

ML No Supervisado - Obligatorio

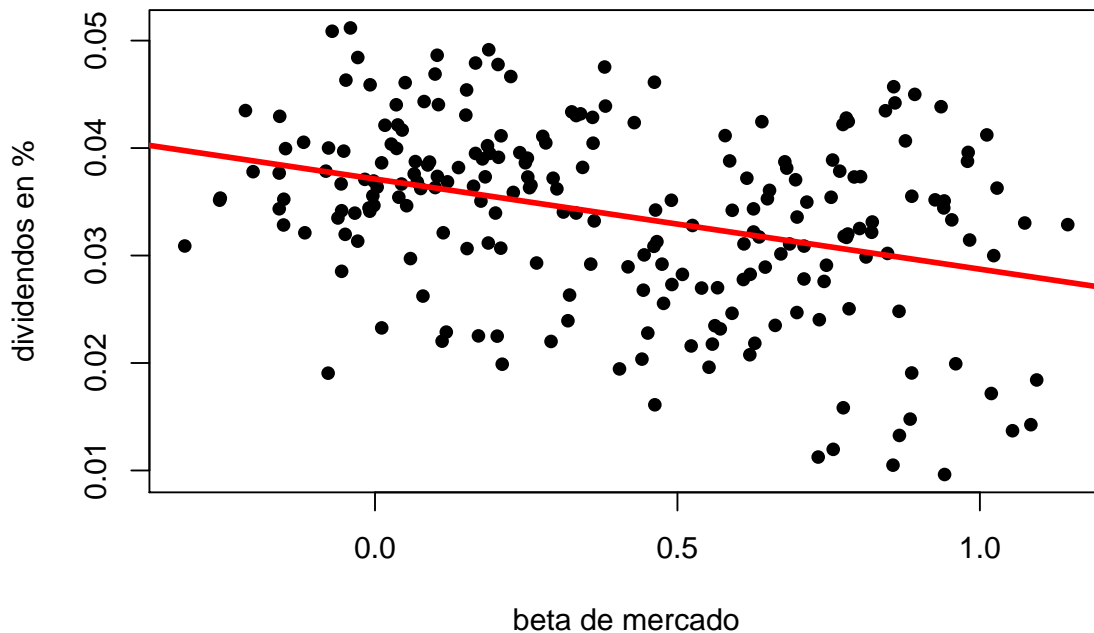
del Palacio - González

12/04/2019

Contents

1. Conocimiento Específico	1
2. Datos	2
3. Selección de Variables	3
4. Clustering inicial	4
5. Rating Crediticio	6
6. Clustering Avanzado	7
7. Comparación de Algoritmos de Clustering	8
8. Visualización	9
9. Asignación	10
10. Conclusiones	11

1. Conocimiento Específico



Gráficamente se puede confrimar el conocimiento específico adquirido. Para las acciones que presentan una menor volatilidad que el mercado (menor *beta*) se espera un pago de *dividendos* más alto. Para acciones que presentan mayor nivel de volatilidad, se observa una caída en los *dividendos* pagados, dado que su rentabilidad viene dada por la suba *esperada* del precio de la misma.

2. Datos

Descripción de los datos y variables continuas:

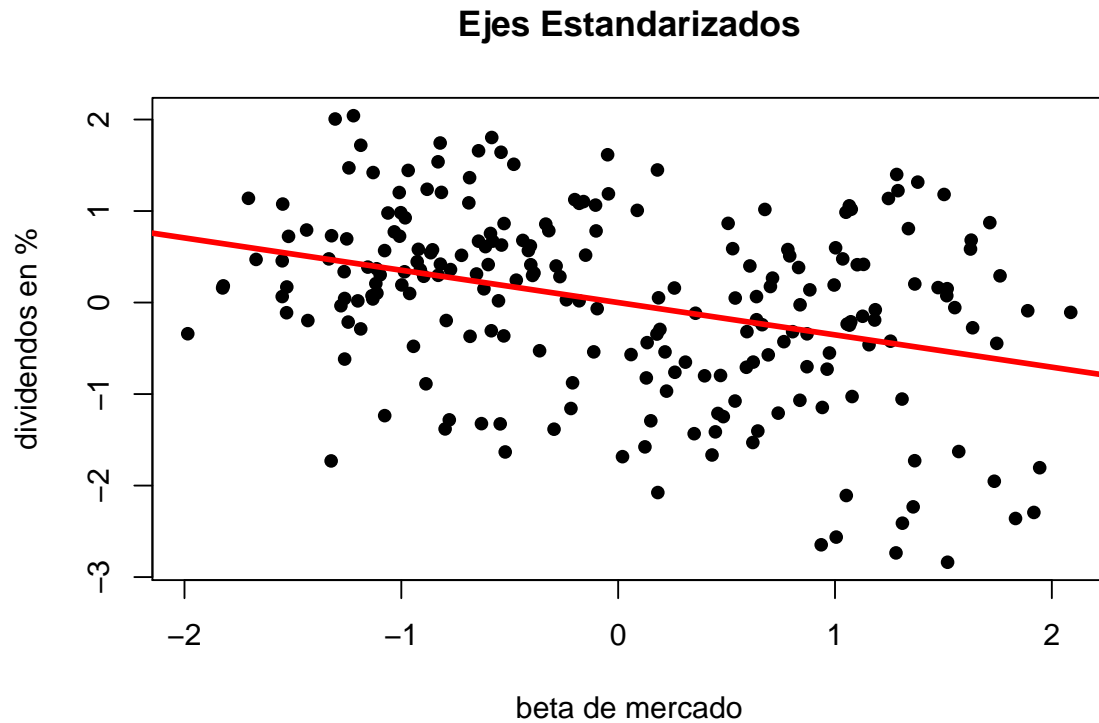
```
## Observations: 222
## Variables: 7
## $ X      <int> 7, 18, 22, 33, 34, 36, 37, 38, 40, 43, 62, 64, 65, 69,...
## $ country <fct> Canada, Canada, China, Netherlands, Netherlands, Nethe...
## $ sector  <fct> Energy, Basic Industries, Basic Industries, Finance, F...
## $ beta    <dbl> 0.4286, 0.1881, 0.0586, 0.2020, 0.3325, 0.4413, 0.1710...
## $ div     <dbl> 0.042, 0.049, 0.030, 0.022, 0.034, 0.020, 0.023, 0.022...
## $ vol     <dbl> 6.9e+07, 5.9e+07, 3.4e+07, 1.1e+06, 7.6e+06, 9.6e+06, ...
## $ cap     <dbl> 4.8e+09, 3.3e+09, 7.9e+09, 1.1e+09, 7.9e+09, 8.9e+09, ...
```

