

Elementos Fundamentales de SQL

Cuarta parte

Group by (agrupando registros)

Podemos agrupar registros de nuestra tabla, de manera que obtengamos un único registro por cada grupo.

Para establecer los grupos lo haremos especificando los campos que lo generan.

Aparecerá un único resultado para cada valor diferente de los campos del grupo.

Formato:

select

from

group by campo1, campo2,...

Partimos de un conjunto de registros. Para tres fechas, se contó el paro y la población de todas las comunidades autónomas, todas sus provincias y cada uno de sus municipios. 24351 filas.

Resultados Mensajes								
	id	CA	Provincia	ISLA	Municipio	Fecha	Padron	TotalParoRegistrado
3514	3514	Cataluña	Barcelona	NULL	Palma de Cervelló, ...	2013-01-01 00:00:00.000	3023	3023
3515	3515	Cataluña	Barcelona	NULL	Palma de Cervelló, ...	2013-02-01 00:00:00.000	3023	3023
3516	3516	Cataluña	Barcelona	NULL	Palma de Cervelló, ...	2013-03-01 00:00:00.000	3023	3023
3517	3517	Castilla y León	Burgos	NULL	Abajas	2013-01-01 00:00:00.000	29	29
3518	3518	Castilla y León	Burgos	NULL	Abajas	2013-02-01 00:00:00.000	29	29
3519	3519	Castilla y León	Burgos	NULL	Abajas	2013-03-01 00:00:00.000	29	29
3520	3520	Castilla y León	Burgos	NULL	Adrada de Haza	2013-01-01 00:00:00.000	243	243
3521	3521	Castilla y León	Burgos	NULL	Adrada de Haza	2013-02-01 00:00:00.000	243	243
3522	3522	Castilla y León	Burgos	NULL	Adrada de Haza	2013-03-01 00:00:00.000	243	243


```
--Mostrar las comunidades autónomas
select ca
from datoscompletotabla
group by ca;
go
```

Al agrupar por CA saca sólo 19 resultados

Resultados Mensajes	
	ca
1	Cantabria
2	Melilla
3	Galicia
4	Castilla y León
5	Madrid, Comunidad de
6	Aragón
7	Andalucía
8	Canarias
9	Asturias, Principado de
10	País Vasco
11	Rioja, La
12	Extremadura
13	Balears, Illes
14	Murcia, Región de
15	Ceuta
16	Castilla - La Mancha
17	Cataluña
18	Navarra, Comunidad ...
19	Comunitat Valenciana

```
--podemos ordenarlas  
select ca  
from datoscompletostabla  
group by ca  
order by ca;  
go
```

Resultados		Mensajes
	ca	
1	Andalucía	
2	Aragón	
3	Asturias, Principado de	
4	Balears, Illes	
5	Canarias	
6	Cantabria	
7	Castilla - La Mancha	
8	Castilla y León	
9	Cataluña	
10	Ceuta	
11	Comunitat Valenciana	
12	Extremadura	
13	Galicia	
14	Madrid, Comunidad de	
15	Melilla	
16	Murcia, Región de	
17	Navarra, Comunidad Foral de	
18	País Vasco	
19	Rioja, La	

```
--podemos filtrarlas
--sólo las que tengan algún dato con padron >300000
select ca
from datoscompletotabla
where padron>300000
group by ca
order by ca;
go
```

Hace los grupos sobre los datos que quedan tras el filtro

Resultados		Mensajes
	ca	
1	Andalucía	
2	Aragón	
3	Balears, Illes	
4	Canarias	
5	Castilla y León	
6	Cataluña	
7	Comunitat Valenciana	
8	Madrid, Comunidad de	
9	Murcia, Región de	
10	País Vasco	

Sólo saca los datos de las CA que tienen algún municipio con más de 300000 habitantes en alguna de las tres fechas con datos.

Es importante el orden de cada cláusula.

