DOCUMENTACIÓN PROYECTO FINAL

### HECHO PARA

Coderhouse

Data Analytics, camada 16215

### HECHO POR

Luis Iannarelli

Ailen Ramirez

CONTENIDO

# 1. Descripción temática de datos

# 2. Diagrama entidad relación de las tablas seleccionadas

# 3. Descripción de tablas y tipo de datos

# 4. Modelo relacional en Power BI

# 5. Transformación de datos

# 6. Análisis funcional del tablero

# 

# 

### 

### DESCRIPCIÓN TEMÁTICA DE DATOS

La temática del proyecto se propone analizar la variación de la movilidad afectada por la pandemia en el caso Argentino. Tomando como punto de partida la fecha del 20 de marzo, cuando se dio inicio a la implementación de la cuarentena en el Área Metropolitana de Buenos Aire (AMBA) y mostrando la variación de actividad con respecto a un día normal, previo a la implementación de el Aislamiento Social, Preventivo y Obligatorio (ASPO).  
El criterio sanitario ha sido la implementación del “aislamiento” el cual se vio reflejado en la movilidad de las sociedades, salvo en el caso de los trabajadores esenciales a los cuales se desalentaba a utilizar medios de transporte masivos, promoviendo la movilidad del transporte privado.   
El objetivo de este trabajo es identificar las distintas fuentes de datos sobre movilidad y

transporte para el AMBA. Para ello se utilizaron los datos disponibles en la web del Gobierno de la Ciudad 1 en conjunto con el Informe de Movilidad Local sobre el COVID-19 de Google 2.

Estos datos están compuestos por 3 datasets: el primero referido a el *flujo vehicular* el cual presenta un listado con información de flujo vehicular detectado por los sensores de la Ciudad. Incluye latitud y longitud del sensor, fecha, hora y cantidad registrada.  
El segundo, presenta el *flujo vehicular captado por radares* desagregado por hora en 2020. Incluye fecha, hora, nombre de la autopista y ubicación del radar.

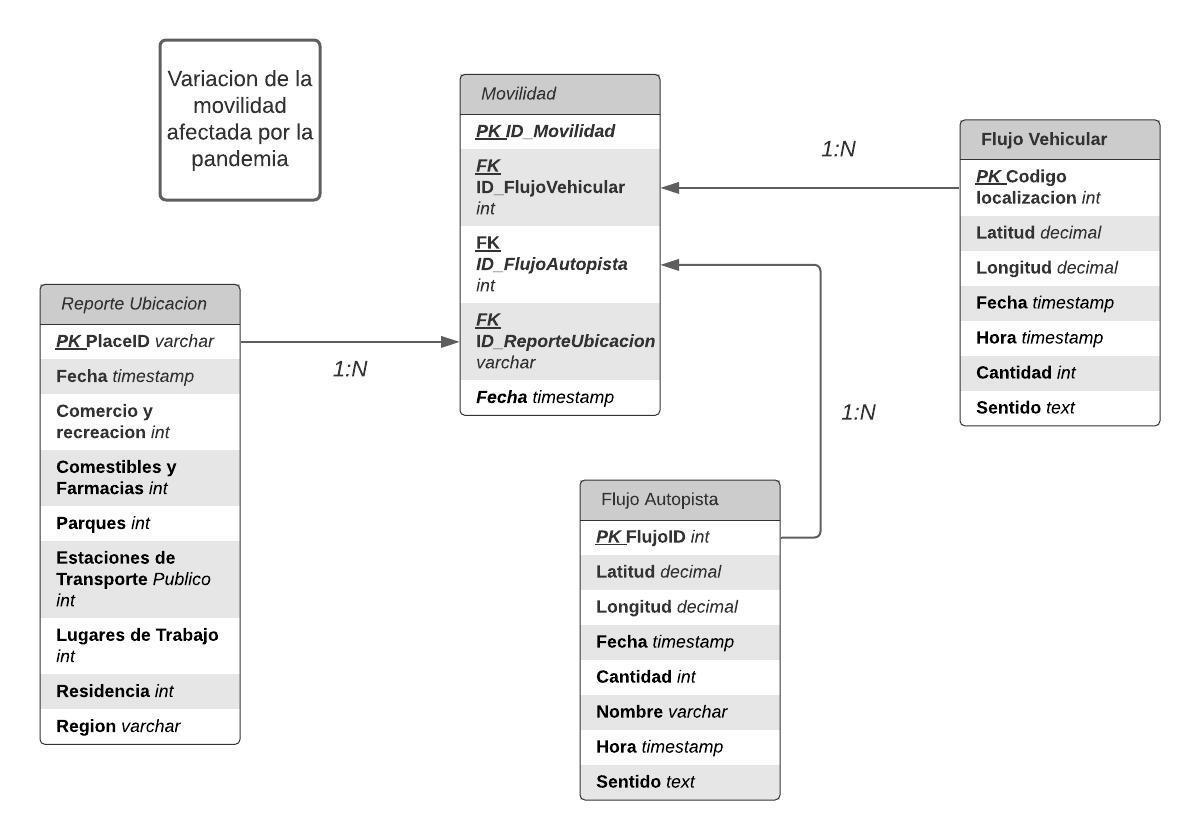
El tercero, presenta *los reportes de ubicación* y muestran las ubicaciones tomadas por el gps de Google ordenadas por zonas geográficas y clasificadas en diversas categorías de lugares, como tiendas y espacios de ocio, supermercados y farmacias, parques, estaciones de transporte, lugares de trabajo y zonas residenciales.

