Introducción

El presente documento tiene como fin explicar el proceso de transformación que se realiza sobre la información proveniente de Telecom, para poder vincularla con la unidad espacial requerida.

# Tablas

Para poder realizar el proceso trasladar la información de las antenas de Telecom a otra unidad espacial (recomendablemente polígonos), son necesarias las siguientes tablas que están disponibles en la DDBB Telecom con los siguientes nombres:

* **public.dispositivos\_caba**= datos crudos provenientes del set de datos que brinda Telecom. Esta tabla se trunca por cada insert que se genera, por tanto solo tendrá un periodo acotado de fechas.
* **aux.calendario\_2020**= calendario 2020 con los feriados clasificados
* **aux.radios\_caba**= radios de CABA.
* **aux.cuadrado\_150**=Grilla de 150\*150 para CABA. Esta tabla puede ser reemplazada por la unidad espacial requerida. (Ver requerimientos)
* **datos.dispositivos\_por\_grilla** = esta tabla contiene para cada id de grilla, hora y día, la cantidad de dispositivos móviles que se registraron (ver estructura en Anexo). Se alimenta de la ejecución de la consulta **001\_Pasaje\_antenas\_a\_voronois**

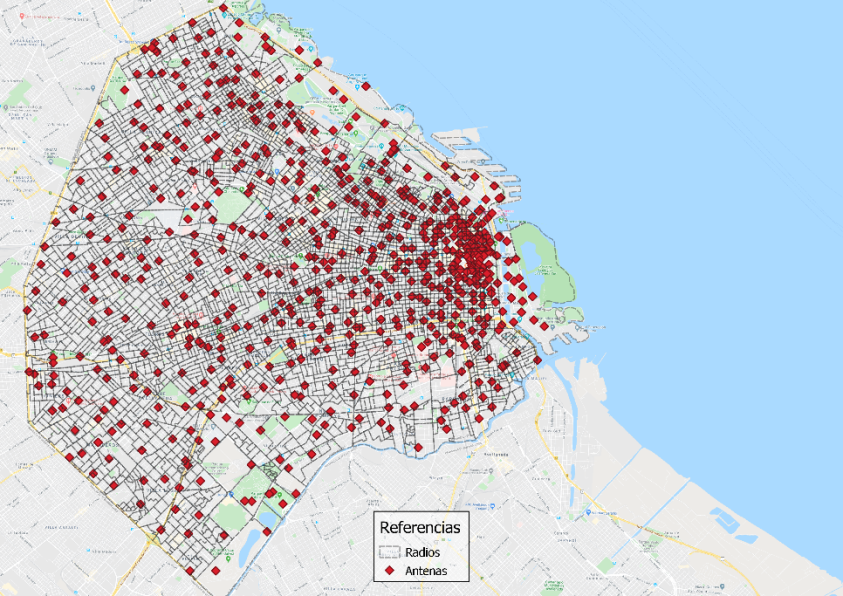
# Repositorio

El repositorio esta disponible en [Github](https://github.com/Florpa/Poligonos-Voronoi-con-Antenas-de-Telecom). La carpeta se encuentra clonada en la VM tambien en la siguiente ruta: F:\repos\_flor\Poligonos-Voronoi-con-Antenas-de-Telecom

# Resumen de la metodología

El procedimiento tiene como fin poder trasladar la información de las Antenas de Telecom (puntos) a otra entidad espacial (polígonos) que permita realizar comparaciones, estimaciones, etc. Para ello se tomó la decisión de construir polígonos de voronois a partir de las Antenas disponibles. (Ver imagen 1).

Estos polígonos tienen como centroide a la antena en cuestión y tomarán todos los valores que se encuentren vinculados a esta última. Es necesario crear polígonos por cada rango horario, puesto que las antenas registran dispositivos de manera intermitente, provocando que en las diferentes horas, no sean siempre las mismas que se encuentran activas. Luego de construir los voronois, se traslada la información a la unidad espacial seleccionada, en base al % de ocupación de la superficie de esta última sobre los polígonos construidos.



**Imagen 1. Polígonos de Voronoi construidos a partir de las antenas de Telecom**

# Consideraciones

La información que provee Telecom contiene los datos de una semana completa (lunes a domingos) y en las franjas horarias de 00:30 a 23:30. En las horas de la madrugada, al registrarse muy pocos dispositivos, ocurre que hay zonas que quedan en 0.3 dispositivos. Esto no es un error, sino que tiene que ver con cómo se distribuye 1 dispositivos registrado en n % de superficies de ocupación.

# Procedimiento

# Carga de datos

La información que provee Telecom, debe descargarse de la Carpeta CABA ([CABA-Datos-Telecom](https://drive.google.com/drive/folders/1pDad_q05I1ne-UAQNlCK65vNXlvunTRI?usp=sharing)), Allí dos veces por semana cargan un set de datos llamado: TRAFICO\_CABA+FECHA.

Habitualmente los días de carga son los Lunes por la tarde, y los viernes por la mañana. Puede ocurrir que se atrasen pero no es usual. Este archivo, es un histórico de los registros, es decir contiene información desde el 01/03/2020, si bien se descarga completo, solo se sumará al set de datos almacenado en Postgres, aquellos días que falten.

