

Centro de Estudios Sociales y de Opinión Pública

¿Por qué automatizar? Propuesta preliminar mínima de innovación de la investigación parlamentaria

Documento de trabajo núm. 346



Marzo 2021

www.diputados.gob.mx/cesop

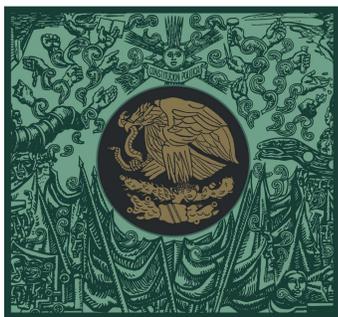


**CÁMARA DE
DIPUTADOS**
LXIV LEGISLATURA

CESOP

Centro de Estudios Sociales y de Opinión Pública

Información que fortalece el quehacer legislativo



**CÁMARA DE
DIPUTADOS**
LXIV LEGISLATURA

Información que fortalece
el quehacer legislativo

CESOP

Centro de Estudios Sociales y de Opinión Pública

Centro de Estudios Sociales y de Opinión Pública

Organización Interna

Netzahualcóyotl Vázquez Vargas

Director de Estudios Sociales encargado del
despacho de la Dirección General del CESOP

Enrique Esquivel Fernández
Asesor General

Ricardo Martínez Rojas Rustrian
Director de Estudios de Desarrollo Regional

Katia Berenice Burguete Zúñiga
Coordinadora Técnica

Investigadores

Gabriel Fernández Espejel
José de Jesús González Rodríguez
Salvador Moreno Pérez
Felipe de Alba Murrieta

Apoyo en Investigación

Natalia Hernández Guerrero
Karen Nallely Tenorio Colón
Ma. Guadalupe S. Morales Núñez
Nora Iliana León Rebollo
Alejandro Abascal Nieto
Abigail Espinosa Waldo

Alejandro López Morcillo
Editor



Felipe de Alba
Winik I. Morales
Aeelen Miranda

Tabla de materias

Presentación	3
1. ¿Qué es automatizar?	6
a) “Las máquinas que hablan”	9
2. Objetivos de la investigación parlamentaria.....	12
3. Ejemplo de procesos factibles de automatización	15
4. La incidencia delictiva con procesos de automatización	17
a) Visualización de datos	18
b) Algoritmos y automatización	19
i) Extracción de datos.....	20
ii) Recopilación y almacenamiento de datos.....	21
iii) Agrupamientos y tendencias.....	22
iv) Presentación de la información (tabla de datos)	24
A manera de conclusión.....	25
Anexo	266



¿Por qué automatizar? Propuesta preliminar mínima de innovación de la investigación parlamentaria

Felipe de Alba
Winik I. Morales
Aeelen Miranda¹

Presentación

En este documento de trabajo presentamos una reflexión sobre la innovación institucional en la Cámara de Diputados, para proponer una línea de cambios que hoy es indispensable. Esta propuesta está relacionada con la innovación: hacer que la investigación parlamentaria sea cada vez más profesional y basada en evidencia. Sobre todo, la idea aquí es organizarla de tal manera que aporte insumos oportunos y expeditos.

Con esta propuesta intentamos responder, en general, a la pregunta ¿cómo construir procesos cada vez más eficientes en el Poder Legislativo mexicano? Por ello, en el documento nos enfocamos en necesidades de innovación en la tarea del investigador parlamentario, específicamente en necesidades de adaptarla a los desafíos de futuro. Pero antes de entrar en el tema tenemos que preguntarnos cuáles son los desafíos de futuros que identificamos aquí.

¹ Felipe de Alba es doctor en Planeación Urbana por la Universidad de Montreal (2004-2008) con estancias de dos años en el Massachusetts Institute of Technology (MIT) (2009-2011) y de un año en l'École normale supérieure (ENS) de Lyon (Francia) (2012). También fue profesor invitado de tiempo completo "C" en la Universidad Autónoma Metropolitana (Cuajimalpa) (2012-2014). Es Investigador "A" del Centro de Estudios Sociales y Opinión Pública (CESOP) de la Cámara de Diputados. Investigador SNI-1. Contacto fdealbam@gmail.com. Aeelen Miranda es licenciada en Estudios Socioterritoriales, por la UAM Cuajimalpa, experta en SIG y Python, con experiencia profesional en consultoría TI. Ha participado en proyectos de investigación para transformación e innovación tecnológica en empresas y sector público. Finalmente, Winik Ivanovich Morales es ingeniero ambiental por el Instituto Tecnológico de Misantla (Veracruz), ha colaborado en varias publicaciones anteriores tanto del CESOP como de la Cámara de Diputados. Actualmente es consultor ambiental.



Traducción de la cita: *La primera regla de cualquier tecnología usada en un negocio es que la automatización aplicada para una operación eficiente magnificará la eficiencia. La segunda es que la automatización aplicada a la ineficiencia magnificará la ineficiencia.*

En primer lugar, se trata de pensar que la modernidad, al menos en términos tecnológicos, ha llegado a casi todas las áreas de la acción social o económica. Las instituciones gubernamentales a veces son más resistentes a cambios que modifiquen su hacer cotidiano. En segundo lugar, se trata de innovar no sólo por el hecho de “cambiar”, sino porque innovar representa o puede desencadenar sinergias de mejora continua.

Entonces, no se trata de enfocarnos en una propuesta que cambie un procedimiento o un proceso dentro del ámbito legislativo, sino más bien de “insertar”, involucrar al Poder Legislativo en un proceso de cambios que puedan crear sinergias, que desarrollen dinámicas de equipo y no medidas aisladas para “madurar” procesos y, por tanto, se trata de buscar cambios que beneficien el quehacer legislativo en general y no a un área o equipo específico.

Para ello tenemos que reflexionar en qué es necesario y en qué no es necesario innovar. La pregunta no tiene desperdicio porque existen seguramente muchas áreas en la práctica legislativa que no requieren de modificación alguna. No nos detendremos aquí en ello, más bien, en un espíritu creativo, buscaremos orientarnos en una necesidad que

involucra la práctica legislativa cotidiana como, por ejemplo, el quehacer de los centros de estudios en relación con las comisiones legislativas.

Un primer eje rector de esta reflexión está fundado en que en algunas áreas o procesos del quehacer legislativo las tareas o procedimientos son repetitivos y eso consume un gran tiempo. Esta “repetición” es lo que permite suponer que podríamos innovar. En este caso podemos decir que el concepto de *innovación* tiene en un primer sentido —al definirse como una intervención con el objetivo de evitar procesos repetitivos— el propósito de mejorar la velocidad de los procesos, lo cual no sólo genera un gran impacto en las políticas nacionales, sino que permite una mayor competitividad.

