

# Propuesta proyecto

## Integrantes:

Javier Navarro Espíndola.

Saúl Armas Gamiño.

## ¿Qué se va a hacer?

Se van a analizar datos referentes a delitos ocurridos en la ciudad de México. Pensamos hacer clustering para ver si hay perfiles distinguibles de delitos con los atributos seleccionados y también reglas de asociación para ver si hay patrones latentes.

El dataset inicialmente contaba con 18 columnas, de las cuales, nos quedamos con 6, ya que pensamos que algunas no nos son de utilidad para el análisis, los 6 atributos que dejamos son:

1. Fecha-hora hechos, la cual se refiere a la fecha-hora en que ocurren los hechos.
2. Delito, el cual describe el delito reportado
3. Agencia, la cual significa donde se reportaron los hechos.
4. Unidad de investigación, la cual hace referencia a quien tomó el caso.
5. Alcaldía de los hechos, localidad (delegación)
6. Fecha\_inicio, se refiere a cuando se reportaron los hechos
7. Usaremos un atributo calculado, el cual será la diferencia de cuando ocurren los hechos, y cuando son reportados.
  - Con los atributos anteriores pensamos construir otros atributos, por ejemplo: hora del día representado solamente por 3 valores (mañana, tarde y noche por ejemplo).

La idea es hacer varios experimentos cambiando atributos, quitando otros o construyendo nuevos para lograr sacar información e incluso poder hacer algunas hipótesis con ella.

## ¿De dónde sacamos los datos?

Los datos que obtuvimos del gobierno de México, del siguiente [link](#), son datos públicos y abiertos.