

PROYECTO FASE 1

DMD941-G01T



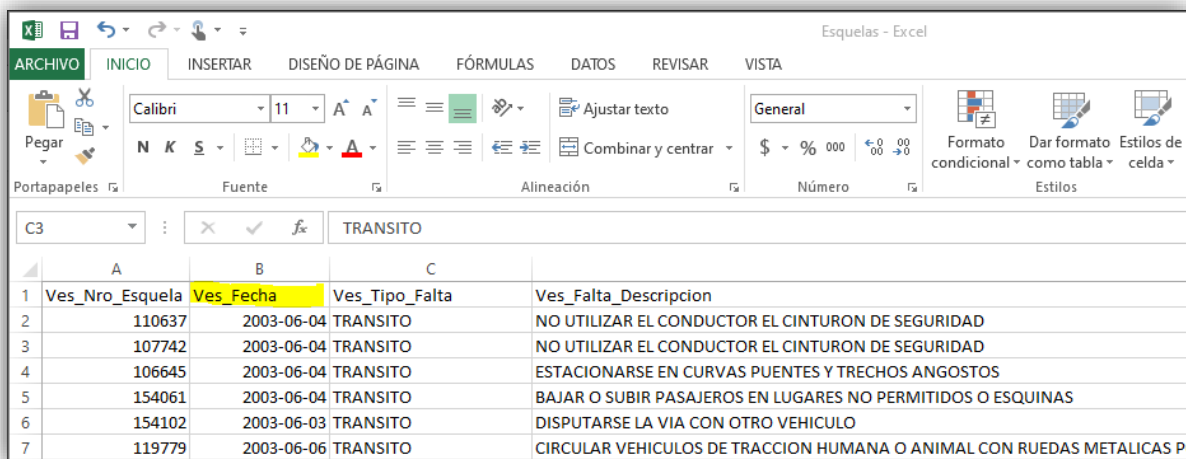
23 ABRIL

GERARDO ANTONIO CABEZAS VAQUERO
Carnet: CV152055

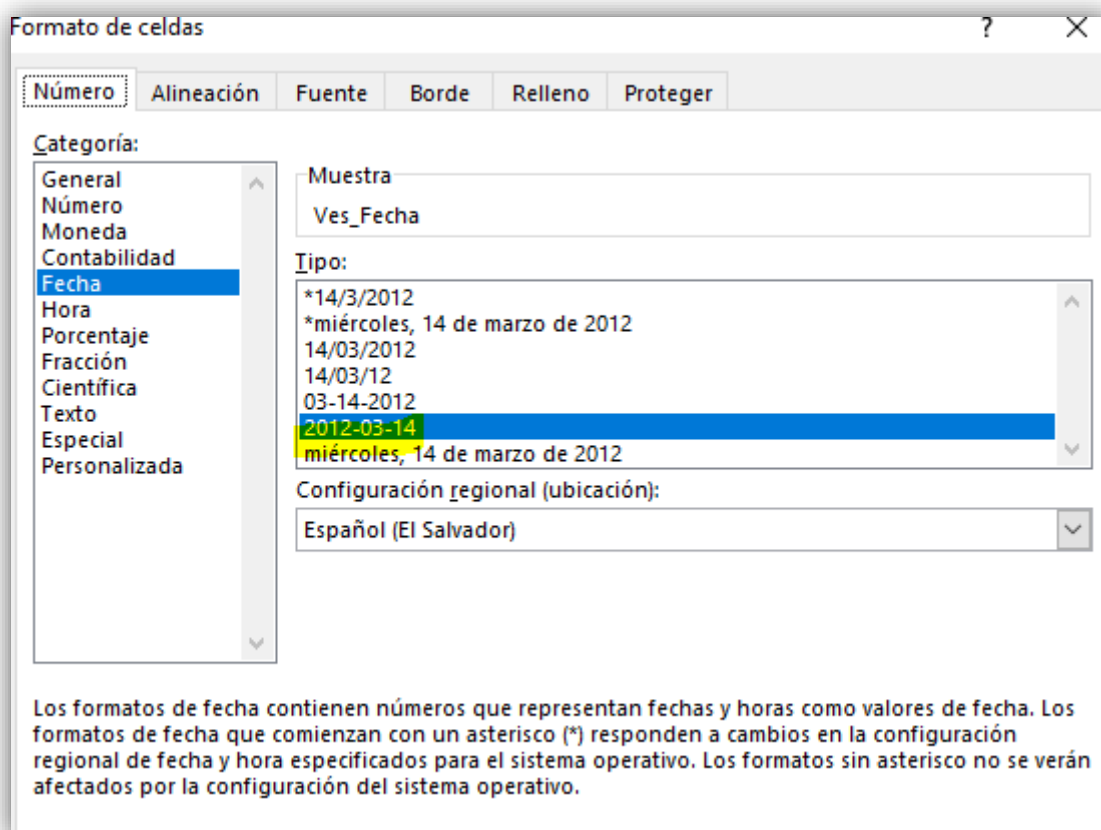
DESARROLLO DEL PROYECTO:

PREPARACIÓN DE DOCUMENTOS:

En el documento de Esquelas proporcionado para el desarrollo del proyecto cambiamos el formato del campo Ves_Fecha para poder guardarlo correctamente en la base de datos:



	A	B	C	
	Ves_Nro_Esquila	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcion
1	110637	2003-06-04	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD
2	107742	2003-06-04	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD
3	106645	2003-06-04	TRANSITO	ESTACIONARSE EN CURVAS PUENTES Y TRENCHOS ANGOSTOS
4	154061	2003-06-04	TRANSITO	BAJAR O SUBIR PASAJEROS EN LUGARES NO PERMITIDOS O ESQUINAS
5	154102	2003-06-03	TRANSITO	DISPUTARSE LA VIA CON OTRO VEHICULO
6	119779	2003-06-06	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIMAL CON RUEDAS METALICAS



Formato de celdas

Número | Alineación | Fuente | Borde | Relleno | Proteger

Categoría:

- General
- Número
- Moneda
- Contabilidad
- Fecha**
- Hora
- Porcentaje
- Fracción
- Científica
- Texto
- Especial
- Personalizada

Muestra:

Ves_Fecha

Tipo:

- *14/3/2012
- *miércoles, 14 de marzo de 2012
- 14/03/2012
- 14/03/12
- 03-14-2012
- 2012-03-14**
- miércoles, 14 de marzo de 2012

Configuración regional (ubicación):

Español (El Salvador)

Los formatos de fecha contienen números que representan fechas y horas como valores de fecha. Los formatos de fecha que comienzan con un asterisco (*) responden a cambios en la configuración regional de fecha y hora especificados para el sistema operativo. Los formatos sin asterisco no se verán afectados por la configuración del sistema operativo.

De igual manera preparamos el archivo parque_vehicular_datos:

C	D	E	F	G	H	I	J	K	L	M	N	O	P
CILINDRADA	CANTIDAD_DE_CILI	CANTIDAD_C	VALOR_DEL	COLORES	FECHA_DE_I	IMP_VALOR	FECHA_DE_I	ANIO_INGRES	MES_INGRES	CLASE	PERTENENCI	MARCA	MODELO
1800	0	4	4094.56	AMARILLO	1994-09-16	4094.56	1994-11-17	1994	11	AUTOMOVIL	PROPIEDAD	NISSAN	N/D
0	0	0	0	AMARILLO		0	1989-02-16	1989	2	AUTOMOVIL	PROPIEDAD	TOYOTA	TIARA
1700	0	0	0	AMARILLO F/C B/N		0	1985-03-20	1985	3	ALQUILER	PROPIEDAD	TOYOTA	CORONA
1600	0	0	0	AMARILLO F/C B/N		0	1988-02-18	1988	2	ALQUILER	PROPIEDAD	HYUNDAI	STELLAR
0	0	0	0	AMARILLO		0	1982-10-26	1982	10	ALQUILER	PROPIEDAD	DATSUN	210
1600	0	0	0	AMARILLO		0	1982-10-26	1982	10	AUTOMOVIL	PROPIEDAD	DATSUN	1600
0	0	0	0	AMARILLO		0	1984-05-23	1984	5	ALQUILER	PROPIEDAD	TOYOTA	CORONA
0	0	0	0	AMARILLO F/C B/N		0	1988-06-03	1988	6	ALQUILER	PROPIEDAD	TOYOTA	COROLLA
0	0	0	0	AMARILLO		0	1988-02-04	1988	2	ALQUILER	PROPIEDAD	TOYOTA	COROLLA

Procedemos a crear la base de datos Data_MOP donde almacenaremos la informacion de las dos tablas para trabajar el proyecto:

New Database

Select a page: General, Options, Filegroups

Script Help

Database name: Data_MOP

Owner: <default>

☒ Use full-text indexing

Database files:

Logical Name	File Type	Filegroup	Initial Size (MB)	Autogrowth / Maxsize	Path
Data_MOP	ROWS...	PRIMARY	8	By 64 MB, Unlimited	C:\Program Files\Microsoft SQL Server\MSSQL9\SAMPLES\DATA_MOP.mdf
Data_MOP...	LOG	Not Applicable	8	By 64 MB, Unlimited	C:\Program Files\Microsoft SQL Server\MSSQL9\SAMPLES\DATA_MOP.ldf

Connection:

Server: DESKTOP-BUMRL9S

Connection: DESKTOP-BUMRL9S\vlan

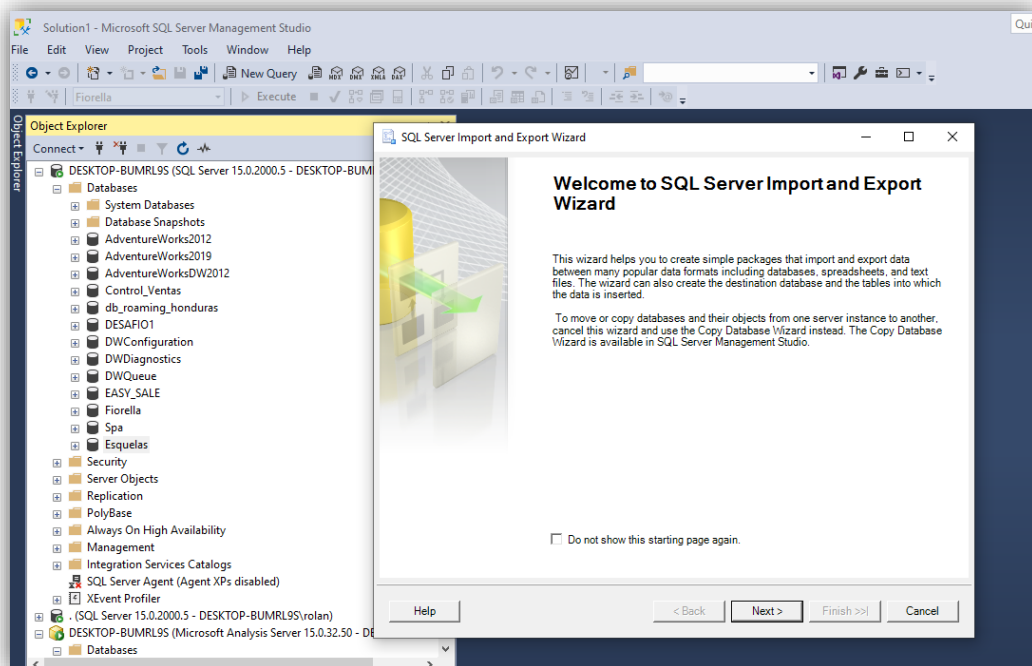
[View connection properties](#)

Progress: Ready

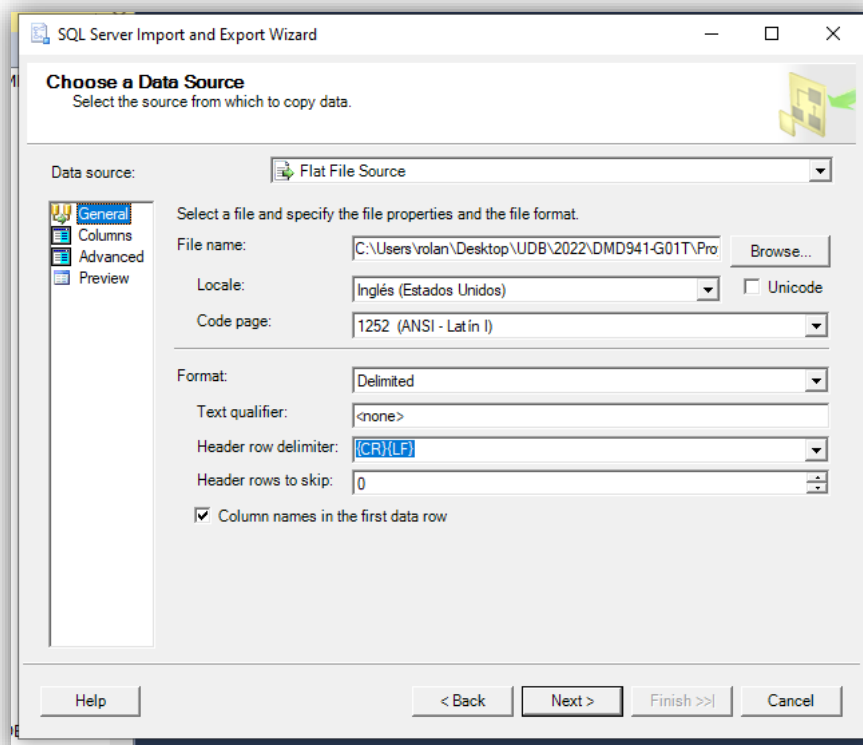
Add Remove

OK Cancel

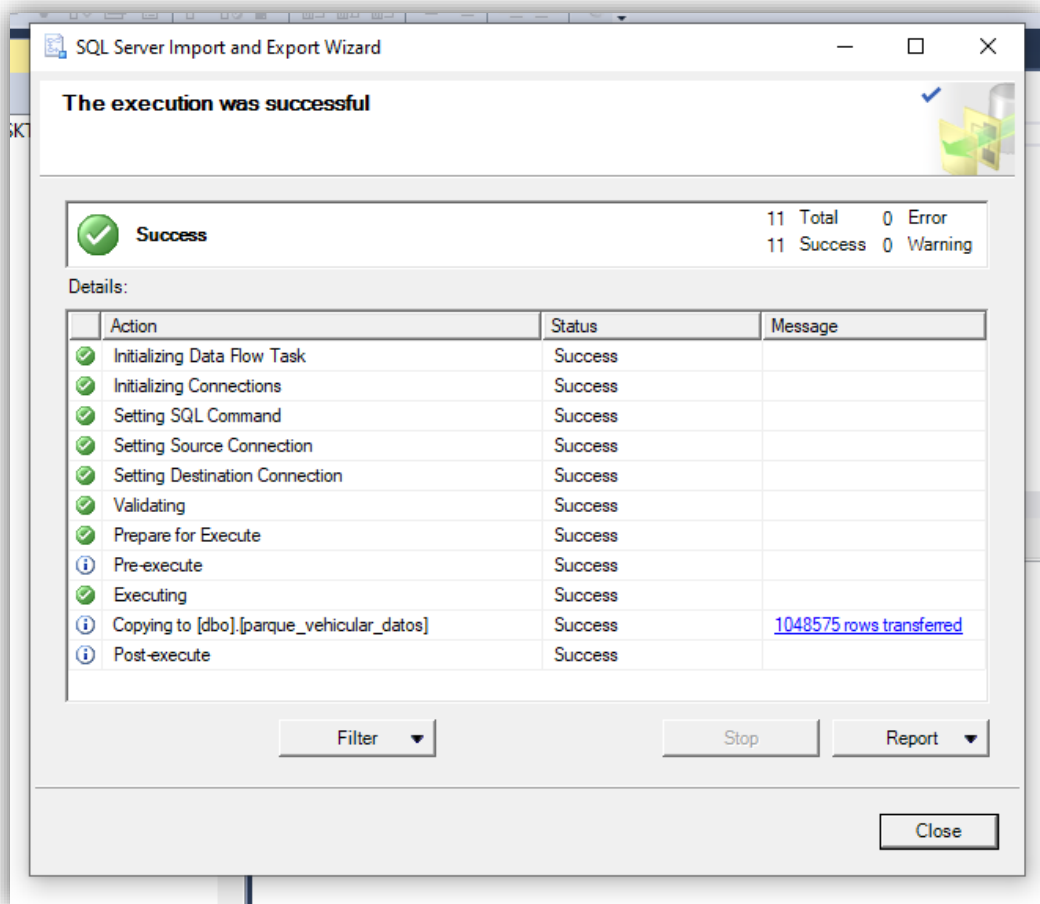
Ahora procederemos a carga los archivos desde el asistente de SQL Server Import:



A continuacion haremos el siguiente procedimiento, cargaremos los 2 archivos donde definiremos los campos todos como tipo texto, seran nuestras tablas pivote, el objetivo es poder manipular la informacion desde su origen para realizar una limpieza desde sentencias SQL:



Archivos cargados exitosamente:



Tablas pivot:

Select Top 5 * From Esuelas

Select TOP 5 * From parque_vehicular_datos

100 %

ResultsMessages

	Ves_Nro_Esuela	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcion	Ves_Departamento	Ves_Estado	Ves_Estado_Descripcion	Ves_Valor	Ves_Interes
1	110637	2003-06-04	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO	CBR	CANCELADA	34.29	0
2	107742	2003-06-04	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO	CBR	CANCELADA	34.29	0
3	106645	2003-06-04	TRANSITO	ESTACIONARSE EN CURVAS PUENTES Y TRECOS ANGOSTOS	SAN SALVADOR	CBR	CANCELADA	34.29	0
4	154061	2003-06-04	TRANSITO	BAJAR O SUBIR PASAJEROS EN LUGARES NO PERMITIDOS O ...	SAN SALVADOR	CBR	CANCELADA	34.29	0
5	154102	2003-06-03	TRANSITO	DISPUTARSE LA VIA CON OTRO VEHICULO	SAN SALVADOR		CARGADA	57.14	36.57

	TIPO_PLACA	ANIO_DE_FABRICACION	CILINDRADA	CANTIDAD_DE_CILINDROS	CANTIDAD_DE_PUERTAS	VALOR_DEL_VEHICULO	COLORES	FECHA_DE_IMPORTACION	IMP_VALOR_DEL_VEHICULO
1	CAMION	1984	9999	0	0	6649.71	BLANCO VERDE		6649.71
2	CAMION	1985	9999	0	0	5125.71	BLANCO		5125.71
3	CAMION	1978	9999	0	0	10813.03	AZUL BLANCO ROJO		10813.03
4	CAMION	1986	9999	0	0	12246.51	BLANCO		12246.51
5	CAMION	1989	9999	0	0	10599.66	BLANCO ROJO AZUL CAFE		10599.66

Creacion de tablas con sus campos y respectivos tipos de datos:

Creación de tabla DB_Esquelas:

```
CREATE TABLE DB_Esquelas (Id INT IDENTITY(1,1),
                           Ves_Nro_Esquila BIGINT,
                           Ves_Fecha DATETIME,
                           Ves_Tipo_Falta VARCHAR(50),
                           Ves_Falta_Descripcion VARCHAR(500),
                           Ves_Departamento VARCHAR(50),
                           Ves_Estado VARCHAR(25),
                           Ves_Estado_Descripcion VARCHAR(50),
                           Ves_Valor NUMERIC(12,2),
                           Ves_Interes NUMERIC(12,2))
```

Insert donde filtramos los campos para tener la información limpia:

```
INSERT INTO DB_Esquelas (Ves_Nro_Esquila,Ves_Fecha,Ves_Tipo_Falta,Ves_Falta_Descripcion,Ves_Departamento,Ves_Estado,Ves_Estado_Descripcion,Ves_Valor,Ves_Interes)
SELECT 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_Departamento IN ('USulután','CHALATENANGO','MORAZÁN','CUSCATLÁN','SAN SALVADOR','SANTA ANA','SONSONATE',
                           'CABAÑAS','LA PAZ','SAN VICENTE','SAN MIGUEL','AHUACHAPAN','LA LIBERTAD','LA UNION') AND
Ves_Estado IN ('IPT','SUS','RAT','ANL','APL','IRI','ANA','IPR','CBR','PPZ','ANC','INC','INI','REV','ANU','') AND
Ves_Estado_Descripcion IN ('ANALISIS APELADA','SUSPENSIVO','RECURSO INCONFORMIDAD INICIADO','PENDIENTE DE PAGO',
                           'ANALISIS','RATIFICADA','CANCELADA','IMPROCEDENTE','REVOCADA','EN APELACION',
                           'INCONSISTENTE','IMPUESTA','ANULADA','ANALISIS COBRADA','CARGADA','EN PAGOS A PLAZO') AND
ISNUMERIC(Ves_Valor)=1 AND ISNUMERIC(Ves_Interes)=1
```

