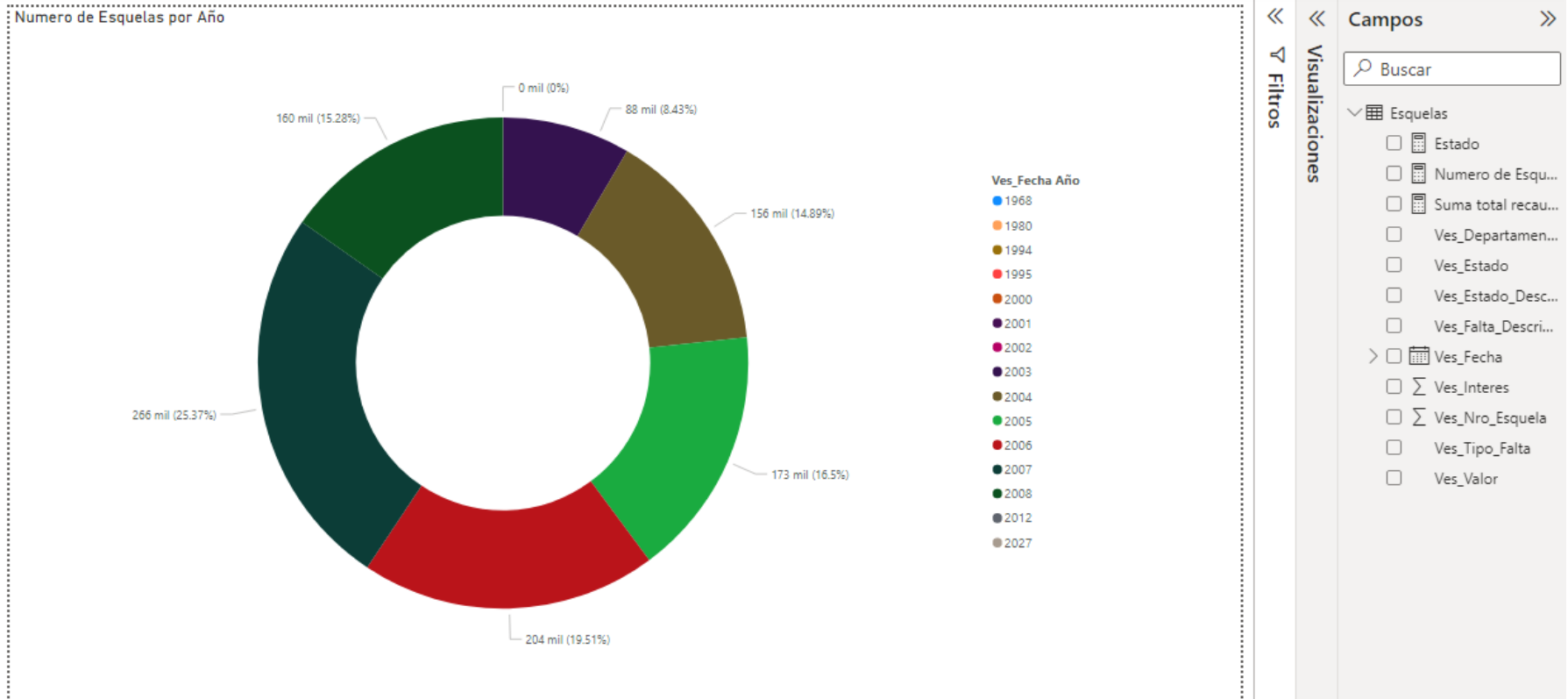
Los primeros datos que mostraremos serán las cantidades de esquelos impuestas en cada departamento del país, siendo San Salvador el departamento que más multas acumula.



