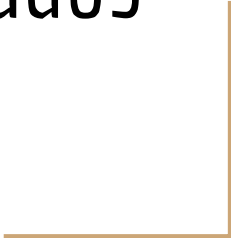


Clasificación de reportes falsos de Incidentes Viales reportados en el C5



José Roberto Pérez Chávez
Mariana Lugo Ibarra
Elena Villalobos Nolasco

C5

- El Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5).
- Dependencia del Gobierno de la Ciudad de México.
- Encargada de captar información integral para la toma de decisiones en materia de seguridad pública, urgencias médicas, movilidad, entre otras.
- Canales:
 - Video monitoreo.
 - Atención 911.
 - Denuncia anónima 089.
 - Locatel.

Problemas


- El C5 sólo cuenta con 20 ambulancias para enviar en caso de un incidente.
- Existe una cantidad considerable de reportes falsos, 20% del total de incidentes reportados.
- Lo anterior provoca la posibilidad de pérdida de vidas de personas que realmente necesitan de una ambulancia.
- Por lo tanto, es necesario identificar los reportes falsos con la mayor precisión para no enviar ambulancias a incidentes falsos.

Objetivo:


- Clasificar si los reportes de incidentes viales realizados al C5, son falsos o verdaderos.



Análisis Descriptivo Exploratorio y data profiling (EDA)



Base de datos

- Conjunto de datos de los incidentes viales reportados por el C5.
- Observaciones: **1383138**.
- Variables: **18** 

Categórica	Fecha/tiempo	Localización
folio	fecha_creacion	latitud
codigo_cierre	hora_creacion	longitud
delegacion_inicio	dia_semana	geopoint
incidente_c4	fecha_cierre	--
clas_con_f_alarma	año_cierre	--
tipo_entrada	mes_cierre	--
delegacion_cierre	hora_cierre	--
--	mes	--

Label: codigo_cierre

(A) La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos y confirmó la emergencia reportada: **43.4%**.

(D) El incidente reportado se registró en dos o más ocasiones procediendo a mantener un único reporte (afirmativo, informativo, negativo o falso) como el identificador para el incidente: **35.8%**.

(N) La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos, pero en el sitio del evento nadie solicitó el apoyo de la unidad: **19.6%**.

(F) El operador/a o despachador/a identifican, antes de dar respuesta a la emergencia, que ésta es falsa. O al ser despachada una unidad de atención a emergencias en el lugar de los hechos se percatan que el incidente no corresponde al reportado inicialmente: **0.8%**.

(I) El incidente reportado es afirmativo y se añade información adicional al evento: **0.4%**.

Variables categóricas (principales)

- [`delegacion_inicio`]: Ubicación delegacional inicial del reporte (top 3):
 1. Iztapalapa (16%)
 2. Gustavo A. Madero (10.5%)
 3. Cuauhtemoc (10.2%)
- [`incidente_c4`]: Tipo de incidente reportado (top 3):
 1. Accidente-choque sin lesionados (55%)
 2. Accidente-choque con lesionados (22.8%)
 3. Lesionado atropellado (14%).
- [`tipo_entrada`]: Tipo de entrada en que se dio el reporte (top 3):
 1. Llamada 911 (53.5%)
 2. Llamada 055 (33.5%)
 3. Botón de auxilio (5.7%)

Observaciones principales de variables categóricas

- Existe un folio único para cada una de las observaciones.
- El 20% de los envíos de asistencia registrados en la base son inútiles debido a que la ambulancia llega al lugar pero no hay emergencia que atender.
- El 55% de los incidentes no (necesariamente) requieren que se envíe una ambulancia de asistencia y sólo el 0.8% de los reportes falsos son detectados a tiempo.

Variables temporales

- `[fecha_creacion]`: Fecha de creación del reporte.
formato: dd/mm/aaaa
- `[hora_creacion]`: Hora de creación del reporte.
formato: hh:mm:ss
- `[dia_semana]`: Día de la semana en la que se creó el reporte (top 3):
Viernes, Sábado y Jueves
- Existen otras variables como `[fecha_cierre]`, `[hora_cierre]`, `[año_cierre]` entre otras, que registran el momento del cierre del reporte, que son *aposteriori* al evento de interés: determinar si un reporte es falso o no.

Observaciones principales de variables temporales

- Los datos de fechas están bastante completos. El atributo hora_creacion es el que presenta mayor falta de información mostrando 28,323 registros (ó 2.05% del total) inválidos.
- Hay una distribución equilibrada de los datos anuales. Adicionalmente, el número de registros por mes y día del mes muestran una distribución balanceada.
- Como era de esperarse, hay días de la semana y horas del día en que se observa un mayor registro de incidentes: los días jueves, viernes y sábado; y las 18, 19 y 20 horas.

Variables de localización

- `[latitud]`: Latitud del reporte.
formato: dd/mm/aaaa
- `[longitud]`: Longitud del reporte.
formato: hh:mm:ss
- `[geopoint]`: Misma información que latitud y longitud pero en formato string.

