

CU45_MODEL_DEVELOPMENT_01_CLUSTER

June 13, 2023

#

CU45_Planificación y promoción del destino en base a los patrones en origen de los turistas

1 IV. Model development

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

1.1 Modelo CLUSTER

1.1.1 Paquetes

```
[1]: library(readr)
library(dplyr)
library(tidyr)
library(stringr)
library(purrr)
library(tibble)
```

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

1.1.2 Datos

```
[2]: receptor <- read_csv("CU_45_05_03_receptor.csv")
      interno_prov <- read_csv("CU_45_05_04_interno_prov.csv")
      interno_muni <- read_csv("CU_45_05_05_interno_mun.csv")
      valoraciones <- read_csv("CU_45_05_02_valoracion_sim.csv")

Rows: 50294 Columns: 7
-- Column specification
-----
Delimiter: ","
chr (5): mes, pais_orig_cod, pais_orig, mun_dest, CMUN
dbl (2): mun_dest_cod, turistas

i Use `spec()` to retrieve the full column specification for this
data.
i Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 521280 Columns: 8
-- Column specification
-----
Delimiter: ","
chr (6): total_nacional, total_ccaa, provincia, municipio_destino,
cmun, mes
dbl (2): turistas, secreto

i Use `spec()` to retrieve the full column specification for this
data.
i Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 79246 Columns: 10
-- Column specification
-----
Delimiter: ","
chr (7): mes, mun_orig_cod, mun_orig, dest, prov_orig_cod, prov_orig,
prov_dest
dbl (3): dest_cod, turistas, prov_dest_cod

i Use `spec()` to retrieve the full column specification for this
data.
i Specify the column types or set `show_col_types = FALSE` to quiet
this message.
Rows: 24780 Columns: 9
-- Column specification
-----
Delimiter: ","
chr (5): grupo, tipo, nombre, CMUN, CDIS
dbl (4): X, Y, t3_1, puntos
```

i Use ``spec()`` to retrieve the full column specification for this data.

i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.

1.1.3 Clustering

```
[3]: turistas_origen_mes <-
  receptor |>
  filter(pais_orig == "Total") |>
  select(mes, CMUN, mun_dest, turistas) |>
  mutate(receptor = turistas) |>
  select(-turistas) |>
  inner_join(
    interno_prov |>
      filter(!is.na(total_ccaa),
             is.na(provincia)) |>
      select(mes, cmun, total_ccaa, turistas) |>
      pivot_wider(names_from = total_ccaa,
                  values_from = turistas,
                  values_fn = ~sum(.x, na.rm = TRUE)),
    by = c("CMUN" = "cmun", "mes" = "mes"))

val <- valoraciones |>
  group_by(CMUN, grupo) |>
  summarise(puntos = median(puntos, na.rm = TRUE), .groups = "drop") |>
  pivot_wider(names_from = grupo,
              values_from = puntos) |>
  mutate(across(2:4, ~replace_na(.x, median(.x, na.rm = TRUE))))

kdata <- turistas_origen_mes |>
  inner_join(val, by = "CMUN") |>
  # select(-municipio_destino) |>
  mutate(anyo = str_sub(mes, 1, 4)) |>
  group_by(anyo, mun_dest) |>
  summarise(across(3:(ncol(turistas_origen_mes) - 1), sum),
            across((ncol(turistas_origen_mes) - 1):(ncol(turistas_origen_mes) +
↪2), median))

by_anyo <- split(kdata, kdata$anyo)

by_anyo[[1]] |>
  column_to_rownames("mun_dest") |>
  select(-anyo) |>
  kmeans(10)
```

```
## PARÁMETRO DEL PASO 2
NCLUS <- 4

cluster_anyos <- map(by_anyo, function(x) {
  m <- x |>
  column_to_rownames("mun_dest") |>
  select(-anyo) |>
  kmeans(NCLUS)
  res <- x |>
  mutate(cluster = m$cluster)
}) |> bind_rows() |>
ungroup()

write_csv(cluster_anyos, "cluster_anyos.csv")
```

`summarise()` has grouped output by 'anyo'. You can override using the
`.groups` argument.