	country	sector	beta	div	vol	cap
	Canada :60	Finance :48	Min. :-0.31	Min. :0.010	Min. :1.13e+06	Min. :1.02e+09
	China :39	Basic Industries :28	1st Qu.: 0.09	1st Qu.:0.029	1st Qu.:2.03e+07	1st Qu.:3.25e+09
	Brazil :29	Consumer Services:26	Median : 0.36	Median :0.035	Median :4.60e+07	Median :5.40e+09
	uk :27	Public Utilities :26	Mean : 0.40	Mean :0.034	Mean :4.83e+07	Mean :5.74e+09
	Netherlands:17	Energy :23	3rd Qu.: 0.71	3rd Qu.:0.040	3rd Qu.:7.28e+07	3rd Qu.:8.42e+09
	Bermuda :16	Technology :15	Max. : 1.15	Max. :0.051	Max. :1.01e+08	Max. :1.10e+10
	(Other) :34	(Other) :56	NA	NA	NA	NA

Estandarización de Variables Continuas

	country	sector	beta	div	vol	cap
	Canada :60	Finance :48	Min. :-1.98	Min. :-2.84	Min. :-1.57	Min. :-1.62
	China :39	Basic Industries :28	1st Qu.: -0.86	1st Qu.: -0.54	1st Qu.: -0.94	1st Qu.: -0.85
	Brazil :29	Consumer Services:26	Median : -0.11	Median : 0.16	Median : -0.08	Median : -0.12
	uk :27	Public Utilities :26	Mean : 0.00	Mean : 0.00	Mean : 0.00	Mean : 0.00
	Netherlands:17	Energy :23	3rd Qu.: 0.86	3rd Qu.: 0.67	3rd Qu.: 0.82	3rd Qu.: 0.92
	Bermuda :16	Technology :15	Max. : 2.09	Max. : 2.04	Max. : 1.76	Max. : 1.79
	(Other) :34	(Other) :56	NA	NA	NA	NA

Cuando se trabaja con algoritmos de Cluster las distintas agrupaciones son definidas según la distancia de las observaciones en el espacio matemático. Trabajar con variables estandarizadas elimina la posible distorsión en el cálculo de la distancia cuando las unidades de medida de las variables no son comparables o existen diferencias de escala.