El resultado de estas 3 documentaciones, nos permitirá trazar la movilidad social en

AMBA.

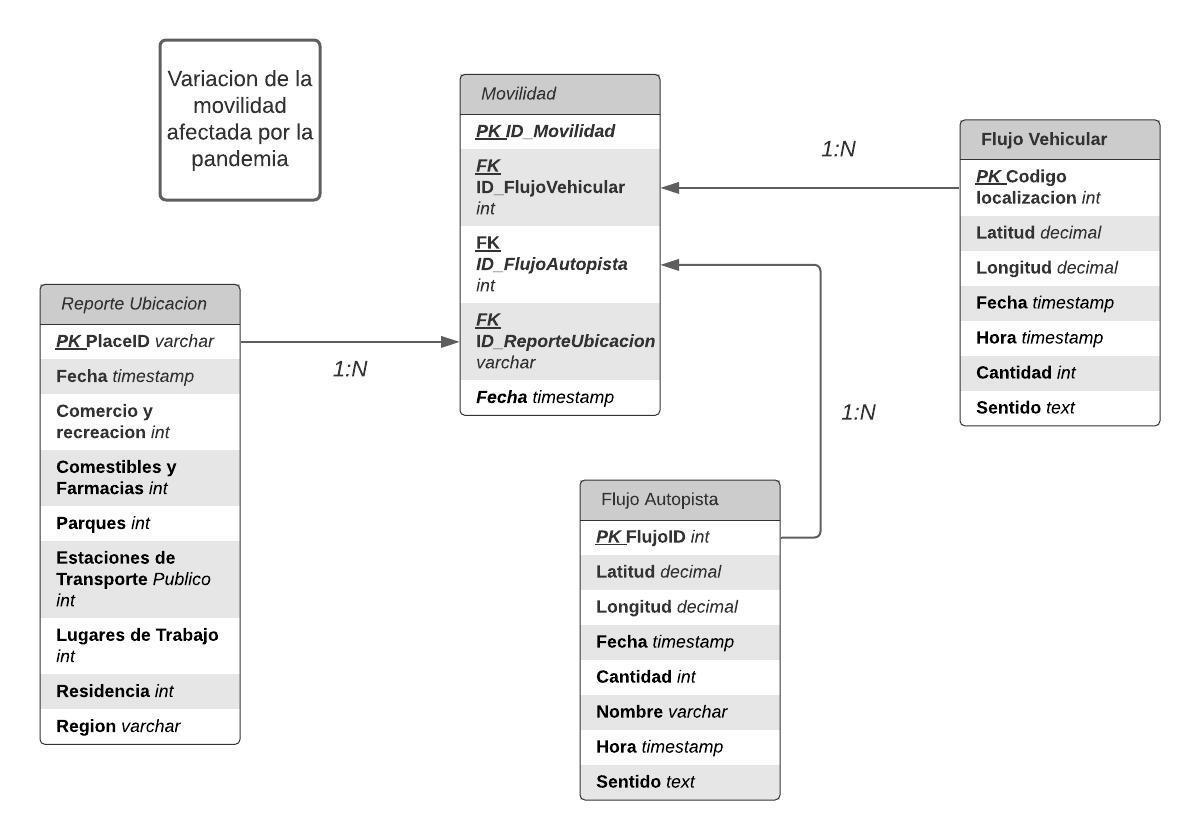
1 https://www.google.com/covid19/mobility/?hl=es 2 <https://data.buenosaires.gob.ar/dataset/?q=transporte>