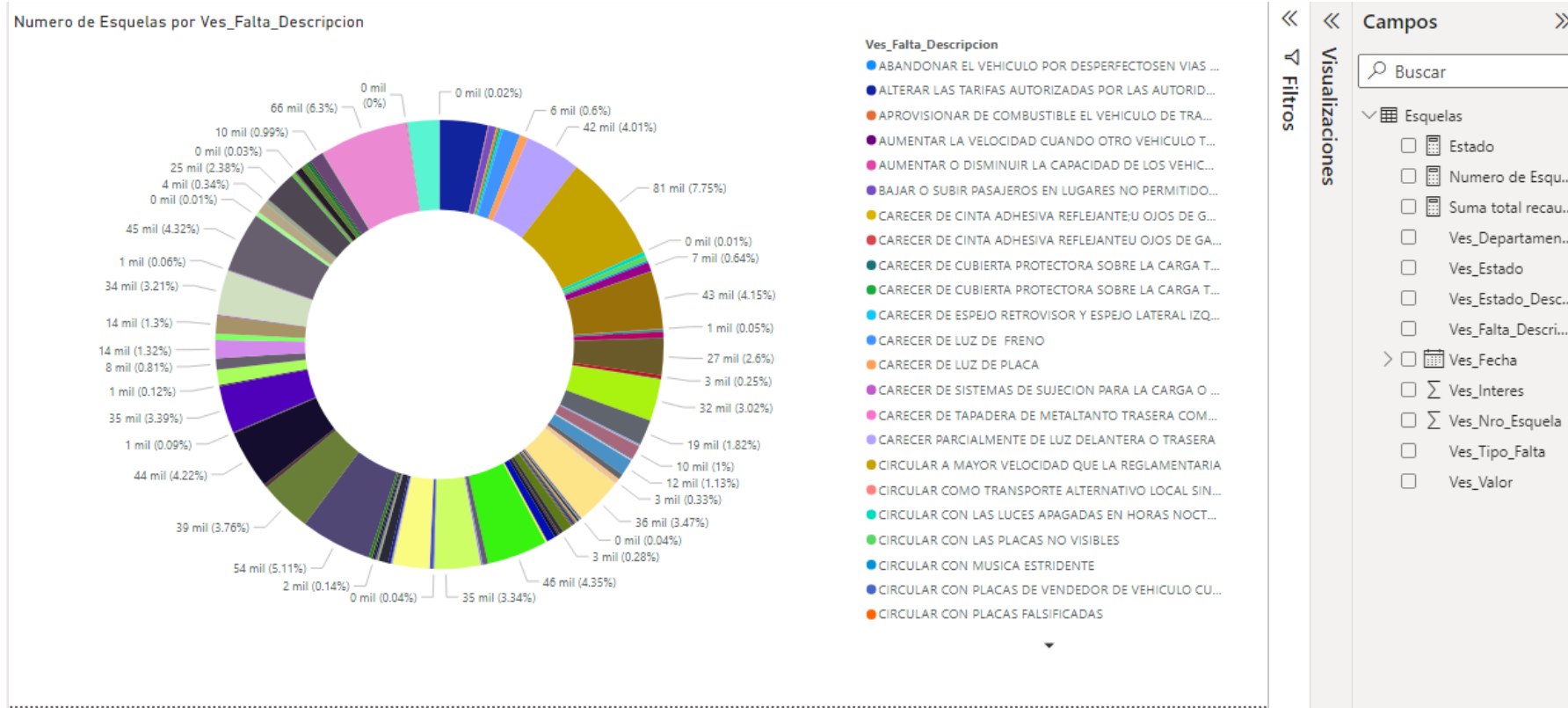
La siguiente consulta la hemos enfocado en conocer los estados de las más de un millón de multas impuestas, como podemos observar, mas de 800 mil multas ya han sido canceladas, representando la mayoría.