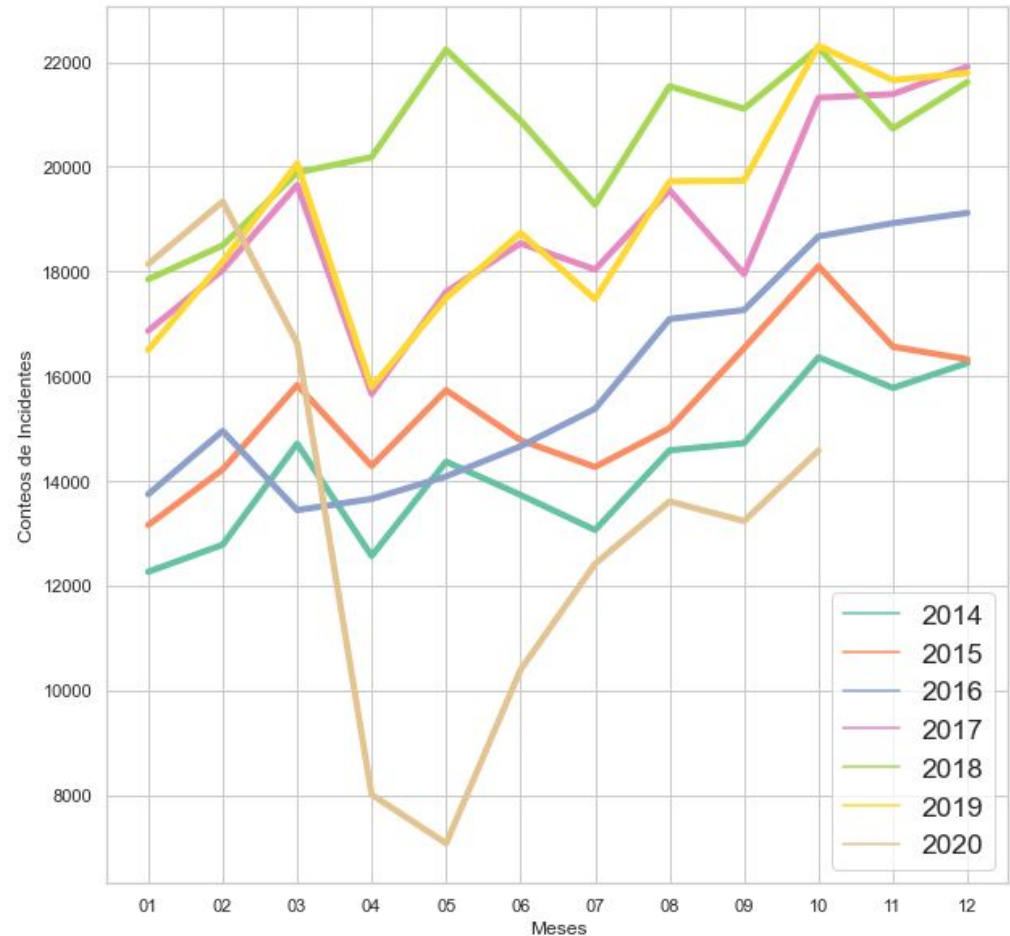
Observaciones principales

- Tenemos coordenadas para la gran mayoría (99%) de los incidentes registrados.
- Poco más del 70% de todas las coordenadas tienen una precisión de 8 decimales. La mayoría de las coordenadas restantes tienen una menor precisión.
- La mayoría de las coordenadas que observamos en el mapa sí corresponden a la CDMX, excepto por algunas que se salen del área de la ciudad.

Análisis Gráfico Exploratorio (GEDA)

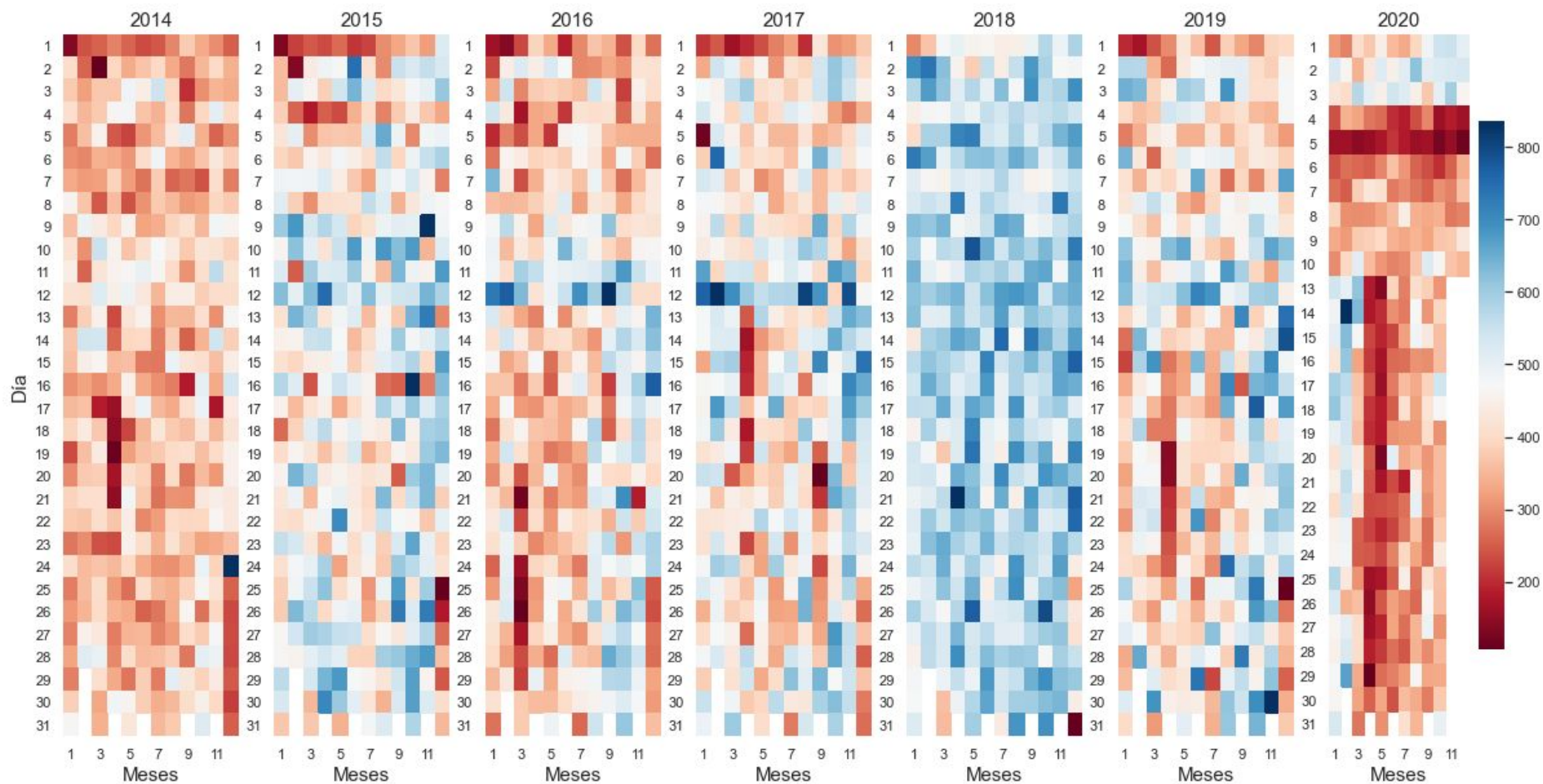
1. Comportamiento de todos los incidentes a través de los años...

- Se observa que el 2018 tuvo un incremento en incidentes en comparación con los otros años.
- Del año 2020, se observa la caída de incidentes que se asume es debido a la pandemia.
- Los demás años parecen tener un comportamiento un poco similar entre ellos. Conforme pasan los meses parece haber un incremento en el número de incidentes. Sin embargo, cada año está en diferentes niveles.

