CU34_MODEL_DEVELOPMENT_01_CLUSTER

June 13, 2023

#

CU34_Predicción de demanda de servicios

1 IV. Model development

En este anexo se incluye el código utilizado durante el desarrollo de los modelos incluidos en el caso de uso.

```
[1]: Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8")
```

'es_ES.UTF-8/es_ES.UTF-8/es_ES.UTF-8/C/es_ES.UTF-8/C'

1.1 Modelo CLUSTER

1.2 Paquetes

```
[2]: library(readr)
     library(FactoMineR)
     library(factoextra)
     library(dplyr)
     library(tidyr)
     library(janitor)
     library(cluster)
     library(DT)
```

Loading required package: ggplot2

Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test

1.3 Datos

```
[3]: datos <- read_csv("CU_34_05_05_servicios_completo.csv")

    # fechaclus <- "2022-01-01"
    # set.seed(1)
    # municlus <- sample(datos$CMUN, 1)

    # apply(datos, 2, \(x) sum(is.na(x)))

    datosclust <- datos |>
      # filter(Fecha == as.Date(fechaclus)) |>
      # filter(CMUN == municlus) |>
      drop_na() |>
      # select(-c(Fecha:CSEC, ccaa, CPRO, NSEC)) |>
      remove_constant()
```

Rows: 274792 Columns: 33
  Column specification

Delimiter: ","
chr   (5): Servicio, CMUN, CDIS, CSEC, NSEC
dbl (27): Futbol, nservicios, capacidad, tmed, prec, velmedia,
presMax, cca…
date (1): Fecha

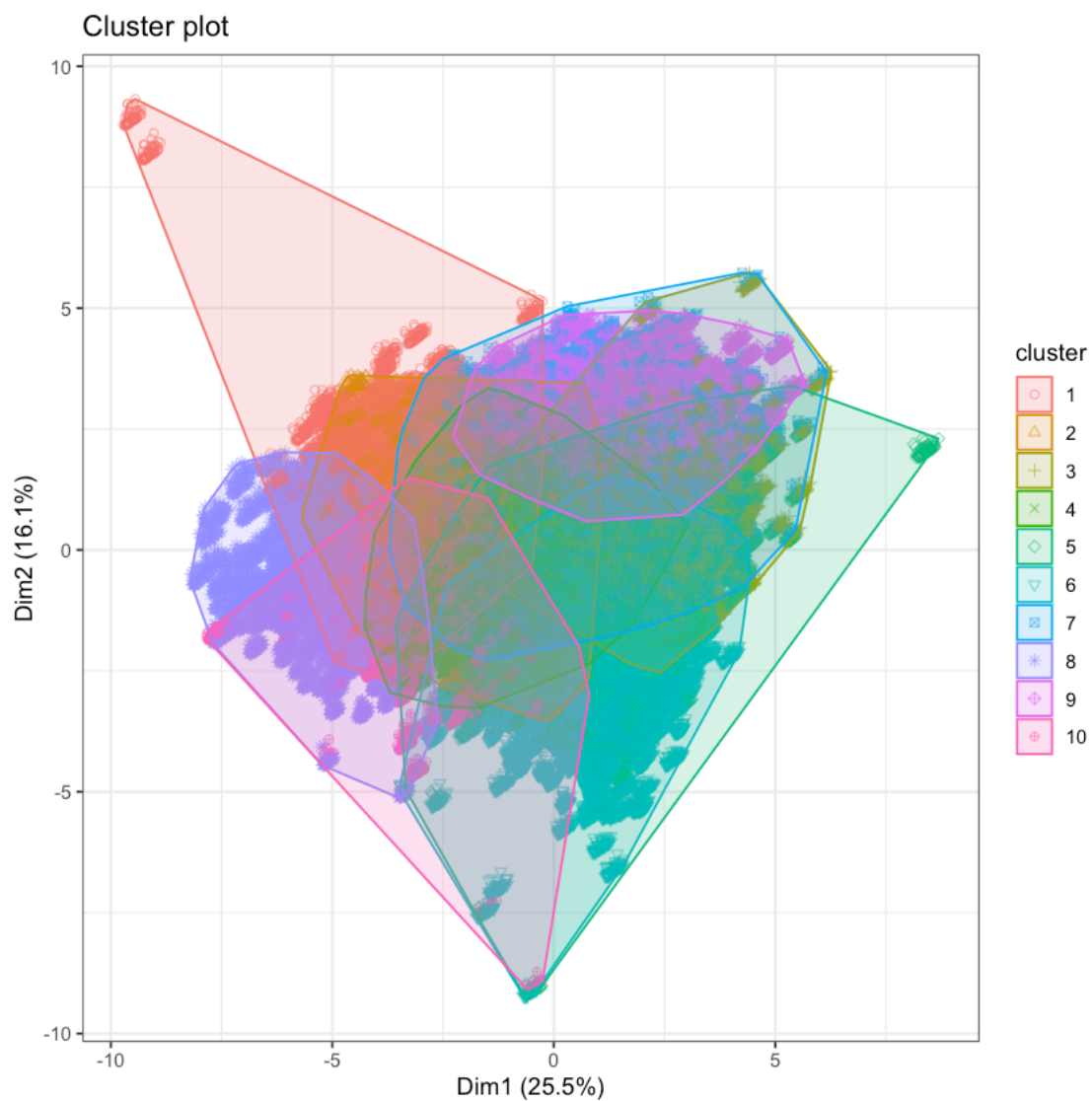 Use `spec()` to retrieve the full column specification for this
data.
  Specify the column types or set `show_col_types = FALSE` to quiet
this message.