Luego de descargar el set de datos, debe almacenarse en la siguiente ruta, con el nombre de TRÁFICO.

storage = 'abfss://movilidad@gcdd.dfs.core.windows.net/ 'telecom/raw-data/'TRAFICO\_Caba\_25oct.txt'

## Carga de datos

Luego que se descargaron los datos y se almacenaron en la ruta indicada, se debe ejecutar el Notebook **001.copy-new-data-to-postgres** (disponible en [Notebook Datos-Telecom](https://adb-188860186850355.15.azuredatabricks.net/?o=188860186850355#notebook/2242103246482473/command/2242103246482480)).

Este notebook recibe como parámetro las fechas de los días deseados.

Lo que hace este notebook es lo siguiente:

* Se conecta al storage, y trae el archivo de Telecom que fue almacenado previamente
* Lee el archivo y lo filtra según los parámetros de fecha que se indiquen de manera manual.
* Recorta el archivo a los campos que se utilizan y realiza las conversiones a los campos pertinentes
* Inserta en el Postgres disponible en Azure los días solicitados. Son alrededor de 40 mil registros por día. La tabla donde se realiza el insert es **tabla datos.dispositivos\_caba.**
* Luego que se insertaron los nuevos registros, comienza la ejecución de scripts.

## Transformación de los datos

Luego de haber insertado los nuevos registros en la **tabla datos.dispositivos\_caba,** lo siguiente por realizar es trasladar la información de las Antenas a polígonos de Voronoi primero, y luego trasladar eso a las grillas de 150\*150. Para ello se requiere que se ejecute el script : **001\_Pasaje\_antenas\_a\_voronois.**

**#000 Agrego la columna Geom.** Esta consulta agrega la columna geom a la tabla datos.dispositivos\_caba.

**#001 Index a las columnas:** Esta consulta indeza las columnas date, hora y geom de la tabla datos.dispositivos\_caba.Es necesario indexar las columnas ya que son demasiados registros y las consultas (principalmente los cruces especiales) demoran.

**#002 Construyo la geometría de las antenas:** en esta consulta se construye la geometría de las antenas de Telecom a partir de los lat/long disponibles.

**#003 Consulta de fechas:** Esta consulta compara fechas. Consulta en la tabla **datos.dispositivos por grilla** por la última fecha, y la compara con las fechas disponibles en la tabla **datos.dispositivos\_caba**. Se almacenan en el **objeto fechas** todas aquellas que no se encuentren procesadas. Para las fechas almacenadas, también almacena las horas.

**#004 Iteración por dia y hora. Construcción de voronoi y pasaje a las grillas. Insert de resultados:**

Para cada dia y hora almacenados en las variables se ejecuta las siguientes consultas dentro de un with:

**Antenas:** Dentro de esta consulta se seleccionan de la tabla **datos.dispositivos\_caba** los siguientes campos: hora,dia, cantidad de dispositivos y el campo geom. A la cantidad de dispositivos la multiplica por 3\*3 para extrapolar la muestra al total de la población. A la geometría le cambia el sistema de proyección para poder realizar a posterior transformaciones y cálculos de superficies.Posee un where de fecha y hora donde se itera según lo almacenado.

**Calendario**: En esta consulta se trae el calendario construido para el 2020.

**Voro:** en esta consulta se construyen los voronois a partir de las geometrías de las antenas.

***ST\_VoronoiPolygons*** calcula un diagrama de Voronoi bidimensional a partir de los vértices de la geometría proporcionada (puntos). El resultado es una GeometryCollection por tanto es necesario aplicarle un st\_dump para simplificar el resultado del geom obtenido.

**CABA:** En este bloque de la consulta se traen los radios de caba, se tranforma la geometria al sistema de proyeccion 5347.Se anida el St\_transform con un St\_union porque solo se requieren el polígono de la Ciudad.

**Vorointer:** En esta consulta se recortan los polígonos Voronoi obtenidos anteriormente con el límite de caba. Esto es así dado que en los limites de CABA, los polígonos generados tienden a excederse. Este recorte se realiza realizando un ST\_Intersection. Por último se calculan las superficies de los polígonos recortados.

**Fracción:** en esta consulta se trae la unidad espacial de destino. En este caso se utilizaron grillas de 150\*150. Pero se han realizado para otros usos el traslado de la información a radios censales o fracciones. Lo importante es que la unidad espacial seleccionada debe tener SI O SI dos campos: un ID único que identifique a las geometrías y un campo geom en proyección 5347.

**Vorofrac:** en esta consulta se realiza la intersección entre la geometría de los polígonos de Voronoi y la geometría de las grillas. Se calculan las superficies para cada una de las intersecciones.

**Combi:** calculo la superficie ocupada de un polígono dentro de otro y sus superficies totales.

**Presalida:** en esta consulta uno las geometrías de las grillas por id, ya que cuando se generaron las intersecciones se particionaron. Sumo la cantidad de dispositivos que le corresponde a cada parte de la grilla particionada.

**Salida:** Esta consulta trae los campos que deseo insertar en mi tabla destino. En el caso de ejemplo la tabla donde se realizará el insert de los resultados es **datos.dispositivos\_por\_grilla.**

## Consideraciones Finales

Todas las tablas que usa el script **001\_Pasaje\_antenas\_a\_voronois** se encuentran disponibles en la DDBB Telecom. El mismo fue pensado para tomar, procesar, insertar los datos de Telecom utilizando como unidad espacial las Grilla DE 150\*150. Es importante que en caso de cambiar la unidad espacial, se realicen los cambios pertinentes en los lugares ya mencionados y los resultados se inserten en otra tabla destino.

La tabla **datos.dispositivos\_por\_grilla,** no posee geometría. El campo fracción contiene el identificador de las geometrías implicadas. Con este campo se pueden unir los resultados con la geometría pertinente.

# 