



Tecnológico de Monterrey

Maestría en Inteligencia Artificial Aplicada

Proyecto Integrador

**Septiembre – Diciembre
2025**

Actividad 3

Avance 1. Análisis exploratorio de datos

Docentes:

Dra. Grettel Barceló Alonso

Dra. Verónica Sandra Guzmán De Valle

Asesor:

Dr. Luis Eduardo Falcón

Alumno:

Carlos Enríquez Gorgonio

A01793102

Septiembre 2025



Esta información es de uso exclusivo de la Secretaría de Recursos Naturales y Medio Ambiente del Estado de Durango y tiene carácter

C O N F I D E N C I A L

La visualización, distribución y almacenamiento de este documento se encuentran sujetos a lo dispuesto en la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental. Cualquier persona que carezca de autorización expresa y por escrito para visualizar o almacenar esta información deberá abstenerse de su uso y eliminar este documento de inmediato.

ÍNDICE

Avance 1 - parte 1 - análisis exploratorio de datos (EDA - Exploratory Data Analysis)	4
Exploración	5
Avance 1 - Parte 2 - Conclusiones	8
Análisis	9
Preguntas Clave	10
Fuentes Bibliográficas	17



Avance 1 - parte 1 - análisis exploratorio de datos (EDA - Exploratory Data Analysis)

Exploración

En esta fase inicial de exploración centrada en examinar y analizar los datos, y entre otras cosas “inspeccionar, limpiar y modelar datos con el objetivo de descubrir información útil, llegar a conclusiones y apoyar la toma de decisiones. El análisis de datos puede involucrar diversas técnicas estadísticas y computacionales, y se lleva a cabo en diversas áreas, desde la investigación científica hasta los negocios. La técnica y los métodos empleados dependen del tipo de datos, el contexto y los objetivos del análisis, y pueden incluir desde análisis descriptivos básicos hasta modelos predictivos complejos.” (Choudhury, S., 2021).

Para poder tener acceso al notebook, es necesario entrar al siguiente vinculo de github:

https://github.com/KarltonBotics/Integrador_CentinelAI

A pesar de que se cuenta con registros desde el año 2015, esa información no se encuentra homologada, por lo que únicamente se utilizaran datos homologados consistentes en información de 3 años, es decir 2023-2025, consistente en 499 registros. No obstante, se procurará incorporar información de años anteriores, previo análisis de

Clave de Incendio
Fecha de Reporte
Longitud Grados
Longitud Minutos
Longitud Segundos
Latitud Grados
Latitud Minutos
Latitud Segundos
Cve de Municipio
Municipio
Predio/Paraje
Causa
Causa Específica
Fecha de Inicio
Fecha de Liquidación
Tipo de Incendio
Tipo de Vegetación
Zona Bioclimática
Arbolado Adulto
Renuevo
Arbustivo
Herbaceo
Hojarasca
Total (ha)
Costo Estimado del Incendio

Imagen 8. Encabezados homologados 2023-2025

Fuente: archivo denominado "REPORTES POR AÑO"

Fuente: Secretaría de Recursos Naturales y Medio Ambiente
del Estado de Durango

A continuación, en la imagen 9 se muestran en color verde los datos en las tablas no homologadas, así como con registros faltantes, los cuales se esta analizando la viabilidad de su incorporación al análisis.

Clave de Incendio
Fecha de Reporte
Longitud Grados
Longitud Minutos
Longitud Segundos
Latitud Grados
Latitud Minutos
Latitud Segundos
Cve de Municipio
Municipio
Predio/Paraje
Causa
Causa Específica
Fecha de Inicio
Fecha de Liquidación
Tipo de Incendio
Tipo de Vegetación
Zona Bioclimática
Arbolado Adulto
Renuevo
Arbustivo
Herbaceo
Hojarasca
Total (ha)
Costo Estimado del Incendio

