Octubre 2021

1er Datatón: Tu Ciudad, Tu Dinero

Eligiendo el eje temático

Equidad de género

Con el proyecto titulado

Mujeres:
Merecemos ser más que una estadística

Presenta el equipo

Los mariachis

Conformado por:
- Cisneros Esparza Brenda
- Flores Hernández Efraín Ismael
- Ramírez Zarco Héctor Iván

Índice:

Introducción 1

¿Qué vamos a lograr? 1

Justificación 1

Desarrollo del proyecto 2

¿Cómo obtenemos valor de la información? 2

¿Qué descubrimos? 3

Propuestas y conclusiones 7

Gráficas e información auxiliar 8



Introducción

Desafortunadamente, no sorprende que los problemas de desigualdad y violencia de género sean una constante en nuestra sociedad y como consecuencia de la pandemia provocada por el COVID-19 más mujeres se hayan visto afectadas. Todas merecemos ser más que una estadística, merecemos ser más que un número que se suma a la cantidad alarmante de casos diarios, somos personas con un nombre y rostro, somos mujeres con historias de vida, con logros, sueños por cumplir, esperanza y valentía por hacer las cosas diferente.

Por ello, enfocaremos este proyecto en dos problemas sociales con alta urgencia de resolver:

- Administración de recursos dada la **demanda futura** de llamadas a la Línea Mujeres (LM)
- **Personalización de campañas** para prevenir los embarazos no deseados, ocupando información de las personas gestantes que acuden a la Interrupción Legal del Embarazo (ILE)

En esta época de resiliencia y cambio, queremos dejar huella con nuestras propuestas para que las futuras generaciones no vivan con lo que hasta ahora hemos adoptado como “normal”. Este es el momento indicado para cuestionar y reformular las reglas preestablecidas, porque somos nosotros quienes le ponemos el color al mundo, somos quienes construimos el mañana.



¿Qué vamos a lograr?

Los beneficios que la Inteligencia Artificial nos brinda, deben ser utilizados a nuestro favor y en el presente trabajo usaremos diferentes herramientas para cumplir con los siguientes objetivos:

Por un lado, tomando en cuenta los datos de la LM:

- Analizar el comportamiento histórico de cada uno de los servicios que proporciona la línea: jurídico, médico y psicológico.
- Generar un modelo de series de tiempo para cada servicio y pronosticar la cantidad de llamadas semanales que se recibirán el siguiente año.

Por otro lado, respecto a la ILE:

- Agrupar estadísticamente a las personas gestantes que deciden interrumpir su embarazo.
- Empatizar con la vulnerabilidad de cada grupo para priorizar a través de un semáforo.
- Abrir la posibilidad de generar campañas personalizadas e innovadoras enfocadas en las características que hacen diferente a cada grupo y sobre todo, considerando también a la pareja.



Justificación

Si bien, de primera instancia los dos temas parecen tener un enfoque diferente, ambos se centran en la desigualdad que una persona padece simplemente por haber nacido mujer. Es inaceptable concebir los altos grados de violencia que suceden diariamente. Existen oficinas especializadas en la atención a las mujeres como SEMUJERES con múltiples campañas para frenar a esta pandemia tan peligrosa que la vivida en los últimos años. Dicha institución lanzó en el 2019 su campaña #YoDecidoMiFuturo con el objetivo de evitar el embarazo antes de los 19 años. Adicionalmente, el gobierno federal retomó en 2020, la Estrategia Nacional para la Prevención del Embarazo en Adolescentes (ENAPEA) buscando erradicar los embarazos de niñas de 14 años o menos. Incluso en este año, durante septiembre, se presentaron las campañas: ¡Yo decido! y ¡Yo exijo respeto!, que se enfocan principalmente a la población en entornos rurales e indígenas.

Sin embargo, no podemos quitar el dedo del renglón, el problema de la violencia de género debe ser atacado desde nuevas perspectivas, apoyados de los avances estadísticos de última tecnología para desarrollar soluciones que presenten empatía y ataquen de raíz la problemática de origen.



Desarrollo del proyecto

¿Cómo obtenemos valor de la información?

La información es importada desde archivos en formato .csv* y se emplea el lenguaje de programación Python para la limpieza, estructuración y modelado estadístico. El tratamiento consta de lo siguiente:

Llamadas realizadas a la Línea Mujeres, cabe señalar que para la correcta generación del pronóstico los siguientes pasos se realizaron para cada uno de los tres servicios proporcionados: médico, jurídico y psicológico.

1. **Agrupación semanal** del número de llamadas recibidas.
2. **Tratamiento de datos atípicos** mediante el método Hampel Filter, el cual considera la mediana de las observaciones tomando una ventana de tiempo. Si una muestra difiere de la mediana en más de *k* desviaciones estándar, se considera un dato atípico y se reemplaza por la mediana. Para el presente trabajo consideraremos *k=3* y una ventana de tiempo de 10 periodos. Para visualizar el resultado de este tipo de imputación véase Gráfica 1.
3. **Entrenamiento del modelo** de series de tiempo para la etapa de evaluación, se utiliza el modelo de Prophet, el fue desarrollado por la comunidad de Facebook. Las últimas 52 semanas de las ya sucedidas no serán consideradas al momento de entrenar para poder evaluar el ajuste que tiene el modelo ante la historia, obteniendo el porcentaje de error (MAPE) y disminuirlo.
4. **Re-entrenamiento del modelo** con todas las observaciones disponibles, esto después de haber obtenido un porcentaje de error aceptable.
5. **Generación del pronóstico y bandas de confianza**, con una probabilidad del 80%, de las siguientes 52 semanas.

Interrupción Legal del Embarazo

1. **Limpieza de datos**. Se cuenta con 47 columnas de diferentes tipos, sin embargo al encontrar valores nulos en al menos un valor para todas ellas, se decide aplicar una transformación categórica y así, con todo el respeto que cada registro merece, no omitir ni imputar información:
   a. Variables numéricas serán representadas en rangos. Ej: *edad en años=23 → "22 a 25 años".*
   b. Variables categóricas serán normalizadas, es decir, agrupamos sus opciones.
   c. Los registros que tengan valores nulos, es decir, que cierta pregunta no fue contestada, se le asignará la etiqueta *"DESCONOCIDO"* así todas las variables ahora son categóricas y no se omite ni imputa ningún valor.
2. **Agrupación de personas gestantes con decisión de ILE** en 10 clústers con algoritmo KModes al tratarse de variables categóricas. Después de interpretar las diferencias entre cada grupo, se propone un semáforo de vulnerabilidad: rojo, naranja y amarillo. El color verde no se utilizará porque ninguna persona gestante tendría que pasar por una ILE: la maternidad debería ser deseada.
3. **Frecuencia por localidad**, se contesta la pregunta: ¿Cómo se distribuye cada uno de los 10 grupos en esta alcaldía o municipio? Con el objetivo de agrupar alcaldías con frecuencias similares.
4. **Personalización de campañas**. Con la información limpia, organizada y segmentada en grupos, tanto de ILE como de alcaldías, tenemos la oportunidad de dirigir diferentes campañas a donde más lo necesita y a quien más lo necesita.

* Nota: Se intenta importar la información vía API pero la cantidad de registros para cada tabla, supera el límite de 32 mil registros permitidos

¿Qué descubrimos?



Sin duda los hallazgos generaron un impacto, no siempre son agradables pero se tiene la responsabilidad de difundirlos y proponer estrategias para mitigar la desigualdad que puedan presentar, directa o indirectamente. Para consultar el análisis desde diferentes enfoques, acceda al tablero interactivo en el siguiente Data Studio.

Por parte de la LM, se notó que la cantidad de llamadas para atender temas relacionados con la psicología van en aumento con el paso de los años, además se pronostica que para el 2022 se continúe con una tendencia positiva. Esto puede ser interpretado desde dos perspectivas: existe mayor conciencia respecto a las emociones y que los servicios psicológicos sean más requeridos por el aumento en casos de depresión y ansiedad en los últimos años, debido al confinamiento y otros problemas sociales que nos rodean.

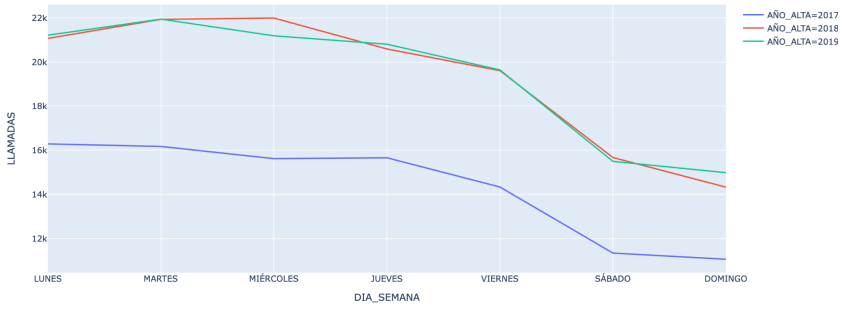
Gráfica 2. Distribución de llamadas por tipo de servicio



Fuente: Elaboración propia con datos de LM

Partiendo del hecho de que el servicio de asesorías está disponible los 365 días del año y las 24 horas, los días con mayor cantidad de llamadas son de lunes a jueves, mientras que durante el fin de semana, la tendencia disminuye de forma gradual, siendo el domingo el día con menor llamadas. El comportamiento por día de la semana nos da visibilidad para asignar de forma correcta los recursos de atención en la Línea Mujeres.

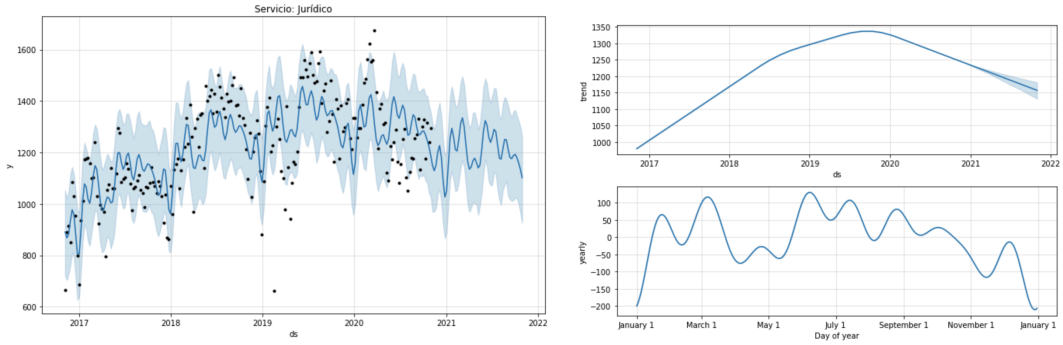
Gráfica 3. Distribución de llamadas por día de la semana y año



Fuente: Elaboración propia con datos de LM

Para el modelo **Jurídico**, se percibe una tendencia a la baja para el siguiente año, adicionalmente se notó que existe una estacionalidad muy marcada, pues en el mes de enero se presenta el punto más bajo en la recepción de llamadas, mientras que marzo y junio son los meses con mayor cantidad. Esto puede explicarse por los periodos ordinarios de vacaciones para el Poder Judicial.

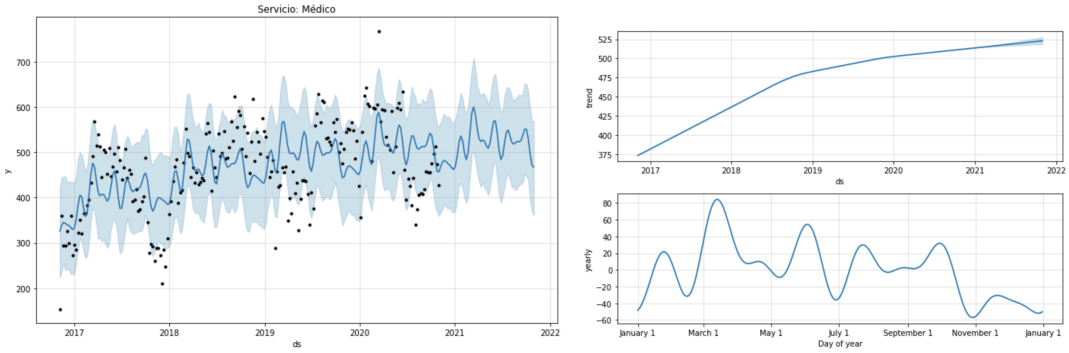
Gráfica 4. Pronóstico y tendencias del modelo Jurídico



Fuente: Elaboración propia con resultado del modelo Jurídico

El modelo **Médico** es diferente, se tiene una tendencia a la alza. Respecto a la tendencia por mes, para finales de octubre se presenta el periodo con menor llamadas recibidas y el punto más alto es a mediados de marzo. Al menos para el 2020, lo anterior puede ser una consecuencia del impacto de la pandemia ante el COVID-19.

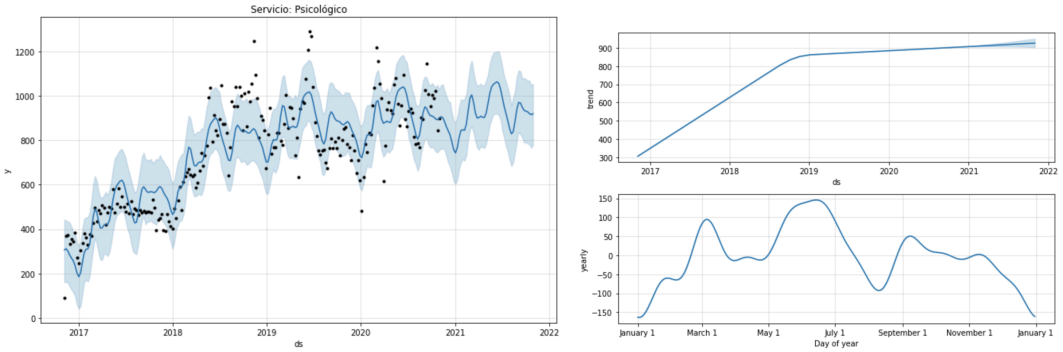
Gráfica 5. Pronóstico y tendencias del modelo Médico



Fuente: Elaboración propia con resultado del modelo Médico

Por último y no por ello menos importante, el modelo **Psicológico** también tiene una tendencia a la alta, siendo diciembre y enero el mes con menos llamadas, probablemente debido al periodo vacacional y mayo junto con junio los periodos en los que más mujeres se comunican para recibir apoyo con temas psicológicos.

Gráfica 6. Pronóstico y tendencias del modelo Psicológico



Fuente: Elaboración propia con resultado del modelo Psicológico

La métrica utilizada para evaluar el comportamiento de cada modelo de series de tiempo fue el error porcentual absoluto medio (MAPE por sus siglás en inglés: Mean Absolute Percentage Error). Cabe señalar que los datos atípicos pueden afectar dicho error, es por ello que para el modelo Médico se excluyeron dichos registros inusuales debido al inicio de la pandemia en 2020.

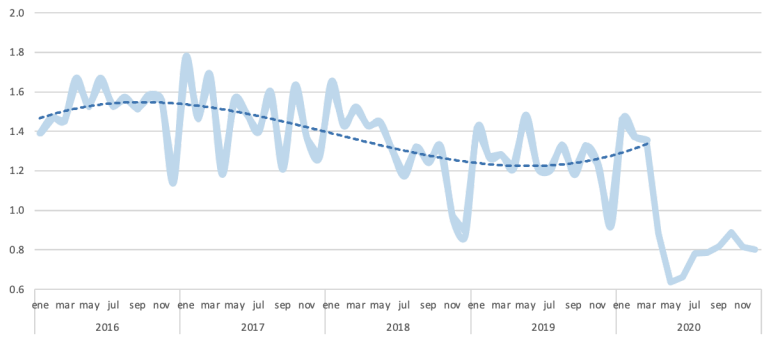
Tabla 1. MAPE por modelo, sin tomar en cuenta registros atípicos

	Jurídico	Médico	Psicológico
MAPE	17%	21%	16%

Fuente: Elaboración propia con resultados de modelación LM

Ahora, respecto a los **hallazgos de la ILE**, descubrimos que parece haber una tendencia a la baja desde 2016, sin embargo a partir de 2019 se tiene un ascenso en la tendencia. En definitiva a partir de abril 2020, la tendencia cambia radicalmente lo que nos hace preguntarnos: ¿Cuántos embarazos no deseados ocurrieron sin oportunidad de asistir a interrumpir su embarazo?

Gráfica 7. Cantidad de personas gestantes (en miles) pre-pandemia



Fuente: Elaboración propia con datos de ILE

Adicionalmente, el 64% de las personas gestantes que acuden a la ILE provienen de la CDMX y el 31% del Estado de México. El 5% restante está centralizado geográficamente en el país (para más detalle, véase Imagen 1), esto puede evidenciar la falta de difusión a los estados que están alejados de la capital, tal es el caso de Campeche: desde 2016 han asistido al procedimiento solamente 5 personas.

Es muy importante dejar claro que aplicar un algoritmo de clustering para encontrar grupos con características en común no significa solamente catalogar registros, si no que se respeta la individualidad de cada persona gestante y confiamos en las herramientas que la estadística nos ofrece para generar mejores soluciones. Dicho esto, las características de cada grupo son:

Semáforo de vulnerabilidad ROJO

A. Con alta incidencia en 2016, provenientes en su mayoría del Estado de México y Ciudad de México siendo empleadas, con edades entre 26 a 29 años y un hijo en promedio, son quienes usaban condón y no especifican Método de Planificación Familiar (MPF) posterior, además se desconocen abortos, cesáreas, ILE previas, menarca e Inicio de Vida Sexual Activa (IVSA). Acudieron con siete Semanas De Gestación (SDG) y recibieron terapia dual, además se desconocen complicaciones en el procedimiento.
   a. Ahora bien, hay mujeres que pertenecen a un grupo minoritario y aunque tengan características similares con el clúster, no representan a la mayoría de mujeres en él. Sin

embargo, es importante hacer mención de estas minorías que pueden interpretarse como un subconjunto del clúster. En este grupo, la minoría de mujeres tienen una o más de las siguientes características:

i. Máximo primaria o incluso sin acceso a la educación.
ii. Mantienen el mismo método anticonceptivo antes y después del ILE.
iii. Tienen algún seguro (IMSS, ISSSTE, etc.).
iv. Mujeres casadas.
v. No firmaron o se desconoce si firmaron el consentimiento informado.
vi. IVSA menor a los 8 años.
vii. Hay quienes han tenido 2 o más gestas.

B. Acudieron en su mayoría durante 2019, del total de este grupo, son estudiantes de preparatoria con 19 a 21 años y sin hijos. Primera menstruación (Menarca) a los 12 años, IVSA cinco años después por lo que es muy probable que el procedimiento sea de su primera gesta.
   a. Es muy importante recalcar que del universo de mujeres que tuvieron complicaciones en el procedimiento, la mayoría está en este clúster.

C. Trabajadoras del hogar no remuneradas con preparatoria, de 22 a 29 años en unión libre con 1 hijo. Mantienen el condón como MPF después del procedimiento. Acuden con diez semanas o más SDG sin citas previas. Presencia de dolor después del procedimiento por lo que se prescriben analgésicos.

Semáforo de vulnerabilidad NARANJA

D. Mujeres foráneas que acuden en 2018, o son estudiantes o no contestaron ocupación, con edades de 19 a 21 años y sin hijos. Recibieron dos citas previas, consejería, terapia dual y se prescribió analgésico.

E. Trabajadoras del hogar no remuneradas de 22 a 25 años con secundaria y sin hijos. Menarca ligeramente tardía respecto al promedio: a los 13 años. Recibieron terapia dual y se desconoce si hubo complicaciones.

F. Trabajadoras del hogar no remuneradas de 22 a 35 años con secundaria y dos hijos en promedio. No ocupan MPF y acuden con una cita previa y siete semanas de gestación, recibe terapia dual y sin dolor después del procedimiento.

G. Trabajadoras del hogar no remuneradas de 30 a 35 años con secundaria, en unión libre con uno o más hijos y acuden acompañadas por su pareja. Es referida de otra unidad con tres o más citas previas, hubo dolor por lo que se prescribió analgésico.

H. Alta frecuencia en 2020, personas de diferentes niveles educativos (pocos sin acceso a la educación) con edades entre 22 a 25 sin hijos ni MPF previo. Recibieron consejería y no se complica el procedimiento ni tuvieron dolor después de él.
   a. La minoría de mujeres separadas y/o desempleadas están en este grupo.

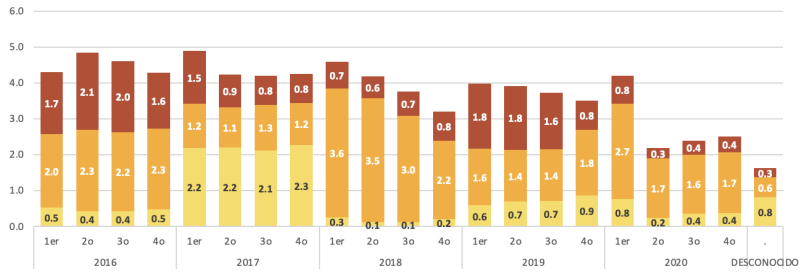
Semáforo de vulnerabilidad AMARILLO

I. No se conoce la fecha de la ILE, estudiantes o empleadas entre 19 y 29 años sin hijos. Llega acompañada de alguien de confianza, ya sea familiar o amigo. Acude directamente a ser atendida por especialidad de gineco-obstetricia con seis a ocho semanas de gestación, no hay dolor después del procedimiento.

J. Mujeres mexiquenses que acuden en 2017 de 22 a 25 años de edad, empleadas y sin hijos. Sin MPF previo y después se deciden por implante subdérmico. Recibieron consejería y terapia dual, no se complica el procedimiento.

Aún cuando no se utilizó la variable fecha (ni en ninguna división como año, trimestre, mes) para generar los clústers, es muy interesante cómo los 10 grupos obtenidos (y a su vez agrupados por semáforo de vulnerabilidad) tienen tendencias notables a lo largo del tiempo.

Gráfica 8. Personas gestantes (en miles) por año-trimestre y semáforo de vulnerabilidad



Fuente: Elaboración propia con resultado del modelo ILE

Propuestas y conclusiones

Por un lado, tomando en cuenta el modelo de series de tiempo sobre la recepción de llamadas en la Linea Mujeres para las siguientes 52 semanas se propone el uso de datos en dos perspectivas:

1) Administración de los recursos: Al conocer las futuras necesidades de las mujeres que ocupan el servicio, tenemos visibilidad para la toma de decisiones en cuanto a la cantidad de profesionales que deberán atender dicha demanda, ya sea para la redistribución de ellos o para futuras contrataciones.
2) Visibilidad para las instituciones: Dado que la Línea Mujeres también ayuda a canalizar los casos con instituciones que ayuden a darle seguimiento, el modelo de series de tiempo podría ser una indicador sobre los futuros casos que estas podrían recibir.

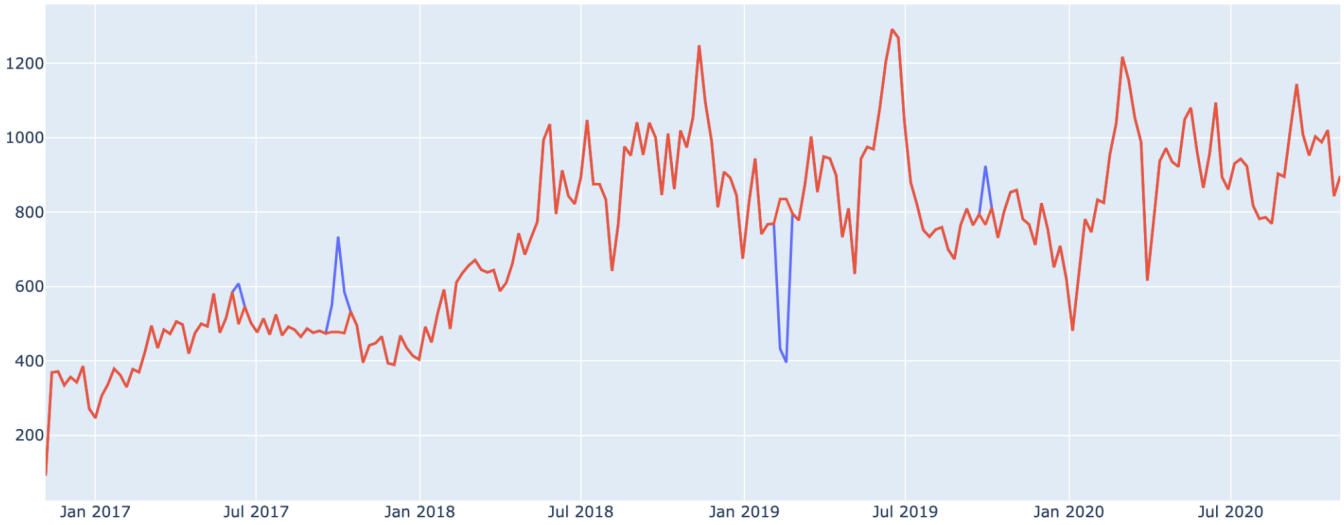
Por otro lado para el modelo ILE, la agrupación de personas gestantes con decisión de interrupción del embarazo y a su vez, la distribución en las diferentes alcaldías y municipios (véase Gráfica 9), facilita una personalización de campañas con un enfoque directo, por ejemplo: planificación familiar en jóvenes, seguridad para menores de edad, medicina preventiva para evitar complicaciones, propuesta de MPF según la edad y otras características que el modelo ya captura en su generalidad. Además, conocer dónde se ubican las minorías con problemas sociales importantes nos permite atacarlos de raíz, evitando los embarazos no deseados ocupando las herramientas estadísticas que hoy en día la tecnología pone al alcance de todos, tanto en este modelo como en el modelo de la Línea Mujeres, para brindar la atención que cada mujer merecemos: **ser más que una estadística**.

Los siguientes pasos para incrementar aún más el valor de los datos públicos son:

- Para LM: Generar clusters de las llamadas recibidas con el objetivo de crear campañas para la prevención de los casos, es decir, buscar reducir la cantidad de llamadas como consecuencia de campañas efectivas y no por desconocimiento de la existencia de la línea. Dicha segmentación deberá realizarse para cada uno de los tres servicios y así obtener una distribución de grupos mucho más efectiva.
- Para ILE: Modelar el pronóstico para cada grupo obtenido y así como con los servicios de LM, se podrían distribuir los recursos y difusión oportunamente, anticipando la demanda y necesidad de cada persona gestante que decide interrumpir su embarazo, teniendo así un impacto positivo en temas de Salud Pública.

Gráficas e información auxiliar

Gráfica 1. Aplicación de Hampel Filter para la serie de tiempo de llamadas con servicio Psicológico (atípicos: azul)

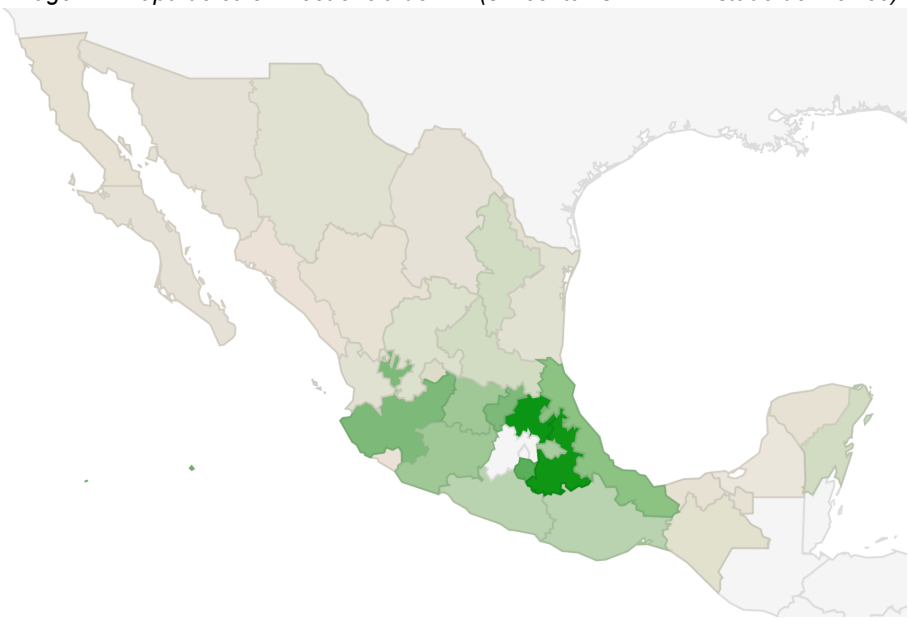


Fuente: Elaboración propia con imputación de valores atípicos

---

Repositorio del código para ambos modelos → aquí

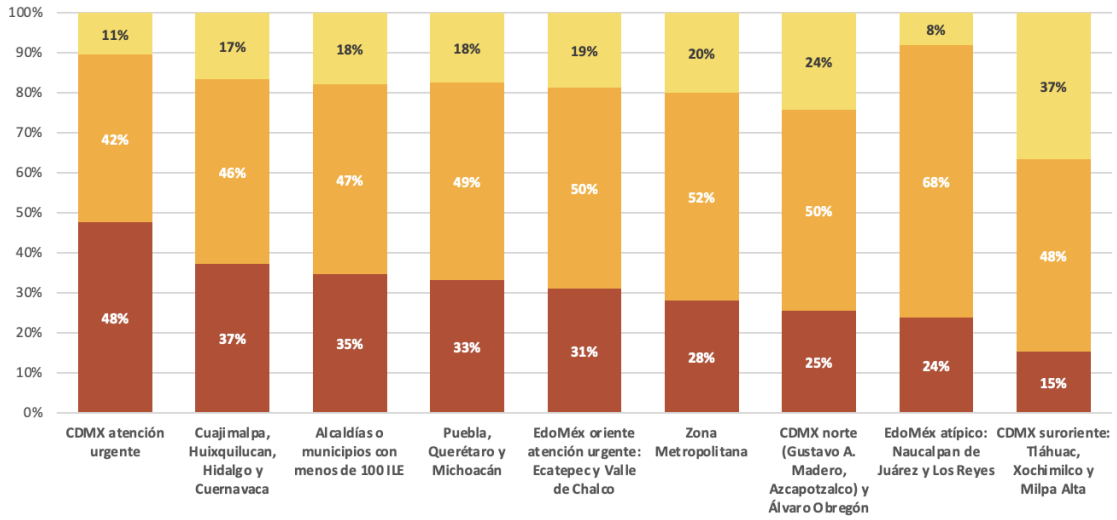
Imagen 1. Mapa de calor: frecuencia de ILE (sin contar CDMX ni Estado de México)



Fuente: Elaboración propia en sitio web: DanielPinero.com

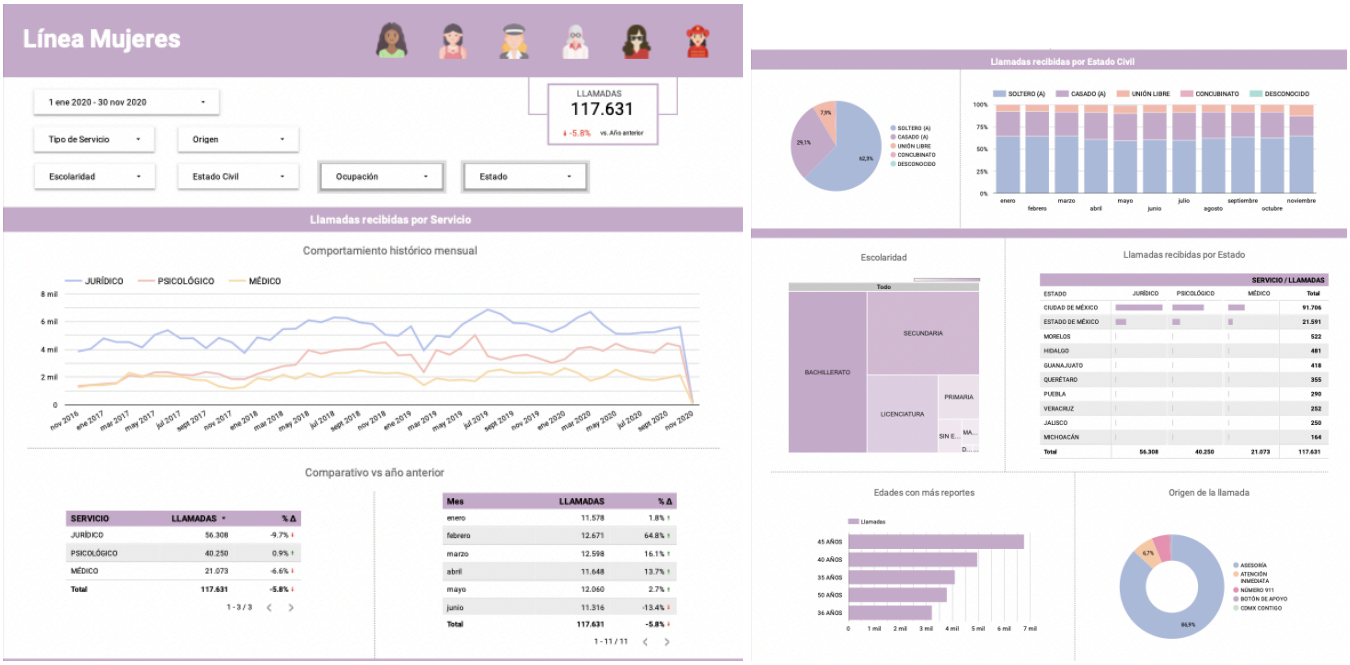
—

Gráfica 9. Distribución porcentual de semáforo por clusters de localidades ILE



Fuente: Elaboración propia con resultado de clustering por localidad según distribución de grupos ILE

Dashboard 1. Línea Mujeres



Fuente: Elaboración propia en [DataStudio]