### DIAGRAMA ENTIDAD-RELACIÓN DE LAS TABLAS SELECCIONADAS

****

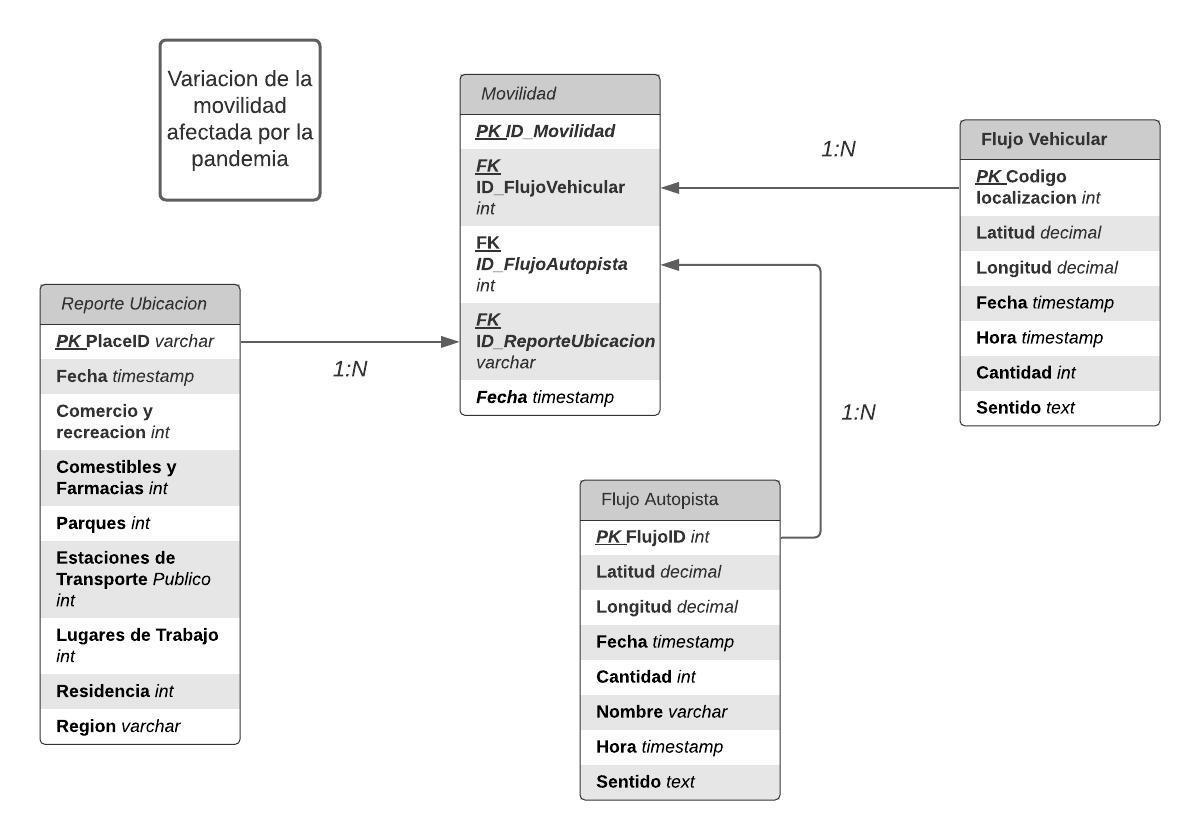
### 

### DESCRIPCIÓN DE TABLAS Y TIPOS DE DATOS



El *reporte de ubicación* muestra la información que se crean a partir de conjuntos de datos anónimos y agregados de los usuarios que activaron el Historial de ubicaciones de Google ordenadas por zonas geográficas y clasificadas en diversas categorías de lugares, como tiendas y espacios de ocio, supermercados y farmacias, parques, estaciones de transporte, lugares de trabajo y zonas residenciales.

