

# Población Penitenciaria de Argentina

*Censos 2002 - 2017*

Nahuel Almeira - Milagros Ávila - Lucía Pappaterra



Ministerio de  
Justicia y Derechos Humanos  
Presidencia de la Nación

Dirección Nacional de Política Criminal en materia de Justicia y Legislación Penal

# SNEEP

Sistema Nacional de Estadística sobre Ejecución de la Pena

# Variables

**50 variables**

*La mayoría son  
Categóricas!!!!*

**Datos demográficos previos:**

edad - género - nacionalidad - ...

**Datos de la vida cotidiana:**

trabaja - estudia - hace deporte - ...

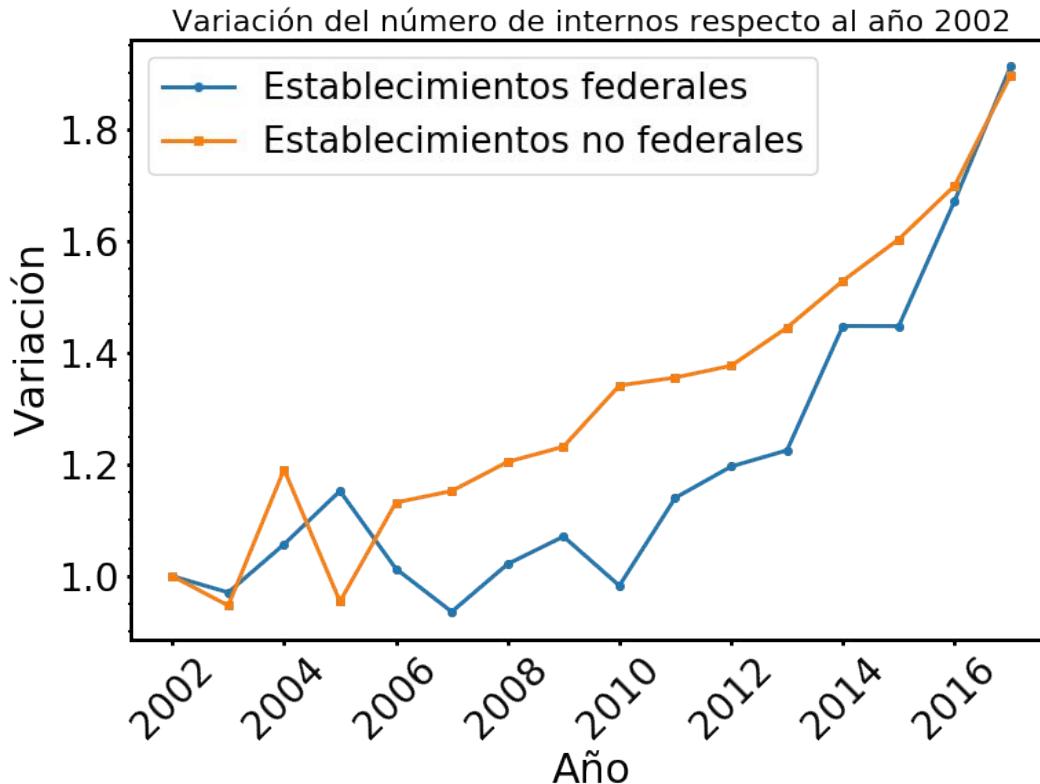
**Situación judicial del interno:**

fecha detención - fecha condena - delito - ...

# Visualización y Análisis

# Evolución temporal

## Cantidad de internos



2002: 45000

...

2017: 85000

Casi el  
doble!!!