1. select
2. from
3. where
4. group by
5. order by

```
--podemos agrupar sobre varios campos
select ca, provincia
from datoscompletostabla
group by ca, provincia
order by ca,provincia;
go
```

Nos dará la lista de todas las
CA y sus provincias.
52 resultados.

	ca	provincia
33	Cataluña	Lleida
34	Cataluña	Tarragona
35	Ceuta	Ceuta
36	Comunitat Valenciana	Alicante/A...
37	Comunitat Valenciana	Castellón/...
38	Comunitat Valenciana	Valencia/...
39	Extremadura	Badajoz
40	Extremadura	Cáceres
41	Galicia	Coruña, A
42	Galicia	Lugo
43	Galicia	Ourense
44	Galicia	Pontevedra
45	Madrid, Comunidad de	Madrid
46	Melilla	Melilla
47	Murcia, Región de	Murcia
48	Navarra, Comunidad ...	Navarra
49	País Vasco	Araba/Álava
50	País Vasco	Bizkaia
51	País Vasco	Gipuzkoa
52	Rioja, La	Rioja, La


```
--también filtrar los datos previamente  
select ca, provincia  
from datoscompletotabla  
where ca like '%c[íi]a%'  
group by ca, provincia  
order by ca,provincia;  
go
```

Nota: el like diferencia las
letras acentuadas de las que
no lo están

Resultados		Mensajes
	ca	provincia
1	Andalucía	Almería
2	Andalucía	Cádiz
3	Andalucía	Córdoba
4	Andalucía	Granada
5	Andalucía	Huelva
6	Andalucía	Jaén
7	Andalucía	Málaga
8	Andalucía	Sevilla
9	Comunitat Valenciana	Alicante/Alacant
10	Comunitat Valenciana	Castellón/Castelló
11	Comunitat Valenciana	Valencia/Valén...
12	Galicia	Coruña, A
13	Galicia	Lugo
14	Galicia	Ourense
15	Galicia	Pontevedra
16	Murcia, Región de	Murcia

```
--otro tipo de filtro
select ca, provincia
from datoscompletostabla
where padron>300000
group by ca, provincia
order by ca,provincia;
go
```

Resultados		Mensajes
	ca	provincia
1	Andalucía	Córdoba
2	Andalucía	Málaga
3	Andalucía	Sevilla
4	Aragón	Zaragoza
5	Balears, Illes	Balears, Illes
6	Canarias	Palmas, Las
7	Castilla y León	Valladolid
8	Cataluña	Barcelona
9	Comunitat Valenciana	Alicante/Alacant
10	Comunitat Valenciana	Valencia/València
11	Madrid, Comunidad de	Madrid
12	Murcia, Región de	Murcia
13	País Vasco	Bizkaia

Primero filtra los datos a aquellos con padron > 300000. Después agrupa lo restante. Con lo cual mostrará las CA y Provincias que tengan algún municipio con más de 300000 habitantes en alguna de las tres tomas de datos.

No podremos colocar en el select otros campos diferentes de los que agrupan.

Sólo muestra grupos, no el contenido de los grupos.

```
--error al mostrar campo  
no agrupador  
select ca,  
provincia, padron  
from datoscompletotabla  
group by ca, provincia  
go
```

```
Mens. 8120, Nivel 16, Estado 1, Línea 1  
La columna 'datoscompletotabla.Padron' de la lista de  
selección no es válida, porque no está contenida en una  
función de agregado ni en la cláusula GROUP BY.
```

Ejercicios:

En la base de datos parobasicas:

- Mostrar los municipios de la ca de la rioja
- Mostrar las provincias de Canarias, Cantabria y Galicia. Incluir el nombre de la CA en la salida.
- Mostrar los municipios que están en Islas. Incluir el nombre de la Isla en la salida.

En la base de datos turismobasicas:

- Sacar los países con datos, ordenado por el nombre del país.
- Sacar los países que hayan tenido algún dato de visitas de mujeres De 25 a 44 años de más de 30000.