CLAVE INCENDIO
LONGITUD
LATITUD
MUNICIPIO
PREDIO
FECHA DE INICIO
FECHA LIQUIDACION
TIPO DE INCENDIO
TIPO DE VEGETACION
ARBOREO
RENUEVO
ARBUSTIVO
HERBACEO
HOJARASCA
AREA TOTAL

Imagen 9. Encabezados no homologados 2023-2025

Fuente: archivo denominado "REPORTES POR AÑO"

Fuente: Secretaría de Recursos Naturales y Medio Ambiente
del Estado de Durango

Esta información es de uso exclusivo de la Secretaría de Recursos Naturales y Medio Ambiente del Estado de Durango y tiene carácter

C O N F I D E N C I A L

La visualización, distribución y almacenamiento de este documento se encuentran sujetos a lo dispuesto en la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental. Cualquier persona que carezca de autorización expresa y por escrito para visualizar o almacenar esta información deberá abstenerse de su uso y eliminar este documento de inmediato.

Como ejemplo de información faltante, se muestra la siguiente tabla correspondiente al año 2018:

PARAJE	AREA TOTAL	HOJARASCA	SUELO ORGANICO	HERBACEO	ARBUSTIVO	ARBOREO	RENUEVO
El Panteon	2.4			2.42			
El Cerro Bola	50.0			10.00	40.00		
El Alto del Toro	4.0	4.00					
La Playa	160.0			110.00	50.00		
Remate de Moromocho, Arroyo La Quebradita	2006.0	241.00		500.00	1,200.00	25.00	40.00
Calichitos, El Potrero y Comalito	1101.8	330.53		330.52	440.70		
Cordón de Tapias	384.0			114.00	270.00		
Arroyo de la Barranca	753.0	80.00		203.00	400.00	25.00	35.00
El Comalito	60.0			18.00	24.00	6.00	12.00
Cordón del Burro	9.0			3.00	5.00		1.00
El Peñol	287.0			140.00	147.00		
Los Cardos	1090.0			420.00	500.00	50.00	40.00
Cristales	20.0			10.00	20.00		
Cajón de las Bayas	11.0			5.00	6.00		

Imagen 10. Ejemplo de registros faltantes, año 2018

Fuente: archivo denominado "REPORTES POR AÑO"

Fuente: Secretaría de Recursos Naturales y Medio Ambiente
del Estado de Durango

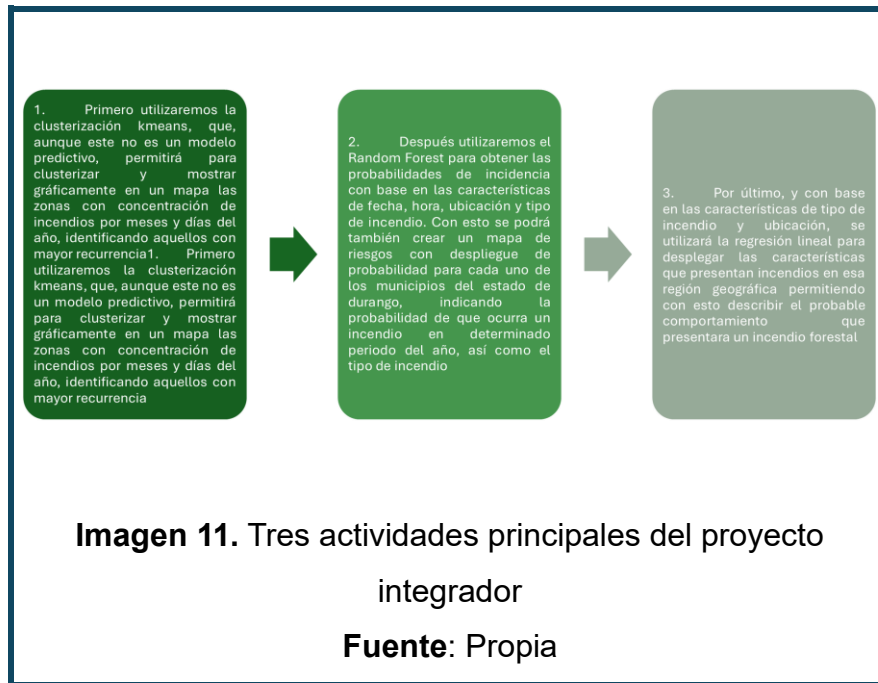
De los años 2016 al 2013, los registros son con mucho peor calidad y se presenta una tercera variación de encabezados, por lo que se determina que serán datos despreciables para el análisis.