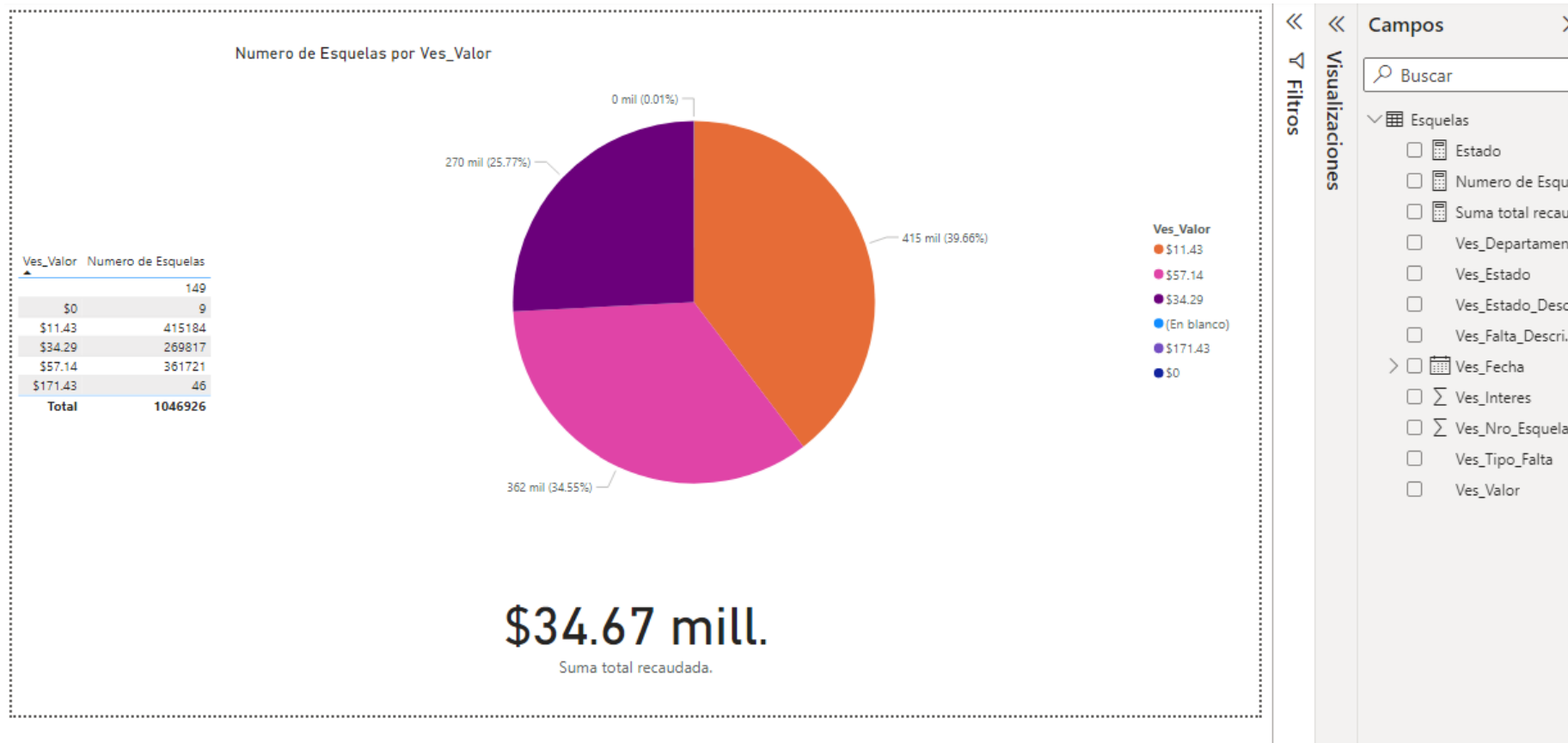
En cuanto al numero de esquelas por años, 2006, 2007 y 2008 encabezan la lista.



En cuanto a los motivos de las multas, podemos observar que las razones son muy diversas, sin embargo, conducir a una velocidad mayor a la permitida es la razón más frecuente de multas.