En consecuencia, la mayor velocidad en la respuesta a una demanda hace que dicha respuesta sea más eficiente. Entonces, la innovación la definiremos como una herramienta útil para ayudar a resolver intervalos de “tiempo muerto” (*deadtime*) producto de la repetición. Así, las diferencias en las tasas de progreso institucional están fundadas en las diferencias en la velocidad de sus resultados. Concluamos que, con la mejora de un proceso repetitivo, la respuesta es por definición más eficiente y, por tanto, expedita.

Un segundo eje rector de esta reflexión está fundado en el hecho de que para innovar es necesario crear o recrear respuestas expeditas o reemplazar procesos repetitivos a partir de la **utilización de sistemas tecnológicos nuevos**. Aquí es donde entra la **automatización**, vista como un proceso de “tecnificación” de procesos que normalmente se hacen de manera manual, o bien la automatización es la tecnificación con dispositivos que generalmente reemplazan el trabajo de un determinado número de personas o que eliminan un número de procesos que hacen más largo el tiempo de respuesta a una exigencia.

Dicho de esta manera, la innovación funciona entonces como el motor de sinergias creativas, las cuales están íntimamente ligadas a la tasa de innovación tecnológica, siguiendo nuestro ejemplo.

En suma, los países que han logrado tasas de crecimiento importantes han experimentado una reestructuración de sus aparatos productivos o de sus procesos institucionales de atención, en los cuales se hace uso de tecnologías más eficientes, tal

como sugiere Schumpeter, quien afirmaba desde hace décadas que “el desarrollo se asocia a los efectos tecnológicos y sociales de la innovación”.²

En cualquier caso, la innovación es un concepto útil para identificar y analizar los distintos elementos que conforman el comportamiento tecnológico de empresas o instituciones. Es un concepto que permite ordenar la diversa información proveniente de distintos casos, y ayuda a la elaboración de hipótesis que, para autores como Schumpeter, explicaría la diferencia en los resultados económicos de distintos países.³

Con estos dos ejes rectores, varios autores sugieren que la innovación lograría disminuir el tiempo de respuesta y “tecnificar” los pasos que dicha respuesta tendría a través de la automatización. La innovación es en sí misma una herramienta que permite derivar recomendaciones de política pública para aquellos países en desarrollo que han podido alcanzar altas tasas de crecimiento económico. Esta última idea proviene de J. H. Rima, para quien “el crecimiento económico se origina en los cambios en la producción generados por la disponibilidad de factores productivos que resultan de la innovación impulsada por los empresarios que tienen la capacidad de liderazgo para ello”.⁴

Aunque esta definición nos resuelve una primera etapa, necesitamos profundizar en el concepto de *automatizar*, tal como pretendemos hacerlo enseguida.

Esperamos que este documento pueda generar interés en nuestros destinatarios directos: los representantes populares de este Poder Legislativo.

1. ¿Qué es automatizar?

El camino de la automatización surge como principio de la evolución del hombre que, limitado en su movimiento y en su fuerza, buscaba simplificar sus tareas y expandir su

² J. A. Schumpeter (1939), *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, translated from the German by Redvers Opie, New Brunswick (U.S.A) and London (U.K.): Transaction Publishers, Harvard University Press, Cambridge, pp. 290.

³ J. A. Schumpeter (1939), *Ibid.* p. 290.

⁴ J. H. Rima (1972), *Development of Economic Analysis*, Richard Irving, p. 97.

dominio, es decir, el control de las fuerzas naturales. Así la automatización con la adquisición del conocimiento y su comprensión de las leyes físicas, lo que "constituyó el primer paso en la realización de su sueño de volar por el aire como un pájaro, nadar en el mar como un pez, y para convertirse en gobernante de toda la naturaleza."⁵

La automatización, como objetivo de innovación tiene una larga historia. Puede anotarse que los primeros sistemas de control surgen en la Revolución Industrial a finales del siglo XIX, y continuó sobre todo a principios del siglo XX.

Sin embargo, dicha automatización se refiere hoy no a procesos mecánicos y electromagnéticos, básicamente. Antes se trataba de engranajes, de palancas y de pequeños motores. Dicho específicamente, la diferencia entre un proceso físico y el proceso técnico radica en el hecho de que el proceso físico no necesariamente tiene que ser controlado desde el exterior, mientras que el proceso técnico incluye tratamiento de la información con fines de control.⁶ Esa automatización "técnica" es la temática de este documento.

Una de las principales ventajas de la automatización es su gran versatilidad y una gran variedad de aplicaciones,⁷ aunque nos concentraremos en aquella automatización que tiene lugar en las últimas seis o siete décadas, organizada y realizada por computadoras personales o en red: la automatización por procesos informáticos.

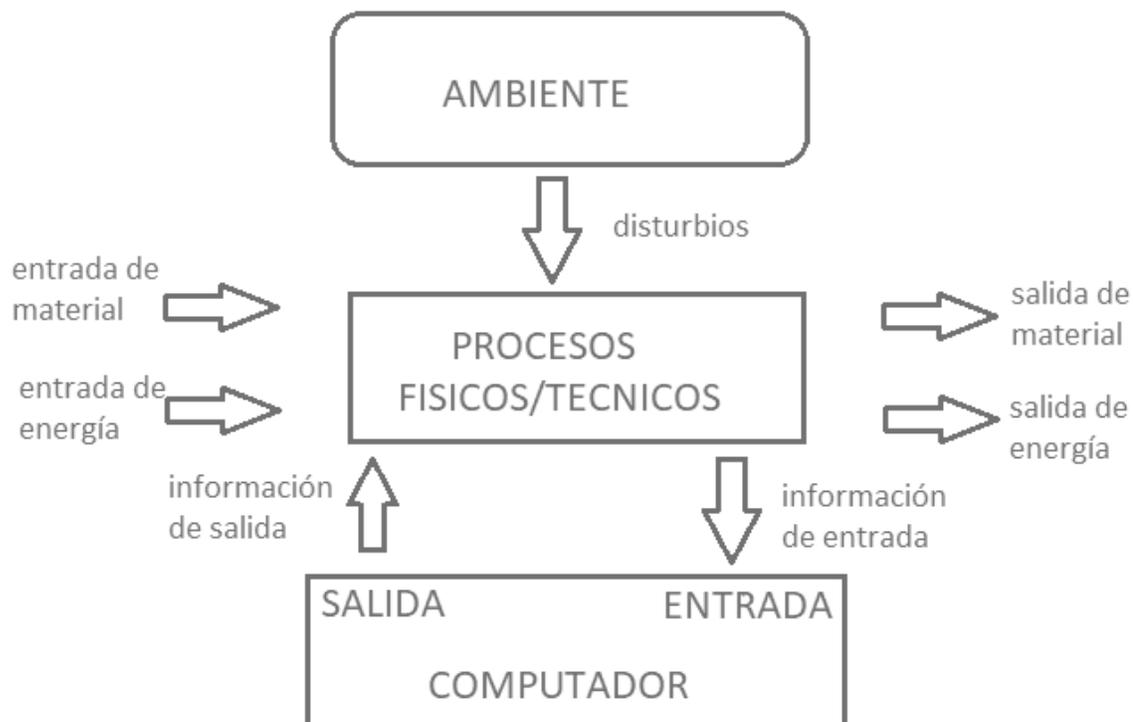
El proceso de automatización es aquel que incluye las partes físicas y técnicas de un proceso de investigación tal como se muestra en la **Infografía 1**.