3. Selección de Variables

Subconjuntos posibles de variables continuas

beta	cap
beta	div
beta	vol
cap	div
cap	vol
div	vol

beta	cap	div
beta	cap	vol
beta	div	vol
cap	div	vol

Resultados test de Hopkins para los distintos sub-conjuntos de variables

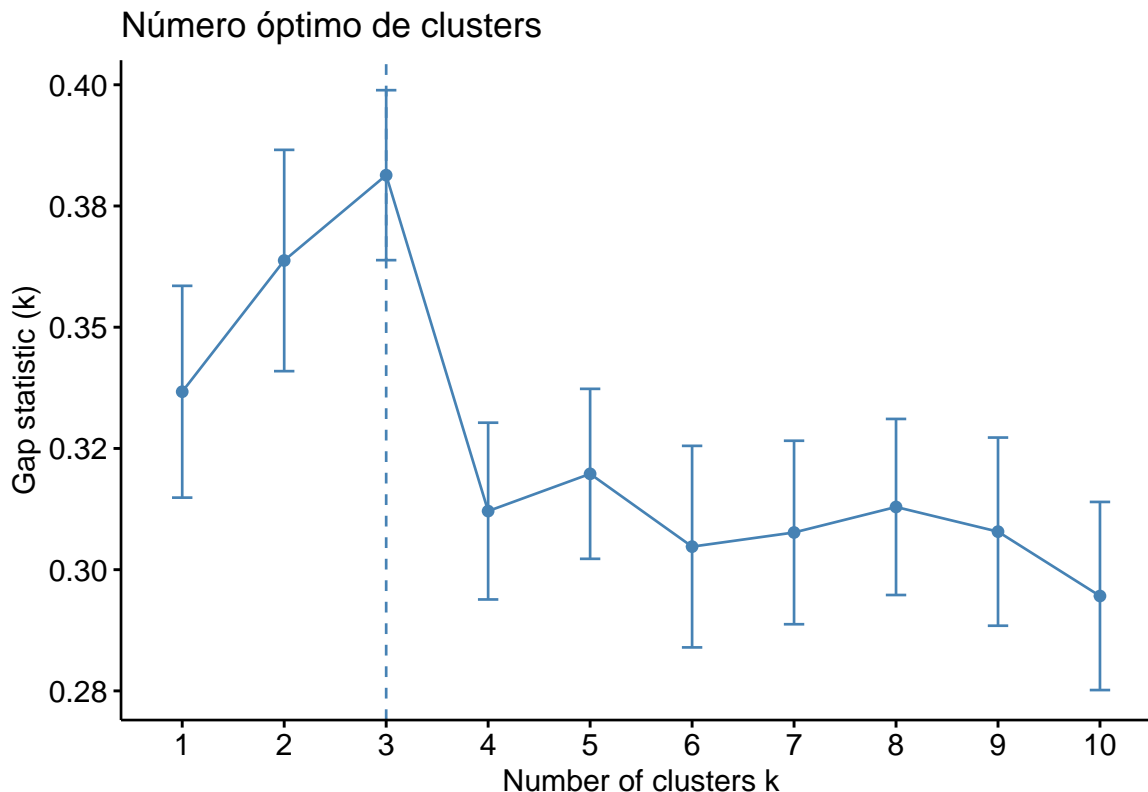
X1	X2	hopkins_stat
beta	cap	0.44
beta	div	0.36
beta	vol	0.55
cap	div	0.47
cap	vol	0.5
div	vol	0.49

X1	X2	X3	hopkins_stat
beta	cap	div	0.41
beta	cap	vol	0.49
beta	div	vol	0.41
cap	div	vol	0.42

En función de los resultados de los distintos test de hopkins realizados, la mejor alternativa es conservar las variables que presentan mayor estructura: **beta** y **div**.

