Analisis de Capacidad de absorción

## Introduccion

El siguiente informe de la capacidad de absorcion tiene como fin agrupar los departamentos mediante el metodo de aprendizaje no supervisado K-means con el fin de observar los departamentos que son similares y distintos segun el comportamiento que tienen en las distintas metricas.

Para el analisis de la capacidad de absorcion se utilizara los 24 departamentos para los cuales se tiene toda la informacion de las encuestas del Dane, excluyendo asi departamentos como Amazonas, Guania, Putumayo, Guaviare entre otros.

Las metricas o variables son las siguientes:

|  |  |  |  |
| --- | --- | --- | --- |
| **POBLACIÓN** | **BARRERAS DE ACCESO A SERVICIOS DE SALUD** | **TRABAJO INFANTIL** | **TASA DESEMPLEO** |
| **POBLACIÓN MIGRANTE** | **DESEMPLEO DE LARGA DURACIÓN** | **TRABAJO INFORMAL** | **DEMANDA LABORAL ANUAL** |
| **PORCENTAJE SOBRE POBLACION** | **HACINAMIENTO CRÍTICO** | **POBREZA MONETARIA** | **OFERTA LABORAL ANUAL** |
| **IDH** | **INADECUADA ELIMINACIÓN DE EXCRETAS** | **POBREZA EXTREMA** | **OFERTA/DEMANDA** |
| **PIB PER CAPITA PRECIOS CORRIENTES** | **INASISTENCIA ESCOLAR** | **DÉFICIT CUANTITATIVO** |  |
| **IPM** | **MATERIAL INADECUADO DE PAREDES EXTERIORES** | **DÉFICIT CUALITATIVO** |  |
| **INTENSIDAD IPM** | **MATERIAL INADECUADO DE PISOS** | **DÉFICIT HABITACIONAL** |  |
| **ANALFABETISMO** | **REZAGO ESCOLAR** | **REGIMEN SUBSIDIADO** |  |
| **BAJO LOGRO EDUCATIVO** | **SIN ACCESO A FUENTE DE AGUA MEJORADA** | **REGIMEN CONTRIBUTIVO** |  |
| **BARRERAS A SERVICIOS PARA CUIDADO DE LA PRIMERA INFANCIA** | **SIN ASEGURAMIENTO EN SALUD** | **COBERTURA** |  |

El modelo K-Means es realizado para conocer los departamentos con caracteristicas similares y asi observar cuales tienen una Buena capacidad de obsorcion y cuales no para recibir migrantes.

## K-means

K-means es un método de agrupación, que busca y clasifica K grupos de observaciones llamados clusters. Los elementos u observaciones con las mismas características se agruparán en un mismo grupo y se separarán de otros grupos con características diferentes. El algoritmo de K-means determina si los datos son similares o diferentes por la distancia entre los datos. Aquellas observaciones que sean similares tendrán una distancia menor entre ellas, se utilizará como medida la distancia euclidiana

## Caso 1: Agrupacion solo con las variables con valores porcentuales

## Matriz de distancia entre departamentos.

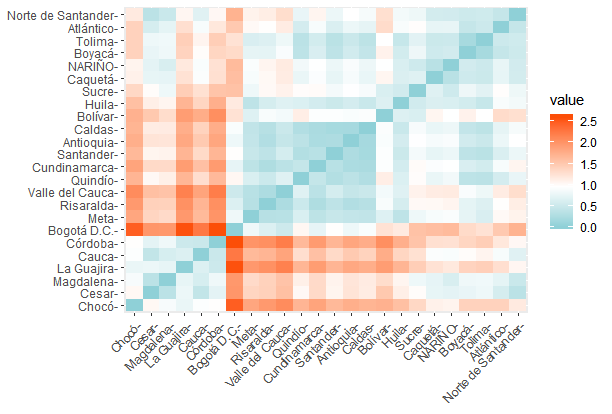


Figura 1: Matriz de distancia

De la matriz de distancia se observa que existen departamentos con caracteristicas similares y otros con caracteristicas distintas por lo que es factible agruparlos segun sus distancias.

Se realizan dos metodos con el fin de obtener el numero optimo de k, es decir el numero de clusters o grupos que minimiza las distancias entre observaciones y maximiza las distancias entre clusters.

