

# Análisis computacional basado en datos de la reinserción social de los Internos en el CERESO Apodaca y en el Centro Preventivo de Reinserción Topo Chico

Evely Gutiérrez Noda

Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Nuevo León

## Abstract

Para comprender e intentar mejorar las condiciones de reinserción social de los internos en el Cereso Apodaca y en el Centro Preventivo de Reinserción Topo Chico se hace necesario conocer e identificar los factores y condiciones sociales de cada interno. En este trabajo se evalúan las condiciones de vida, económicas, educativas y familiares de estas personas, así como las condiciones de los penitenciarios. Los datos utilizados provienen de encuestas realizadas a mujeres y hombres por separado, que se encuentran privados de libertad, se tienen datos básicos como edad, tiempo de sentencia, delito cometido, nivel educacional, estado civil, ingresos mensuales antes de estar internos, si tienen hijos y cuantos de ellos nacen y viven en la prisión (para el caso de las mujeres), dibujos echos a mano y visitas familiares. Utilizando bibliotecas de Python se logra construir escenarios para identificar causas y factores que pueden ser causa de cometer un delito mediante modelos estadísticos y de aprendizaje automático. Los resultados arrojan que el tener ingresos bajos y una numerosa familia provoca cometer un delito de grado mayor como homicidio, así como el nivel educacional bajo.

## 1. Introducción

Existe poca bibliografía sobre el tema de las prisiones en México y aún es más escasa para el caso de las mujeres en prisión. El siguiente trabajo busca conocer la situación de los reclusos en los penales del Estado de Nuevo León, desde el punto social y económico antes de estar privados a la libertad. El contexto de la investigación son los centros de privación de libertad Cereso Apodaca y Topo Chico. El estudio busca detectar e identificar los factores socioeconómicos de los internos, abordando temas sobre la problemática de los internos y sus posibilidades de reinserción social, siendo esta la contribución de este trabajo.

## 2. Antecedentes

El sistema penitenciario durante años estuvo enfocado en la privación de la libertad como castigo y fue hasta fechas recientes que el mandato constitucional lo encaminó a la reinserción social. Es un eslabón sustancial en el proceso de seguridad pública y uno de los más criticados por el estado de crisis que en él permanece.

Actualmente de la población penitenciaria en el mundo, 714 mil son mujeres y niñas internadas. Los análisis internacionales en torno al tema indican que la población penitenciaria ah crecido de una forma superior, incluso la de mujeres mas que hombres a partir del año 2000. La población masculina creció en estos años un 20 % y la femenina un 50 %.

El Instituto Nacional de Estadística y Geografía en el 2017 señala que al cierre del año los centros penitenciarios estatales

de México tenían 188,262 personas privadas a la libertad, de las cuales un 95 % eran hombres y un 5 mujeres. No toda esta población estaba ya bajo sentencia, siendo un 35 % de los internos los que se encuentran bajo un estatus jurídico sin sentencia ya que aun están en ese proceso.

En general, los servicios penitenciarios se rigen por los principios que se describen en las legislaciones penales de cada país. El sistema penitenciario es el cúmulo de disposiciones legales en las que se establecen las penas, delitos y procesos; además de designar a las autoridades estatales encargadas de su ejecución y administración [1].

### 2.1. Estructura del sistema penitenciario mexicano

La Constitución Política de los Estados Unidos Mexicanos (CPEUM) ha encomendado distintas funciones al sistema penitenciario. Actualmente, el artículo 18 señala que privar de la libertad aspira a reinsertar al sentenciado a la sociedad. Sin embargo, dicho objetivo, advierte UNODC, generalmente no puede alcanzarse durante el encarcelamiento debido a que en la prisión se tiende a criminalizar a los individuos, lo cual los orilla a delinquir nuevamente <sup>1</sup>.

Las características del sistema penitenciario en México dificultan la reinserción social, motivo por el cual centrar la ejecución de la pena en el respeto a los derechos humanos parece un camino viable [5]. La CPEUM establece en su artículo 18 los objetivos y principios sobre los cuales tiene que organizarse el sistema penitenciario, alguno de ellos son: respetar los derechos humanos, buscar la reinserción de las personas sentenciadas a

Email address: evely.gutierrez.noda89@gmail.com (Evely Gutiérrez Noda)

<sup>1</sup>UNODC: [www.unodc.org/mexicoandcentralamerica/es/index.html](http://www.unodc.org/mexicoandcentralamerica/es/index.html)

través del trabajo, capacitación y educación y garantizar que las mujeres cumplan sus condenas en lugares distintos a los destinados para los hombres [3].

El sistema penitenciario propone medios alternativos para la solución de controversias, como el acuerdo reparatorio y la suspensión condicional del proceso, así como formas de terminación anticipada. Además, establece que las audiencias sean públicas, orales y en español [9]. Todo lo anterior tiene como objetivo garantizar un mejor desempeño de las instituciones involucradas en el proceso de justicia penal a través de la liberación de las cargas de trabajo, la promoción de la transparencia, el combate efectivo a la impunidad lograr establecer la justicia de manera pronta, eficiente y efectiva [3].

