

Plataforma de Seguridad del Municipio de Chía |

Chía | Equipo #34 | LUNARLAB

Álvarez Andrés, Bermudez Camila, Cadena Laura, Castellanos Diego, Higuaran Alejandro, Rubiano Sebastián, Zarabanda Alvaro

PROBLEMA EMPRESARIAL

La Alcaldía Municipal, a través de distintas fuentes, dispone de un conjunto de datos relevantes para los que es necesario realizar un análisis cruzado que permita identificar, estratificar, segmentar y generar espacios de condiciones ecosistémicas, económicas y de seguridad social para el municipio sustentado en condiciones que permitan la retroalimentación y toma de decisiones a partir de planes para establecer programas y estrategias. La percepción de riesgo de seguridad de la población del municipio es del 41% lo cual es un valor alto y representa uno de los mayores problemas para el municipio.

IMPACTO EMPRESARIAL

Desde el contexto gubernamental y social incide en mejorar las condiciones para programas especiales orientados a poblaciones vulnerables y desde el aspecto de seguridad mejorará la percepción tomando en cuenta las medidas necesarias para lograr un municipio más seguro. Además le permitiría al municipio convertirse en un referente a nivel tecnológico y de administración dentro de los municipios de la Sabana

DATOS

1. **ENTIDAD_ATIENDE_EMERGENCIAS:** Registros de entidades y personas atendidas.
2. **ATENCION_TIPO_POLICIA:** Registros de situaciones en el departamento de policía.
3. **CASOS_MOVILIDAD:** Incidencias reportadas por tránsito.
4. **USUARIOS_REGISTRADOS:** Usuarios registrados en la aplicación 'ventanilla única virtual'.
5. **EMERGENCIAS_CHIAPP:** Casos generados a través de la aplicación CHIAPP, que permite identificar las principales zonas donde se presentan emergencias en el municipio.



- 6. TIPO_RECEPCION_EMERGENCIAS:** Número de emergencias reportadas por día, agrupadas por el canal por el cual fueron reportadas.
- 7. TIPOS_EMERGENCIAS_REPORTADAS:** Tipos de emergencias notificadas al centro de emergencias en un periodo de tiempo.
- 8. REP_CAMARAS_EMERGENCIA:** Casos reportados o identificados por cámaras de seguridad.
- 9. ZONIFICACION_CHIA:** Zonificación del municipio en base a tipo de zona, región zonal y barrio/aldea.
- 10. LLAMADAS123:** Llamadas realizadas al centro de emergencias.
- 11. FOES:** Referencia a trámites específicos del “Fondo de educación Superior”.

Cada uno de los once conjuntos de datos tiene sus atributos correspondientes como se muestra a continuación:

Tabla	ENTIDAD_ATIENDE_EMERGENCIAS	
Etiqueta	Tipo	Descripción
FECHA	Fecha	Fecha de la emergencia
ENTIDAD	Cadena	Nombre de la institución
C_ATENDIDOS	Entero	Número de personas atendidas en la emergencia
C_NO_ATENDIDOS	Entero	Número de personas no atendidas en la emergencia
TOTAL_CASOS	Entero	Número total de personas involucradas en la emergencia

Tabla	ATENCION_TIPO_POLICIA	
Etiqueta	Tipo	Descripción
FECHA	Fecha	Fecha de la situación
CASOS	Cadena	Tipo de la situación
CANTIDAD	Entero	Suma total de situaciones reportadas



Tabla	CASOS_MOVILIDAD	
Etiqueta	Tipo	Descripción
FECHA	Fecha	Fecha del caso
CASO	Cadena	Tipo del caso
CANTIDAD	Entero	Suma del total de casos reportados

Tabla	USUARIOS_REGISTRADOS	
Etiqueta	Tipo	Descripción
FECHA_CREACIÓN	Fecha	Fecha de creación
TIPO_DOCUMENTO	Cadena	Tipo de documento
FECHA:NACIMIENTO	Fecha Fecha	usuario
FECHA_EXPEDICIÓN	Fecha	expedición
GENERO	Cadena	Género del usuario
ESTADO_CIVIL	St Anillo	Usuario Estado civil
GRUPO_POBLACIONAL	Cadena	Grupo de población
TIPO_ZONA	Cadena	Tipo de zona
SECTOR	Cadena	Sector
DIRECCIÓN	Cadena	Dirección del informe
ZONA	Cadena	Zona
REGIÓN	Cadena	Región
MUNICIPIO_EXPEDICIÓN	Cadena	Municipio de expedición



Tabla	EMERGENCIAS_CHIAPP	
Etiqueta	Tipo	Descripción
ID_EMERGENCIA	Entero	Identificador único de emergencia
LATITUD	Doble	Coordenada de latitud
LONGITUD	Doble	Coordenada de longitud
FECHA	Fecha	Fecha de ocurrencia de

Tabla	TIPOS_EMERGENCIAS_REPORTADAS	
Etiqueta	Tipo	Descripción
FECHA	Fecha	Fecha de emergencias reportadas
CASOS	String	Tipo de emergencia
CANTIDAD	Entero	Suma de emergencias reportadas por tipo para cada fecha.

Tabla	REP_CAMARAS_EMERGENCIA	
Etiqueta	Tipo	Descripción
MES	Cadena	Mes en que se detectó la emergencia y se reportó en una cámara
FECHA	Fecha	Fecha en que se reportó la emergencia
HORA	Hora	Hora en que se reportó la emergencia
CÁMARA	Cadena	Nombre de la cámara en que se detectó la emergencia
ENTIDAD QUE ATIENDE	Cadena	Institución a la que se reportó la emergencia
CATEGORÍA	Cadena	Categoría de la emergencia



DESCRIPCIÓN	Cadena	Descripción detallada de la emergencia
RESULTADO CASO	Cadena	Resultado del procedimiento desencadenado por el reporte
TIPO CASO ATENCIÓN	Cadena	Tipo de atención que se brindó la emergencia

Tabla	ZONIFICACION_CHIA	
Etiqueta	Tipo	Descripción
ID SECTOR	Entero	Identificador único de sector
SECTOR	Cadena	Nombre del sector
ID ZONA	Entero	único de zona
ZONA	Cadena	Nombre de la zona
ID_TIPO_ZONA	Entero	Tipo de identificador único de zona
TIPO_ZON	Cadena	Tipo de zona
ID_REGION	Entero	Identificador único
REGION	Cadena	Nombre de la región

Tabla	LLAMADAS123	
etiqueta	Tipo	Descripción
MA RCA TEMPORAL	Fecha	Fecha desde la llamada
DIRECCIÓN	Cadena	Dirección del evento
ENTIDAD_ATIENDE	Cadena	Entidad que atiende la llamada
TIPO DE EMERGENCIA	Cadena	Tipo de emergencia de la llamada
DESCRIPCIÓN	Cadena	Descripción de la llamada
ENTIDAD	Cadena	Entidad responsable de la llamada



SEGUIMIENTO CASO	DEL	Cadena	Pista de la case
---------------------	-----	--------	------------------

Tabla	FOES	
Etiqueta	Tipo	Descripción
FECHA_CREACIÓN	Fecha Fecha	creación
ID_TRÁMITE	String	Identificador único del procedimiento
FECHA_NACIMIEN TO	Fecha Fecha	de nacimiento del usuario
ID_TIPO_TRÁMITE	Integer	Tipo de procedimiento identificador único
TRAMITE	String	Descripción del procedimiento
ID_ESTADO	Integer	procedimiento Identificador único
ESTADO	String	Descripción del estado del procedimiento

Ventajas:

- Proporciona información rica sobre los usuarios, trámites y emergencias que ocurren en la ciudad en tiempo real.
- El esquema del conjunto de datos está bien definido por la entidad, lo que podría hacer que el análisis sea más comprensible y es un buen punto de partida.
- Las emergencias se ubican geoespacialmente y en el tiempo para que podamos analizar los datos por tiempo (Año, mes, día de la semana, período) y por ubicación.
- La información sobre qué entidad atendió a las víctimas es por día, por lo que nos permitirá entender los picos de servicio y evidenciar la falta de personal en caso de ser necesario
- Las características categóricas parecen tener un formato bien establecido
- Las fechas parecen tener un formato , como '%d/%m/%Y' o día/mes/año en todas las tablas, lo que nos permite analizar claramente los datos a lo largo del tiempo.

- La mayoría de los usuarios registrados tienen su fecha de nacimiento, esto definirá la población que necesita fortalecer la necesidad de reportar incidentes.

Desventaja:

- El conjunto de datos contiene relativamente pocos, solo tres puntos de datos, que es una tabla 'USUARIO' de tamaño de muestra muy pequeño
- No se cuenta con una identificación única que nos proporcione una clave para unir procesos de datos y tablas.
- En la tabla 'EMERGENCIAS' hay valores de longitud y latitud que no corresponden dentro de los límites de Chía y podrían generar problemas al momento de realizar los reportes.
- Hay columnas con una cantidad significativa de valores NULL y representan un desafío durante el proceso de interpolación.
- Las cámaras están listadas pero no están identificadas con una identificación única ni tienen información geoespacial para ubicarlas con precisión, lo que ayudaría a analizar su efectividad o inferir si hay una falta de cámaras en diferentes áreas de Chía.
- La zonificación es en formato tipo texto y no con las respectivas coordenadas, divisiones o límites, sería útil que la información esté en algún formato que permita extraer su geometría y forma para complementar los datos mediante análisis espacial.
- La tabla 'llamadas123' debe estar limpia, contiene mucha información sobre los casos, pero la mayoría de ellos solo contienen comentarios. Es necesario conocer los códigos policiales para lograr una mejor clasificación de los casos.
- No existe un esquema o diagrama de relaciones que nos permita unir la información de las diferentes tablas y realizar un análisis más profundo.
- Hay pocos registros recientes sobre el canal donde se reportaron las emergencias, por lo que no vamos a poder definir con certeza estrategias para impulsar la comunicación por otro tipo de canales

- . En el conjunto de datos de llamadas123, hay varias llamadas donde se clasificó la emergencia. como “OTRAS”, por lo que puede inducir a error en nuestras conclusiones, o vamos a tener que reclasificar algunas dependiendo de la descripción de la situación de la llamada
- La mayoría de los usuarios registrados no tienen estado civil y su zona. Sería un desafío establecer campañas de financiación de herramientas

MÉTODOS

EDA

Un componente clave para comprender de Chia es proporcionar visualizaciones de alta calidad de su información criminal histórica e identificar características críticas en los informes policiales. visualizaciones interactivas que proporcionaremos:

Tabla: Usuarios registrados

- Histogramas que perfilan a los usuarios de Chiapp (edad, región, género y grupo poblacional)(univariante)

Tabla: Entidad atiende emergencia

- Ratio mensual de emergencias atendidas por entidad

Tabla: Atención tipo policia

- Un gráfico de barras que muestra mensualmente la proporción de emergencias atendidas por cada entidad y los tipos de situaciones reportadas

Tabla: Casos movilidad

- Gráfico de barras que muestra mensualmente la proporción de casos de movilidad reportados

Tabla: Emergencias Chiapp

- Heatmap donde ocurrieron las emergencias para identificar hotspots por tipo de emergencia (univariante)

Tabla: Tipo recepción emergencias

- Un gráfico de flujo para darte cuenta de la evolución del uso de canales para reportar incidentes (univariante)

Tabla: Tipo de emergencias reportadas



- Gráfico de barras que muestra mensualmente la proporción de situaciones reportadas dada su naturaleza

Tabla: Rep cámaras emergencia + Llamadas 123

- Tabla cruzada de la entidad que atendió la emergencia y el tipo de solución dada y su clasificación de emergencias reportadas por cámaras de seguridad
- Un gráfico de barras que muestra mensualmente la proporción de emergencias atendidas por cada entidad y los tipos de situaciones reportadas
- Histograma para definir cuándo son las emergencias reportadas principalmente (hora y día) (univariante)
- Line plots de la cantidad de delitos reportados mensualmente para identificar estacionariedad (univariante)
- Un mapa para identificar dónde se encuentran las cámaras de seguridad incluyendo la capa de zonas del municipio

