

Diccionario de datos

Proyecto GiroZero

(Descripción base de datos)

*Centro del Transporte y Logística
Facultad de Ingeniería
Universidad Andrés Bello*

INTRODUCCIÓN

El presente documento constituye el Diccionario de Datos del proyecto "Giro Zero", una iniciativa en la cual CTL-UNAB pone a disposición su plataforma tecnológica para el procesamiento y visualización de datos GPS con el propósito de ofrecer información esencial para comprender el movimiento y las dinámicas del transporte de carga. El proyecto Giro Zero representa un paso significativo en la comprensión y optimización del transporte de mercancías, y este diccionario de datos desempeña un papel fundamental al proporcionar una descripción detallada de los elementos de datos utilizados en esta plataforma.

La movilidad y el transporte de carga son componentes críticos en la economía y el desarrollo de cualquier país. La capacidad de obtener información precisa y oportuna sobre la gestión de flotas, rutas de transporte, eficiencia energética y otros aspectos relacionados con la logística y la distribución de mercancías es esencial para tomar decisiones informadas y estratégicas. El proyecto Giro Zero se esfuerza por ofrecer soluciones avanzadas en esta área, aprovechando la tecnología GPS y los datos geoespaciales para brindar una visión más clara y completa de la industria del transporte de carga.

Este Diccionario de Datos tiene como objetivo proporcionar a los usuarios, desarrolladores y analistas una guía completa de los elementos de datos que se utilizan en el proyecto Giro Zero. Cada entidad y campo se describe detalladamente, especificando su nombre y tipo de dato. Esto permitirá a todos los involucrados comprender la estructura de la base de datos y garantizar la integridad y calidad de los datos recopilados y procesados.

A medida que avanzamos en este documento, se explorarán las tablas clave como: (i) grilla, (ii) red vial, (iii) detenciones, (iv) tabla de velocidades de operación a nivel de red vial ("Stage 2"), (v) tabla de velocidades de operación a nivel de cuadrantes ("Stage 3") y (vi) Map Matching (*mm_output*). Lo anterior, proporciona un marco sólido y estructurado para comprender la gestión de datos en este proyecto, y su comprensión y uso adecuado son esenciales para el éxito y la efectividad de esta iniciativa.

1 Red Vial

La red vial utilizada en este estudio corresponde a la red de transporte de Colombia, la cual cuenta con un total de 7.207.144 nodos y 2.280.742 arcos. Su área de cobertura abarca 90% de los barrios y localidades de Colombia.

La base de datos de red vial es open source, con licencia ODbL¹, y se obtuvo directamente desde OpenStreetMaps, proyecto colaborativo de creación de mapas con participación en todo el mundo. La Figura 2, presenta la cobertura y expansión de la red vial en la zona bajo estudio. Por su parte, la Figura 1 presenta la composición de arcos y nodos para un extracto del área bajo estudio.

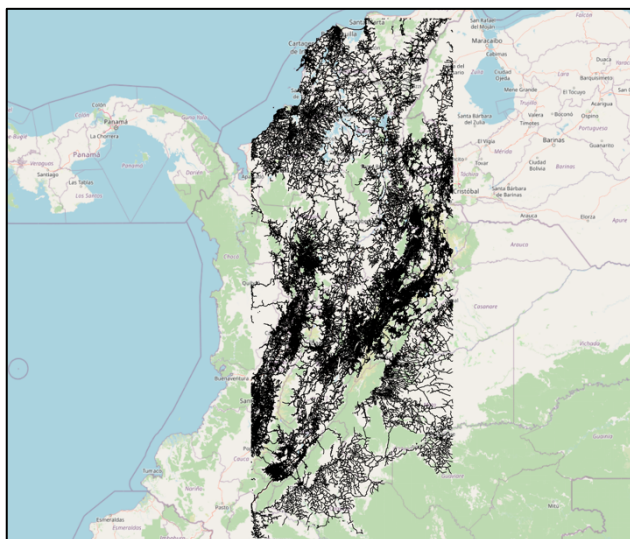


Figura 1: Composición arcos y nodos red vial



Figura 2: Cobertura red vial

Esta red representa la totalidad de calles presentes en el área geográfica definida para el estudio, y cuenta con información relacionada a longitud/distancia de cada arco, nombre, tipo de calle, dirección, entre otros atributos. La Tabla 1 presenta un extracto de esta base de datos.