Existen dos tipos de delitos por tipo de fuero: común y federal. Los delitos del fuero común son aquellos en los que la persona es afectada por el delincuente (homicidio, amenazas, robo en cualquier modalidad, entre otros). Mientras que los delitos del fuero federal son, en general, los que perjudican los intereses de la Federación o la seguridad del país, como la defraudación fiscal, los delitos electorales, los daños o robos al patrimonio arqueológico, el lavado de dinero, el narcotráfico, etcétera. Estas ofensas son perseguidas y sancionadas por las autoridades de la federación<sup>2</sup>.

La reforma constitucional de Seguridad y Justicia del año 2008 y la de Derechos Humanos del año 2011 transformaron el sistema penitenciario en uno garante que busca la reinserción social de la persona sentenciada mediante el trabajo, la capacitación, la educación, la salud, el deporte y el respeto a sus derechos fundamentales [8]. Las reformas legales han provisto de derechos a las personas imputadas. Ahora tienen la garantía de una defensa adecuada y derecho a ser consideradas inocentes hasta que se compruebe lo contrario. Según un estudio del Instituto de Investigación sobre Políticas Criminales, el uso excesivo del encarcelamiento conduce al hacinamiento, condiciones de detención degradantes e inhumanas y pobres resultados de rehabilitación [6].

Además, se incorporaron la reinserción social de la persona sentenciada como fin del encarcelamiento y la modificación de penas a cargo de la autoridad judicial. Ahora un juez de ejecución podrá disminuir el tiempo en prisión, imponer medidas de seguridad, así como resolver controversias sobre las condiciones de internamiento. Finalmente, el artículo 22 de la CPEUM establece que las penas deberán ser proporcionales con el delito cometido y bien jurídico afectado. La aplicación de estas medidas, junto con las que se contemplan en la Ley Nacional de Ejecución Penal, centran el sistema penitenciario en el respeto a los derechos humanos de las personas privadas de la libertad. Puesto que busca mejorar las condiciones de vida de quienes viven en los centros penitenciarios a través de: la supervisión de los centros penitenciarios; capacitación y profesionalización del personal penitenciario; mejora de la infraestructura, estandarización de procesos y procedimientos y; el desarrollo de un sistema integral de reinserción social.

<sup>2</sup>UNODC: [www.unodc.org/mexicoandcentralamerica/es/index.html](http://www.unodc.org/mexicoandcentralamerica/es/index.html)

## 2.2. Ciencia de datos

La ciencia de datos es el estudio de dónde proviene la información que permite la extracción de grandes cantidades de datos estructurados y no estructurados para identificar patrones que pueden ayudar a una organización a controlar los costos, aumentar la eficiencia, reconocer nuevas oportunidades de mercado y aumentar la ventaja competitiva de la organización. El campo de la ciencia de datos emplea matemáticas, estadística y disciplinas informáticas e incorpora técnicas como el aprendizaje automático, el análisis de conglomerados, la extracción de datos y la visualización [11].

La Ciencia de Datos permite extraer, clasificar y procesar grandes cantidades de información. Es un área de conocimiento que busca agregar valor a los datos por medio de la confluencia de análisis estadísticos, métodos computacionales y visualización de información.

La principal ventaja de contar con la ciencia de datos en una organización es el apoderamiento y la facilitación de la toma de decisiones. Las organizaciones con científicos de datos pueden tener en cuenta las pruebas cuantificables basadas en datos en sus decisiones comerciales. Estas decisiones basadas en datos pueden, en última instancia, conducir a una mayor rentabilidad y una mayor eficiencia operativa, rendimiento de negocio y flujos de trabajo. En las organizaciones orientadas al cliente, la ciencia de datos ayuda a identificar y refinar las audiencias objetivo. La ciencia de datos también puede ayudar al reclutamiento, por ejemplo, el procesamiento interno de aplicaciones y las pruebas de aptitud basadas en datos pueden ayudar al equipo de recursos humanos de una organización a realizar selecciones más rápidas y precisas durante el proceso de contratación [7].

Los beneficios de la ciencia de datos pueden variar según el objetivo de la empresa y la industria. Para el caso de los departamentos de ventas y marketing pueden extraer datos de clientes para mejorar las tasas de conversión, para el caso de las instituciones bancarias se puede utilizar para mejorar la detección de fraudes, servicios de transmisión como Netflix pueden usarla para determinar en que están interesados en ver sus usuarios y así proponerles programas de televisión o películas. Las compañías de envíos como DHL, FedEx y UPS pueden utilizar la ciencia de datos para encontrar las mejores rutas y horarios de entrega, así como los mejores modos de transporte para sus envíos.

## 3. Literatura relacionada

Una investigación realizada para evaluar psicológica y neuropsicológicamente a un grupo de internos violentos de una Penitenciaría Federal en México y caracterizarlos de acuerdo con la clasificación de impulsividad y premeditación de Barratt y utilizando la clasificación de Hare (1991) de psicopatía. Tuvo como objetivo analizar las diferencias entre los grupos de acuerdo a diferentes clasificaciones. La muestra fue integrada por 75 participantes, 50 internos y 25 controles para el estudio 1 (impulsivos vs premeditados) y 54 sujetos para el estudio 2 (no psicópatas vs psicópatas). Obtuvieron resultados que indican que las poblaciones violentas independientemente de su

clasificación pueden ser diferenciadas de la población no violenta, pero sin embargo, no se encontraron diferencias significativas en las evaluaciones neuropsicológicas entre los grupos impulsivo y premeditado, mientras que los resultados entre los no psicópatas y psicópatas muestran diferencias significativas en distintos dominios cognoscitivos [4].