ANÁLISIS DE DATOS Y HALLAZGOS

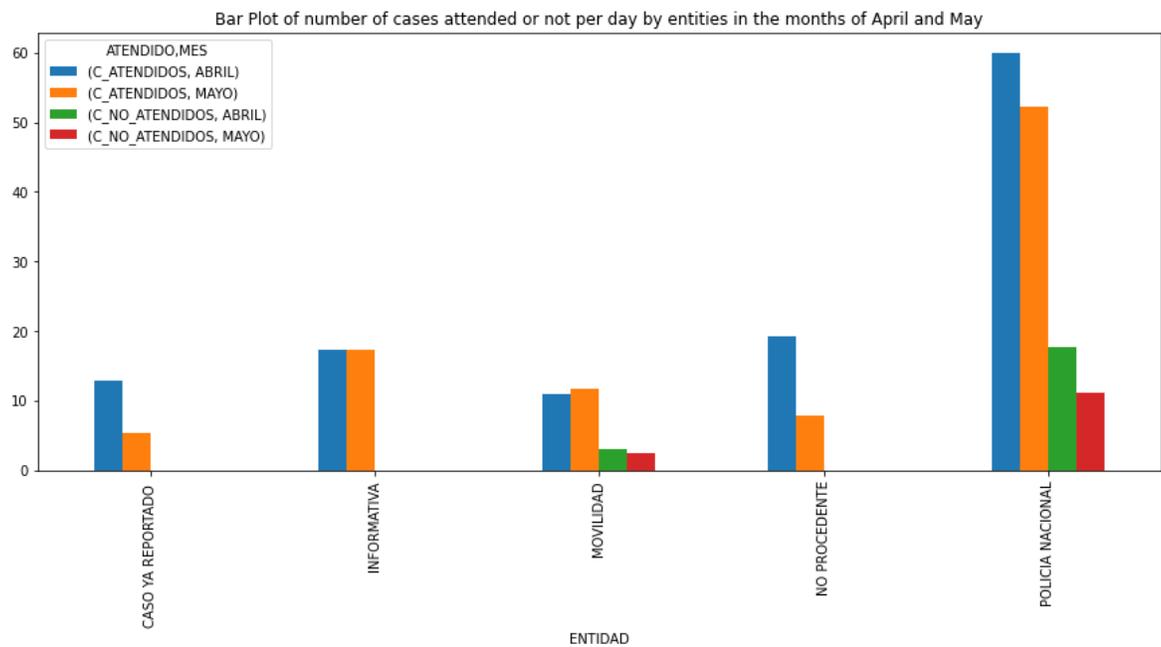
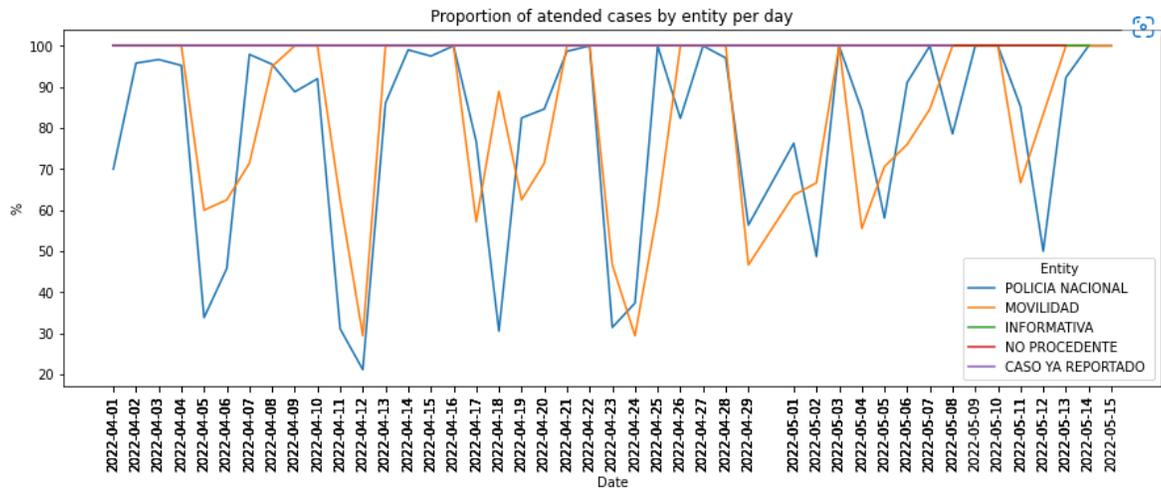
1. ENTIDAD_ATIENDE_EMERGENCIAS:

Número de registros: 810

Análisis gráfico:

La tabla tiene información sobre el número de casos atendidos o no atendidos por diferentes entidades (18 entidades más “otras” entidades valor ty) en los meses de abril y mayo de 2022. El

86% de las emergencias reportadas entre el 01-04-2022 y el 15-05-2022 fueron asignadas a 5 *entidades*. El 48% de los casos fueron asignados a la Policía, el otro 9% fueron asignados a *movilidad*. Algunas categorías no están relacionadas directamente con una entidad como INFORMATIVA, NO PROCEDENTE y CASO YA REPORTADO, por lo que podemos decir que la mayoría de los casos son asignados a la Policía. Según el gráfico de líneas, la Policía ha atendido más casos denunciados al final del período observado que al principio. Hubo un pico el 2022-04-12, cuando la policía asistió a poco más del 20% de los casos denunciados. Podría ser interesante revisar qué sucedió con el personal de la Policía durante los picos para averiguar si tenían poco personal en esos días.

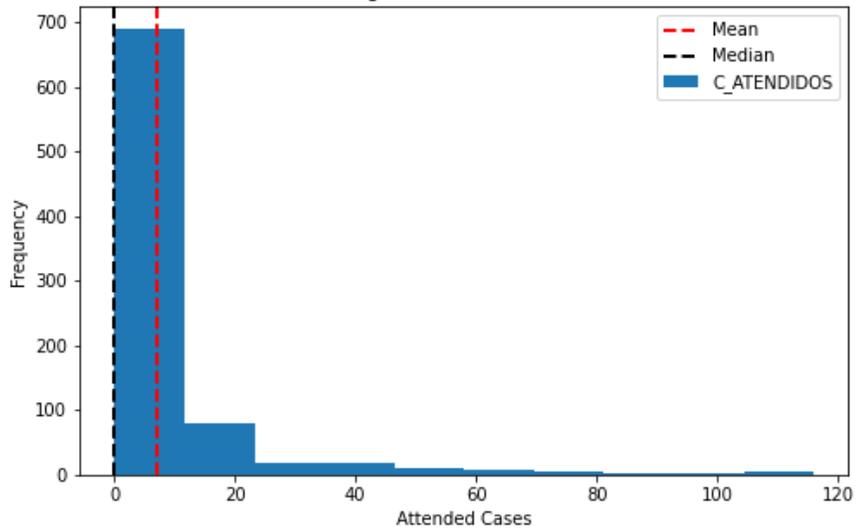


Procedemos a ver las diferencias en el número de casos atendidos y no atendidos y observamos la cobertura general de las diferentes entidades, esto lo haremos observando el porcentaje de casos atendidos sobre el total de casos de las diferentes entidades durante el mes de abril y mayo. Hay días donde no hubo casos y tomaremos el porcentaje como 100

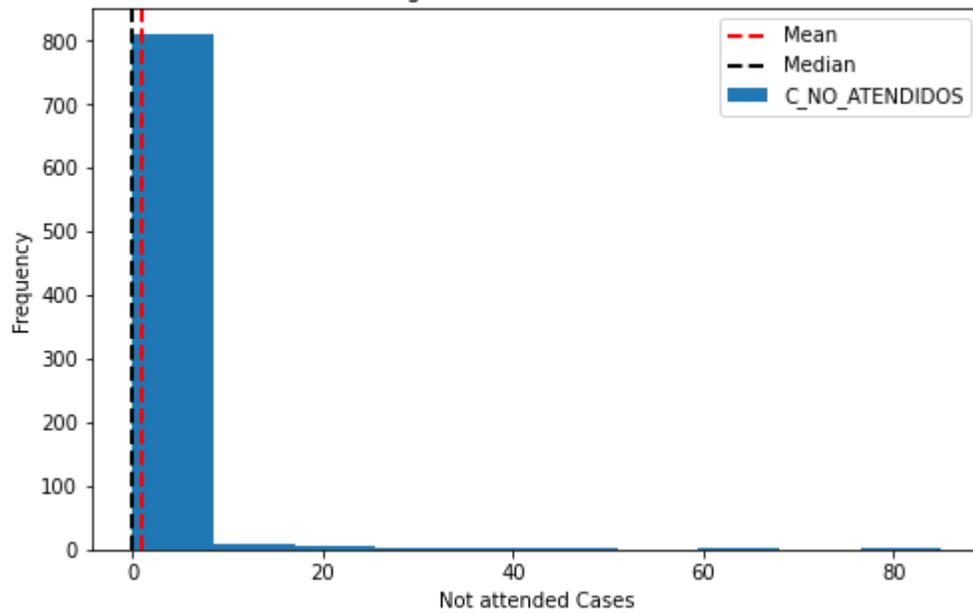
Vemos la distribución de casos atendidos, casos no atendidos, total de casos y desempeño.

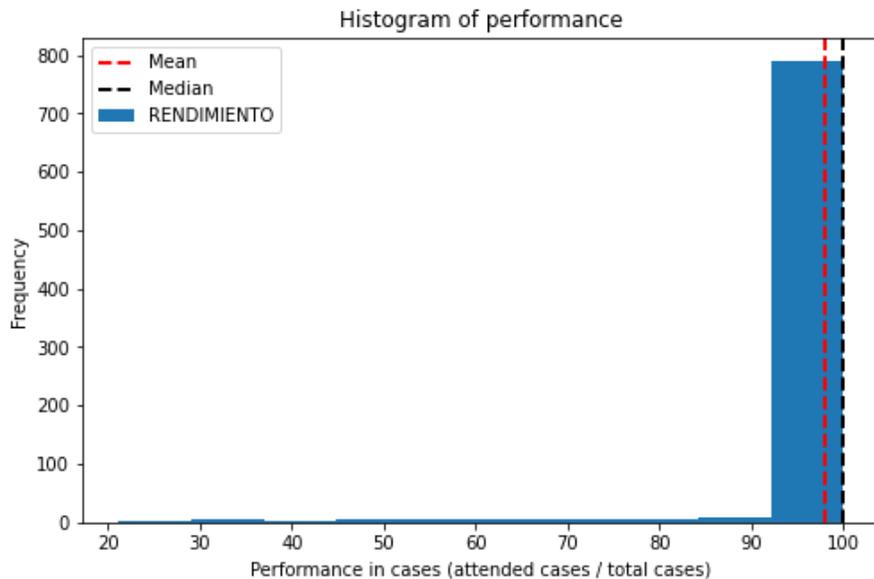
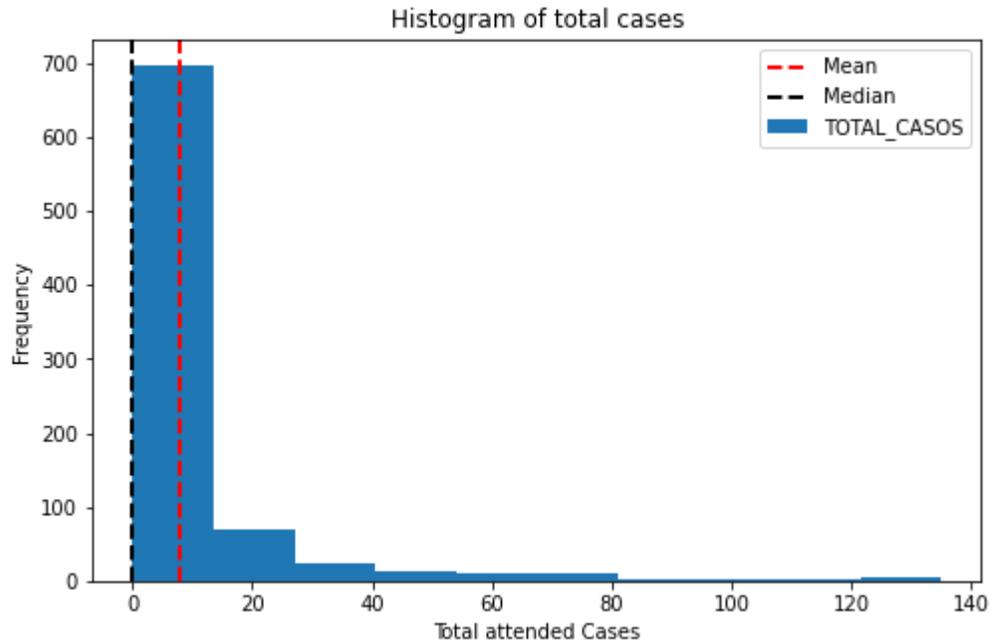


Histogram of attended cases



Histogram of no attended cases

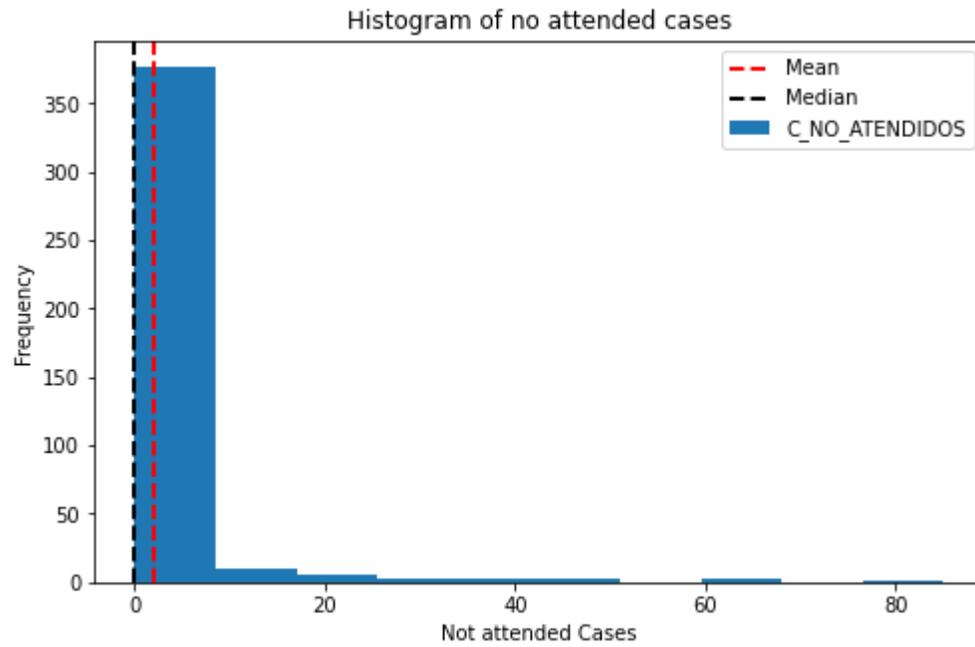
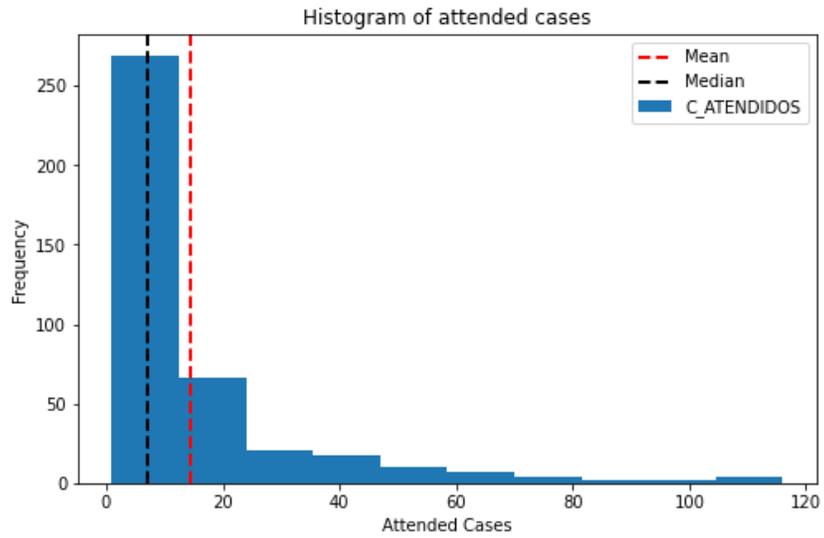