4. Clustering inicial

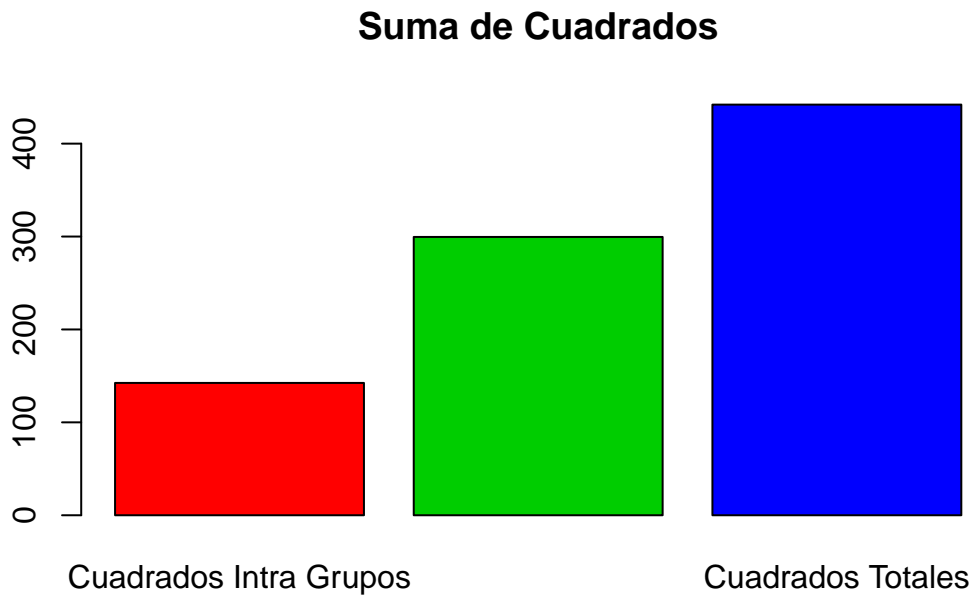
Criterio Gap



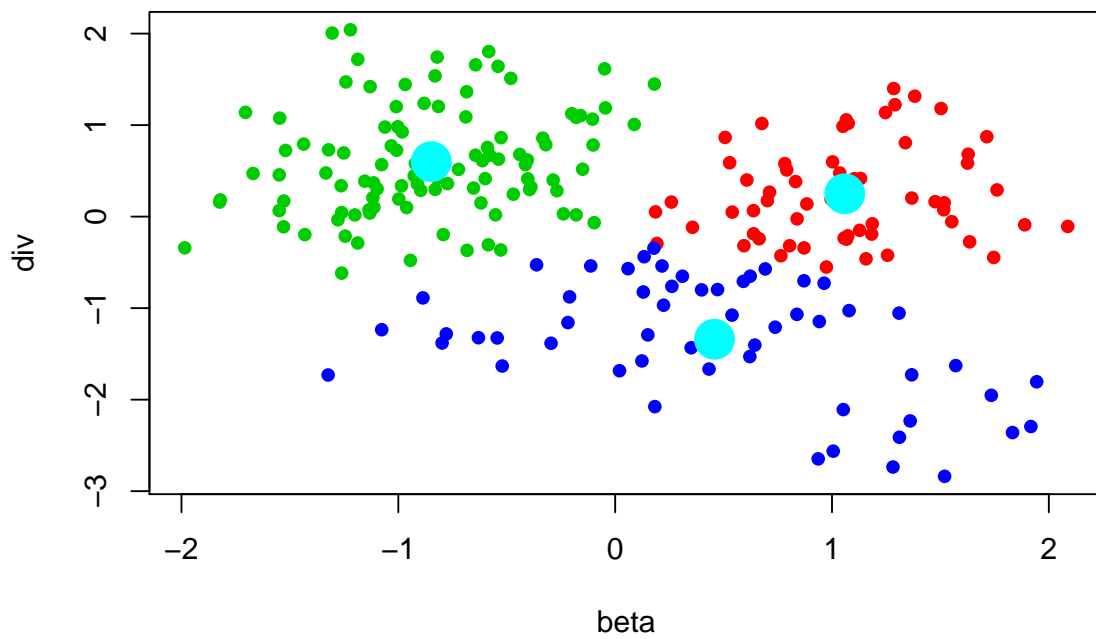
Kmeans

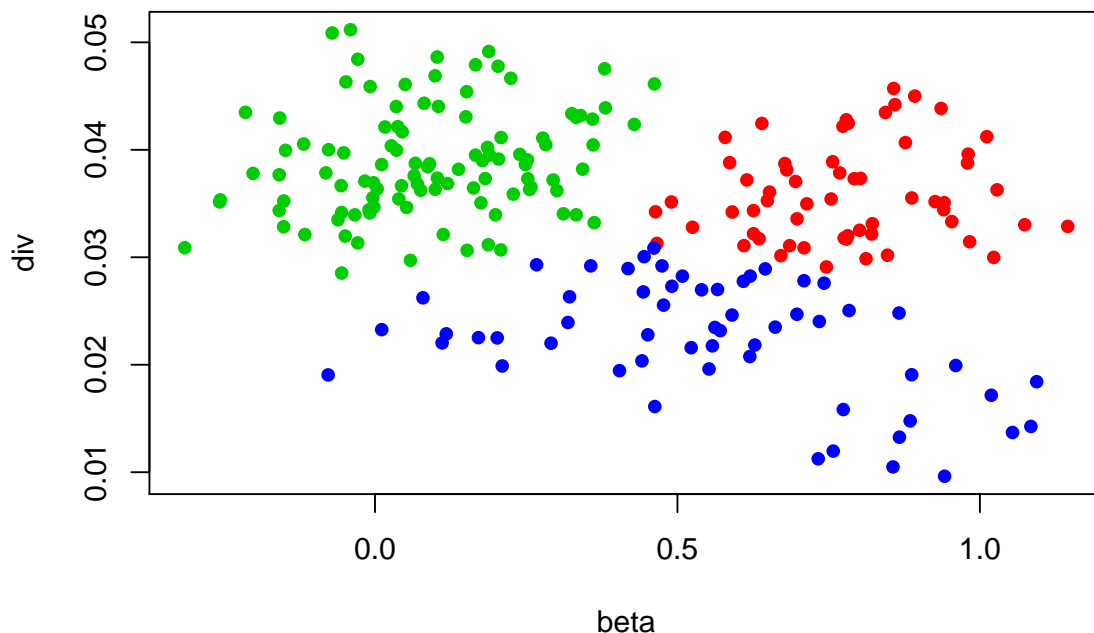
Una vez definido el número de clusters, según el criterio de **Estadístico Gap**, el algoritmo **kmeans** generará los centroides iniciales de cada cluster en forma aleatoria. Esto implica que la solución que se encontrará será un óptimo local dependiente del punto de partida aleatorio de los centroides. Una buena práctica consiste en definir la semilla generadora del algoritmo kmeans optando por alguna de las que genera menor distancia entre las observaciones de un mismo cluster (es decir, los individuos de un cluster más parecidos entre sí). Se presentan los resultados para una muestra aleatoria de 100 y la representación gráfica para alguna de las semillas que minimizan las obsevaciones intra-cluster.