Otro estudio realizado compara las curvas de crecimiento de la población interna de diversos países americanos respecto de la provincia de Buenos Aires, analizando las semejanzas y diferencias en términos generales y realizan un análisis descriptivo en términos sexuales y delictivos de dicho crecimiento en la Provincia de Buenos Aires. Los resultados muestran que mientras que en países centrales la tasa de crecimiento es constante, en Latinoamérica las políticas de encarcelamiento dependen de un factor limitante de hacinamiento. Para el caso de las mujeres en la provincia de Buenos Aires, además de presentar una tendencia al encarcelamiento mayor que los hombres, no son objeto de las políticas de excarcelación y presentan una curva de crecimiento exponencial de la población [10].

En México se realizó un estudio que ofrece un panorama amplio acerca de la situación general del sistema penitenciario mexicano. Se documenta el crecimiento desbordado que ha tenido la población penitenciaria durante los últimos años y el deterioro que se observa en las condiciones de vida en prisión. Ofrecen un panorama detallado sobre la población penitenciaria y las características de los establecimientos penitenciarios del centro de la República Mexicana y opinan sobre los desaciertos de las políticas que han incrementado de manera indiscriminada la población en reclusión [2].

## 4. Metodología

La siguiente sección describen las características de los datos utilizados para este estudio así como la naturaleza de los mismos. Se describen además las herramientas utilizadas para todo el procesamiento de datos que se realiza.

### 4.1. Datos utilizados

Se cuenta con dos encuestas, una aplicada a los hombres y otra a las mujeres, con alrededor de 287 preguntas cada una y una sesión de autoevaluación del estado de ánimo del encuestado. Las dos encuestas son muy parecidas, hacen diferencias en algunas preguntas referidas a las condiciones que tienen las mujeres en el penitenciario para tener a sus hijos, en caso que hayan dado a luz a su bebé estando bajo cumplimiento de condena, dado que viven con ellas sus primeros 2 años de vida.

Las preguntas de los cuestionarios giran sobre temas como: la infraestructura y los servicios de la penitenciaria, la situación familiar de los reclusos, si se realiza trabajo social con ellos, si los atienden clínicamente en caso de adicciones así como atención psicológica, sobre los servicios educativos y de capacitación, contiene temas de cultura y deporte, del ámbito laboral antes de ingresar al penitenciario y dentro del mismo, si cuentan o no con servicios de asesoría penal y jurídica, abordan el caso de la reinserción social de los reclusos y se tienen una sesión de preguntas referentes al estado de anímico del recluso.

Se conoce que estas encuestas fueron respondidas por aproximadamente 312 hombres y 172 mujeres que se encuentran privados de libertad en CERESO en Apodaca. Fueron realizadas en papel y las respuestas se digitalizaron mediante formato .CSV. Estos datos se encuentran parcialmente limpios, por tanto, se procederá a revisar y limpiar nuevamente en caso de ser necesario.

### 4.2. Herramientas utilizadas para procesamiento de los datos

Se utilizan herramientas de Bash, que es un lenguaje de consola para el manejo de distintos tipos de archivos. Se comienza haciendo copias de los archivos originales que se van a utilizar, en este caso, backup-male.csv para las respuestas a las encuestas de los hombres, backup-female.csv para las de mujeres y backup-uniform.csv para la unión de los datos de hombres y mujeres. Con estas copias se garantiza no dañar los datos originales en caso de llegar a cometer errores y así no perder información valiosa. Para hacer las copias de los archivos se usa el comando de bash cp.

---

```
cp uniform.csv backup-uniform.csv
cp female.csv backup-female.csv
cp male.csv backup-male.csv
```

---

Otra forma más abreviada de hacer las copias de los 3 archivos al mismo tiempo es:

---

```
for encuesta in "female" "male" "uniform"; do awk
'{{print "union" $0}}' backup-${encuesta}.csv >
backup${encuesta}.csv; done
```

---

Con las herramientas Bash wc -l se puede conocer la cantidad de filas en cada uno de los archivos de datos. Esta herramienta cuenta las líneas de los archivos, por tanto agrega una línea más correspondiente al encabezado de la columna, por tanto, la cantidad total de mujeres y hombres es uno menos de los totales que se muestran.

---

```
wc -l backup-*
173 backup-female.csv
313 backup-male.csv
485 backup-uniform.csv
971 total
```

---

Con otra de las herramientas Bash se pueden separar las columnas que se consideran útiles para analizar. Para esto se utiliza la herramienta awk, en este caso son las columnas 12: Nivel Escolar y 29: Primera vez Interno. Esta información se obtiene del fichero backup-male.csv y se almacenan en otro fichero pregunta1male.csv por el comando de salida >.

---

```
awk -F',' ' {{print $12,"$29}} ' backup-male.csv >
pregunta1male.csv
```

---

Siguiendo con el uso del comando awk y las herramientas sort y uniq se puede ordenar las columnas deseadas y unificar por respuestas, para facilitar el análisis de los datos. En este ejemplo se trabaja con datos de los hombres y unifican por el nivel de escolaridad.