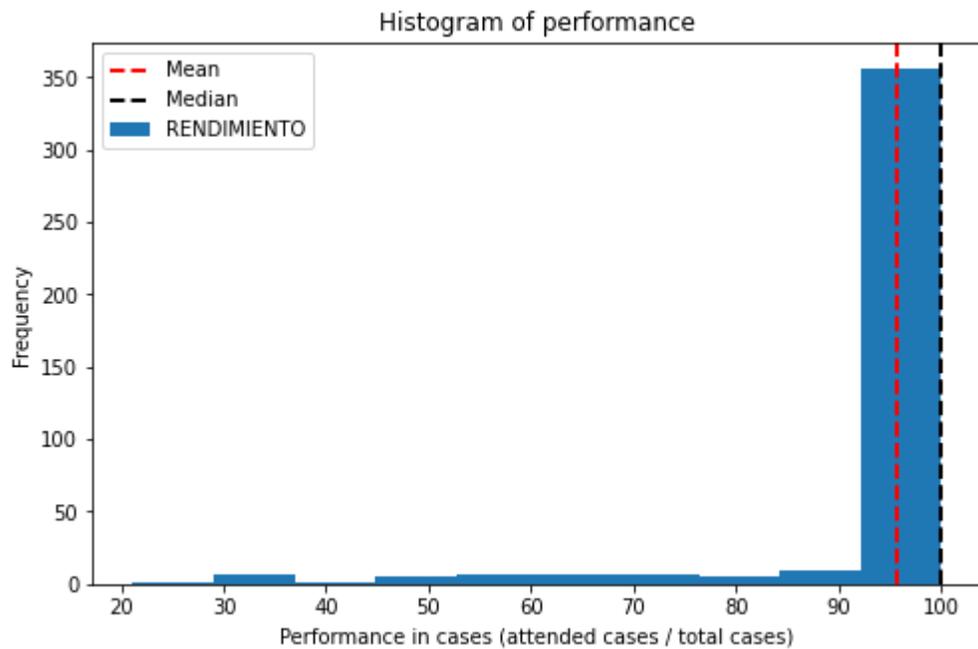
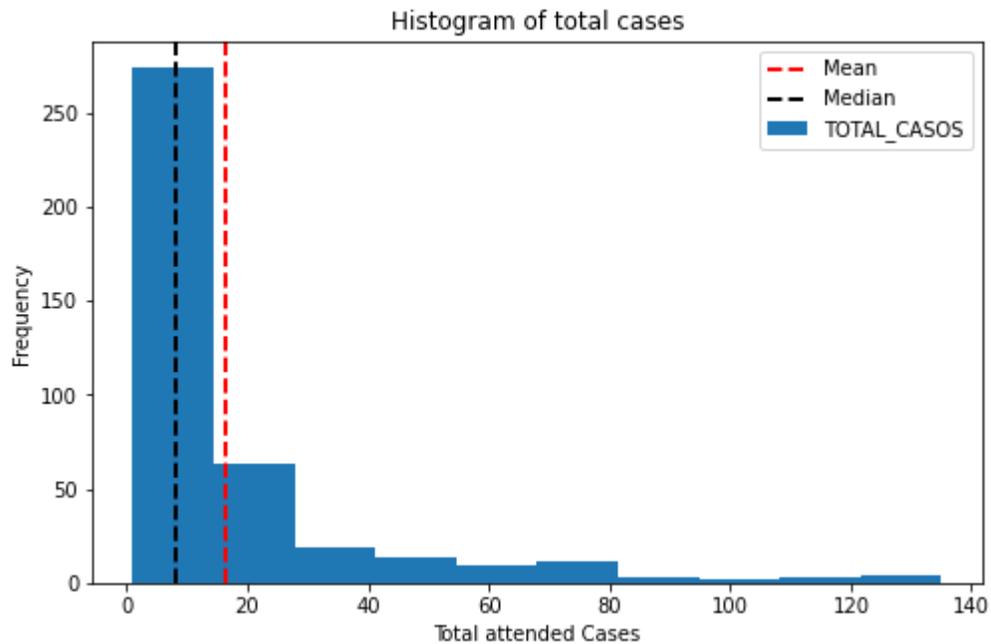


Como podemos ver, la distribución de los casos desatendidos, atendidos y totales están sesgadas a la derecha, y en el caso de la actuación a la izquierda. Se puede ver que así, la mayoría de las veces no hay casos o el número es muy mínimo, pero cuando los hay, su número es mayor. Esto puede ocurrir por otra razón, quizás los casos se acumulan y al cabo de ciertos días ya se reciben en las distintas oficinas. Así como tomamos el retorno de 100 para los días en los que no hay casos, vemos que la distribución está sesgada hacia la izquierda.



Veamos el desempeño cuando no hay casos





Sin contar los registros donde no hubo casos, podemos ver que las distribuciones se mantienen. Aunque la mediana es 100, la media es 95 y se ven valores hasta por debajo del 50%.

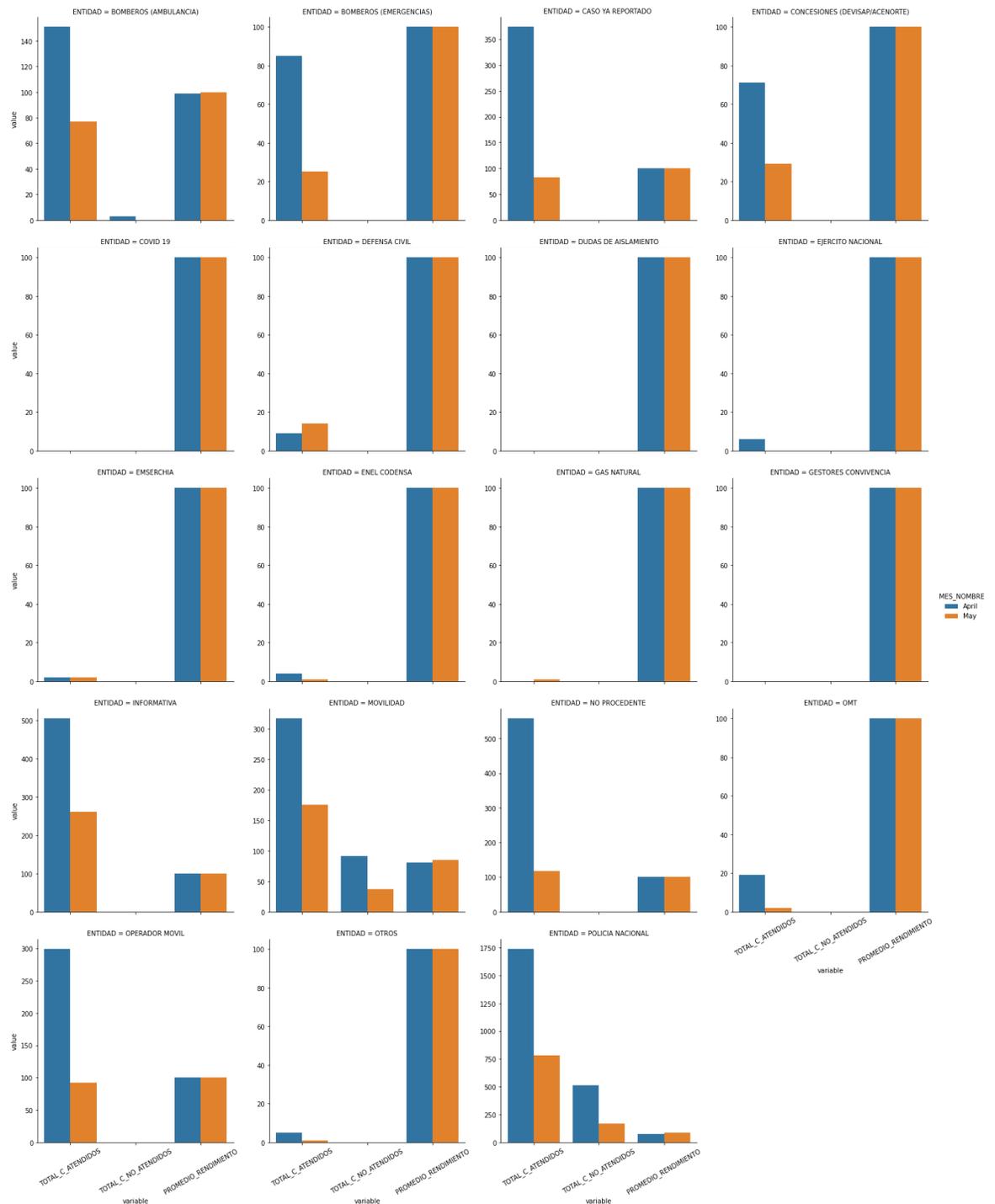
Como solo tenemos 2 meses, vamos a ver la variación entre los casos atendidos, no atendidos y el desempeño para cada una de las 18 entidades.



lunarlab
Team 34

correlation.
one

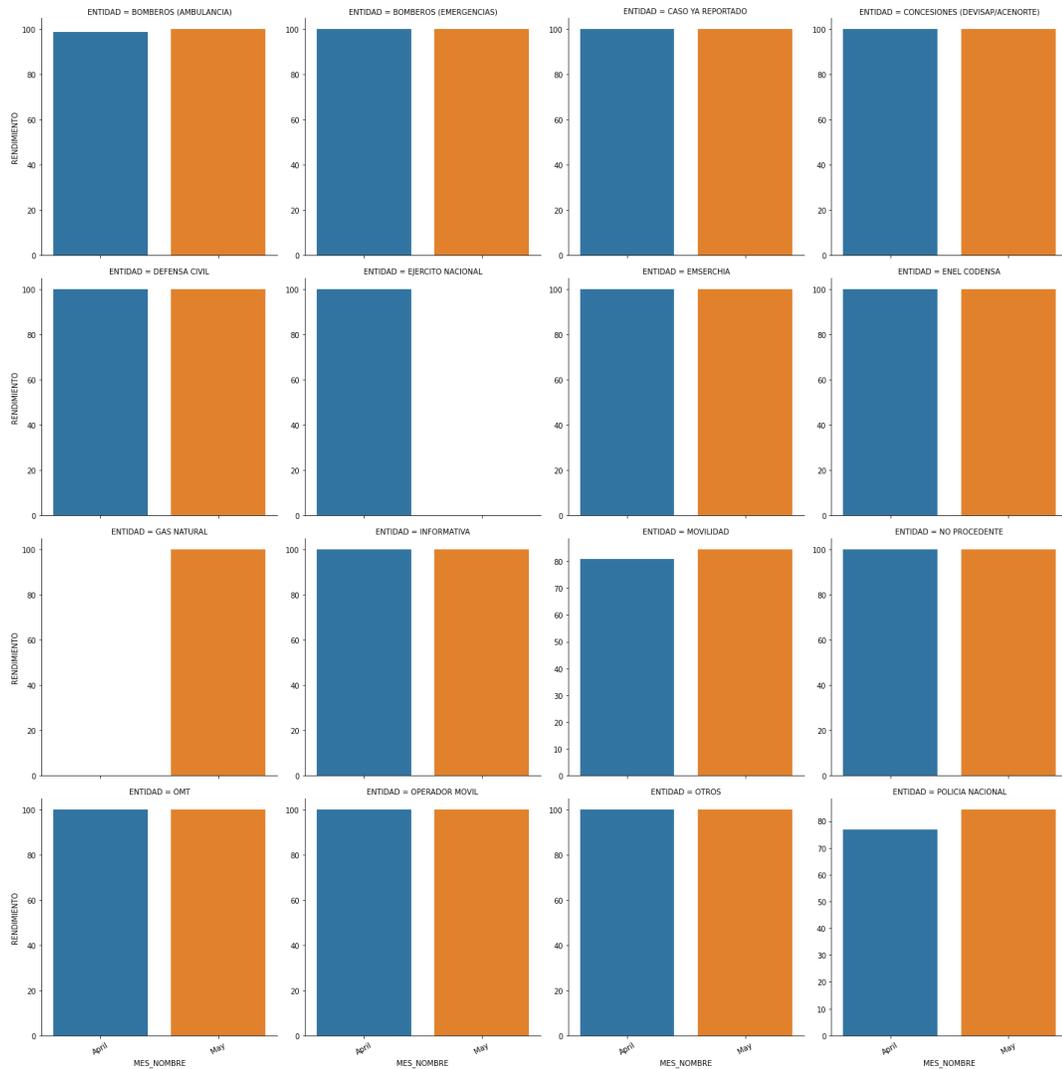
La idea es ver si hay algún factor que nos ayude a identificar por qué hay valores bajos de rendimiento en determinados registros.