Funciones agrupadoras con el Group by

Las funciones agrupadoras que hemos visto (count, sum, max, min, avg, stdev y var) aumentan sus capacidades al usarlas en sentencias sql con group by. Podemos usarlas para sacar los mismos resultados que antes pero para cada agrupación de datos que generemos. Por ejemplo, contará el número de registros por CA si agrupamos por la CA.

```

--Sacar el número de registros (municipios) por cada
--CA ordenadas descendente por ese valor.
set dateformat dmy
select ca as 'Comunidad Autónoma', count(*) as NRegistros
from datoscompletostabla
where fecha='01-01-2013'
group by ca
order by NRegistros desc
go

```

Resultados		Mensajes
	Comunidad Autónoma	NRegistros
1	Castilla y León	2248
2	Cataluña	947
3	Castilla - La Mancha	919
4	Andalucía	771
5	Aragón	731
6	Comunitat Valenciana	542
7	Extremadura	386
8	Galicia	315
9	Navarra, Comunidad Foral de	272
10	País Vasco	251
11	Madrid, Comunidad de	179
12	Rioja, La	174
13	Cantabria	102
14	Canarias	88
15	Asturias, Principado de	78
16	Baleares, Illes	67
17	Murcia, Región de	45
18	Ceuta	1
19	Melilla	1

```
--contar las islas por CA
set dateformat dmy
select ca as 'Comunidad Autónoma',
       count(distinct isla) as NIsas
from datoscompletostabla
where fecha='01-01-2013'
group by ca
order by CA
go
```

	Comunidad Autónoma	NIsas
1	Andalucía	0
2	Aragón	0
3	Asturias, Principado de	0
4	Baleares, Illes	4
5	Canarias	7
6	Cantabria	0
7	Castilla - La Mancha	0
8	Castilla y León	0
9	Cataluña	0
10	Ceuta	0
11	Comunitat Valenciana	0
12	Extremadura	0
13	Galicia	0
14	Madrid, Comunidad de	0
15	Melilla	0
16	Murcia, Región de	0
17	Navarra, Comunidad ...	0
18	País Vasco	0
19	Rioja, La	0


```

--Nº de habitantes y de parados por CA
--filtrando a datos de enero
set dateformat dmy
select ca as 'Comunidad Autónoma',
       SUM(padron) as NHabit,
       SUM(TotalParoRegistrado) as Nparo
from datoscompletostabla
where fecha='01-01-2013'
group by ca
order by nParo desc
go

```

	Comunidad Autónoma	NHabit	Nparo
1	Andalucía	8449985	1102398
2	Cataluña	7570908	661817
3	Comunitat Valenciana	5129266	586420
4	Madrid, Comunidad de	6498560	561919
5	Galicia	2781498	290790
6	Canarias	2118344	289517
7	Castilla - La Mancha	2121888	271610
8	Castilla y León	2546078	245877
9	País Vasco	2193093	175281
10	Murcia, Región de	1474449	158288
11	Extremadura	1110660	152100
12	Aragón	1349467	114705
13	Asturias, Principado de	1077360	107608
14	Balears, Illes	1119439	95509
15	Cantabria	593861	58650
16	Navarra, Comunidad ...	644566	55597
17	Rioja, La	323609	28613
18	Ceuta	84018	12197
19	Melilla	80802	11882

```
--nº de provincias y municipios por CA
--valor mayor y menor de habitantes en un municipio de la CA
set dateformat dmy
select ca as 'Comunidad Autónoma',
       count(distinct provincia) as NProv,
       count(distinct Municipio) as NMuni,
       Max(padron) as MasHabitantesEnMunicipio,
       Min(padron) as MenosHabitantesEnMunicipio
from datoscompletostabla
where fecha='01-01-2013'
group by ca
order by CA
go
```

