

Priorización de ambulancias de respuesta a incidentes viales en la CDMX

Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5)

Ángel Rafael Ortega Ramírez
José Antonio Lechuga Rivera
Cecilia Avilés Robles

Objetivo

Disminuir los reportes falsos atendidos por ambulancias en la Ciudad de México.

Análisis exploratorio

Se tienen 1,383,138 registros de incidentes viales en CDMX, con las siguientes características:

- Años 2014 a 2020 (`año_cierre`)
- 5 códigos de cierre (`codigo_cierre`)
- 4 clasificaciones de incidentes (`clas_con_f_alarma`)
- 9 medios por el cual se dio aviso del incidente (`tipo_entrada`)
- 26 tipos de incidentes (`incidente_c4`)

- En la base de datos, tenemos alrededor de 1,622 elementos faltantes, distribuidos de la siguiente forma:

Missing Values	
geopoint	446
latitud	443
longitud	435
delegacion_inicio	158
delegacion_cierre	140

Variables categóricas

	Día semana	Código cierre	Año cierre	Mes cierre	Delegación inicio	Incidente c4	Clas con f alarma	Tipo entrada	Delegación cierre	Mes
Tipo	category	category	category	category	category	category	category	category	category	category
Núm. Categorías	7	5	7	12	16	26	4	9	16	12
Núm. Obs.	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138	1,383,138
Obs. nulas	0	0	0	0	158	0	0	0	140	0
% Obs. nulas	0	0	0	0	0.000114233	0	0	0	0.000101219	0
Moda 1 / Veces / %	Viernes	(A)	2018	Octubre	IZTAPALAPA	Accidente choque sin lesionados	EMERGENCIA	LLAMADA DEL 911	IZTAPALAPA	10
	229,681 16.61%	600,108 43.39%	246,577 17.83%	133,641 9.66%	223,687 16.17%	760,506 54.98%	799,037 57.77%	739,339 53.45%	225,202 16.28%	133,641 9.66%
Moda 2 / Veces / %	Sábado	(D)	2019	Agosto	GUSTAVO A. MADERO	Accidente choque con lesionados	URGENCIAS MEDICAS	LLAMADA DEL 066	GUSTAVO A. MADERO	8
	214,134 15.48%	495,631 35.83%	229,467 16.59%	121,123 8.76%	144,620 10.46%	315,031 22.78%	572,619 41.4%	463,432 33.51%	147,582 10.67%	121,123 8.76%
Moda 3 / Veces / %	Jueves	(N)	2017	Septiembre	CUAUHTEMOC	Lesionado atropellado	FALSA ALARMA	BOTÓN DE AUXILIO	CUAUHTEMOC	9
	198,147 14.33%	271,677 19.64%	226,535 16.38%	120,581 8.72%	141,391 10.22%	193,553 13.99%	10,630 0.77%	78,630 5.68%	141,781 10.25%	120,581 8.72%

Variables categóricas (continuación)

- Las variables `delegacion_inicio` y `delegacion_cierre` presentan 158 y 140 valores faltantes, respectivamente*.
- La delegación con mayor número de incidentes registrados es Iztapalapa, con una diferencia de ~6% con la siguiente.
- El tipo de accidente que más se registra es “Accidente, choque sin lesionados”, con un ~55%.

***NOTA:** Los registros con faltantes de alguna de estas dos variables serán imputados con la moda de la categoría correspondiente para el modelo. Dado que representan un % muy bajo del total de los registros, no esperamos tener un gran impacto de sesgo y equidad, pero es importante considerar esta acción al momento de cuantificar estos elementos.



Variables de fecha

Métrica	fecha_creacion	fecha_cierre
Tipo	datetime64[ns]	datetime64[ns]
Núm. de observaciones	1,383,138	1,383,138
Mínimo	31/12/13	01/01/14
Máximo	31/10/20	31/10/20
Núm. de observaciones únicas	2,497	2,496
Núm. de faltantes	0	0
Top1 / veces / %	Timestamp('2020-02-14 00:00:00') 1,163 0.08%	Timestamp('2020-02-14 00:00:00') 1,132 0.08%
Top2 / veces / %	Timestamp('2018-10-26 00:00:00') 1,028 0.07%	Timestamp('2017-12-02 00:00:00') 999 0.07%
Top3 / veces / %	Timestamp('2019-11-30 00:00:00') 1,011 0.07%	Timestamp('2017-12-08 00:00:00') 999 0.07%

- Primera fecha con registro: 31/diciembre/2013.
- Última fecha con registro: 31/octubre/2020.
- En el periodo del que se tienen registros, el 14/febrero/2014 es la fecha en la que más incidentes registrados hubo.