Infografía 1. Proceso de automatización

⁵ Noble, D. (2017). *Forces of production: A social history of industrial automation*. Routledge, p. 302.

⁶ Olsson, G. & Piani, G. (1992). *Computer systems for automation and control* (pp. 155-194). New York: Prentice Hall, p. 1.

⁷ Mitsubishi, El libro de la automatización. Un mundo lleno de soluciones. En línea: <https://www.ncsolutions.es/wp-content/uploads/2016/04/El-libro-de-la-automatizacion.pdf> (consulta: 1 de febrero de 2021).



Fuente: Elaboración propia con Información de Olsson G., Piani, G., p. 32.

La **innovación** tanto quizá como la **automatización** es una más y no la única de las actuales estrategias de modernización en la Cámara de Diputados. Al respecto, desde nuestro punto de vista, existe una serie de necesidades impostergables en materia de innovación legislativa, en rubros específicos. Enseguida identificamos cinco puntos.

Primero, es necesario desarrollar análisis con mayor profundidad en todo el proceso legislativo; segundo, es necesaria la creación de bases de datos especializadas para consulta de investigadores; tercero, es necesario mejorar el tiempo de respuesta a las exigencias de información y análisis que las comisiones legislativas tienen; cuarto, es necesario dotar a la Cámara de información certera, con mayor profundidad y análisis exhaustivo, a partir de la consulta del mayor número de fuentes disponibles; finalmente, quinto, es necesario usar la automatización como herramienta durante el proceso de creación de iniciativas de ley, para que ésta se fundamente en investigaciones con evidencia registrada en datos duros, fortaleciendo así el debate legislativo.

a) “Las máquinas que hablan”

Steve Jobs —el fallecido dirigente tecnológico de lo que hoy es **Apple**, una de las compañías informáticas más importantes en el mercado—, siempre soñó con una “máquina amiga” que facilitara las tareas cotidianas de las personas, en una oficina, en cualquier lugar. Esta máquina amiga era una revolución en sí misma.⁸

En dicho sentido, aquí se tratará de una automatización definida por el *Machine Learning* (o aprendizaje de las máquinas) y por la Inteligencia Artificial (AI) que ha cambiado el rostro de la investigación científica, industrial e institucional en el mundo y, por ende, el proceso de toma de decisiones.

Por una parte, hay muchas definiciones del concepto *Machine Learning*, pero acordemos aquí que “es la aplicación de técnicas y algoritmos

capaces de aprender a partir de distintas y nuevas fuentes de información, construyendo algoritmos que mejoren de forma autónoma con la experiencia”.⁹ Esto permite disponer de métodos capaces de detectar automáticamente patrones en los datos, y usarlos para predecir sobre datos futuros en un periodo de incertidumbre.

Si algo caracterizó durante toda su vida al fallecido Steve Jobs fue ser capaz de imaginar lo que nadie sabía cómo fabricar y después fabricarlo y hacerlo humano.

Jobs soñó allá por 1974 con hacer una máquina amiga; una máquina para cuyo manejo no fuesen necesarios conocimientos de informática, ni haber oído nunca hablar de lenguajes de programación.

Su unión con Steve Wozniak, la otra mitad de la primitiva Apple, redondeó el proyecto. Steve Jobs imaginaba cosas imposibles y Wozniak conseguía hacer que funcionasen.

En 1976 en un garaje - ¿qué hubiera sido de la innovación americana si no hubiesen existido los garajes? - nació Apple I, que no era más que un teclado con un procesador conectados a un monitor.

Fuente: La Vanguardia

⁸ Puede verse el video original que hizo historia “The Lost 1984 Video: young Steve Jobs introduces the Macintosh”. En línea: <https://www.youtube.com/watch?v=2B-XwPjn9YY> (consulta: 28 de febrero 2021). También puede consultarse una reseña interesante sobre el trabajo que este dirigente tecnológico realizó en toda su vida, en “Steve Jobs, una máquina de imaginar”, *La Vanguardia*, 6 de octubre de 2011. En línea: <https://www.lavanguardia.com/tecnologia/20111006/54226605783/steve-jobs-una-maquina-de-imaginar.html#:~:text=Jobs%20so%C3%B1%C3%B3%20all%C3%A1%20por%201974,primitiva%20Apple%20redonde%C3%B3%20el%20proyecto> (consulta: 28 de febrero de 2021).

⁹ Calvo, Javier (2018), “Machine Learning, una pieza clave en la transformación de los modelos de negocio”, *Revista Management Solutions*, Madrid. En línea: <https://www.managementsolutions.com/es/publicaciones-y-eventos/informes-sectoriales/white-papers/machine-learning-una-pieza-clave-en-la-transformacion-de-los-modelos-de-negocio> (consulta: 2 de marzo de 2021).

Por otra parte, existe una extensa bibliografía sobre el tema de la inteligencia artificial, sobre lo cual contentémonos por ahora con señalar que Andreas Kaplan y Michael Haenlein (2019) consideran que la **AI** es “la capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible”.¹⁰

Los autores del presente documento, con un equipo de investigadores, hemos desarrollado varias investigaciones que incluyen procesos automatizados,¹¹ utilizando algoritmos de *Machine Learning*, o aprendizaje de máquinas.



Entonces, para la esfera de nuestro interés profesional, la automatización tiene como objetivo la utilización de técnicas y algoritmos matemáticos que *faciliten* el quehacer legislativo al dar acceso veloz y un procesamiento expedito de grandes volúmenes de datos,¹² lo cual también aumenta la eficiencia en el análisis del propio investigador y, por ende, incrementaría la calidad de los insumos destinados a las comisiones legislativas.

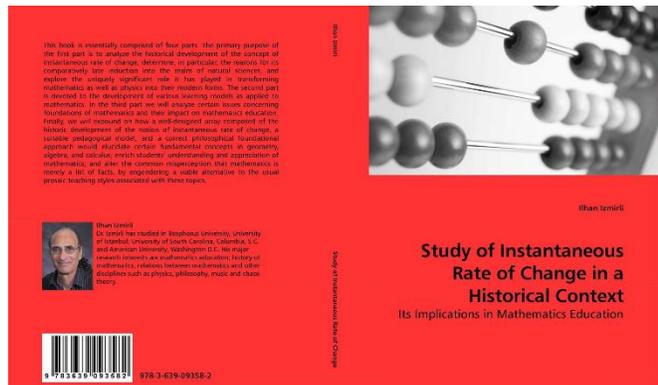
¹⁰ Andreas Kaplan; Michael Haenlein (2019), Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence, *Business Horizons*, 62(1), 15-25 (consulta: 1 de enero de 2021).