## Metodo 1: WSS

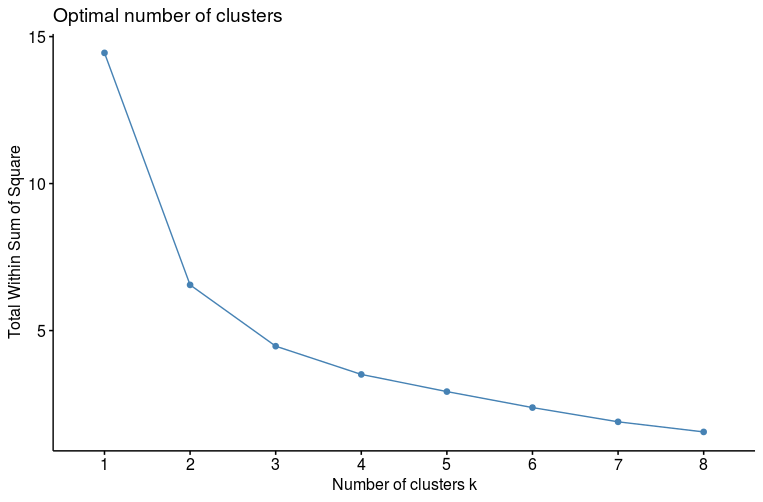


Figura 2: Metodo 1

## Method 2: The silhouette

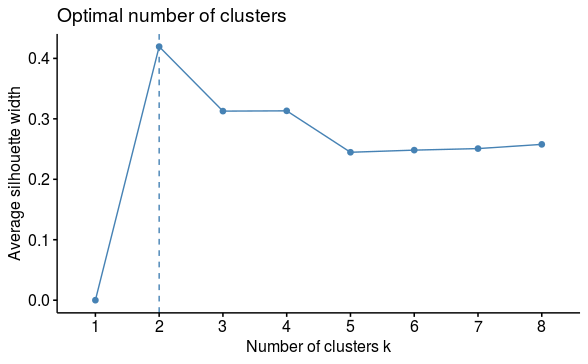


Figure 3: Metodo 2

## Cluster k=2

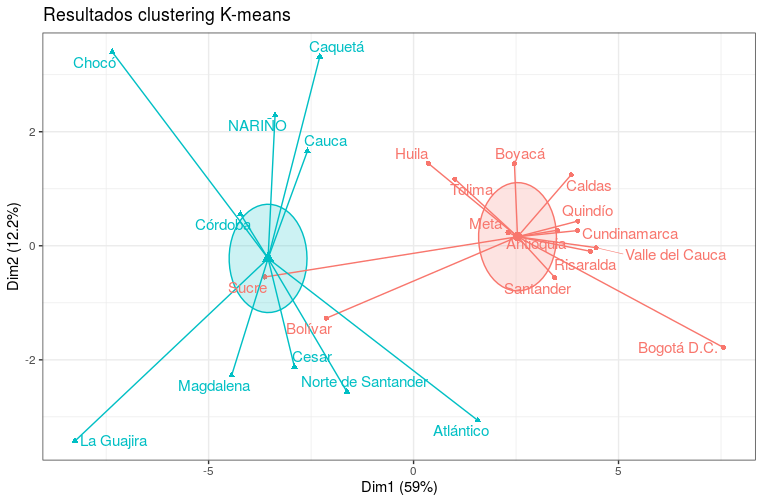


Figura 4: Caso 1 Modelo con 2 clusters