Variables geoespaciales

	Latitud	Longitud
tipo	float64	float64
Núm. de observaciones	1,382,695	1,382,703
Media	19.38	-99.14
Mínimo	19.09	-991.76
Máximo	195.3	-98.95
Núm. de observaciones únicas	82,501	78,984
Núm. de faltantes	443	435
Top 1 / Veces / %	19.30431996 2,173 0.16%	-99.08024004 2,162 0.16%
Top 2 / Veces / %	19.37168001 2,162 0.16%	-99.08714016 2,100 0.15%
Top 3 / Veces / %	19.347020999999998 1,730 0.13%	-99.180646 1,731 0.13%

- Hay 443 y 435 registros donde faltan `latitud` y `longitud`, respectivamente.*
- Se aprecia que hubo “errores de captura” (ej. punto decimal incorrecto) al registrar los incidentes. Se tienen 13 registros capturados incorrectamente:
 - 3 de latitud
 - 10 de longitud
- En éstos, su coordenada complementaria no está capturada.

***NOTA:** Los registros con faltantes de alguna de estas dos variables no serán tomados en cuenta para el modelo ya que representan un % muy bajo del total de los registros.

Variables de texto

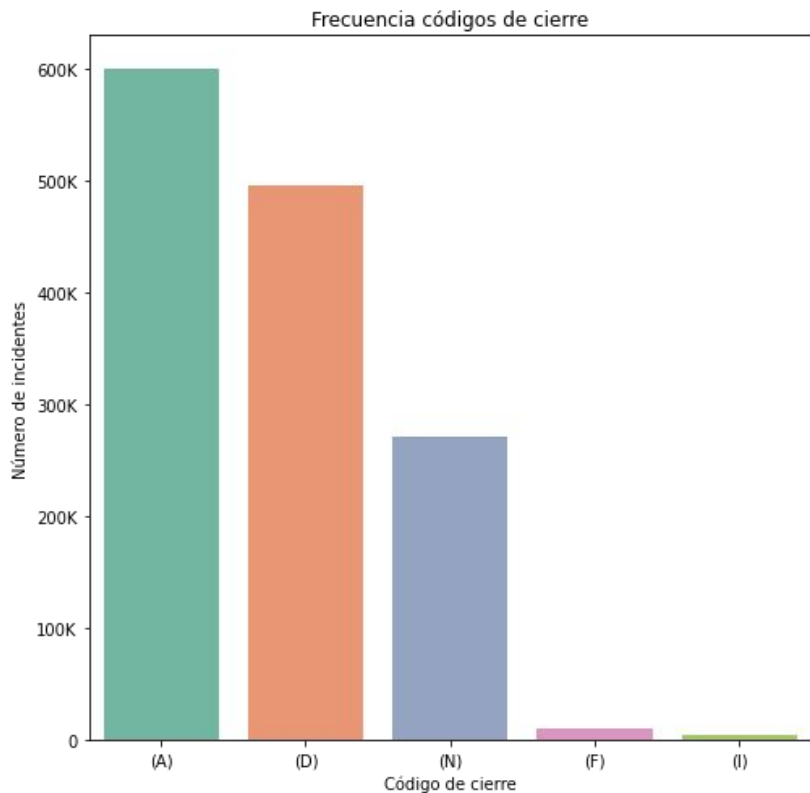
	Folio	Geopoint
Tipo	object	object
Núm. de observaciones	1,383,138	1,383,138
Observaciones únicas	1,383,138	134,288
% Observaciones únicas	1	0.097
Tamaño promedio	15	22.5
Tamaño mínimo	15	14
Tamaño máximo	15	24

- La variable de `folio` identifica a cada registro con un valor único.
- El folio siempre se construye de 15 caracteres.
- En cuanto a la variable `geopoint` (par de coordenadas) se tienen 134,288 ubicaciones distintas registradas

Variables numéricas

- Para este análisis, las variables `latitud` y `longitud` fueron consideradas como “geoespaciales”, por lo que no se cuentan con variables numéricas
- Adicionalmente, la variable `mes` fue considerada categórica.