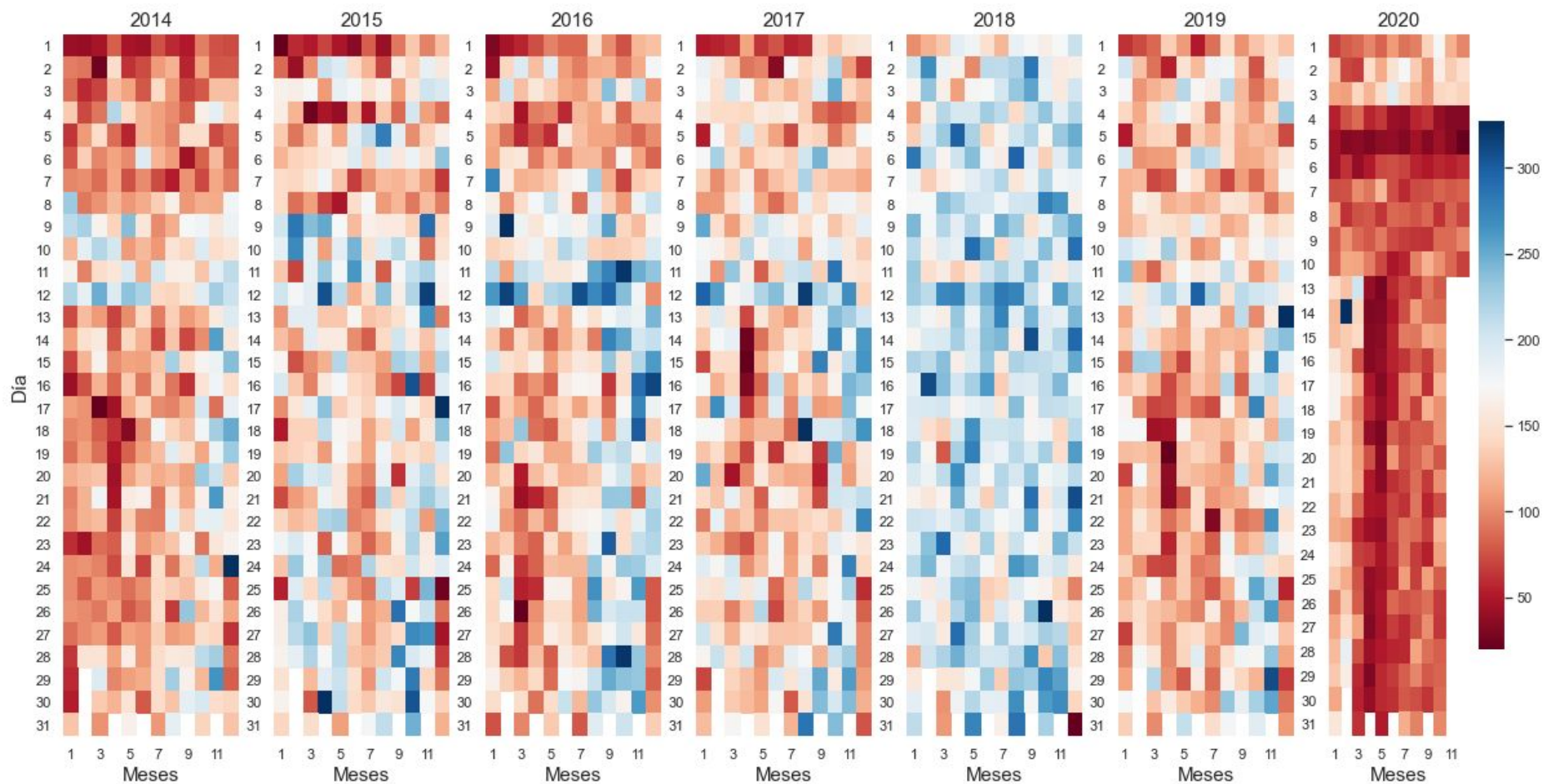


2. Existe algún factor temporal de año, mes o día que afecte la ocurrencia de reportes falsos...

Conteos (aprox) por día, mes y año de los reportes considerados como VERDADEROS

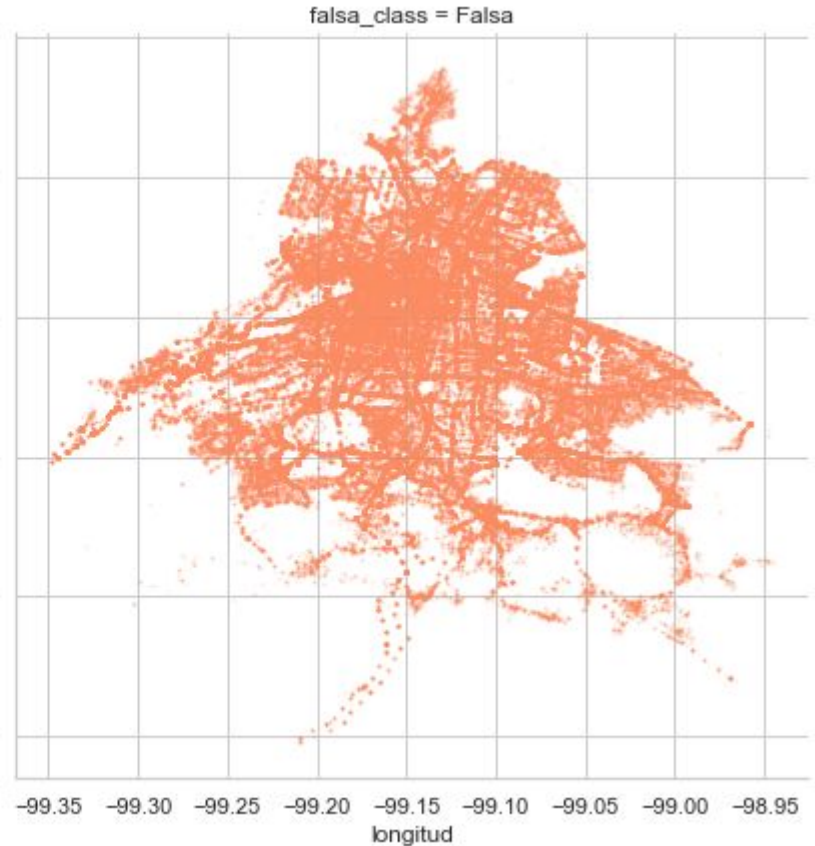
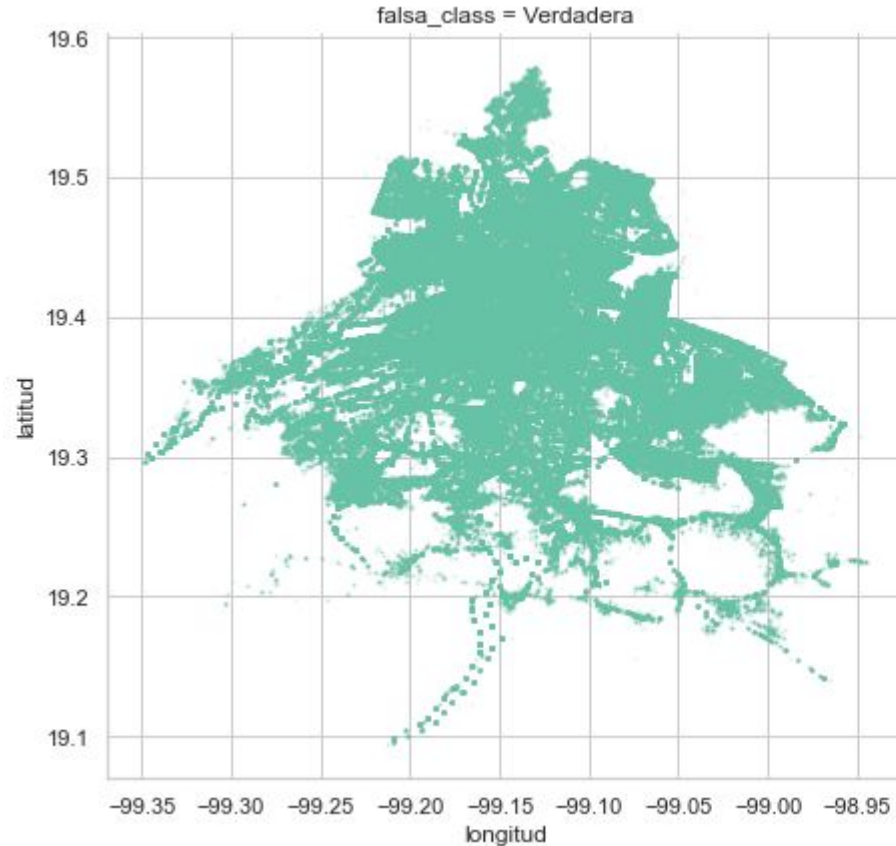