Avance 1 - Parte 2 - Conclusiones

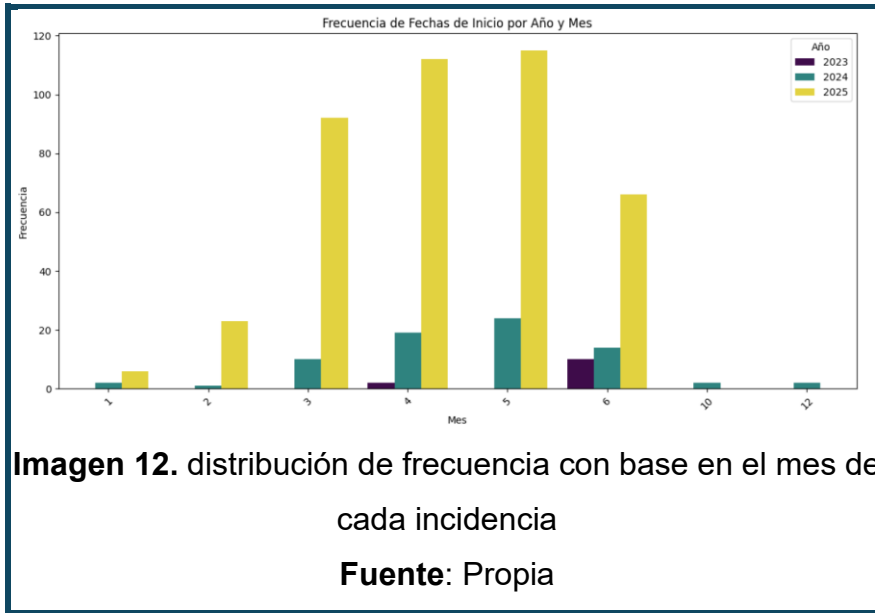
Analisis

Con un enfoque en el cumplimiento de las 3 actividades clave descritas en el entregable anterior (actividad 1), como se muestra en la imagen:



Para lo cual nos enfocaremos en todas las características que nos permitan trabajar con valores de;

1. Fecha
2. Hora
3. Estado
4. Municipio
5. Latitud
6. Longitud
7. Tipo de incendio



Preguntas Clave

¿Hay valores faltantes en el conjunto de datos? ¿Se pueden identificar patrones de ausencia?

Si, son mínimos, únicamente en 4 columnas

Causa	5
Causa Especifica	6
Fecha de Inicio	0
Fecha de Liquidación	0
Tipo de Incendio	4
Tipo de Vegetación	0
Zona Bioclimática	5

Imagen 13. Valores faltantes en el dataset

Fuente: Propia

Al tratarse de 500 registros de datos, estos datos no afectan principalmente porque no tienen relación con la fecha o la ubicación geográfica.

¿Cuáles son las estadísticas resumidas del conjunto de datos?

Se cuenta con 500 registros

25 columnas

Las columnas principales se encuentran en formato fecha hora, lo que facilita considerablemente el análisis de la información, así como la extracción de valores como numero de mes o año en que ocurre cada evento.

El segundo valor principal al momento es el nombre del municipio el cual no se ve muy afectado si se cambia a cadena ordinaria de texto, sin embargo, se explorará al momento de ubicar en un mapa, la viabilidad de trabajar con coordenadas o con un rango de coordenadas que abarquen el perímetro de cada municipio.