Códigos de cierre

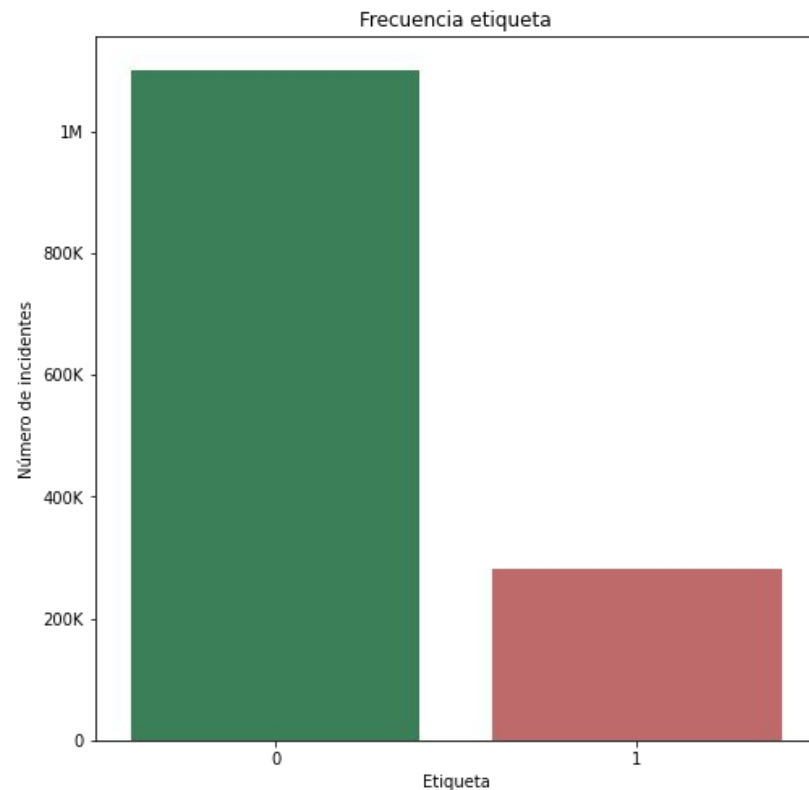


Se cuentan con 5 códigos de cierre:

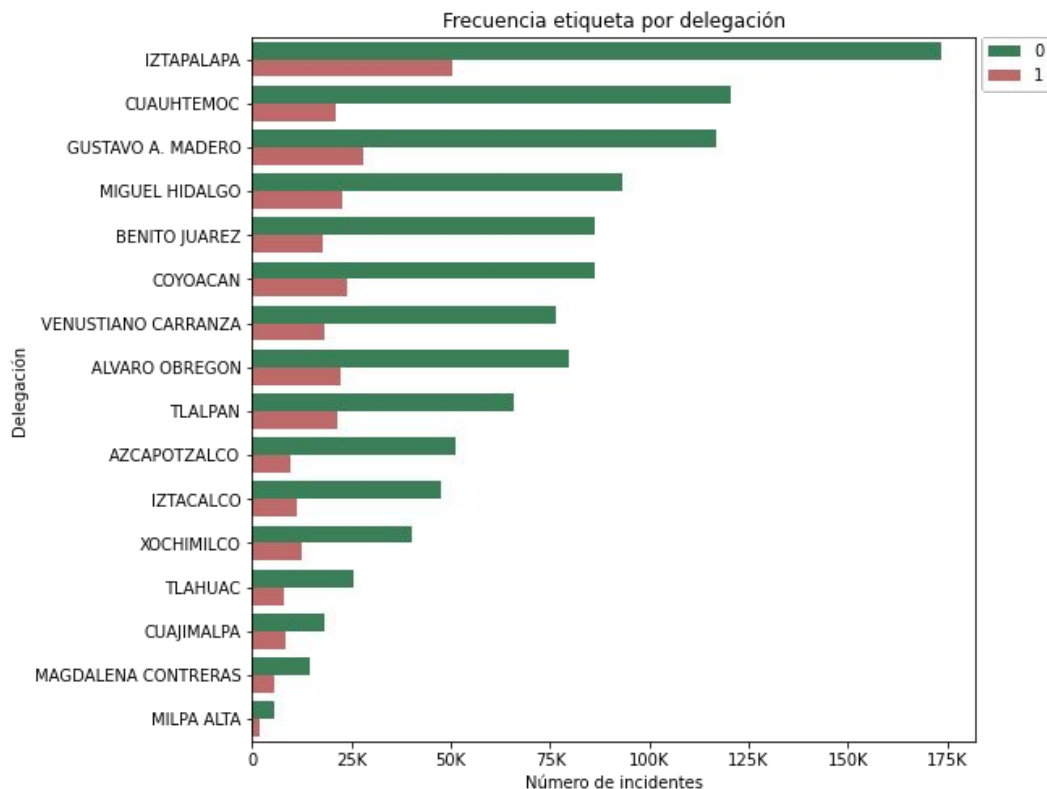
- **(A):** La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos y confirmó la emergencia reportada
- **(D):** El incidente reportado se registró en dos o más ocasiones procediendo a mantener un único reporte (afirmativo, informativo, negativo o falso) como el identificador para el incidente
- **(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
- **(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
- **(I):** El incidente reportado es afirmativo y se añade información adicional al evento

