



**MÓDULO**

**ANÁLISIS CLUSTER**

## ***Ejercicio 1***

DuPont es una de las empresas líder mundial en desarrollo y producción de genética y biotecnología de avanzada para cultivos de maíz. Impulsó el cambio biotecnológico en el mercado argentino de semillas e introdujo las primeras variedades de híbridos de maíz tolerantes al glifosato e insectos. Persigue incrementar la productividad y el desarrollo sustentable de los sistemas de agricultura en todo el mundo.

DuPont llevó a cabo un experimento en 55 campos de la región pampeana para estudiar el comportamiento agronómico de 12 variedades experimentales de maíz. Con el fin de agrupar y caracterizar las variedades estudiadas, lo cual es de gran importancia para los investigadores decidir si descartar una variedad o seguirla cultivando (según si son más resistentes a las plagas, tienen mayor crecimiento, etc.).



Se realizó un diseño de bloques parcialmente aleatorizado con 12 parcelas de 20m de longitud y una distancia de siembra de 0,90 x 0,35 m. Cada parcela constituyó una variedad. Se muestrearon 30 plantas de cada bloque y se observaron los valores promedios de los siguientes descriptores agronómicos y morfológicos:

VARIEDAD: identificación de la variedad de maíz.

LM: longitud de la mazorca (cm), desde la inserción hasta la punta de la espiga.

DM: diámetro de la mazorca (cm), medida sobre el ancho de la parte central de la espiga.

NGH: número de granos por hilera.

DT: diámetro de la tusa (cm).

AG: ancho del grano (cm).

PESO: peso de 100 granos tomados al azar.

En el archivo MAIZ-CLUSTER.sas se encuentra la información disponible para realizar el análisis y dar respuesta a la empresa.



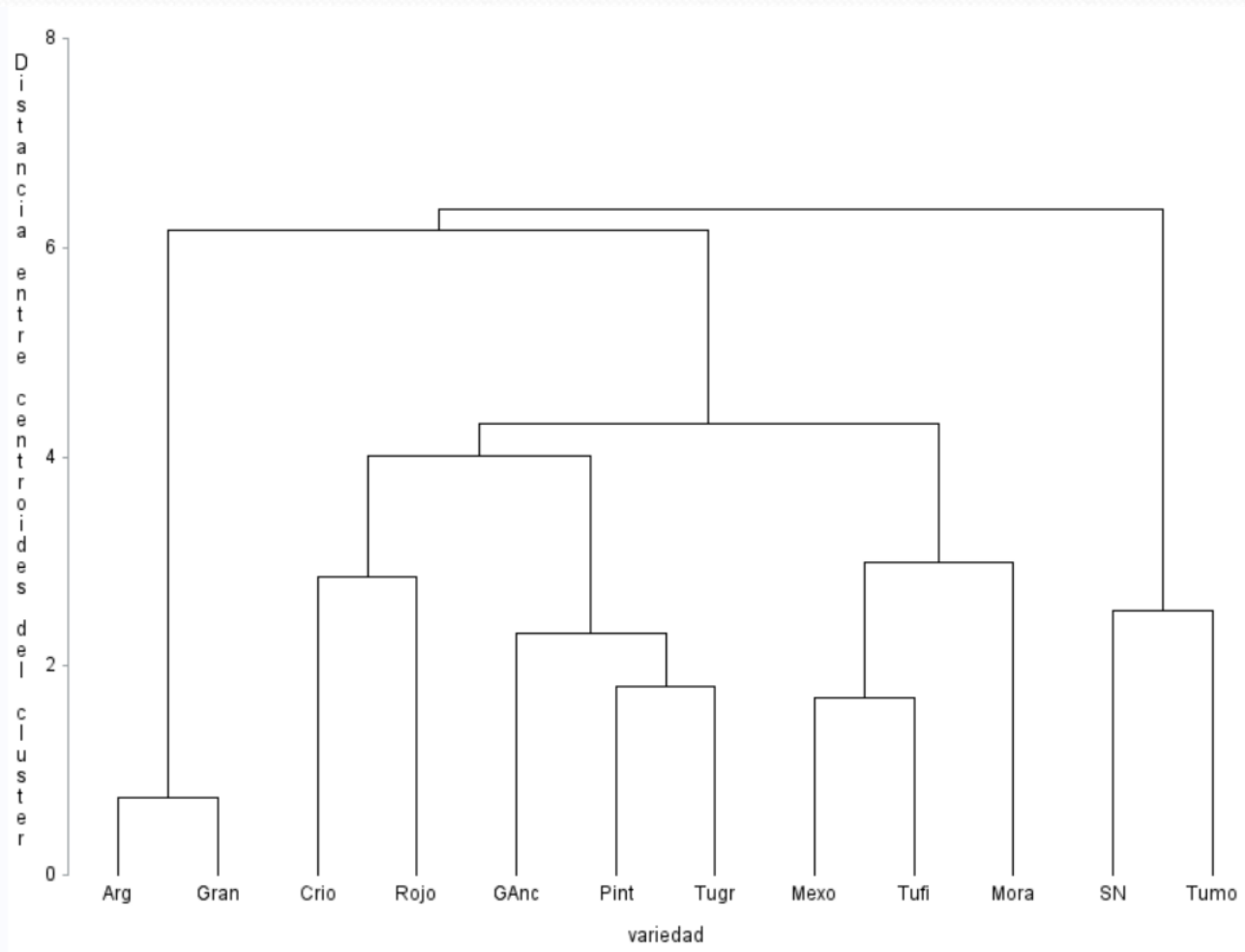
```
proc cluster data=MAIZ method=centroid noeigen simple
```

```
nonorm outtree=tree;
```

```
id variedad;
```

```
var LM DM NGH DT AG PESO;
```

```
run;
```





## ***Ejercicio 2***

El conocimiento sobre la estacionalidad de la vegetación natural en el país es muy escaso, especialmente en lo referente a las comunidades herbáceas distribuidas a lo largo de un gradiente de elevación en la reserva natural “El Rico” (Rosario, Santa Fe).