Conteos (aprox) por día, mes y año de los reportes considerados como FALSOS



- En los gráficos anteriores no se observó algún patrón que surja a partir del día del mes, mes o año, ni si estos cambian en función de la llamada considerada como verdadera o falsa.
- Existe algún patrón de incremento que se observa sobre los días 12, sin embargo, se desconoce la razón de este específico incremento.
- Se confirma que el año 2018 tiene un mayor número de incidentes en comparación a los otros años.

3. Existe alguna diferencia en la localización de los reportes falsos o verdaderos...

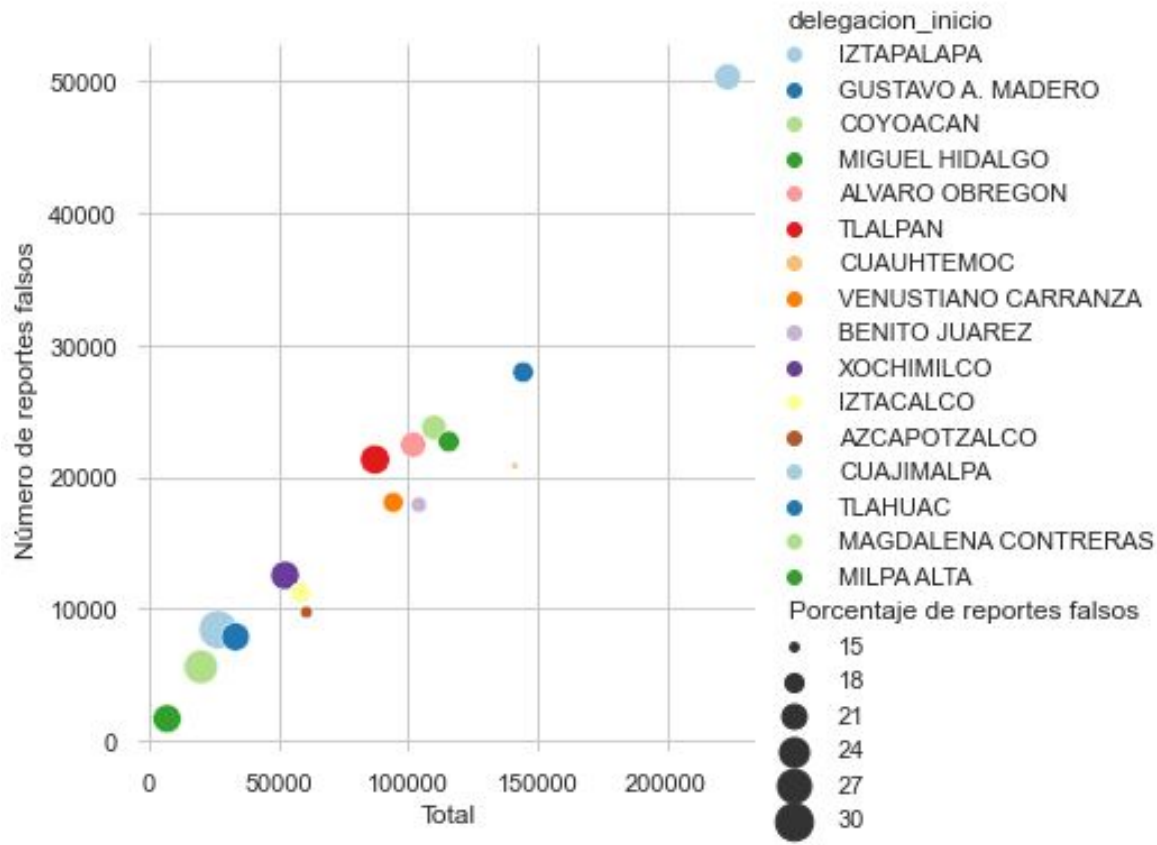


- En los gráficos de latitud y longitud anteriores, no se observó alguna concentración clara que sugiera que la localización de los reportes falsos incrementen más de acuerdo a ciertas zonas de la CDMX.
- Se observa la concentración de incidentes en las vialidades principales de la ciudad.

4. ¿Qué delegaciones tienen más reportes falsos?

Este gráfico permite ver tanto los conteos como los porcentajes de reportes falsos por cada delegación (tamaño de la burbuja).