K-means clustering with 10 clusters of sizes 11, 18, 11, 15, 3, 14, 36, 10, 25, 1

Cluster means:

	receptor	Andaluc<U+00ED>a	Arag<U+00F3>n	Asturias, Principado de
1	12363.3636	2526.46145	1083.2727273	1379.272727
2	770.1111	344.38889	42.4444444	42.500000
3	5386.8182	3373.94927	603.7272727	921.000000
4	1802.0667	1085.13333	236.6000000	257.333333
5	47009.3333	28.35633	2794.3226667	3549.771667
6	23805.7143	15.82829	2393.6428571	2870.594786
7	398.8611	102.11111	0.9166667	1.944444
8	2002.8000	1130.40000	207.0000000	133.700000
9	1167.8800	765.20000	52.2400000	91.720000
10	2602822.0000	854.18400	163.6530000	177.368000
	Balears, Illes	Canarias	Cantabria	Castilla - La Mancha
1	690.1818182	877.8181818	745.545455	345.2752
2	3.8333333	8.0555556	6.166667	2715.7839
3	342.0000000	369.8181818	438.727273	1509.8508
4	49.2666667	89.4666667	71.333333	1325.8213
5	2494.0000000	3410.3333333	2808.666667	114.6107
6	1364.4285714	1828.2142857	1568.642857	52.9965
7	0.9166667	0.8888889	4.861111	584.0551
8	18.4000000	42.8000000	62.000000	3987.8337
9	11.4400000	33.0400000	25.200000	506.6774
10	125.8520000	186.8080000	88.249000	1183.6420
	Castilla y Le<U+00F3>n	Catalu<U+00F1>a	Ceuta	Comunitat Valenciana
1	2539.13891	2812.545455	0.000000	2367.62164
2	328.95411	145.500000	0.000000	330.33333
3	2826.46864	1411.454545	2.818182	2899.91318

4	2372.66667	555.733333	0.000000	892.73333
5	22.09633	13.736667	202.000000	21.26500
6	531.52129	1841.078357	34.714286	191.22450
7	217.25722	8.777778	0.000000	79.13889
8	756.10000	500.400000	0.000000	1406.50000
9	707.72000	240.960000	0.000000	627.32000
10	738.77200	525.048000	712.229000	615.44900
	Extremadura	Galicia	Madrid, Comunidad de	Melilla
1	1708.63636	2796.727273	0.00000	12.90909
2	48.88889	100.055556	0.00000	0.00000
3	1421.90909	1612.454545	91.63636	13.27273
4	269.26667	526.400000	0.00000	0.00000
5	1647.49900	641.107000	0.00000	152.66667
6	2114.36114	2554.543000	0.00000	67.21429
7	11.61111	8.472222	0.00000	0.00000
8	256.30000	386.700000	0.00000	0.00000
9	217.96000	201.680000	0.00000	0.00000
10	204.37200	354.562000	0.00000	969.50800
	Murcia,	Regi<U+00F3>n de Navarra,	Comunidad Foral de	Pa<U+00ED>s Vasco
1	1464.272727	425.6363636	1533.72727	
2	27.333333	1.7777778	77.61111	
3	741.000000	184.0000000	792.00000	
4	178.733333	31.5333333	264.46667	
5	1708.727333	1888.3333333	2603.97867	
6	3160.080857	953.5000000	2727.36486	
7	6.416667	0.8888889	19.97222	
8	353.300000	11.0000000	293.50000	
9	91.560000	2.5200000	146.12000	
10	163.752000	85.4480000	244.09300	
	Rioja, La	comercio hosteleria	turismo	
1	271.727273	2.909091	2.818182	2.909091
2	0.000000	2.750000	2.833333	3.055556
3	64.181818	2.772727	3.090909	3.045455
4	8.133333	3.533333	3.000000	2.966667
5	1210.000000	3.000000	3.000000	2.500000
6	582.571429	3.142857	3.142857	3.178571
7	0.000000	3.250000	3.486111	2.916667
8	6.100000	3.000000	3.050000	2.800000
9	1.240000	3.120000	3.240000	2.880000
10	46.747000	4.000000	4.000000	4.000000

Clustering vector:

Ajalvir	Alcal<U+00E1> de Henares
8	5
Alcobendas	Alcorc<U+00F3>n
5	6
Aldea del Fresno	Algete
7	3

