Introducción

El presente documento tiene como función ser una guía detallada para poder llevar adelante la actualización del tablero de DGCINFO-Regreso-Espacio-Público que es utilizado por Regeneración y Antropología Urbana.

Resumen de la Metodología

El tablero de Regreso Espacio al Público tiene como fin brindar información sobre la cantidad de personas que se encuentran en las diferentes zonas de la ciudad. Para ello utiliza la información proveniente de las antenas de Telecom, la cual contiene por antena, la cantidad de dispositivos que se registran conectados, en intervalos de 30 min.

Para poder trasladar la información que registran las antenas a la grilla de 150\*150, es necesario realizar transformaciones y procesamientos en Postgres. Finalizados los diferentes procesos, el resultado será lo que se utilice para actualizar el set de datos vinculado al tablero.

# Tablas

Para poder llevar adelante la actualización del tablero es necesario un conjunto de tablas, que se encuentran disponibles en el la DDBB Telecom, dentro del servicio AZURE.

Las tablas son:

* **public.dispositivos\_caba**= datos crudos provenientes del set de datos que brinda Telecom. También es el input para generar los datos que se insertaran en **datos.dispositivos\_por\_grilla.**
* **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**
* **aux.calendario 2020**= calendario 2020 con los feriados clasificados
* **aux.radios\_caba**= radios de CABA.
* **tablero\_regeneracion\_urbana.area\_interes\_grilla=** Contiene las zonas de interés que figuran en el tablero (Centros de Trasbordo, Corredores comerciales,etc)
* **aux.cuadrado\_150**=Grilla de 150\*150 para CABA.
* **tablero\_regeneracion\_urbana.data** =Tabla input para el tablero. Se actualiza con los datos de **datos.dispositivos\_por\_grilla.**

# 

# Repositorio

El repositorio esta disponible en [Github](https://github.com/Florpa/Telecom--Actualizacion-de-tableros-Power-Bi). La carpeta se encuentra clonada en la VM tambien en la siguiente ruta:

**F:\repos\_flor\Telecom--Actualizacion-de-tableros-Power-Bi\Consultas para actualizar Tablero de Regeneracion**

# 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.

El procesamiento de esta información se hace para el conjunto de datos completo, pero para actualizar el tablero se definieron los siguientes criterios:

**Semana hábil:** Lunes a Jueves

**Fin de semanas y feriados**: Viernes a Domingo

**Franjas Horarias:**

* Mediodía: 11 a 15
* Tarde: 15:30 a 17:00
* Noche: 17:30 a 23:00

Las zonas de interés que se utilizan para actualizar la información de este tablero no son las mismas que se utilizan para el tablero de Concentración de Personas.

Este tablero se actualiza los Lunes por la tarde en la medida de lo posible. Sin embargo, no es tan prioritario como el Tablero de Concentración de Personas que utiliza Telecom.

# Metodología

## Descarga 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.

Tener en cuenta que solo hay que sumar la información del fin de semana anterior, dado que para actualizar el tablero de Concentración de Personas ya se han cargado los datos de la semana anterior.

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.**

**Lo que el script ejecuta es lo siguiente:**

**#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:** En este gran bloque se realizan los siguientes procesos:

Para cada dia y hora identificados y almacenados en pasos anteriores se multiplican la cantidad de dispositivos registrados \*3.3. Luego con la geometría de las antenas se construyen los polígonos de voronois, y se le trasladan todos los atributos que la primera posee. Luego se trasladan los valores de los polígonos de voronois a la unidad espacial elegida , grillas de 150\*150 en este caso, según la superficie de ocupación de estas últimas sobre los polígonos construidos. **(para más detalle ver** [**Documentación Traspaso Antenas Telecom a Polígonos**](https://drive.google.com/file/d/1iAmcL1ftofHqyyDL7x4hQAVLs3JquGSy/view?usp=sharing)**)**. Luego de ejecutarse el traspaso, se insertan en **datos.dispositivos\_por\_grilla** los resultados.

