

Informe de Inconsistencias en la Base de Datos de Conteo Vehicular

Periodo analizado: Enero 2025

Cobertura: Intersecciones y sentidos viales de la ciudad (conteos horarios).

Registros múltiples para un mismo cruce, sentido y día

Durante la revisión de la base de datos se identificó que, para varios cruces y sentidos de circulación, aparecen **dos o más registros distintos para un mismo día**. Esto resulta problemático, ya que, en teoría, debería existir un único registro por hora y por dirección de circulación.

En particular, observamos que el campo **ALIAS**, que debería identificar de manera única a cada registro, presenta diferencias mínimas: al mismo identificador de día, cruce y sentido se le añade una letra (ej. *T* o *C*). Sin embargo, no existe documentación que explique el significado de estas variaciones.

Ejemplo encontrado

Cruce: **12 de Octubre y Luis Cordero**

Sentido: **NORTE-SUR**

Fecha: **01/01/2025**

- Registro 1: 01-01-2025PM335-0-01
- Registro 2: 01-01-2025PM335-0-01T

Ambos registros pertenecen al mismo día y lugar, pero presentan diferencias en los conteos horarios.

ALIAS	CRUCE	SENTIDO	00	01	02	03	04	05	06	07	08	09	10	11	12	13
01-01-2025PM335-0-01	12 de Octubre y Luis Cordero	NORTE-SUR	128	242	230	125	68	72	39	1	0	22	0	0	2	1
01-01-2025PM335-0-01T	12 de Octubre y Luis Cordero	NORTE-SUR	163	313	300	172	96	91	54	1	0	35	0	0	2	1
02-01-2025PM335-0-01	12 de Octubre y Luis Cordero	NORTE-SUR	23	17	9	6	24	58	54	0	2	1	4	2	1	1
02-01-2025PM335-0-01T	12 de Octubre y Luis Cordero	NORTE-SUR	37	23	14	9	29	81	66	0	2	1	4	2	1	1
03-01-2025PM335-0-01	12 de Octubre y Luis Cordero	NORTE-SUR	64	46	19	12	25	46	40	0	0	0	0	0	0	0

Consulta a aclarar

- ¿Cuál es la diferencia entre un registro con sufijo (*T* o *C*) y uno sin sufijo?
- ¿Corresponden a **distintas fuentes de conteo** (por ejemplo, equipos diferentes en la misma intersección)?
- ¿Se trata de **versiones preliminares y ajustadas** de los datos?
- ¿Cuál debería considerarse como el registro válido para el análisis?

2. Registros con conteos totalmente nulos en todas las horas del día

Descripción del problema

En varios cruces y sentidos de circulación, los registros correspondientes a determinados días aparecen con **valores en cero en todas las horas (00 a 23)**. Este patrón se repite de manera consecutiva durante varios días, lo cual sugiere un problema sistemático más que un comportamiento real del tráfico.

Ejemplo encontrado

Cruce: **12 de Octubre y Luis Cordero**

Sentido: **SUR-NORTE**

- **01 de enero de 2025:** 01-01-2025PM335-0-07 y 01-01-2025PM335-0-07T → todos los conteos horarios en 0.
- **02 de enero de 2025:** 02-01-2025PM335-0-07 y 02-01-2025PM335-0-07T → todos los conteos horarios en 0.
- **03 al 07 de enero de 2025:** se repite la misma situación

ALIAS	CRUCE	SENTIDO	00 -	01 -	02 -	03 -	04 -	05 -	06 -	07 -	08 -	09 -	10 -	11 -	12 -	13 -	14 -	15 -	16 -	17 -
01-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
01-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
02-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
02-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
03-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
03-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
04-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
04-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
05-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
05-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
06-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
06-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
08-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
08-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
09-01-2025PM335-0-07	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
09-01-2025PM335-0-07T	12 de Octubre y Luis Cordero	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Consulta a aclarar

- ¿A qué se debe que existan registros **completamente nulos en todas las horas**?
- ¿Se trata de una **falla del sensor** en esos días?
- ¿Corresponde a un **problema en la captura o transferencia de datos** hacia la base?
- ¿Estos registros deberían considerarse **como datos perdidos** y descartarse en el análisis?

3. Presencia de letras en el identificador (*ALIAS*) sin documentación sobre su significado

Descripción del problema

Se observa que en la columna **ALIAS**, que debería funcionar como identificador único del cruce, sentido y fecha, existen registros que terminan con letras adicionales (ejemplo: **C, B, T**). No obstante, la base de datos no contiene información que explique el significado o la función de estas letras, lo cual genera incertidumbre al momento de utilizar los datos.