Los campos de la tabla está compuesta por su clave primaria “PlaceID” tipo de dato *varchar,* “Fecha” tipo de dato *timestamp,* “Comercio y recreación”, *“*Comestibles y farmacias”, *“*Parques“, *“*Estaciones de transporte público”, “Lugares de trabajo”, “Residencia” y “Región” todos estos con tipo de dato *int.*



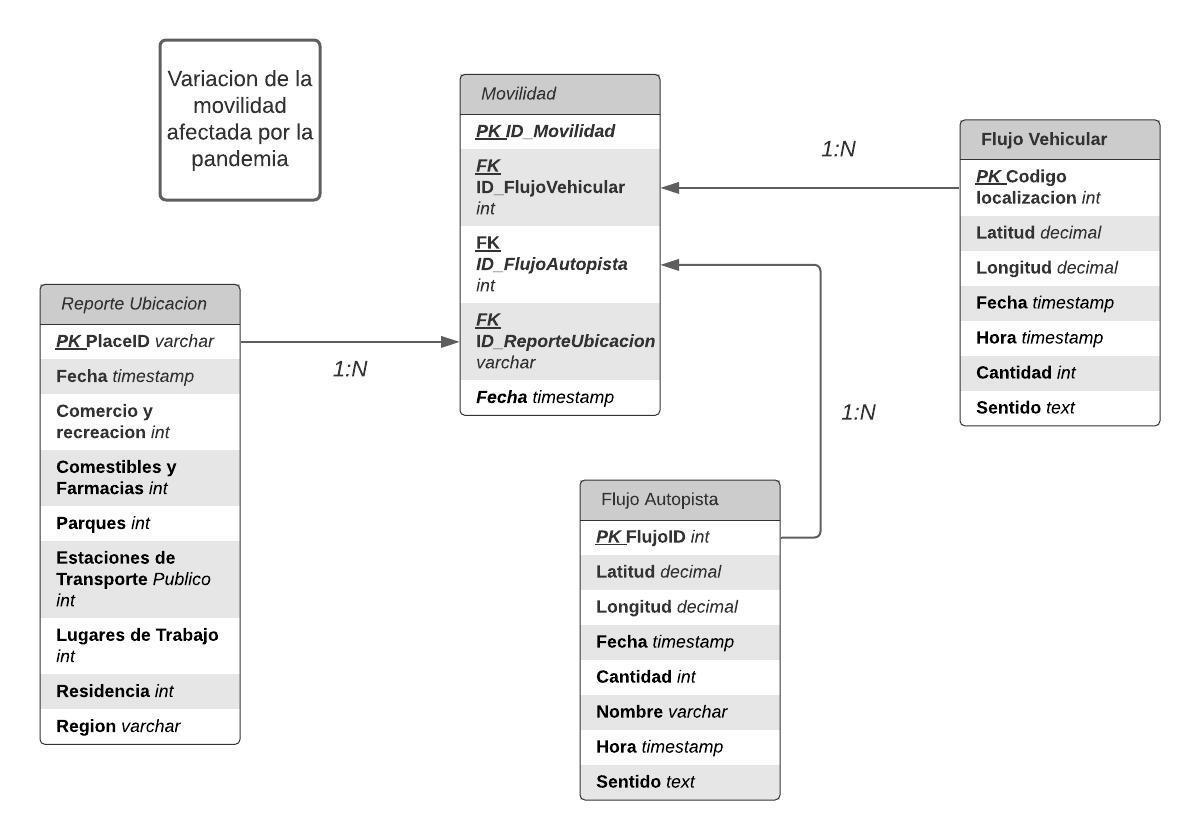
La tabla *flujo autopista* muestra el flujo vehicular captado por radarescon información desagregada por hora en 2020.

