Introducción

El presente documento tiene como función ser una guía detallada para poder llevar adelante la actualización del tablero de Concentración de Personas que se realiza todos los viernes por la tarde.

Resumen de la Metodología

El tablero de Concentración de Personas 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.
* **Aux.area\_interes\_grilla=** Contiene las zonas de interés que figuran en el tablero (Barrios vulnerables, Centros de Trasbordo, Corredores comerciales,etc)
* **aux.cuadrado\_150**=Grilla de 150\*150 para CABA.
* **datos.tablero\_transporte**=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 Transporte

# 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 solo se utilizan los días hábiles de lunes a viernes en los horarios de 10 a 17 hrs. La actualización se efectuará los días Viernes con los valores registrados hasta el Jueves de la semana corriente inclusive.

# 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.

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 Poligonos**](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 **002\_ Inserts\_ updates.**

Este es el último script a ejecutar. El mismo toma los datos generados por la consulta anterior y realiza los siguientes procesos en distintas consultas a enumerar:

**#0001 primer insert:** en este bloque se insertan en la tabla datos.tablero\_transporte\_ , el valor de la mediana de los dispositivos calculados para la Ciudad,para cada entidad de la grilla. Los parámetros que recibe es que los días a insertar no sean fin de semana o feriados y solo en la franja horaria de 10 a 17.

**#002 updates semana Pre Cuarentena**: en este bloque se actualizan los campos relacionados con la precuarentena (n° de semana, cantidad de dispositivos conectados). Dicho periodo se estableció que sea la semana n° 2, que comprende los días del 09/03 al 15/03. Este periodo sirve para comparar cómo incrementó o disminuyo la cantidad de población actual en relación a la semana anterior al establecimiento del aislamiento social preventivo y obligatorio.

**#003 Updates semana anterior:** en este bloque se actualizan los campos relacionados a la semana anterior a la actual. Los datos que se actualizan son el número de la semana y la cantidad de dispositivos que se registraron para la semana en cuestión.

**#004 Update de población:** En este bloque se actualiza el campo de “Población”. La misma corresponde a la mediana de los dispositivos registrada para esta semana, pero en el horario de las 4 am. Se tomó esta decisión para establecer relaciones entre la cantidad de personas que “duermen” en la Ciudad, para luego comparar en otros rangos horarios cuántas personas ingresaron a la Ciudad.

**#005 Update de fechas y valores en null:** Actualizo los valores en null para que no me rompan la visualización o me generen valores inválidos.

**#006 Insert Final a la Tabla que alimenta el tablero**: Este bloque inserta todos los resultados hasta acá generados en la tabla **datos.tablero\_transporte. Además** de insertarse los resultados de los updates anteriores, a cada lugar de la grilla se le agrega la información disponible en la **tabla aux.area\_interes\_grilla**. El join entre las tablas permite identificar las zonas de interés sobre las que se realiza un seguimiento particular y la cantidad de personas que estuvo presente allí para la semana en cuestión.

## 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 viernes pasado el mediodía. Esto provoca que la información solo incluya los días de L a J.
* 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.

## Cuestiones que quedaron Obsoletas

## 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.

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

('F:/repos\_flor/Actualizacion-Tableros/querys\_tablero\_telecom/query\_tablero\_telecom\_new/import\_to\_postgres')

## Carga de datos

Luego que se descargaron los datos y se almacenaron en la ruta indicada, se ejecuta el Script 000\_cargadedatos\_postgres . El mismo está en la siguiente ruta:

F:\repos\_flor\Actualizacion-Tableros\querys\_tablero\_telecom\query\_tablero\_telecom\_new\000\_cargadedatos\_postgres