1.4 Cluster

```
[4]: zdatosclust <- datosclust |>
       select(-c(Fecha:CSEC, NSEC)) |>
       scale()

     # km <- kmeans(datosclust, 4)
     km <- clara(zdatosclust, 10)
     # plot(km)
     fviz_cluster(km, geom = "point", alpha = 0.5) +
       theme_bw()

     datos2 <-  datosclust |>
       mutate(cluster = km$clustering) |>
       relocate(cluster, NSEC, .before = 1)
```


Cluster plot

1.5 Output

```
[5]: ## TABLA CARACTERIZACIÓN
     datos2 |>
       group_by(cluster) |>
       summarise(across(Futbol:densidad_hab_km2, median)) |>
       datatable()

     ## TABLA DETALLE MUNICIPIO
     datos2 |>
       filter(CMUN == "079")

     ## DATOS PARA MAPA

     datos2 |>
       group_by(NSEC, CMUN, CDIS, CSEC, cluster) |>
       summarise(n = n()) |>
       rowwise() |>
       slice_max(n) |>
       ungroup() |>
       datatable()

     ## guardar datos

     write_rds(datos2, "output_cluster.rds")
     write_rds(km, "modelo_cluster.rds")
```

HTML widgets cannot be represented in plain text (need html)

cluster	NSEC	Fecha	Servicio	CMUN	CDIS	CSEC	Futbol	ns
<int>	<chr>	<date>	<chr>	<chr>	<chr>	<chr>	<dbl>	<
7	Madrid - 01.001	2022-01-01	Taxi	079	01	001	1	50
6	Madrid - 01.002	2022-01-01	Taxi	079	01	002	0	3
6	Madrid - 01.003	2022-01-01	Taxi	079	01	003	0	1
3	Madrid - 01.004	2022-01-01	Taxi	079	01	004	0	0
4	Madrid - 01.006	2022-01-01	Taxi	079	01	006	0	5
3	Madrid - 01.007	2022-01-01	Taxi	079	01	007	0	6
3	Madrid - 01.008	2022-01-01	Taxi	079	01	008	0	5
7	Madrid - 01.009	2022-01-01	Taxi	079	01	009	1	51
4	Madrid - 01.011	2022-01-01	Taxi	079	01	011	0	2
6	Madrid - 01.012	2022-01-01	Taxi	079	01	012	0	2
6	Madrid - 01.013	2022-01-01	Taxi	079	01	013	0	4
7	Madrid - 01.014	2022-01-01	Taxi	079	01	014	1	53
6	Madrid - 01.015	2022-01-01	Taxi	079	01	015	0	11
4	Madrid - 01.016	2022-01-01	Taxi	079	01	016	0	1
7	Madrid - 01.018	2022-01-01	Taxi	079	01	018	1	50
3	Madrid - 01.019	2022-01-01	Taxi	079	01	019	0	1
4	Madrid - 01.020	2022-01-01	Taxi	079	01	020	0	2