¹ **Licencia Abierta de Bases de Datos:** La Licencia Abierta de Bases de Datos es un acuerdo de licencia copyleft diseñado para permitir a los usuarios compartir, modificar y usar libremente una base de datos, manteniendo la misma libertad para los demás.

Tabla 1: Estructura base de datos Red Vial

	osm_id	idcalle	name	type	one_way	oneway	maxspeed_f	maxspeed_b	idcuadrant	idtramo	largo_km	x1	y1	x2	y2
1	373458898	143984	Via Cienaga -...	trunk	0	UNKNOWN	110,0000000...	110,0000000...	1262448	63268	0,067804157...	-74,2836796...	10,9926265...	-74,2841784...	10,99226210...
2	373458898	152027	Via Cienaga -...	trunk	0	UNKNOWN	110,0000000...	110,0000000...	1262448	63269	0,008151053...	-74,2836195...	10,99267010...	-74,2836796...	10,9926265...
3	818583570	239097	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63270	0,06800694...	-74,2831912...	10,99222410...	-74,2836195...	10,99267010...
4	818583571	247116	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63271	0,03163994...	-74,2829942...	10,9924336...	-74,2831912...	10,99222410...
5	373458898	152026	Via Cienaga -...	trunk	0	UNKNOWN	110,0000000...	110,0000000...	1262448	63272	0,09904645...	-74,2826938...	10,99335310...	-74,2834184...	10,99281530...
6	818583569	239096	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63273	0,064581851...	-74,2823592...	10,99287185...	-74,2826938...	10,99335310...
7	818583569	239096	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1264848	63274	0,37877058...	-74,2821629...	10,99258960...	-74,2823592...	10,99287185...
8	798526890	222851	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63275	0,05948471...	-74,2842734...	10,99076250...	-74,2839407...	10,9903369...
9	798526889	222850	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63276	0,076016851...	-74,2849487...	10,99059789...	-74,2842734...	10,99076250...
10	798526891	222852	NULL	residential	0	UNKNOWN	50,0000000...	50,0000000...	1262448	63277	0,04758568...	-74,2848668...	10,9903005...	-74,2844592...	10,99045160...

Fuente: Centro de Transporte y Logística

Descripción:

- **osm_id:** Identificador proveniente de OpenStreetMaps, el cual indica la representación de cada calle.
- **idcalle:** Representa el identificador único que caracteriza cada arco en la red vial.
- **name:** Indica el nombre de cada calle o arco de la red.
- **type:** Especifica a qué tipo de categoría corresponde cada arco de la red.
- **one_way:** Especifica la dirección de la calle, utilizando códigos (ej: 0 = desconocido).
- **oneway:** Especifica la dirección de la calle (Ej: Un sentido, doble sentido, desconocido).
- **maxspeed_f:** indica la velocidad de diseño de cada arco de la red vial representada en kilómetros por hora.
- **maxspeed_b:** indica la velocidad de diseño de cada arco de la red vial representada en kilómetros por hora.
- **idcuadrant:** Identificador único de la porción de área de cuyo arista es de 500x500 metros.
- **idtramo:** Porción de red vial que se encuentra cortada por un cuadrante o pixel de 500x500 metros.
- **Largo_km:** Indica la distancia de cada arco. Esta medida está representada en kilómetros.
- **X1:** Representan las coordenadas X de los nodos de origen (source) cada arco de la red.
- **Y1:** Representan las coordenadas Y de los nodos de origen (source) cada arco de la red.
- **X2:** Representan las coordenadas X de los nodos de destino (target) cada arco de la red.
- **Y2:** Representan las coordenadas Y de los nodos de destino (target) cada arco de la red.

2 Grilla 500x500

La base de datos utilizada tiene un total de 2.289.600 cuadrantes de 500x500 metros los cuales abarcan toda la región abarcada por la Red Vial. En los acápites posteriores se describe el proceso de construcción de la grilla y la asignación de calles (arcos de la red) y zonas de Colombia. La Tabla 2 presenta un extracto de esta base de datos.