Alpedrete	Ambite
4	2
Aranjuez	Arganda del Rey
1	1
Arroyomolinos	Atazar, El
3	7
Becerril de la Sierra	Berrueco, El
9	7
Boadilla del Monte	Boalo, El
1	9
Brunete	Buitrago del Lozoya
4	9
Bustarviejo	Cabrera, La
2	2
Cadalso de los Vidrios	Camarma de Esteruelas
4	9
Campo Real	Caraba<U+00F1>a
9	7
Casarrubuelos	Cercedilla
9	4
Chinch<U+00F3>n	Ciempozuelos
8	1
Cobe<U+00F1>a	Collado Mediano
2	4
Collado Villalba	Colmenar Viejo
1	1
Colmenar de Oreja	Colmenar del Arroyo
9	7
Colmenarejo	Coslada
9	6
Cubas de la Sagra	Daganzo de Arriba
9	8
Escorial, El	Fresnedillas de la Oliva
3	7
Fresno de Torote	Fuenlabrada
2	6
Fuente el Saz de Jarama	Fuentidue<U+00F1>a de Tajo
9	2
Galapagar	Garganta de los Montes
3	7
Getafe	Gri<U+00F1><U+00F3>n
5	9
Guadalix de la Sierra	Guadarrama
9	3
Horcajo de la Sierra-Aoslos	Hoyo de Manzanares
9	9
Humanes de Madrid	Legan<U+00E9>s
3	6

Loeches	Lozoya
8	7
Lozoyuela-Navas-Sieteiglesias	Madrid
9	10
Majadahonda	Manzanares el Real
6	9
Meco	Mejorada del Campo
4	3
Miraflores de la Sierra	Molar, El
2	4
Molinos, Los	Montejo de la Sierra
9	7
Moraleja de Enmedio	Moralzarzal
9	4
Morata de Taju<U+00F1>a	M<U+00F3>stoles
2	6
Navacerrada	Navalafuente
9	7
Navalagamella	Navalcarnero
7	3
Navas del Rey	Nuevo Bazt<U+00E1>n
7	2
Paracuellos de Jarama	Parla
1	6
Patones	Pedrezuela
2	7
Pelayos de la Presa	Perales de Taju<U+00F1>a
4	8
Pezuela de las Torres	Pinto
7	6
Pozuelo de Alarc<U+00F3>n	Pozuelo del Rey
6	7
Pr<U+00E1>dena del Rinc<U+00F3>n	Puebla de la Sierra
7	7
Puentes Viejas	Quijorna
9	7
Rascafr<U+00ED>a	Redue<U+00F1>a
4	7
Ribatejada	Rivas-Vaciamadrid
7	1
Robledillo de la Jara	Robledo de Chavela
7	4
Rozas de Madrid, Las	Rozas de Puerto Real
6	7
San Agust<U+00ED>n del Guadalix	San Fernando de Henares
8	6
San Lorenzo de El Escorial	San Mart<U+00ED>n de Valdeiglesias
3	2

San Mart<U+00ED>n de la Vega	San Sebasti<U+00E1>n de los Reyes
3	6
Santa Mar<U+00ED>a de la Alameda	Santorcaz
4	2
Serna del Monte, La	Sevilla la Nueva
7	2
Somosierra	Soto del Real
7	4
Talamanca de Jarama	Tielmes
7	7
Titulcia	Torrej<U+00F3>n de Ardoz
7	6
Torrej<U+00F3>n de Velasco	Torrej<U+00F3>n de la Calzada
9	9
Torrelaguna	Torrelodones
2	3
Torres de la Alameda	Tres Cantos
8	1
Valdelaguna	Valdemanco
7	7
Valdemaqueda	Valdemorillo
7	4
Valdemoro	Valdeolmos-Alalpardo
6	2
Valdepi<U+00E9>lagos	Valdilecha
7	9
Velilla de San Antonio	Vell<U+00F3>n, El
8	2
Venturada	Villa del Prado
7	9
Villaconejos	Villalbilla
2	8
Villamanrique de Tajo	Villamanta
2	7
Villanueva de Perales	Villanueva de la Ca<U+00F1>ada
7	1
Villanueva del Pardillo	Villar del Olmo
4	7
Villarejo de Salvan<U+00E9>s	Villaviciosa de Od<U+00F3>n
8	1
Zarzalejo	<U+00C1>lamo, El
7	9

Within cluster sum of squares by cluster:

[1] 178990492 19730258 101354640 54378445 62184119 397608240 16087381

[8] 20252277 25627787 0

(between_SS / total_SS = 100.0 %)

Available components:

[1]	"cluster"	"centers"	"totss"	"withinss"	"tot.withinss"
[6]	"betweenss"	"size"	"iter"	"ifault"	