## Insert y actualización de resultados

Luego que terminó de ejecutarse el script **001\_Pasaje\_antenas\_a\_voronois,** lo siguiente que resta hacer es ejecutar el script **001\_ Inserts\_ updates**

Este es el último script a ejecutar. El mismo toma los datos generados por la consulta anterior y realiza los siguientes procesos dentro de la consulta:

**#001Insert y actualizacion para la semana durante la franja del mediodía:** dentro de este bloque se insertan y actualizan los valores de los dispositivos registrados en la semana, durante la franja del mediodía, en la tabla **tablero\_regeneracion\_urbana.data\_**

**#1A Actualización de registros para el fin de semana, franja del mediodía**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de semana, en la franja del mediodía.

**#1b Actualización de registros con los valores de la semana anterior**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante la semana anterior en la franja del mediodía.

**#1c Actualización de registros con los valores del fin de semana anterior**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de anterior en la franja del mediodía.

**#02 Insert y actualización para la semana durante la franja de la tarde:** dentro de este bloque se insertan y actualizan los valores de los dispositivos registrados en la semana, durante la franja de la tarde, en la tabla **tablero\_regeneracion\_urbana.data\_**

**#2A Actualización de registros para el fin de semana, franja de la tarde**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de semana, en la franja de la tarde.

**#2b Actualización de registros con los valores de la semana anterior**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante la semana anterior en la franja de la tarde.

**#2c Actualización de registros con los valores del fin de semana anterior**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de anterior en la franja de la tarde.

**#03 Insert y actualización para la semana durante la franja de la noche:** dentro de este bloque se insertan y actualizan los valores de los dispositivos registrados en la semana, durante la franja de la noche, en la tabla **tablero\_regeneracion\_urbana.data\_**

**#3A Actualización de registros para el fin de semana, franja de la noche**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de semana, en la franja de la noche.

**#3b Actualización de registros con los valores de la semana noche**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante la semana anterior en la franja de la noche.

**#3c Actualización de registros con los valores del fin de semana noche**: dentro de este bloque se actualizan los valores de los dispositivos registrados durante el fin de anterior en la franja de la noche.

**#04 Actualizo los valores de la población de esa semana para las 3 franjas horarias de la semana.**

**#05 ACTUALIZAR LOS CAMPOS DE LA SEMANA PRE CUARENTENA**: en este bloques de consultas se actualizan para cada franja horaria, distinguiendo por semana y fin de semana la cantidad de población que estuvo en la ciudad.

**#5A Población de la precuarentena:** en esta consulta se actualizan los valores para las tres franjas horarias de la población que hubo en la ciudad antes de la cuarentena.

#**5B Actualizo los valores que se registraron durante la semana y el fin de semana para la franja horaria DEL MEDIODÍA**

#**5C Actualizo los valores que se registraron durante la semana y el fin de semana para la franja horaria DE LA TARDE**

#**5D Actualizo los valores que se registraron durante la semana y el fin de semana para la franja horaria DE LA NOCHE.**

**#006 Update de fechas: en este bloque se actualizan las fechas y el periodo de la semana comprendida en la actualización.**

**#07 Insert Final:** en esta consulta se realiza el insert a la tabla que alimenta el tablero de power bi con los valores ya actualizados.

## Consideraciones Finales

* Los resultados obtenidos de la ejecución del script **001\_Pasaje\_antenas\_a\_voronois,** son utilizados también para alimentar el tablero de Regeneracion Urbana.
* No es habitual, pero ha ocurrido que los sets de datos de Telecom que envían vienen con menos registros o con una distorsión en la cantidad de dispositivos que registran. Por lo general, lo arreglan en cuestión de días.
* si se quieren modificar los rangos horarios, o los días que competen a este tablero es en este script  **002\_ Inserts\_ updates donde se deben cambiar los parámetros.**
* El calendario que se utiliza en este tablero comprende desde Marzo (inicio de pandemia y del set de datos) hasta Diciembre 2020. En caso de prolongarse luego de este periodo deben insertarse los registros.
* El tablero se actualiza todos los lunes pasado el mediodía.
* Todas las tablas que se utilizan en este proceso se encuentran indexadas. Esto es así porque dada la complejidad de las consultas espaciales y el caudal de información, es necesario para alivianar la carga de procesamiento que se indexen.