

# **Fastbook 04**

## **Analítica de Cliente & Predictive Analytics**

Práctica de Segmentación (II).  
Trabajamos con R



Edix Educación

## 04. Práctica de Segmentación (II). Trabajamos con R

Ha llegado el momento de ampliar la parte práctica con este último fastbook sobre segmentación.

1. Veremos ejemplos de **segmentaciones anonimizadas**.

Sin más, procedamos a ver la motivación de cada diferencia.

*Autor: Miguel Ángel Fernández*



Introducción



Caso 1: K-Means



Caso 2: Aglomerativo



Caso 3: DBSCAN



Caso 4: Patrones de segmentación

# Introducción



Python y R son los dos **lenguajes más empleados en ciencia de datos**. Cada uno tiene sus virtudes y, según nuestra formación académica, podemos encontrar uno más atractivo que el otro. Los matemáticos y estadísticos suelen encontrarse más cómodos con R, mientras que perfiles con un corte más informático (ingenieros y desarrolladores) prefieren Python.

Independientemente de las preferencias que tengamos, a lo largo de nuestra trayectoria profesional nos vamos a encontrar con **unos equipos que usen R y otros Python** para hacer sus análisis o compañías en las que establezca uno de los dos lenguajes todos los equipos dedicados a analizar los datos. También se da el caso en que un mismo equipo afronte diferentes tipos de proyectos con cada lenguaje.

**Sea cual sea la circunstancia en la que nos vayamos a encontrar, nuestro perfil es mucho más completo si sabemos manejarnos en ambos lenguajes.**

Y recuerda: en realidad no hay dos segmentaciones iguales: podemos encontrarnos con multitud de casuísticas diferentes en cada segmentación de clientes a la que nos enfrentemos.

Por esta razón, en este fastbook vamos a ver ejemplos de segmentaciones más cortas y que dibujen patrones más variados para ampliar nuestra experiencia sin dedicar casi esfuerzo a tratar y analizar los datos. Para este objetivo, **las variables van a estar anonimizadas**, es decir, van a llamarse  $x_1, x_2, \dots, x_N$ ; de esta manera no habrá interpretación posible sobre los datos.

Para hacer una segmentación correctamente hay que seguir siempre las mismas fases que vimos en los ejemplos del fastbook 03:

- 1 Leer datos.
- 2 Limpieza de los datos.
- 3 EDA.
- 4 Ajustar modelo.

No obstante, en este tema queremos centrarnos en saber **ajustar modelos de segmentación con R y observar ejemplos más variados**. Para cumplir este objetivo vamos a ver cuatro casos, cada uno en un notebook distinto.

En los tres primeros simplemente veremos **cómo segmentar en R usando K-Means, Aglomerativo y DBSCAN**, mientras que en el último aprenderemos que cada modelo tiene sus **puntos fuertes y débiles**, de modo que en algunas situaciones nos conviene usar uno u otro.

# Caso 1: K-Means



En el primer caso, tenemos un dataset con 2000 filas (clientes) y disponemos de 7 variables donde veremos cómo **segmentar en R usando K-Means**.



Descárgate el [fichero](#) para acceder al 'Caso 1: K-Means'.

## Caso 2: Aglomerativo



En el segundo caso tenemos un dataset con aproximadamente 9000 filas (clientes) y disponemos de 17 variables donde veremos cómo **segmentar en R usando un clustering aglomerativo**.



Descárgate el [fichero](#) para acceder al 'Caso 2: Aglomerativo'.

## Caso 3: DBSCAN



En el tercer caso tendremos un dataset muy pequeño con 178 clientes, y 13 donde veremos cómo **segmentar en R usando DBSCAN**. Podemos pensar en este caso como una segmentación de los mejores 178 clientes de nuestra compañía. Como vimos en teoría, es muy útil estudiar segmentos concretos de nuestra cartera de clientes en vez de todos de golpe. Si fuésemos un banco, estos 178 clientes podrían ser las empresas con mayor volumen de facturación.



Descárgate el [fichero](#) para acceder al 'Caso 3: DBSCAN'.

## Caso 4: Patrones de segmentación



Una vez hemos visto los tres primeros casos, que muy intencionadamente han sido elegidos por su sencillez, podemos ver este último caso 4. Aquí estudiaremos **qué patrones pueden dibujar nuestros clientes** y, en función de ellos, **entenderemos cuándo es mejor usar K-Means, Aglomerativo o DBSCAN**.

Cada modelo tiene unas características que le permiten segmentar bien unos patrones determinados. Por ejemplo, el diseño de K-Means solo le permite generar segmentos con forma de circunferencia.



Descárgate el [fichero](#) para acceder al 'Caso 4: Patrones de segmentación'.

¡Enhorabuena! Fastbook superado



edix