Tabla 2: Estructura base de datos Grilla 500x500

	idcuadrant	left	top	right	bottom
1	1	-8532574,403999999165535	1254889,339900000020862	-8532074,403999999165535	1254389,339900000020862
2	2	-8532574,403999999165535	1254389,339900000020862	-8532074,403999999165535	1253889,339900000020862
3	3	-8532574,403999999165535	1253889,339900000020862	-8532074,403999999165535	1253389,339900000020862
4	4	-8532574,403999999165535	1253389,339900000020862	-8532074,403999999165535	1252889,339900000020862
5	5	-8532574,403999999165535	1252889,339900000020862	-8532074,403999999165535	1252389,339900000020862
6	6	-8532574,403999999165535	1252389,339900000020862	-8532074,403999999165535	1251889,339900000020862
7	7	-8532574,403999999165535	1251889,339900000020862	-8532074,403999999165535	1251389,339900000020862
8	8	-8532574,403999999165535	1251389,339900000020862	-8532074,403999999165535	1250889,339900000020862

Fuente: Centro de Transporte y Logística

Descripción:

- **Idcuadrant:** Identificador único de la porción o pixel de área de cuyo arista es de 500x500 metros.
- **top:** Representan las coordenadas del nodo superior izquierdo de cada cuadrante.
- **right:** Representan las coordenadas del nodo superior derecho de cada cuadrante..
- **bottom:** Representan las coordenadas del nodo inferior derecho de cada cuadrante.
- **left:** Representan las coordenadas del nodo inferior izquierdo de cada cuadrante.

2.1 División de red topológica en cuadrantes simétricos

Este procedimiento busca acotar el espacio de búsqueda y realizar análisis con un mayor nivel de detalle. Dividir la red topológica en cuadrantes simétricos genera grandes ventajas, siendo además una importante innovación en esta área. Dentro de las ventajas está la posibilidad de representar cada sector o zona geográfica de una ciudad en pequeños cuadrantes de igual tamaño, permitiendo almacenar información en cada uno de ellos y realizar análisis detallados, conociendo cada segmento de red contenido sobre él, logrando así una fácil y rápida asignación de datos. Su implementación se lleva a cabo realizando dos procedimientos:

1. Definir un área que contenga a la totalidad de la zona bajo estudio.
2. División del área definida, en cuadrantes simétricos.

El definir un área de trabajo, tiene como finalidad el abarcar toda la zona geográfica en estudio, es decir, que no quede ningún arco sin ser contemplado. Para esto se requiere de la ubicación cartesiana de aquellos arcos que se encuentren en los extremos de la red, para posteriormente unirlos formando un polígono. Con esto se limita el área en estudio.

El segundo paso consiste en dividir el área en cuadrantes simétricos. Consiste en subdividir el área en cuadrantes iguales, de 500 metros de largo por 500 metros de ancho, para emular la estructura de damero que típicamente utilizan muchas ciudades de Sudamérica, incluida Santiago, para clasificar y almacenar información. Para llevar a cabo este proceso, se utiliza un vértice del polígono creado en el paso (1), y se determinan los próximos 3 vértices del cuadrante interior, asignando 500 unidades de largo y otros 500 de ancho. El paso siguiente consiste en tomar el vértice superior derecho del cuadrante predecesor, y determinar los vértices de los cuadrantes restantes mediante un algoritmo programado. Finalmente, se procede a unir cada uno de estos vértices, obteniendo así un área geográfica totalmente fragmentada.

2.2 Asignación de segmentos de red y zonas geográficas a cuadrantes simétricos

Una vez que se establece el cuadrículado para toda el área dispuesta a analizar, corresponde la asignación de segmentos de la red y zonas geográficas en cada uno de los cuadrantes correspondientes, la cual representa nuestra segunda metodología.