Frecuencia etiqueta

- Se crea una etiqueta para clasificar los datos de la siguiente forma:
 - 1, cuando el código de cierre es (F) o (N)
 - 0, en otro caso
- **282,303** registros clasificados como 1 (falsos)
- **1,100,835** registros clasificados como 0 (verdaderos)



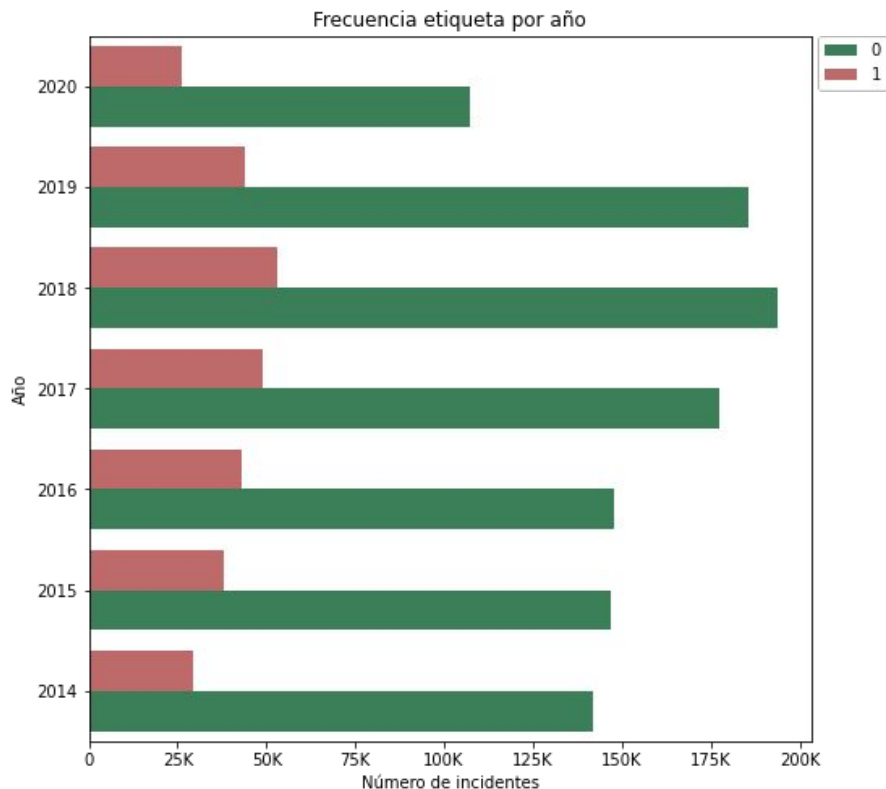
Etiqueta por delegación



- Principales delegaciones con mayor número de incidentes clasificados tanto “1” como “0” son:

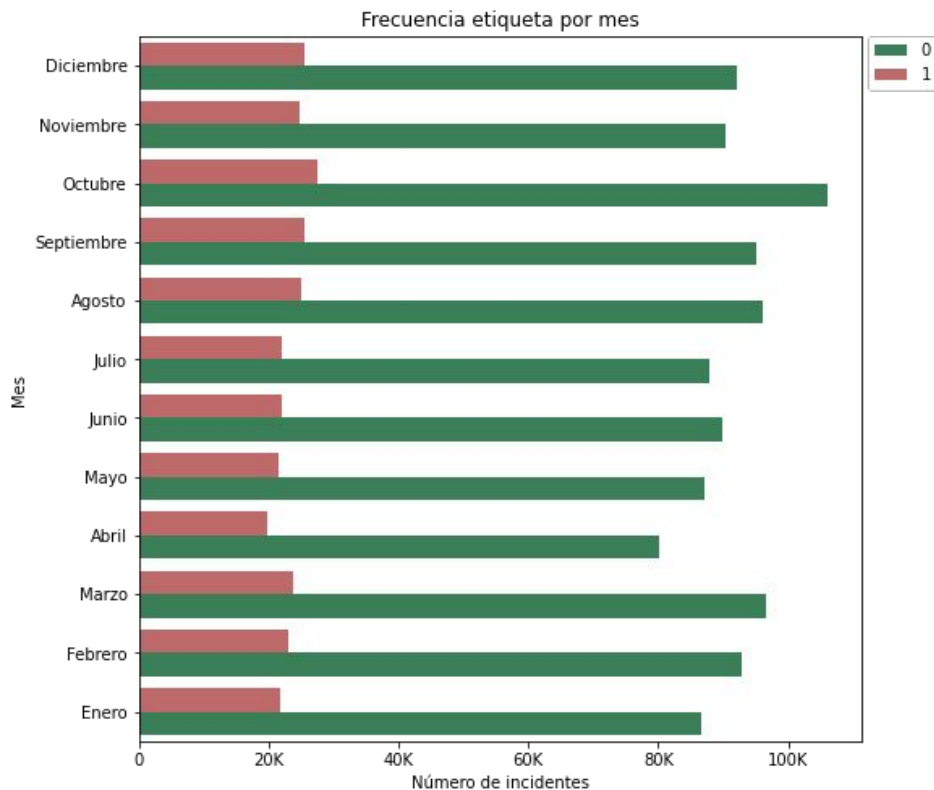
- Iztapalapa
- Cuauhtémoc
- Gustavo A. Madero

Etiqueta por año



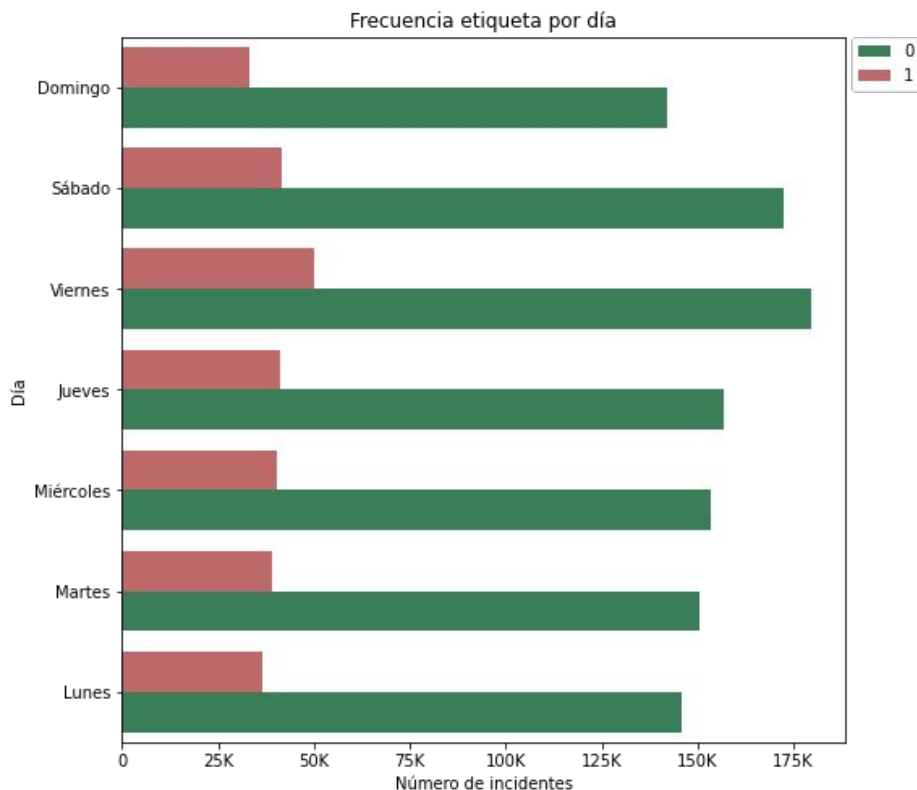
- 2018: año con mayor número de incidentes registrados (y clasificados por etiqueta)
- 2017, 2018 y 2019 presentan considerable mayor número de incidentes registrados en contraste al resto de los años.
- 2020: se nota el efecto de la pandemia por COVID19, pues a pesar de que al corte de captura faltan 2 meses de registros el número total es considerablemente menor al de los otros años.