Además, se identificó que en algunos de estos registros con sufijo en el *ALIAS*, los conteos aparecen totalmente nulos en todas las horas, mientras que en otros casos los valores son distintos de cero.

Ejemplo encontrado

- 01-01-2025PM003-0-012C → **todos los conteos en 0.**
- 01-01-2025PM004-0-071C → **todos los conteos en 0.**
- 01-01-2025PM009-0-07B → presenta valores distintos de cero en varias horas.
- 01-01-2025PM015-0-04T → presenta conteos horarios consistentes y distintos de cero.

Consulta a aclarar

- ¿Qué representan las letras adicionales al final del campo *ALIAS* (**T, C, B, etc.**)?
- ¿Estas letras corresponden a **distintos dispositivos de conteo** en la misma intersección, o a **procesos de validación/corrección** de los datos?
- En los casos donde el registro termina en letra y presenta **todos los valores en cero**, ¿se debe interpretar como datos faltantes por falla técnica, o como ausencia real de tráfico?
- ¿Cuál de los registros (con o sin sufijo) debería considerarse como **oficial** para el análisis?

ALIAS	CRUCE	SENTIDO	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
01-01-2025PM003-0-012C	Av. 10 de Agosto y Falconi	NORTE-SUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
01-01-2025PM004-0-071C	Av. 10 de Agosto y El Inca	SUR-NORTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
01-01-2025PM009-0-07B	Av. 6 de Diciembre y Granados	SUR-NORTE	0	0	0	2	1	0	2	8	7	18	8	8	9	10	15	10	7	71
01-01-2025PM015-0-04T	Av. Amazonas y Gaspar de Villarroel	OCCIDENTE-ORIENTE	140	248	190	87	47	52	116	124	158	199	258	385	439	502	453	471	438	431

4. Presencia de valores atípicos extremadamente elevados en los conteos horarios

Descripción del problema

Al calcular el valor máximo de vehículos contabilizados por hora en toda la base de datos, se observan cifras inusualmente altas que no parecen coherentes con la capacidad de las intersecciones o el flujo vehicular esperado.

Por ejemplo, se obtuvo la siguiente serie de máximos por hora (00 a 23):

Hora	00	01	02	03	04	05	06	07	08	09	10	11
Valos maximo	1969	1910	2306	934	719	1906	3848	19005	8111	7279	17694	7115
Hora	12	13	14	15	16	17	18	19	20	21	22	23
Valos maximo	8344	7347	7301	8177	18036	21098	12705	6216	5705	4707	3855	2896

Se destacan algunos picos atípicos, como **19005, 17694, 18036 y 21098 vehículos en una sola hora**, los cuales resultan extremadamente altos para condiciones normales de circulación en la ciudad.

Ejemplo encontrado

- En un día específico, se registró un conteo de **21098 vehículos en una sola hora**, cifra que excede ampliamente lo esperado en la mayoría de intersecciones.

Consulta a aclarar

- ¿Estos valores corresponden a **conteos reales** de tránsito o se trata de **errores en la captura/almacenamiento de datos**?
- ¿Es posible que se trate de una **duplicación de registros** o que se estén sumando varias fuentes/dispositivos en un mismo campo?
- ¿Existe algún **valor de referencia máximo por intersección/hora** que permita validar si estos picos son verídicos o deben tratarse como anomalías?
- En caso de confirmarse como errores, ¿qué procedimiento recomienda la entidad para el **tratamiento de estos outliers** en el análisis posterior?

El análisis preliminar de la base de datos de conteo vehicular correspondiente al mes de enero revela la presencia de varias inconsistencias relevantes: registros duplicados, conteos nulos o incompletos, inconsistencias en los sentidos viales y valores extremadamente elevados que sugieren la existencia de datos atípicos.

Dado que estas situaciones afectan de manera directa la **validez y confiabilidad** de los análisis que se realicen sobre la información, solicitamos de manera formal se aclare:

1. **El origen de las inconsistencias detectadas**, en particular los valores atípicos de conteo horario (ej. hasta 21.098 vehículos en una hora) y los registros duplicados o incompletos.

2. **Si la base de datos entregada cuenta con algún proceso de depuración o preprocesamiento previo** antes de su entrega, o si corresponde a datos en crudo directamente exportados de los dispositivos de conteo.
3. **Qué lineamientos recomienda la entidad** respecto al tratamiento de estos registros en los análisis posteriores (p. ej., eliminación de duplicados, corrección manual, imputación de valores faltantes o descarte de outliers).

La aclaración de estas dudas es fundamental para garantizar que el trabajo posterior de análisis y modelamiento se base en información confiable y metodológicamente consistente.