

Buscando patrones frecuentes en delitos CDMX

Navarro Espindola
Javier
ENES Morelia
javojavojavojavo@gmail.
com

Armas Gamiño Saúl
ENES Morelia
luasikirfl@gmail.com



Figure 1: Escuela Nacional de Estudios Superiores Unidad Morelia

ABSTRACT

Este es un trabajo escolar en el cual llevaremos a cabo tareas de búsqueda de patrones, en nuestro caso, buscaremos reglas de asociación y patrones frecuentes, dentro de un conjunto de datos conformado por delitos reportados en la ciudad de México.

CCS CONCEPTS

• Análisis exploratorio de datos; • Patrones Frecuentes; • Reglas de Asociación;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Patrones Frecuentes, 2021, ENES Morelia

© 2021 Copyright held by the owner/author(s).

ACM ISBN 0.

<https://doi.org/0>

KEYWORDS

Delitos, CDMX, Patrones, Reglas, Análisis

ACM Reference Format:

Navarro Espindola Javier and Armas Gamiño Saúl. 2021. Buscando patrones frecuentes en delitos CDMX. In *Proceedings of Patrones Frecuentes*. ACM, New York, NY, USA, 6 pages. <https://doi.org/0>

1 INTRODUCCION

La ciudad de México cuenta con alrededor de 8.85 millones de habitantes, con tal cantidad de personas viviendo en el mismo lugar es inevitable que ocurran delitos en el área, nuestra misión en este trabajo es encontrar patrones de incidencia en los delitos, esperando que con base a eso se puedan reducir, evitar o predecir los eventos antes de que sea tarde.

2 DATOS

Empecemos por lo primordial, de donde obtuvimos los datos utilizados, estos provienen del gobierno, en este caso del gobierno federal,

se pone a nuestra disposición un conjunto de datos con alrededor de 800 mil registros, los atributos con los que cuenta son:

- Año hechos el cual se refiere a el año en que sucede el delito.
- Mes hechos hace referencia al mes en que ocurrió.
- Categoría delito hay 16 clasificaciones de delito.
- Geopoint coordenadas donde ocurrió el suceso.
- Calle hechos, calle hechos2 y colonia dan más información del lugar.
- Mes inicio nos dice el mes en que se denuncia el delito.
- Año inicio se refiere a en qué año se denuncia.
- Fiscalía nos dice en que fiscalía se reporta el delito
- Agencia indica la agencia encargada del caso.
- Unidad investigación nos dice en dónde se guarda el expediente.
- Delito es una descripción de lo ocurrido.
- Fecha hechos es una columna de tipo fecha, contiene cuándo ocurrió el delito.
- Fecha inicio es cuando se inicia la investigación.
- Alcaldía hechos nos dice en qué alcaldía pasó el incidente.
- longitud y latitud vienen en columnas separadas, son para poder mapear.

Para nuestro trabajo, desechamos muchos atributos que consideramos irrelevantes, como las calles de los hechos, no porque no importe saber esa información, sino que, al ser un atributo del tipo "var char", los que capturaban la información escribían cualquier cosa, por lo cual no nos sirve, otros datos los ignoramos ya que nos parecían repetitivos, si ya tenemos el campo "Fecha hechos" no necesitamos de "Año hechos", "Mes hechos", etc. Al final nos quedamos con los siguientes atributos:

- Fecha hechos
- Delito
- Alcaldía hechos
- Fecha inicio
- Latitud
- Longitud

3 EXPERIMENTOS

Para comenzar, usamos el atributo de Fecha inicio para obtener una división de día, con esto nos referimos a obtener clasificaciones como "día", "tarde" y "noche", hicimos dos pruebas, en una dividimos el día en 3, considerando que "día" es desde las 0:00 hasta las 8:00 horas, tarde desde las 8:00 hasta las 16:00 horas, y la noche el resto, y en el segundo experimento dividimos solo en dos el día, siendo en este caso el día desde las 0:00 hasta las 12:00 horas, y la noche las horas restantes.

Lo primero que hicimos fue hacer una gráfica, en cada experimento, la cual muestra la relación entre los patrones frecuentes encontrados y el umbral usado, esto para darnos una ayuda visual acerca de cuál sería el umbral más apropiado para nuestro trabajo.

3.1 Dividiendo las horas del día en día y noche

Vamos a comenzar hablando de los resultados obtenidos con este experimento. Lo primero es la gráfica de umbrales y patrones mencionada al inicio, en este caso la gráfica es la siguiente:

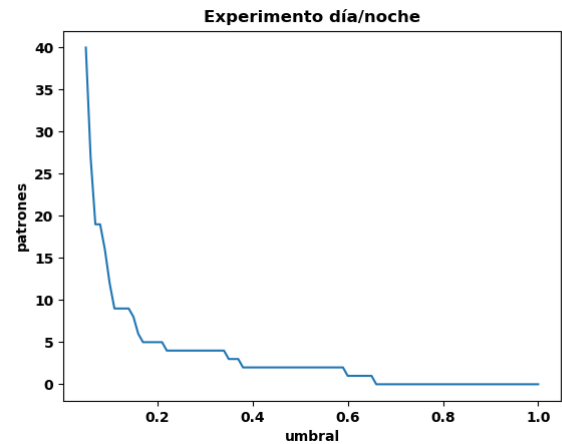


Figure 2: Grafica que representa la relación entre la cantidad de patrones frecuentes encontrados con el umbral utilizado, en la cual consideramos el día en dos secciones.

Como mencionamos antes, esta gráfica es una ayuda para determinar el umbral que usaríamos en el trabajo, pero no consideramos solamente eso, tomamos en cuenta la cantidad de registros con la cual estamos trabajando, además de cuántas categorías diferentes pueden tener, por ejemplo, solo del campo "delito" hay 274 posibles valores que puede tener, sumado a esto, el campo de "Alcaldía hechos" tiene 486 valores posibles.

Tomando en cuenta lo anterior, además de recordar que el conjunto de datos cuenta con más de 800 mil registros, consideramos que un buen umbral para buscar patrones frecuentes es 6%, ya que sería el equivalente a que el patrón aparezca 48 mil veces.

3.1.1 Resultados patrones frecuentes. Con base a todo lo anterior, una vez que determinamos un umbral para experimentar, lo siguiente que hicimos fue obtener los patrones frecuentes "generales" de cualquier tipo de delito en cualquier alcaldía, los resultados se presentan en las siguientes tablas.

En la primer tabla ponemos los delitos que se repiten al menos 48 mil veces en el registro.

Delitos más frecuentes		
Soporte en %	Patrón encontrado	casos aproximados
6.35	Robo a negocio sin violencia	50,862
6.45	Robo de objetos	51,641
8.59	Violencia Familiar	68,754

Table 1: En esta tabla se puede observar que los delitos más comunes en la CDMX son el robo a negocio sin violencia, el robo de objetos, y la violencia familiar.

A continuación enlistaremos las alcaldías donde se registran más delitos.

Alcaldías con más incidencia		
Soporte en %	Alcaldía	Casos aproximados
6.49	Álvaro Obregon	51,991
8.72	Benito Juárez	69,824
6.70	Coyoacán	53,640
16.24	Cuauhtémoc	129,955
9.90	Gustavo A Madero	79,218
14.82	Iztapalapa	118,610
6.63	Miguel Hidalgo	53,099

Table 2: Esta tabla muestra las alcaldías donde ocurren más delitos, por lo cual consideramos que se debería poner más atención en ellas.

La siguiente tabla es acerca de los delitos ocurridos en el día y en la noche, además de decirnos cuánto tiempo se tardan en denunciar un delito.

Delitos y denuncias		
Soporte en %	Tiempo	Casos aproximados
34.97	Día	279,813
65.02	Noche	520,186
59.02	Denuncia en 1 día	472,162
15.73	Más de un mes en denunciar	125,905
10.16	Un mes en denunciar	81,291
15.08	Denuncian en una semana	120,640

Table 3: Aquí podemos ver las horas donde ocurren más delitos, y cuánto es lo más común que se tarden en denunciarlos, resaltando que en este experimento, la noche es cuando ocurren más delitos, y que lo más comun es que se denuncien en un día o menos

Ahora veamos una tabla con información respecto a las denuncias en algunas alcaldías.

Denuncias en alcaldías			
Soporte en%	Alcaldía	Información del reporte	Casos
8.89	Cuauhtémoc	Tardaron un día en denunciar	71,132
6.40	Gustavo A Madero	Delito ocurrido en la noche	51,262
6.23	Gustavo A Madero	Denunciado en 1 día o menos	49,904
9.47	Iztapalapa	Delito ocurrido en la noche	75,829
9.25	Iztapalapa	Denunciado en 1 día	74,076

Table 4: Esta tabla nos da la información relacionada a en cuánto tiempo se denuncia un delito en diferentes alcaldías, y en qué momento ocurren los delitos, no es de extrañar que aparezcan las mismas alcaldías que en la Tabla 2, y que en este caso nos diga que los delitos ocurrieron en la noche y se denunciaron en menos de un día, pues ya sabíamos que es lo más común, en base a la Tabla3

Por último, obtuvimos información que nos relaciona el tiempo que tardó la gente en denunciar un delito, con la hora en la que ocurrió éste.

Relación tiempo denuncia, hora delito			
Soporte en%	Tiempo hechos	Tiempo en denunciar	Casos estimados
21.27	Día	Denuncia en 1 día	170,185
37.47	Noche	Denuncia en 1 día	301,976
10.73	Noche	Más de un mes	85,887
6.72	Noche	Tardaron un mes	53,783
9.81	Noche	Denuncia en una semana	78,539

Table 5: En esta tabla se puede apreciar el porcentaje de casos en el que se hacen las denuncias, dependiendo de en qué momento se llevaron a cabo los delitos, nuevamente aparece más el momento de "Noche" que el de "Día", y como era de esperarse por la información obtenida antes, el tiempo de denuncia con más porcentaje es "un día o menos"

3.1.2 *Resultados Reglas de asociación.* Lo siguiente que realizamos en nuestro trabajo fue buscar reglas de asociación, para esto continuamos con el umbral de al menos 6% de los registros, y utilizamos una confianza de al menos 50%, debido a que consideramos importante encontrar estas asociaciones, ya que en el tema de seguridad, creemos que si algo es al menos 50% probable que pase, debe ser atendido y no ignorado, además buscamos la mayor cantidad de asociaciones posibles pero sin usar una confianza muy baja, a fin de evitar que nos de resultados de cosas muy poco probables.

las reglas de asociación que encontramos son las que aparecen en la siguiente tabla.

Reglas De Asociación		
Confianza%	Antecedente	Consecuencia
66.07	Delito en Cuauhtémoc	Pasó en la noche
54.73	Delito en Cuauhtémoc	Denuncia en 1 día
64.71	Delito en Gustavo A Madero	Pasó en la noche
62.99	Delito en Gustavo A Madero	Denuncia en 1 día
63.93	Delito en Iztapalapa	Pasó en la noche
62.45	Delito en Iztapalapa	Denuncia en 1 día
60.82	Delito ocurrido de Día	Denuncia en 1 día
58.05	Delito ocurrido de Noche	Denuncia en 1 día
68.21	Más de 1 mes en denunciar	Pasó en la noche
66.16	Más de 1 mes en denunciar	Pasó en la noche
65.10	Denunciaron en una semana	Pasó en la noche

Table 6: Reglas de asociación encontradas, podemos obtener de éstas que, un delito en Cuauhtémoc, Iztapalapa o Gustavo A Madero, implica que pasó de noche y se denunció en menos de un día, además de que los delitos ocurridos en la noche son más propensos a tardar más en ser denunciados.

3.2 Dividiendo las horas del día en: Día, Tarde y Noche

Hemos hecho lo mismo que en la sección anterior, pero en este caso el día lo partimos en 3 partes, con esto buscamos ver si hay variación notable entre los periodos seleccionados.

Comenzamos igual haciendo una gráfica de la relación entre patrones encontrados y el umbral utilizado.

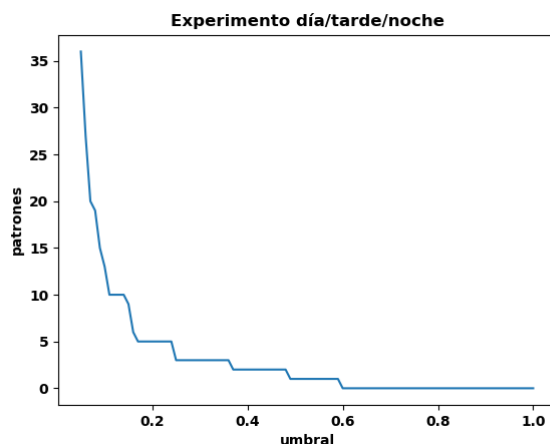


Figure 3: Gráfica que representa la relación entre la cantidad de patrones frecuentes encontrados con el umbral utilizado, tomando en cuenta el día en 3 secciones.

Como se puede apreciar, el cómo dividimos el día sí tiene un impacto a la hora de buscar patrones frecuentes, recordemos que estamos considerando 6% como umbral, considerando eso, y que ahora el día son 3 partes, los resultados para patrones frecuentes son los siguientes.

3.2.1 Resultados Patrones Frecuentes. Delitos más comunes.

Delitos más frecuentes		
Soporte en %	Patrón encontrado	casos aproximados
6.35	Robo a negocio sin violencia	50,862
6.45	Robo de objetos	51,641
8.59	Violencia Familiar	68,754

Table 7: Delitos más frecuentes en la CDMX considerando el día en 3 secciones, como se puede apreciar, no se afectó el resultado obtenido considerando el día 2 partes, ya que los delitos más comunes siguen siendo el robo a negocios sin violencia, el robo de objetos y la violencia familiar.

Tabla con los delitos por alcaldía.

Alcaldías con más incidencia		
Soporte en %	Alcaldía	Casos aproximados
6.49	Álvaro Obregón	51,991
8.72	Benito Juárez	69,824
6.70	Coyoacán	53,640
16.24	Cuauhtémoc	129,955
9.90	Gustavo A Madero	79,218
14.82	Iztapalapa	118,610
6.63	Miguel Hidalgo	53,099

Table 8: Comparando esta tabla con su homóloga en el apartado 3.1, podemos darnos cuenta que considerar el día en 3 partes no afecta tampoco el resultado de las alcaldías con más delitos, lo cual es lógico, ya que siguen teniendo la misma cantidad de casos sin importar en qué momento del día pasan.

Tabla con las horas de los delitos y los tiempos de denuncia, además de los casos que reportan.

Hora del delito		
Soporte %	Hora	Casos
15.25	Día	122,002
36.43	Noche	291,495
48.31	Tarde	386,502
Tiempo de denuncia		
58.02	Un día	472,162
15.73	Más de un mes	125,905
15.08	Una semana	120,640
10.16	Un mes o menos	81,291

Table 9: Esta tabla tiene dos apartados, en el primero, nos describe el porcentaje de delitos ocurridos en cada momento del día, destacando que en la tarde es cuando más suceden éstos, seguidos de la noche, y al final el día, el segundo apartado nos muestra los porcentajes de denuncia de cada tiempo, aquí sobresale un día, lo que significa que los delitos son en su mayoría denunciados casi al momento de ocurrir.

Tabla del tiempo en que ocurre el delito, y el tiempo que se tarda en denunciar.

Relación tiempo delito y denuncia			
Soporte %	Hora	Tiempo en denunciar	casos
10.36	Día	Un día	82,921
24.14	Noche	Un día	193,140
24.51	Tarde	Un día	196,100
10.87	Tarde	Mayor a un mes	87,025
7.20	Tarde	Una semana	57,678

Table 10: En esta tabla claramente se puede notar un cambio a su homóloga en la sección 3.1, pues en este caso el día son 3 partes, lo cual claramente afectaría los resultados que involucren el momento del día, ahora aparece más el momento "tarde" que los otros dos, ya que es el momento donde ocurren más delitos.

Por último, tenemos una tabla de la relación de alcaldías con horas de delitos y tiempos en denunciar.

Relación alcaldías horas delito			
Soporte%	Alcaldía	Hora	casos
8.82	Cuauhtémoc	Tarde	70,568
6.59	Iztapalapa	Tarde	52,747
Relación alcaldía, tiempo denuncia			
Soporte%	Alcaldía	Tiempo denuncia	casos
8.89	Cuauhtémoc	Un día	71,132
6.23	Gustavo A Madero	Un día	49,904
9.25	Iztapalapa	Un día	74,076

Table 11: Esta tabla tiene dos secciones, en la primera nos muestra la relación entre la alcaldía y la hora en que sucede el delito, se destacan las alcaldías de Cuauhtémoc e Iztapalapa con más casos, la segunda sección nos dice la relación entre dónde ocurrió el evento y cuánto tiempo tomó su denuncia, resalta el hecho de que en cualquier alcaldía el tiempo de denuncia fue un día o menos.

3.2.2 Resultados Reglas de Asociación. Para obtener las reglas de asociación, usamos una confianza de al menos 50%, y el mismo umbral de 6% o más, los resultados son los siguientes.

Reglas de Asociación encontradas		
Confianza	Antecedente	Consecuencia
54.30%	Sucedió en Cuauhtémoc	Pasó en la tarde
54.73%	Sucedió en Cuauhtémoc	Se denunció en un día
62.99%	Sucedió en Gustavo A Madero	Se denunció en un día
62.45%	Sucedió en Iztapalapa	Se denunció en un día
67.96%	Sucedió en el Día	Se denunció en un día
66.25%	Sucedió en la Noche	Se denunció en un día
50.73%	Sucedió en la Tarde	Se denunció en un día
69.11%	Tiempo en denunciar más de 1 mes	Sucedió en la Tarde

Table 12: Éstas son la reglas de asociación que se encontraron, como hay 2 que involucran a Cuauhtémoc como antecedente, podemos decir que, si pasó en dicha alcaldía, es probable que haya pasado en la tarde y se denunciara en un día o menos.

3.3 Analizando un solo delito

Por falta de tiempo, no pudimos hacer un análisis delito por delito, lo cual consideramos nos permitiría establecer umbrales más grandes, pero si pudimos hacer este ejercicio con un delito, en específico analizamos el delito de Fraude. Hicimos una gráfica de los umbrales y los patrones encontrados, y nos dio el siguiente resultado.

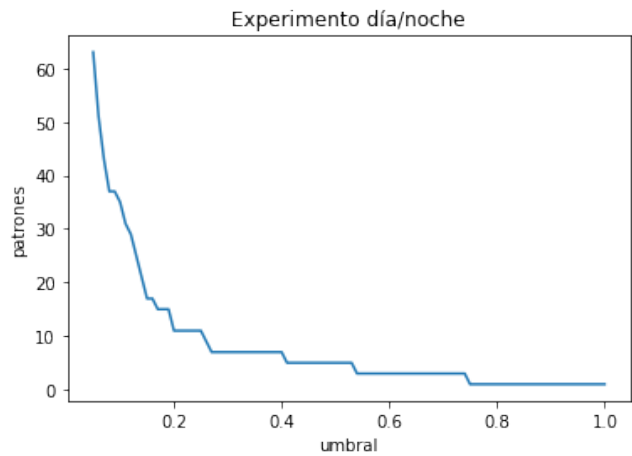


Figure 4: Grafica que representa la relación entre la cantidad de patrones frecuentes encontrados con el umbral utilizado, en este caso solo para el delito del Fraude

3.3.1 Patrones frecuentes del Fraude. Para obtener dichos patrones, como solo estamos haciendo una demostración de que al analizar delitos individuales podemos usar umbrales más altos, consideramos un 30% de umbral, 5 veces más alto que el umbral utilizado al analizar todos los delitos a la vez. Es oportuno mencionar que, el delito de fraude cuenta con 45,349 registros, por lo cual, no obtuvimos información de este delito en lo particular en los experimentos anteriores, aun al usar un umbral de solo el 6%,

Reglas encontradas en el Fraude		
Confianza	Antecedente	Consecuencia
74.27%	Fraude	Cometido en la noche
100%	Pasó de noche	Fraude
53.90%	Fraude	La denuncia tardó más de un mes
100%	La denuncia tardó más de un mes	Se hizo Fraude
74.61%	La denuncia tardó más de un mes	Pasó de noche
100%	La denuncia tardó más de un mes y el delito pasó de noche	Se hizo fraude

Table 14: Para finalizar el análisis individual, hemos obtenido las reglas de asociación descritas con anterioridad, este análisis individual nos describe al delito en particular, lo cual consideramos importante, ya que se requiere información concreta y específica de un problema para poder solucionarlo.

debido a como comentamos al inicio, la cantidad de registros totales de todos los delitos existentes. Los patrones obtenidos son los siguientes.

Patrones encontrados		
Soporte %	Patron	Casos
74.27	Pasó en la noche	33,680
53.9	La denuncia tardó más de un mes	24,443
40.21	pasó de Noche y se tardó más de un mes la denuncia	18,234

Table 13: Llevando a cabo el análisis de un solo delito nos fue posible pedir umbrales más altos para determinar patrones, en este caso 5 veces mas grande, y aun así obtuvimos 3 patrones que sobrepasan considerablemente dicho umbral, es destacable que casi todos los fraudes pasen de noche, además de que su denuncia toma mucho tiempo en la mayoría de casos.

3.3.2 Reglas de asociación del Fraude. Las siguientes reglas las obtuvimos usando un umbral de 30%, y pedimos una confianza de al menos 50%, pero cabe resaltar que casi todos los resultados superan ese porcentaje considerablemente.

4 ANÁLISIS DE LOS RESULTADOS

Lo primero en llamarnos la atención fue, que efectivamente, como pensamos en un inicio, la mayoría de los delitos ocurren en la noche, o al menos en horas con poca luz, lo cual quiere decir, evidentemente, que es más probable ser víctima de algún delito en lugares con poca luz, así como que es más seguro si se está en un lugar a plena luz del día.

Lo segundo que notamos, y nos pareció algo muy curioso, y hasta cierto punto controversial, es respecto a el tiempo que tardan

las personas en denunciar un delito, ya que, lo más común es que denuncien en menos de un día (59%), seguido de que denuncian hasta después de un mes (15.7%) sumando más de 74% del total de delitos. Nos llama la atención las personas que denuncian en más de un mes. ¿Será que necesitan tiempo para procesar lo que pasó? ¿Igual y tienen miedo a represalias? ¿Probablemente no saben el proceso de denuncia por falta de información, no saben a dónde ir? ¿Igual no están muy convencidos de que vaya a servir que denuncien? Esto nos hace preguntarnos, ¿Cuántos delitos cometidos no son denunciados?

Nos llama la atención que el delito más frecuente entre las 274 categorías existentes fue violencia familiar. Viendo los resultados nos preguntamos ¿por qué nos sorprenden? No hemos sabido de muchas personas que denuncien o sufran este delito. Probablemente es porque es algo vergonzoso o ¿tabú? que no le cuentas a las personas, mientras que las denuncias de robo, secuestro o violaciones salen incluso en las noticias, y parece algo cotidiano, por lo cual esperamos que alguno de estos estuviera en el top.

Por último, pero no menos importante, el hecho de que casi todos los fraudes se cometan en la noche, nos hace pensar que tal vez, quien comete este delito, busca la menor cantidad de testigos, por lo cual, esperan un momento en que la mayoría de la gente no esté atenta a lo que hacen, para evitar mayores consecuencias, ya que cuando pensamos en fraude, lo primero que se nos viene a la mente son fraudes de tipo políticos o institucionales, y al parecer, dado el tiempo que lleva denunciarlos (más de un mes en su mayoría), se escondió bien la evidencia. Probablemente el tiempo de denuncia también sea largo porque el proceso para producir documentación y pruebas necesarias para levantarla es laborioso (burocracia). Podríamos sugerir que se revise el proceso de denuncia, para ver si no resulta perjudicial, o lento.

5 CONCLUSIONES Y TRABAJO FUTURO

En conclusión, investigar los patrones generales de los delitos nos da una orientación de lo que se puede hacer para evitarlos o prevenirlos, pero para poder realmente hacerle frente a la problemática social que representan, es necesario hacer un análisis más específico, también consideramos que para algo tan complejo como evitar que sucedan delitos, no es suficiente con saber el cuándo, qué y dónde, sino que además se requiere saber el porqué, y lamentablemente no contamos con esa información.

Un posible trabajo a futuro sería hacer el análisis individual de cada uno de los 274 delitos, además de hacerlos para cada una de las alcaldías, ya que así encontraremos patrones más locales y específicos, y no tan generales, debido a que, como vimos en algún momento en la clase, hay problemas que no se pueden resolver con medidas generales, y se deben tomar acciones que funcionen en el contexto de cada localidad.

REFERENCES

- [1] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *IEEE Annals of the History of Computing*, 9(03), 90-95.
- [2] Raschka, S. (2016). *MLxtend*.
- [3] Bressert, E. (2012). *SciPy and NumPy: an overview for developers*.
- [4] Procuraduría General de Justicia de la Ciudad de México. (2019, 7 agosto). *Carpetas de investigación PGJ de la Ciudad de Mexico (archivo)*. <https://datos.cdmx.gob.mx/explore/dataset/carpetas-de-investigacion-pgj-cdmx/table/>