Los datos fueron recolectados en otoño (Marzo) y en primavera (Noviembre) del mismo año. Se extrajeron sobre una transecta perpendicular al río, donde las parcelas de 6,25 m<sup>2</sup> (2,5 x 2,5) se encuentran en línea recta cada 20 metros, abarcando 600 metros. Las 30 parcelas están enumeradas en forma correlativa a lo largo del gradiente. Todas son fijas y están marcadas, ya que el análisis se repite año tras año.

**Parcela**: superficie sobre la cual se toman los datos de vegetación.

**Transecta**: línea geográfica sobre la que se extrae la muestra.

**Abundancia**: número de individuos por especie dentro de la parcela.

**Cobertura** (de una especie): proporción de terreno ocupado por la proyección perpendicular de las partes aéreas de los individuos de la superficie considerada. Se mide en porcentaje.

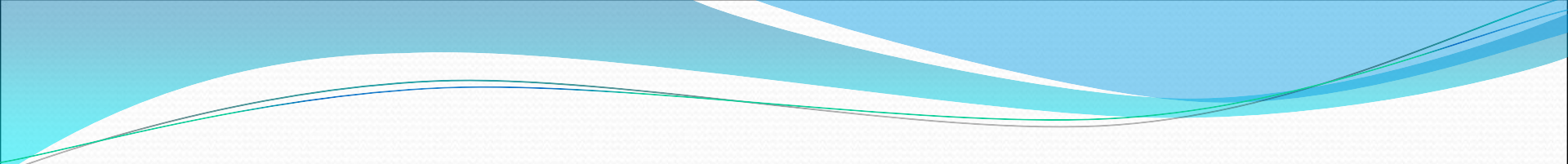
**Densidad (d)**: N° de individuos (n) en un área determinada (a).  $d = n/a$ .

**Especie dominante**: se considera a aquella que es la más notable en la comunidad. En este estudio es la que presenta mayor cobertura.

**Comunidad**: conjunto de individuos de distintas especies que interactúan entre sí y con el medio ambiente. En este estudio las comunidades están determinadas por la especie dominante. Por ejemplo: donde predomina el “canutillo”, es un “canutillar”.

**Ecotono**: zona de transición entre dos comunidades por invasión mutua (desde el punto de vista vegetal).





En cada parcela se realizó un relevamiento exhaustivo de la vegetación, anotando los nombres de todas las especies que aparecen en la misma, acompañadas de una medida de la abundancia y cobertura.

Para poder realizar el análisis cuantitativo se pasó a una escala de porcentajes de cobertura debida a Braun-Blanquet.

Las especies que aparecieron en 3 ó menos parcelas fueron eliminadas por considerarse que no influyen en la caracterización de las parcelas y que sólo complican el análisis (ruido).

La matriz usada en el análisis es de 60 individuos (30 parcelas en otoño y 30 en primavera) por 30 variables (especies).



Debido a la complejidad de los nombres de las especies, se utilizan abreviaturas para hacer referencias a las mismas, las cuales son:

<i>Abreviatura</i>	<i>Especies</i>	<i>Abreviatura</i>	<i>Especies</i>
Aph	<i>Alternanthera philoxeroides</i>	Gsp	<i>Gymnocoronis spilanthoides</i>
Bph	<i>Baccharis phyteumoides</i>	Hal	<i>Hemarthria altissima</i>
Bpi	<i>Baccharis pingraea</i>	Ham	<i>Hymenachne amplexicaulis</i>
Bla	<i>Bidens laevis</i>	Lhe	<i>Leersia hexandra</i>
Bve	<i>Borreria verticillata</i>	Lpe	<i>Ludwigia peploides</i>
Csp	<i>Cleome</i> sp.	Mpe	<i>Mikania periplocifolia</i>
Cbo	<i>Conyza bonariensis</i>	Osp	<i>Oxalis</i> sp.
Cvi	<i>Cyperus virens</i>	Pel	<i>Panicum elephantipes</i>
Ecr	<i>Echinochloa crusgalli</i>	Ppr	<i>Panicum prionitis</i>
Egr	<i>Echinodorus grandiflorus</i>	Pta	<i>Plagiocheilus tanacetoides</i>
Epo	<i>Echinochloa polystachya</i>	Pac	<i>Polygonum acuminatum</i>
Ecl	<i>Eleocharis elegans</i>	Pho	<i>Polygonum hóp</i>
Eno	<i>Eleocharis nodulosa</i>	Sam	<i>Solanum amygdalifolium</i>
Ean	<i>Enhydra anagallis</i>	Tge	<i>Thalia geniculata</i>
Eeb	<i>Eryngium ebracteatum</i>	Vin	<i>Vernonia incana</i>

# Objetivos

En este trabajo se pretende describir las distintas comunidades vegetales de la reserva natural El Rico. Específicamente, se desea:

- agrupar las parcelas de similar vegetación y
- describir las características de los grupos formados.



# Resultados

```
data espe;
input parce $ estacion $ parcela $ Aph Bph Bpi Bla Bve Csp Cbo Cvi Ecr Egr Epo Eel Eno
Ean Eeb Gsp Hal Ham Lhe Lpe Mpe Osp Pel Ppr Pta Pac Pho Sam Tge Vin;
cards;
1 o o1 0.1 5 0 0 0.1 0 0.1 0 0 0 0 0 0 0 0 0 0 0 5 5 37.5
0 0.1 0 0 0 0.1 0 0 0
2 o o2 0.1 17.5 17.5 0 0.1 0 0.1 0 0 0 0 0 0 0 5 0 0
0 0 0.1 17.5 0 0.1 17.5 0 0.1 0 0 0 0
.
.
.
60 + +30 0 0 0 0 0 0 0 0 0 0 0 0 5 0 0 0 0
0 37.5 0 0 0 0 0 17.5 0 0 0 0
;

proc cluster data=espe method=average noeigen nonorm outtree=dendogra ;
id parcela;
var Aph Bph Bpi ... Tge Vin;
run;
```

Se tomó la “distancia de la cuerda” como medida de la distancia entre dos individuos. Primero normaliza los vectores y luego calcula la distancia euclídea. Es decir la distancia medida entre la parcela 1 y la parcela 2 es:

$$d(\text{parcela I}, \text{parcela II}) = \sqrt{\left[ \sum_{i=1}^{30} \left( e_i^{(I)} - e_i^{(II)} \right)^2 \right]}$$

donde  $e_i^{(I)}$  es el coeficiente  $i$  (que corresponde a la especie  $i$ ) del vector normalizado de las coberturas de la parcela I.



