

PROGRAMA DE
ESPECIALIZACIÓN ANALÍTICA

ADVANCED DATA SCIENCE

Métodos No Supervisados

SESIÓN II

Docente: Jimmy Salazar

REGLAS



Se requiere **puntualidad** para un mejor desarrollo del curso.



Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.



Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.



Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.



Identificarse en la sala Zoom con el primer nombre y primer apellido.

ITINERARIO

*07:00 PM – 07:30 PM **Soporte técnico DMC***

*07:30 PM – 08:30 PM **Módulo 1***

*08:30 PM – 09:30 PM **Módulo 2***

Horario de Atención Área Académica 09:00 am a 10:00 pm

AGENDA

1. Reducción de Dimensionalidad

2. Segmentación DBSCAN

- Metodología
- PCA
- Perfil y Visualización
- Caso práctico real

EVALUACIÓN

Asistencia (Curso):

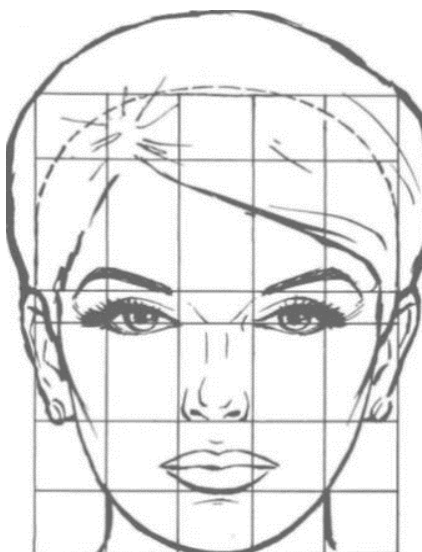
mínimo 80% sesiones para
recibir la certificación

Examen Final
(30%)

+

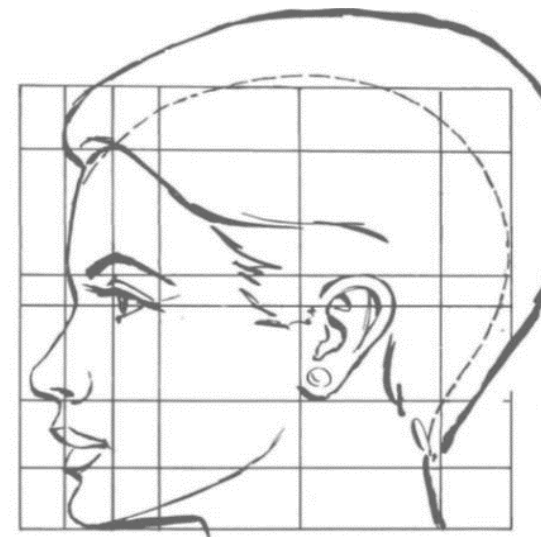
Trabajo
(70%)

Reducción de Dimensionalidad



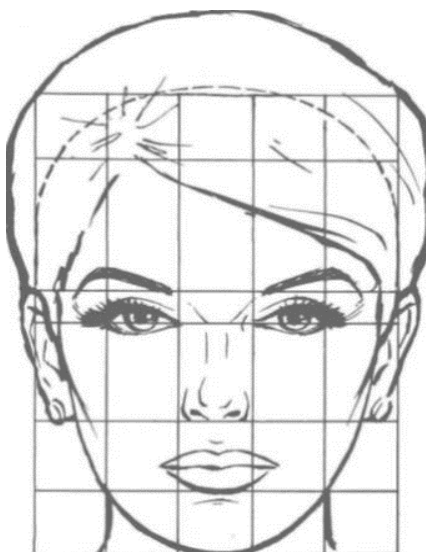
6 VARIABLES

- Ojo derecho
- Ojo izquierdo
- Nariz completa
- Boca completa
- Oreja derecho
- Oreja izquierda