	Comunidad Autónoma	NProv	NMuni	MasHabitantesPorMunicipio	MenosHabitantesPorMunicipio
1	Andalucía	8	771	702355	65
2	Aragón	3	731	679624	10
3	Asturias, Principado de	1	78	277733	166
4	Balears, Illes	1	67	407648	267
5	Canarias	2	88	382296	805
6	Cantabria	1	102	178465	71
7	Castilla - La Mancha	5	919	172472	2
8	Castilla y León	9	2248	311501	4
9	Cataluña	4	947	1620943	28
10	Ceuta	1	1	84018	84018
11	Comunitat Valenciana	3	542	797028	17
12	Extremadura	2	386	152270	68
13	Galicia	4	315	297355	213
14	Madrid, Comunidad de	1	179	3233527	53
15	Melilla	1	1	80802	80802
16	Murcia, Región de	1	45	441354	562
17	Navarra, Comunidad ...	1	272	197604	16
18	País Vasco	3	251	351629	121
19	Rioja, La	1	174	153402	9

```
--nº de Municipios por provincia
set dateformat dmy
select Provincia,
       count(distinct Municipio) as NMuni
from datoscompletostabla
where fecha='01-01-2013'
group by provincia
order by provincia
go
```

Resultados		Mensajes
	Provincia	NMuni
1	Albacete	87
2	Alicante/Alacant	141
3	Almería	102
4	Araba/Álava	51
5	Asturias	78
6	Ávila	248
7	Badajoz	165
8	Balears, Illes	67
9	Barcelona	311
10	Bizkaia	112
11	Burgos	371
12	Cáceres	221
13	Cádiz	44

Y más...

Para trabajar los conceptos probamos con otra Base de Datos

```
--probando con otra base de datos
--trabajaremos con la base de datos audienciasbasicas
--contendrá datos de las audiencias de un conjunto de
--programas
use audienciasbasicas
go
--Trabajamos con la tabla datosprogramas
--contiene la audiencia en nº y en share(%)
--de los programas de más audiencia
--especificando la fecha/hora de la emisión
SELECT Id,Programa,Cadena,Espectadores,Share,FechaHora
FROM Datosprogramas
GO
```

	Id	Programa	Cadena	Espectadores	Share	FechaHora
1	1	El Peliculón: Red	ANTENA 3	4196000	21,1	2013-05-19 22:07:00.000
2	2	Motociclismo: G.P. Francia	TELECINCO	3958000	35,3	2013-05-19 14:01:00.000
3	3	Post Motociclismo: G.P. Francia	TELECINCO	3600000	27,2	2013-05-19 14:50:00.000
4	4	Tenis Masters 1.000: Nadal-Federer	La1	2893000	19,6	2013-05-19 16:15:00.000
5	5	Salvados: Sin transparencia	LA SEXTA	2859000	14,5	2013-05-19 21:36:00.000
6	6	Multicine 2 / la trampa del amor	ANTENA 3	2718000	20,1	2013-05-19 17:41:00.000
7	7	Masters 1000	La1	2693000	18,3	2013-05-19 16:00:00.000

--Mostrar las cadenas y el nº de elementos que
--aparece de cada una

```
select cadena, COUNT(*) as NProgramas  
from Datosprogramas  
group by cadena  
order by Nprogramas desc
```

Resultados		Mensajes
	cadena	NProgramas
1	ANTENA 3	96
2	TELECINCO	85
3	La1	59
4	LA SEXTA	27
5	CUATRO	11
6	La2	1

```
--Media de audiencia por día de la semana
select datename(dw, fechahora) as ds,
       DATEPART(dw, fechahora) as ND,
       AVG(espectadores) as media
from Datosprogramas
group by datename(dw, fechahora), DATEPART(dw, fechahora)
order by nd
```

Resultados		Mensajes	
	ds	ND	media
1	Lunes	1	2037050
2	Martes	2	2073125
3	Miércoles	3	2012150
4	Jueves	4	2154475
5	Viernes	5	2221775
6	Sábado	6	1887600
7	Domingo	7	2386666,66666667

```
--Nº de veces de cada programa
select programa, COUNT(*) as NProgramas
from Datosprogramas
group by programa
order by Nprogramas desc
```

	programa	NProgramas
1	Informativos t5 15:00	12
2	Los simpsons	11
3	Amar es para siempre	10
4	Antena 3 noticias 1	10
5	El secreto de puente viejo	10
6	Pasapalabra	10
7	Salvame:diario	10
8	Telediario 1	10
9	Telediario 2	8
10	Antena 3 noticias 2	8
11	El Intermedio	8