## Cluster k=3

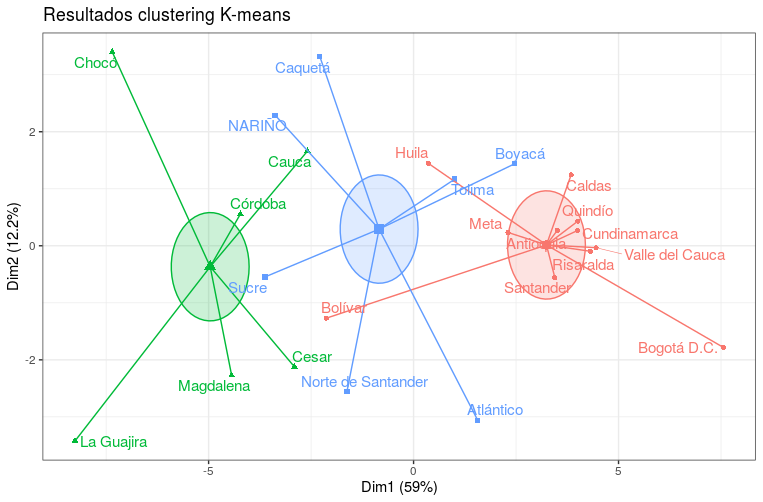


Figura 5: Modelo con 3 clusters

## Cluster k=4

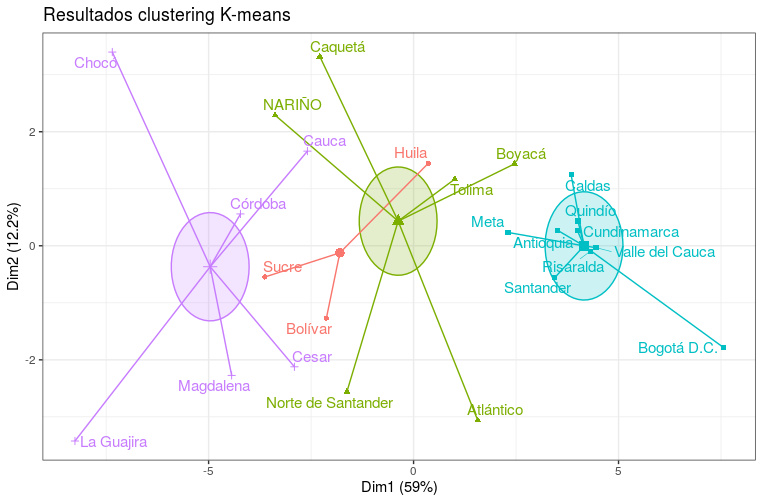


Figura 6: Modelo con 4 clusters

## Dendograma

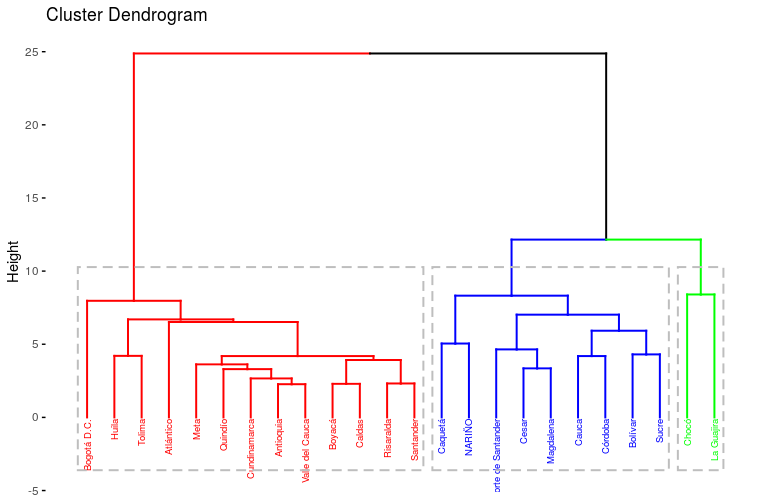


Figura 7: Dendograma

# Metricas

## Tamaño de clusters

K2 = [10,14]

K3= [11,6,7]