---

```
awk -F',' '{print $3}' pregunta1male.csv | sort |
uniq -c
```

---

```
1 Comercial
9 Licenciatura
12 Ningun estudio
30 Preparatoria
86 Primaria
158 Secundaria
5 Tecnico
```

---

Gracias a estas herramientas se puede ver que la mayoría de los hombres internos en este caso, son de bajo nivel educacional, siendo 158 con nivel Secundaria. Para cargar los datos a un Data Frames se utiliza la librería Pandas, ya que los datos están en formato .CSV y es una gran cantidad de información que al abrirlos en un Excel se perderían columnas. Se presentan los ejemplos de trabajos con la librería Pandas:

- Agregar columnas en común para hombres y mujeres en el data frame y guardarlas en un fichero .CSV nuevo usando la librería pandas. Algunas columnas fueron: gender (correspondiente al género), stayYears, stayMonths, stayDays, (correspondiente al tiempo que llevan en el penitencinario), crime, (correspondiente al crimen que cometieron), sentenceYears, sentenceMonths, sentenceDays (correspondiente al tiempo de sentencia a partir del delito cometido).
- Determinar qué cantidad de hombres y mujeres llevan privados de la libertad y aun están esperando sentencia.
- Conocer el estado civil de los internos antes y después de entrar en el penal. Indagar sobre que cantidad de hijos menores de edad tienen los internos y cuantos de ellos cuentan con seguro IMSS o Popular.
- Trabajar sobre temas relacionados con los hijos que viven con las internas en el penal y los que no viven con ellas.

Además, se refleja el trabajo realizado con la librería Pandas en algunos ejemplos de:

- Leer data frame de archivos .CSV.
- Guardar el data frame en un fichero con formato .CSV.
- Agregar columnas de pruebas en el data frames.
- Combinar dos o más data frames (ej. el de hombres con el de mujeres).
- Realizar filtros a los renglones de un data frame para practicar el trabajo con la librería Pandas.

Para comenzar la lectura de datos y cargarlos en un Data Frame, se necesita importar la librería Pandas, esto se logra con el siguiente código en la consola.

---

```
pip3 install pandas
```

---

Ya teniendo instalado pandas, se comienza la lectura de datos con el siguiente código en Python 3. Primero se crea una clase llamada ReadData para la lectura del fichero uniform.csv usando la librería pandas mediante la variable pd. Luego se seleccionan las columnas que se van a utilizar y se guardan en un fichero nuevo llamado UniformDataFrame.csv.

Ya teniendo el archivo con los datos extraídos para analizar, se comienza a realizar algunas pruebas con ellos, por ejemplo, se busca la cantidad de hombres que responden detalladamente el tiempo que llevan dentro del penitencinario (respuestas con días, meses y años), de igual modo para las mujeres, y se busca también la cantidad de hombres y mujeres por independientes no responden a esa pregunta con tanta especificación. El código en Python para esto y los resultados obtenidos se muestran a continuación:

---

```
from ReadData import ReadData
from UniformMaleFemale import UniformMaleFemale
rd = ReadData()
def buscaCantidad(sexo, fichero, i, j, k):
    with open(fichero, 'r') as uniform:
        todo = 0
        next(uniform)
        for lineas in uniform:
            temp = lineas.split(',')
            if str(temp[1]) == sexo:
                if int(temp[i]) > 0 and
                   int(temp[j]) >= 0 and
                   int(temp[k]) > 0:
                    todo = todo + 1
                #print(temp)
            else:
                if int(temp[i]) > 0 and
                   int(temp[j]) > 0 and
                   int(temp[k]) > 0:
                    todo = todo + 1
        uniform.close()
    return todo
```

---

Hombres que responden detalladamente tiempo de estancia:

2

Mujeres que responden detalladamente tiempo de estancia:

0

Hombres que responden detalladamente tiempo de sentencia:

10

Mujeres que responden detalladamente tiempo de sentencia:

1

---

El trabajo anterior se pudiera resumir a dos líneas de código en Python solamente leyendo y filtrando el Data Frame y analizando el fichero resultante con herramientas de Bash. El siguiente ejemplo es para determinar el caso de los internos hombres cuántos de ellos responden detalladamente el tiempo de sentencia, primero se filtra el data frame por estas colum-

nas, solo guardando las que sean mayores que cero, y luego se guardan los resultados en un fichero .CSV nuevo. En Bash se cuentan estos resultados y se sabe cuántos hombres responden detalladamente con día, mes y año el tiempo de sentencia que tienen. Es bueno hacer esta comparación para ver las utilidades y facilidades que brindan las herramientas Bash y el trabajo con Data Frame y la librería Pandas.

---

```
$ awk -F ',' '{print $2}' TiempoSentenciaHombres.csv
| sort | uniq -c

10 male
```

---

Otro trabajo realizado con esta misma herramienta es calcular cuántos hombres y mujeres están privados de la libertad y aun están esperando sentencia, arrojando que 68 hombres y 86 mujer están bajo estas condiciones.

---

```
Hombres sin sentencia:
68
Mujeres sin sentencia:
86
```

---

Otra herramienta utilizada es Plotly, la cual es una de las mejores librerías de uso libres para crear una variedad de gráficas interactivas y ricas en funcionalidad. Se crean algunas gráficas de diferentes tipos que permitirán concluir algunas cosas sobre los datos que se obtuvieron de las encuestas. Con esta librería se pueden graficar diagramas del tipo caja-bigote, Histograma, Histograma horizontal, Histograma superpuesto y de barra.