Se utiliza esta distancia con el objetivo que las distancias entre parcelas no estén muy influenciadas por las diferencias de cobertura en una misma especie y se le da más peso a la presencia de especies diferentes.

Por ejemplo:

$$A = (15, 12, 0)$$

$$B = (0, 0, 5)$$

$$C = (3, 5, 0)$$

La distancia Euclidea:

$$d(A,B) = 19.85$$

$$d(A,C) = 13.89$$

$$d(B,C) = 5.83$$

Cálculo de la distancia de la cuerda, primero normalizar:

$$e_A = (0,781; 0,624; 0)$$

$$e_B = (0; 0; 1)$$

$$e_C = (0.514; 0.857; 0)$$

Luego se calcula la distancia euclídea:

$$d(A,B) = 1.41$$

$$d(A,C) = 0.100$$

$$d(B,C) = 1.143$$

```

data espe;
input parce $ estacion $ parcela $ Aph Bph Bpi Bla Bve Csp Cbo Cvi Ecr Egr Epo Eel Eno
Ean Eeb Gsp Hal Ham Lhe Lpe Mpe Osp Pel Ppr Pta Pac Pho Sam Tge Vin;
cards;
1 o o1 0.1 5 0 0 0.1 0 0.1 0 0 0 0 0 0 0 0 0 0 0 5 5 37.5
0 0.1 0 0 0 0.1 0 0 0
2 o o2 0.1 17.5 17.5 0 0.1 0 0.1 0 0 0 0 0 0 0 5 0 0
0 0 0.1 17.5 0 0.1 17.5 0 0.1 0 0 0 0
.
.
.
60 + +30 0 0 0 0 0 0 0 0 0 0 0 0 0 5 0 0 0 0
0 37.5 0 0 0 0 0 0 17.5 0 0 0 0
.
/* -Cálculo de las distancias normalizadas (distancia de la cuerda) - */
proc iml; use espe;
read all var {Aph Bph Bpi Bla Bve Csp Cbo Cvi Ecr Egr Epo Eel Eno Ean Eeb Gsp Hal Ham
Lhe Lpe Mpe Osp Pel Ppr Pta Pac Pho Sam Tge Vin}
into matrix;
sc=matrix*matrix`; t=vecdiag(sc);
d=diag(sc); i=inv(d);
dist= sqrt(i);
dnorm= matrix`*dist;
uno=vecdiag(dnorm`*dnorm);
cuerda=dnorm`;
create newdata from cuerda;
append from cuerda;
close newdata;

data distan;
merge newdata espe ;
keep parce estacion parcela Aph Bph Bpi Bla Bve Csp Cbo Cvi Ecr Egr Epo Eel Eno Ean
Eeb Gsp Hal Ham Lhe Lpe Mpe Osp Pel Ppr Pta Pac Pho Sam Tge Vin coll1 coll2
coll3 coll4 coll5 coll6 coll7 coll8 coll9 coll10 coll11
coll12 coll13 coll14 coll15 coll16 coll17 coll18 coll19
coll20 coll21 coll22 coll23 coll24 coll25 coll26 coll27 coll28
coll29 coll30;

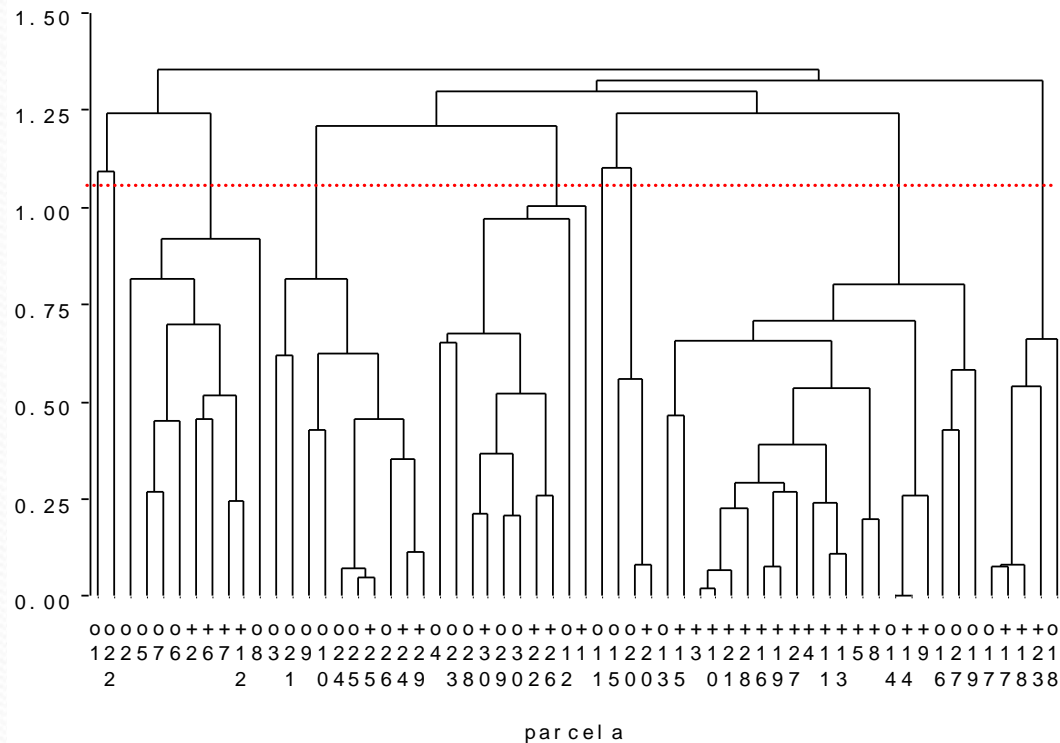
proc cluster data=distan method=average noeigen nonorm outtree=dendo;
id parcela;
copy Aph Bph Bpi Bla Bve Csp Cbo Cvi Ecr Egr Epo Eel Eno Ean Eeb Gsp Hal Ham Lhe Lpe
Mpe Osp Pel Ppr Pta Pac Pho Sam Tge Vin;
var coll1 coll2 coll3 coll4 coll5 coll6 coll7 coll8 coll9 coll10 coll11 coll12 coll13 coll14 coll15
coll16 coll17 coll18 coll19 coll20 coll21 coll22 coll23 coll24 coll25 coll26 coll27 coll28 coll29
coll30; run;