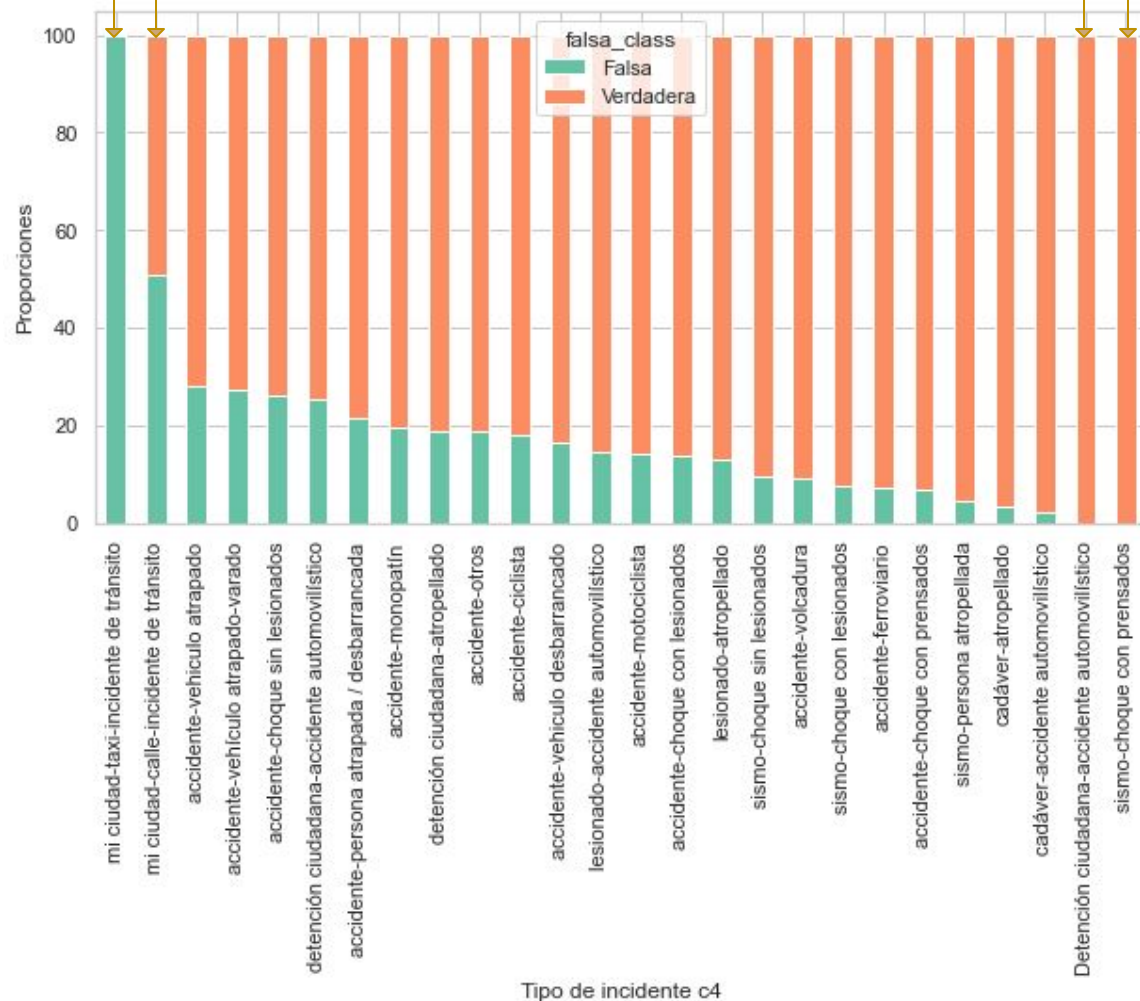
Se observa una relación lineal entre el número de incidentes totales y los considerados falsos. Lo importante de este gráfico es el cuadrante inferior izquierdo, donde parece que hay mayor porcentaje de reportes falsos a pesar de que el número de incidentes es menor.



5. Existe algún incidente que tenga más reportes falsos...

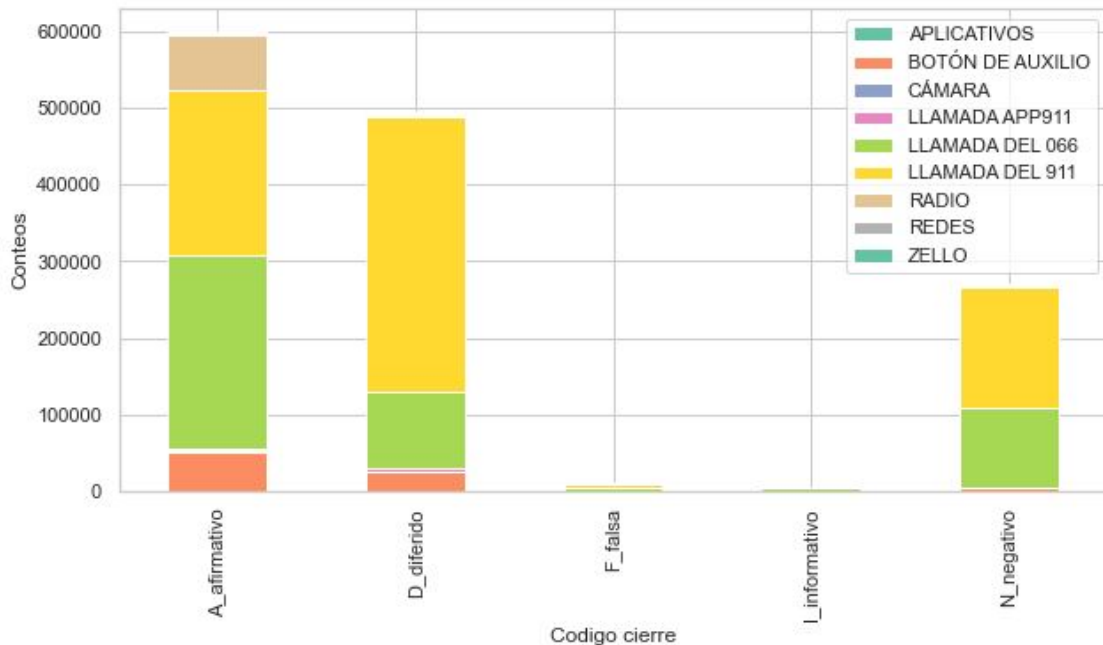
En este gráfico se observa que las proporciones de acuerdo al tipo de incidente difieren entre ellas.