Censo 2001: 36 millones

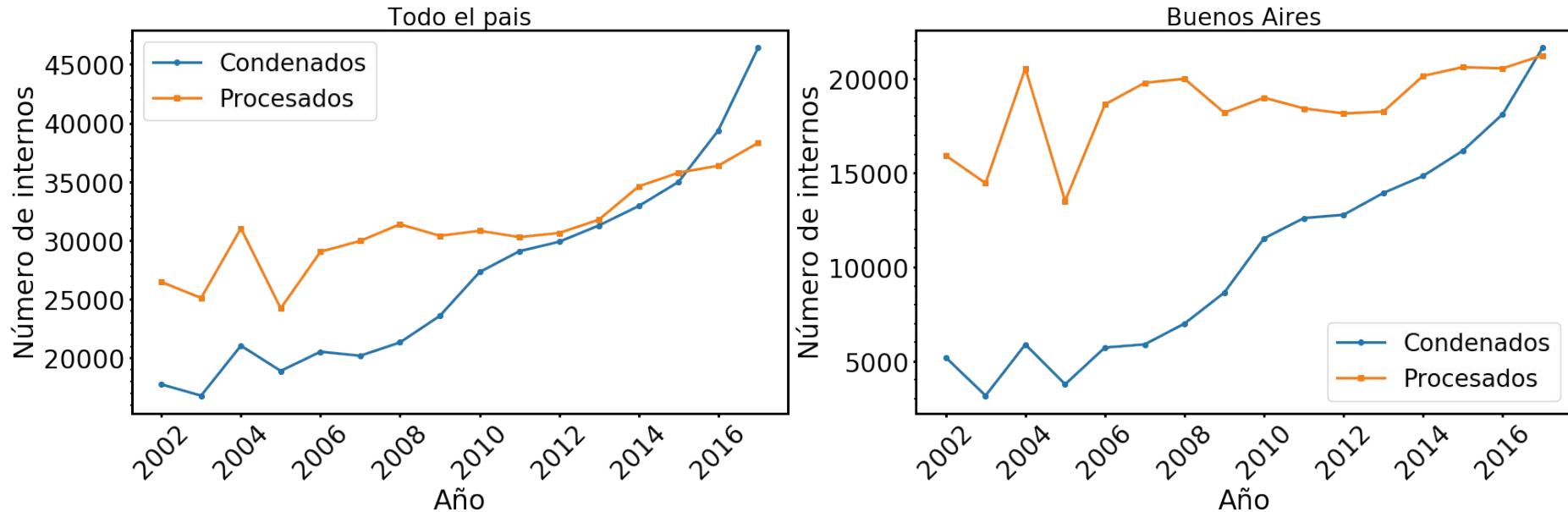
...