```



El orden de parecidos en el que fueron formados los clusters se puede observar en el dendograma:

**Gráfico 1:** Secuencia de las fusiones de las parcelas según el orden de parecidos (Dendograma).



La línea de puntos indica donde se decidió detener el proceso. Se obtuvieron 9 clusters bien diferenciados, ya que si se formaban 8, se unen grupos de parcelas caracterizados por distintas especies.

Los 9 grupos representan distintas especies dominantes, estos son:

**Grupo 1: Cataisal** (*Polygonum acuminatum*). Formado por las parcelas 3, 4, 5, 8, 9, 10, 11, 13, 14, 15, 16, 19, 21, 27 y 28 de primavera y las parcelas 13, 14, 16, 19 y 27 de otoño. Es interesante resaltar que hay una mayor presencia de parcelas que pertenecen a la estación de primavera.

**Grupo 2: Canutillar** (*Panicum elephantipes*). Constituido por las parcelas 3, 9, 10, 21, 24, 25 y 26 de otoño y las 24, 25 y 29 de primavera.



**Grupo 3: Canutillar** (*Echinochloa polystachya*). Forman otro canutillar la parcela 23 de primavera y las parcelas 17 y 18 en las dos estaciones.

**Grupo 4: Achiral** (*Thalia geniculata*). Integrados por las parcelas 15 y 20 de otoño y la 20 de primavera.

**Grupo 5: Verdolagal** (*Ludwigia peploides*). Lo constituyen las parcelas 4, 12, 23, 28, 29 y 30 de otoño y las 1, 22, 26 y 30 de primavera.

**Grupo 6: Pajonal** (*Panicum prionitis*). En él se encuentra las parcelas 2, 6, 7 y 12 de primavera y las 2, 5, 6, 7 y 8 de otoño. En estas cinco últimas parcelas también sobresale una considerable cobertura de *Mikania periplocifolia*.

Los **Grupos 7, 8 y 9** están formados sólo por una parcela y no se asemejan a ninguna de las otras comunidades mencionadas. Ellos son la parcela 1 de otoño donde la especie dominante es *Mikania periplocifolia* (**Cataisal**), la parcela 22 de otoño donde predomina *Hymenachne amplexicaulis* (**Canutillar**) y en la parcela 11 de otoño no hay una clara dominancia de una única especie (**Ecotonal**).

## Conclusiones

- Los grupos obtenidos en el Análisis Cluster, corresponden a distintas comunidades herbáceas, en donde se observa que las comunidades no varían su composición florística en otoño y primavera.
- Generalmente en parcelas próximas, se encuentran las mismas comunidades. Esto estaría evidenciando que las comunidades se ordenan de acuerdo a un gradiente ambiental (elevamiento topográfico).



## ***Ejercicio 3***

Se dispone de una base de datos que contiene información de 303.586 titulares de tarjetas de crédito de un banco nacional. La información suministrada contiene variables que indican comportamientos de consumos y de pagos correspondientes a los meses de Abril, Mayo y Junio del 2009.

Se plantea la necesidad de asesorar en la toma de decisión sobre la compra de esta cartera de clientes de tarjetas de crédito de una entidad bancaria que es ofrecida a otras entidades financieras. Para tal fin, se intentará identificar grupos o conglomerados de clientes que permitan medir la potencialidad de la cartera y ayuden a decidir sobre la compra de la misma; analizar las diferencias entre los grupos definidos y derivar recomendaciones acerca de posibles acciones comerciales a seguir con los clientes de cada grupo definido, en caso de decidirse la compra de la cartera.

Se evaluaron variables socio-demográficas de cada cliente; variables que caracterizan las tarjetas de crédito y variables que indican comportamiento de consumo y de pago (Abril, Mayo y Junio de 2009). Algunas de ellas son:

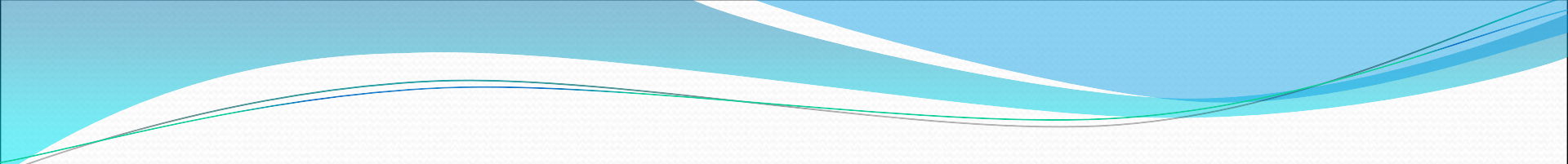
- Número de identificación único de cada cliente.
- Género del titular de la tarjeta de crédito.
- Región donde está radicado el titular de la tarjeta de crédito, según las siguientes 3 regiones: Capital Federal, Gran Buenos Aires e Interior.
- Marca de tarjeta de crédito que compró el cliente (Master, Visa, Makro, Amex).
- Categoría de edad del titular de la tarjeta (menores de 25 años; entre 25-35 años; entre 36-45 años; entre 46-55 años; mayores a 56 años).