Plotly ofrece un API excelente para crear gráficas interactivas que pueden ser incluidas en webs y blogs. El paquete no viene incluido, pero se puede instalar de manera muy sencilla con pip en consola con la siguiente línea:

---

```
$pip install plotly
```

---

Para poder utilizar Plotly se necesitan credenciales, por lo que es necesario registrarse. Este registro es gratuito, sin embargo, el número de archivos privados que se pueden almacenar es limitado en la versión gratuita. Luego de tener ya lista la librería, se leen los datos guardados en los ficheros .CSV, se importan las funcionalidades a utilizar y se cambian las credenciales por el APY KEY que se genera luego de crear un usuario en Plotly.

Se realizó un agrupamiento de datos o Clustering, técnica de minería de datos (data mining) dentro de la disciplina de Inteligencia Artificial, que identifica de forma automática agrupaciones o clústeres de elementos de acuerdo a una medida de similitud entre ellos <sup>3</sup>.

El objetivo fundamental de las técnicas de clustering consiste en identificar grupos o clústeres de elementos tal que:

- La similitud media entre elementos del mismo clúster sea alta. Similitud intra-clúster alta.

- La similitud media entre elementos de distintos clústeres sea baja. Similitud inter-clúster baja.
- La identificación de clústeres o grupos de elementos se basa en una medida de similitud. Diferentes medidas de similitud dan lugar a diferentes clústeres.

En la mayoría de los análisis hechos anteriormente se ha trabajado con datos numéricos y categóricos, pero dado que los datos provienen de encuestas donde aparecen frecuentemente preguntas abiertas, y estas son contestadas por personas, en lenguaje natural, surge la necesidad de analizarlos también. Para esto se trabaja con nltk, librería de Python que se utiliza para realizar análisis de textos y también permite la visualización de los resultados obtenidos mediante cuadros, gráficas (y o) extractos de texto. Se trabaja con respuestas a mano que tienen que ver con la Reinserción Social, ya que es un tema muy importante para los internos en el momento de obtener la libertad y retomar sus vidas en libertad.

## 5. Resultados Obtenidos

Se realizó una estadística descriptiva básica para obtener información de los datos de los internos con el objetivo de realizar conteos, promedios, correlaciones, junto con limpieza, categorización y cuantificación de la información. Los ficheros resultantes del procesamiento anterior que ya quedaron limpios y procesados son los que se utilizan. Con los ficheros correspondientes a los años de sentencia que tienen los hombres y las mujeres se puede conocer que en promedio de los años de sentencia de las mujeres es de **16.52** y el de los hombres de **14.94**.

A continuación se evidencian los porcentajes de los delitos cometidos por los internos. Los promedios sacados antes de limpiar los datos no son agradables de revisar ya que la respuesta a el tipo de delito cometido por el interno es una respuesta a mano, es necesario limpiar la información y unificar los delitos más generales como: delitos de homicidios, de robos, de violaciones, de violencia familiar, de drogas, de salud, de posesión de armas, de secuestro u otros.

---

```
Robo
10.23469387755102
Homicidio
22.035714285714285
Violacion
19.814814814814813
Violencia Familiar
7.25
Armas y Drogas
11.0
Armas
11.384615384615385
Secuestro
32.473684210526315
Accidente
6.0
Chantaje
9.0
Federal
```

---

<sup>3</sup>Técnicas de Clustering: [www.cs.us.es/~fran/curso\\_unia/clustering.html](http://www.cs.us.es/~fran/curso_unia/clustering.html)

12.5  
Privacion de la libertad  
32.4  
Narcomenudeo  
8.0  
Lesiones  
5.5  
Delicuencia organizada  
8.5  
Empapelado  
14.0  
Drogas  
8.75  
Dano a instituciones publicas  
8.0  
Fraude  
9.0  
Contra la salud  
11.0  
Trata de personas  
12.0

Para visualizar la información a analizar se representaron los datos referentes a los tipos de delitos y a los años de sentencia impuestos a los internos. Luego se crea la gráfica de caja-bigote con el código siguiente que se puede correr en Jupyter, dando como resultado el gráfico 1:

```
import plotly as plotly
import plotly.plotly as py
import plotly.graph_objs as go
import pandas as pd
import ssl
plotly.tools.set_credentials_file(username='evelyn',
    api_key='2e7wTiiQwgd8kwBZbDOE')
if getattr(ssl, '_create_unverified_context', None):
    ssl._create_default_https_context =
        ssl._create_unverified_context
d = pd.read_csv("practica3ok.csv")
datos = []
for sem in d.crime.unique():
    if sem != '':
        datos.append(go.Box(y = d.loc[d['crime'] ==
            sem].sentenceYears, name = sem))
py.iplot(datos, filename='Crimen_Años de sentencia')
```

Del gráfico anterior se puede observar que los años de sentencia mayores son en el delito de Secuestro, y privados a la libertad, cosa que es muy rara y puede ser que la respuesta que hayan dado los internos a esta pregunta no sea correcta, ya que el valor de años de sentencia para estos homicidios es de 90 o 92 años. Seguido se representa información referente a la edad de los internos y al delito cometido donde se puede apreciar que personas de mayores edades cometen delitos del tipo violación y fraude (ver gráfico 2).

Otro gráfico importante fue el de nivel educacional con respecto al crimen cometido en el gráfico 3, en este caso se limpian un poco más los datos de nivel de educación ya que provienen de respuestas escritas a mano y se cometen muchos errores en la digitalización de estas respuestas.