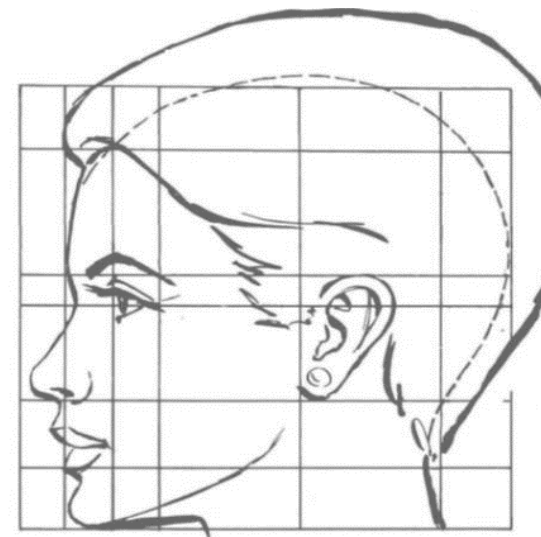
4 VARIABLES

- Ojo izquierdo
- Nariz media
- Boca media
- Oreja izquierda



6 VARIABLES

- Ojo derecho
- Ojo izquierdo
- Nariz completa
- Boca completa
- Oreja derecho
- Oreja izquierda



4 VARIABLES

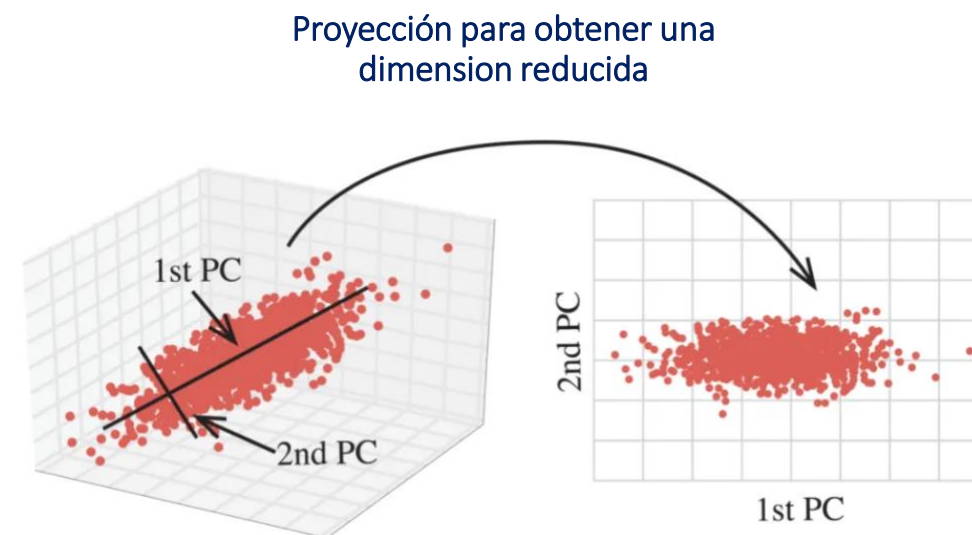
- Ojo izquierdo
- Nariz media
- Boca media
- Oreja izquierda



Podemos Identificar a la
imagen original

Características

- ✓ Procedimiento matemático que reduce un conjunto de variables posiblemente correlacionadas a un conjunto de variables no correlacionadas.
- ✓ El objetivo es obtener un conjunto de variables que expliquen la mayor variabilidad posible de las variables originales.
- ✓ Es una técnica solo aplicable para variables numéricas.
- ✓ Sensible a la escala de las variables (transformación para homologar sus dimensiones)



Saigayatri Vadali , ene-18

Agrupamiento basado en Densidad

Características

- ✓ Los métodos basados en la distancia tienden a funcionar bien con clusteres esféricos y mal con clusteres de otras formas.
- ✓ Para solucionar este problema otros métodos han desarrollado el concepto de **densidad**, el cual permite descubrir clusteres con **formas arbitrarias** (conjunto de datos que determinan un volumen).
- ✓ La idea es hacer crecer un cluster siempre cuando la densidad en el entorno del objeto exceda de un **umbral**.