Censo 2010: 40 millones

Aumentó  
un 10%

# Evolución temporal

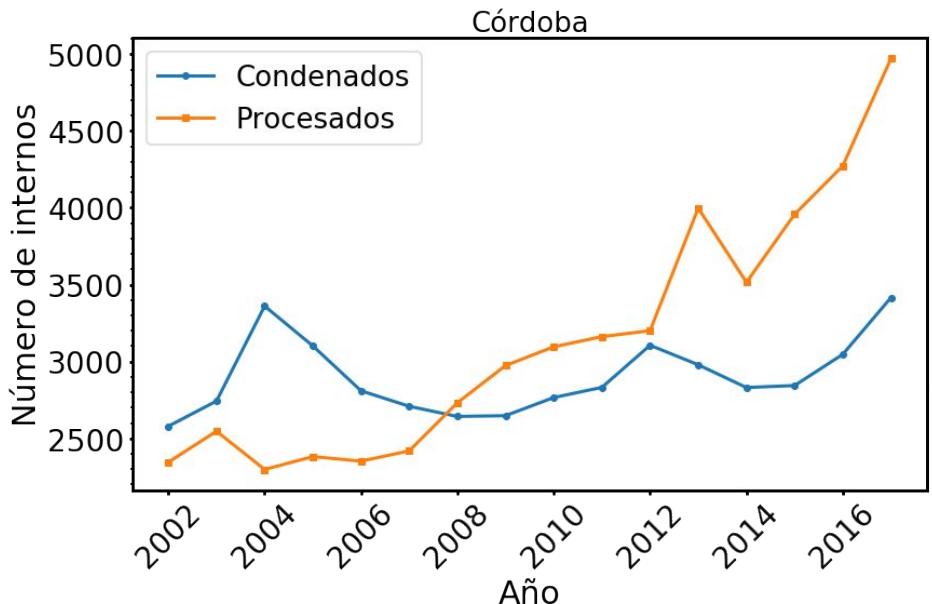
## Situación Legal



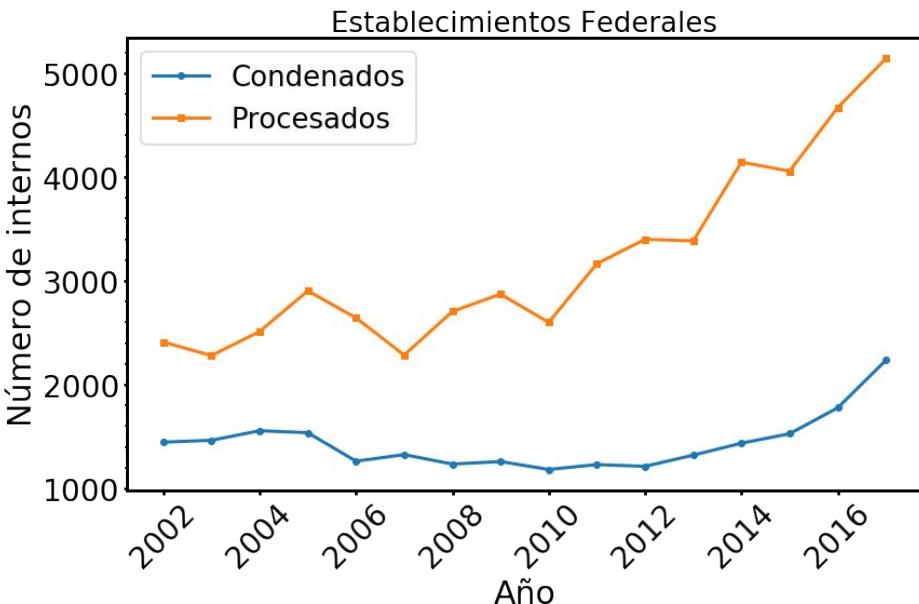
51% de la población penal  
(censo 2017)

# Evolución temporal

## Situación Legal



10% de la población penal  
(censo 2017)



9% de la población penal  
(censo 2017)

# Población Penitenciaria de Argentina

## Nacionalidad

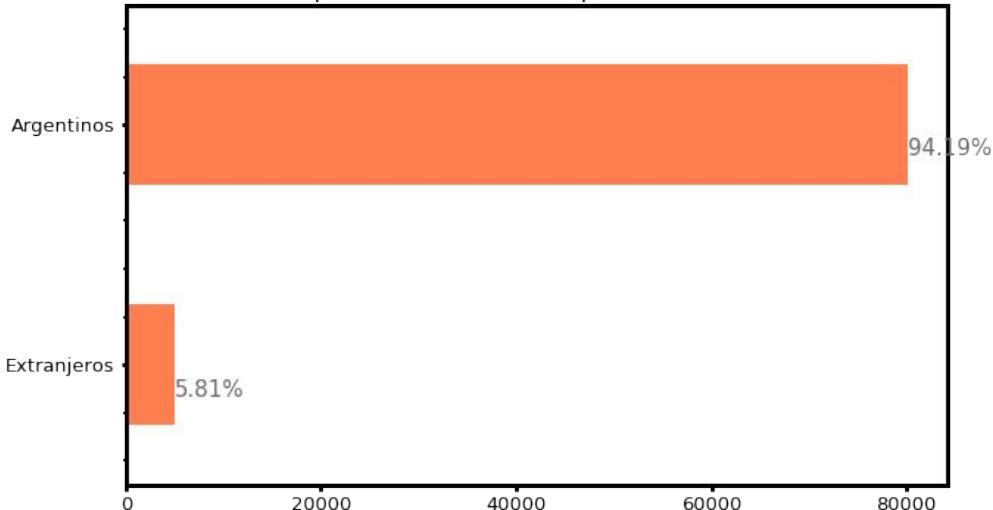
“El 20% de los  
presos en  
cárcel  
argentinas son  
extranjeros”



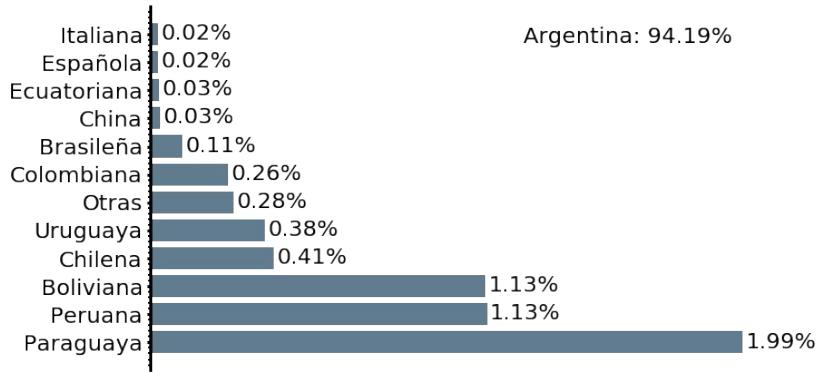
# Población Penitenciaria de Argentina 2017