Los tipos de incidentes que se encuentran en ambos extremos del gráfico son aquellos que tienen muy pocas observaciones (los indicados por las flechas superiores).

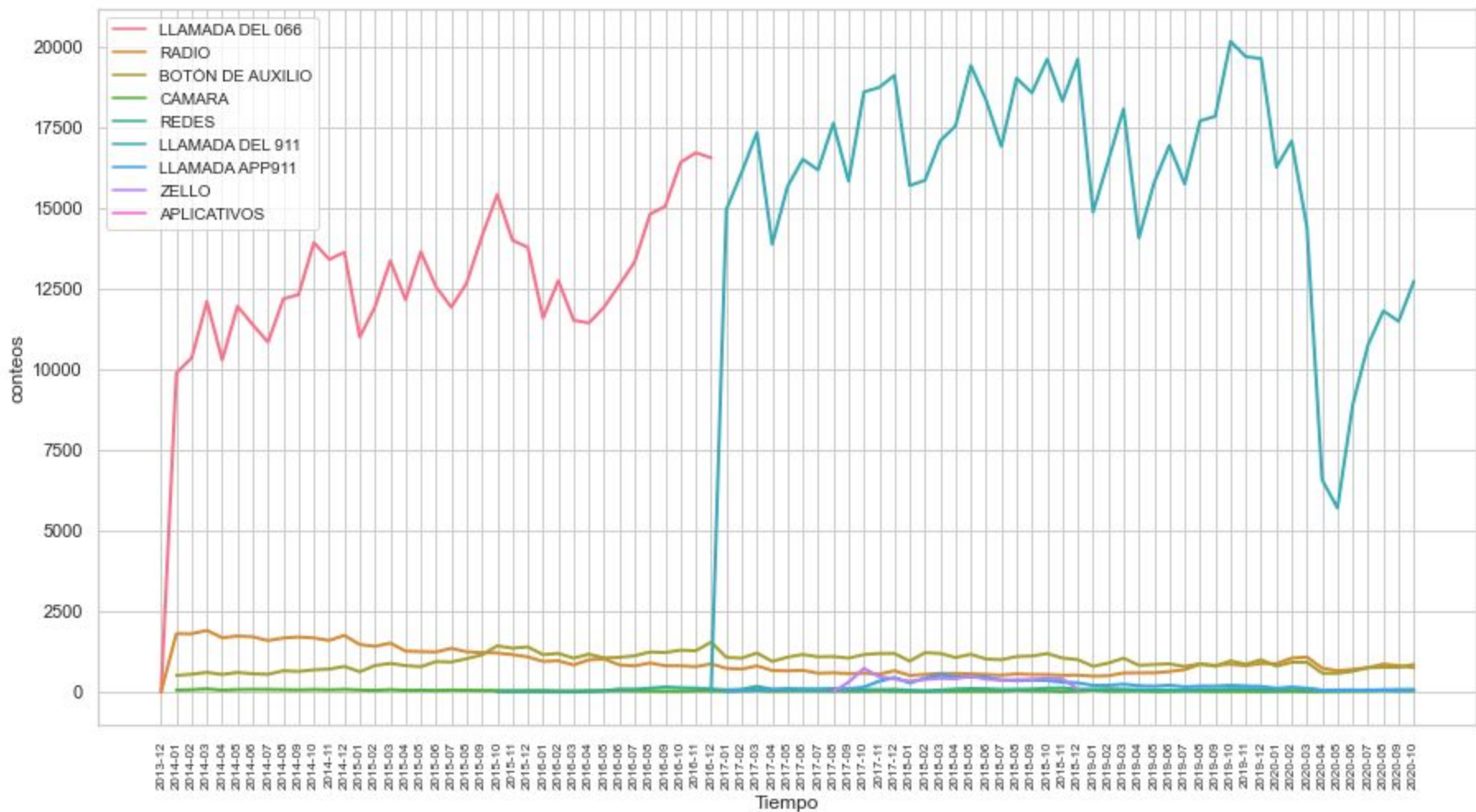


6. Existe relación entre el tipo de entrada y el código de cierre, ¿qué canal es más propenso a que se den falsas alarmas?

- Parece que los canales que más propensión a falsas alarmas o reportes negativos son aquellos que provienen del 911 y/o del 066. Sin embargo, también son los que más se utilizan para reportar incidentes verdaderos.
- El radio y el botón de auxilio son los siguientes canales en preferencia. Éstos reportan muy pocos incidentes falsos.