Sólo se asignan aquellos segmentos de red que se encuentren al interior de dicho cuadrante, sin tomar en cuenta los tramos más allá del área del cuadrante en cuestión, ya que éstos están contenidos en algún cuadrante vecino (Ver **Figura 3**). La lógica empleada en esta etapa se detalla a continuación:

1. En caso de que un arco sobrepase el área de un cuadrante, el algoritmo sólo asignará el segmento contenido en él. Donde los segmentos restantes serán asignados a los cuadrantes vecinos. Lo anterior se determina a través de los puntos de corte o intersección existentes entre el arco de red y las aristas del cuadrante.
2. Por el contrario, cuando el segmento de red se encuentra completamente contenido por el cuadrante, la asignación se realiza inmediatamente. Esto puede determinarse directamente usando las coordenadas de los vértices de cada cuadrante, así como de los puntos inicio y fin de los arcos.

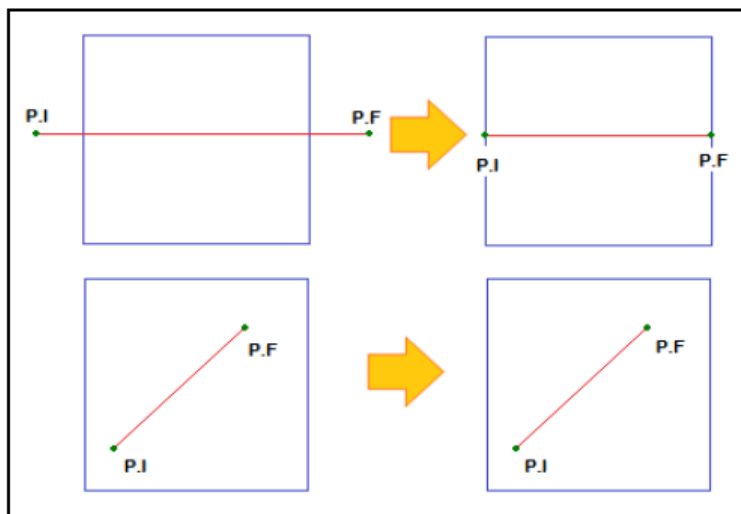


Figura 3: Asignación de segmentos de red a cuadrantes

Fuente: Centro de Transporte y Logística

La relación cuadrante-zona geográfica, tiene por finalidad obtener la siguiente asociación: cada arco pertenece a uno o más cuadrantes, y a su vez cada cuadrante compone una o más zonas geográficas. Pueden existir innumerables zonas geográficas las cuales podrán ser representadas mediante un conjunto de cuadrantes. De esta manera consideramos que la unidad base de análisis sea el cuadrante.

Los criterios utilizados para realizar la asignación de cuadrantes a una zona se detallan a continuación:

1. Cada cuadrante pertenece sólo a una zona.
2. Se asigna aquel cuadrante que contenga más del 50% del área, contenida en la zona analizada.

El primer criterio tiene la particularidad de eliminar duplicidad de información, es decir, que no existan zonas con un mismo cuadrante en común. El segundo es para decidir si se asigna o no un cuadrante a la zona en cuestión. En la Figura 4, se presenta gráficamente este criterio de asignación, donde la zona geográfica original y transformada se muestra en amarillo. Podemos ver que los cuadrantes 1 y 2 son asignados, dado que la zona analizada contempla más del 50% de su área. Mientras que los cuadrantes 3 y 4 no lo son, ya que no cumplen con los criterios anteriores.

Este proceso de asignar de arcos y zonas geográficas a cada cuadrante permite almacenar la información grandes volúmenes de información mejorando la performance y análisis de los datos asociados.

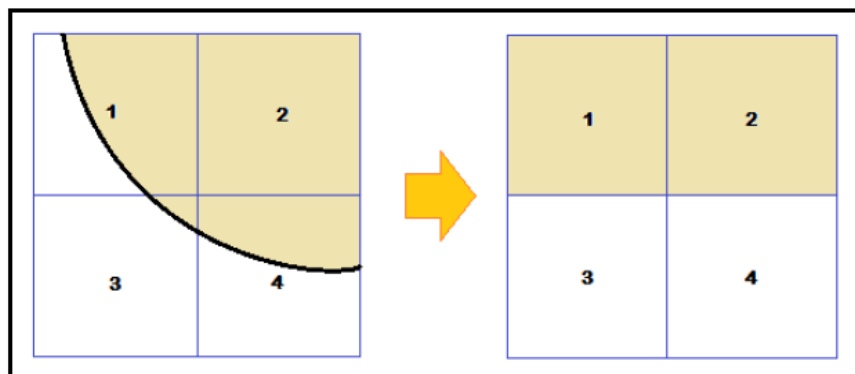


Figura 4: Asignación de cuadrantes a comunas
Fuente: Centro de Transporte y Logística

3 Detenciones

La Tabla 3 presenta un extracto de la base de datos relacionada con las detenciones en ralentí producidas por los vehículos de carga monitoreados por medio de dispositivos GPS.