## Withinss Desempeño del modelo

K2 = 6.55

K3= 4.46

## Distancia entre clusters

K2 = 7.89

K3= 9.97

## Distancia entre observaciones

K2 = 3.340236, 3.212881

K3= 2.007200, 1.366046, 1.095056

# Cluster center coordinates

**K2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **PORCENTAJE SOBRE POBLACION** | **IDH** | **IPM** | **Intensidad IPM** | **Analfabetismo** | **Bajo logro educativo** |
| 1 | 0,04724197 | 0,7206 | 0,3475 | 0,4299 | 0,1752 | 0,5668 |
| 2 | 0,02147227 | 0,7622857 | 0,1789286 | 0,4030714 | 0,09735714 | 0,4732857 |
| **Cluster** | **Desempleo de larga duración** | **Hacinamiento crítico** | **Inadecuada eliminación de excretas** | **Inasistencia escolar** | **Material inadecuado de paredes exteriores** | **Material inadecuado de pisos** |
| 1 | 0,0978 | 0,1559 | 0,2589 | 0,0555 | 0,0623 | 0,1482 |
| 2 | 0,1265 | 0,07621429 | 0,09785714 | 0,03092857 | 0,02035714 | 0,05071429 |
| **Cluster** | **Sin aseguramiento en salud** | **Trabajo infantil** | **Trabajo informal** | **POBREZA MONETARIA** | **POBREZA EXTREMA** | **Déficit cuantitativo** |
| 1 | 0,1362 | 0,0338 | 0,8706 | 0,4464 | 0,1565 | 0,2102097 |
| 2 | 0,1002143 | 0,02278571 | 0,7388571 | 0,2511429 | 0,05728571 | 0,08001737 |
| **Cluster** | **REGIMEN SUBSIDIADO** | **REGIMEN CONTRIBUTIVO** | **COBERTURA** | **TASA DESEMPLEO** | **OFERTA/DEMANDA** | **Rezago escolar** |
| 1 | 0,6790769 | 0,2106853 | 0,8897622 | 0,1014 | 1,809 | 0,36 |
| 2 | 0,4819398 | 0,4365834 | 0,9185233 | 0,1058669 | 0,8211569 | 0,2853571 |
| **Cluster** | **Déficit habitacional** | **Sin acceso a fuente de agua mejorada** | **Barreras de acceso a servicios de salud** | **Déficit cualitativo** | **Barreras a servicios para cuidado de la primera infancia** |  |
| 1 | 0,610947 | 0,2428 | 0,1085 | 0,4007373 | 0,1112 |  |
| 2 | 0,349695 | 0,1126429 | 0,05764286 | 0,2696776 | 0,08242857 |  |

**K3**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **PORCENTAJE SOBRE POBLACION** | **IDH** | **IPM** | **Intensidad IPM** | **Analfabetismo** | **Bajo logro educativo** |
| 1 | 0,0233334 | 0,7672727 | 0,1551818 | 0,3997273 | 0,08072727 | 0,4533636 |
| 2 | 0,04505945 | 0,709 | 0,3895 | 0,44 | 0,2105 | 0,5881667 |
| 3 | 0,0351439 | 0,7405714 | 0,2765714 | 0,415 | 0,13771429 | 0,5397143 |
| **Cluster** | **Desempleo de larga duración** | **Hacinamiento crítico** | **Inadecuada eliminación de excretas** | **Inasistencia escolar** | **Material inadecuado de paredes exteriores** | **Material inadecuado de pisos** |
| 1 | 0,1270909 | 0,07018182 | 0,088 | 0,02963636 | 0,02081818 | 0,037 |
| 2 | 0,1011667 | 0,17666667 | 0,3415 | 0,05666667 | 0,0635 | 0,21216667 |
| 3 | 0,1062857 | 0,11342857 | 0,1345714 | 0,046 | 0,04257143 | 0,07314286 |
| **Cluster** | **Sin aseguramiento en salud** | **Trabajo infantil** | **Trabajo informal** | **POBREZA MONETARIA** | **POBREZA EXTREMA** | **Déficit cuantitativo** |
| 1 | 0,1013636 | 0,01890909 | 0,7152727 | 0,23 | 0,05281818 | 0,07378949 |
| 2 | 0,1476667 | 0,03216667 | 0,8838333 | 0,4983333 | 0,204 | 0,23378653 |
| 3 | 0,1091429 | 0,03657143 | 0,8398571 | 0,3514286 | 0,08028571 | 0,14399096 |
| **Cluster** | **REGIMEN SUBSIDIADO** | **REGIMEN CONTRIBUTIVO** | **COBERTURA** | **TASA DESEMPLEO** | **OFERTA/DEMANDA** | **Rezago escolar** |
| 1 | 0,4432095 | 0,4785893 | 0,9217988 | 0,1028306 | 0,732326 | 0,276 |
| 2 | 0,711423 | 0,1854858 | 0,8969088 | 0,1038333 | 2.056.558 | 0,3753333 |
| 3 | 0,6277263 | 0,2630892 | 0,8908156 | 0,106 | 1.313.096 | 0,3295714 |
| **Cluster** | **Déficit habitacional** | **Sin acceso a fuente de agua mejorada** | **Barreras de acceso a servicios de salud** | **Déficit cualitativo** | **Barreras a servicios para cuidado de la primera infancia** |  |
| 1 | 0,3177646 | 0,1038182 | 0,04454545 | 0,2439751 | 0,08472727 |  |
| 2 | 0,6902392 | 0,2986667 | 0,0875 | 0,4564527 | 0,11983333 |  |
| 3 | 0,4811934 | 0,153 | 0,12528571 | 0,3372024 | 0,08785714 |  |