- Canal de venta por el cual se emitió la tarjeta de crédito (Sucursales del banco emisor, Carrefour, Comercializadoras externas, Comercializadora Makro, Venta directa, Call-center internos y Otros).
- Forma de **utilización** de la tarjeta de crédito, según el porcentaje de pago usual del consumo, puede ser: Transactor (pago superior al 85% del saldo), Revolving (efectiviza menos del 85% del saldo) o Inactivo (no registra consumos en los últimos 3 meses).
- Cantidad de tarjetas adicionales que posee la tarjeta de crédito.
- **Límite de compra máximo** en efectivo que fue otorgado a ese cliente.
- Monto total acumulado en pesos que el cliente le adeuda a la entidad financiera por haber realizado compras en distintas cantidades de cuotas. Es decir, es el total de **saldo deudor** que tiene acumulado el cliente.

- Monto que **pagó** el cliente en los últimos tres meses (Abril, Mayo y Junio).
- Monto en pesos del saldo de la última liquidación.
- Saldo en pesos de la tarjeta de crédito correspondiente a los últimos tres meses (Abril, Mayo y Junio).
- **Compras**, total consumido en los últimos tres meses (Abril, Mayo y Junio).
- **Pago mínimo** que podrá realizar el cliente de acuerdo a sus últimos consumos en los últimos tres meses (Abril, Mayo y Junio).
- **Antigüedad**, tiempo transcurrido desde el alta de la tarjeta de crédito hasta el día de la fecha.
- Porcentaje de límite que utilizó el cliente en Junio (ídem Abril y Mayo). Esta variable se calcula realizando el cociente entre las compras de Junio y el límite otorgado de la tarjeta de crédito.
- Etc.
- Etc.





En primer lugar se selecciona el subconjunto de tarjetas de crédito que brindan información comercial sobre el comportamiento de pagos y de consumos; decidiéndose trabajar solamente con las tarjetas “Activas” con más de 3 meses de antigüedad en el banco (159.452 ).

Se realizó una selección de variables. El conjunto resultante está compuesto por cinco mediciones sobre uso y modalidad de pago: Compras, Pago mínimo, Pago, Utilización y Saldo, promediadas a través de los tres últimos meses de uso, Abril, Mayo y Junio de 2009. Además, dos características de las tarjetas de crédito: Límite otorgado y Antigüedad.

Todas las variables son cuantitativas continuas, por lo cual se eligió como medida de parecido entre los individuos (tarjetas de crédito) la distancia euclídea. Previamente, dado que estas variables están medidas en distintas escalas y/o tienen rangos de variación muy diferentes se realizó un proceso de estandarización utilizando los criterios del: Rango, Desviación estándar y Máximo absoluto.

Se aplica el algoritmo de agrupación no jerárquico de k-medias seleccionando al azar los K centros iniciales para distintas cantidades de conglomerados que van de 3 a 9, combinándolos con cada uno de los procesos de estandarización planteados.

Para la elección de la mejor combinación, número de conglomerados - criterio de estandarización, se usaron los estadísticos F,  $R^2$  y CCC (criterio de agrupación cúbico). Los valores deseables para ellos son los más altos.



N° de conglomerados	Criterios de estandarización											
	Ninguno			Rango			Desviación estándar			Máximo absoluto		
	<i>F</i>	R <sup>2</sup>	CCC	<i>F</i>	R <sup>2</sup>	CCC	<i>F</i>	R <sup>2</sup>	CCC	<i>F</i>	R <sup>2</sup>	CCC
9	132525.8	0.88	-34.4	41918.7	0.5868	297.0	37357.4	0.47	796.7	39421.5	0.58	267.5
8	147162.5	0.87	-14.5	41472.8	0.5717	232.9	38681.3	0.45	777.9	43461.4	0.57	261.1
7	151264.7	0.86	-27.9	44633.6	0.5537	204.9	39596.8	0.41	778.4	44064.6	0.55	216.1
6	150671.8	0.84	-56.4	48937.5	0.5314	207.1	<b>43817.0</b>	<b>0.37</b>	<b>862.7</b>	48378.8	0.52	221.7
5	172588.6	0.82	-31.7	49510.4	0.5026	140.2	37812.6	0.32	644.6	49411.0	0.49	163.9
4	188612.6	0.79	-31.8	54431.6	0.4631	102.3	42211.1	0.26	704.5	50580.5	0.45	94.2
3	174799.8	0.74	-110	62531.8	0.4029	75.1	45695.0	0.19	701.6	59802.6	0.39	80.9

Algoritmo de agrupación de  $k$ -medias para 6 conglomerados con las variables estandarizadas.

```
proc fastclus data=cluster_std maxc=6 maxiter=50 drift  
replace=full out=cluster_std_6;  
var compras Pago_Mínimo Pago Util Saldo limite_otorgado  
antigüedad;  
run;
```



**Tabla:** Promedios de las variables para cada *cluster*.

N° de Conglome-rado	N° de tarjetas	Com-pras	Pago Mínimo	Pago	Utiliza-ción en %	Saldo	Límite Otorga-do	Antigüe-dad
1	83337	228.9	81.9	277.6	9.2	505.9	2995.5	2.4
2	24826	326.0	450.5	546.2	8.2	2511.0	4096.6	2.6
3	32852	799.7	111.3	767.0	33.0	1036.9	2726.5	2.9
4	4326	725.1	1258.0	1299.0	9.1	6054.4	8518.6	3.7
5	3276	3525.3	356.0	3628.6	31.2	4226.6	13573.6	7.2
6	10835	812.3	110.5	884.7	8.8	1126.5	11492.8	7.7