almacen	Freq
142.445374983235	85
186.53563006236	6
186.906183355112	8
187.349540537448	1



Gráficos con asignación a clusters sobre Variables estandarizadas y sin estandarizar





La estructura de Clusters obtenida no se ajusta del todo al conocimiento específico. Se observa que el algoritmo clasifica de acuerdo a lo esperado para las acciones que, en promedio, presentan la menor volatilidad respecto al mercado (representadas con el color *Verde*). Las mismas pagan en promedio los más altos dividendos. El comportamiento de los otros agrupamientos es más difuso y resulta menos intuitivo en términos del conocimiento específico. El cluster representado por el color *Azul* muestra una gran dispersión respecto a la volatilidad de las acciones agrupadas y, en promedio, es el grupo de acciones que paga menores dividendos. Por su parte, el grupo representado en color *Rojo* es aquel que agrupa, en promedio, las acciones con mayor nivel de volatilidad respecto al mercado pero, contrariamente a lo que se podría esperar, pagan en promedio altos dividendos, similares al grupo de menor volatilidad respecto al mercado.

De lo anterior se desprende que, a partir del conocimiento específico, este primer agrupamiento no consigue el objetivo buscado. Si bien se llega a un segmento de acciones de empresas que se ajusta al perfil *Conservador* de los clientes de XYZ, los perfiles de *Crecimiento* y *Especulativo* no se encuentran correctamente segmentados.

5. Rating Crediticio

	country	sector	beta	div	vol	cap	cluster	score
AAV	Canada	Energy	0.43	0.04	6.9e+07	4.8e+09	2	3
ABX	Canada	Basic Industries	0.19	0.05	5.9e+07	3.3e+09	2	2
ACH	China	Basic Industries	0.06	0.03	3.4e+07	7.9e+09	2	3
AEB	Netherlands	Finance	0.20	0.02	1.1e+06	1.1e+09	3	2
AED	Netherlands	Finance	0.33	0.03	7.6e+06	7.9e+09	2	2
AEG	Netherlands	Finance	0.44	0.02	9.6e+06	8.9e+09	3	2

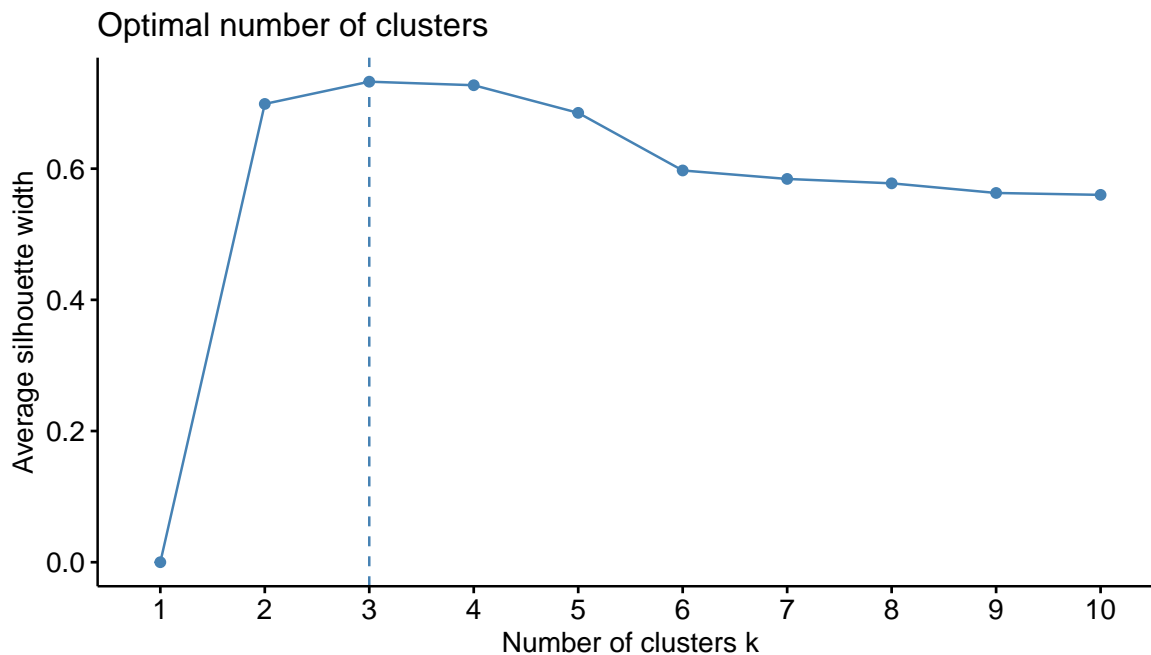
La calificación crediticia es una variable categórica ordinal. Sus valores responden a una clasificación del riesgo de imposibilidad de pago, por parte de una empresa, de cumplir con sus obligaciones financieras contractuales a su vencimiento, y las pérdidas económicas estimadas en caso de incumplimiento o incapacidad. El valor A representa el menor riesgo y C el mayor.