Y más...

```
--Siete Programas con mayor Media de audiencia
select top 7 programa, avg(espectadores) as MProgramas
from Datosprogramas
group by programa
order by Mprogramas desc
```

<div>Resultados Mensajes</div>		
	programa	MProgramas
1	Promoga futbol:c.rey / r.madrid-at.madrid	11449000
2	Futbol:copa del rey / r.madrid-at.madrid	10062000
3	Eurovision:votaciones	6336000
4	Post Fórmula 1: G.P. España	6064000
5	Fórmula 1: G.P. España	5651000
6	Festival eurovision	5369000
7	Águila Roja	5062000

Ejercicios:

Para la BD parobasicas:

- Total de habitantes y parados por isla para datos del 1/3/2013.
- Mostrar los 3 municipios de Canarias con mayor media de parados.
- Cuántos municipios por provincia y CA tienen el paro por encima del 20% de la población para el dato de 1/1/2013.

Para la BD audienciasbasicas:

- Cuántos programas diferentes tiene cada cadena con algún share >20
- Media de espectadores que han visto programas punteros en la sexta los lunes, por hora de comienzo.

Para la BD turismobasicas:

- Contar cuántos datos hay por grupo de edad.
- Dar los 3 países con más turistas en el periodo de 2013.

Filtrar grupos: Having

Relacionado con el group by tenemos la posibilidad de realizar filtros de grupos.

Es decir, decidir qué grupos salen en la consulta y cuáles no salen.

La sintaxis es como la del where, pero afectando a campos agrupadores o a funciones agrupadoras.

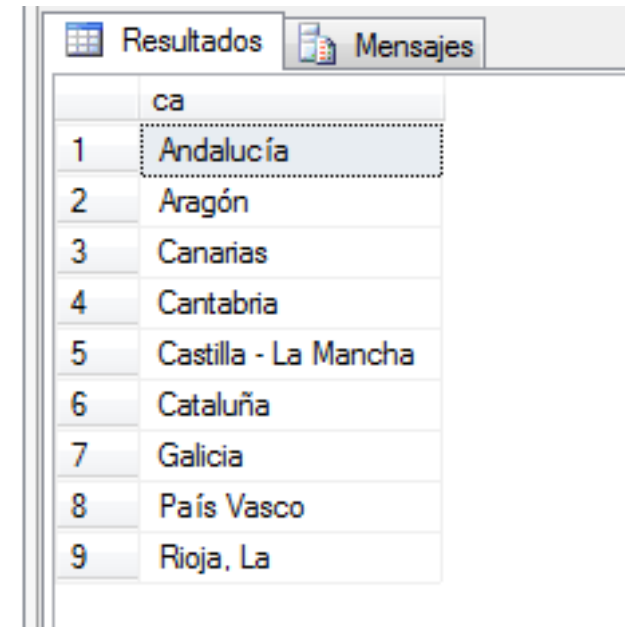
Es importante el orden de cada cláusula.

1. select
2. from
3. where
4. group by
5. having
6. order by

```
--Mostrar las comunidades autónomas que no  
--contengan una e
```

```
select ca  
from datoscompletostabla  
group by ca  
having ca not like '%e%'  
order by ca;  
go
```

```
--funcionaría igual con where  
select ca  
from datoscompletostabla  
where ca not like '%e%'  
group by ca  
order by ca;  
go
```



The screenshot shows a database application window with two tabs: 'Resultados' (Results) and 'Mensajes' (Messages). The 'Resultados' tab is active, displaying a table with two columns: an index and the name of the autonomous community ('ca'). The table contains 9 rows of data. The first row, '1 Andalucía', is highlighted with a dashed border.