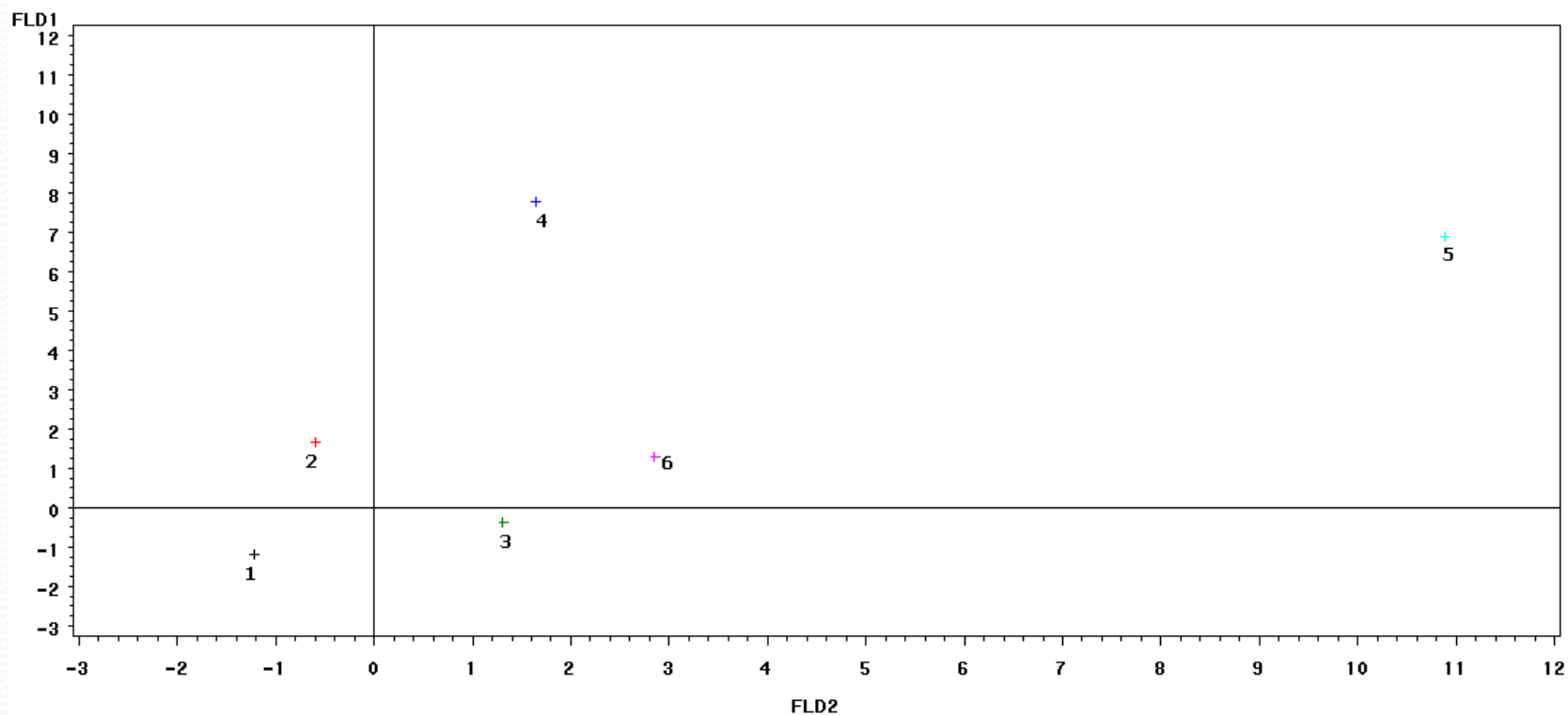
Pasos del Test Secuencial	Hipótesis nula	Correlación canónica	Correlación canónica al cuadrado	Valor de F	p-asociada
1	Ninguna FLD es significativa	0.871	0.759466	29262.1	<.0001
2	Sólo es significativa la 1° FLD	0.843	0.710849	23935.4	<.0001
3	Sólo son significativas la 1° y la 2° FLD	0.742	0.550015	16538.6	<.0001
4	Sólo son significativas la 1°, la 2° y la 3° FLD	0.587	0.345043	9695.9	<.0001
5	Sólo son significativas la 1°, la 2°, la 3° y la 4° FLD	0.110	0.012187	655.7	<.0001