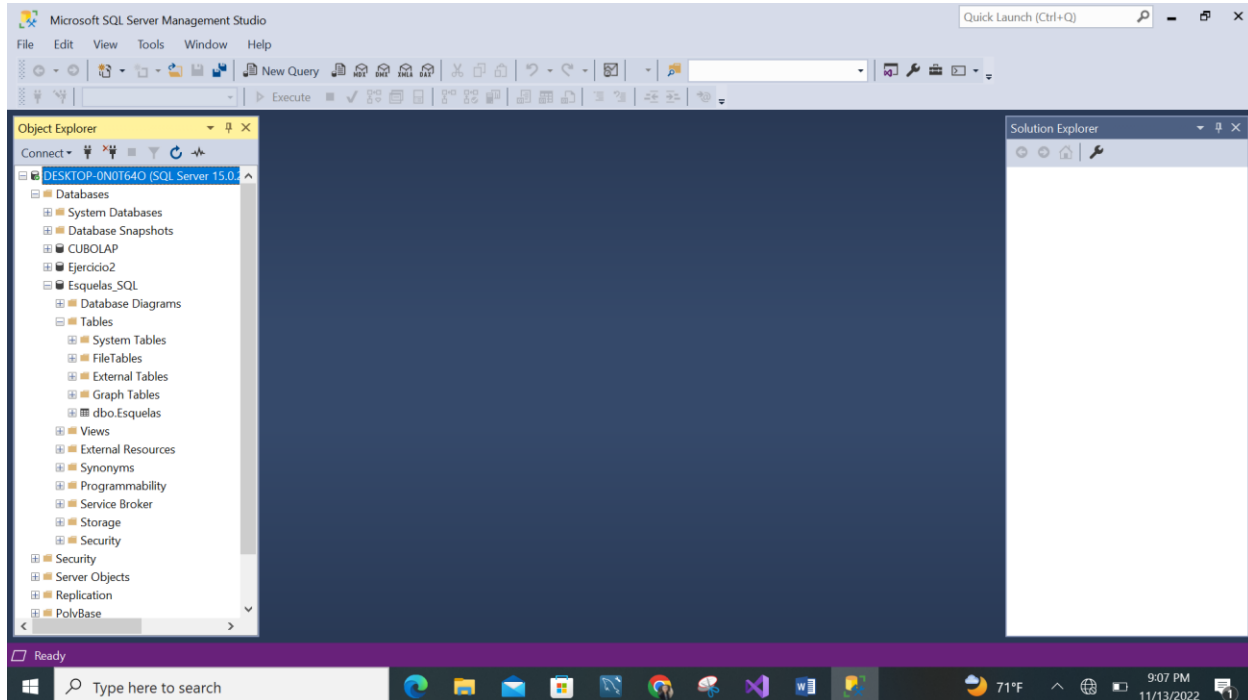


En cuanto a valor monetario, hemos clasificado el grafico de pastel por los montos en que se puede recibir una multa y vemos que la multa mas aplicada es la de \$57.14 y también hemos totalizado el valor de las multas, por lo que obtuvimos que el gobiernos ha recaudado más de \$34 millones de dólares en multas.