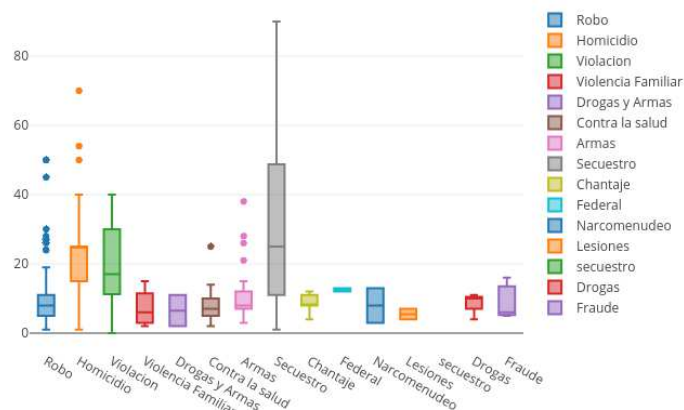


Figura 1: Crimen vs Años de sentencia.

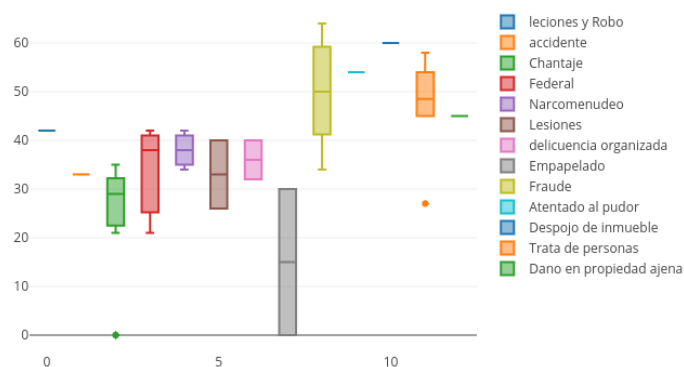


Figura 2: Crimen vs Edad.

Se realizó otro gráfico con el nivel de educación y los años de sentencia de los internos (ver gráfico 4). Se puede ver por ejemplo que los internos que respondieron a **ningún estudio** cometen el delito contra la salud más frecuentemente y la media de casi todos los internos cometen homicidios, independientemente del nivel educacional que tengan.

Se realizan además modelos lineales separando los datos en dos grupos, mujeres y hombre, es decir, los modelos lineales se realizan primero para los datos de los hombres y luego de las mujeres. El objetivo de esto consiste en encontrar aspectos significativos que condicionen la necesidad de las personas de cometer delitos. Es necesario filtrar los años de sentencia de los los internos hombres con edades mayores a 15 años para conocer si se comportan con una distribución normal. Luego de verificar esto se comienzan a modelar los años de sentencia como una función de la edad de los internos hombres, y se obtienen



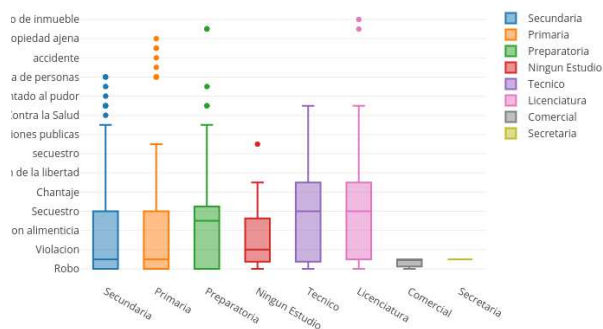


Figura 3: Educación vs Crimen.

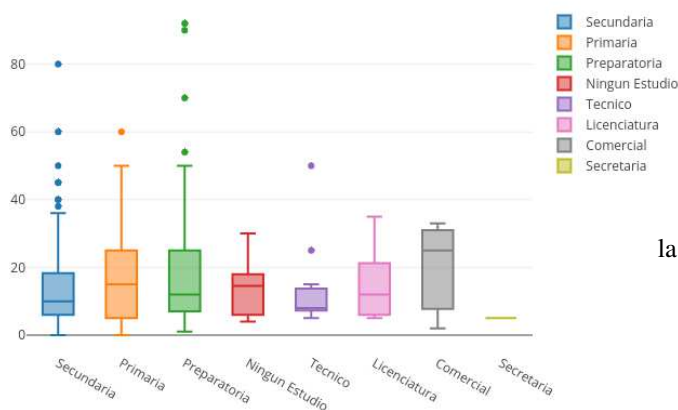


Figura 4: Educación vs Sentencia.

los siguientes valores y se representa la gráfica 5.

```
y = f(x) = 0.0817 x + 12.0093
error 0.06292450975594598
valor p 0.19544903608298309
pendiente no significativo
R^2 0.006944806502868976
```

En el caso anterior se obtiene una pendiente no significativa, el error es pequeño y se nota que van aumentando los años de sentencia con la edad, o que la edad de los hombres condiciona de algún modo la cantidad de años que les impongan. A medida que van creciendo los años de edad, va creciendo levemente los años de sentencia.