## Nacionalidad

Proporción de internos por nacionalidad

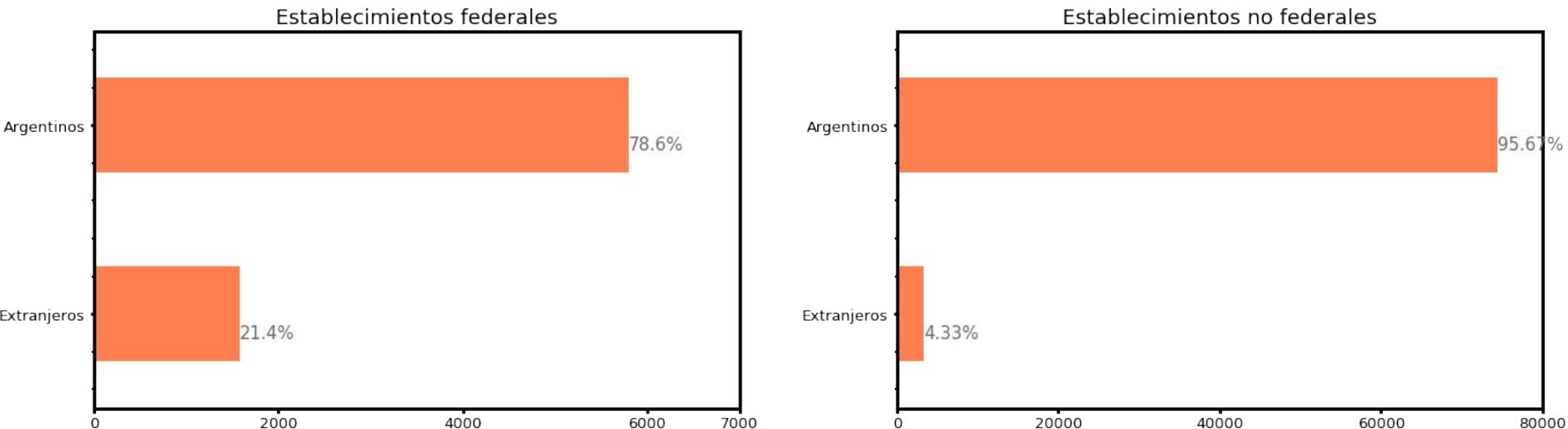


Todos los establecimientos



# Población Penitenciaria de Argentina 2017

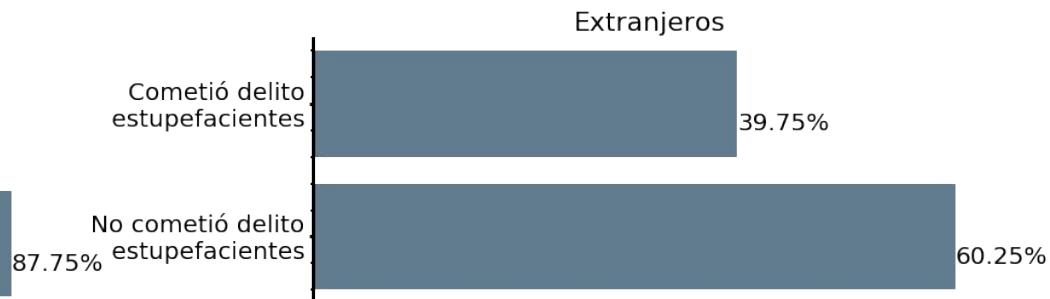
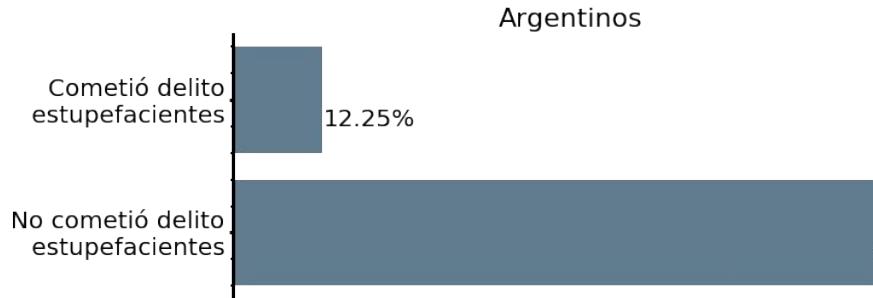
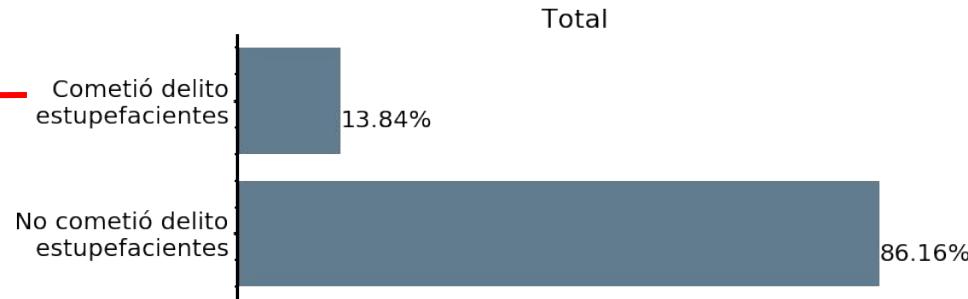
## Nacionalidad | Establecimiento