Esta variable puede resultar útil para enriquecer los datos originales de acciones aportando información

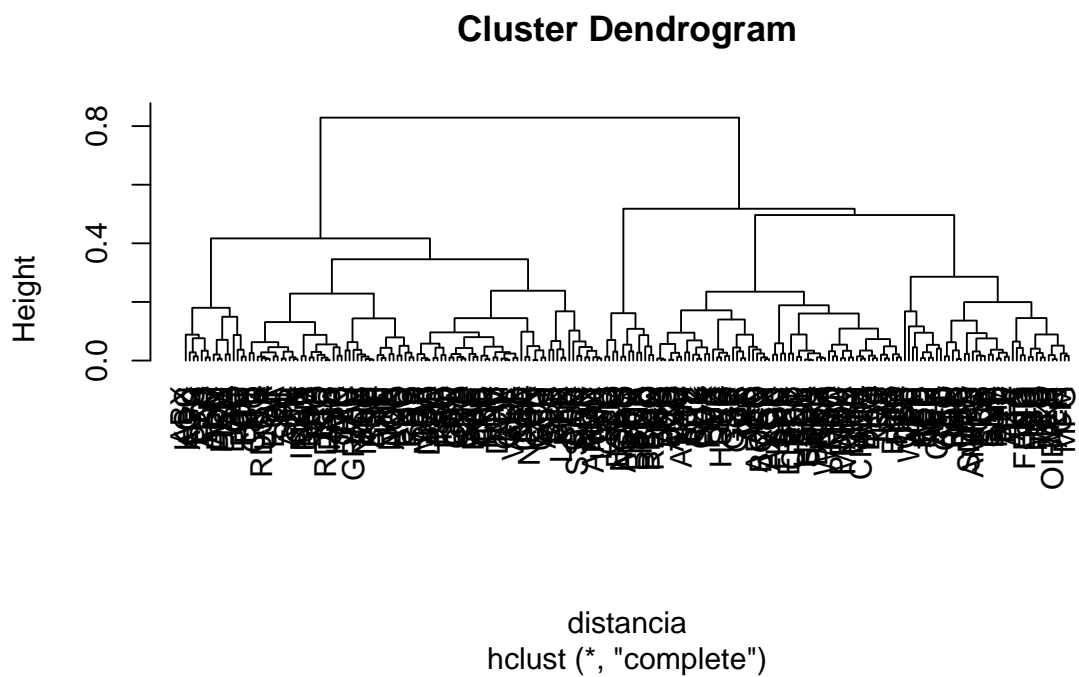
para el análisis. Es esperable que las calificaciones más riesgosas correspondan al perfil especulativo, mientras que las de menor riesgo puedan ser asignadas al perfil conservador.

6. Clustering Avanzado

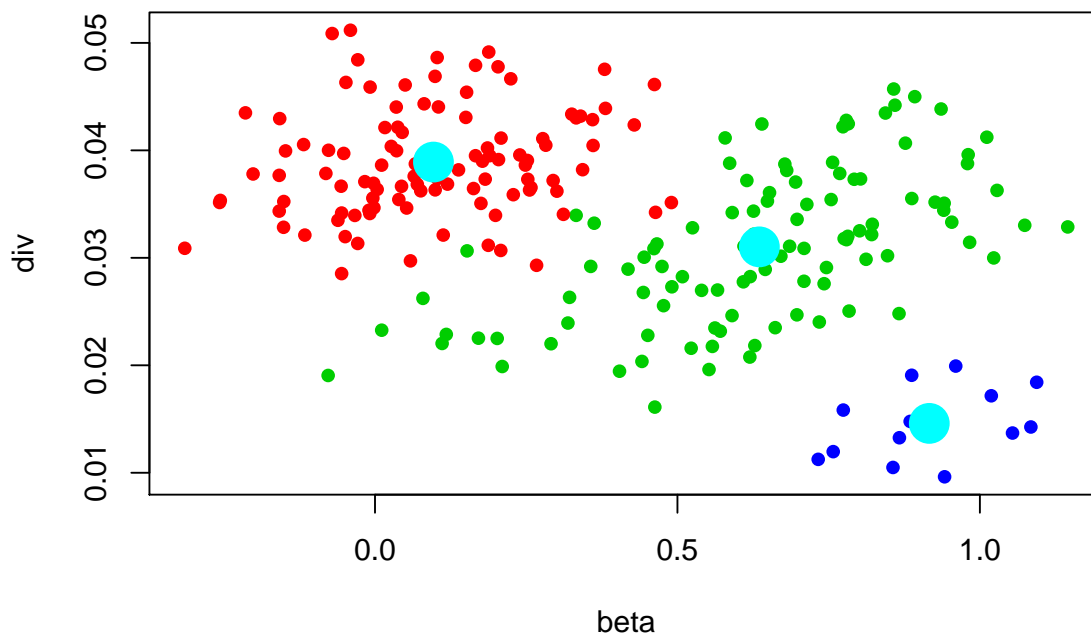
Potencial número de clusters según Silueta



Dendrograma Jerárquico



Gráficos con asignación a cluster jerárquico agregando Credit Score



La estructura obtenida tras la aplicación de clustering jerárquico, considerando las calificaciones crediticias de los distintos sectores y países, es consistente con la intuición previa formada a partir del conocimiento específico. El segmento que en promedio presenta acciones menos volátiles respecto al mercado es el que presenta los dividendos pagos más altos. Mientras tanto, el que presenta en promedio acciones más volátiles es el que menores dividendos paga.