Otros datos de interés para trabajar son los ingresos mensuales y los años de sentencia de los internos hombres, dado que puede ser que tenga valor significativo el hecho de tener menos dinero a cometer delitos más o menos graves y por consecuencia mayor o menor cantidad de años. Los ingresos mensuales antes de entrar a la prisión se registran en la encuesta de

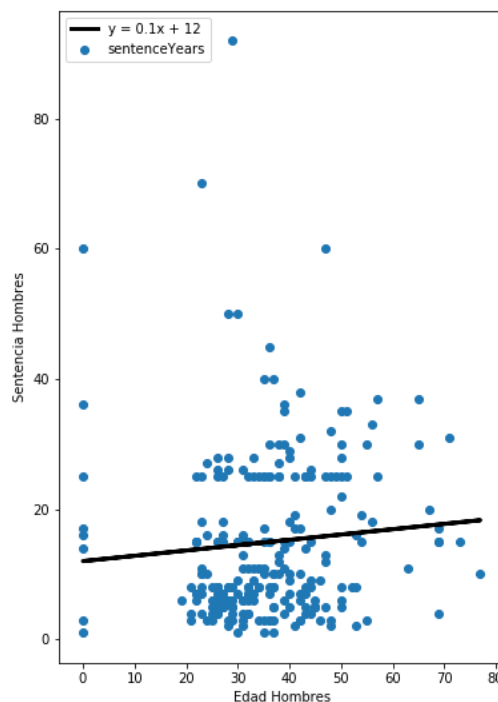


Figura 5: Sentencia vs Edad internos hombres.

la siguiente forma:

- Menos de 3,000
- De 3,001 a 6,000
- De 6,001 a 9,000
- Más de 9,000

Se intenta modelar los años de sentencia en función de la cantidad de ingresos que tenían los internos hombres mensualmente (ver diagrama 6).

```
y = f(x) = -1.6490 x + 18.0927
error 0.7573076672582115
valor p 0.030488607756852505
pendiente significativo
R^2 0.020637081133893283
```

Se obtiene una pendiente significativa, se nota que la pendiente baja a medida que más ingresos mensuales tenga el interno, parece que mientras más ingresos mensuales tengan los hombres, el delito a cometer es más leve, o pudiera decirse que los años de sentencia de los hombres se pueden estimar como -1,6 veces del ingreso mensual más 18, aproximadamente.

Para el caso de las internas mujeres sucede que la distribución de sus datos no están nada normales, dando como resultados no significativos al realizar el mismo procesamiento realizado a los datos de internos hombres.

En trabajos con respuestas a mano dadas por los internos en las encuestas, que tienen que ver con la Reinserción Social,

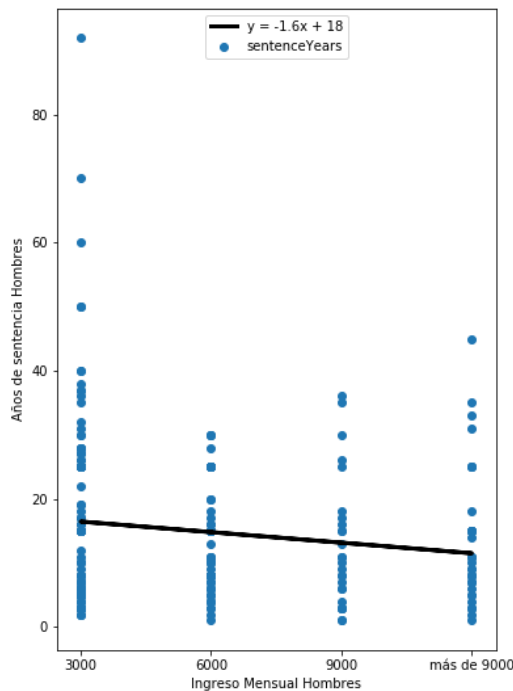


Figura 6: Sentencia vs Ingresos mensuales internos hombres.

se estudiaron las preguntas de la encuesta que se muestran a continuación ya que forman parte de un tema muy importante para los internos en el momento de obtener la libertad y retomar sus vidas.

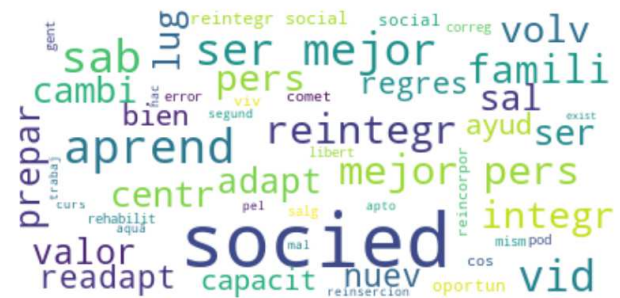
- Conocimientos sobre reinserción social
- Consejos que darían a los jóvenes para evitar cometer un delito
- Planes futuros al salir de la prisión
- Conocimientos sobre organizaciones de apoyo como enlace laboral (conseguir empleo)

Los resultados obtenidos se muestran en cuadros de imágenes por cada una de las preguntas de interés antes mencionadas en la figura 7.

Para el caso de las respuestas de que conocen por **Reinserción Social**, se ve en la imagen anterior que muchos hablan de **readaptarse, familia, valores, integridad, sociedad, aprender, ser mejor, vida, mejor persona, ayuda, regresar**, esto al parecer indica que tienen conocimientos sobre este tema.

Los internos **aconsejan a los jóvenes: estudiar, evitar, mala compañía, valores, familia, amistad** entre otros, donde se ve que la mayoría piensa en la amistad, las malas compañías y estudiar, con esto se puede valorar la posibilidad de ser estas las causas principales que pueden conllevar a cometer un delito.