¹¹ Algunas de estas publicaciones se encuentran en la página del Centro de Estudios Sociales y de Opinión Pública (CESOP), otras pueden encontrarse en la Biblioteca Virtual de la Cámara de Diputados. En línea: <http://www.diputados.gob.mx/>

¹² García, Juan Antonio (2020), Introducción a la Automatización Industrial, Universidad de Málaga. En línea <https://dademuch.com/2020/04/22/introduccion-a-la-automatizacion-industrial-guia-pdf/> (consulta: 2 de marzo de 2021).

En la actualidad, con técnicas de automatización, es posible diseñar una gran cantidad de modelos de visualización de datos interactivos (*Dashboard*), de acelerar de manera increíble el procesamiento de grandes volúmenes de información (*Big Data*) y, sobre todo, tener conclusiones expeditas (*text automation*) sobre casi cualquier fenómeno social.

La automatización (*automation*) se refiere en sí misma a **un fenómeno tecnológico definido por la instantaneidad** (véase la imagen al lado),¹³ es decir, por la velocidad de los resultados, sobre todo, por el vasto manejo de fuentes a un mismo tiempo y del tratamiento de los fenómenos de toda índole.



Hemos visto por ejemplo la necesidad de tal procesamiento de datos durante la pandemia del Covid-19, una información coyuntural que, como *Dashboard* aún no existe —y que, por cierto, estamos desarrollando para su pronta aparición—, sin olvidar que se han producido muchos documentos al respecto por parte de investigadores de los Centros de Estudio de la Cámara de Diputados (para mayor detalle, los documentos del CESOP sobre el tema Covid-19 pueden consultarse en el siguiente link: <http://www5.diputados.gob.mx/index.php/camara/Centros-de-Estudio/CESOP/Tema-Covid-19>).¹⁴

¹³ En la segunda imagen se presenta la portada del libro clásico de Ilhan Izmirlı (2010), *Study of Instantaneous Rate of Change in a Historical Context. It's implications in Mathematic's Education*. VDN Verlag Dr. Muller, 2010, 612 pp. En este libro se analiza el desarrollo histórico de diversos modelos de aprendizaje aplicados a las matemáticas, así como la inclusión del concepto de “instantaneidad” y su impacto en la educación. El autor sugiere que con un modelo pedagógico adecuado se podrían dilucidar ciertos conceptos fundamentales de la geometría, el álgebra y el cálculo, para enriquecer la comprensión y la apreciación de las matemáticas por parte de los estudiantes, y alterar la percepción errónea común de que las matemáticas son simplemente una lista de hechos, engendrando una alternativa viable a los estilos de enseñanza prosaicos habituales.

¹⁴ Todos los centros de estudio han abordado el tema del Covid-19 dada su importancia. El CESOP no es la excepción. Por ahora, nos conformamos en enumerar algunos documentos que sobre el Covid-19 hemos desarrollado a lo largo de la pandemia: “A un año de la pandemia de Covid-19: cinco reflexiones acerca de los cambios necesarios” (En contexto, febrero 2021); “¿Cómo cambia el tejido socio-delictivo durante el Covid-19?: Diez entidades con descenso en violencia de género” (Documento de trabajo, octubre 2020); “Numeralia del Covid-19” (Carpeta informativa, septiembre 2020); “¿Cómo cambia la movilidad en los tiempos del Covid-19 en cuatro entidades más pobladas del país?” (Documento de trabajo, agosto 2020);

2. Objetivos de la investigación parlamentaria

Cualquier proceso de innovación, cuyo fin último es obtener resultados certeros y rápidos, pasa por revisar un tercer eje de la acción legislativa: los procesos de investigación parlamentaria.

Estamos en una **carrera contra el tiempo**, en la que, por definición, “el tiempo nos gana”. El tiempo, ese recurso infinito, es hoy un recurso valioso para el Poder Legislativo, por lo que se requeriría revisar antiguos procesos de recopilación de información, basados en tratamientos manuales. De allí que intentaremos reflexionar sobre dicha urgencia porque “el tiempo apremia”, porque la inmensidad de datos nos abruma y porque tenemos que dar un nuevo rostro al **cómo hacer (know how) de la investigación parlamentaria**.

Eso suena bien, pero el problema tampoco es simple, por lo que ha sido un tema tratado por la investigación académica, porque es de la mayor relevancia y ha sido abordado por diferentes autores. Para Díaz Santana (2000), por ejemplo, los legisladores del país “se enfrentan a un cúmulo de temas en los que, además de información, precisan de un tipo de investigación específica”, por lo que “la investigación parlamentaria” debe ser “oportuna, veraz y objetiva”, que “coadyuve a una adecuada toma de decisiones en beneficio de la ciudadanía”. De allí que, “la relación entre las fuentes de información y asesoría parlamentaria va interiormente ligada a su desarrollo y efectividad”.¹⁵

La investigación parlamentaria en México —si nos apuramos con una definición rápida, aunque quizá injusta—, es una práctica que se construyó a la par de la evolución del sistema político mexicano, bajo las exigencias de cada nueva legislatura. Sea porque el partido mayoritario no cambió durante todo el siglo XX, sea porque cambió repetidas

“La fase no tratada del Covid-19: los municipios” (Documento de trabajo, julio de 2020); “¿Cómo resignificar el durante y el después de la pandemia del Covid-19? (Carpeta informativa, junio 2020); entre muchos otros.

¹⁵ Héctor Díaz Santana (2000), “Los Servicios de Información y Asesoría en el Poder Legislativo Mexicano”, en Germán Pérez y Antonio Martínez (comps.), *La Cámara de Diputados en México*, FLACSO-Cámara de Diputados-Porrúa, México 2000, p. 184. En línea <http://www3.diputados.gob.mx/camara/content/download/21259/105709/file/PB3002%20Sistemas%20de%20informacion%20e%20investigacion.pdf> (consulta: 2 de marzo de 2021).

veces durante las dos décadas pasadas, o sea porque las alianzas *in situ* cambiaron definitivamente la noción de mayorías legislativas.

Esa última idea, la de nuevas mayorías, es probablemente el fenómeno que requiere un ambiente de innovación, debido a que, en algunos aspectos, el funcionamiento actual de la Cámara de Diputados tiene desfases en la disponibilidad y actualización de la información, o en el carácter expedito de sus análisis.¹⁶

Al respecto, sin ánimo de controversia, baste un ejemplo: algunas iniciativas de ley presentan su justificación basada en fuentes periodísticas, lo que podría ir en detrimento de su calidad y profundidad. Al respecto, podrían mencionarse muchos ejemplos fundados en la experiencia personal y en conversaciones con otros colegas con mucha más antigüedad en la Cámara. Queremos aquí simplemente sugerir que el tema puede ser materia de revisión de las prácticas legislativas y puede, sobre todo, abordarse desde muy diversos ángulos dada las problemáticas específicas que enfrentan los investigadores parlamentarios para satisfacer las necesidades de información institucionales.