**Tabla:** Test secuencial de significación de las funciones lineales discriminantes.



**Tabla: Estructura de las tres primeras funciones lineales discriminantes**

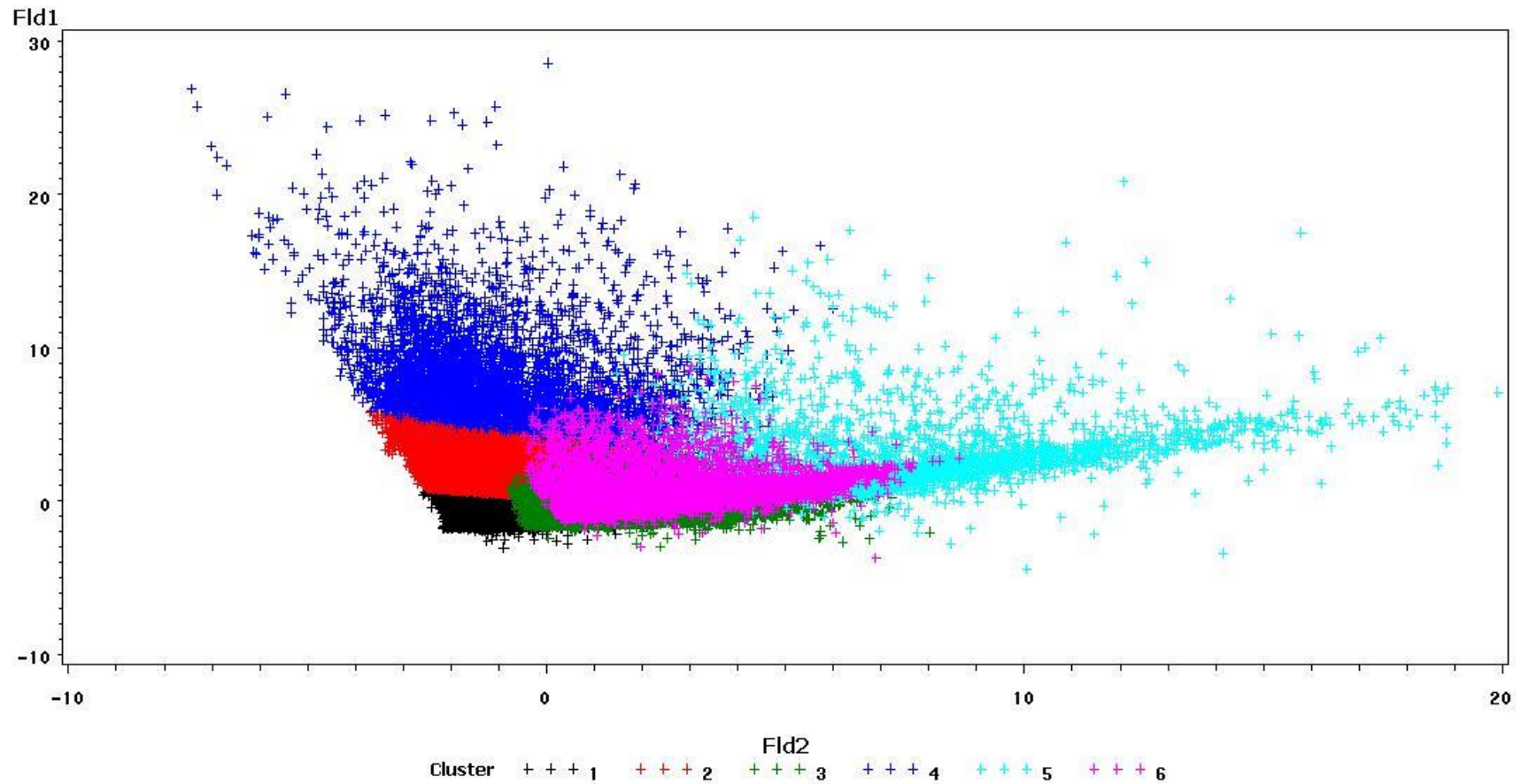
<b>Variables</b>	<b>FLD1</b>	<b>FLD2</b>	<b>FLD3</b>
<b>Compras</b>	-0.266	0.440(*)	-0.204
<b>Pago Mínimo</b>	0.580(*)	-0.147	0.072
<b>Pago</b>	0.111	0.248	0.112
<b>Utilización</b>	0.0.10	0.430(*)	0.770(*)
<b>Saldo</b>	0.669(*)	-0.219	0.200
<b>Límite otorgado</b>	0.195	0.345	-0.273
<b>Antigüedad</b>	0.082	0.414(*)	-0.487(*)



**Gráfico:** Promedios de los *clusters* sobre los dos primeros ejes discriminantes.



## Gráfico: Tarjetas de crédito proyectadas sobre los dos primeros ejes discriminantes.



**Tabla: Participación porcentual de las marcas de tarjetas de crédito en los conglomerados.**

Marca	Conglomerado						Total
	1	2	3	4	5	6	
<b>Visa</b>	50.3%	58.4%	54.6%	51.9%	48.0%	45.1%	52.1%
<b>Master</b>	25.3%	32.8%	29.2%	41.7%	49.4%	54.8%	30.2%
<b>Makro</b>	18.0%	3.1%	13.3%	1.1%	1.2%	0.0%	12.7%
<b>Amex</b>	6.4%	5.7%	3.0%	5.3%	1.3%	0.1%	5.0%
<b>Total</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
<b>Estadístico Chi-cuadrado=10923.1</b>			<b>G.L.= 15</b>		<b>p-asociada&lt;.0001</b>		

**Tabla:** Participación porcentual de las marcas de tarjetas de crédito en los conglomerados.



**Tabla: Participación porcentual de la cantidad de tarjetas adicionales en cada conglomerado**

Tarjetas de crédito adicionales	Conglomerado						Total
	1	2	3	4	5	6	
<b>0</b>	51.7%	42.9%	39.3%	36.4%	14.9%	24.1%	44.7%
<b>1</b>	38.4%	43.2%	43.4%	42.6%	39.6%	42.8%	40.6%
<b>2</b>	6.2%	8.4%	10.6%	11.0%	19.0%	16.4%	8.5%
<b>3</b>	2.7%	4.0%	4.6%	6.6%	13.6%	10.0%	4.1%
<b>4 y más</b>	1.0%	1.5%	2.1%	3.3%	12.9%	6.7%	2.0%
<b>Total</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
<b>Estadístico Chi-cuadrado=10764.3                      G.L.= 20                      p-asociada&lt;.0001</b>							

**Tabla:** Participación porcentual de la cantidad de tarjetas adicionales en cada conglomerado.

**Tabla: Participación porcentual de los canales de venta de tarjetas de crédito en los conglomerados**

Canal de venta	Conglomerado						Total
	1	2	3	4	5	6	
<b>Sucursales del banco emisor</b>	29.3%	23.5%	38.2%	39.7%	83.2%	92.8%	35.9%
<b>Carrefour</b>	31.0%	35.4%	28.2%	21.4%	3.6%	1.4%	28.3%
<b>Comercializadoras externas</b>	18.1%	23.7%	18.5%	20.9%	5.5%	3.3%	17.9%
<b>Comercializadora Makro</b>	13.1%	6.6%	7.3%	7.0%	1.6%	0.2%	9.6%
<b>Venta directa</b>	3.9%	5.9%	4.4%	5.4%	2.9%	1.7%	4.2%
<b>Call-center interno</b>	4.6%	4.7%	3.3%	5.5%	2.8%	0.6%	4.1%
<b>Otros</b>	0.1%	0.1%	0.1%	0.2%	0.3%	0.0%	0.1%
<b>Total</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
<b>Estadístico Chi-cuadrado=23853.57                      G.L.= 30                      p-asociada&lt;.0001</b>							



