Relevamiento geolocalizado de Techos solares en la Ciudad Autónoma de Buenos Aires.

Alan Maggio

Ciencia de Datos - Ingeniería Industrial UTN BA

**Abstract**

Se realizará un análisis de los techos Solares de la Ciudad de Buenos Aires partiendo de un dataset de acceso libre en el cual se detalla la geolocalización de las instalaciones solares fotovoltaicas y las instalaciones solares térmicas.

**Keywords**

Techo solar, fotovoltaico, térmica, geolocalización.

# Introducción

Se utilizará un dataset de "BA data" en el cual se encuentra el relevamiento geo localizado de las instalaciones solares existentes en la ciudad de Buenos Aires. En dicho dataset se encuentran tanto las instalaciones solares fotovoltaicas como las instalaciones solares térmicas.

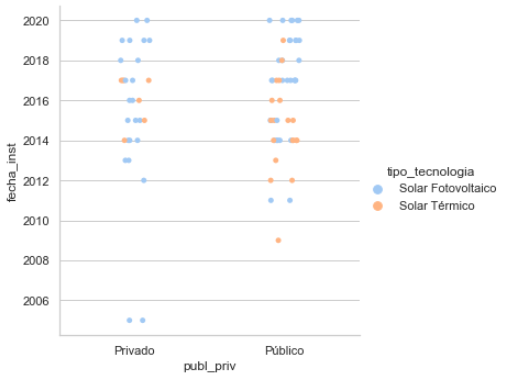
## Análisis complementario

Para poder complementar el análisis del dataset se utilizará la herramienta "Geopandas" la cual trabaja con datos Geoespaciales. Se combinará la geolocalización de los techos disponible en el dataset con mapas de barrios y comunas también a disposición en BA data.

# Analisis exploratorio de datos

## Análisis de tipos de techos

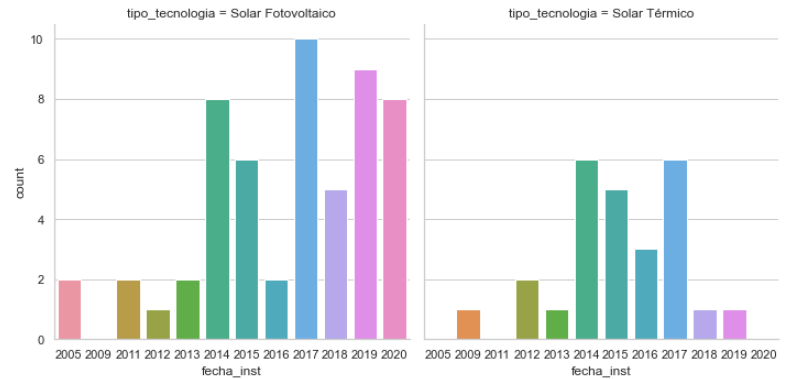
Se analiza los distintos tipos de techos que se encuentran instalados en la ciudad de Buenos aires.



Se puede observar que la mayor cantidad de los techos, como era de esperarse, fueron instalados en los últimos años. Algo que también se puede destacar es que hay más densidad de techos en edificios públicos que en privados.

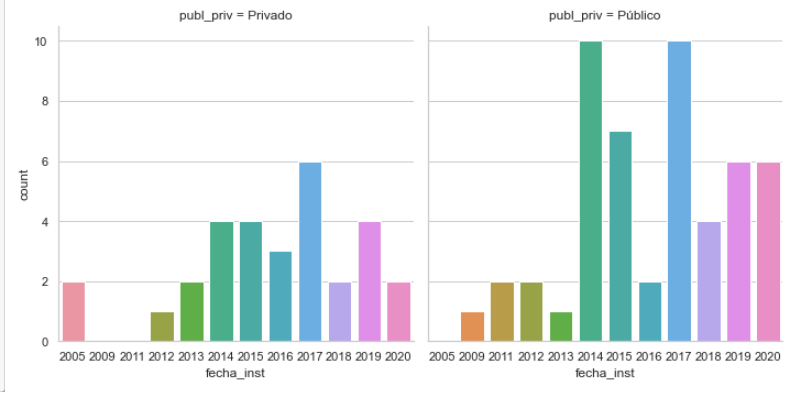
## Análisis por año de instalación

Contabilización de techos por año en que fueron instalados, separado en Solar Fotovoltaico y Solar Térmico.



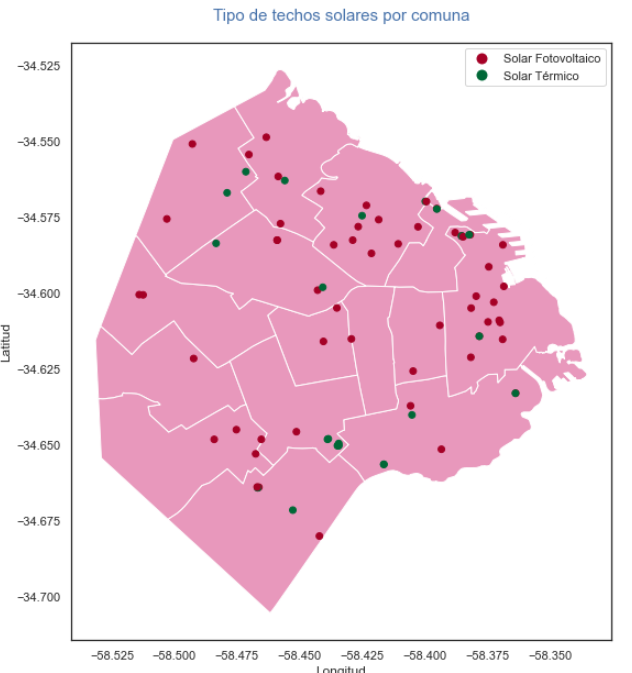
Se puede observar que en la mayoría de los años se instalaron más techos del tipo Solar Fotovoltaico que Solar Térmico. También se puede observar la tendencia a la baja, hasta volverse nula en 2020, la instalación de techos del tipo Solar Térmico

