

Proyecto final

Iván Arriola, Federico Miquelerena, Damián Rovetta

12-07-2023

Introducción

Esto es un análisis descriptivo de los datos del tráfico de Montevideo, Uruguay. Hemos tomado los registros desde enero de 2021 hasta mayo de 2023 y nuestro interés es saber el comportamiento de la velocidad y el volumen de tráfico (variables explicativas) dependiendo de varias variables que iremos desarrollando a lo largo de la investigación.

Datos

Descripción general de los datos

Todos los datos fueron sacados de Catalogo de Datos Abiertos de **gub.uy**. En particular, los datos elegidos son los siguientes:

- Conteo vehicular en las principales avenidas de Montevideo
- Velocidad promedio vehicular en las principales avenidas de Montevideo
- Ubicación de sensores de medición de conteo vehículos

Los tres dataset son mantenidos por la Intendencia de Montevideo.

Descripción de variables

Originalmente los datos vienen presentados de la siguiente forma:

Conteo vehicular en las principales avenidas de Montevideo

- **cod_detector**: Numérico - ID de la cámara que monitorea un carril específico para detectar vehículos.
- **id_carril**: Numérico - Número del carril monitoreado (1, 2, 3, ...).
- **fecha**: Fecha, AAAA-MM-DD - Día en que se realizó la medición.
- **hora**: hh:mm:ss - Hora en que se realizó la medición.
- **dsc_avenida**: Texto - Nombre de la avenida donde se mide el tráfico.
- **dsc_int_anterior**: Texto - Nombre de la vía desde donde vienen los vehículos.
- **dsc_int_siguiente**: Texto - Nombre de la vía hacia donde se dirigen los vehículos.
- **latitud**: Float - Latitud del lugar de medición.
- **longitud**: Float - Longitud del lugar de medición.
- **volumen**: Numérico - Cantidad de vehículos detectados en el carril en los últimos 5 minutos.
- **volumen_hora**: Numérico - Cantidad de vehículos detectados en el carril en la última hora.

Velocidad promedio vehicular en las principales avenidas de Montevideo

- **cod_detector:** Numérico - ID de la cámara que monitorea un carril específico para detectar vehículos.
- **id_carril:** Numérico - Número del carril monitoreado (1, 2, 3, ...).
- **fecha:** AAAA-MM-DD - Día en que se realizó la medición.
- **hora:** hh:mm:ss - Hora en que se realizó la medición.
- **dsc_avenida:** Texto - Nombre de la avenida donde se mide el tráfico.
- **dsc_int_anterior:** Texto - Nombre de la vía desde donde vienen los vehículos.
- **dsc_int_siguiente:** Texto - Nombre de la vía hacia donde se dirigen los vehículos.
- **latitud:** Float - Latitud del lugar de medición.
- **longitud:** Float - Longitud del lugar de medición.
- **velocidad_promedio:** Numérico - Promedio de las velocidades de los autos que circularon por el carril durante los últimos 5 minutos.

Ubicación de sensores de medición de conteo vehículos

- **dsc_avenida:** Texto - Nombre de la avenida donde se encuentra el sensor o cámara y donde se mide el tránsito.
- **dsc_int_anterior:** Texto - Nombre de la vía que forma el cruce desde donde vienen los vehículos.
- **dsc_int_siguiente:** Texto - Nombre de la vía que forma el cruce donde está el sensor. En general, el sensor se encuentra un poco antes de esta vía. El sentido de circulación será desde el cruce con **dsc_int_anterior** hacia el cruce con **dsc_int_siguiente**.
- **latitud:** Float - Coordenada que indica la latitud de la ubicación del sensor.
- **longitud:** Float - Coordenada que indica la longitud de la ubicación del sensor.

Sobre estos datos en particular, son *100 sensores* que se van cambiando de ubicación mes a mes.

Base de datos

Debido a que los datos están estrechamente relacionados y a su vez son sumamente masivos, hemos decidido utilizar una base de datos quedando de la siguiente manera.