Etiqueta por mes



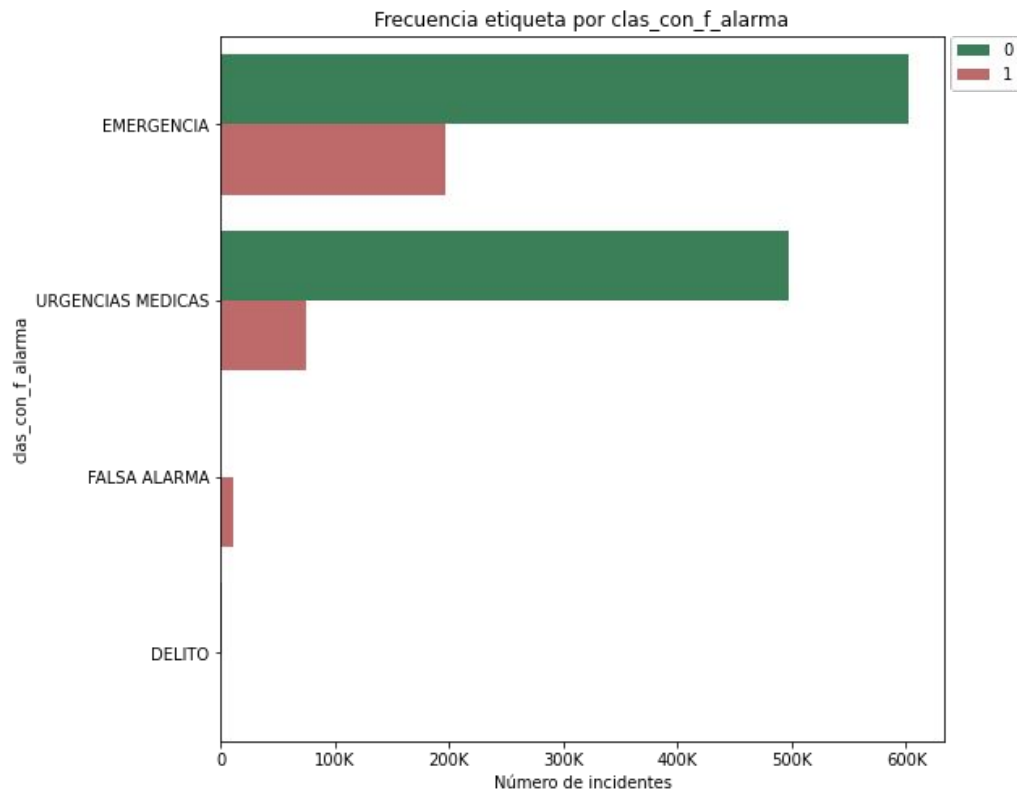
- El mes con mayor número de incidentes registrados es Octubre.
- No se aprecia alguna distribución específica de la etiqueta por mes de clasificación.