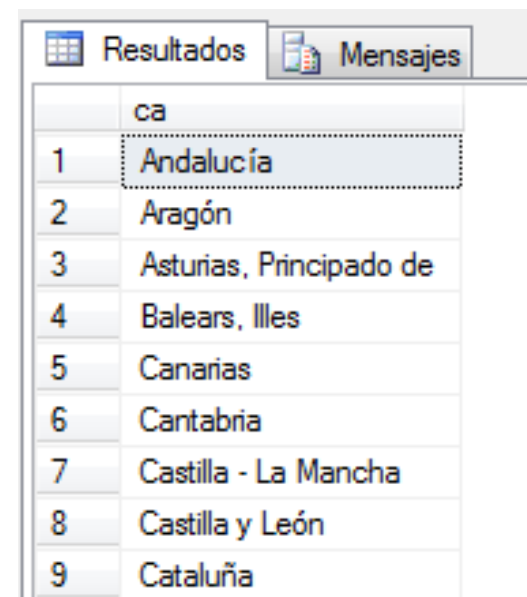
	ca
1	Andalucía
2	Aragón
3	Canarias
4	Cantabria
5	Castilla - La Mancha
6	Cataluña
7	Galicia
8	País Vasco
9	Rioja, La

```
--Daría error filtrar algo en el having
--que no sea agrupador ni función agrupadora
set dateformat dmy
select ca
from datoscompletostabla
group by ca
having fecha='01/01/2013'
order by ca;
go
```

Mens. 8121, Nivel 16, Estado 1, Línea 5

La columna 'datoscompletostabla.Fecha' de la cláusula HAVING no es válida, porque no está contenida en una función de agregado ni en la cláusula GROUP BY.

```
--Para filtrar de campos no agrupadores ni de
--funciones agrupadoras tendría que hacerse
--obligatoriamente en el where
set dateformat dmy
select ca
from datoscompletostabla
where fecha='01/01/2013'
group by ca
order by ca;
go
```



The screenshot shows a window titled 'Resultados' (Results) with a 'Mensajes' (Messages) tab. It displays a table with two columns: an index and the region name ('ca'). The regions are listed in ascending order: Andalucía, Aragón, Asturias, Principado de, Balears, Illes, Canarias, Cantabria, Castilla - La Mancha, Castilla y León, and Cataluña. The first row, '1 Andalucía', is highlighted with a dotted border.

	ca
1	Andalucía
2	Aragón
3	Asturias, Principado de
4	Balears, Illes
5	Canarias
6	Cantabria
7	Castilla - La Mancha
8	Castilla y León
9	Cataluña

Y más...

```
--Mostrar los días de la semana y las cadenas que tienen
--programas detallados ese día, indicando el nº de programas
--Sólo ANTENA 3, LA SEXTA y La2.
select datename(dw, fechahora) as diasem,
       cadena, COUNT(*) as Nprogramas,
       DATEPART(dw, fechahora) as ndia
FROM Datosprogramas
group by datename(dw, fechahora), cadena, DATEPART(dw, fechahora)
having cadena in ('ANTENA 3', 'LA SEXTA', 'La2')
order by ndia, cadena;
GO
```

	Resultados	Mensajes				
	diasem	cadena	Nprogramas	ndia		
1	Lunes	ANTENA 3	14	1		
2	Lunes	LA SEXTA	4	1		
3	Martes	ANTENA 3	14	2		
4	Martes	LA SEXTA	4	2		
5	Miércoles	ANTENA 3	13	3		
6	Miércoles	LA SEXTA	5	3		
7	Jueves	ANTENA 3	13	4		
8	Jueves	LA SEXTA	5	4		
9	Viernes	ANTENA 3	14	5		
10	Viernes	LA SEXTA	2	5		
11	Viernes	La2	1	5		
12	Sábado	ANTENA 3	14	6		

Y más...

El uso más importante del having es mediante las funciones agrupadoras.