Clave de Incendio	object
Fecha de Reporte	datetime64[ns]
Longitud Grados	int64
Longitud Minutos	int64
Longitud Segundos	float64
Latitud Grados	int64
Latitud Minutos	int64
Latitud Segundos	float64
Cve de Municipio	int64
Municipio	object
Predio/Paraje	object
Causa	object
Causa Especifica	object
Fecha de Inicio	datetime64[ns]
Fecha de Liquidación	datetime64[ns]
Tipo de Incendio	object
Tipo de Vegetación	object
Zona Bioclimática	object
Arbolado Adulto	float64
Renuevo	float64
Arbustivo	float64
Herbaceo	float64
Hojarasca	float64
Total (ha)	float64
Costo Estimado del Incendio	int64

Imagen 14. Tipo de datos de cada columna del dataset

Fuente: Propia

	Clave de Incendio	Fecha de Reporte	Longitud Grados	Longitud Minutos	Longitud Segundos	Latitud Grados	Latitud Minutos	Latitud Segundos	Cve de Municipio	Municipio	...	Arbolado Adulto	Remuevo	Arbustivo	Herbáceo	Hojasaca	Total (ha)	Estado del Incendio	Costo
count	500	500	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500	—	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
unique	500	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	16	—	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	23-10-0529	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Mezquital	—	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	184	—	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	2025-04-20 09:55:24.960000	104.750000	30.71000	30.411160	23.108000	34.656000	29.170780	16.960000	NaN	—	1.887349	11.661049	70.216661	90.880269	85.368770	263.014098	21036.324000	;
min	NaN	2025-05-18 16:14:00	103.000000	0.000000	0.000000	22.000000	0.000000	0.000000	1.000000	NaN	—	0.000000	0.000000	0.000000	0.000000	0.000000	0.300000	0.000000	;
25%	NaN	2025-05-06 01:41:00	104.000000	20.00000	15.517500	22.000000	17.000000	15.000000	14.000000	NaN	—	0.000000	0.000000	0.267500	1.362475	1.799725	5.009590	4980.000000	;
50%	NaN	2025-06-11 12:27:30	105.000000	31.50000	30.495000	23.000000	40.000000	29.000000	17.000000	NaN	—	0.000000	0.000000	5.573200	9.621750	10.115200	28.599300	10470.000000	;
75%	NaN	2025-07-29 10:27:30	105.000000	42.00000	44.615000	24.000000	50.000000	44.000000	23.000000	NaN	—	0.000000	2.255750	33.134125	47.861750	51.235800	145.000000	17415.000000	;
max	NaN	2025-08-18 10:30:50	107.000000	59.00000	59.990000	26.000000	59.000000	59.000000	37.000000	NaN	—	190.000000	701.907500	2978.642800	3700.000000	2647.682600	9409.864000	611960.000000	;
std	NaN	NaN	0.807964	15.13923	17.132735	1.017057	18.473229	16.846152	6.762189	NaN	—	11.489577	58.840439	255.781721	329.129148	261.319005	888.893422	45017.067681	;

Imagen 15. Descripción del dataset

Fuente: Propia

¿Hay valores atípicos en el conjunto de datos?

Es difícil determinar cual es el enfoque de nuestro valor atípico, por lo que nos enfocamos en las fechas de las incidencias, y los valores atípicos corresponden a los meses con 0 eventos, ubicándose mayormente al último trimestre de cada año