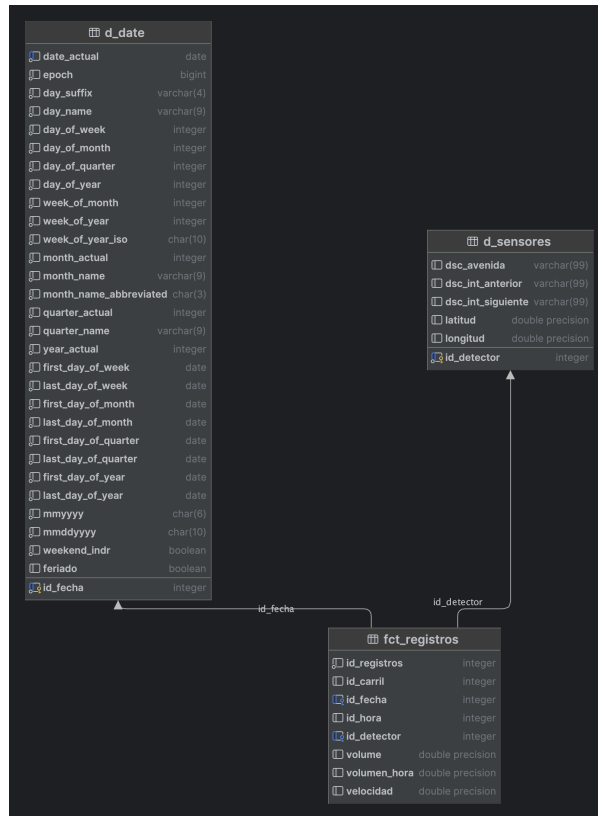


Figure 1: Diagrama de la base de datos

Nuestra tabla principal será `fct_registros`.

Tabla: `fct_registros`

- Cantidad de datos: 85386695.
- Variables de la tabla:
 - `id_registros`: *Numérico (Primary Key)*.
 - `id_carril`: *Numérico*.
 - `id_fecha`: *Numérico (Foreign Key, vinculado con d_sensores)*. La fecha de la que fue tomada el registro, tiene el formato `YYYY-MM-DD`
 - `id_hora`: *Numérico*. Hora en la que fue tomado el registro con formato `HH:MM`.
 - `id_detector`: *Numérico (Foreign Key, vinculado con d_date)*.
 - `volume`: *Numérico*. Cantidad de autos que pasaron en los últimos 5 minutos.
 - `volumen_hora`: *Numérico*. Cantidad de autos que pasaron en un lapso de una hora.
 - `velocidad`: *Numérico*. Velocidad registrada, su unidad son kilómetros por hora

Tabla: `d_sensores`

- Cantidad de datos: 273
- Variables de la tabla:
 - `id_detector`: *Numérico (Primary Key)*.
 - `dsc_avenida`: *Caracter*. Calle donde se encuentra el sensor.
 - `dsc_int_anterior`: *Caracter*. Cruce previo de la calle en `dsc_avenida`.

- `dsc_int_siguiente`: *Character*. Cruce posterior de la calle en `dsc_avenida`. Estas dos juntas nos dirá que cada sensor se encuentra en *Avenida* entre *Anterior* y *Siguiente*.
- `latitud`: *Numérico continuo*.
- `longitud`: *Numérico continuo*. Junto a `latitud` nos indica las coordenadas geograficas del sensor.
- `barrio`: *Character*. Esta variable fue creada a partir del paquete `geouy`