Etiqueta por día de la semana



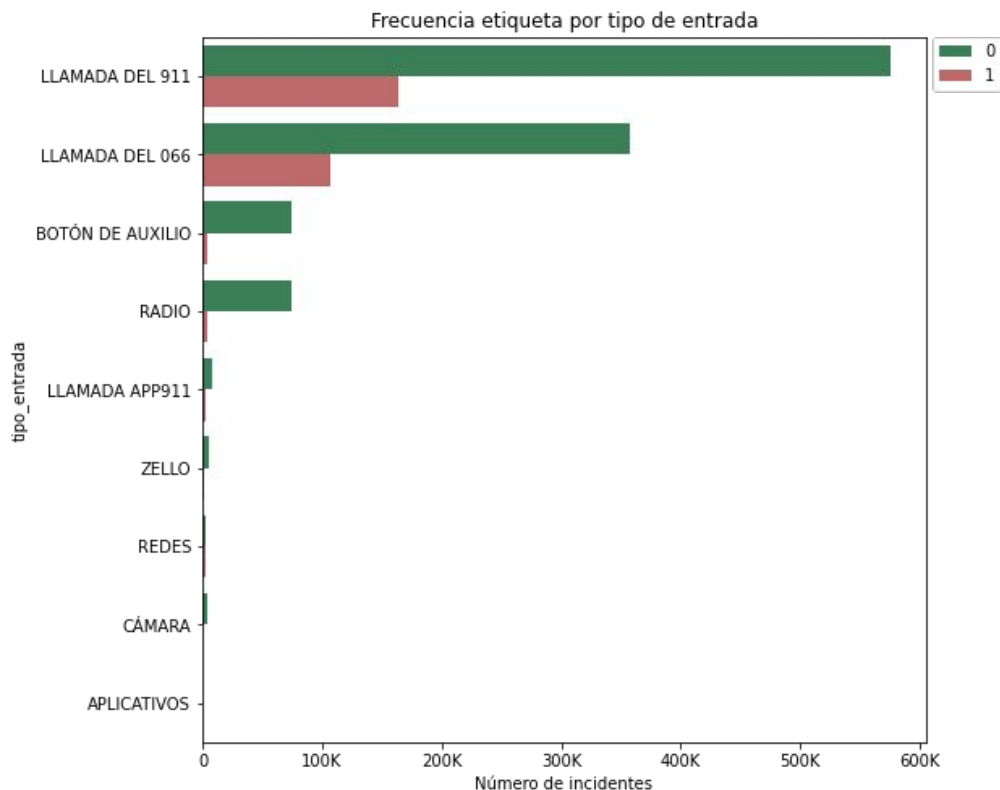
- Días con mayor número de registros:
 - Viernes
 - Sábado
 - Jueves
- Se aprecia que conforme se acerca el fin de semana, los incidentes registrados aumentan.

Etiqueta por clasificación del incidente



- El número de incidentes `clas con f alarma` clasificados como “FALSA ALARMA” (10,630) está relacionado con el número de `codigo de cierre` clasificados como (F) (10,626).