100 %
Messages
(890667 rows affected)
Completion time: 2022-04-21T19:35:43.9650754-06:00

Creacion de tabla DB_paquete_vehicular_datos:

```
CREATE TABLE DB_parque_vehicular_datos (Id INT IDENTITY(1,1),
TIPO_PLACA VARCHAR(50),
ANIO_DE_FABRICACION INT,
CILINDRADA INT,
CANTIDAD_DE_CILINDROS VARCHAR(10),
CANTIDAD_DE_PUERTAS VARCHAR(10),
VALOR_DEL_VEHICULO NUMERIC(12,2),
COLORES VARCHAR(100),
FECHA_DE_IMPORTACION VARCHAR(25),
IMP_VALOR_DEL_VEHICULO NUMERIC(12,2),
FECHA_DE_INGRESO DATE,
ANIO_INGRESO INT,
MES_INGRESO INT,
CLASE VARCHAR(50),
PERTENENCIA VARCHAR(50),
MARCA VARCHAR(50),
MODELO VARCHAR(50),
CAPACIDAD NUMERIC(12,2),
COMBUSTIBLE VARCHAR(25),
ADUANA VARCHAR(50),
CONDICION_INGRESO VARCHAR(50),
PROPIETARIO_DEPARTAMENTO VARCHAR(50),
PROPIETARIO_MUNICIPIO VARCHAR(50),
ESTADO VARCHAR(20))
```

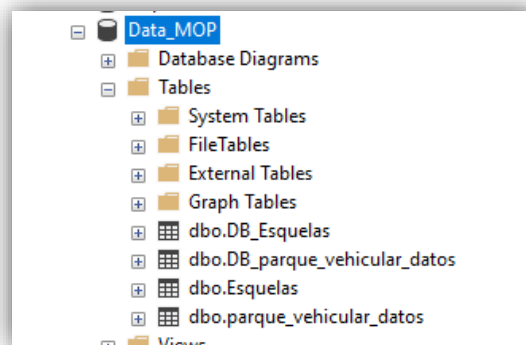
Insert donde filtramos los campos para tener la información limpia:

```
INSERT INTO DB_parque_vehicular_datos (TIPO_PLACA, ANIO_DE_FABRICACION, CILINDRADA, CANTIDAD_DE_CILINDROS, CANTIDAD_DE_PUERTAS, VALOR_DEL_VEHICULO, COLORES,
FECHA_DE_IMPORTACION, IMP_VALOR_DEL_VEHICULO, FECHA_DE_INGRESO, ANIO_INGRESO, MES_INGRESO, CLASE, PERTENENCIA, MARCA,
MODELO, CAPACIDAD, COMBUSTIBLE, ADUANA, CONDICION_INGRESO, PROPIETARIO_DEPARTAMENTO, PROPIETARIO_MUNICIPIO, ESTADO)
SELECT TIPO_PLACA, ANIO_DE_FABRICACION, CILINDRADA, CANTIDAD_DE_CILINDROS, CANTIDAD_DE_PUERTAS, VALOR_DEL_VEHICULO, COLORES,
FECHA_DE_IMPORTACION, IMP_VALOR_DEL_VEHICULO, FECHA_DE_INGRESO, ANIO_INGRESO, MES_INGRESO, CLASE, PERTENENCIA, MARCA,
MODELO, CAPACIDAD, COMBUSTIBLE, ADUANA, CONDICION_INGRESO, PROPIETARIO_DEPARTAMENTO, PROPIETARIO_MUNICIPIO, ESTADO
FROM parque_vehicular_datos
```

100 %
Messages
(1048575 rows affected)
Completion time: 2022-04-21T20:41:26.9991516-06:00

Las tablas que utilizaremos para elaborar las estrategias de minerías de datos son las siguientes:

- DB_Esquelas
- DB_parque_vehicular_datos

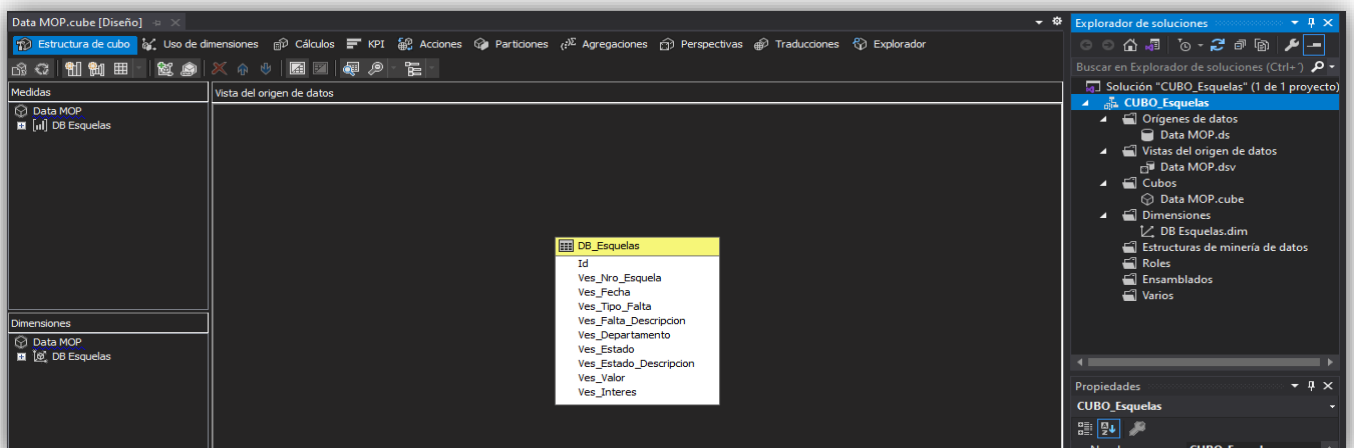
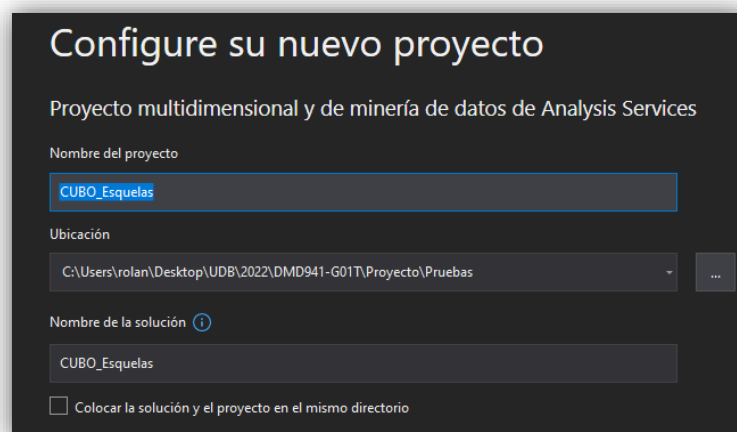
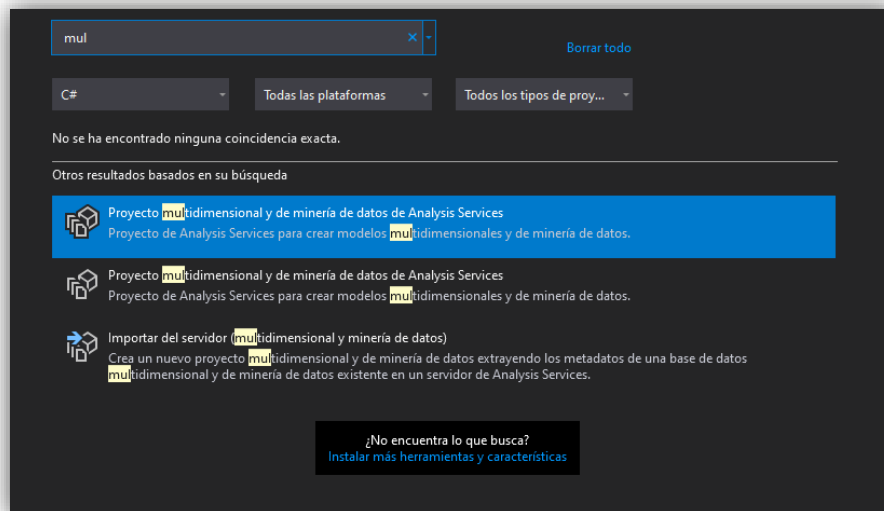


Estrategia de minería de Datos CUBOS OLAP

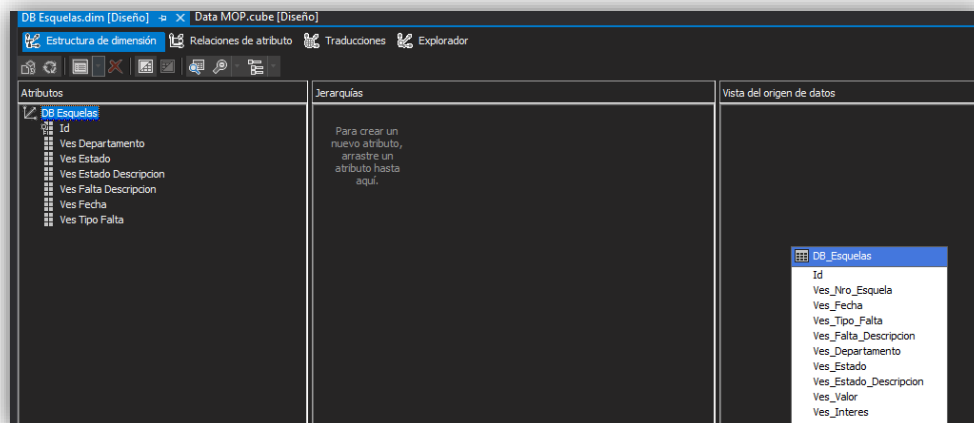
Esquelas:

Creación de proyecto:

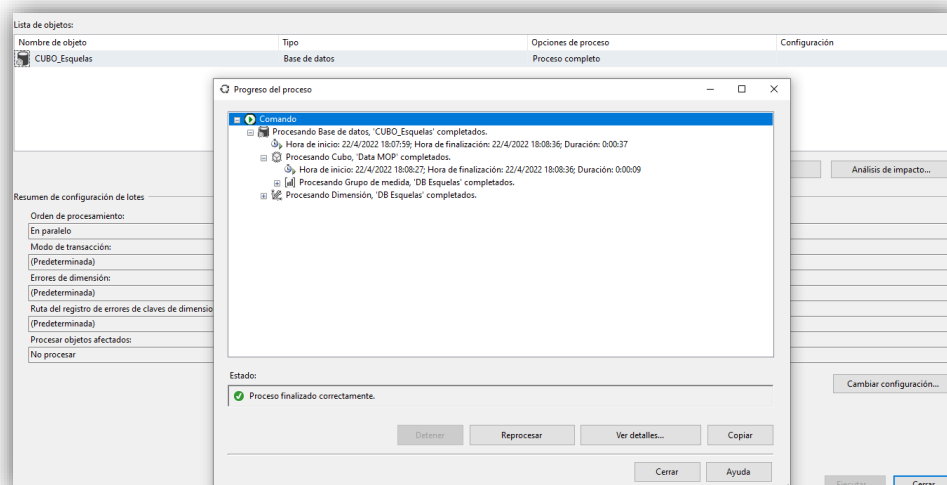
El cubo nos permite poder revisar los campos que mejor nos ayude a tomar un KPI para poder aplicar una estrategia:



Modificamos las dimensiones para poder tener un mejor control de toda la información:



Procesamos el cubo para verificar que todo se encuentre bien:



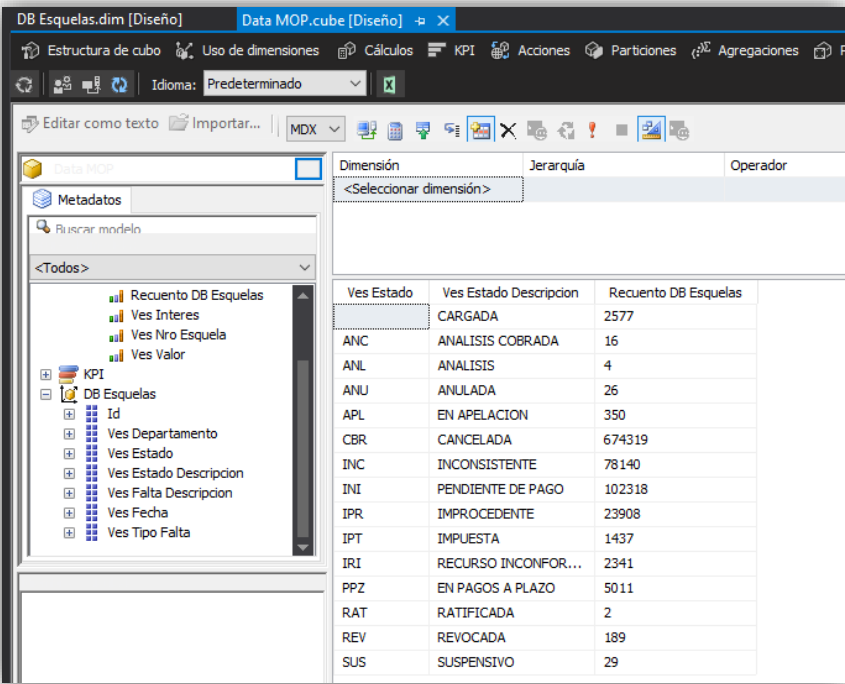
Al modificar la dimension y agregar los demas campos para podemos realizar consultas mas exactas y rapidas gracias al CUBO:

Ejemplo: Calculo de intereses por departamento:

Dimension	Jerarquia
<Seleccionar dimension>	
Ves Departamento	Ves Interes
AHUACHAPAN	20082.71
CABAÑAS	5493.03
CHALATENANGO	13943.21
CUSCATLAN	42232.59
LA LIBERTAD	108942.89
LA PAZ	89941.5899999...
LA UNION	11726.26
MORAZAN	3766.11
SAN MIGUEL	83245.8099999...
SAN SALVADOR	111380.97999...
SAN VICENTE	27123.97
SANTA ANA	58080.0599999...
SONSONATE	57282.2799999...
USulután	30182.12

Ejemplo 2:

Cantidad de esquelas por su estado y descripcion:

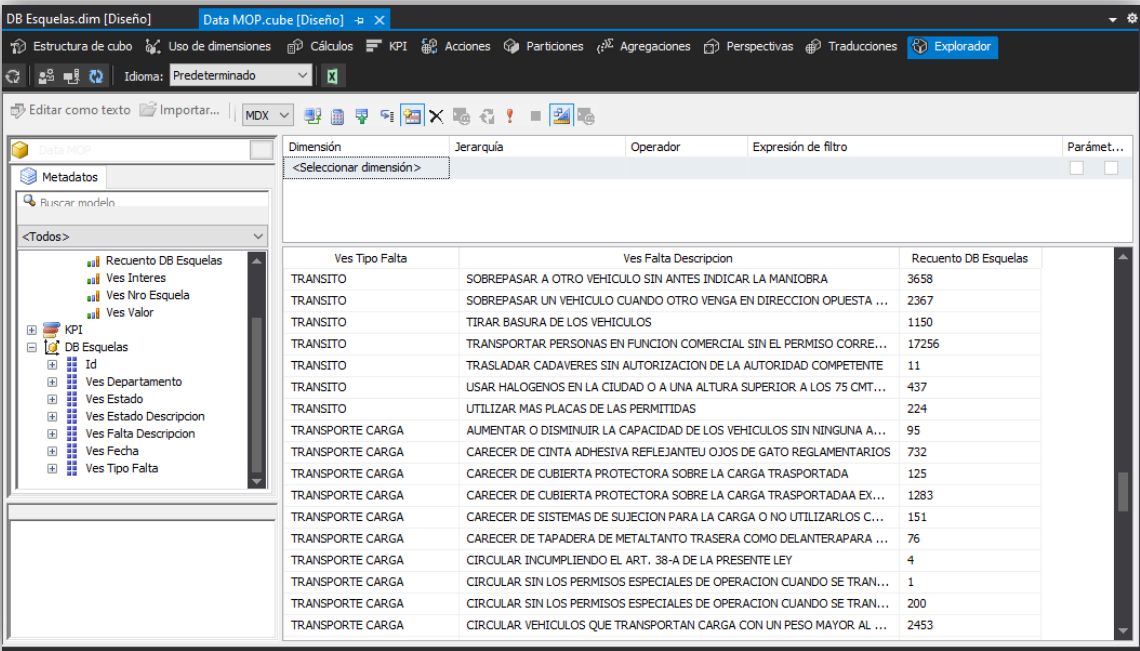


Dimension	Jerarquia	Operador
<Seleccionar dimension>		

Ves Estado	Ves Estado Descripcion	Recuento DB Esquelas
	CARGADA	2577
ANC	ANALISIS COBRADA	16
ANL	ANALISIS	4
ANU	ANULADA	26
APL	EN APELACION	350
CBR	CANCELADA	674319
INC	INCONSISTENTE	78140
INI	PENDIENTE DE PAGO	102318
IPR	IMPROCEDENTE	23908
IPT	IMPUESTA	1437
IRI	RECURSO INCONFOR...	2341
PPZ	EN PAGOS A PLAZO	5011
RAT	RATIFICADA	2
REV	REVOCADA	189
SUS	SUSPENSIVO	29

Ejemplo 3:

Cantidad de esquelas por tipo de esquila:

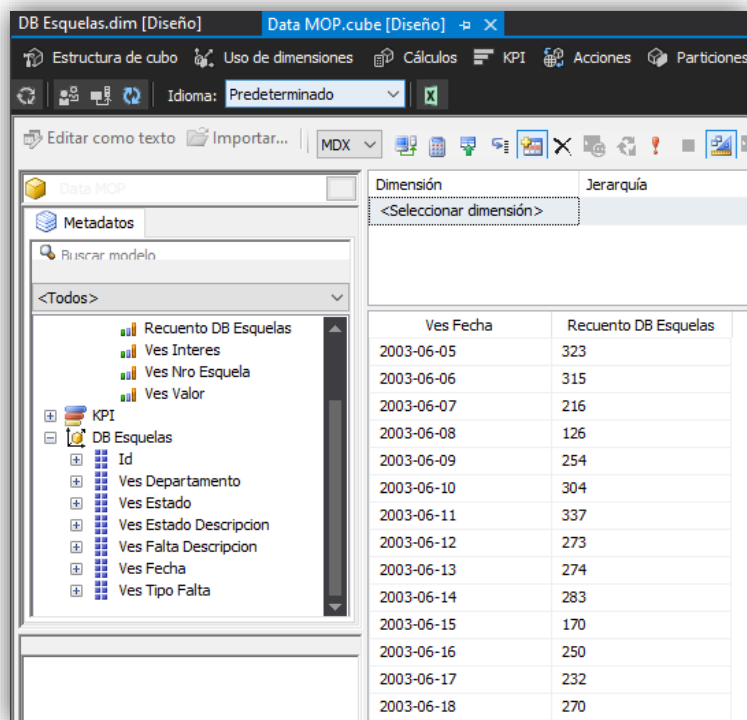


Dimension	Jerarquia	Operador	Expresion de filtro	Paramet...
<Seleccionar dimension>				