Las claves primarias fueron asignadas en función de la ubicación de cada muestra de datos.

Las muestras que tienen la misma ubicación dentro de una comuna se agruparon dentro de un mismo ID.

Incluye “fecha y hora” tipo de dato *timestamp*,

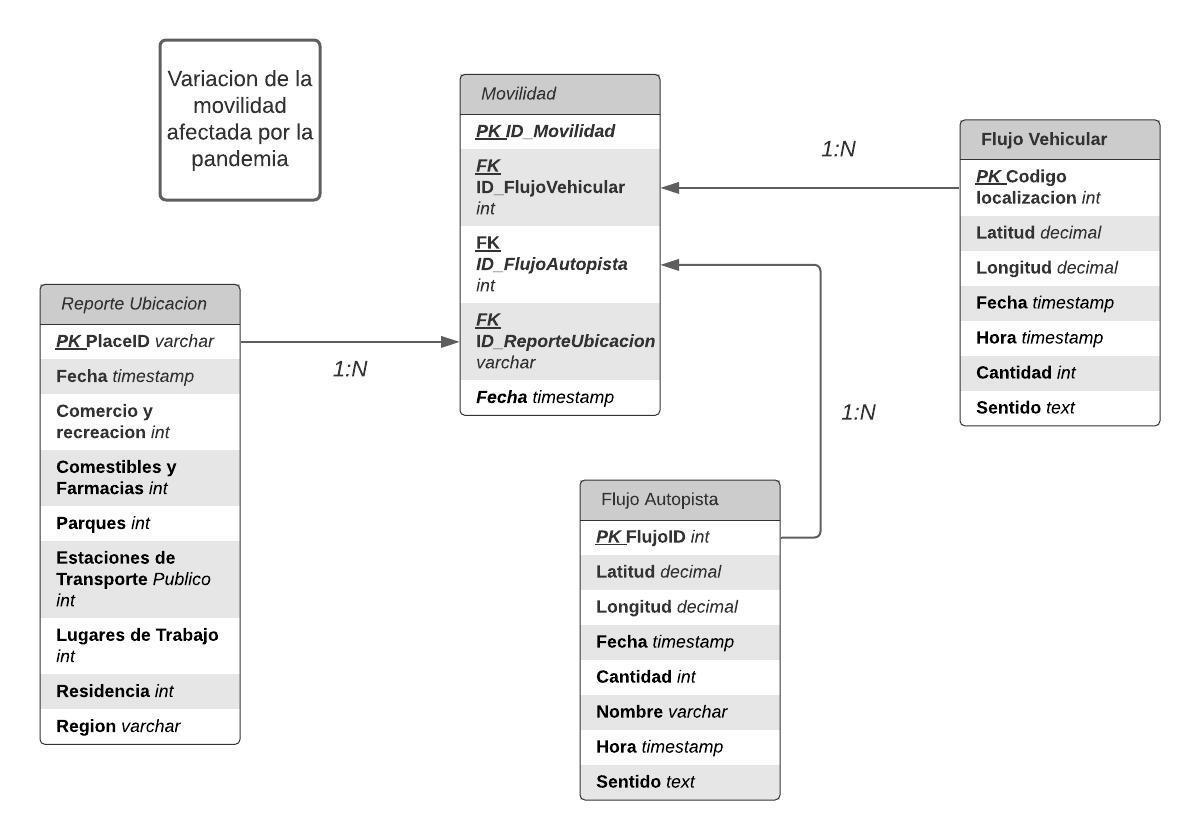
“nombre de la autopista” tipo de dato *varchar* y “Latitud y Longitud” tipo de dato *decimal, “*Sentido” tipo de dato *text*.