	Año	Mes_num	Eventos	Mes
0	2025	1	6.0	Enero
1	2025	2	23.0	Febrero
2	2025	3	92.0	Marzo
3	2025	4	112.0	Abril
4	2025	5	115.0	Mayo
5	2025	6	66.0	Junio
6	2025	7	0.0	Julio
7	2025	8	0.0	Agosto
8	2025	9	0.0	Septiembre
9	2025	10	0.0	Octubre
10	2025	11	0.0	Noviembre
11	2025	12	0.0	Diciembre
12	2024	1	2.0	Enero
13	2024	2	1.0	Febrero
14	2024	3	10.0	Marzo
15	2024	4	19.0	Abril
16	2024	5	24.0	Mayo
17	2024	6	14.0	Junio
18	2024	7	0.0	Julio
19	2024	8	0.0	Agosto
20	2024	9	0.0	Septiembre
21	2024	10	2.0	Octubre
22	2024	11	0.0	Noviembre
23	2024	12	2.0	Diciembre
24	2023	1	0.0	Enero
25	2023	2	0.0	Febrero
26	2023	3	0.0	Marzo
27	2023	4	2.0	Abril
28	2023	5	0.0	Mayo
29	2023	6	10.0	Junio
30	2023	7	0.0	Julio
31	2023	8	0.0	Agosto
32	2023	9	0.0	Septiembre
33	2023	10	0.0	Octubre
34	2023	11	0.0	Noviembre
35	2023	12	0.0	Diciembre

Imagen 16. Detalles de valores atípicos

Fuente: Propia

¿Cuál es la cardinalidad de las variables categóricas?

La principal es la “Fecha de Inicio”, representando las categorías por meses del año, aunque se puede escalar también a días, sería muy ambicioso desear predecir por día.

La cardinalidad por lo tanto es 12, correspondientes a los meses del año.

también existen otras que para el análisis principal no son relevantes como el tipo de incendio que al momento es binario porque solo muestra “Superficial” o “De copa”, sin embargo, pueden existir otro tipo de incendios en los registros.

también los municipios muestran comportamiento categórico:

```
Valores únicos de 'Tipo de Incendio':  
['Superficial' nan 'De copa']  
  
Valores únicos de 'Municipio':  
['Pueblo Nuevo' 'Topia' 'Durango' 'Tepehuanes' 'Canelas'  
'Santiago Papasquiaro' 'San Dimas' 'Mezquital' 'Tamazula' 'Guanaceví'  
'El Oro' 'Otáez' 'Ocampo' 'Canatlán' 'Nombre de Dios' 'Cuencamé']
```

Imagen 17. Variables categóricas relevantes

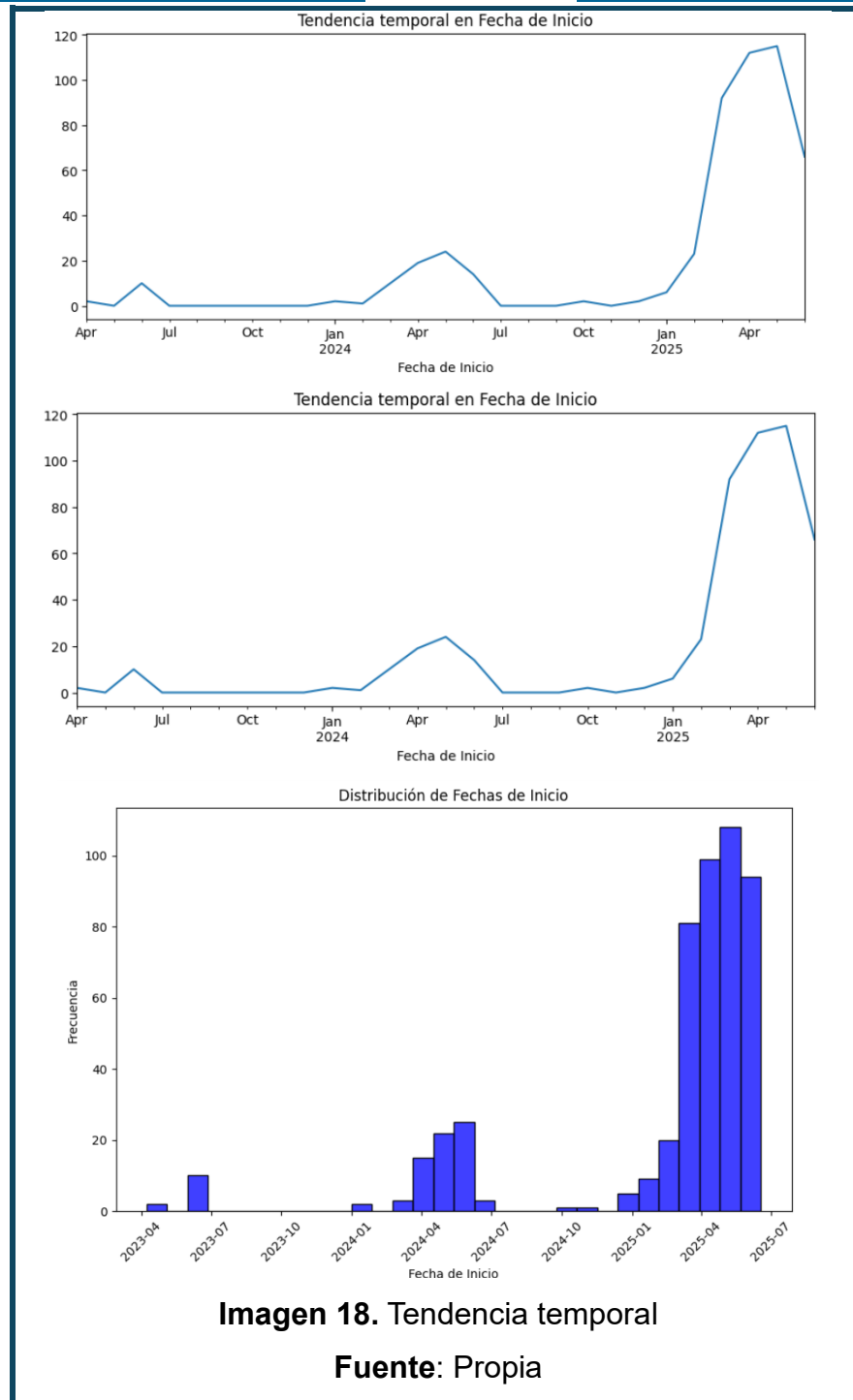
Fuente: Propia

¿Existen distribuciones sesgadas en el conjunto de datos? ¿Necesitamos aplicar alguna transformación no lineal?

No.

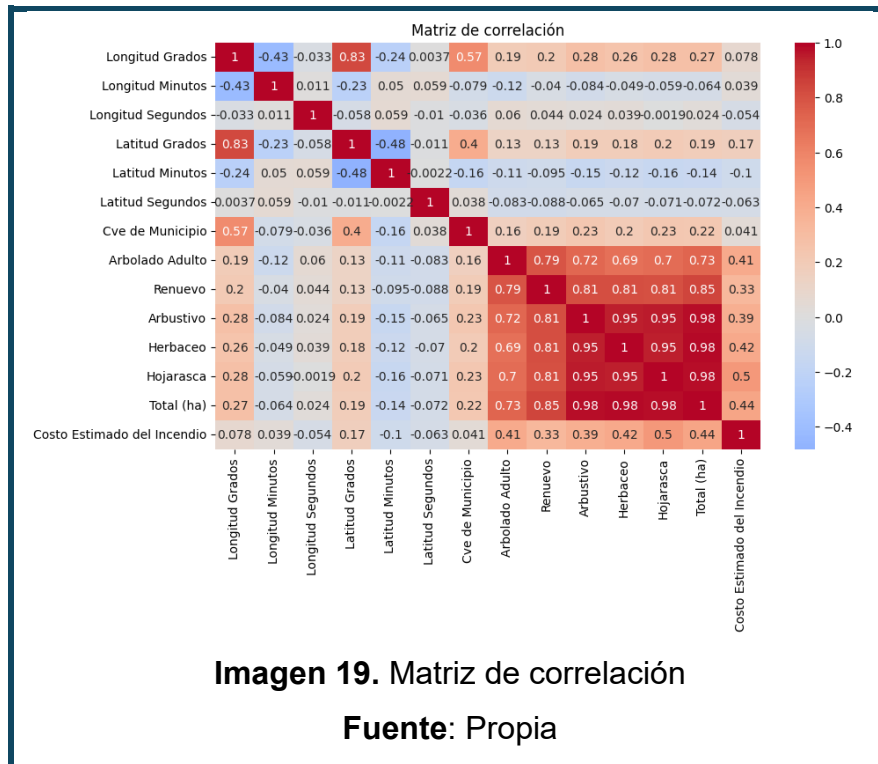
¿Se identifican tendencias temporales? (En caso de que el conjunto incluya una dimensión de tiempo).

Si, en los meses de verano, iniciando en abril de cada año.



¿Hay correlación entre las variables dependientes e independientes?

Considerando que mi variable dependiente es la cantidad de registros en un determinado periodo de tiempo. Y que mis variables independientes son las características del incendio, pero principalmente el municipio al tener enfoque en mostrar gráficamente la ubicación de las predicciones, se obtuvo el siguiente headmap



¿Cómo se distribuyen los datos en función de diferentes categorías?
(análisis bivariado)

Considerando la agrupación mes de la incidencia y ubicación, que sería el municipio, se observó la siguiente relación

	Sumatoria 2023	Sumatoria 2024	Sumatoria 2025	Total
Municipio				
Mezquital	0.0	0.0	184.0	184.0
Pueblo Nuevo	0.0	74.0	75.0	149.0
Tamazula	0.0	0.0	41.0	41.0
Durango	0.0	0.0	39.0	39.0
San Dimas	0.0	0.0	19.0	19.0
Tepehuanes	12.0	0.0	6.0	18.0
Canelas	0.0	0.0	15.0	15.0
Santiago Papasquiaro	0.0	0.0	10.0	10.0
Guanaceví	0.0	0.0	7.0	7.0
Otáez	0.0	0.0	5.0	5.0
Ocampo	0.0	0.0	4.0	4.0
Topia	0.0	0.0	4.0	4.0
Canatlán	0.0	0.0	2.0	2.0
Nombre de Dios	0.0	0.0	1.0	1.0
Cuencamé	0.0	0.0	1.0	1.0
El Oro	0.0	0.0	1.0	1.0

Top 10 Municipios - Año 2023												
Municipio	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Tepehuanes	0	0	0	2	0	10	0	0	0	0	0	0

Top 10 Municipios - Año 2024												
Municipio	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Pueblo Nuevo	2	1	10	19	24	14	0	0	0	2	0	2

Top 10 Municipios - Año 2025												
Municipio	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Canelas	0	1	5	1	7	1	0	0	0	0	0	0
Durango	4	6	6	10	9	4	0	0	0	0	0	0
Guanaceví	0	0	2	4	1	0	0	0	0	0	0	0
Mezquital	0	8	47	46	45	38	0	0	0	0	0	0
Otáez	0	0	1	1	1	2	0	0	0	0	0	0
Pueblo Nuevo	1	2	16	23	24	9	0	0	0	0	0	0
San Dimas	0	2	6	5	5	1	0	0	0	0	0	0
Santiago Papasquiaro	0	2	1	1	2	4	0	0	0	0	0	0
Tamazula	0	1	6	13	17	4	0	0	0	0	0	0
Tepehuanes	0	1	1	1	2	1	0	0	0	0	0	0

Imagen 20. Conteo para distribución de análisis bivariado
Fuente: Propia

¿Se deberían normalizar las imágenes para visualizarlas mejor?

No, en caso de que se lleguen a incorporar imágenes en algún momento del proyecto, estas serán descompuestas a otro tipo de valor, pero no se procesaran para ser visualizadas como imágenes.

¿Hay desequilibrio en las clases de la variable objetivo?

Como se observó en el análisis de incidencias basado en la ubicación, y considerando que nuestra variable objetivo es la ubicación geográfica, la probabilidad de que ocurra un evento en un determinado municipio o ubicación geográfica, el desequilibrio es fuerte, ya que se centra en la mitad de los municipios



del estado de durango, dejando áreas sesgadas con 1 incidencia al año o menos, mientras que solo 7, presentan mas de 10 incidencias al año

Fuentes Bibliográficas

- INEGI. (2023). *Clasificadores — Catálogo SCIAN 2023*. Instituto Nacional de Estadística y Geografía. <https://www.inegi.org.mx/scian/>
- Choudhury, S. (2021). *Data analysis and statistical methods for science and engineering* (2nd ed.). Wiley. <https://www.wiley.com/en-us/Data+Analysis+and+Statistical+Methods+for+Science+and+Engineering+%2C+2nd+Edition-p-9781119774797>
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media. <https://www.oreilly.com/library/view/python-data-science/9781491912127/>
- Instituto Nacional de Estadística y Geografía. (2023). *Aspectos geográficos de Durango: Compendio 2022*. INEGI.
- Enríquez Gorgonio, C. (2011). *Diseño y construcción de prototipo (CEG11) para un sistema reutilizable de detección de fuego por sensado redundante en el Parque Nacional Cofre de Perote [Tesis de licenciatura, Universidad de Xalapa]*. Xalapa, Ver.

- Belcic, I., & Stryker, C. (2024, 28 de diciembre). ¿Qué es el aprendizaje supervisado? IBM. <https://www.ibm.com/mx-es/think/topics/supervised-learning>
- Secretaría de Recursos Naturales y Medio Ambiente. (s. f.). Misión y visión. Gobierno del Estado de Durango. Recuperado el 10 de septiembre de 2025 de <https://medioambiente.durango.gob.mx/mision-y-vision%E2%82%AC%A2>
- Cámara de Diputados del H. Congreso de la Unión. (2015). Ley General de Transparencia y Acceso a la Información Pública. Última reforma publicada en el Diario Oficial de la Federación el 13 de mayo de 2021. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGTAIP.pdf>
- Cámara de Diputados del H. Congreso de la Unión. (2002). Ley Federal de Transparencia y Acceso a la Información Pública. Última reforma publicada en el Diario Oficial de la Federación el 20 de mayo de 2021. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFTAIP.pdf>
- Congreso de la Unión. (2017). Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados. Diario Oficial de la Federación, 26 de enero de 2017. Recuperado de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPPSO.pdf>
- Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). (2015). Lineamientos Generales en Materia de Clasificación y Desclasificación de la Información, así como para la Elaboración de Versiones Públicas. Diario Oficial de la Federación, 15 de junio de 2015. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5391143&fecha=15/06/2015

- Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). (2018). Lineamientos Generales de Protección de Datos Personales para el Sector Público. Diario Oficial de la Federación, 26 de enero de 2018. Recuperado de https://www.dof.gob.mx/nota_detalle.php?codigo=5512450&fecha=26/01/2018
- Congreso del Estado Libre y Soberano de Durango. (2016). Ley de Transparencia y Acceso a la Información Pública del Estado de Durango. Última reforma publicada el 25 de noviembre de 2022. Recuperado de <https://congresodurango.gob.mx/Archivos/legislacion/LEY%20DE%20TRANSPARENCIA%20Y%20ACCESO%20A%20LA%20INFORMACION%20PUBLICA.pdf>
- Congreso del Estado Libre y Soberano de Durango. (2018). Ley de Protección de Datos Personales del Estado de Durango. Recuperado de <https://congresodurango.gob.mx/Archivos/legislacion/LEY%20DE%20PROTECCION%20DE%20DATOS%20PERSONALES.pdf>
- Bonaccorso, G. (2018). Mastering Machine Learning Algorithms (2.^a ed.). O'Reilly Media. <https://www.oreilly.com/library/view/mastering-machine-learning/9781788621113/f240dfcf-f0cf-4e13-9707-2f19695e94e8.xhtml>
- Boehmke, B. (2018). Advanced machine learning with R. O'Reilly Media. <https://www.oreilly.com/library/view/advanced-machine-learning/9781838641771/>