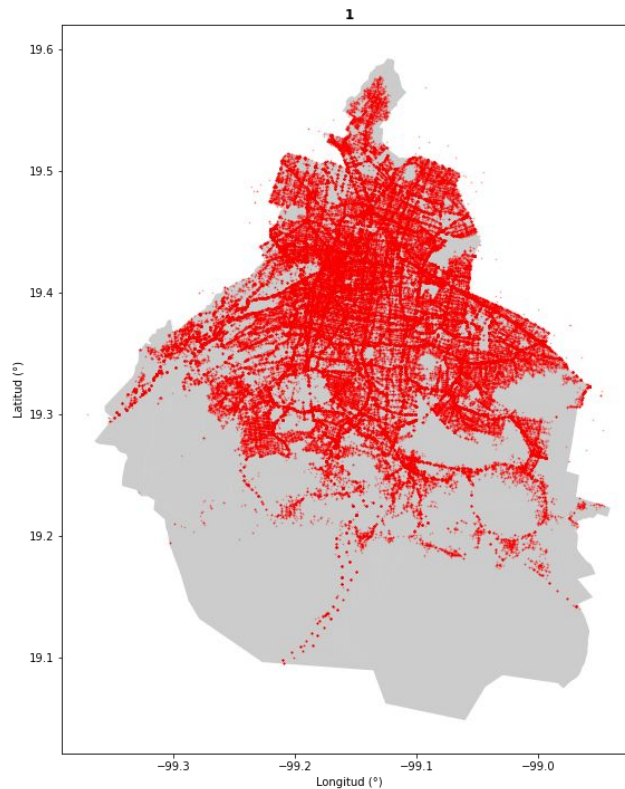
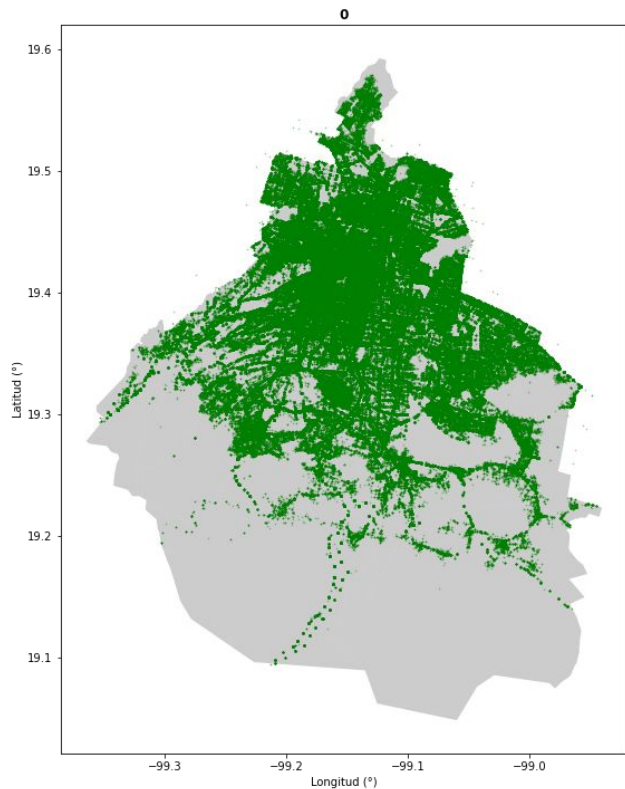
Etiqueta por tipo de entrada



- Los tipos de entrada que involucran “llamadas” (del 911, del 066 y app911) son los que presentan mayor número de registros clasificados como “1” (falsos).
- El resto de tipos de entrada no muestran un número considerable de registros clasificados como “1” (falsos).

Etiqueta en el mapa

Etiqueta en el mapa



- Los incidentes registrados se distribuyen de forma similar con la clasificación de la etiqueta.
- A pesar de usar el parámetro α , no se aprecia alguna zona con mayor concentración de registros.

Preguntas selectas de proyecto

Variables a eliminar

La siguiente lista de variables serán eliminadas debido a que no contaremos con ellas al momento de la predicción:

- `codigo_cierre`
- `fecha_cierre`
- `año_cierre`
- `mes_cierre`
- `hora_cierre`
- `clas_con_f_alarma`
- `delegacion_cierre`

Variables relevantes a crear

Adicional a las transformaciones y limpieza de datos, se crearon las siguientes variables al considerarse de alto interés:

1. Time stamp de creación: se genera a partir de los datos de la fecha y hora de creación del reporte de incidente
2. Tipo de día: se generó la variable binaria `day_type` la cual toma un valor de 1 si el día que se genera el incidente es jueves, viernes o sábado (días de alta demanda histórica) o 0 en caso contrario
3. Trimestre: con el objetivo de reducir la dimensionalidad, se crearon las variables `sin_trim` y `cos_trim` a partir del mes de creación.
4. Llamada: para reducir dimensionalidad, a partir de la variable `tipo_entrada`, se creó una variable binaria que toma el valor de 1 si la entrada fue por llamada y 0 de lo contrario.
5. Tipo de incidente: la variable `incidente_c4` se redujo a las siguientes 5 categorías para reducir dimensionalidad: **accidente**, **cadáver**, **sismo**, **mi ciudad**, **detención ciudadana** y **lesionado**.

Resultados

La lista de variables a mantener es:

- Llamada
- Accidente

La respectiva importancia de cada feature al final fue:

Importancia	Feature
0.5369	Llamada
0.1207	Accidente