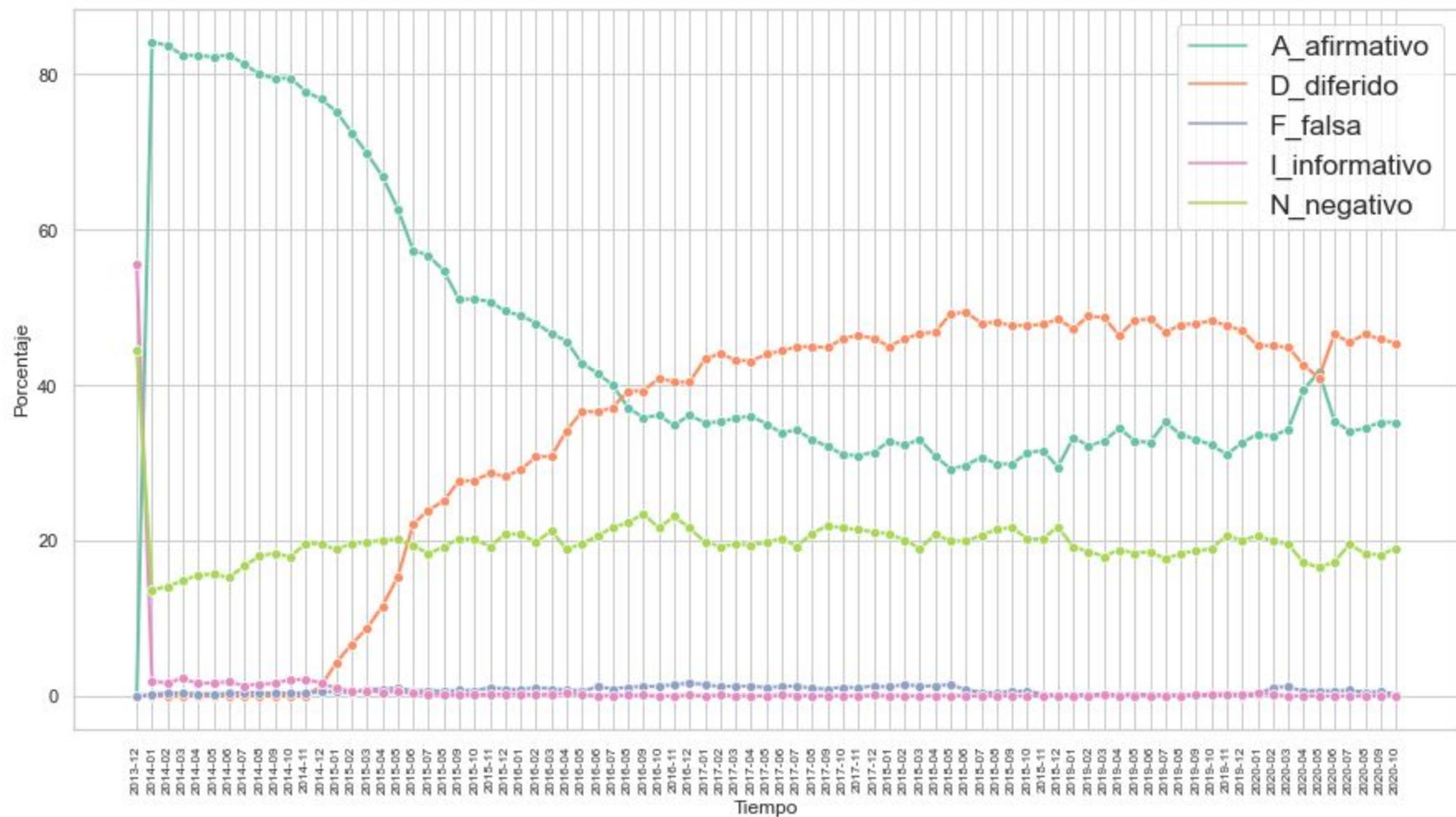


7. ¿Cómo se ha comportado el uso de canales para reportar emergencias a lo largo del tiempo? ¿Existe algún canal que se esté volviendo particularmente relevante?



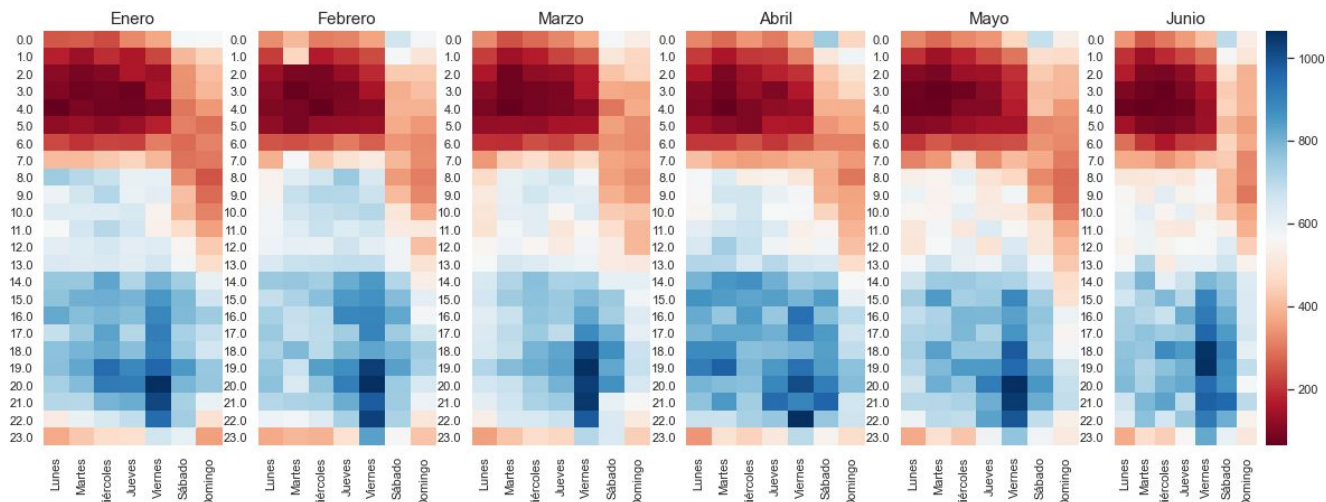
- Se observa que efectivamente el 066 y el 911 es el principal canal de entrada para reportes de incidentes. Debido a que hubo un cambio en el 2016 del 066 al 911, se asumirá que este es un único tipo de entrada.
- Con respecto a los otros canales parece que se han mantenido constantes a lo largo del tiempo, en comparación con los otros tipos de entrada.
- Canales más recientes como el Zello y la APP911, comenzaron a operar y registrar incidentes, sin embargo, comenzando el 2019 decrementaron su uso.

8. ¿Cómo ha evolucionado la proporción llamadas falsas a lo largo del tiempo?



- Con el gráfico anterior observamos que las llamadas falsas, sobre todo las negativas, han tenido un comportamiento estable a lo largo de los años.
- Las llamadas afirmativas o verdaderas, comenzaron con alto porcentaje, sin embargo, al paso del tiempo parece que se estableció en cierto porcentaje.
- Para el reporte diferido parece que aumentó con el paso del tiempo y después se estableció.

Horas

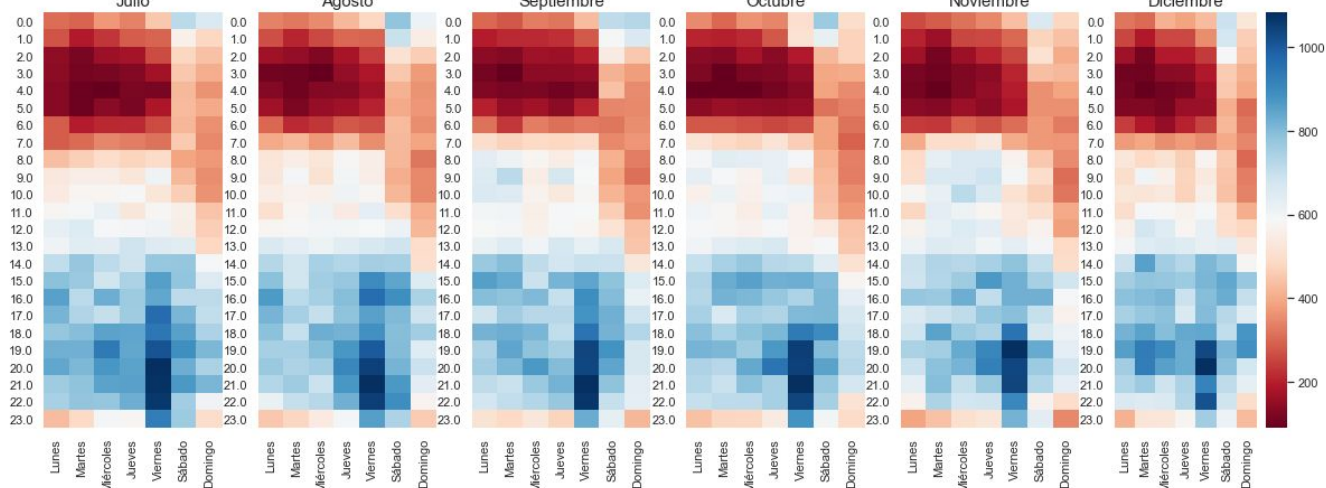


9. ¿Cuáles son los días de la semana y las horas del día en las que más se presentan reportes VERDADEROS?