## Sector Privado vs Publico



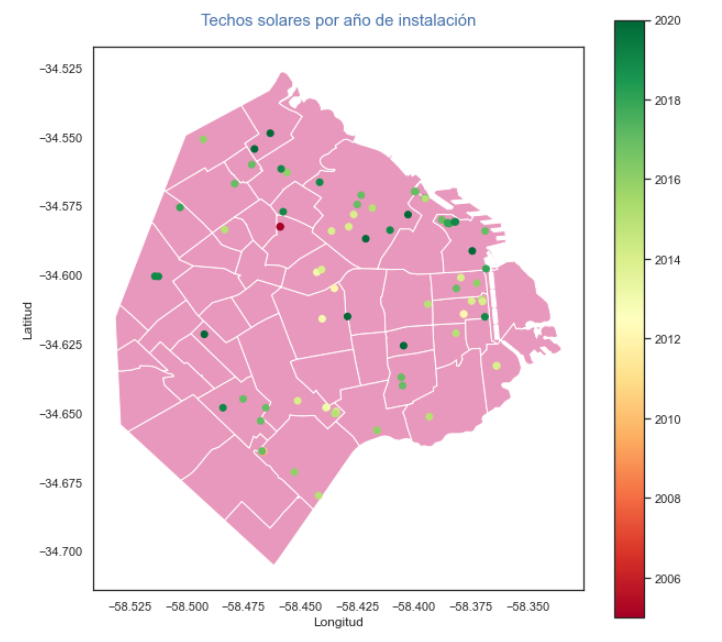
En concordancia con lo analizado en el punto 2.1 en este countplot se puede observar que se desarrolló más la instalación de techos Solares en el sector Público que en el sector Privado.

## Tipo de techos por comuna



Podemos observar el mapa de la ciudad de Buenos Aires dividido por comunas en donde se localizaron la totalidad de los techos solares (térmicos y fotovoltaicos) de la ciudad.

## Avance de la instalación de techos solares



Podemos visualizar como se fue desarrollando la instalación de los techos solares en la ciudad de Buenos Aires, siendo el primero antes del 2006. Se puede observar cómo fueron tomando más relevancia con el correr de los años.

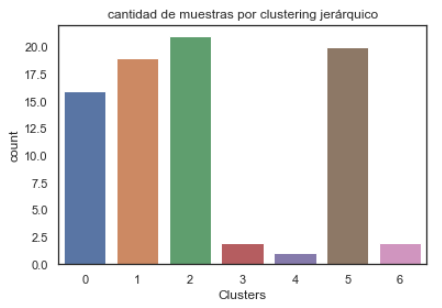
# Aprendizaje no supervisado

Inicialmente realizamos un clustering con los datos de nuestro data set original. Al no obtener un valor de Silhouette Score que nos resulte satisfactorio, sumamos más información al análisis.

## Clustering

Realizamos un clustering por el método “K Mean” y por el método de “Clustering Jerárquico” obteniendo valores de Silhouette Score de 0.34 y 0.33 respectivamente.

El valor óptimo fue obtenido con la formación de 7 cluster distintos.



## Conclusión parcial

Luego de correr el modelo en varias ocasiones realizando la selección de distintos Hiperparametros el mayor resultado de Silhouette Score obtenido fue de 0.34 aproximadamente para ambos modelos. Se considera que dicho valor es muy bajo con lo cual los distintos clústeres generados no están bien definidos o que tienen una baja similaridad respecto a su mismo cluster.

# Aprendizaje no supervisado 2

Como hemos mencionado antes, para intentar mejorar la performance del modelo, se suma otro data set. En este caso se seleccionó uno que contiene la cantidad de población por cada barrio de la ciudad de Buenos Aires.

## Clustering

Al igual que antes, realizamos el clustering por el método “K Mean” y por el método de “Clustering Jerárquico” obteniendo valores de Silhouette Score de 0.61 y 0.60 respectivamente.



Podemos observar los valores de Silhouette obtenidos y concluir que las muestras están aceptablemente definidas. A su vez, también podemos observar que los cluster se encuentran muy dispares, obteniendo solo 1 muestra para el Cluster 1 y 2 muestras para el Cluster 2.

# Conclusion

Concluimos que, con los datos trabajados en este análisis, realizar un proceso de aprendizaje No supervisado en el cual se intenta realizar la segmentación de un conjunto de muestras, no es suficiente.

Para poder continuar con desarrollo del aprendizaje no supervisado y obtener un resultado más contundente, se sugiere agregar más información al análisis, como por ejemplo el de "Calidad de aire" donde se brinda información diaria de los niveles medios de monóxido de carbono, dióxido de nitrógeno y material particulado respirable menor a 10 micrones medidos en las 4 estaciones de monitoreo de Buenos Aires.