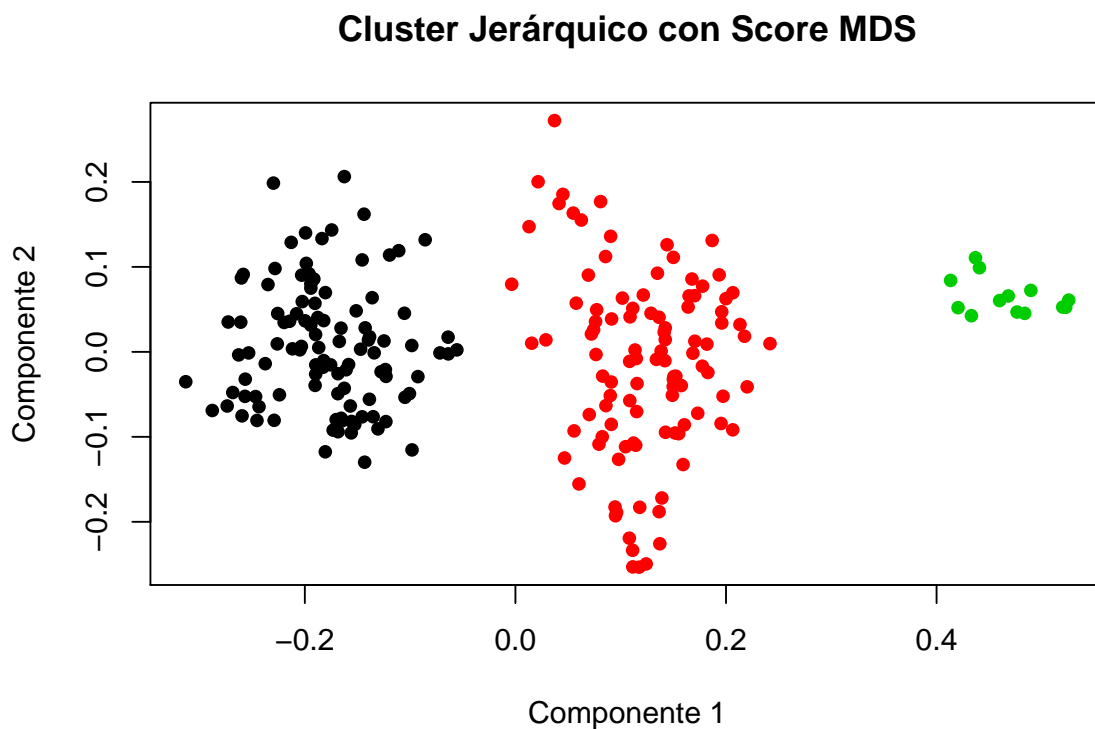
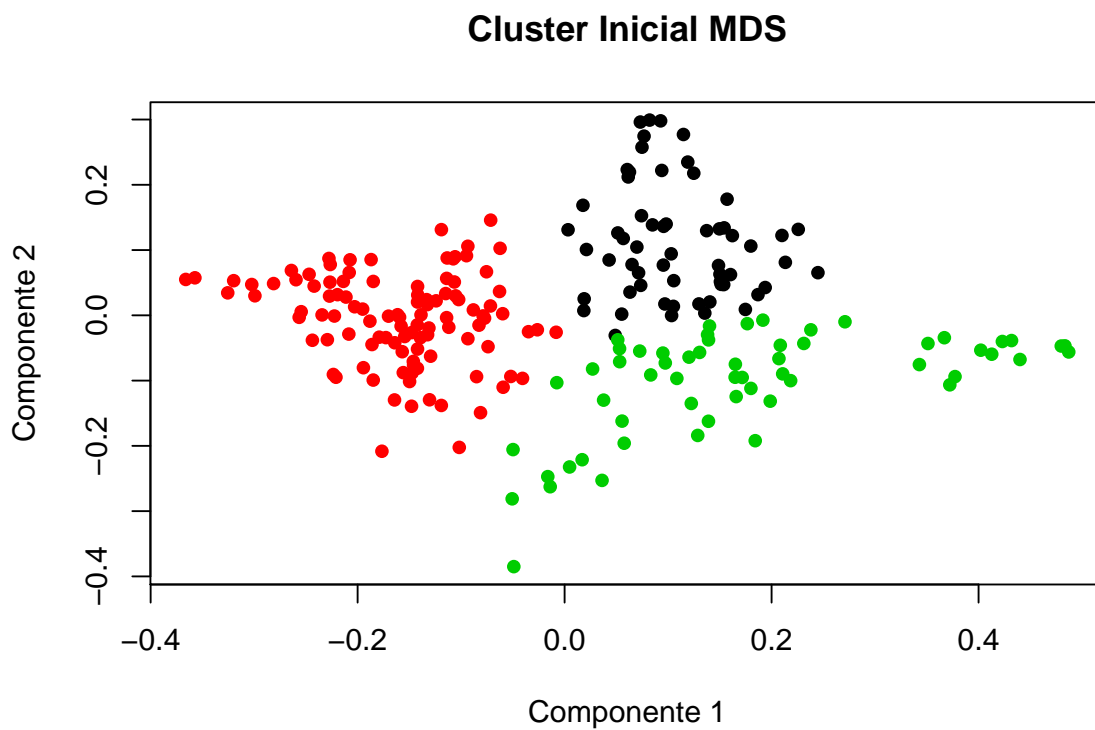
Así, se observa una segmentación que puede ser asociada a los distintos perfiles de clientes de XYZ. El perfil *Conservador* se ve representado en las acciones de color *Rojo*. El perfil *Especulativo* está segmentado en color *Azul*. Fianlemente las acciones representadas con el color *Verde* se ajustan al perfil de los clientes de *Crecimiento*