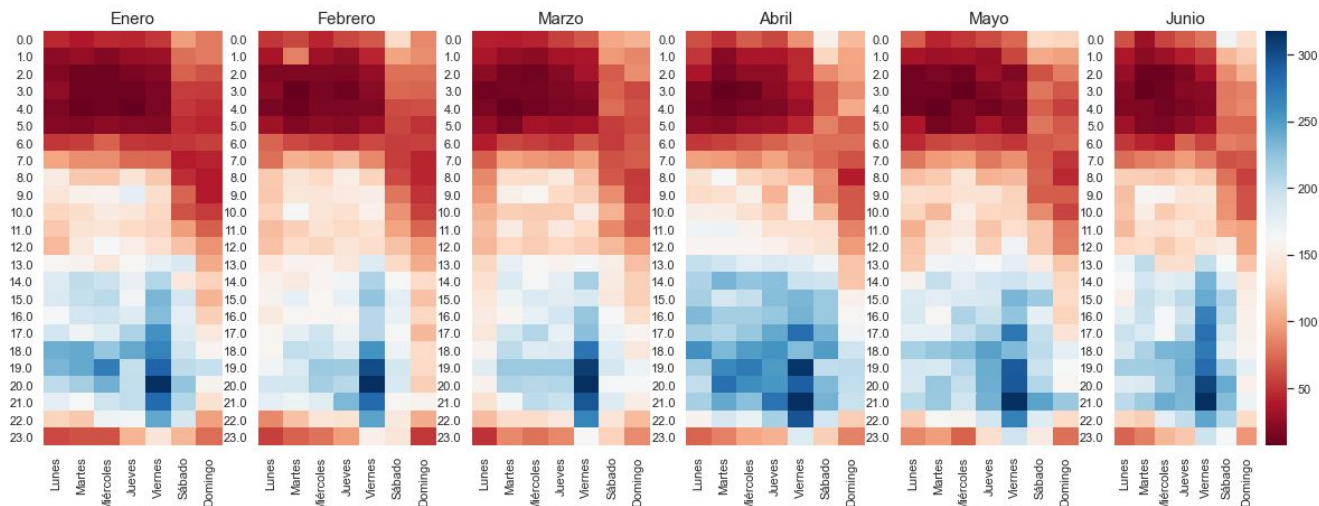
Se observa que en general las horas en las que más se presentan incidentes es en las tardes/noches (entre las 2 pm y las 10 pm).

En cuanto a los días de la semana parece que se distribuye equitativamente a través de la semana, excepto por los viernes que tienen más conteos.

Horas

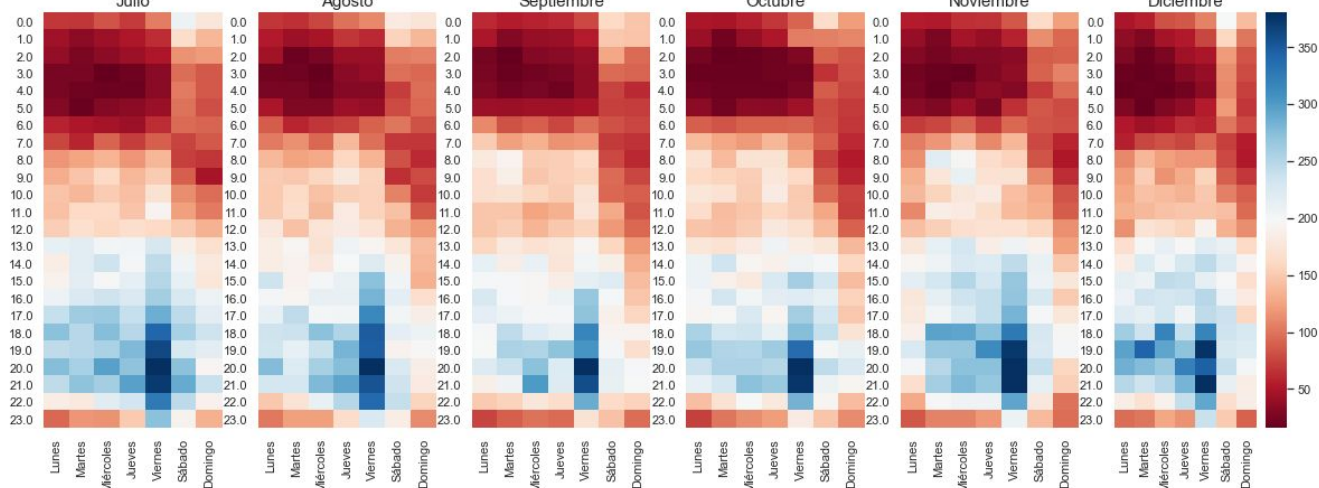


Horas



10. ¿Cuáles son los días de la semana y las horas del día en las que más se presentan reportes FALSOS?

Horas



Se observa una mayor concentración de incidentes falsos en los viernes por la tarde (entre las 5 pm y las 10 pm). En comparación con el gráfico anterior, los conteos aproximados para los reportes falsos se concentran más en dicho intervalo de tiempo.

Conclusiones...

- Las llamadas son el tipo de entrada más común para reportar incidentes.
- El día de la semana y la hora, son features importantes para la clasificación de reportes falsos.
- Geográficamente no se mostró que una delegación particular que tuviera una mayor concentración de reportes falsos. Sin embargo, por las proporciones de reportes falsos, respecto al total, en algunas delegaciones no podemos descartar el feature como importante para el modelo.
- El tipo de incidente parece ser también un feature a considerar en la evaluación de los modelos.