Tabla: `d_date`

- Cantidad de datos: 3652
- Variables de la tabla:
 - `id_fecha`: *Numérico (Primary Key)*
 - `date_actual`: *Fecha*. Secuencia de fechas desde el 01-01-2021 con formato *YYYY-MM-DD*
 - `epoch`
 - `day_suffix`: *Character*. Fecha del dia abreviado.
 - `day_name`: *Character*. Nombre del día
 - `day_of_week`: *Numérico*. Dia de la semana que indica 1 como lunes, 2 como martes, etc.
 - `day_of_month`: *Numérico*. Fecha del mes, va desde 1 hasta 31.
 - `day_of_quarter`: *Numérico*. Dia del cuatrimestre.
 - `day_of_year`: *Numérico*. Dia del año, del 1 al 366.
 - `week_of_month`: *Numérico*. Semana de cada mes, valores del 1 al 5.
 - `week_of_year`: *Numérico*. Semana del año, valores del 1 al 53.
 - `week_of_year_iso`: *Character*. Variable que combina el año, la semana del año y el día de la semana.
 - `month_actual`: *Numérico*. Mes del año tomado como numero, enero como 1, febrero como 2 y así sucesivamente.
 - `month_name`: *Character*. Mes del año traducido en texto, de enero a diciembre
 - `month_name_abbreviated`: *Character*. Mes del año en formato abreviado.
 - `quarter_actual`: *Numérico*. Indica el cuatrimestre correspondiente con numeros del 1 al 4.
 - `quarter_name`: *Character*. Indica el cuatrimestre en formato de texto, primero, segundo, tercero y cuarto.
 - `year_actual`: *Numérico*. Indica el año.
 - `first_day_of_week`: *Fecha*. Indica el primer día de la semana que corresponde tal fecha.
 - `last_day_of_week`: *Fecha*. Indica el ultimo día del rango de la semana correspondiente.
 - `first_day_of_month`: *Fecha*. Limite inferior que indica a que mes corresponde cada fecha.
 - `last_day_of_month`: *Fecha*. Limite superior que indica a que mes corresponde cada fecha.
 - `first_day_of_quarter`: *Fecha*. Limite inferior que indica a que cuatrimestre corresponde cada fecha.
 - `last_day_of_quarter`: *Fecha*. Limite superior que indica a que cuatrimestre corresponde cada fecha.
 - `first_day_of_year`: *Fecha*. Limite inferior que indica a que año corresponde cada fecha.
 - `last_day_of_year`: *Fecha*. Limite superior que indica a que año corresponde cada fecha.
 - `mmyyyy`: *Character*. Secuencia de caracteres que indica el mes y el año en formato *MMYYYY*
 - `mmddyyyy`: *Character*. Secuencia de caracteres que indica el mes, la fecha y el año en formato *MMDDYYYY*.
 - `weekend_indr`: *Lógico*. TRUE si la fecha tiene como dia de la semana sabado o domingo, FALSE en caso contrario.
 - `feriado`: *Lógico*. TRUE si la fecha correspondiente coincide con dias feriados en Uruguay, FALSE en caso contrario.

Análisis exploratorio

Para empezar mostraremos la superficie de Montevideo alcanzada por la información de los sensores de la Intendencia. Esta información no alcanza a todos los barrios, en el siguiente mapa se puede visualizar la cantidad por barrio.

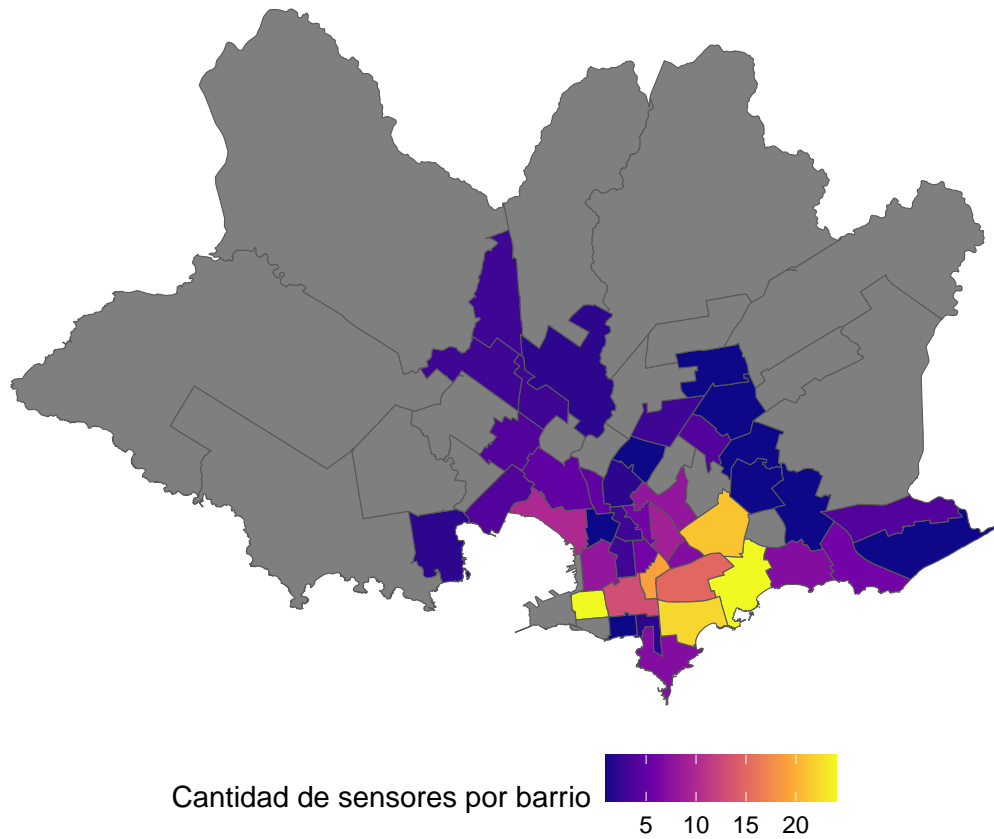


Figure 2: Los barrios que tienen sensores son 42 sobre 62 siendo los barrios de Buceo, Centro, Pocitos y Unión con mas de 20 sensores.

