



IIC2433 Minería de Datos

Guía Tarea 3

Pasos generales :

1. Preprocesamiento de la base de datos (limpieza, elección atributos relevantes).
Separamos por candidatos electos, luego por tipo de elección.
2. Desarrollo algoritmo KMeans usando sólo librerías permitidas.
Para el caso que se ocupe PCA, este puede ser usado desde sklearn.
Todas las demás funciones auxiliares que se usen deben ser implementadas desde cero.
3. Probar con la base de datos lo obtenido con distintos k (por lo menos 2 k diferentes).
Explicar brevemente lo observado.
4. Presentar visualización para cada elección.
 - a) Queremos responder la pregunta sobre cómo se distribuye territorialmente el espectro político.
Una forma sencilla es usar las coordenadas geográficas de cada comuna (ver figura 1).
 - b) Aquí se puede elegir la visualización que más acomode, que sea clara para responder la pregunta en a).
 - c) Pueden agrupar comunas y partidos políticos, según les sea más cómodo trabajar.
 - d) En la página del curso se encuentra base de datos con la ubicación geográfica de cada comuna: *Comunas_Lat-Long.csv*.

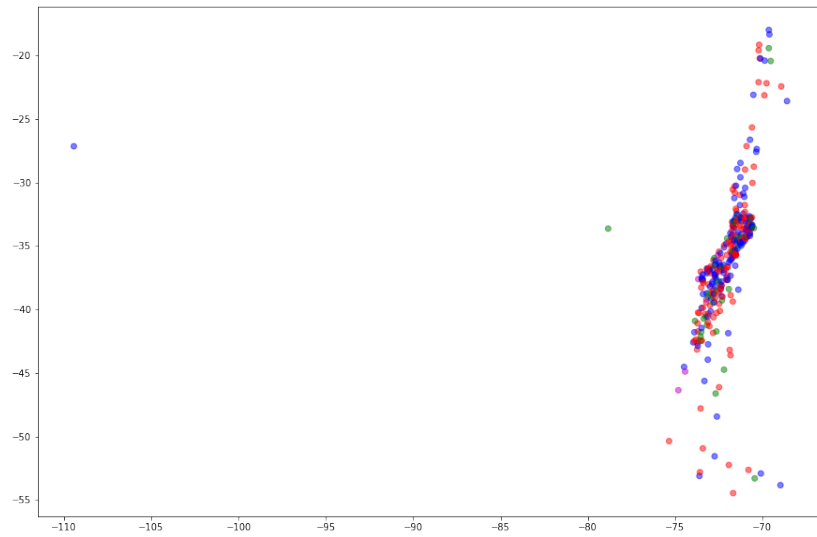


Fig 1. Clusters obtenidos con KMeans $k=4$, basado en la votación total por elección presidencial, normalizada por comuna. Gráfico simple considerando distribución territorial lat/long, usando scatter de matplotlib.
Otra librería útil para graficar mapas es basemap (opcional).