Las respuestas de los internos a los **planes futuros** luego de quedar en libertad giran alrededor de **trabajar, negocios, familia, hijos, hogar** destacando la necesidad de los internos de volver a recuperar su trabajo y hogar.



(a) Reinserción Social



(b) Consejos a los jóvenes



(c) Planes futuros



(d) Organizaciones de apoyo

Figura 7: Cuadros de textos que representan la información de datos escritos con lenguaje natural



Se evidencia que conocen algunas **organizaciones de apoyo** donde la mayoría responden **Sedesol, Rece, Hyrse, Jyrsa**, donde la más frecuente respuesta fue **Sedesol**.

## 6. Conclusiones

La ciencia de los datos sigue siendo un campo emergente dentro de las empresas ya que la identificación y el análisis de grandes cantidades de datos no estructurados puede resultar demasiado complejo, costoso y lento para las empresas.

Gracias al trabajo con las herramientas `bash` se pudo limpiar un poco más los datos, ya que al unificar respuestas con el `uniq -c`, se descubrieron errores como por ejemplo: quedaban algunos NO CONTESTO y se cambiaron por NA.

Se puede decir según los resultados anteriores que edad de los internos y los años de sentencia tienen una correlación positiva, lo cual podría decir que influye en la edad de la persona el delito que valla a cometer y por esto le darían más o menos años de sentencia, pero se necesitaría hacer pruebas de hipótesis para concluir esto.

Se puede plantear que el nivel escolar de la persona puede tener relación con el delito a cometer, ya que también tienen un coeficiente de correlación positivo. Sin embargo, el nivel escolar va en dirección contraria con respecto a los años de sentencia, ya que es negativo el valor de correlación, lo cual se podría decir que a menor nivel escolar mayor cantidad de años de sentencia, lo cual comprueba que mientras más bajo sea la educación de la persona, peores y mayor cantidad de delitos puede cometer.

Con respecto al delito cometido y la edad de la persona también da un valor de correlación positivo, se pudiera decir que la edad condiciona de alguna manera el delito a cometer.

El uso de los diagramas facilita la interpretación de la información, la relaciona y permite llegar a conclusiones, en el caso de los datos que se están analizando, por medio de los diagramas también se pudo ver la relación existente entre las columnas seleccionadas para el análisis de esta información.

Los modelos lineales realizados en el trabajo no fueron muy buenos, ya que los datos no se encuentran normalmente distribuidos y esto no ayuda los modelos. A pesar de esto se pudo apreciar que en el caso de los hombres puede existir una relación de dependencia de los años de sentencia con respecto a la edad de los internos, y un poco más con la cantidad de dinero que tenían como ingreso mensual, cosa que en el caso de las mujeres no se comportó de esta manera, por lo cual valdría la pena seguir revisando que factores condicionan el hecho de que las mujeres cometan ciertos delitos.

## Referencias

- [1] Contreras Miguel Angel. 10 temas de derechos humanos. <http://www.codhem.org.mx/LocalUser/codhem.org/difus/libros/10temas.pdf>.
- [2] Marcelo Bergman. Cárceles en México: cuadros de una crisis. *URVIO: Revista Latinoamericana de Estudios de Seguridad*, (1):74–87, 2007.
- [3] Constitución Política de los Estados Unidos Mexicanos. Constitución política de los estados unidos mexicanos, 2017.
- [4] Nallely Arias García and Feggy Ostrosky Solís. Evaluación neuropsicológica en internos penitenciarios mexicanos. *Revista chilena de Neuropsicología*, 5(2):113–127, 2010.
- [5] Morales Julieta González Luis. Los derechos humanos de las personas privadas de la libertad: fundamento de la ejecución de la pena y la reinserción social. <https://archivos.juridicas.unam.mx/www/bjv/libros/7/3104/22.pdf>, 2011.
- [6] Jacobson Jessica. Catherine heard y helen fair, prison: Evidence of its use and over-use from around the world. [http://www.prisonstudies.org/sites/default/files/resources/downloads/global\\_imprisonment\\_web2c.pdf](http://www.prisonstudies.org/sites/default/files/resources/downloads/global_imprisonment_web2c.pdf), 2017.
- [7] Gary King, James Honaker, Anne Joseph, and Kenneth Scheve. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American political science review*, 95(1):49–69, 2001.
- [8] Cobo Sofía M. Derecho de ejecución de la pena. <http://www.inacipe.gob.mx/stories/investigacion/descargas/CAP%C3%8DTULO%2016.%20MANUAL%20Reforma%20Enero2014.pdf>, 2014.
- [9] Aguilar Luis María. Reforma constitucional en materia penal de 2008. antecedentes, objetivos y ejes rectores. page 33, 2016.
- [10] Fabián A Quintero and Servicio Penitenciario Bonaerense. Características poblacionales del encarcelamiento en la provincia de buenos aires en contexto global. *Revista electrónica de la Especialización en Derecho Penal de la Universidad Nacional de La Plata*, (13), 2008.
- [11] James O Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

## 7. Agradecimientos

Es un honor para mí agradecer a las personas e instituciones que jugaron un papel importante durante la elaboración de este artículo, agradezco a:

- A mis profesores, en especial a la Dr. Elisa por impartir la asignatura relacionada con el tema y brindarme los datos suficientes para realizar este estudio.
- A la facultad y a CONACyT por darme el beneficio de una beca que me permitió estudiar en este país.