La tabla *flujo vehicular* presenta un listado con información de flujo vehicular detectado por los sensores de la Ciudad.  
Las claves primarias fueron asignadas en función de la ubicación de cada muestra de datos.

Las muestras que tienen la misma ubicación dentro de una comuna se agruparon dentro de un mismo ID

Incluye “latitud y longitud“ del sensor tipo de dato *decimal*, “fecha y hora“ tipo de dato *timestamp,* “cantidad registrada” tipo de dato *int* y “sentido” con el que ingresaba o egresaba el automóvil, tipo de dato *text.*



La tabla *Movilidad* contiene la información de las tres tablas anteriores, de esta forma, refiere a el grupo de columnas de aquellas, con sus respectivas *Foreign keys, “*IdFlujoVehicular” y “IdFlujoAutopista” tipo de dato *int,* “IdReporteUbicacion” tipo de dato *varchar .* A su vez, su *Primary key* será “IDMovilidad” tipo de dato *decimal.* Y contiene la

“fecha” de cada grupo tipo de dato *timestamp.*

### MODELO RELACIONAL EN POWER BI

Las fuentes de origen de datos utilizadas para conectar Power Bi son bases de datos de archivo plano (xlsx) mencionadas anteriormente. Incluyen datos de Flujo Vehicular (Gobierno de la Ciudad), Flujo Vehicular por radar(Gobierno de la Ciudad) y Reportes de ubicación (Google).   
A la base Flujo Vehicular por radar se le hizo una conversión de datos de la columna “fecha” que contenían valores erróneos en algunas de sus filas (144237 a 150188) a tipo de fecha que se igualara con sus otros valores de la columna. De esta forma, se logró que Power Bi deje de presentar errores en la totalidad de la columna “fecha”.

**TRANSFORMACIÓN DE DATOS**

Generamos el modelo relacional en Power Bi agregando una *tabla fáctica* y las *tablas puentes* necesarias para la correcta vinculación entre ellas y las tablas utilizadas para el proyecto.

Para crear la tabla fáctica partimos de la tabla flujo vehicular. Duplicamos esta tabla y la renombramos como tabla “Movilidad”. El campo fecha y hora se transformó a fecha. Aplicamos un filtro para fecha y ID. Posteriormente eliminamos los elementos duplicados. Para agregar los IDs de las tablas restantes se combinaron las consultas por fecha y se visualizaron los IDs.

Quitamos duplicados seleccionando todos los campos para asegurarnos que todos los registros sean únicos.

Al hacer la vinculación de los ID de la *tabla fáctica* con sus correspondientes tablas nos generó una cardinalidad de “varios a varios”.   
  
Para que esto no suceda a fin de evitar la duplicidad de datos, creamos *tablas puente*. Duplicando cada una de las tablas de nuestro proyecto (excepto la tabla “Movilidad”). Cambiando así los nombres y quedándonos con las columnas que contienen únicamente los ID. En esta instancia, quitamos también posibles duplicados.  
Tabla Base2020ReporteMovilidadAR corresponde a “Puente Reporte”  
Tabla Flujo vehicular corresponde a “Puente Flujo”  
Tabla Flujo vehicular por radar corresponde a “Puente Radar”  
Filtramos los datos nulos de las 3 *tablas puente.*

Manualmente generamos las relaciones entre las tablas puente, la tabla fáctica y todas las tablas de nuestro proyecto y generamos el DER.