Ves Tipo Falta	Ves Falta Descripcion	Recuento DB Esquelas
TRANSITO	SOBREPASAR A OTRO VEHICULO SIN ANTES INDICAR LA MANIOBRA	3658
TRANSITO	SOBREPASAR UN VEHICULO CUANDO OTRO VENGA EN DIRECCION OPUESTA ...	2367
TRANSITO	TIRAR BASURA DE LOS VEHICULOS	1150
TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL PERMISO CORRE...	17256
TRANSITO	TRASLADAR CADAVERES SIN AUTORIZACION DE LA AUTORIDAD COMPETENTE	11
TRANSITO	USAR HALOGENOS EN LA CIUDAD O A UNA ALTURA SUPERIOR A LOS 75 CMT...	437
TRANSITO	UTILIZAR MAS PLACAS DE LAS PERMITIDAS	224
TRANSPORTE CARGA	AUMENTAR O DISMINUIR LA CAPACIDAD DE LOS VEHICULOS SIN NINGUNA A...	95
TRANSPORTE CARGA	CARECER DE CINTA ADHESIVA REFLEJANTEU OJOS DE GATO REGLAMENTARIOS	732
TRANSPORTE CARGA	CARECER DE CUBIERTA PROTECTORA SOBRE LA CARGA TRASPORTADA	125
TRANSPORTE CARGA	CARECER DE CUBIERTA PROTECTORA SOBRE LA CARGA TRASPORTADAA EX...	1283
TRANSPORTE CARGA	CARECER DE SISTEMAS DE SUJECION PARA LA CARGA O NO UTILIZARLOS C...	151
TRANSPORTE CARGA	CARECER DE TAPADERA DE METALTANTO TRASERA COMO DELANTERAPARA ...	76
TRANSPORTE CARGA	CIRCULAR INCUMPLIENDO EL ART. 38-A DE LA PRESENTE LEY	4
TRANSPORTE CARGA	CIRCULAR SIN LOS PERMISOS ESPECIALES DE OPERACION CUANDO SE TRAN...	1
TRANSPORTE CARGA	CIRCULAR SIN LOS PERMISOS ESPECIALES DE OPERACION CUANDO SE TRAN...	200
TRANSPORTE CARGA	CIRCULAR VEHICULOS QUE TRANSPORTAN CARGA CON UN PESO MAYOR AL ...	2453

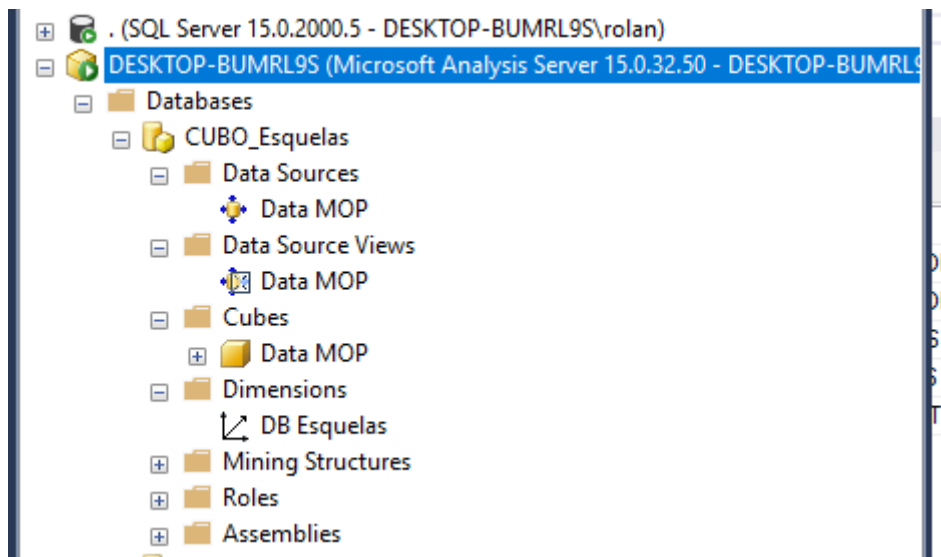
Ejemplo 4:

Podemos ver cuantas esquelas se colocaron por fecha:



Ves Fecha	Recuento DB Esquelas
2003-06-05	323
2003-06-06	315
2003-06-07	216
2003-06-08	126
2003-06-09	254
2003-06-10	304
2003-06-11	337
2003-06-12	273
2003-06-13	274
2003-06-14	283
2003-06-15	170
2003-06-16	250
2003-06-17	232
2003-06-18	270

Desde SQL server podemos ver en microsoft analysis server la estructura de nuestro CUBO:



De la misma manera podemos seguir realizando consultas según lo requerido de una manera mas ágil.

Estrategia de minería de Datos Reglas de Asociación

Esquelas:

Crearemos una tabla para realizar este analisis, nuestro KPI sera el estado y el deparamento para poder analizar la información:

```
--CREAREMOS UNA TABLA PARA REALIZAR EL ANALISIS DE REGLAS DE ASOCIACION  
  
CREATE TABLE Reglas_Asociacion_Esquelas (Ves_Estado VARCHAR(20),  
SanSalvador BIT,  
Usulután BIT,  
LaLibertad BIT,  
SanVicente BIT,  
LaPaz BIT,  
Ahuchapan BIT,  
Morazan BIT,  
SanMiguel BIT,  
Cuscatlan BIT,  
LaUnion BIT,  
SantaAna BIT,  
Chalatenango BIT)
```

```
--CON ESTE QUERY SABES POR DEPARTAMENTO CUALES ESTADOS DE ESQUELAS SON LAS MAS COMUNES Y SI APARECEN EN TODOS LOS ESTADOS  
SELECT Ves_Departamento , Ves_Estado, COUNT(Ves_Estado) AS 'COUNT ESTADO' FROM DB_Esquelas GROUP BY Ves_Departamento , Ves_Estado
```

	Ves_Departamento	Ves_Estado	COUNT ESTADO
1	LA LIBERTAD	CBR	43969
2	LA PAZ	INC	5253
3	SANTA ANA	CBR	38327
4	AHUACHAPAN	INC	1961
5	LA PAZ	IPR	927
6	SONSONATE		80
7	AHUACHAPAN	INI	1093
8	LA LIBERTAD	PPZ	249
9	CHALATENANGO	IPT	24
10	SAN SALVADOR	REV	130
11	SAN VICENTE	REV	9
12	SONSONATE	CBR	36252
13	LA PAZ	CBR	70314
14	CHALATENANGO	INC	2443
15	SAN SALVADOR		1880
16	SAN VICENTE		16

```
--INSERTAMOS EN NUESTRA TABLA PARA EL ANALISIS DE REGLAS DE ASOCIACION
--PRIMERO INSERTAREMOS TODOS LOS ESTADOS Y LUEGO HAREMOS UN UPDATE A CADA REGISTRI

INSERT INTO Reglas_Asociacion_Esquelas (Ves_Estado)
VALUES ('ANC')
INSERT INTO Reglas_Asociacion_Esquelas (Ves_Estado)
VALUES ('IPR')
INSERT INTO Reglas_Asociacion_Esquelas (Ves_Estado)
VALUES ('INI')
INSERT INTO Reglas_Asociacion_Esquelas (Ves_Estado)
VALUES ('INC')
INSERT INTO Reglas_Asociacion_Esquelas (Ves_Estado)
VALUES ('')
```

Messages

(1 row affected)

(1 row affected)

(1 row affected)

(1 row affected)

Nuestras tabla quedaria de la siguiente manera:

```
SELECT * FROM Reglas_Asociacion_Esquelas
```

	Ves_Estado	SanSalvador	Usulután	LaLibertad	SanVicente	LaPaz	Ahuchapan	Morazan	SanMiguel	Cuscatlan	LaUnion	SantaAna	Chalatenango
1	ANC	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
2	IPR	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
3	INI	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
4	INC	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
5		NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
6	CBR	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Procedemos a llenar las columnas con 0 o 1:

Para ellos metemos nuestro group by en un tabla para poder manipular los datos:

```
CREATE TABLE DATOS_GROUP_BY (Ves_Departamento VARCHAR(30),Ves_Estado VARCHAR(30),CountEstado INT)

INSERT INTO DATOS_GROUP_BY
--CON ESTE QUERY SABES POR DEPARTAMENTO CUALES ESTADOS DE ESQUELAS SON LAS MAS COMUNES Y SI APARECEN EN TODOS LOS ESTADOS
SELECT Ves_Departamento , Ves_Estado, COUNT(Ves_Estado) AS 'COUNT ESTADO' FROM DB_Esquelas GROUP BY Ves_Departamento , Ves_Estado
```

Empezamos a realizar los update:

Este update lo realizaremos con cada uno de los estados:

```

UPDATE Reglas_Asociacion_Esuelas SET SanSalvador = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'SAN SALVADOR' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
SantaAna = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'SANTA ANA' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
Cuscatlan = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'CUSCATLAN' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
Morazan = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'MORAZAN' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
SanMiguel = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'SAN MIGUEL' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
Ahuchapan = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'AHUACHAPAN' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
SanVicente = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'SAN VICENTE' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
LaPaz = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'LA PAZ' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
LaUnion = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'LA UNION' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,
Usulután = CASE WHEN (SELECT COUNT(*) FROM DATOS_GROUP_BY WHERE Ves_Departamento = 'USULUTAN' AND Ves_Estado='ANC')>=1 THEN 1
ELSE 0 END,

```

(1 row affected)

Completion time: 2022-04-23T10:17:15.9538485-06:00

Ya tenemos lista nuestra tabla para realizar el analisis siguiente:

	Ves_Estado	SanSalvador	Usulután	LaLibertad	SanVicente	LaPaz	Ahuchapan	Morazan	SanMiguel	Cuscatlan	LaUnion	SantaAna	Chalatenango
1	ANC	1	0	1	0	1	0	1	0	0	1	1	0
2	IPR	1	1	1	1	1	1	1	1	1	1	1	1
3	INI	1	1	1	1	1	1	1	1	1	1	1	1
4	INC	1	1	1	1	1	1	1	1	1	1	1	1
5		1	1	1	1	1	1	1	1	1	1	1	1
6	CBR	1	1	1	1	1	1	1	1	1	1	1	1
7	IPT	1	1	1	1	1	1	1	1	1	1	1	1
8	SUS	1	0	0	0	0	0	0	0	0	0	0	0
9	RAT	1	0	0	0	0	0	0	1	0	0	0	0
10	APL	1	0	1	0	1	0	0	1	1	0	1	1
11	ANL	1	0	0	1	0	0	0	0	1	0	0	0
12	PPZ	1	1	1	1	1	1	1	1	1	1	1	1
13	IRI	1	1	1	1	1	1	1	1	1	1	1	1
14	REV	1	0	1	1	1	0	0	0	1	0	1	1
15	ANU	1	1	1	0	0	0	0	0	0	0	1	0

Pasaremos los datos anteriores a un archivo excel y nos quedara de esta manera en formato csv.