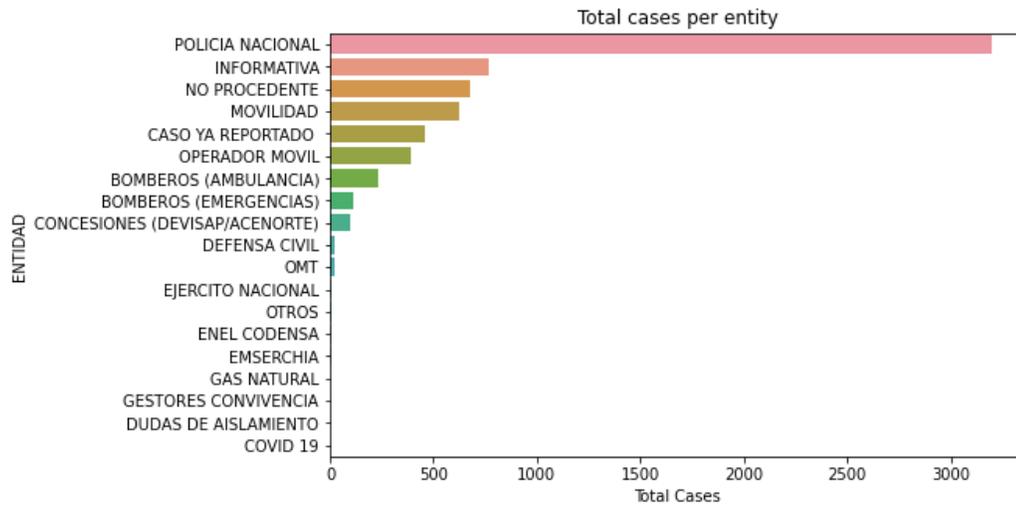
Como podemos ver, el desempeño en varias entidades, en promedio, alcanza o alcanza el valor de 100. Y vemos como efectivamente el total de casos es casi 0 para los 2 meses. En el caso de la policía nacional vemos como hay aproximadamente 800 casos no atendidos, de por sí el gráfico es



difícil de entender. Veamos qué pasa si no contamos en la gráfica los registros donde no hubo casos ese día y solo graficamos el desempeño.

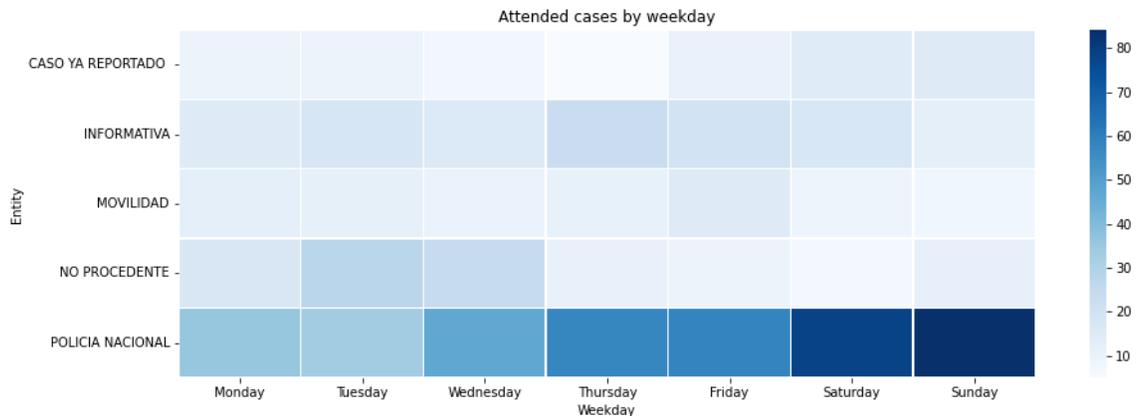


Ahora podemos ver claramente que el rendimiento del 100% está fluctuado por la POLICÍA NACIONAL y por la MOVILIDAD. Vistos ya los 2 gráficos, vamos a ver el porcentaje de casos que atiende cada entidad, a ver si hay algún condicionante.



Efectivamente, la Policía Nacional podría tener ese bajo índice de desempeño en comparación con las demás entidades debido a la gran cantidad de casos que recibe en comparación con las demás. En el caso de Movilidad no se conocería con exactitud el factor, pero podría haber problemas dentro de la entidad porque comparte el total de casos con entidades como informativas, que tiene aproximadamente un 100% en su desempeño.

Podemos ver que la mayoría de los casos atendidos por la Policía son atendidos principalmente los fines de semana. *Movilidad* parece tener más casos el viernes. *No procedente* durante los tres primeros días hábiles. *Caso ya informado* tiene más casos los fines de semana e *Informativa* tiene una cantidad similar de casos durante toda la semana.



Corrección/transformación/limpieza de datos:



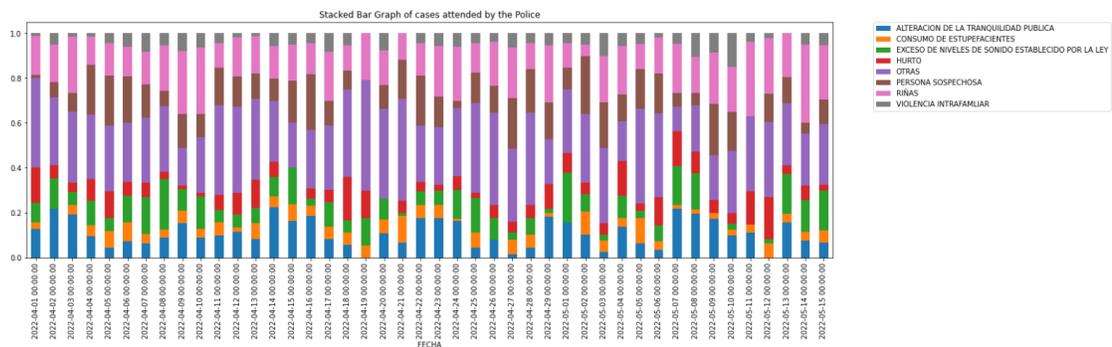
- Las fechas almacenadas en la columna "FECHA" estaban en formato String (Texto), por lo que se convirtieron a datos de fechahora con formato día(2 dígitos)/mes(2 dígitos)/año(4 dígitos).
- Se obtuvo una columna denominada "MES" mediante el proceso de análisis de la columna "FECHA" hasta la fecha y hora para obtener el mes de las fechas.
- Se agruparon los datos según entidad y mes para ver el total de casos atendidos o no en esos 2 meses.
- Escala logarítmica para fines gráficos.

2. ATENCION_TIPO_POLICIA:

Número de registros: 1232

Análisis gráfico:

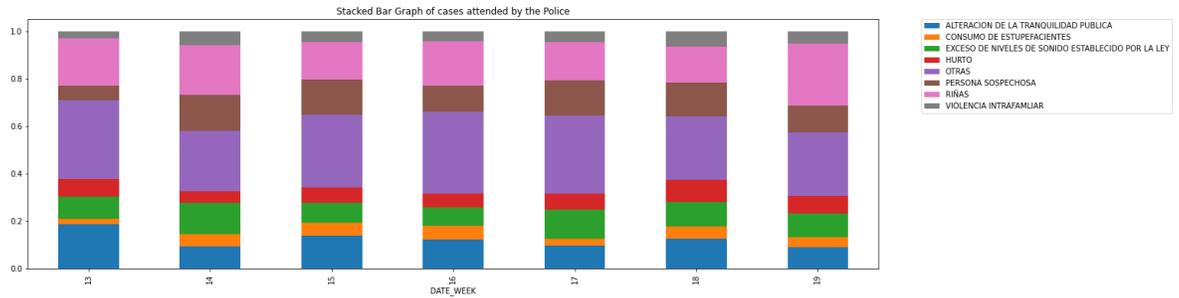
Los datos se presentan desde el 1 de abril de 2022 al 15 de mayo del mismo año. Podemos ver que los casos relacionados con el ruido excesivo se dan principalmente los viernes, sábados y domingos, así como las peleas. El hurto y el consumo de drogas no parecen tener una temporalidad tan clara.



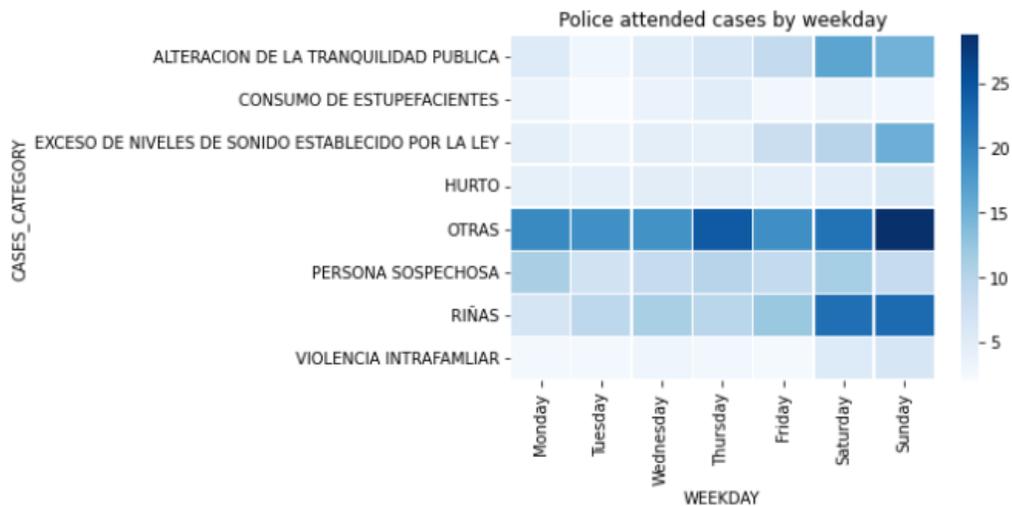
Entre semanas parece haber diferencias en cuanto a la cuantía de algún tipo de delito. Por lo tanto, realizamos una prueba de Chi y con un valor de p de $6.7e-08$ podemos decir que hay suficiente evidencia estadística para decir que la semana no es independiente del delito. Entonces, como intuitivamente dijimos antes, hay algunas semanas que tienen más delitos de una tipología especial.

Como sabemos el día de la madre es uno de los días más violentos del

año, y vemos que la semana 19, que corresponde a la segunda semana de mayo, tuvo un mayor número de peleas en comparación con las semanas anteriores.



Los casos de perturbación de la tranquilidad pública parecen darse principalmente los fines de semana, lo que estaría alineado con los casos de riñas y ruido excesivo. Las búsquedas de personas sospechosas no tienen un horario claro pero ocurren mayormente durante la semana. En cuanto a los casos de violencia doméstica, no se observa un patrón entre los días.



Corrección/transformación/limpieza de datos:

Se arregló el formato de fecha, se eliminaron los espacios en blanco que tenían algunos tipos de casos y se seleccionó el top 7 de los tipos de casos que se reportan con mayor frecuencia (con estas categorías el 70% de los casos reportados se acumulan) para facilitar la interpretación.

3. CASOS_MOVILIDAD:

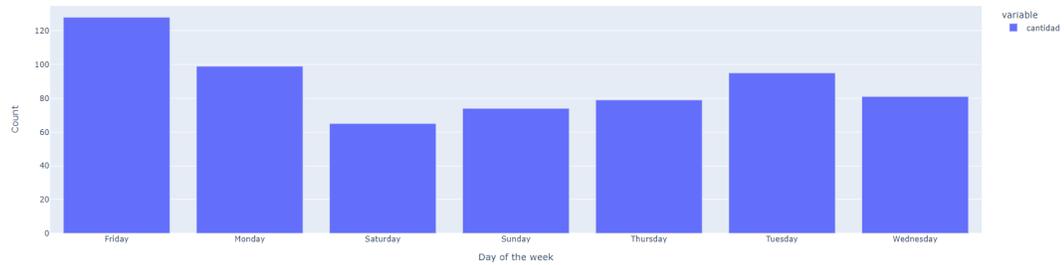
Número de registros: 352



Análisis gráfico:

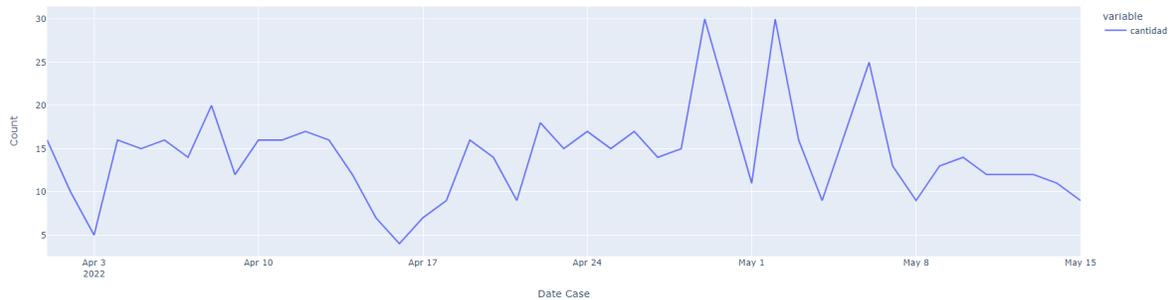
La mayoría de los casos relacionados con la movilidad se están reportando los días viernes, lo que se puede justificar considerando que Chía está muy cerca de la ciudad capital, Bogotá, y al final de la semana es cuando hay una mayor afluencia de habitantes a las regiones aledañas a la "Sabana".

Cases per Day

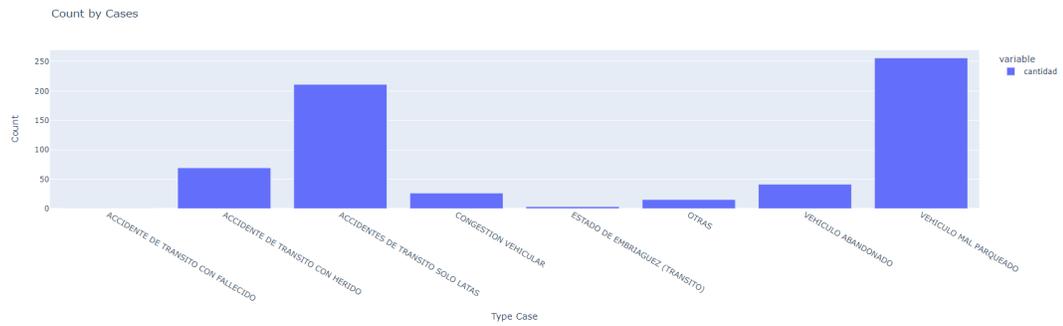


Hay un ligero patrón entre las fechas, a principio de mes y a mitad de mes los casos tienden a disminuir, hay algunas fechas excepcionales, pero de cualquier manera se debe complementar el dato con el tipo de caso de cada mes para obtener más análisis preciso y extenso.

Cases per Date



Los casos más denunciados se están registrando en "Vehículo mal parqueado" y "Accidentes de tránsito solo latas". Aproximadamente a partir del 2021 el Municipio ha ido incrementando los operativos contra accidentes ([Accidentes](#)) es claro que la falta de accidentes fatales representan la forma en que se están tomando los accidentes y las medidas.



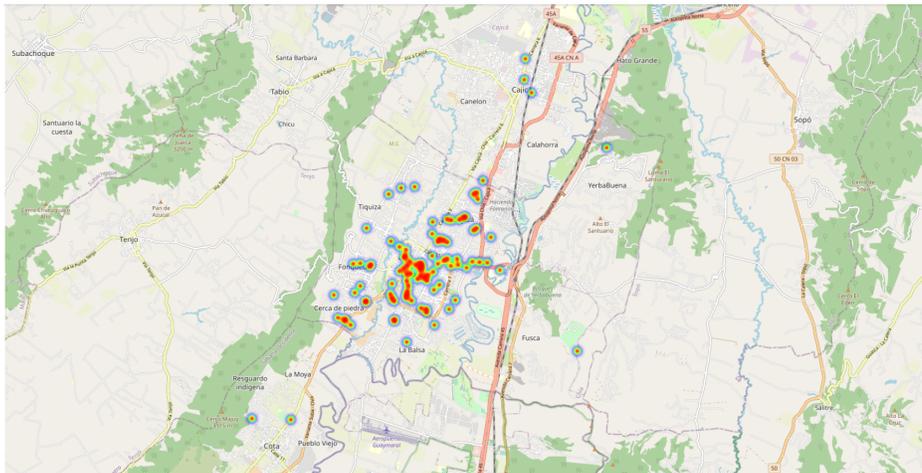
Corrección/transformación/limpieza de datos:

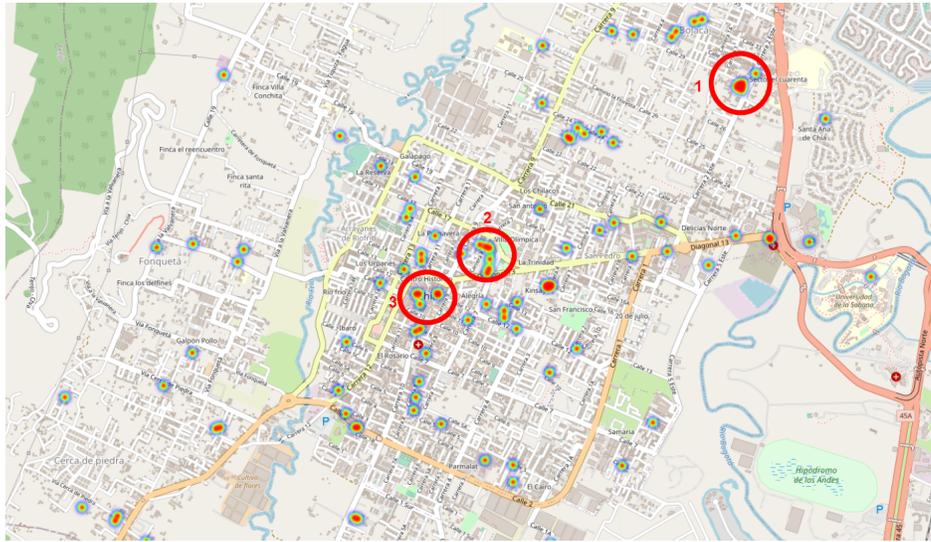
Se corrigió el formato de fecha.

4. EMERGENCIAS_CHIAPP:

Número de registros: 191

Análisis gráfico:





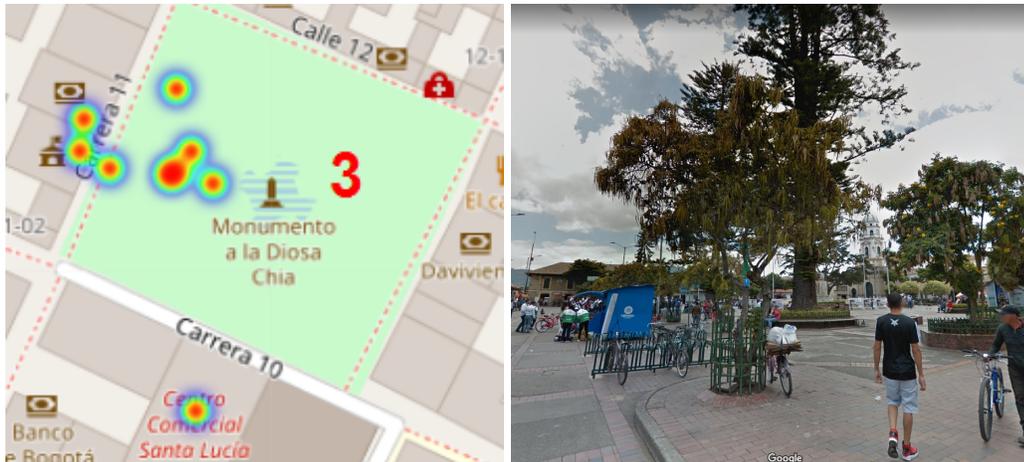
Dado que las emergencias reportadas están ubicadas geográficamente, es posible, por ahora, identificar tres puntos críticos.



El primero está ubicado al norte de la ciudad en una zona rural cercana al campus de la Universidad de Cundinamarca, junto a un parque con grandes árboles.



El segundo también está ubicado cerca de un parque, frente al Coliseo y junto a la Biblioteca.



Y el tercero está ubicado junto al ayuntamiento en el parque central de la ciudad.

Corrección/transформación/limpieza de datos:

No se aplicó ninguna corrección a los datos.

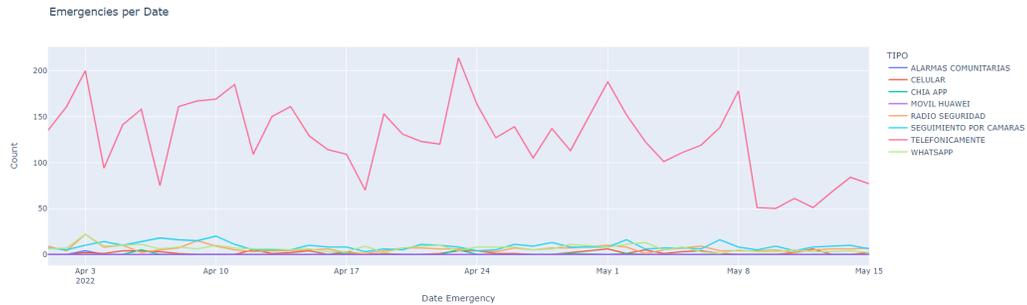
5. TIPO_RECEPCION_EMERGENCIAS:

Número de registros: 352

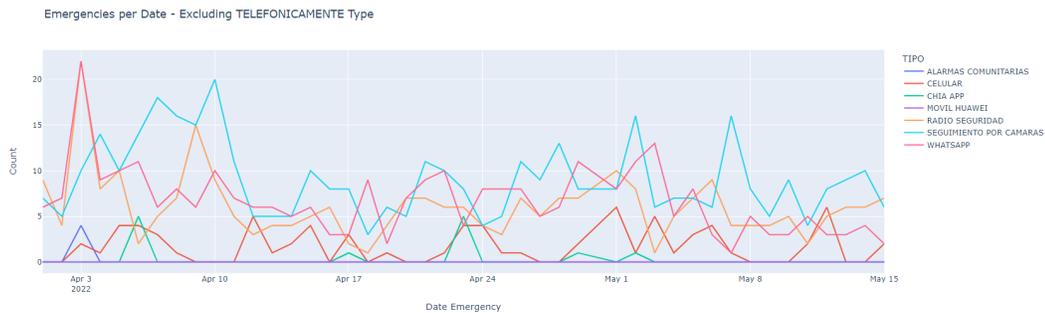
Análisis gráfico:



De este conjunto de datos se puede concluir que la mayoría de los reportes se brindan a través de llamadas al servicio de emergencia. Por este motivo, se ha generado un nuevo gráfico excluyendo este tipo de informes.



Además de las llamadas telefónicas, es posible concluir que las cámaras fueron la principal fuente de denuncias seguidas del whatsapp y las radios de seguridad utilizadas por los vigilantes. Hubo una tendencia a disminuir el número de casos al final del mes.



Corrección/transформación/limpieza de datos:

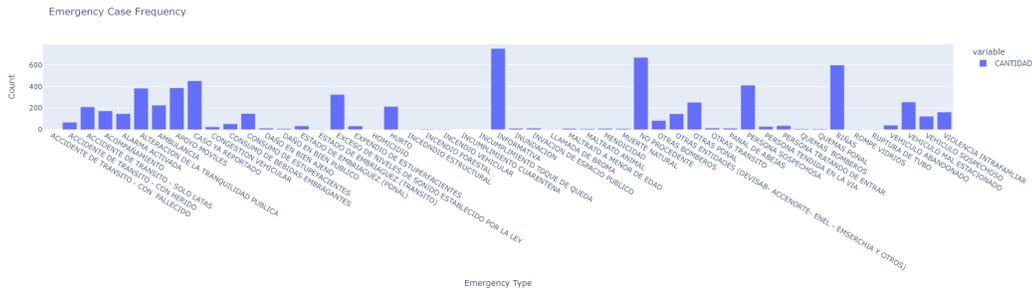
Se corrigió el formato de fecha.

6. TIPOS_EMERGENCIAS_REPORTADAS

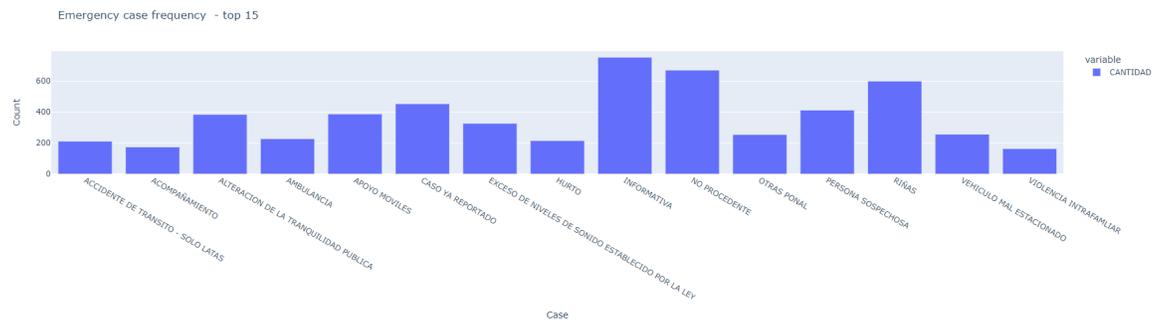
Número de registros: 2244

Análisis gráfico:

En este conjunto de datos tenemos información sobre el tipo de emergencia, la fecha y el número de informes en la misma emergencia y la fecha. Con base en esos datos, es posible graficar la frecuencia del tipo de emergencia:

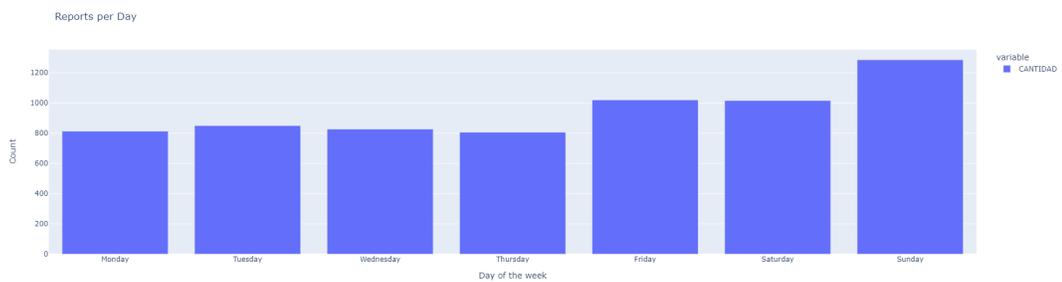


hay más de 50 categorías de emergencias, luego filtramos las 15 principales que envuelven el 82% de las emergencias reportadas.



Las emergencias reportadas con mayor frecuencia son categorías que requirieron una revisión más profunda para aclarar su naturaleza, por ejemplo , *informativo, no procedente y caso ya informado. Riñas, apoyo móviles y percibimos de la tranquilidad pública* son las otras categorías que tienen más reportes

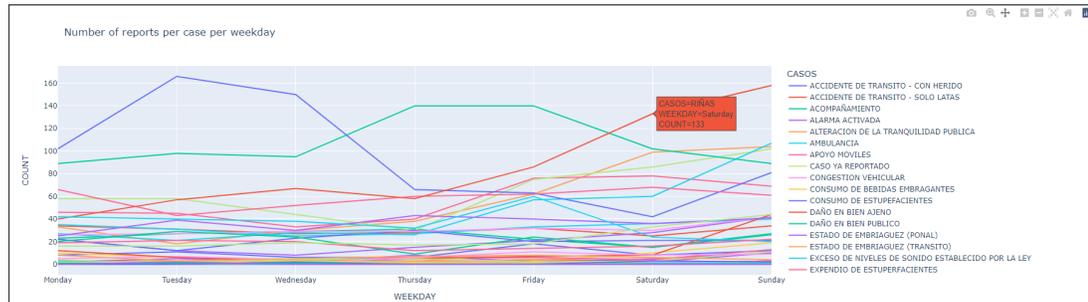
Por otro lado, usando las fechas podemos ver el número de casos en un día de la semana, para darnos una idea de qué día tiene más reportes :



Podemos ver que los reportes comúnmente se hacen durante los fines de semana, y con esa información es posible hacer una tabla de contingencia para ver qué tipo de emergencia es la más reportada por día de la semana y hacer un plan para tratar de reducir significativamente los reportes, y para ello se utilizó un diagrama de líneas para evidenciar los tipos de

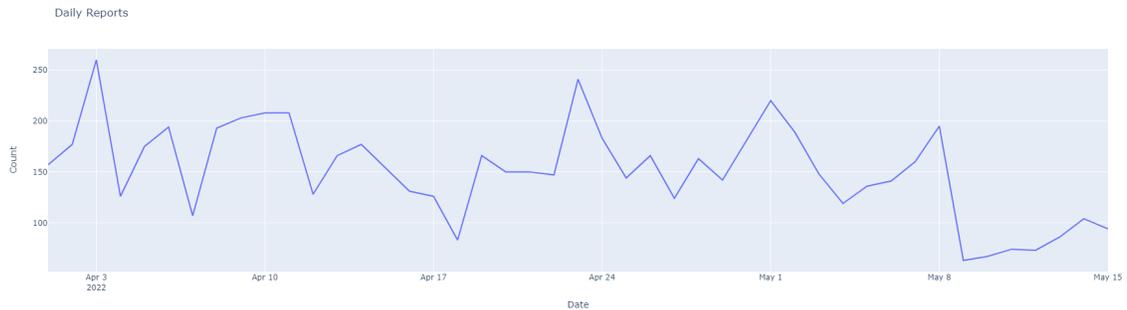


emergencia relevantes.



Los fines de semana el tipo de emergencia más común es RIÑAS, y entre semana son INFORMATIVA y NO PROCEDENTE.

Con las características de este conjunto de datos, es factible hacer un análisis mensual y buscar los meses con una mayor tasa de informes y hacer un análisis más profundo para esos meses, sin embargo, con los datos proporcionados solo tenemos informes de abril y mayo de 2022. y no es suficiente información para sacar conclusiones firmes. En lugar de esto, haremos un análisis diario, veremos el comportamiento y los posibles patrones.



A simple vista con la información disponible parece que el número de reportes aumenta al inicio de cada mes, pero necesitamos más datos para corroborarlo. En segundo lugar, podemos ver picos de informes en algunos fines de semana, y eso parece ayudarnos a corroborar la información encontrada en el gráfico anterior.

Corrección/transformación/limpieza de datos:

- Se arregló el formato de fecha
- Una instancia en la columna CANTIDAD tiene valor NaN, en este caso se reemplazó por 0

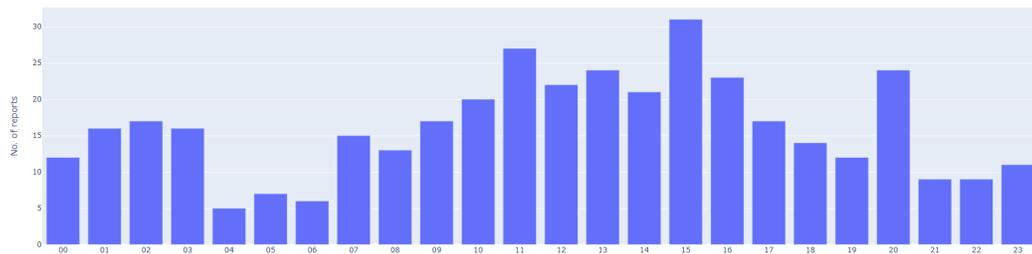


7. REP_CAMARAS_EMERGENCIA:

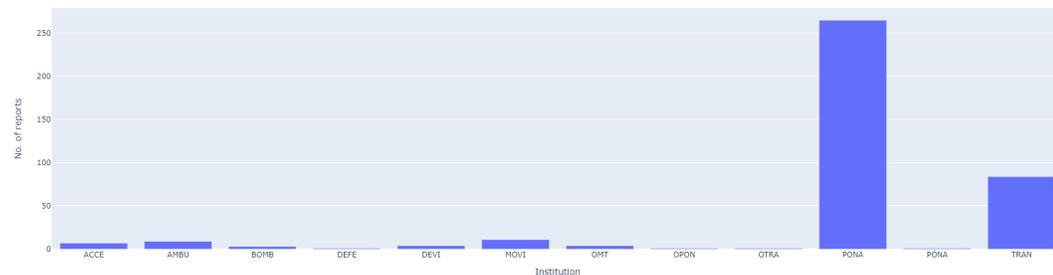
Número de registros: 391

Análisis gráfico:

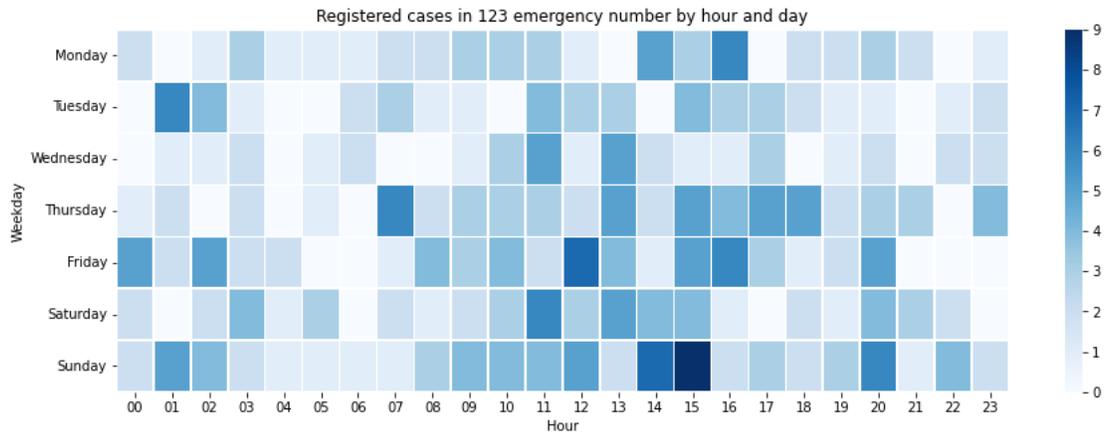
Con la información entregada relacionada a las cámaras se logró agrupar las denuncias realizadas por los ciudadanos tanto por hora del día como por institución que atendió la denuncia.



Según el subconjunto proporcionado, es posible concluir que la hora del día en que se dan más informes es alrededor de las 3:00 p. m. y la hora en que se dan menos informes es entre las 04:00 y las 06:00 p. m



. son recibidas y/o atendidas por la PONA (Policía Nacional) y parece lógico ya que todas las denuncias se comunican primero al número de emergencias 123.



Corrección/transformación/limpieza de datos:

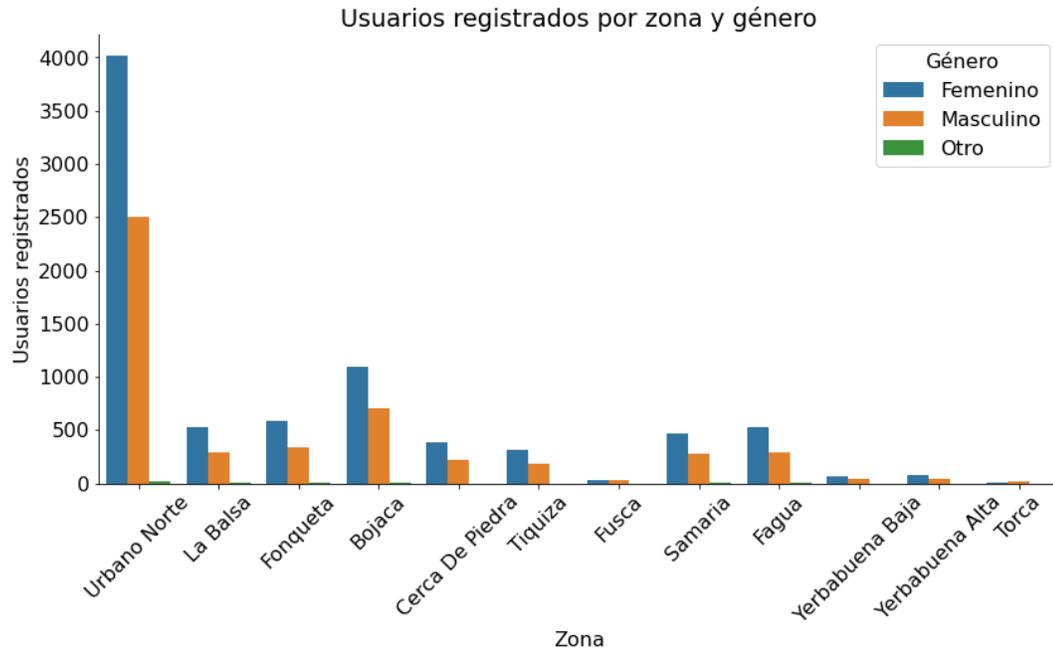
Se arregló el formato de fecha, se arregló el dato de la columna “ENTIDAD_ATIENDE” ya que venía con caracteres especiales como comas, punto y coma, parentesis y espacios en blanco.

10. USUARIOS_REGISTRADOS:

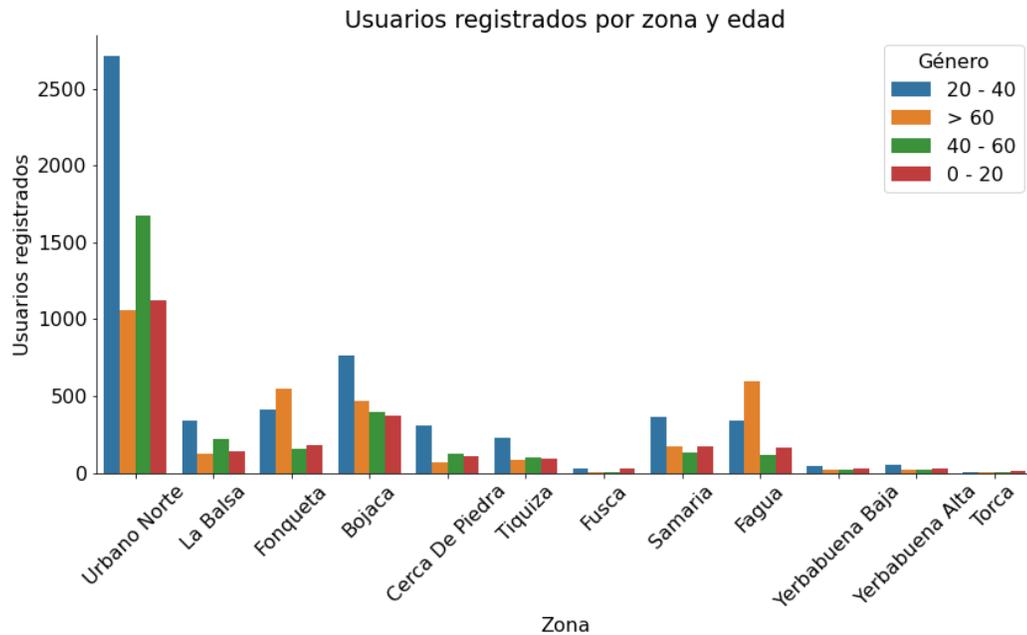
Número de registros: 22894

Gráfico de Análisis:

El siguiente gráfico muestra que la mayoría de los usuarios registrados son mujeres en casi todas las zonas de chiá. Esos usuarios pueden informar problemas en las plataformas. También podemos ver que la mayoría de las personas registradas son de la zona “Urbano Norte”. Si el “Urbano Norte” es la zona más poblada, podemos correlacionar este resultado con los datos de emergencias por localidad. Esas parcelas coinciden porque la zona urbana es la que más casos registra en chiapp. Por lo tanto, las zonas más pobladas pueden estar sujetas a más emergencias, lamentablemente, la base de datos proporcionada por chiapp no tiene qué tipo de emergencias se informan. El hecho de que la mayoría de los usuarios registrados sean mujeres, podría llevarnos a la conclusión errónea de que la mayoría de las víctimas de emergencias son mujeres. Para sacar esa conclusión, necesitaríamos más datos, por ejemplo, cómo se mueven las mujeres por la ciudad.



La parcela por edad muestra que la mayoría de las personas registradas tienen entre 20-40 años y también son registradas de la zona “Urbano Norte”. Necesitamos más análisis para ver si hay correlación entre la edad y las emergencias reportadas de las otras tablas. Un análisis más profundo puede llevarnos a la conclusión equivocada de que las personas entre 20 y 40 años están más involucradas. Para eso necesitaríamos más información, por ejemplo de las llamadas de la policía que informen de la edad de los afectados para dibujar mejor.



El gráfico muestra que la segmentación de las personas en las zonas rurales es mayor que la presente en las zonas urbanas. Las áreas urbanas y rurales son muy similares en el número total de usuarios registrados.

De acuerdo con la información del sitio web oficial de Chia, el área total de la zona urbana es de 17 km y el área total del municipio es de 79 km, es claro en el gráfico que una mayor cantidad de la población vive en el área urbana ya que la mayor parte de los registros de usuario se registran allí.

Corrección/transformación/limpieza de datos:

- Los datos presentaron cuatro valores categóricos para el género ("Masculino", "Femenino", "Otro", "Intermedio") y NAN. Para la limpieza de datos, las NAN se configuraron en "Otro" y tanto "Intermedio" como "Otro" también se configuraron en "Otro". Esto se hizo para tener un solo género además de masculino y femenino.
- La fecha de nacimiento se utilizó para calcular la edad de las personas registradas. Sin embargo, algunos de los cálculos fueron negativos y superiores a 100, lo que no tiene mucho sentido. Esos datos se configuraron como "No disponible".
- Las columnas "SECTOR", "DIRECCION", "FECHA_EXPEDICION_DOCUMENTO", "MUNICIPIO_EXPEDICION", "GRUPO_POBLACIONAL", "ESTADO_CIVIL" no se utilizaron para este análisis básico de datos exploratorios.

11. FOES:

Número de registros: 10648

Análisis gráfico:

Dado que el enfoque del proyecto es la seguridad, hemos concluido que la información del “FOES- Fondo de Educación Superior” no es relevante o no complementaría los otros conjuntos de datos. El conjunto de datos representa la información sobre la asignación de un subsidio relacionado con la educación superior por parte de la alcaldía.

Corrección/transformación/limpieza de datos:

No se aplicó ninguna corrección a los datos.

-

HIPÓTESIS

Ha habido algunas hipótesis que el equipo quisiera probar una vez que la entidad publique los datos completos:

- Una mayor cantidad de cámaras está relacionada con una disminución en las tasas de criminalidad
- Las llamadas son una forma más confiable y efectiva de informar emergencias que las cámaras.
- Las áreas rurales tienden a ser más vulnerables y susceptibles de verse afectadas por robos o emergencias.

MODELOS

Nuestro proyecto está enfocado en determinar cuánto se relaciona con la inseguridad ciudadana en diferentes sectores o barrios.

Algunos métodos a utilizar:

- **Regresión Lineal:** Predecir posibles casos de inseguridad creciente o situaciones en las que sea necesaria la presencia policial visible.
- Determinación de la asociación entre inseguridad y espacios públicos.
- **Clustering:** Agrupación de áreas o sectores para determinar altos índices de criminalidad.
- **Clasificación:** Clasificar los casos informados para determinar una mejor gestión de la situación y comprender cómo mitigar esos casos.
- **Regresión logística:** Uso para determinar la probabilidad de eventos con llamadas de emergencia y saber cómo debe actuar la policía.

INTERFAZ

El producto final será una Aplicación de Página Única (SPA) que muestre mapas de Chia y un tablero para visualizar los datos proporcionados más fácilmente. Algunas de las capas que mostrará este mapa son el mapa de calor de los datos de incidentes, un mapa de las cámaras utilizadas para la vigilancia por parte de las entidades y su cobertura estimada. Con la zonificación de los datos del municipio sería posible identificar los barrios más seguros e inseguros. También incluirá una opción para localizar al usuario y devolverle el índice de vulnerabilidad (por definir) de su barrio o zona más cercana.

La interfaz del tablero permitirá a los usuarios consultar los datos históricos, filtrando por fecha, cámara o incluso la entidad implicada en la resolución de la emergencia, así como mostrar las entidades con más casos dirigidos o resueltos y el canal desde el que se reportaron las emergencias. . Con el filtro de fecha implementado, sería posible ver diferentes mapas para cada fecha y controles deslizantes para ver la progresión de las métricas a lo largo del tiempo. Por acuerdo con el cliente, la aplicación podría mejorarse posteriormente para incluir encuestas para recopilar datos de los ciudadanos sobre su percepción de seguridad o para recibir sugerencias de ellos para mejorar el sitio.

MOCKUP FRONT END

Paletas

Cabecera, panel, menús



#2A4438 #467256 #B9E6CA #EBEBEB



#031926 #468189 #77ACA2 #F4E9CD

Gráficos y diseños



#372C2E #DE9E48 #705E62 #EBF0F2



#D9A384 #BF6B63 #8C6148 #EBF0F2



#DF5F3 #557373 #1E1E1E #F2EFEA



#0F1626 #AB987A #FF533D #F4F4F4



#572F5E #AD71A2 #FFC6CD #FFEE66



#2A4438 #467256 #B9E6CA #EBEBEB



#E1B382 #C89666 #2D545E #12343B



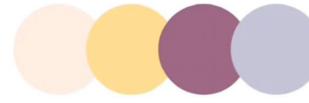
#2B2658 #FFB4A0 #FFE7E1 #F6CF65



#423150 #FFC7B9 #30233B #F8DB88



#F3ECE4 #D88B79 #A7382D #A71F13



#FFECE3 #FFDE94 #986282 #C3C3D5



#031926 #468189 #77ACA2 #F4E9CD



#3F0D0C #D9B061 #8D6F57 #D8D0C5



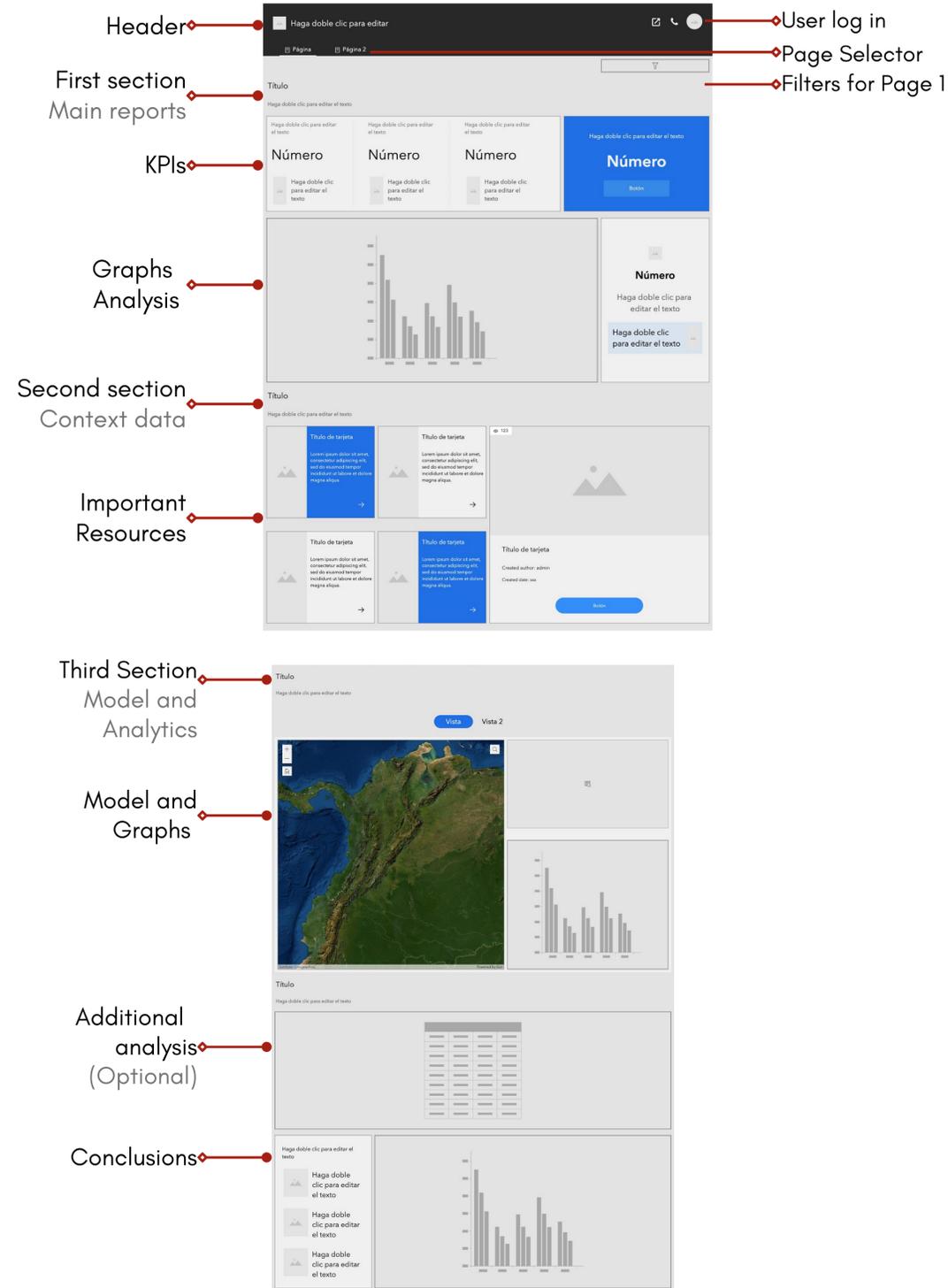
#F6EDE9 #BB544C #353331 #FFC68B



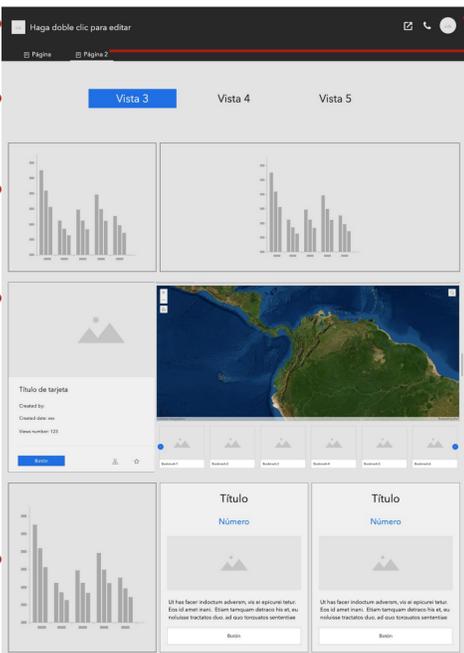
#202020 #DFB577 #FFC7B9 #F2EFEA

Diseño General

Página 1:



Página 2 (Opcional):



Annotations for the dashboard page:

- Header: Haga doble clic para editar
- User log in
- Page Selector
- Views from data: Vista 3, Vista 4, Vista 5
- Graphs for the view
- Important resources and analysis
- Additional analysis (Optional)

Diseño Funcional



Functional design annotations:

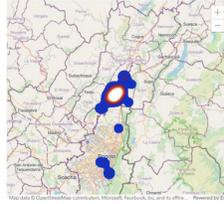
- Header: Security Platform for Chia's Municipality
- User log in
- Page Selector: Emergencias Chiapp
- Filters for Page 1
- First section Main reports: Reportes e Indicadores de interés
- KPIs: EMERGENCIAS REPORTADAS (191), RECEPCIÓN DE EMERGENCIAS (352), TOTAL DE CASOS POR ENTIDAD (836)
- Graphs Analysis: Emergencias reportadas por mes y año
- Second section Context data: Información de cotexto
- Important Resources: Emergencias Chiapp, Chiapp, AppChi, Mapa AppChi, Mockup, Mockup 2, Chiapp, Chia Data

Third Section
Model and
Analytics

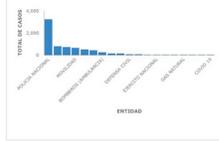
Model and
Graphs

Modelo de seguridad y análisis de la información

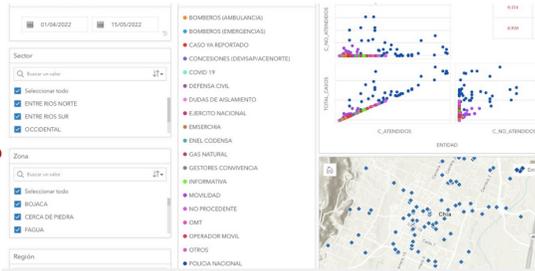
Mapa AppCHI



Casos atendidos por entidad



Additional
analysis
(Optional)



Diseño Final

El diseño final está en español debido a la funcionalidad y el lenguaje nativo de los usuarios finales.