Tabla 3: Estructura base de datos Detenciones

patente	cluster_id	idpunto	id_detencion	inicio_detencion	fin_detencion	Fecha	duracion_en_segundos_sum	x_detencion	y_detencion
GY909	0	2	1	23:59:55	23:59:55	2022-10-31	10277	-73,40809	7,79715
GY909	0	13364	2	07:18:34	07:18:34	2022-11-29	120	-73,40807	7,79719
GY909	1	8	3	00:08:22	00:08:22	2022-11-01	120	-73,42205	7,82547
GY909	1	13372	4	07:31:27	07:31:27	2022-11-29	119	-73,42219	7,82569
GY909	1	14305	5	15:04:52	15:04:52	2022-12-02	120	-73,4223	7,8255
GY909	2	11	6	00:14:22	00:14:22	2022-11-01	120	-73,45071	7,88209

Fuente: Centro de Transporte y Logística

Descripción:

- **patente:** Identificador único de cada vehículo monitoreado.
- **cluster_id:** Id interno del algoritmo de clusterización. Esta variable es interna, no utilizada en los análisis y dashboard desarrollados.
- **idpunto:** Identificador único de cada detención.
- **inicio_detencion:** Representan la hora en la cual se inicia una detención en ralentí.
- **Fin_detencion:** Representan la hora en la cual finaliza una detención en ralentí.
- **Fecha:** Representan la fecha en la cual se genera una detención en ralentí.
- **Duración_en_segundos_sum:** Representan el tiempo, expresado en segundos, de la detención en ralentí.
- **x_detencion:** Representan las coordenadas x (Latitud) de la detención.

- **y_detencion:** Representan las coordenadas y (Longitud) de la detención.

4 Velocidad de operación a nivel de red vial (Stage 2)

La Tabla 4 presenta un extracto de la base de datos relacionada con las velocidades de operación a nivel de red vial (stage 2), obtenidas por los vehículos de carga monitoreados por medio de dispositivos GPS.

Tabla 4: Estructura base de datos Stage 2

osm_id	idcalle	idtramo500	type	fecha1	hora1	n_movil	n_puntos	velocidad_promedio
23592322		1245427	secondary	2022-11-23	11	1	1	14
23953563		1546763	trunk	2022-11-03	13	1	1	66
23953563		1546763	trunk	2022-11-16	7	1	1	46
23953563		1546763	trunk	2022-11-25	21	1	1	71
23953563		1546763	trunk	2022-11-29	22	1	1	67

Fuente: Centro de Transporte y Logística

Descripción:

- **osm_id:** Identificador proveniente de OpenStreetMaps, el cual indica la representación de cada calle.
- **idcalle:** Representa el identificador único que caracteriza cada arco en la red vial.
- **idtramo500:** Porción de red vial que se encuentra cortada por un cuadrante o pixel de 500x500 metros.
- **type:** Especifica a qué tipo de categoría corresponde cada arco de la red.
- **fecha1:** Representan la fecha en la cual existe información sobre velocidad de operación en el tramo o red vial.
- **hora1:** Representan la hora en la cual existe información sobre velocidad de operación en el tramo o red vial. La base de datos cuenta con agrupaciones de horas de cada 1 hora.
- **n_movil:** Número de vehículos que representan la velocidad de operación.
- **n_puntos:** Número de puntos GPS que conforman la velocidad de operación.
- **velocidad_promedio:** Velocidad de operación promedio en rangos de 1 hora (60 minutos).

5 Velocidad de operación a nivel de grilla (Stage 3)

La Tabla 5 presenta un extracto de la base de datos relacionada con las velocidades de operación a nivel de grilla 500x500 (stage 3), obtenidas por los vehículos de carga monitoreados por medio de dispositivos GPS.

Tabla 5: Estructura base de datos Stage 3

idcuadrante500	fecha1	hora1	n_movil	n_puntos	velocidad_promedio
577	2022-12-14	8	1	1	13
716	2022-12-14	10	1	1	60
721	2022-12-14	10	1	2	73.2
761	2022-12-14	14	1	1	37
761	2022-12-14	15	1	2	31
762	2022-12-14	14	1	4	3.5

Fuente: Centro de Transporte y Logística

Descripción:

- **Idcuadrante500:** Identificador único de la porción o pixel de área de cuyo arista es de 500x500 metros.
- **fecha1:** Representan la fecha en la cual existe información sobre velocidad de operación en el tramo o red vial.
- **hora1:** Representan la hora en la cual existe información sobre velocidad de operación en el tramo o red vial. La base de datos cuenta con agrupaciones de horas de cada 1 hora.
- **n_movil:** Número de vehículos que representan la velocidad de operación.
- **n_puntos:** Número de puntos GPS que conforman la velocidad de operación.
- **velocidad_promedio:** Velocidad de operación promedio en rangos de 1 hora (60 minutos).

6 Map Matching (mm_output)

La Tabla 6 presenta un extracto de la base de datos relacionada con las rutas y/o recorridos generados por los vehículos de carga monitoreados por medio de dispositivos GPS.

Tabla 6: Estructura base de datos Map Matching (mm_output)

patente	fecha_hora	t_inicio	t_fin	arco	id	path_ruta	path_punto	type	osm_id	name	kmh	km	hora	vel	cum_km
1	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227626	1	1	1	trunk	884210266	Vía a la Costa...	110,0000000...	0,48350735...	00:00:22.000	63,0000000...	0,48350735...
2	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227627	1	2	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,06484431...	NULL	NULL	0,54835166...
3	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227632 222...	1	3	2	trunk	884210266	Vía a la Costa...	110,0000000...	0,54844669...	00:02:22.000	0	1,09679836...
4	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227636	1	4	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,50918279...	NULL	NULL	1,605981157...
5	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227635	1	5	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,03925702...	NULL	NULL	1,645238180...
6	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227485	1	6	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,54822839...	NULL	NULL	2,193466579...
7	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227487	1	7	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,53608061...	NULL	NULL	2,729547195...
8	GY909	2022-11-01 00:00:22.000	01:35:58.000	2227490	1	8	NULL	trunk	884210266	Vía a la Costa...	110,0000000...	0,08762473...	NULL	NULL	2,817171930...

Fuente: Centro de Transporte y Logística

Descripción:

- **patente:** Identificador único de cada vehículo monitoreado.
- **fecha_hora:** Representan la fecha en la cual se efectúa el recorrido del vehículo.
- **t_inicio:** Representan la hora en la cual se inicia una el recorrido o viaje.

- **t_fin:** Representan la hora en la cual se finaliza el recorrido o viaje.
- **arco:** Porción de red vial que se encuentra cortada por un cuadrante o pixel de 500x500 metros. También llamado *"idtramo500"*
- **id:** Representa el identificador de la ruta
- **path_ruta:** Indica la secuencia de arcos que conforman en la ruta del vehículo.
- **path_punto:** Indica la secuencia de puntos gps que conforman en la ruta del vehículo.
- **type:** Especifica a qué tipo de categoría corresponde cada arco de la red vial.
- **osm_id:** Identificador proveniente de OpenStreetMaps, el cual indica la representación de cada calle.
- **name:** Indica el nombre de cada calle o arco de la red
- **km/h:** Indica la velocidad de diseño del arco o segmento de red vial.
- **Km:** Indica la distancia del arco o segmento de red vial.
- **hora:** Indica la hora en la cual se registra un punto GPS dentro de la ruta seguida por el vehículo.
- **vel:** Indica la velocidad (km/h) de un punto GPS dentro de la ruta seguida por el vehículo.
- **cum_km:** Indica la distancia total acumulada producto del recorrido el vehículo bajo análisis.