Veamos ahora la distribucion de las velocidades registradas

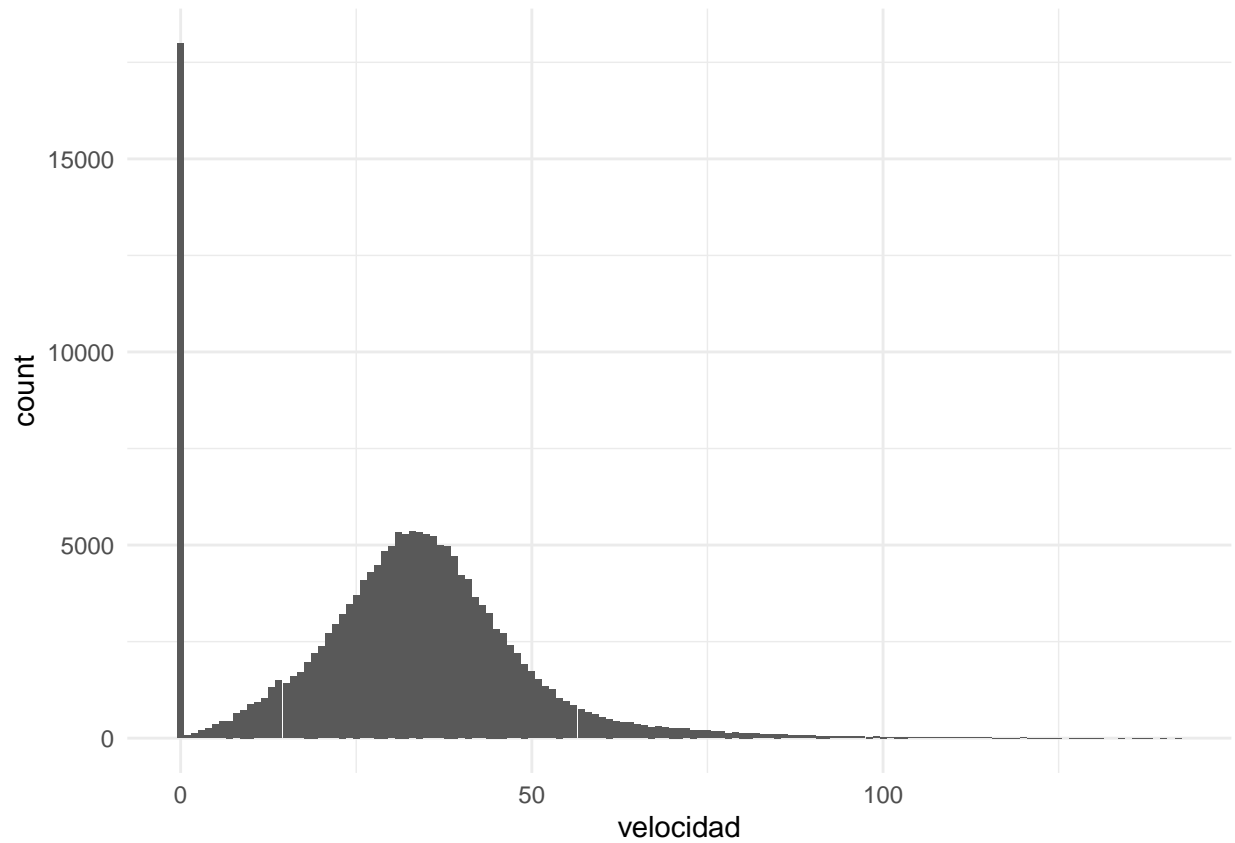


Figure 3: Se distribuye normal a partir de 1 km/h

WICKHAM, HADLEY. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'REILLY MEDIA.