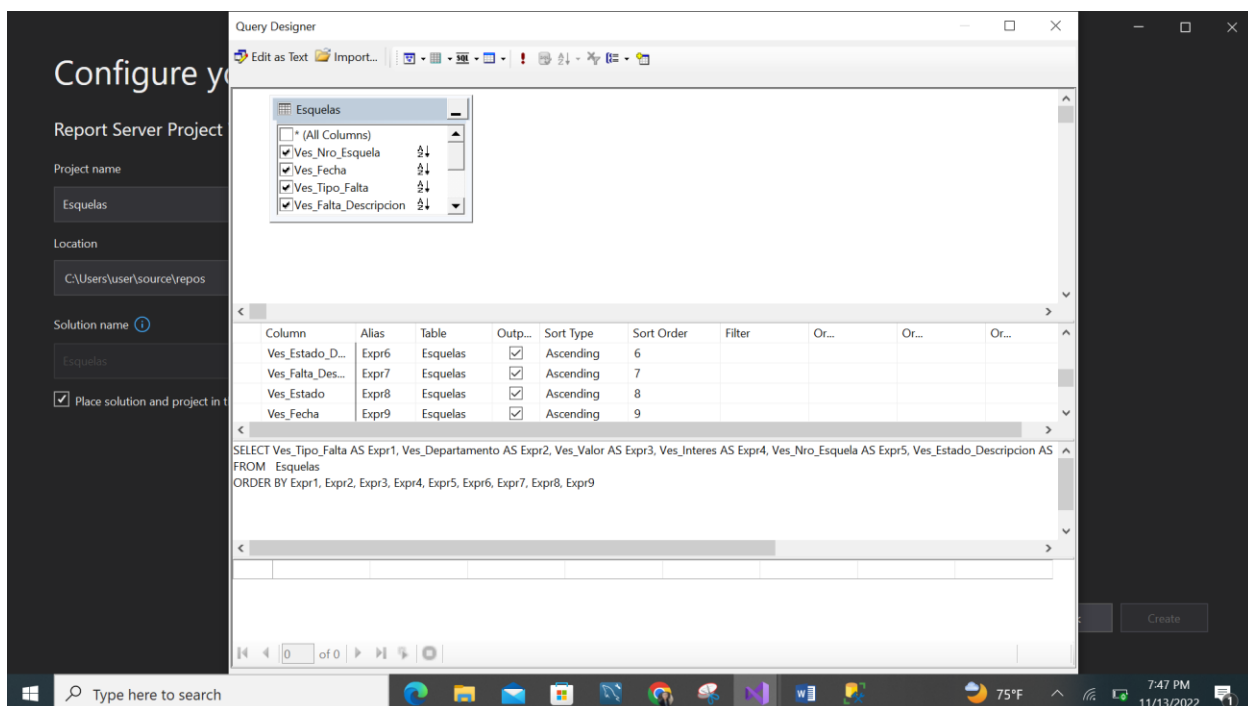




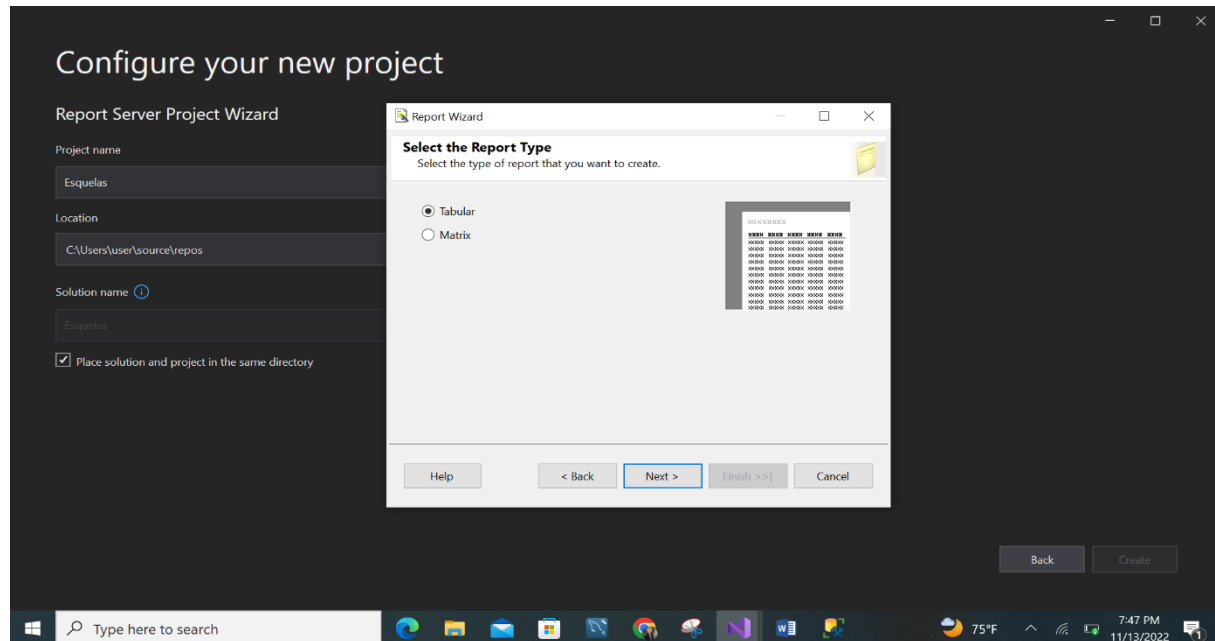
## TEMA: REPORTING SERVICES



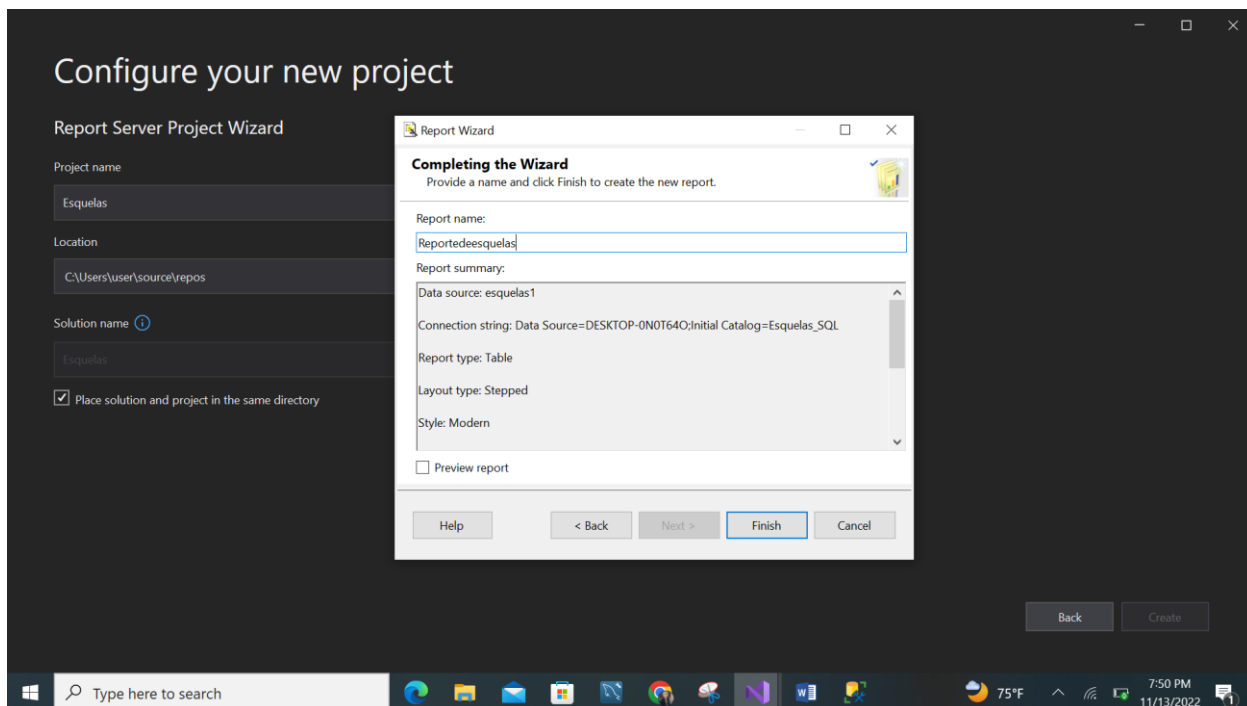
Para poder hacer un mayor análisis y uso de reporting services hemos importado el archivo cvs, el cual nos proporcionara un conocimiento mejor de los datos de Esquelas en el MOP.



Haciendo uso de visual studio creamos un proyecto en el cual diseñamos las características de nuestro reporte, esto para poder organizar los datos e información que se encontraban en SQL Server.



Hacemos una selección del tipo de reporte, en el que tabulamos los datos.



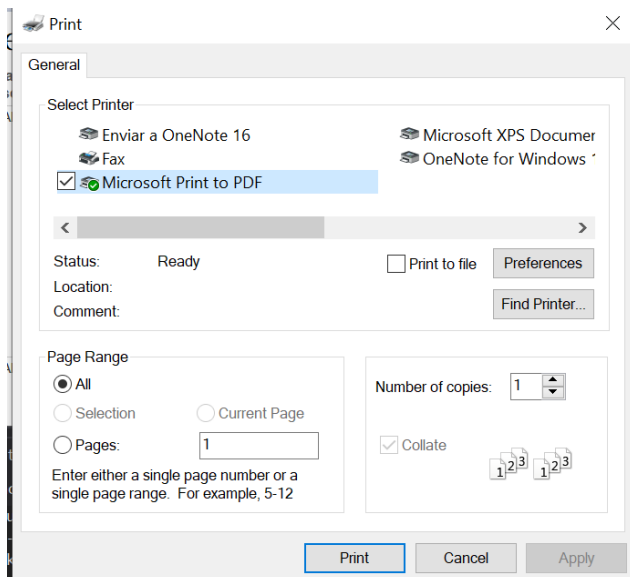
Completamos de organizar nuestro reporte y finalizamos.

SSIS Toolbox Reportedeesquelas.rdl [Design] 75% Find Next

### Reportedeesquelas

falta descripcion	departament o	valor	Interes	Numero estado	falta	estado des	fecha
TRANSITO		11.430000305 1758	0	694 CANCELADA	INCUMPLIR LAS DISPOSICION ES RELATIVAS AL PORCENTAJE MINIMO DE PASO DE LUZ SOLAR EN LOS VIDRIOS POLARIZADO SREGULADOS EN EL ART. 14 -A DE ESTA LEY	CBR	6/12/2006 12:00:00 AM
TRANSITO		11.430000305 1758	0	695 CANCELADA	NO PORTAR LOS TRIANGULOS REFLECTIVO	CBR	6/12/2006 12:00:00 AM

Podemos observar el reporte de esquelas, basado en los datos que se encontraban en nuestro gestor de una manera más gráfica y con alias que ayudan a reconocer los valores internos que se encontraban a modo de dar a conocer la información pertinente.