Security Platform for Chia's Municipality





🏠 Página
📄 Página 2

▼ Emergencias chiapp

Reportes e Indicadores de interés

EMERGENCIAS REPORTADAS

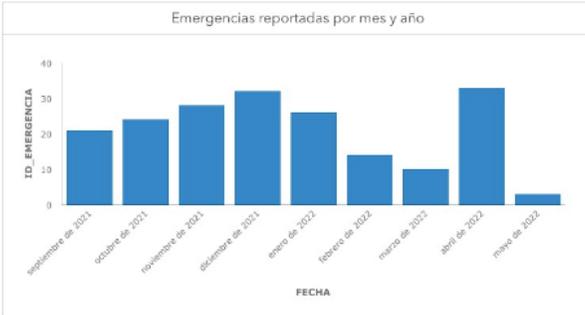
191

RECEPCIÓN DE EMERGENCIAS

352

TOTAL DE CASOS POR ENTIDAD

836



FECHA	ID. EMERGENCIA
septiembre de 2021	21
octubre de 2021	25
noviembre de 2021	28
diciembre de 2021	32
enero de 2022	26
febrero de 2022	15
marzo de 2022	10
abril de 2022	33
mayo de 2022	4

La información de la gráfica representa el total de emergencias que se han reportado en un periodo de 9 meses.

Información de cotexto



Emergencias Chiapp

Información importante emergencias



Chiap



AppChi

Create your own initiative by combining existing applications with a custom site. Use this...



Mapa AppChi



Mockup



Mockup 2



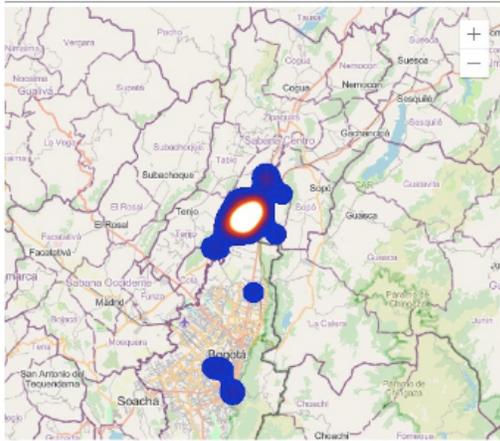
Chiap



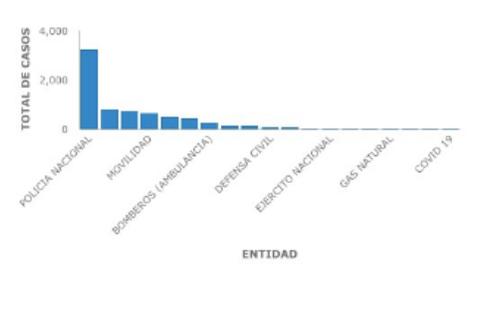
Chia Data

Modelo de seguridad y análisis de la información

Mapa AppChi



Casos atendidos por entidad



Seleccione el rango de fechas

01/04/2022 - 15/05/2022

Sector

Buscar un valor

- Seleccionar todo
- ENTRE RIOS NORTE
- ENTRE RIOS SUR
- OCCIDENTAL

Zona

Buscar un valor

- Seleccionar todo
- BOJACA
- CERCA DE PIEDRA
- FAGUA

Región

ENTIDAD

- BOMBEROS (AMBULANCIA)
- BOMBEROS (EMERGENCIAS)
- CASO YA REPORTADO
- CONCESIONES (DEVISAP/ACENORTE)
- COVID 19
- DEFENSA CIVIL
- DUDAS DE AISLAMIENTO
- EJERCITO NACIONAL
- EMSEERCHIA
- ENEL CODENSA
- GAS NATURAL
- GESTORES CONVIVENCIA
- INFORMATIVA
- MOVILIDAD
- NO PROCEDENTE
- OMT
- OPERADOR MOVIL
- OTROS
- POLICIA NACIONAL

C. NO ATENDIDOS

0.114

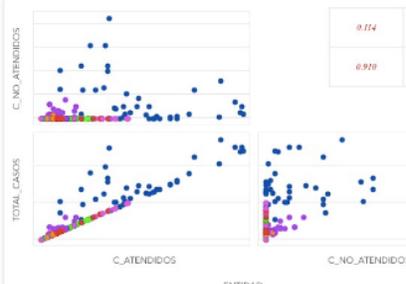
0.970

TOTAL CASOS

C. ATENDIDOS

ENTIDAD

C. NO ATENDIDOS





LINK AL DISEÑO



[Mockup link](#)

ESPECTATIVAS

Lograr una aplicación que ayude a la alcaldía del municipio de Chía a prevenir y predecir emergencias principalmente relacionadas con la seguridad. Esta aplicación tendrá un tablero para ver los datos actualizados y un modelo predictivo para estimar problemas de seguridad.

Esperamos trabajar en cuatro versiones de la aplicación

Versión 1. Una aplicación de tablero para ver cómo varían los informes de seguridad según la fecha y el lugar de las distintas instituciones que reportan las emergencias. Esta aplicación es un análisis de datos exploratorio que pretende tener 5 gráficos interactivos (histogramas, mapas de calor, mapas, etc.).

Versión 2. Cree un modelo predictivo para estimar dónde y cuándo es más probable que ocurra una emergencia de seguridad. El modelo podría ser un modelo de regresión lineal o cualquier modelo de aprendizaje automático con parámetros de bondad de ajuste adecuados.

Versión 3. Construir modelos para ver inferencia causal para evaluar si existe alguna correlación entre factores como la vulnerabilidad de las personas, ubicación, institución, entre otros.

Versión 4. Recopilar datos externos (p. ej.: <https://www.datos.gov.co/> o cualquier otra fuente) para mejorar nuestro análisis y las relaciones sobre nuestros datos).

CRONOGRAMA

Fecha	Entregable	Detalles
Semana 1	Formación del equipo	Nuestro equipo está conformado por siete personas
Semana 2	Se trabaja en la formación de la idea	Selección de proyectos del banco de proyectos
Semana 3	Se debe finalizar la idea y empezar a definir el alcance	Con el proyecto sobre el municipio de Chía
Semana 4	Se completa el alcance del proyecto	Reunión con el cliente para definir el alcance. El alcance del proyecto está relacionado con la seguridad del municipio de Chía.

Semana 5	Conjuntos de datos obtenidos	Los datos fueron proporcionados por el cliente.
Semana 6	EDA básico/Limpieza de conjuntos de datos completada	Realización de análisis exploratorios de datos básicos, incluidos histogramas, mapas de calor, entre otros
Semana 7	EDA en profundidad, análisis de jupyter, maqueta de interfaz	El EDA en profundidad se explicó anteriormente en la sección de métodos.
Semana 8	Diseño de la interfaz	Estructura de una aplicación web.
Semana 9	Diseño de la interfaz completo	Inicio de la creación del framework front-end para comenzar a agregar las tramas y análisis
Semana 10	Infraestructura de aplicaciones completa	Creación de aplicaciones web en plotly-dash
Semana 11	Proyecto y Datafolio completos	Construcción del datafolio y manuscrito del proyecto.
Semana 12	Presentación ensayada y lista	Escribir un guión para un video y grabarlo. Presentación del vídeo.

CONCLUSIONES

- De los 11 conjuntos de datos iniciales, utilizaremos 10, que se han priorizado de acuerdo con el siguiente gráfico de clasificación:

DATA						
<u>ENTIDAD_ATIENDE_EMERGENCIAS</u>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>ATENCION_TIPO_POLICIA</u>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>CASOS_MOVILIDAD</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>USUARIOS_REGISTRADOS</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<u>EMERGENCIAS_CHIAPP</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>TIPO_RECEPCION_EMERGENCIAS</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>TIPOS_EMERGENCIAS_REPORTADAS</u>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>REP_CAMARAS_EMERGENCIA</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>ZONIFICACION_CHIA</u>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>LLAMADAS123</u>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>FOES</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

SCORE FROM 0-5



Después de analizar los conjuntos de datos, concluimos que los datos de "FOES" no se utilizarán para el proyecto, ya que solo representan la información sobre la asignación de un subsidio relacionado con la educación superior por parte de la alcaldía. Y dado que el proyecto trata sobre delitos y seguridad, los datos no están relacionados con el propósito principal.



- No existe un ID o campo que nos permita relacionar directamente cada uno de los conjuntos de datos, además de las fechas, sin embargo estas no garantizan certeza para los casos.
- Se solicitará mayor información a la Alcaldía de Chía, incluyendo la ubicación de las cámaras, así como a hospitales, CAIs, Estaciones de Bomberos.
- Dado que estamos trabajando con una muestra de los datos, será necesaria una mayor limpieza una vez que se obtenga la base de datos final.
- Algunas tablas tienen pocos registros en cuanto al tiempo, por ejemplo, en “entidad_atiende_emergencias” los registros abarcan solo 2 meses (abril y mayo) del año 2022.
- En “llamadas123” en la columna de entidad hay valores que no corresponden a entidades, serían más bien descriptivos o del medio donde se notificó las llamadas.
- Los usuarios registrados se ubican principalmente en la zona urbana, por lo que hay que tener cuidado en el momento de sacar conclusiones para no decir inferencias obvias. Por lo tanto, vamos a necesitar información más granular en el área urbana para resaltar puntos importantes.
- Algunas de las preocupaciones de limpieza de datos encontradas por el equipo se informaron a los propietarios de la base de datos, quienes mejorarán la forma en que los usuarios ingresan datos en una versión posterior.



GLOSARIO – POWER BI UNIDAD 2

Datos oscuros:	Según Gartner, la definición de datos oscuros es "activos de información que las organizaciones recopilan, procesan y almacenan en el curso de su actividad comercial habitual, pero que generalmente no utilizan para otros fines". El 90% de los datos de las empresas son datos oscuros.
Paneles:	Una herramienta de visualización de datos que muestra el estado actual de la empresa, el estado de las métricas y los KPI, y el análisis de datos y las perspectivas actuales.
Analista de datos:	Como lo afirma modernanalyst.com, un analista de datos es un profesional que se encarga de analizar y extraer datos para identificar patrones y correlaciones, mapear y rastrear datos de un sistema a otro para resolver un problema, utilizando BI y datos. Herramientas de descubrimiento para ayudar a los ejecutivos de negocios en la toma de decisiones y realizar análisis estadísticos de datos comerciales, entre otras cosas.
Análisis de datos:	Según TechTarget, "el análisis de datos es el proceso de examinar conjuntos de datos para sacar conclusiones sobre la información que contienen, cada vez más con la ayuda de sistemas y software especializados".
Gobernanza de datos:	Según Boris Evelson, de Forrester Research, la gobernanza de datos "se ocupa de todo el espectro (creación, transformación, propiedad, etc.) de personas, procesos, políticas y tecnologías que gestionan y gobiernan el uso que hace una empresa de sus activos de datos. (Como aplicaciones de administración de gobierno de datos, administración de datos maestros, administración de metadatos y calidad de datos).
Data Mashup:	Una integración de múltiples conjuntos de datos en una representación analítica y visual unificada.
Silos de datos:	Según Tech Target, un silo de datos son "datos que están bajo el control de un departamento o persona y están aislados del resto de la organización". Los silos de datos son un cuello de botella para las operaciones comerciales efectivas.

Fuentes de datos:	La fuente de donde provienen los datos a analizar. Puede ser un archivo, una base de datos, un conjunto de datos, etc. Las soluciones de BI modernas como Necto pueden combinar datos de múltiples fuentes de datos.
Visualización de datos:	Visualización gráfica de datos. Puede incluir formas tradicionales como gráficos y tablas, y formas modernas como infografías.
Almacén de datos:	Una base de datos relacional que integra datos de múltiples fuentes dentro de una empresa.



El futuro digital
es de todos

MinTIC

GLOSARIO – POWER BI UNIDAD 1

Análisis automatizado:	Análisis automático de datos para encontrar información oculta en los datos y mostrar a los usuarios las respuestas a preguntas en las que ni siquiera han pensado.
Analista de BI:	Como lo afirma modernanalyst.com, un analista de datos es un profesional que se encarga de analizar y extraer datos para identificar patrones y correlaciones, mapear y rastrear datos de un sistema a otro para resolver un problema, utilizando BI y datos. Herramientas de descubrimiento para ayudar a los ejecutivos de negocios en la toma de decisiones y realizar análisis estadísticos de datos comerciales, entre otras cosas. (También se puede llamar analista de datos).
Gobernanza de BI:	Según Boris Evelson, de Forrester Research, la gobernanza de BI es una parte clave del gobierno de datos, pero se centra en un sistema de BI y gobierna quién usa los datos, cuándo y cómo.
Big Data:	Conjuntos de datos enormes y complejos que las herramientas tradicionales de procesamiento de datos no pueden manejar.
Cuellos de botella:	Puntos de congestión o bloqueo que dificultan la eficiencia del sistema de BI.
Business Intelligence:	Según Gartner, "Business Intelligence es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento".
Inteligencia empresarial centralizada:	Un modelo de BI que permite a los usuarios trabajar conectados y compartir conocimientos, mientras ven la misma y única versión de la verdad. TI gobierna los permisos de datos para garantizar la seguridad de los datos.
BI colaborativo:	Un enfoque de Business Intelligence en el que la herramienta de BI permite a los usuarios colaborar entre colegas, compartir conocimientos e impulsar el conocimiento colectivo para mejorar la toma de decisiones.

Conocimiento colectivo:

Conocimiento que beneficia a toda la empresa, ya que proviene del intercambio de conocimientos y hallazgos de datos entre grupos y departamentos para enriquecer el análisis.