## Caso 2 Todas las variables escaladas.

## Matriz de distancia.

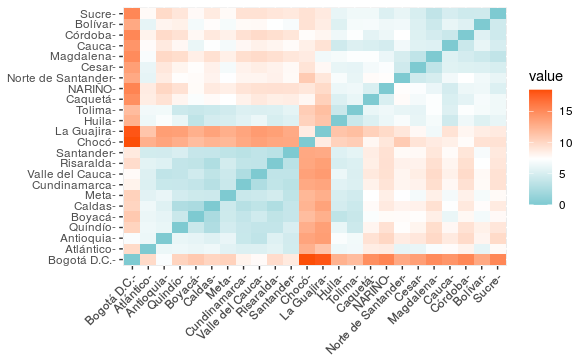


Figura 8Matriz de distancia caso 2

## Metodo 1: WSS

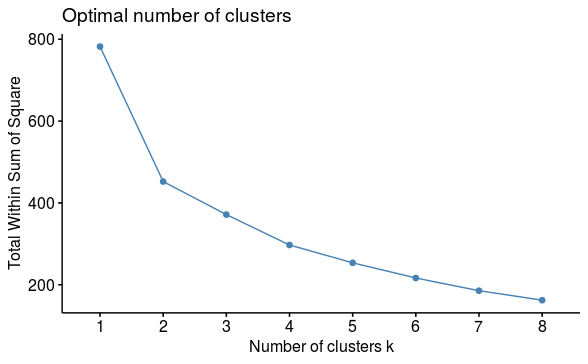


Figura 9: metodo 1: caso2

## Metodo 2 : silhouette

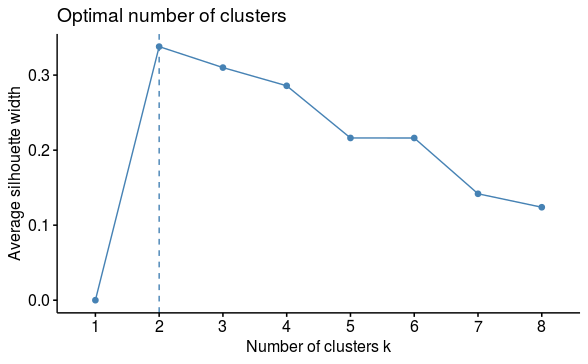


Figura 10: Metodo 3: caso2

## Cluster con K=2

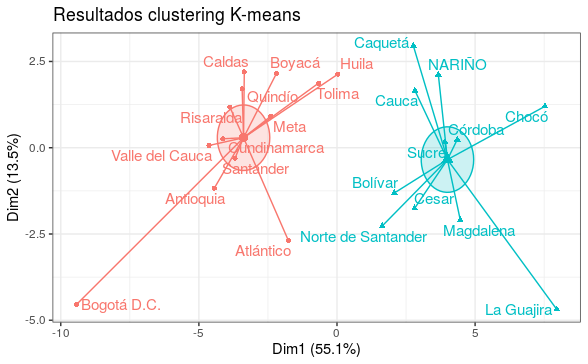


Figure 11: Modelo con 2 clusters, caso2

## Cluster con K=3

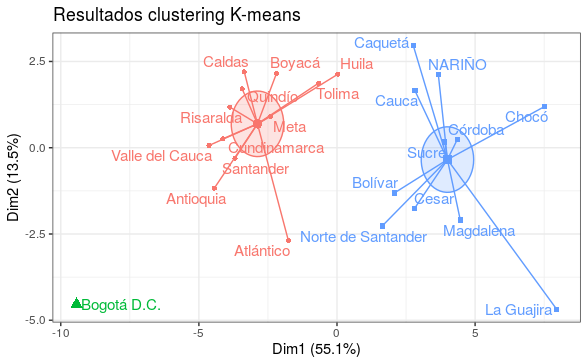


Figure 12: Modelo con 3 clusters, caso2

## Dendograma k=2

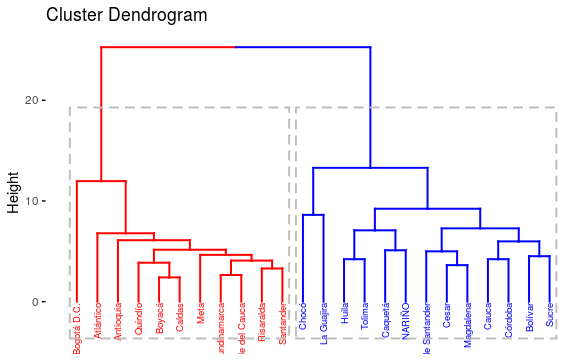


Figura 13: Dendograma caso2

# Metricas

## Tamaño de clusters

K2 = [13,11]

## Withinss Desempeño del modelo

K2 = 452.24

## Distancia entre clusters

K2 = 329.75

## Distancia entre observaciones

K2 = [209.1992, 243.0458]

# Cluster center coordinates

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **POBLACIÓN** | **POBLACIÓN MIGRANTE** | **PORCENTAJE SOBRE POBLACION** | **IDH** | **PIB PER CAPITA PRECIOS CORRIENTES** | **IPM** | **Intensidad IPM** |
| 1 | 0.3295251 | 0.1134470 | -0.2432443 | 0.6979590 | 0.5855694 | -0.7941100 | -0.7022402 |
| 2 | -0.3894387 | -0.1340737 | 0.2874705 | -0.8248606 | -0.6920365 | 0.9384936 | 0.8299203 |
| **Cluster** | **Analfabetismo** | **Bajo logro educativo** | **Barreras a servicios para cuidado de la primera infancia** | **Barreras de acceso a servicios de salud** | **Desempleo de larga duración** | **Hacinamiento crítico** | **Inadecuada eliminación de excretas** |