Y obtener un pdf del reporte, de tal forma que los datos se puedan obtener de una forma más ordenada y presentable.

### **III. ANTECEDENTES**

El análisis y la minería de datos se han convertido en una necesidad para las empresas e instituciones, esto debido a que la toma de decisiones tiene como uno de sus principales factores las predicciones basadas en las consecuencias de las acciones previas, por lo cual, no solamente se trata de llevar un registro de los datos, sino, de saber obtener información de relevancia por medio la interpretación de estos.

### **IV. SITUACION ACTUAL**

El “Ministerio de obras públicas” no cuenta con un análisis de los datos que tienen almacenados, únicamente poseen una base de datos con los registros de las esquelas, dicha base de datos ha sido entregada en un documento CSV el cual consta de una única tabla con diversos campos relacionados a las esquelas impuestas, dicha base consta de más de un millón de registros los cuales deben ser analizados.

### **V. METODOLOGIA**

A metodología utilizada en el presente proyecto será la de investigación descriptiva, ya que se utilizará una base de datos desorganizada, la cual será analizada por medio de diversas técnicas de minería de datos, con el fin de obtener información relevante la cual será presentada de una manera organizada.

### **VI. FORMULACION DEL PROBLEMA**

La actual base de datos que posee el “Ministerio de obras públicas”, cuenta con registros de las esquelas impuestas sobre el territorio nacional, sin embargo, no tienen la capacidad de analizar dichos datos, por lo que no tienen un funcionamiento más que llevar el registro histórico, es necesario poder analizar la información que poseen ya que de esta manera se pueden tomar medidas de prevención y/o corrección.

### **VII. FACTIBILIDAD**

El análisis de los datos se realizará utilizando diversas herramientas de software y sus respectivas licencias, además se utilizará equipo con características de gamas media, esto debido a que el procesamiento de bases de datos tan extensas obliga a utilizar un poder de procesamiento

medio-alto, luego del procesamiento de los datos, se espera que la información obtenida pueda ser utilizada para la creación de nuevas estrategias y planes por parte de la institución.

## **VIII. JUSTIFICACION**

Hemos implementado **estrategias de minería de datos** para poder organizar una cantidad grande de datos y hacer un análisis de las esquelas de tránsito, lo cual es tema del que desean presentar reportes.

Usando **estrategias de minería de datos**, podemos visualizar los datos de una forma más entendible y gráfica, siendo esta una ayuda para el MOP, ya que podrán valorar los datos y dar una respuesta inmediata a situaciones que se presenten dentro de los resultados, por esa razón se presenta esta documentación.

## **IX. IMPORTANCIA**

La importancia de poder hacer uso de estas estrategias es el poder documentar un gran volumen de datos y poder brindar resultados más gráficos, siendo este el uso de Power BI, en el cual hicimos gráficos los resultados y Reporting Services en donde podemos hacer la impresión de los datos para que se puedan tener físicamente por quienes observaran los análisis realizados.

## **X. ALCANCES Y LIMITACIONES**

Se debe tener equipos y herramientas técnicas adecuadas como así también una excelente gestión a la hora de hacer los procesos de ordenamiento de datos mediante las estrategias de minería, razón por la cual se puede tener una complicación cuando se tenga ya presente los documentos, si estos no han tenido un proceso correcto en el desarrollo del análisis de los datos, por eso se dice que la mayor complicación es que las herramientas son demasiadas complejas.

## **XI. PLANIFICACION DE RECURSOS**

A la hora de planificar se debe de tener presente el volumen, tipo de datos, variables, y todos aquellos aspectos necesarios antes de implementar la herramienta para que se puedan detallar la información y datos de una forma más clara, eso nos dará una mejor referencia de planificación.

## **XII. RECOMENDACIONES**

- Tener presente el formato de archivo para el uso de la herramienta.
- Hacer una investigación de la implementación de la herramienta.
- Tener presente la entrega de los datos físicos. Valorar los datos más importantes que serán expuestos de acuerdo al tipo de reporte que se quiere presentar.