**Tabla: Distribución porcentual de las variables socio-demográficas de las tarjetas de**

		demograficas de las tarjetas de							
		Conglomerados							
V a r i a b l e s	R e g i ó n		1	2	3	4	5	6	Total
		Capital Federal	52.7%	52.2%	46.6%	58.9%	62.5%	58.3%	52.6%
		GBA	34.5%	35.2%	34.8%	35.0%	32.6%	34.7%	34.7%
		Interior	12.8%	12.6%	18.6%	8.4%	4.9%	7.9%	12.7%
		Estadístico Chi-cuadrado=1723.2                      G.L.= 10                      p-asociada<.0001							
	E d a d		1	2	3	4	5	6	Total
		Menor a 25 años	3.7%	5.2%	1.2%	0.9%	0.3%	0.3%	3.0%
		Entre 25-35 años	23.6%	28.9%	21.2%	20.4%	14.4%	16.4%	23.1%
		Entre 35-45 años	28.7%	31.0%	30.8%	31.2%	36.1%	36.8%	30.3%
		Entre 45-55 años	26.1%	26.4%	26.4%	27.0%	33.0%	31.4%	26.7%
		Mayor a 55 años	18.0%	8.5%	20.3%	20.5%	16.0%	15.1%	16.8%
		Estadístico Chi-cuadrado=3832.7                      G.L.= 20                      p-asociada<.0001							
	G é n e r o		1	2	3	4	5	6	Total
		Femenino	60.0%	59.5%	60.1%	42.5%	39.0%	41.6%	57.8%
		Masculino	40.0%	40.5%	39.9%	57.5%	61.0%	58.4%	42.2%
		Estadístico Chi-cuadrado=2328.6                      G.L.= 5                      p-asociada<.0001							

**Conglomerado 1:** es el de mayor tamaño en cuanto a cantidad de tarjetas de crédito; está conformado por las tarjetas con menor antigüedad y no tienen una utilización muy frecuente. Además, tienen límites de consumos muy bajos y compras de pequeños montos. Los pagos mensuales son pequeños y, en general, no utilizan el pago mínimo como forma de pago. Es el conglomerado que posee mayor proporción de tarjetas de crédito de las marcas Makro y Amex y está integrado por las tarjetas titulares con menor cantidad de tarjetas adicionales. Estos titulares en su mayoría son mujeres y hay una gran proporción de titulares menores de 35 años. ***Estos perfiles son atractivos para la entidad, aunque tienen la desventaja de tener un bajo nivel de utilización.***

**Conglomerado 2:** realizan pagos mínimos promedios, al igual que su utilización y antigüedad. Tienen una antigüedad promedio de 2,6 años y en promedio tienen un elevado monto de límite otorgado para la realización de compras. Son las tarjetas que menos porcentaje de su límite utilizan y posee mayor proporción de tarjetas Visa. Al igual que el conglomerado 1, posee muy pocas tarjetas con adicionales. adicionales. Tiene la mayor proporción de tarjetas comercializadas por Carrefour y comercializadoras externas; mayor proporción de clientes menores de 35 años y posee una proporción pequeña de clientes mayores de 55 años. ***Las tarjetas de crédito que lo componen no muestran un buen comportamiento en los pagos, por lo tanto son riesgosas para el banco.***



**Conglomerado 3:** compuesto por las tarjetas que poseen la mayor utilización promedio. Estas tarjetas no efectúan pago mínimo sino que pagan su saldo a medida que consumen. En promedio tienen 2,9 años de antigüedad y es el segundo conglomerado en cuanto a cantidad de tarjetas de crédito. La mayor proporción de tarjetas de este conglomerado son Visa. Hay proporción grande de tarjetas titulares que poseen una o ninguna tarjeta de crédito adicional. La mayoría fueron comercializadas por las Sucursales del banco emisor, Carrefour y Comercializadoras externas. Tiene el mayor porcentaje de clientes del Interior y el menor de Capital Federal. La mayoría de los clientes son mujeres. ***Sus titulares registran una buena conducta de pagos.***

**Conglomerado 4:** las tarjetas que lo integran son las que acumulan montos más elevados de saldo deudor debido a que efectúan sus pagos a través del pago mínimo. Tienen un límite otorgado promedio, pero utilizan un porcentaje bajo de éste para realizar sus compras. La mayoría son Visa y Master; tienen mayor cantidad de tarjetas adicionales que los conglomerados anteriores; hay mayor proporción de clientes hombres y existe menor proporción de clientes de Interior. *Las tarjetas que lo componen no presentan un buen comportamiento al momento de cancelar su saldo. Al igual que las tarjetas del conglomerado 2 tienen un alto nivel de riesgo.*



**Conglomerado 5:** lo integran las tarjetas que realizan compras de mayores montos, por lo tanto son las que efectúan los mayores pagos. Poseen un límite disponible alto y utilizan gran parte de éste. Es el conglomerado más pequeño en cuanto a cantidad de tarjetas de crédito, tienen mucha antigüedad y casi la totalidad son Visa y Master. Las tarjetas titulares poseen mayor proporción de adicionales. A diferencia de los 4 anteriores, la mayor proporción de las tarjetas fueron emitidas por las sucursales del banco emisor. La mayoría de los clientes de este conglomerado son mayores de 35 años, masculinos y de Capital Federal. *Estos perfiles serían los más beneficiosos para la entidad financiera.*



**Conglomerado 6:** lo conforman las tarjetas de crédito más antiguas; con un límite otorgado muy alto, pero no utilizan mucho esta disponibilidad de crédito. Cancelan su saldo a medida que éste vence, por lo tanto no utilizan el pago mínimo como forma de financiación. Casi todas son de las marcas Visa y Master; poseen tarjetas adicionales; fueron comercializadas por las sucursales del banco, sus titulares tienen entre 25 y 55 años y más de la mitad son hombres. Casi todos los clientes son de Capital Federal y Gran Buenos Aires. ***Al igual que en el conglomerado 5, estos perfiles serían los más beneficiosos para la entidad financiera.***

- En los grupos de tarjetas de crédito con baja tasa de utilización (Conglomerado 1, 2, 4 y 6) sería recomendable realizar acciones y promociones que incentiven el consumo de las mismas.
- Para las que tienen una gran utilización y su límite de compra no es elevado (Conglomerado 3) se podría ofrecer una actualización de ese límite.
- Las que se encuentran momentáneamente inactivas se podrían iniciar acciones con el fin de incitar a los clientes titulares a que activen la utilización de las mismas.