Este retraso en la innovación nos la da un último ejemplo: el primer intento de crear un servicio de carrera en la Cámara —como un eje central en la innovación en la investigación parlamentaria— fue en el año 2000 y sólo ha sido retomado después de 20 años.

Nótese que hoy existen sólidos pasos para la implementación de un servicio de carrera parlamentario, como inicio de procesos de innovación en la Cámara de Diputados. Pero conviene decir que pese a su carácter relativamente reciente, se advierten claras medidas para la reglamentación específica en la selección de miembros, actuales y futuros, sobre todo en la capacitación de los mismos (Imagen 2).

¹⁶ Esta no es una crítica injusta si podemos reconocer el laborioso trabajo que realizan cotidianamente un ejército de trabajadores e investigadores para hacer funcionar al Poder Legislativo en todas sus improntas. Recuperar estas experiencias sería también un elemento esencial en una estrategia de innovación en el largo plazo. Por otra parte, sobre la discusión de “tecnologizar” la Cámara en sus procesos también puede consultarse otro de nuestros trabajos: “Big Data ¿Cómo innovar para alcanzar la “sacudida” que sugiere Carlos Slim?” (Documento de Trabajo, enero 2020, en línea: <http://www5.diputados.gob.mx/index.php/esl/content/download/169147/843943/file/CESOP-IL-72-14-BigData-310120.pdf>. Consultado el 1 de febrero de 2021.

Imagen 2. Página oficial del Servicio de Carrera de la Cámara de Diputados



Fuente: Sitio web: <http://serviciodecarrera.diputados.gob.mx/>

No sobra decir que reglamentar los procesos de contratación, de selección y condiciones de su permanencia son pasos notables en vías de certificar la innovación parlamentaria con criterios de confiabilidad y eficiencia.

Sobre la importancia del servicio de carrera, por ejemplo, Cruz Villalobos (2009) sugiere que “con la instauración de un servicio profesional de carrera” se puede “profesionalizar a los funcionarios a través de un proceso de capacitación que” tendría “como objetivo la actualización y especialización del personal que se encuentra al servicio

de una dependencia y órgano público, con el propósito de darle continuidad al trabajo de carácter técnico”.¹⁷

Concluamos que en el tema de las limitantes institucionales se requiere de una visión de corto, mediano y de largo plazo para la innovación. Dejemos sentado aquí también que eventuales limitantes en **la forma del cómo se hace investigación** podrían adjudicarse a los retrasos en la formación del personal o en un relativo abandono de la innovación en los procesos de investigación parlamentaria ocurrida durante años, una conclusión personal a la que llegamos después de algunos años de observación directa.

Todo ello parece suficiente para justificar nuestro interés aquí de **desarrollar una propuesta preliminar para innovar un aspecto de la investigación en la Cámara** y lo que ello implica en un servicio parlamentario al que hoy realmente se busca profesionalizar, afortunadamente.

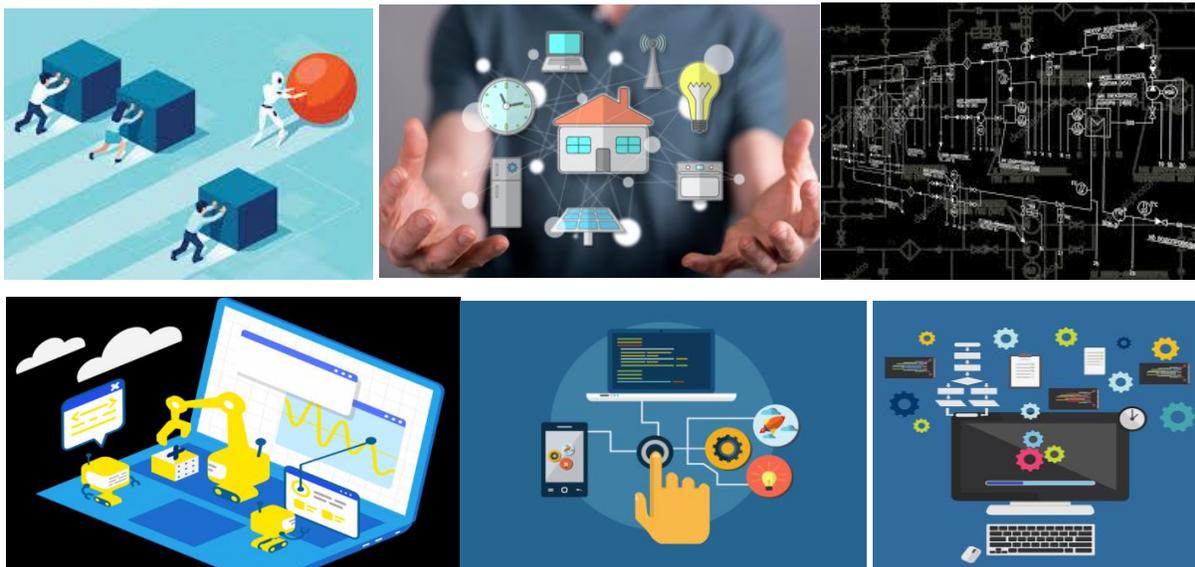
3. Ejemplo de procesos factibles de automatización

¿Cuáles serían los procesos de análisis factibles de automatización? Los temas, las áreas o los procesos que requerirían automatización pueden ser muchos dentro de la Cámara de Diputados, pero no vamos a detallarlos aquí. Hemos seleccionado el tema de la incidencia delictiva, a manera de ejemplo, que hemos trabajado anteriormente, a partir de modelos de análisis basado en algoritmos creados con el lenguaje de programación

¹⁷ Véase Cruz Villalobos, Osvaldo (2009), *El Servicio Profesional de Carrera y los cambios en la Gestión Pública*, IAPEM, México, 2009, p. 214. En línea: <http://iapem.dyndns.org:8080/libros/2009%20120%20ElServicioProfesional%20y%20los%20cambios%20978-968-6452-87-7.pdf> (consulta: 28 de diciembre de 2020).

Python¹⁸ y sus librerías más populares.¹⁹ ¿Por qué esta selección? Porque el tema de la seguridad pública es un eje de análisis de la mayor relevancia en el debate de la agenda pública nacional (Imagen 3).

Imagen 3. Infografías representando procesos algorítmicos de la innovación



Fuente: búsqueda libre en imágenes de internet.

¹⁸ El lenguaje de programación Python hoy popular entre los tecnólogos y grupos interdisciplinarios de investigación, útil para la industria, la comunicación, el gobierno y la enseñanza. Existe también mucha bibliografía al respecto. Por ejemplo, en el popular sitio web *PythonForBeginners* se explica así: “Python es un lenguaje de programación de alto nivel orientado para el desarrollo web y de aplicaciones. Uno de los beneficios más prometedores de Python es que tanto la biblioteca estándar como el **intérprete están disponibles de forma gratuita**. Python es relativamente simple, por lo que es fácil de aprender, ya que requiere una sintaxis única que se centra en la legibilidad. Los desarrolladores pueden leer y traducir código Python mucho más fácil que otros lenguajes. A su vez, esto reduce el costo de mantenimiento y desarrollo del programa, ya que permite a los equipos trabajar en colaboración sin barreras significativas de idioma y experiencia.” *PythonForBeginners*, en línea: <https://www.pythonforbeginners.com/learn-python/what-is-python> (consulta: 28 de febrero de 2021).