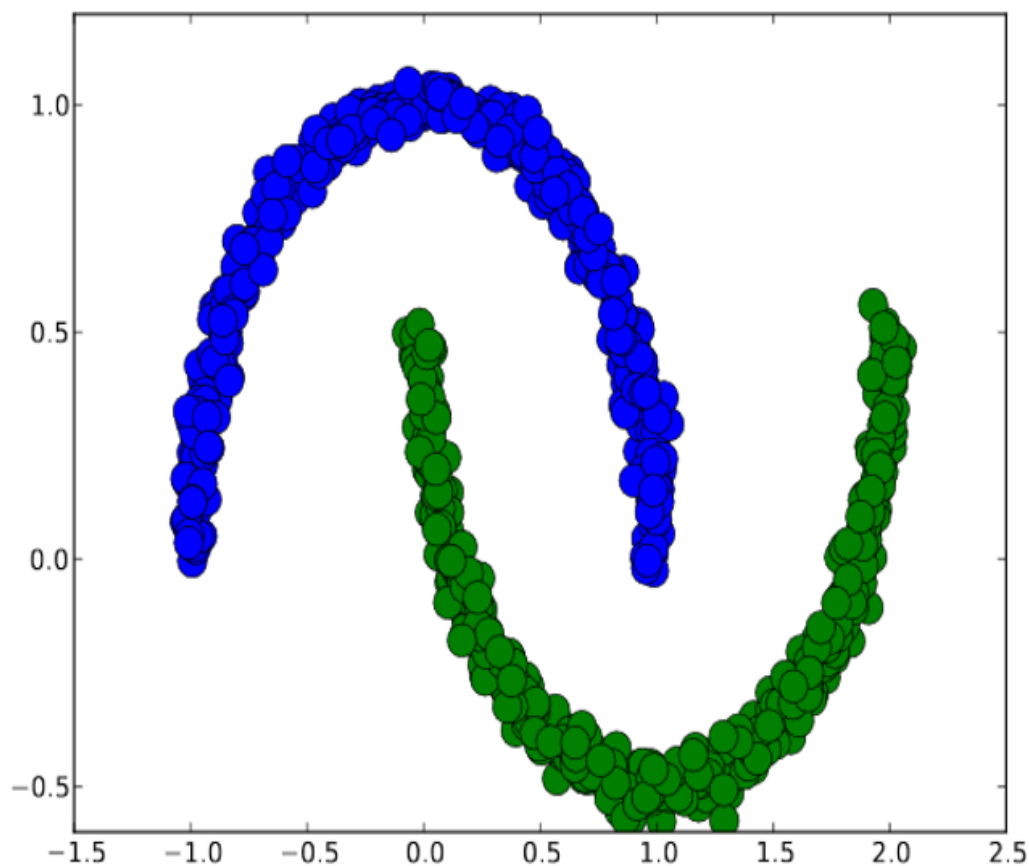
Segmentación DBSCAN

Características

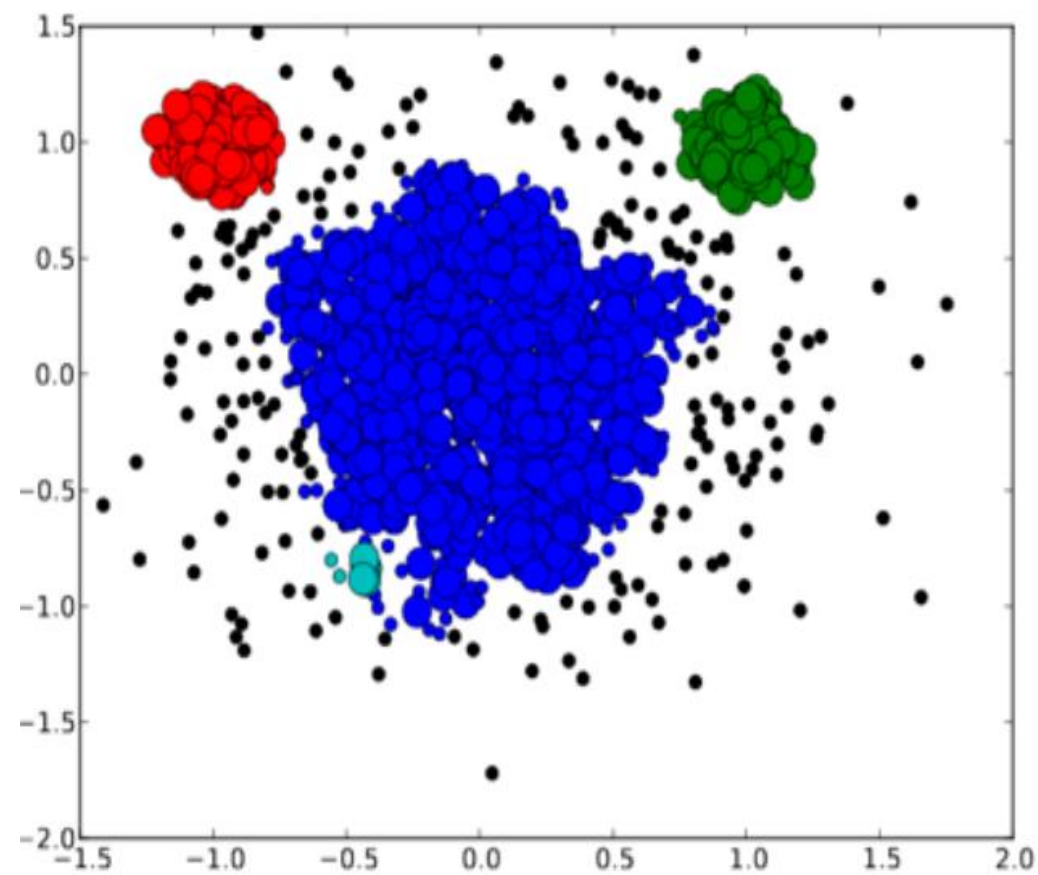
- ✓ Asume que la densidad alrededor de los datos normales es similar a la densidad alrededor de sus vecinos.
- ✓ La densidad alrededor de los valores atípicos es considerablemente diferente a la densidad alrededor de sus vecinos.
- ✓ Hace crecer regiones con suficiente alta densidad en grupos
- ✓ Los grupos que separan las regiones son de baja densidad de objetos (***ruidos***).

DBSCAN

Número estimado de cluster: 2



Número estimado de cluster: 4



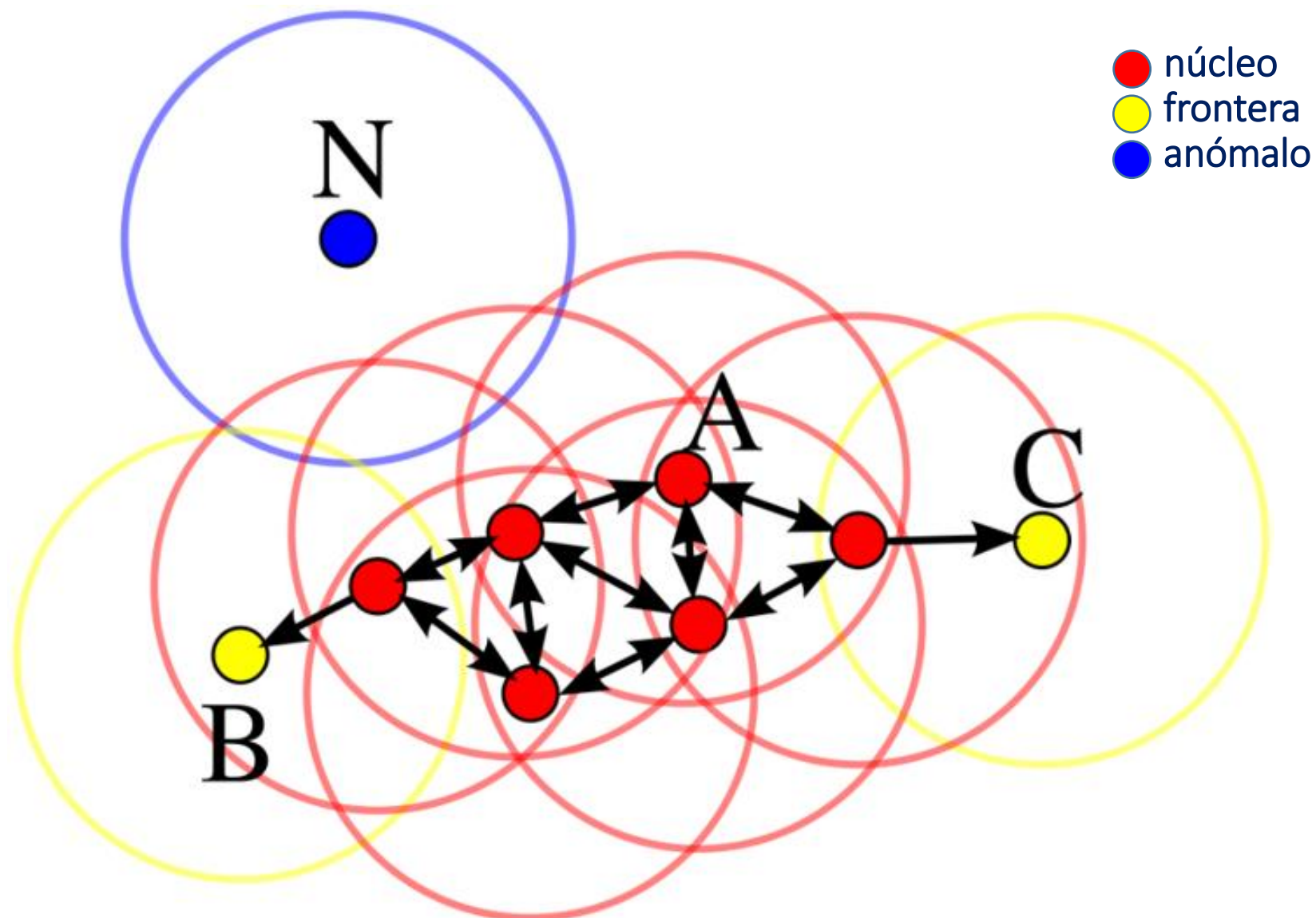
DBSCAN

- ✓ El algoritmo requiere dos parámetros principales:
 - ***El parámetro epsilon (eps)***, define el radio de vecindad alrededor de un punto.
 - ***El número mínimo de puntos (MinPts)*** de vecinos en un radio **eps**.
- ✓ Cualquier punto en el set de datos, con un número mayor o igual que ***MinPts*** se considera un **punto núcleo**.
- ✓ Un punto se considera **punto frontera** si tiene menos de MinPts vecinos pero el es vecino de un punto núcleo.
- ✓ Un punto que no es ni núcleo ni frontera, se considera un **punto de ruido** o valor atípico (outlier).

DBSCAN

¿Cómo funciona?

- Eps: 5u
- Pts min: 3



DBSCAN

✓ **Ventajas:**

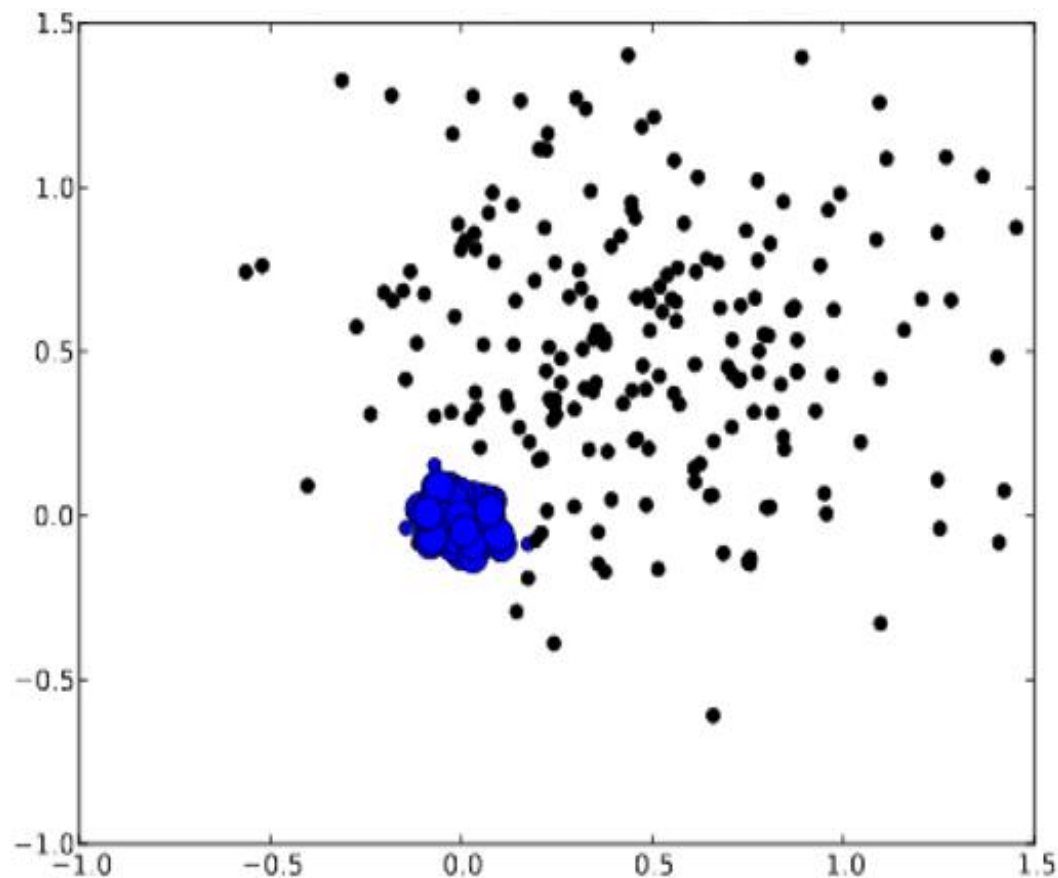
- No necesita asumir un número fijo de clusteres.
- No depende de las condiciones de inicio.

✓ **Desventajas:**

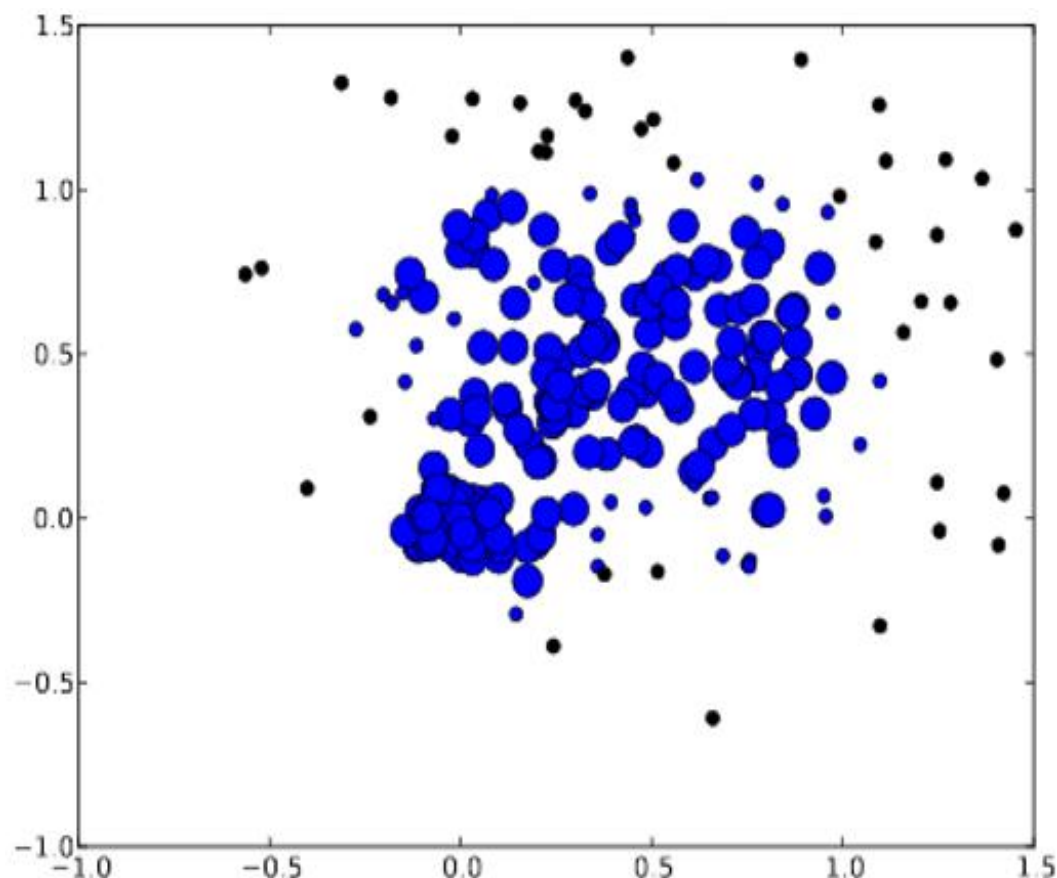
- Asume densidades similares en todos los clusteres.
- Puede tener problemas al separar clusteres.

Para el caso de un cluster muy disperso y otro compacto uno cercano al otro ¿qué pasaría?

Número de cluster estimado: 1

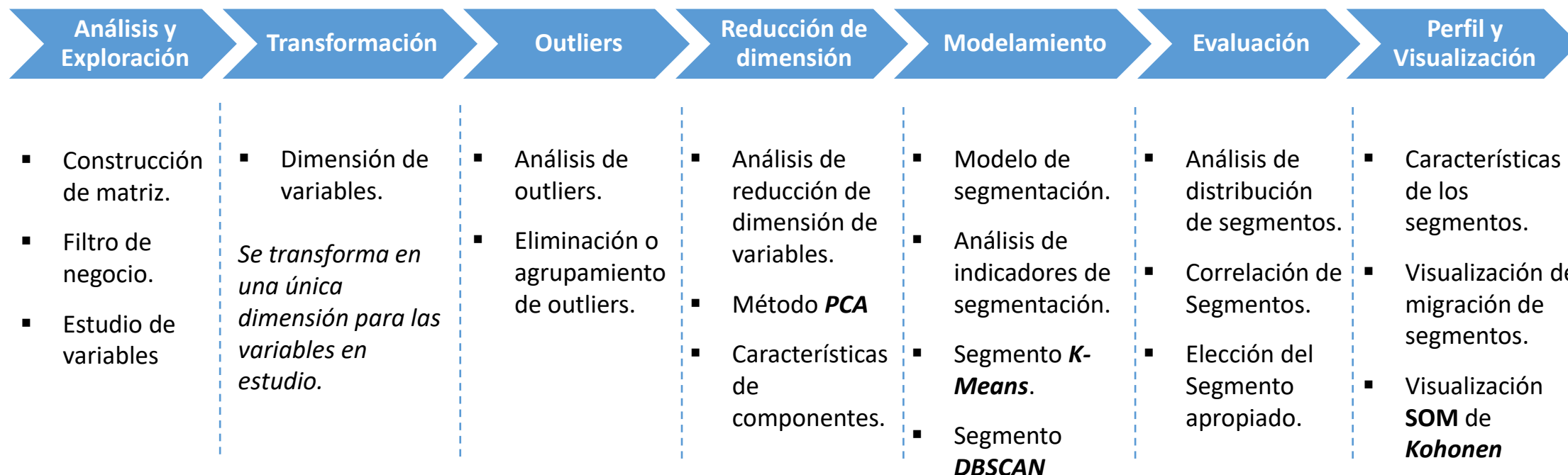


Número de cluster estimado: 1



Metodología

Proceso de Segmentación



Retroalimentación

— PROGRAMA DE —
ESPECIALIZACIÓN ANALÍTICA

ADVANCED DATA SCIENCE