| 1 | -0.7343318 | -0.5550073 | -0.3784753 | -0.5260051 | 0.5292981 | -0.6516696 | -0.6138380 |
| 2 | 0.8678467 | 0.6559177 | 0.4472889 | 0.6216424 | -0.6255341 | 0.7701550 | 0.7254449 |
| **Cluster** | **Inasistencia escolar** | **Material inadecuado de paredes exteriores** | **Material inadecuado de pisos** | **Rezago escolar** | **Sin acceso a fuente de agua mejorada** | **Sin aseguramiento en salud** | **Trabajo infantil** |
| 1 | -0.7300429 | -0.6628377 | -0.6109164 | -0.7223356 | -0.4952822 | -0.2979741 | -0.5521818 |
| 2 | 0.8627779 | 0.7833536 | 0.7219921 | 0.8536693 | 0.5853335 | 0.3521512 | 0.6525785 |
| **Cluster** | **Trabajo informal** | **POBREZA MONETARIA** | **POBREZA EXTREMA** | **Déficit cuantitativo** | **Déficit cualitativo** | **Déficit habitacional** | **REGIMEN SUBSIDIADO** |
| 1 | -0.7250163 | -0.7800891 | -0.5817844 | -0.6456246 | -0.6936937 | -0.7807788 | -0.7335951 |
| 2 | 0.8568375 | 0.9219234 | 0.6875634 | 0.7630109 | 0.8198198 | 0.9227386 | 0.8669760 |
| **Cluster** | **REGIMEN CONTRIBUTIVO** | **COBERTURA** | **TASA DESEMPLEO** | **DEMANDA LABORAL ANUAL** | **OFERTA LABORAL ANUAL** | **OFERTA/DEMANDA** |  |
| 1 | 0.7319695 | 0.1225913 | 0.1055494 | 0.2943336 | 0.4043100 | -0.5530858 |  |
| 2 | -0.8650549 | -0.1448806 | -0.1247402 | -0.3478488 | -0.4778209 | 0.6536468 |  |