¹⁹ Para hacer uso del lenguaje Python se acude a una extensa gama de librerías, que son conjuntos de algoritmos para operaciones especializadas. Algunas de las librerías más comunes en Python son: Numpy, considerada una de las bibliotecas de aprendizaje automático más populares de Python; Pandas, es el soporte para operaciones como reindexación, iteración, clasificación, agregaciones, concatenaciones y visualizaciones se encuentran entre las características más destacadas; Plotly, se utiliza para trazar gráficos fácilmente, que funcionan muy bien en aplicaciones web interactivas y hacer diferentes tipos de gráficas básicas de líneas, pastel, dispersión, mapas de calor, diagramas polares, etcétera; Scrapy: es un marco colaborativo para extraer los datos que se requieren de los sitios web, es una herramienta bastante sencilla y rápida. En línea <https://www.iaartificial.net/librerias-de-python-para-machine-learning/> (consulta: 1 de marzo de 2021).

Ahora bien, el tratamiento que haremos aquí enseguida sobre un reporte automatizado de la seguridad pública en el país es de orden técnico, es decir, entramos en la zona gris de los tecnicismos que requiere de algunos conocimientos básicos en la informática.

Sin embargo, que no se asuste el lector, porque los responsables de esta publicación entienden que un lector neófito tendría dificultades para comprender lo que enseguida se expone sin conocimientos previos de informática. Por ello, para facilitar la lectura de esta sección específica, hemos omitido, tanto como ha sido posible, las fórmulas y algoritmos necesarios para desarrollar cada proceso. En caso de no lograrlo siempre es posible dirigirse a los autores para aclarar cualquier duda (ver email de contacto proporcionado al principio).

Finalmente, conviene insistir que el objetivo de este ejercicio es hacer evidente la trascendencia de resolver problemas de interés público, como parte de nuestra misión como investigadores del servicio de carrera en la Cámara de Diputados.

4. La incidencia delictiva con procesos de automatización

Enseguida se presenta un ejercicio de reporte automatizado sobre el tema de la incidencia delictiva, lo que sería acorde con una necesidad de información proveniente de la Comisión de Seguridad Pública o de la Junta de Coordinación Política (Jucopo). El objeto de dicho reporte automatizado sería que dichas instancias legislativas cuenten con información basada en evidencia, actualizada y expedita.

Con este ejemplo sugerimos que casi todos los temas abordados por las comisiones de la Cámara de Diputados podrían someterse a un esquema similar de optimización, con las variaciones específicas que cada caso exija: de fuentes por consultar, de tipos o fórmulas de procesamiento, de periodos para su actualización, etcétera). En general, estamos seguros que el análisis parlamentario ganaría con esta estrategia de innovación.

El diseño de este modelo algorítmico **puede completarse en una semana o dos**, pero su ejecución y actualización lleva apenas unos minutos.

Ahora bien, si tenemos como objetivo diseñar un *Dashboard* o tabla de datos (o reporte ejecutivo sobre el tema, monografía, etc.) el objeto de todo este procesamiento es que dicha información debe ser **consultable y actualizable en cualquier momento**.

Para ello, podemos tener en mente los siguientes procedimientos que serán abordados a continuación: a) crear una plataforma o medio para la visualización de datos; b) desarrollar los algoritmos²⁰ y procesos de automatización.

a) Visualización de datos

El objetivo aquí sería completar un *Dashboard* (o tabla de datos) para la presentación de información sobre incidencia delictiva, en este caso. Para entender mejor esto que se describe enseguida, puede consultarse al final el *Anexo*, donde cada palabra o frase marcada enseguida con formato (**negritas**) se identifica en la tabla de datos, para indicar al lector algunas de las operaciones que se llevarían a cabo en cada sección de nuestro reporte automatizado.

En dicha tabla de datos se visualizarían datos numéricos, datos de texto e imágenes de diferente tipo (mapas y gráficas), ordenados de tal manera que resalten los “picos” (**máximos y/o mínimos**),²¹ o aquellos datos únicos (promedios o tasas de incidencia por cada cien mil habitantes) para destacar ángulos relevantes de la información tratada.

Así, por ejemplo, se presentarían **sumas** de delitos por periodo (mensual, semestral, anual) y según las escalas a tratar (nacional, estatal o municipal). De igual

²⁰ “Un algoritmo informático es una secuencia de instrucciones finitas que llevan a cabo una serie de procesos para dar respuesta a determinados problemas. Es decir, un algoritmo informático resuelve cualquier problema a través de unas instrucciones y reglas concisas, mostrando el resultado obtenido” Robledano Ángel (2019), Lenguajes de programación con pseudocódigo: ¿Qué es un algoritmo informático?, p. 98. En línea <https://openwebinars.net/blog/que-es-un-algoritmo-informatico/#:~:text=Un%20algoritmo%20inform%C3%A1tico%20es%20una,concisas%2C%20mostrando%20el%20resultado%20obtenido> (consulta: marzo de 2021).

²¹ Este tipo de letra se utiliza para vincular visualmente al lector con las partes que se incluyen en la tabla de datos. Véase al final el **Anexo**.

manera, los delitos podrían ser agrupados por **categorías**, como violencia de género, por sector económico o grupo poblacional afectados por la actividad delictiva. Finalmente, las presentaciones **gráficas o mapas** destacarían *visualmente* las zonas criminógenas y su posible vínculo cultural-geográfico.²²

Hay que insistir en que el tratamiento de los datos se organizaría en tipologías de análisis (tipos de delitos) o por escalas (nacional, estatal o municipal) y en datos absolutos o relativos (sumas totales o subtotales, tasas, promedios) o cualquier otro modo de agrupación pertinente. En dicho sentido, las gráficas y los mapas apoyan estos análisis o su correspondencia geográfica que permitiría hacer evidentes las correlaciones geográficas o afinidades temporales en la incidencia delictiva.

b) Algoritmos y automatización

¿Qué se automatiza de todo esto? Primero, hay que insistir en que el proceso de crear todos los algoritmos llevaría una semana, por única vez. Cada proceso automatizado puede replicarse y modificarse para otros análisis específicos, sea mejorando la profundidad de análisis en periodo de tiempo o en las correlaciones entre variables.²³ Podemos identificar cuatro procesos generales de automatización, que se detallan más adelante:

- i) Extracción de datos (*Request*).
- ii) Recopilación y tratamiento de datos (*Exploratory Data Analysis*).
- iii) Agrupamientos y tendencias (*Model of Analysis*).
- iv) Presentación de la información (*Dashboard* o tabla de datos).