	A	B	C	D	E	F	G	H	I	J	K
1	Ves_Estado,	SanSalvador,	Usulután,	LaLibertad,	SanVicente,	LaPaz,	Ahuchapan,	Morazan,	SanMiguel,	Cuscatlan,	LaUnion,SantaAna,Chalatenango
2	ANC,	1,0,1,0,1,0,1,0,0,1,1,0									
3	IPR,	1,1,1,1,1,1,1,1,1,1,1,1									
4	INI,	1,1,1,1,1,1,1,1,1,1,1,1									
5	INC,	1,1,1,1,1,1,1,1,1,1,1,1									
6	,	1,1,1,1,1,1,1,1,1,1,1,1									
7	CBR,	1,1,1,1,1,1,1,1,1,1,1,1									
8	IPT,	1,1,1,1,1,1,1,1,1,1,1,1									
9	SUS,	1,0,0,0,0,0,0,0,0,0,0,0									
10	RAT,	1,0,0,0,0,0,0,0,0,0,0,0									
11	APL,	1,0,0,0,0,0,0,0,0,0,0,0									
12	ANL,	1,0,0,0,0,0,0,0,0,0,0,0									
13	PPZ,	1,1,1,1,1,1,1,1,1,1,1,1									
14	IRI,	1,1,1,1,1,1,1,1,1,1,1,1									
15	REV,	1,0,1,1,1,0,0,0,0,0,0,0									
16	ANU,	1,1,1,0,0,0,0,0,0,0,0,0									
17											

Este tipo de reglas nos permiten encontrar las combinaciones de artículos o ítems que ocurren con mayor frecuencia en una base de datos transaccional, de igual manera nos permiten medir la importancia de esas combinaciones o relaciones que existen.

Vamos a limitar las reglas de asociación y para hacerlo haremos una métrica:

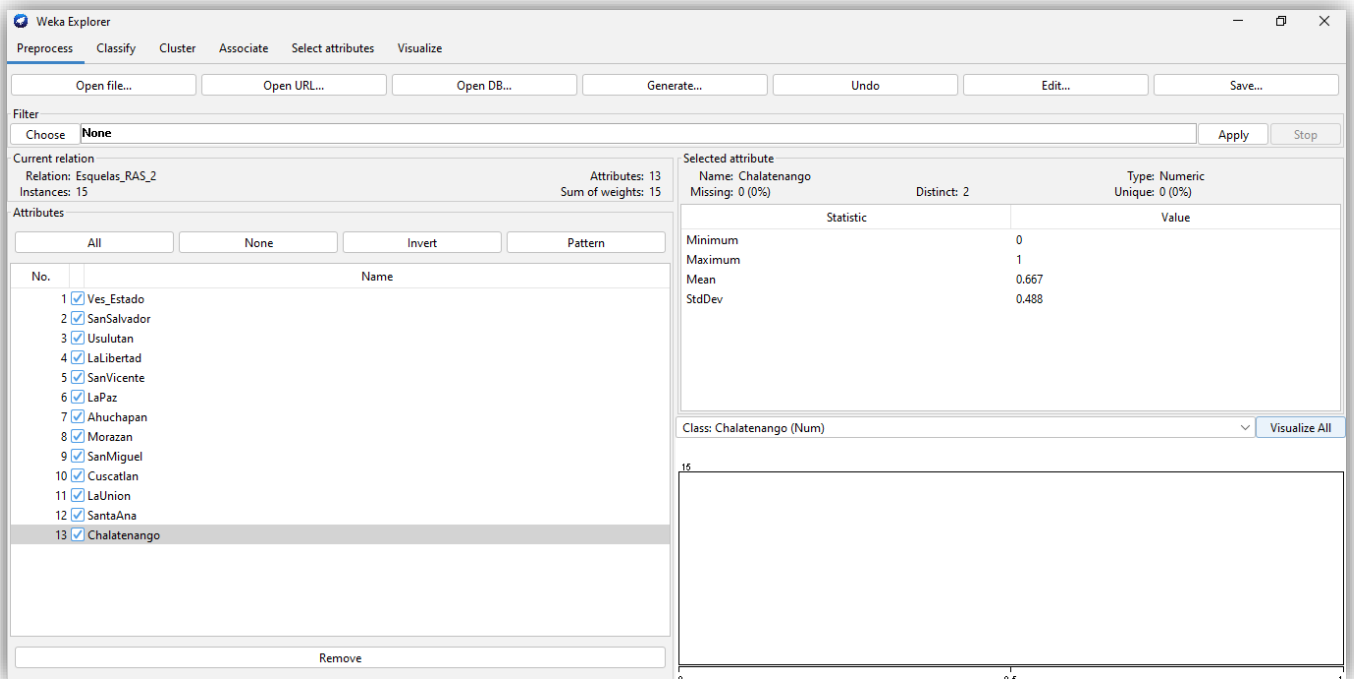
Soporte: Dada una regla “Si A=> B”, el soporte de esta regla se define como el número de veces o la frecuencia (relativa) con que A y B aparecen juntos en una base de datos.

Para las reglas de asociación utilizaremos un programa llamado WEKA:

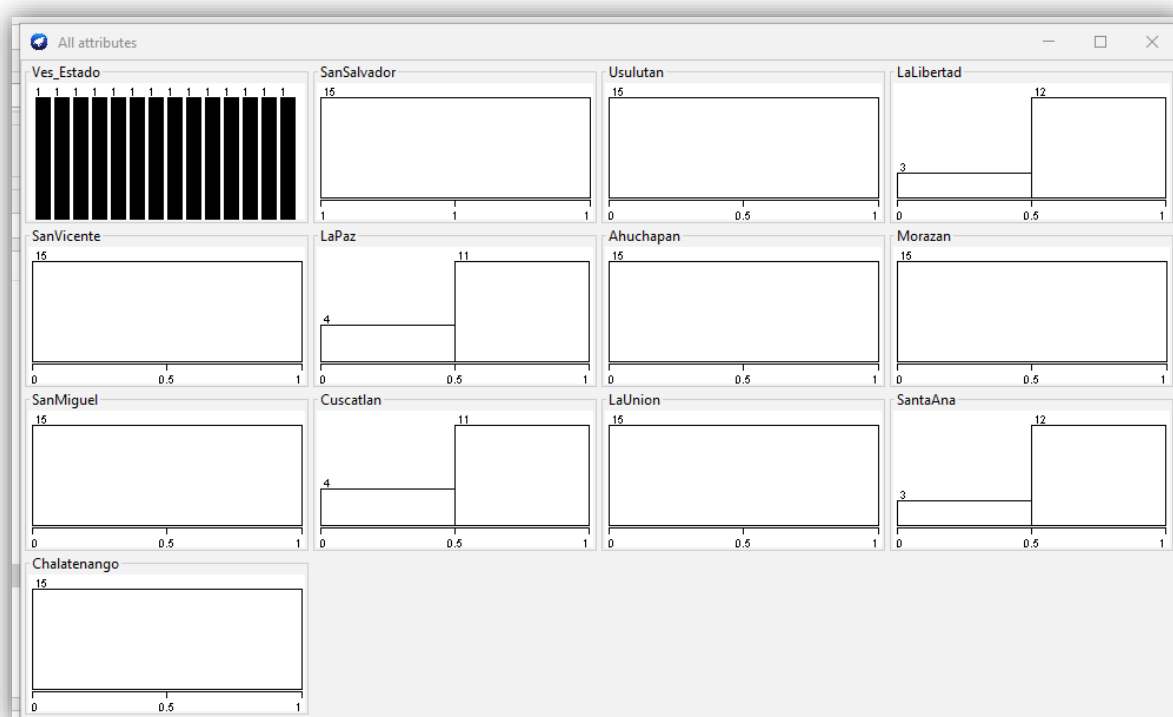


Al darle clic en explorer se nos abrirá la siguiente ventana donde cargaremos nuestros datos:

Cargamos una muestra de nuestro archivo Esquelas:



Podemos observar que podemos analizar los datos por departamento y verificar que estamos son los más comunes, es decir analizar por departamento como se manejan los estados para agilizar el proceso o verificar que esta sucediendo.



Este análisis también podríamos hacer por el tipo de esquelas que se aplican por departamento y verificar que infracciones se comente más.

Estrategia de Agrupamiento con k-means

DB_parque_vehicular_datos:

Preparamos un archivo excel con una muestra de los datos para mostrar el analisis k-means:

Parque_Vehicular - Excel

ARCHIVO

INICIO

INSERTAR

DISÑO DE PÁGINA

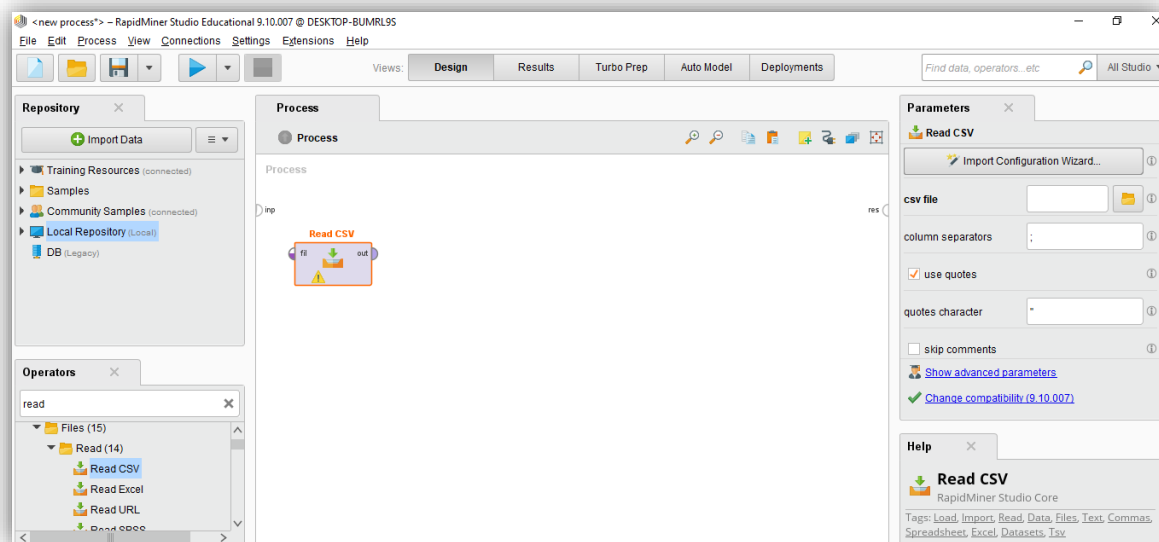
FÓRMULAS

DATOS

REVISAR

VISTA

En esta parte utilizaremos el programa RapiMiner para el tratamiento de la información y lo trabajaremos con el archivo excel:



Vista de los registros:

Import Data - Specify your data format

Specify your data format

☒ Header Row
Start Row
Column Separator

File Encoding
Escape Character
Decimal Character

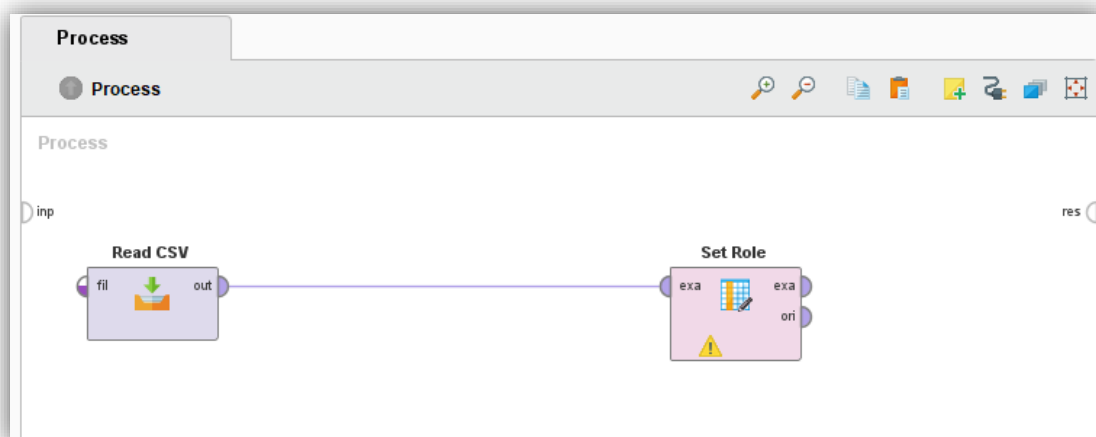
☒ Use Quotes
☐ Trim Lines
☒ Skip Comments

1	TIPO_PL...	ANIO_DE...	CILINDR...	CANTIDA...	CANTIDA...	VALOR_...	COLORES	FECHA_...	IMP_VA...	FECHA...
2	PARTIC...	1990	1800	0	4	4094.56	AMARILLO	1994-09...	4094.56	1994-1
3	PARTIC...	1964	0	0	0	0	AMARILLO		0	1989-0
4	ALQUILER	1984	1700	0	0	0	AMARILL...		0	1985-0
5	ALQUILER	1986	1600	0	0	0	AMARILL...		0	1988-0
6	ALQUILER	1979	0	0	0	0	AMARILLO		0	1982-1
7	PARTIC...	1974	1600	0	0	0	AMARILLO		0	1982-1
8	ALQUILER	1975	0	0	0	0	AMARILLO		0	1984-0
9	ALQUILER	1973	0	0	0	0	AMARILL...		0	1988-0
10	ALQUILER	1975	0	0	0	0	AMARILLO		0	1988-0

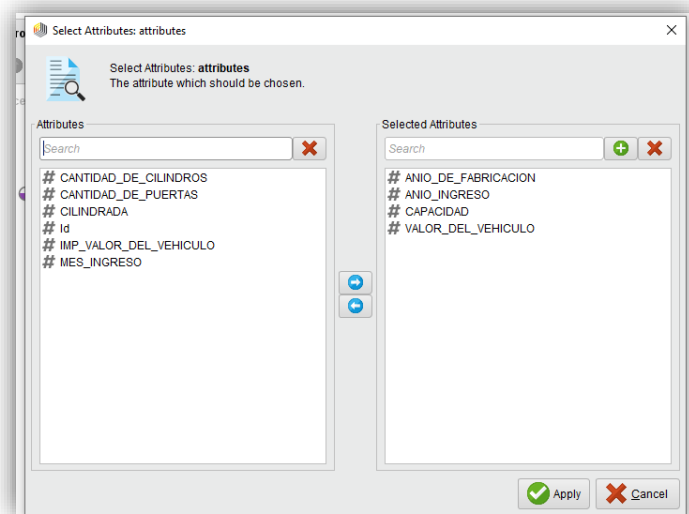
no problems.

Previous Next Cancel

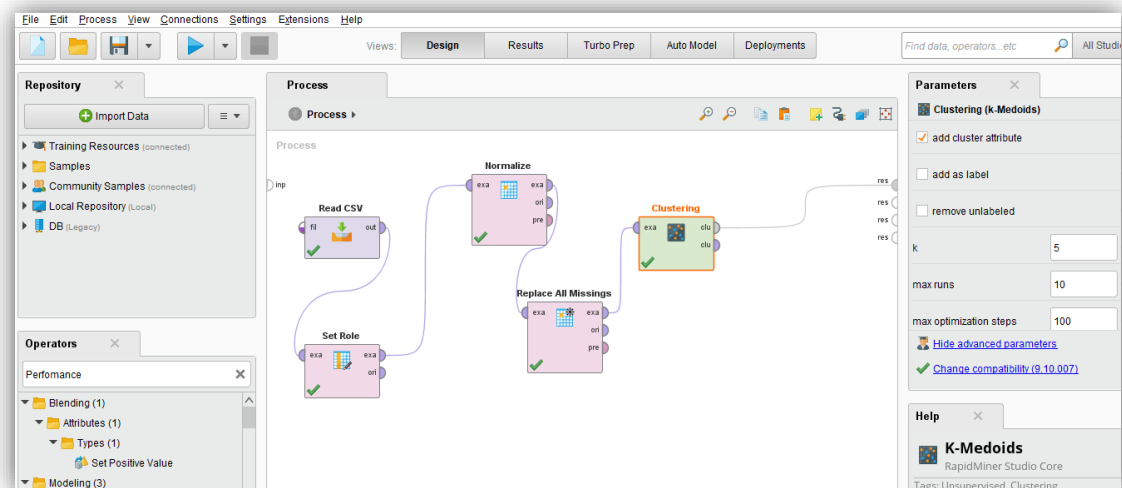
Agregamos un Set Role:



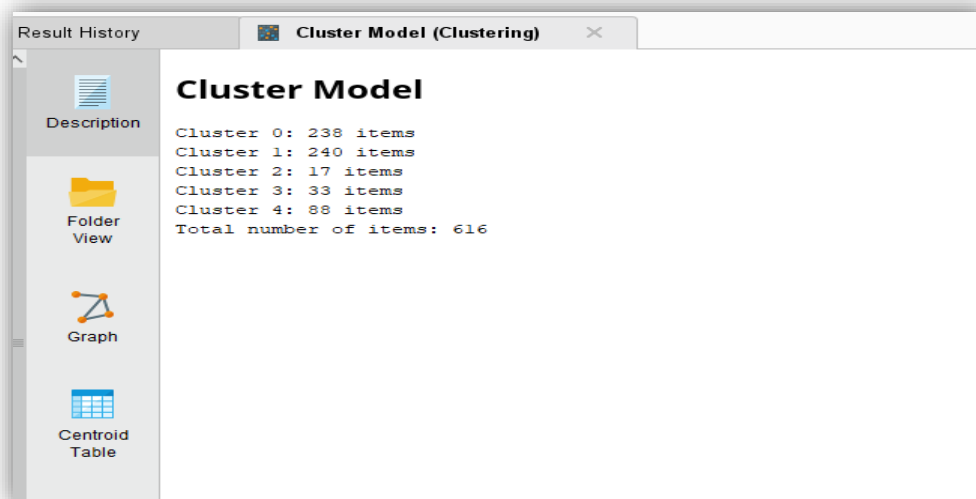
Preparamos los atributos:



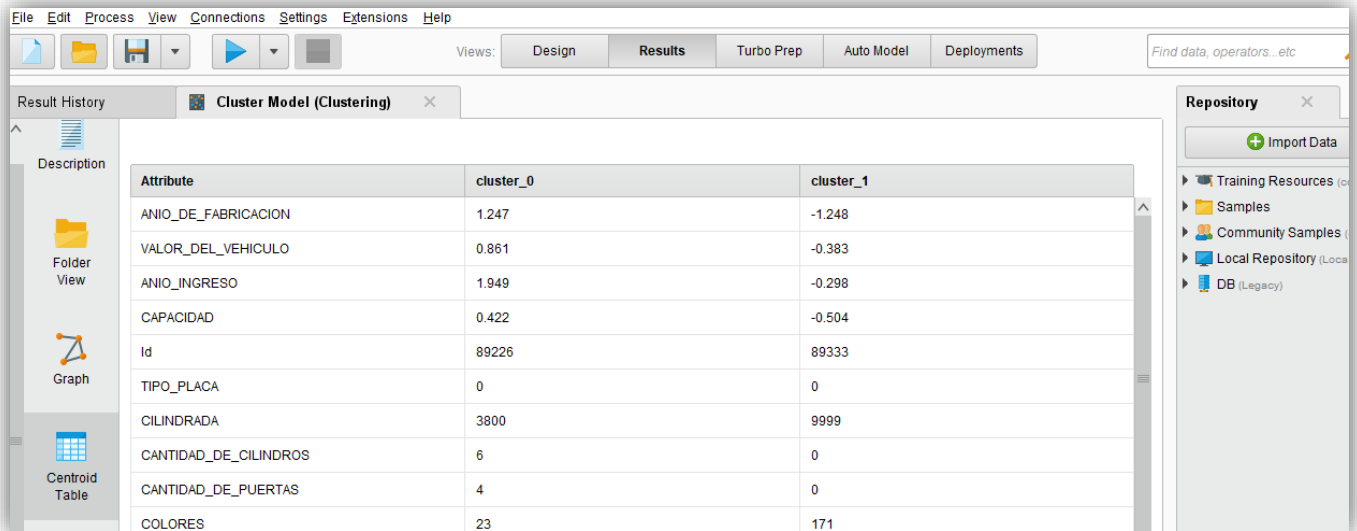
Nuestro proyecto quedaria de la siguiente manera:



Ejecutamos el proceso:



Resultado:

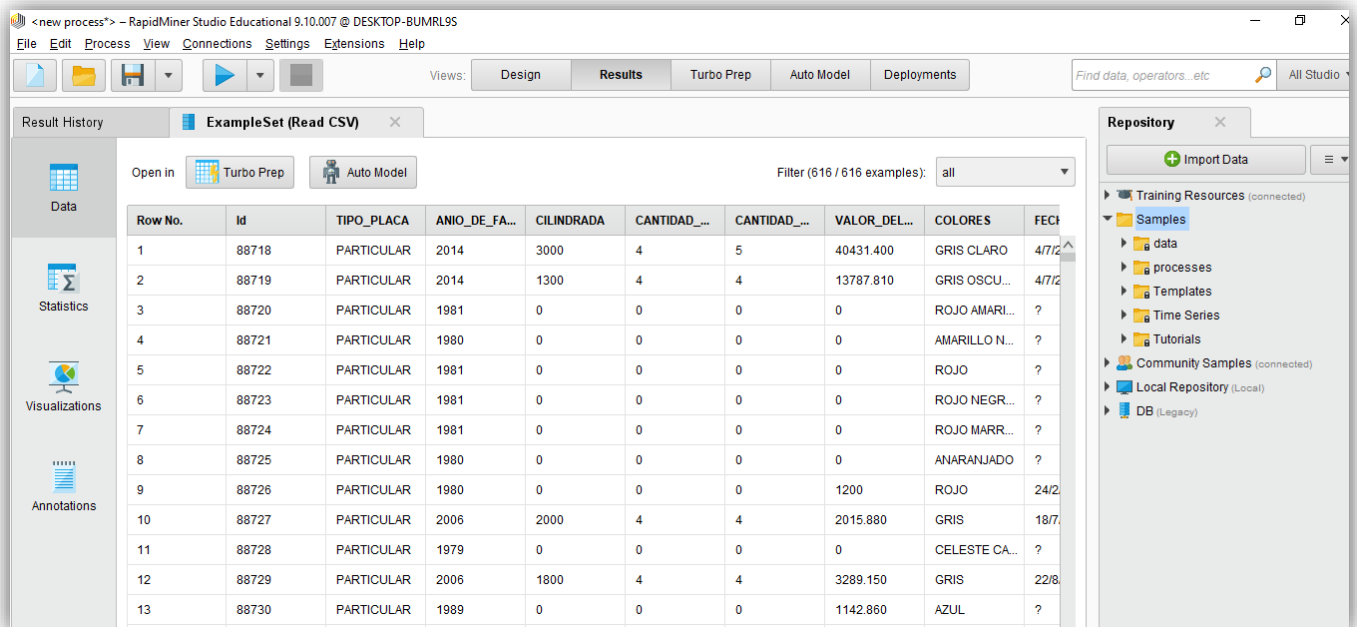


Attribute	cluster_0	cluster_1
ANIO_DE_FABRICACION	1.247	-1.248
VALOR_DEL_VEHICULO	0.861	-0.383
ANIO_INGRESO	1.949	-0.298
CAPACIDAD	0.422	-0.504
Id	89226	89333
TIPO_PLACA	0	0
CILINDRADA	3800	9999
CANTIDAD_DE_CILINDROS	6	0
CANTIDAD_DE_PUERTAS	4	0
COLORES	23	171

Estrategia de Árbol de Decisión

DB_parque_vehicular_datos:

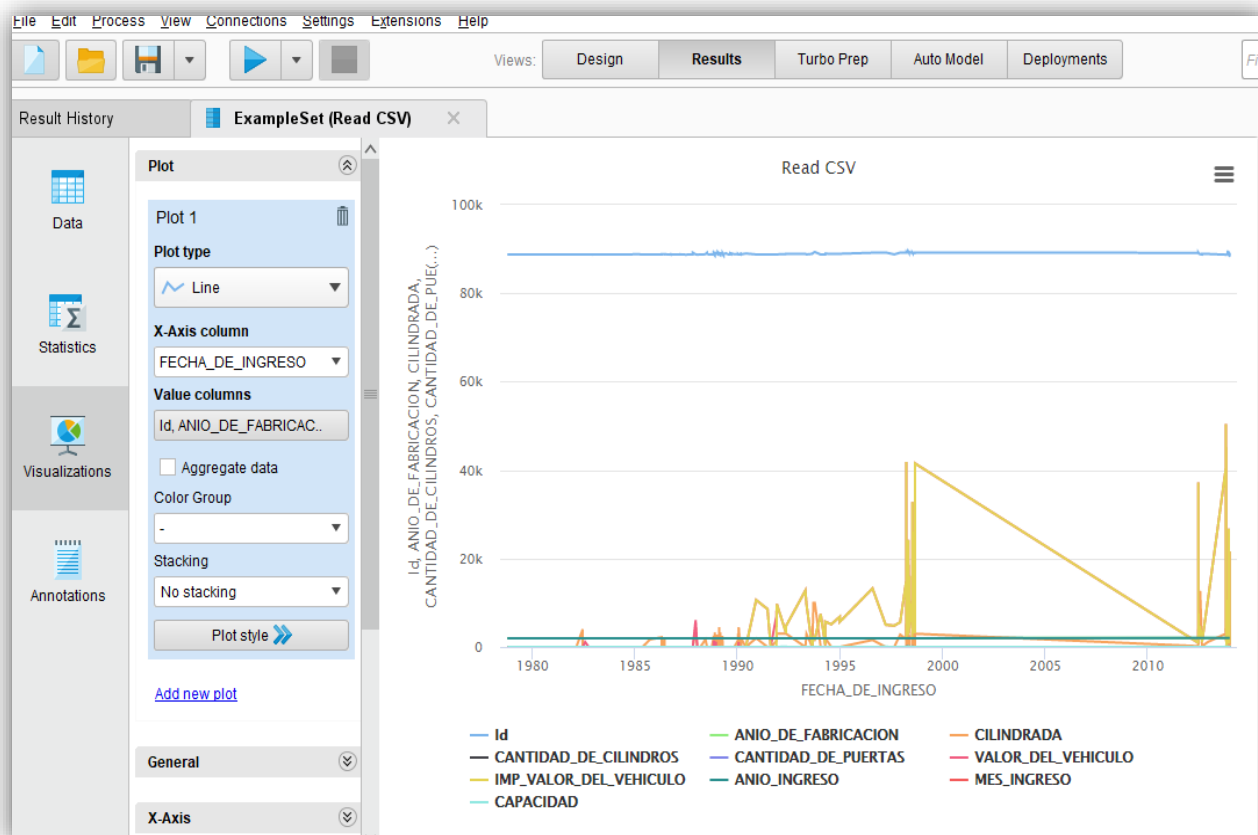
Creamos un nuevo proyecto en Rapidminer y cargamos nuestro archivo:



Row No.	Id	TIPO_PLACA	ANIO_DE_FA...	CILINDRADA	CANTIDAD_...	CANTIDAD_...	VALOR_DEL...	COLORES	FECH...
1	88718	PARTICULAR	2014	3000	4	5	40431.400	GRIS CLARO	4/7/2...
2	88719	PARTICULAR	2014	1300	4	4	13787.810	GRIS OSCU...	4/7/2...
3	88720	PARTICULAR	1981	0	0	0	0	ROJO AMARI...	?
4	88721	PARTICULAR	1980	0	0	0	0	AMARILLO N...	?
5	88722	PARTICULAR	1981	0	0	0	0	ROJO	?
6	88723	PARTICULAR	1981	0	0	0	0	ROJO NEGR...	?
7	88724	PARTICULAR	1981	0	0	0	0	ROJO MARR...	?
8	88725	PARTICULAR	1980	0	0	0	0	ANARANJADO	?
9	88726	PARTICULAR	1980	0	0	0	1200	ROJO	24/2...
10	88727	PARTICULAR	2006	2000	4	4	2015.880	GRIS	18/7...
11	88728	PARTICULAR	1979	0	0	0	0	CELESTE CA...	?
12	88729	PARTICULAR	2006	1800	4	4	3289.150	GRIS	22/8...
13	88730	PARTICULAR	1989	0	0	0	1142.860	AZUL	?

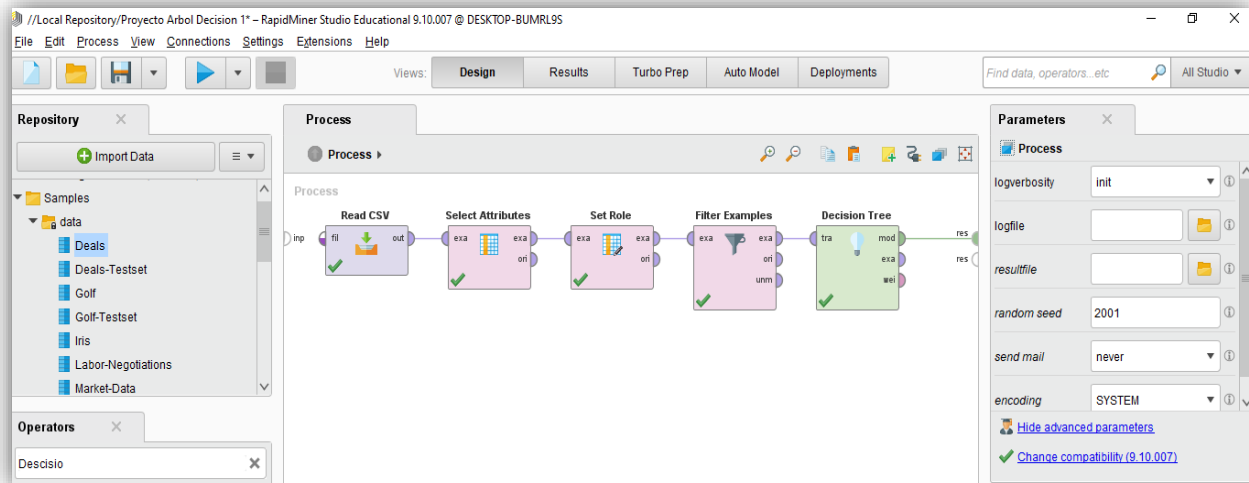
Verificamos las estadísticas:

Name	Type	Missing	Statistics			Filter (24 / 24 attributes):
Id	Integer	0	Min	Max	Average	Search for Attributes
TIPO_PLACA	Nominal	0	Least	Most	Values	
ANIO_DE_FABRICACION	Integer	0	Min	Max	Average	
CILINDRADA	Integer	2	Min	Max	Average	
CANTIDAD_DE_CILINDROS	Integer	17	Min	Max	Average	
CANTIDAD_DE_PUERTAS	Integer	28	Min	Max	Average	
VALOR_DEL_VEHICULO	Real	0	Min	Max	Average	
COLORES	Nominal	0	Least	Most	Values	



Árbol de Decisión:

Con el proyecto anterior le agregamos filtros y el árbol de decisión, en este caso utilizaremos como nuestro kpi el año de ingreso del automóvil.



Y nuestro árbol quedaría de la siguiente manera:

