

Preprocesamiento para Ciencia de Datos

Avances de Proyecto; segunda entrega.

Problemática y entendimiento del negocio: Consumo de sustancias psicoactivas en México.

Fernando R. Valenzuela García de León
Daniel Rojo Mata

Universidad Nacional Autónoma de México
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

18 de septiembre de 2025

Agenda de la Presentación

- 1 1. Contexto y Problemática
- 2 2. Entendimiento del Negocio y Solución
- 3 3. Preguntas que se pueden resolver
- 4 4. Conclusión

Contexto: Una Crisis de Salud Pública

Datos Clave ENCODAT 2016

- Consumo de drogas ilícitas: **9.9 % alguna vez en la vida.**

Contexto: Una Crisis de Salud Pública

Datos Clave ENCODAT 2016

- Consumo de drogas ilícitas: **9.9 % alguna vez en la vida.**
- Adolescentes (12–17): **se duplicó**, de 2.9 % a 6.2 %.

Contexto: Una Crisis de Salud Pública

Datos Clave ENCODAT 2016

- Consumo de drogas ilícitas: **9.9 % alguna vez en la vida.**
- Adolescentes (12–17): **se duplicó**, de 2.9 % a 6.2 %.
- Alcohol excesivo en el último mes: **19.8 % (+50 % desde 2011)**.

El Reto Principal

No falta información, sino que los datos están **altamente fragmentados**.

Fuente: Secretaría de Salud (2017); ENCODAT 2016-2017.

El Problema: Data Incomunicada



1.

Datos Relacionales

- Defunciones y urgencias.
- Problema: históricos, agregados, sin trazabilidad.



2.

Texto No Estructurado

- Expedientes clínicos, reportes.
- Problema: conocimiento atrapado, difícil de analizar.



3.

Conocimiento Experto

- Jerarquías (CIE-10), policonsumo.
- Problema: conceptual, sin vínculo con datos operativos.

Esta desconexión impide una visión integral y la toma de decisiones oportuna.

¿Para Quién Resolvemos Este Problema?

Actores Clave en México

- Secretaría de Salud / CONASAMA: visión macro para políticas públicas.

¿Para Quién Resolvemos Este Problema?

Actores Clave en México

- Secretaría de Salud / CONASAMA: visión macro para políticas públicas.
- Hospitales y clínicas: historiales clínicos y gestión de recursos.

¿Para Quién Resolvemos Este Problema?

Actores Clave en México

- Secretaría de Salud / CONASAMA: visión macro para políticas públicas.
- Hospitales y clínicas: historiales clínicos y gestión de recursos.
- Gobiernos locales: datos geoespaciales para focalizar intervenciones.

¿Para Quién Resolvemos Este Problema?

Actores Clave en México

- Secretaría de Salud / CONASAMA: visión macro para políticas públicas.
- Hospitales y clínicas: historiales clínicos y gestión de recursos.
- Gobiernos locales: datos geoespaciales para focalizar intervenciones.
- Investigadores y academia: microdatos granulares y longitudinales.

Fuente: Conasama, OPS, EMCDDA.

Una Posible Solución: Una Base de Datos Federada

Proponemos integrar las tres fuentes heterogéneas usando el **código CIE-10 (F10-F19)** como llave maestra.



Figura: Arquitectura de integración federada vía CIE-10.

Pregunta 1: Perfil Epidemiológico

La Pregunta

Riesgo Alcohol–Cocaína (F10–F14): **¿Cuál es el perfil demográfico (edad, sexo, entidad) de las muertes con ambos diagnósticos?**

¿Por qué es importante?

- Focalizar **campañas de prevención.**

Pregunta 1: Perfil Epidemiológico

La Pregunta

Riesgo Alcohol–Cocaína (F10–F14): **¿Cuál es el perfil demográfico (edad, sexo, entidad) de las muertes con ambos diagnósticos?**

¿Por qué es importante?

- Focalizar **campañas de prevención**.
- Identificar si el fenómeno es **nacional o regional**.

Pregunta 1: Perfil Epidemiológico

La Pregunta

Riesgo Alcohol–Cocaína (F10–F14): **¿Cuál es el perfil demográfico (edad, sexo, entidad) de las muertes con ambos diagnósticos?**

¿Por qué es importante?

- Focalizar **campañas de prevención**.
- Identificar si el fenómeno es **nacional o regional**.
- Generar evidencia para **guías clínicas**.

¿Cómo se responde?

- **Grafo:** identificar la arista F10–F14.
- **SQL:** filtrar *defunciones* con ambos códigos.
- **Resultado:** perfil demográfico y geográfico detallado.

Pregunta 2: Traducción y Descubrimiento

La Pregunta

¿Cómo el sistema traduce términos como “crack” o “piedra” a F14 y genera estadísticas de mortalidad y urgencias?

¿Por qué es importante?

- **Democratiza el acceso** a los datos.

Pregunta 2: Traducción y Descubrimiento

La Pregunta

¿Cómo el sistema traduce términos como “crack” o “piedra” a F14 y genera estadísticas de mortalidad y urgencias?

¿Por qué es importante?

- **Democratiza el acceso** a los datos.
- Responde rápido a **funcionarios o prensa**.

Pregunta 2: Traducción y Descubrimiento

La Pregunta

¿Cómo el sistema traduce términos como “crack” o “piedra” a F14 y genera estadísticas de mortalidad y urgencias?

¿Por qué es importante?

- **Democratiza el acceso** a los datos.
- Responde rápido a **funcionarios o prensa**.
- Vigila jergas de **drogas emergentes**.

¿Cómo se responde?

- **Texto (NLP)**: mapear “crack/piedra” → F14.
- **SQL**: consultas automáticas en defunciones y urgencias.
- **Resultado**: dashboard con tendencias y mapas.

Pregunta 3: Validación del Modelo

La Pregunta

Cannabis-Estimulantes: **¿Podemos validar estadísticamente esta conexión con datos de México?**

¿Por qué es importante?

- Convierte el sistema en **plataforma de conocimiento.**

Pregunta 3: Validación del Modelo

La Pregunta

Cannabis-Estimulantes: **¿Podemos validar estadísticamente esta conexión con datos de México?**

¿Por qué es importante?

- Convierte el sistema en **plataforma de conocimiento**.
- Ajusta el modelo con **evidencia local**.

Pregunta 3: Validación del Modelo

La Pregunta

Cannabis-Estimulantes: **¿Podemos validar estadísticamente esta conexión con datos de México?**

¿Por qué es importante?

- Convierte el sistema en **plataforma de conocimiento**.
- Ajusta el modelo con **evidencia local**.
- Demuestra que el sistema puede **auto-evaluarse**.

¿Cómo se responde?

- **SQL:** calcular co-ocurrencia F12–F15.
- **Grafo:** comparar frecuencia real vs. teórica.
- **Resultado:** validar o ajustar la fuerza de conexión.

Más Preguntas Resolubles

Además de las ya presentadas, el sistema puede explorar:

- **Análisis de Subtipos (Grafo + SQL):** Para los opioides (F11), ¿qué subtipos específicos (ej. F11.1 'uso nocivo' vs. F11.2 'dependencia') son más frecuentes en **urgencias** versus en **defunciones**?

Más Preguntas Resolubles

Además de las ya presentadas, el sistema puede explorar:

- **Análisis de Subtipos (Grafo + SQL):** Para los opioides (F11), ¿qué subtipos específicos (ej. F11.1 'uso nocivo' vs. F11.2 'dependencia') son más frecuentes en **urgencias** versus en **defunciones**?
- **Tendencias Demográficas (SQL):** ¿Ha cambiado la tasa de mortalidad por estimulantes (F15) de manera diferente para **mujeres y hombres** a lo largo de los años disponibles en los datos?

Más Preguntas Resolubles

Además de las ya presentadas, el sistema puede explorar:

- **Análisis de Subtipos (Grafo + SQL):** Para los opioides (F11), ¿qué subtipos específicos (ej. F11.1 'uso nocivo' vs. F11.2 'dependencia') son más frecuentes en **urgencias** versus en **defunciones**?
- **Tendencias Demográficas (SQL):** ¿Ha cambiado la tasa de mortalidad por estimulantes (F15) de manera diferente para **mujeres y hombres** a lo largo de los años disponibles en los datos?
- **Patrones Geográficos Cruzados (SQL):** ¿Son las mismas entidades federativas las que presentan la mayor carga de urgencias por **alcohol (F10)** y por **cannabis (F12)**, o existen patrones geográficos distintos?

Más Preguntas Resolubles

Además de las ya presentadas, el sistema puede explorar:

- **Análisis de Subtipos (Grafo + SQL):** Para los opioides (F11), ¿qué subtipos específicos (ej. F11.1 'uso nocivo' vs. F11.2 'dependencia') son más frecuentes en **urgencias** versus en **defunciones**?
- **Tendencias Demográficas (SQL):** ¿Ha cambiado la tasa de mortalidad por estimulantes (F15) de manera diferente para **mujeres y hombres** a lo largo de los años disponibles en los datos?
- **Patrones Geográficos Cruzados (SQL):** ¿Son las mismas entidades federativas las que presentan la mayor carga de urgencias por **alcohol (F10)** y por **cannabis (F12)**, o existen patrones geográficos distintos?
- **Rutas de Policonsumo (Texto + Grafo + SQL):** Al buscar "benzodiacepinas" (F13), ¿cuál es el diagnóstico de policonsumo más frecuente en defunciones, según las **conexiones del grafo** (ej. con opioides, F11)?

Conclusión: De Datos Fragmentados a Conocimiento Utilizable

El Problema

La fragmentación de datos en silos impide una visión completa del fenómeno.

Conclusión: De Datos Fragmentados a Conocimiento Utilizable

El Problema

La fragmentación de datos en silos impide una visión completa del fenómeno.

El Negocio

Actores diversos (gobierno, hospitales, investigadores) requieren información distinta pero interconectada.

Conclusión: De Datos Fragmentados a Conocimiento Utilizable

El Problema

La fragmentación de datos en silos impide una visión completa del fenómeno.

El Negocio

Actores diversos (gobierno, hospitales, investigadores) requieren información distinta pero interconectada.

Possible Solución

Una base de datos federada que actúa como puente, habilitando consultas transversales y transformando datos aislados en conocimiento utilizable.

Un modelo de datos integrado es un requisito estratégico para la salud pública.

¡Gracias por su atención!, :)



Referencias I

-  Secretaría de Salud, et al. (2017). *Consumo de drogas: prevalencias...*
-  Unidad de Encuestas y Análisis de Datos. (2018). *Tendencias nacionales...*
-  Johns Hopkins University. (2023). *Substance Abuse Among Children.*
-  Secretaría de Salud, et al. (2017). *Consumo de alcohol: prevalencias...*