A manera de una primera demostración, en la Imagen 4 se muestra una forma simple de hacer la selección de un archivo (ahora sobre Covid-19) con una fecha en específico en

²² No hemos incluido mapas en serie, pero hubiera ayudado al lector a comprender la utilidad de dichas imágenes en la identificación de zonas sensibles de la incidencia delictiva. No obstante, hemos desarrollado estos trabajos en algunas publicaciones, puede consultarse: De Alba, F. *et al.* (2020), *Ranking municipal de los delitos en México (2015-2020)*. *Ejercicios de visualización*, Cámara de Diputados, 264 pp.

²³ Esto se ejecuta literalmente en segundos o minutos, tal como se indica al final de cada inciso, insistimos.

una página en específico y descargarlo cuando se requiera la actualización de la información.

Imagen 4. Demostración de una forma de automatización

```
1 yesterday = datetime.now() - timedelta(1)
2 ident = datetime.strftime(yesterday, '%Y%m%d')
3
4 today = date.today()
5 d2 = today.strftime("Fecha de actualización : %d de %B de %Y")
6
7 contagios=
8     pd.read_csv("https://datos.covid-19.conacyt.mx/Downloads/Files/Casos_Diarios_Municipio_Confirmados_%.csv" %(ident))
9
10 decesos=
11     pd.read_csv("https://datos.covid-19.conacyt.mx/Downloads/Files/Casos_Diarios_Municipio_Defunciones_%.csv" %(ident))
12
```

Fuente: Elaboración propia.

Como se observa en la anterior imagen, la estructura en la programación de Python para la consulta *online* de algunos tipos de archivos es muy intuitiva y, por si fuera poco, la sintaxis está basada en un idioma simple y entendible. Esos simples comandos mostrados ejecutan una función definida, la descarga de un archivo.

i) Extracción de datos

Como hemos visto, en el proceso de programación se crean accesos directos a las fuentes de información, luego dicha información es validada para iniciar su tratamiento, como sería el caso de las bases de datos del Secretariado Ejecutivo Nacional del Sistema Nacional de Seguridad Pública (**SENSNSP**).

Con la Imagen 5 se muestra la rapidez de operación de Python sobre archivos de gran volumen: con el comando se abre el archivo que contendría los delitos del año 2015 al año 2020, un archivo con un tamaño de 1.2+ gigabytes. Su apertura se lleva a cabo en un tiempo exacto de 33.6 segundos. En detalle, **el archivo contiene** 33 columnas y **4,965,059 filas**. *Tiempo de ejecución de este comando: 33.6 segundos* (Imagen 5).

Imagen 5. Demostración de la velocidad de Python

```
1 archivo = pd.read_csv("Delitos SENSNSP.csv", encoding = "Latin-1")
2 archivo.info(verbose = False)
```

executed in 33.6s, finished 21:38:56 2021-01-01

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4965059 entries, 0 to 4965058
Columns: 33 entries, Unnamed: 0 to cve_ent_mun3.1
dtypes: float64(18), int64(1), object(14)
memory usage: 1.2+ GB
```

Fuente: Elaboración propia.

Aunado al acceso directo a las bases de datos también tendríamos que descargar otros elementos de información, como archivos georreferenciados (*shapefiles*), útiles para un *mélange* (cruzamiento) de la incidencia delictiva con variables de localización, importantes en el análisis territorial del delito; o la relación de la incidencia delictiva en zonas de mayor presencia del fenómeno migratorio; o la identificación regional de robos específicos (abigeato, “huachicoleo”); finalmente, para detectar zonas de extrema pobreza con alta incidencia delictiva, entre muchas otras variables.

Antes de iniciar la automatización es preciso reconocer que en bases de datos existe una gran variedad de tipos de archivos, provenientes de diferentes instituciones gubernamentales. A veces es necesario “convertir” esos archivos a un formato más común o a un formato que al guardarlo signifique menor volumen (sobre todo en los casos de archivos de gran tamaño (más de 1 gigabyte). Este “pretratamiento” hace que la automatización represente menos “diferencias” al momento de procesar grandes volúmenes de datos.

ii) Recopilación y almacenamiento de datos

Los datos absolutos sobre incidencia delictiva se priorizarán en la recopilación, pues son éstos la base de la estadística que aquí se desarrollaría. Después, se clasificarían por año o mes en carpetas individuales o específicas al intervalo de tiempo que pertenezcan. Además, se podría agregar una subclasificación por subtipos de delitos (sic). También,

el sistema analizará las diferencias entre bases de datos, tales como el registro de nuevos municipios o nuevas columnas con el fin de homogenizar el proceso de estudio. Es pertinente mencionar que esta información se descarga y almacena de manera automática.

Si bien la información consultada proviene de fuentes oficiales (gubernamentales), ésta no siempre es accesible en línea o no cuenta con facilidades para la descarga mediante procesos automatizados, por ello, partimos de establecer una “Biblioteca online” propia de cada proceso. Con esta nueva técnica de almacenamiento se asegura que el material de trabajo esté presente para su revisión individual y además se permitiría al investigador experimentado diseñar nuevos procesos a partir de la información “cruda”. El almacenamiento en línea o “en la nube” (*cloud*) es una tendencia muy actual en este tipo de procesamientos.

iii) Agrupamientos y tendencias

Después de obtener y **almacenar** las bases de datos y en función de las características de los archivos el modelo algorítmico los agrupará, compilará o separará según criterios específicos,²⁴ además de ejecutar operaciones concretas en sumas, restas, divisiones, obtención de porcentajes, tasas, entre otros.

Con este primer resultado se obtienen tablas con los valores más importantes (**rankings**), así como aquellas **variaciones** en un intervalo contrastado con otro. Las tablas darán pie a gráficas que, para mejor visualización, se presentarán en colores atractivos al carácter del tema en cuestión. En la Imagen 6 se presenta un ejemplo de los valores **máximos** de cada columna, proceso que se obtiene en tan sólo 52 milisegundos.

Imagen 6. Identificación de valores máximos por columna de la base de datos

²⁴ Los criterios específicos, como ya se mencionó, son las escalas geográficas, por variables y temporalidades.

```
In [11]: 1 print (registro_delitos.max())
executed in 52ms, finished 15:56:48 2020-12-29
Unnamed: 0                98839
Cve. Municipio            32058
MPIO_ENT                 Ánimas Trujano, Oaxaca
Municipio                 Ánimas Trujano
Entidad                   Zacatecas
Tipo de delito            Violencia familiar
Enero20                   2292
Febrero20                 2044
Marzo20                   1928
Abril20                   1625
Mayo20                    1677
Junio20                   1736
Julio20                   1847
Agosto20                 1812
Septiembre20             1787
dtype: object
```

Fuente: Elaboración propia.

Por ejemplo, en el análisis de delitos, las operaciones estarían desarrolladas primero a escala nacional, estatal y municipal. También se presentarían sumas totales por entidad (32), tasas de incidencia delictiva (número de delitos por cada 100 mil habitantes) y variaciones, contrastando cifras mensuales de un año con otro (**comparaciones**).

En ejercicios posteriores se podrían incluir subcategorías delictivas (muertes violentas, abigeato, trata de blancas, violencia familiar, secuestro, etc.) que, segregadas por estado, regiones o municipios permiten extraer tendencias, o cifras de acumulados. Además, estos resultados se agrupan según la escala temporal y se analizan las diferencias resultantes entre intervalos anuales, por ejemplo.

El modelo de análisis anterior se repite con periodos más específicos como meses, quincenas y/o semanas, si la información lo permite (aunque no es el caso de la información del SENSNSP). Los mapas generados representarían los datos más extremos en cada uno de los periodos contrastados.

La gran versatilidad de la información manejada y de la facilidad de los procesos automáticos para cada variable o subvariables, como ya se explicó, permite, en el caso de mapas, la interrelación entre elementos físicos y culturales en función del enfoque, tema o pertinencia del análisis.

iv) Presentación de la información (tabla de datos)

Como última etapa, todas las operaciones anteriores se concentran al final en una tabla de datos (véase Anexo), diseñada expresamente, que puedan actualizarse automáticamente en todos los campos. Además, la tabla de datos incluye **texto automatizado** con las conclusiones más relevantes, estructurada y completada con elementos gráficos como mapas, líneas de tiempo y diversas gráficas, así como también podría incluir cuadros resumen, mapas mentales, diagramas de flujo, o cualquier tipo de estructura de análisis, que resuma grandes cantidades de información.

Al presentar esta misma información en un *Dashboard*, los elementos gráficos pasan a ser elementos interactivos que, como es de esperar, exponen mayor cantidad de información para el consultor que decida indagar más a fondo sobre una o más variables en específico. Es aquí donde se encuentra una de las ventajas de estos procesamientos de grandes volúmenes de información: descubrir dependencias, correlaciones, tendencias.

En general, con este procesamiento acucioso, exhaustivo e infinitamente veloz, se alcanzaría una **mayor eficiencia en las capacidades del investigador parlamentario** o del analista. Con este tipo de procesamientos se puede “alcanzar a ver” lo que sería oculto en un análisis “manual”. Con estas técnicas se logran visualizar evoluciones o tendencias, correlaciones geográficas o temporales, tipologías, con lo que las hipótesis típicas sobre fenómenos sociales alcanzarían mayor eficacia en “prefigurar” casi cualquier problemática social. En fin, con este tratamiento, la magnitud de la información y la velocidad de operación, las variables se dilucidan o se abren nuevos puntos de análisis sobre los efectos o las reacciones sociales que tendrían de las políticas públicas.

A manera de conclusión

Aquí podemos recuperar cuatro elementos esenciales de esta propuesta. En primer lugar, concordemos que la Cámara de Diputados requiere de procesos de innovación dada la particularidad de la investigación parlamentaria, por su “apego” a la evolución del sistema político y la evidencia de retardos en su actualización.

En segundo lugar, la actual legislatura está caracterizada por el interés de hacer cambios de raíz, lo que es una coyuntura idónea para proponer cambios fundamentales al quehacer parlamentario y, específicamente, a las tareas de investigación.

En tercer lugar, los procesos de automatización están alcanzando un increíble auge por la gran eficiencia que muestran en la manipulación de grandes bases de datos, lo que facilita la recuperación, el tratamiento y el análisis de la información en procesos que llevan apenas algunos segundos y en resultados de alta precisión. Para ello se requieren funcionarios capacitados profesionalmente y dispuestos a innovar. Es un gran desafío.

En cuarto lugar, esta propuesta contiene un ejemplo mínimo de cómo podríamos desarrollar múltiples procesos que proporcionen a los legisladores análisis expeditos y basados en evidencia constantemente actualizada. Esto, en sí mismo, sería un avance de proporciones mayúsculas.

Anexo. Visualización de todas las automatizaciones

Ciudad de México

Iztapalapa



Mapas

Incidencia delictiva en el confinamiento

Hasta el 14 de noviembre solo estaban disponibles los datos hasta septiembre de 2020

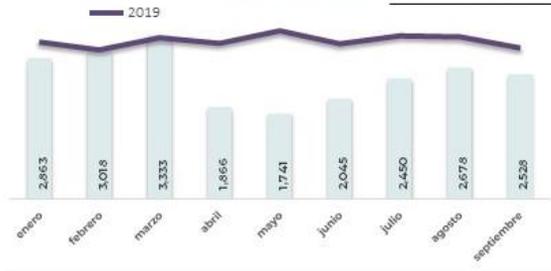
Suma
máximos y
mínimos

Delitos (ene-sep)	17,316
Mayor incidencia	3,333
Menor incidencia	1,741

variaciones

Variaciones 2020-2019			
enero	-10.1	julio	-26.0
febrero	-0.3	agosto	-18.6
marzo	2.0	septiembre	-17.5
abril	-40.9		
mayo	-48.9		
junio	-35.1		

Incidencia delictiva de 2020 en contraste con 2019

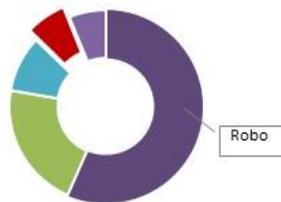


Comparaciones

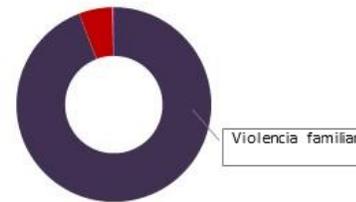
Rankings

Top de incidencia septiembre	enero	febrero	marzo	abril	mayo	junio	julio	agosto	Totales	
Robo	1,259	1,239	1,298	731	754	868	955	1,015	998	9,117
Violencia familiar	351	441	498	296	242	333	417	434	398	3,410
Amenazas	197	181	256	119	94	117	184	205	172	1,525
Narcotráfico	164	117	99	52	145	109	126	163	103	1,078
Lesiones	118	117	158	97	81	81	107	122	111	992

Top de incidencia



Top de violencia de género



Subcategorías

Top de violencia de género										
Violencia familiar	351	441	498	296	242	333	417	434	398	3,410
Violación	24	32	27	18	16	15	29	19	20	200
Feminicidios	0	0	0	2	1	2	4	1	1	11
Otras violencias de género	0	0	0	0	0	0	0	0	0	0

Fuente: SENASISP, 2020

Texto automatizado

Incidencia delictiva

De enero a septiembre del 2020, el municipio de Iztapalapa registró 17,316 delitos y la incidencia más baja ocurrió en el mes de mayo. En ese mismo lapso, los tres delitos con mayor incidencia fueron: robo (41.4%), violencia familiar (14.5%) y amenazas (6.5%).

CENTRO DE ESTUDIOS SOCIALES Y DE OPINIÓN PÚBLICA

www.diputados.gob.mx/cesop

 cesop01

 @cesopmx