4	Madrid - 01.021	2022-01-01	Taxi	079	01	021	0	5
4	Madrid - 01.022	2022-01-01	Taxi	079	01	022	0	6
6	Madrid - 01.023	2022-01-01	Taxi	079	01	023	0	7
6	Madrid - 01.024	2022-01-01	Taxi	079	01	024	1	53
4	Madrid - 01.025	2022-01-01	Taxi	079	01	025	0	2
6	Madrid - 01.026	2022-01-01	Taxi	079	01	026	0	5
4	Madrid - 01.027	2022-01-01	Taxi	079	01	027	0	5
6	Madrid - 01.028	2022-01-01	Taxi	079	01	028	0	1
6	Madrid - 01.029	2022-01-01	Taxi	079	01	029	0	3
6	Madrid - 01.030	2022-01-01	Taxi	079	01	030	0	7
6	Madrid - 01.031	2022-01-01	Taxi	079	01	031	0	8
1	Madrid - 01.032	2022-01-01	Taxi	079	01	032	0	4
6	Madrid - 01.033	2022-01-01	Taxi	079	01	033	0	4
...	...	...	...	...	...	...	...	...
9	Madrid - 21.002	2022-01-31	Delivery	079	21	002	0	12
7	Madrid - 21.003	2022-01-31	Delivery	079	21	003	1	64
9	Madrid - 21.004	2022-01-31	Delivery	079	21	004	0	15
1	Madrid - 21.005	2022-01-31	Delivery	079	21	005	0	16
9	Madrid - 21.006	2022-01-31	Delivery	079	21	006	0	14
4	Madrid - 21.007	2022-01-31	Delivery	079	21	007	0	11
7	Madrid - 21.008	2022-01-31	Delivery	079	21	008	1	63
9	Madrid - 21.009	2022-01-31	Delivery	079	21	009	0	10
9	Madrid - 21.010	2022-01-31	Delivery	079	21	010	0	17
9	Madrid - 21.011	2022-01-31	Delivery	079	21	011	0	15
9	Madrid - 21.012	2022-01-31	Delivery	079	21	012	0	13
7	Madrid - 21.013	2022-01-31	Delivery	079	21	013	1	62
4	Madrid - 21.014	2022-01-31	Delivery	079	21	014	0	13
6	Madrid - 21.015	2022-01-31	Delivery	079	21	015	0	4
6	Madrid - 21.017	2022-01-31	Delivery	079	21	017	0	10
6	Madrid - 21.018	2022-01-31	Delivery	079	21	018	0	9
6	Madrid - 21.019	2022-01-31	Delivery	079	21	019	0	7
7	Madrid - 21.020	2022-01-31	Delivery	079	21	020	1	60
6	Madrid - 21.021	2022-01-31	Delivery	079	21	021	0	9
4	Madrid - 21.022	2022-01-31	Delivery	079	21	022	0	8

A tibble: 151466 x 32

`summarise()` has grouped output by 'NSEC', 'CMUN', 'CDIS', 'CSEC'. You can
override using the `.groups` argument.

HTML widgets cannot be represented in plain text (need html)