Podremos excluir grupos basadas en los cálculos (cuenta, suma, media,...)

```
--Mostar las CA y provincias
--Para aquellas con nº de Municipios distintos sea
--menor que 70
select ca,Provincia,count(distinct municipio) as Nmunic
from datoscompletostabla
group by ca,Provincia
having COUNT(distinct municipio)<70
order by ca,Provincia;
go
```

Resultados		Mensajes	
	ca	Provincia	Nmunic
1	Andalucía	Cádiz	44
2	Balears, Illes	Balears, Illes	67
3	Canarias	Palmas, Las	34
4	Canarias	Santa Cruz de Tenerife	54
5	Ceuta	Ceuta	1
6	Galicia	Lugo	67
7	Galicia	Pontevedra	62
8	Melilla	Melilla	1
9	Murcia, Región de	Murcia	45
10	País Vasco	Araba/Álava	51


```
--Mostrar las CA cuya media de habitantes sea más
--de 20000
select ca,avg(padron) as MedPadron
from datoscompletotabla
group by ca
having avg(padron)>20000
order by ca;
go
```

Resultados		Mensajes
	ca	MedPadron
1	Canarias	24072
2	Ceuta	84018
3	Madrid, Comunidad de	36304
4	Melilla	80802
5	Murcia, Región de	32765

```
--Mostrar CA con más de 200000 de habitantes
--a fecha 1/2/2013
set dateformat dmy
select ca, sum(padron) as Nhabit
from datoscompletotabla
where fecha='01/02/2013'
group by ca
having sum(padron)>2000000
order by ca;
go
```

Resultados			Mensajes	
	ca	Nhabit		
1	Andalucía	8449985		
2	Canarias	2118344		
3	Castilla - La Mancha	2121888		
4	Castilla y León	2546078		
5	Cataluña	7570908		
6	Comunitat Valenciana	5129266		
7	Galicia	2781498		
8	Madrid, Comunidad de	6498560		
9	País Vasco	2193093		

```
--Día de la semana y cadena con más de 5 programas
--en la lista de los más vistos
select DATENAME(dw, fechahora), cadena,
       datepart(dw, fechahora), count(id) as nprog
FROM Datosprogramas
group by DATEPART(dw, fechahora),
         datename(dw, fechahora), cadena
having COUNT(ID) > 5;
GO
```

	(Sin nombre de columna)	cadena	(Sin nombre de columna)	nprog
5	Martes	La1	2	6
6	Martes	TELECINCO	2	13
7	Miércoles	ANTENA 3	3	13
8	Miércoles	La1	3	6
9	Miércoles	TELECINCO	3	11
10	Jueves	ANTENA 3	4	13
11	Jueves	La1	4	9
12	Jueves	TELECINCO	4	13
13	Viernes	ANTENA 3	5	14
14	Viernes	La1	5	10
15	Viernes	TELECINCO	5	13
16	Sábado	ANTENA 3	6	14
17	Sábado	La1	6	12
18	Sábado	TELECINCO	6	11
19	Domingo	ANTENA 3	7	14
20	Domingo	LA SEXTA	7	6

Y más...

```
--Cadena con media de share en sus programas de los
--lunes después de las 20:00 horas sea superior a 12
select cadena, avg(share) as medshare
from Datosprogramas
where (DATEPART(dw, fechahora)=1) and
      (DATEPART(hour, fechahora) >=20)
group by cadena
having avg(share) >12
order by medshare desc;
go
```

Resultados		Mensajes
	cadena	medshare
1	La1	17,82
2	TELECINCO	13,3875

Ejercicios:

Para la BD parobasicas:

- Sumar el padrón de las ca con más de 5 provincias para datos del 1/3/2013
- Sumar el paro de las provincias con 4 islas (usando having) para datos 1/1/2013.

Para la BD audienciasbasicas:

- Suma de audiencia de programas para cada cadena en martes y para las cadenas con tres o menos programas.
- Mostrar las cadenas con media de share mayor que 10 en el horario de las 10, 11 y 12 de la mañana.

Para la BD turismobasicas:

- Mostrar la media de turistas por pais para aquellos países con media mayor de 25000 en 2012
- Mostrar los dos países con mayor suma de turistas en 2013, que tengan todos sus datos >23000.