7. Comparación de Algoritmos de Clustering

##				
##		1	2	3 Sum
##	1	2	57	0 59
##	2	102	3	0 105
##	3	1	44	13 58
##	Sum	105	104	13 222

A partir de esta tabla de contingencia se observa por un lado que, los clusters 1 y 2 de los distintos modelos están cruzados. Esto debería corregirse reasignando los nombres en alguno de los clusters. La conclusión más interesante sin embargo es que, con el clustering jerárquico (que es el que mejor representa el conocimiento específico) la cantidad de acciones que se asocian al grupo de clientes *Especulativo* es solo el 6% del total.

8. Visualización



Aplicando *MDS* para mejorar la visualización, se observa que dos clusters no resultan claramente diferenciados bajo el procedimiento de *KMEANS*. Incorporando la variable de *calificación crediticia* y sustituyendo el método de clustering por el *Jerárquico*, *MDS* muestra una clara separación entre los

clusters definidos. La primera dimensión de MDS puede interpretarse como el componente principal en que puede desagregarse la matriz de distancias del cluster jerárquico respetando lo más posible su estructura.

9. Asignación

Del análisis de la clasificación realizada surge que, para el perfil *Conservador* las acciones que más se adaptan son aquellas pertenecientes a países más desarrollados e industrias con mayor grado de madurez. Por el contrario en el perfil *Especulativo* se concentran en acciones del sector Financiero y economías emergentes o menor grado de desarrollo. Para el perfil de *Crecimiento* se observa un peso importante del sector tecnológico en países emergentes y el financiero para países con economías estables.

Se presentan a continuación las tablas con las asignaciones de cada perfil según país y sector respectivamente.

	Conservador	Crecimiento	Especulativo
Argentina	0	9	3
Bermuda	3	13	0
Brazil	13	12	4
Canada	51	9	0
China	12	22	5
Japan	6	5	0
Mexico	4	6	1
Netherlands	6	11	0
uk	10	17	0

	Conservador	Crecimiento	Especulativo
Basic Industries	28	0	0
Capital Goods	3	7	0
Consumer Non-Durables	3	9	1
Consumer Services	7	19	0
Energy	17	6	0
Finance	0	37	11
Health Care	0	4	1
Miscellaneous	5	0	0
Public Utilities	21	5	0
Services	6	5	0
Technology	3	12	0
Transportation	12	0	0

Las siguientes tablas muestran las asignaciones a cada perfil, detallando la cantidad por país-industria:

Perfil conservador

country	sector	Freq
Canada	Basic Industries	16
Canada	Energy	12
Brazil	Public Utilities	8
Canada	Consumer Services	7
Canada	Public Utilities	5
Brazil	Basic Industries	4

Perfil Crecimiento

country	sector	Freq
China	Technology	10
Canada	Finance	9
Netherlands	Finance	9
Bermuda	Finance	7
uk	Finance	7
China	Consumer Services	6

Perfil especulativo

country	sector	Freq
Brazil	Finance	4
China	Finance	4
Argentina	Finance	2
Argentina	Consumer Non-Durables	1
Mexico	Finance	1
China	Health Care	1

10. Conclusiones

El trabajo realizado cumple con el objetivo trazado: se logra segmentar las acciones de empresas, de acuerdo con los distintos perfiles de los clientes de XYZ. Resulta interesante la importancia de incorporar más de una fuente de datos al análisis, lo que permitió un mejor desempeño que el observado únicamente considerando los datos de las acciones. Una recomendación a realizar es la de incorporar al análisis más acciones pertenecientes a sectores y mercados del segmento **especulativo** en función del desbalance que existe en el tamaño de la segmentación.