9% de la población penal  
(censo 2017)

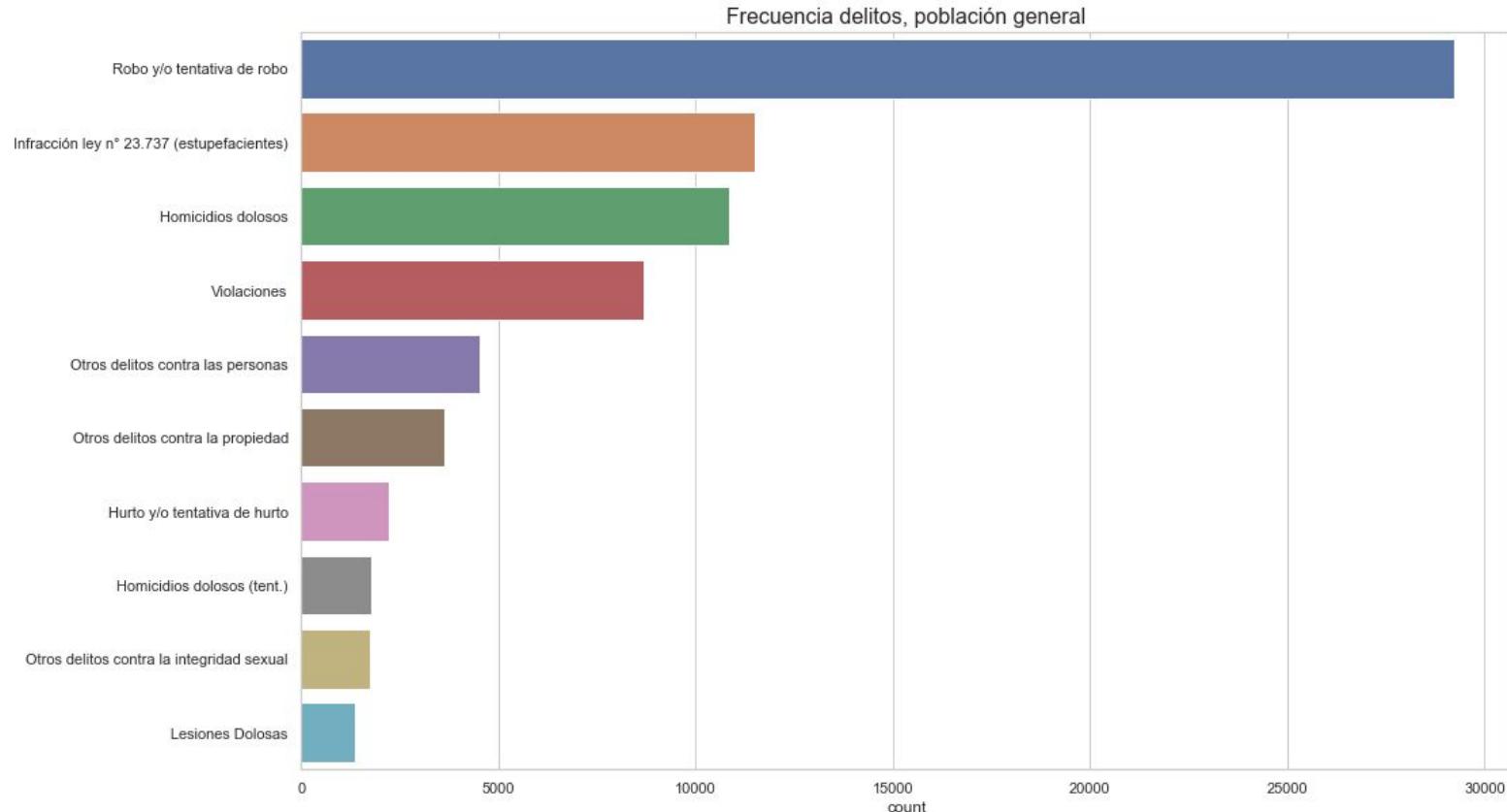
# Población Penitenciaria de Argentina 2017

## Delito | Nacionalidad



# Población Penitenciaria de Argentina 2017

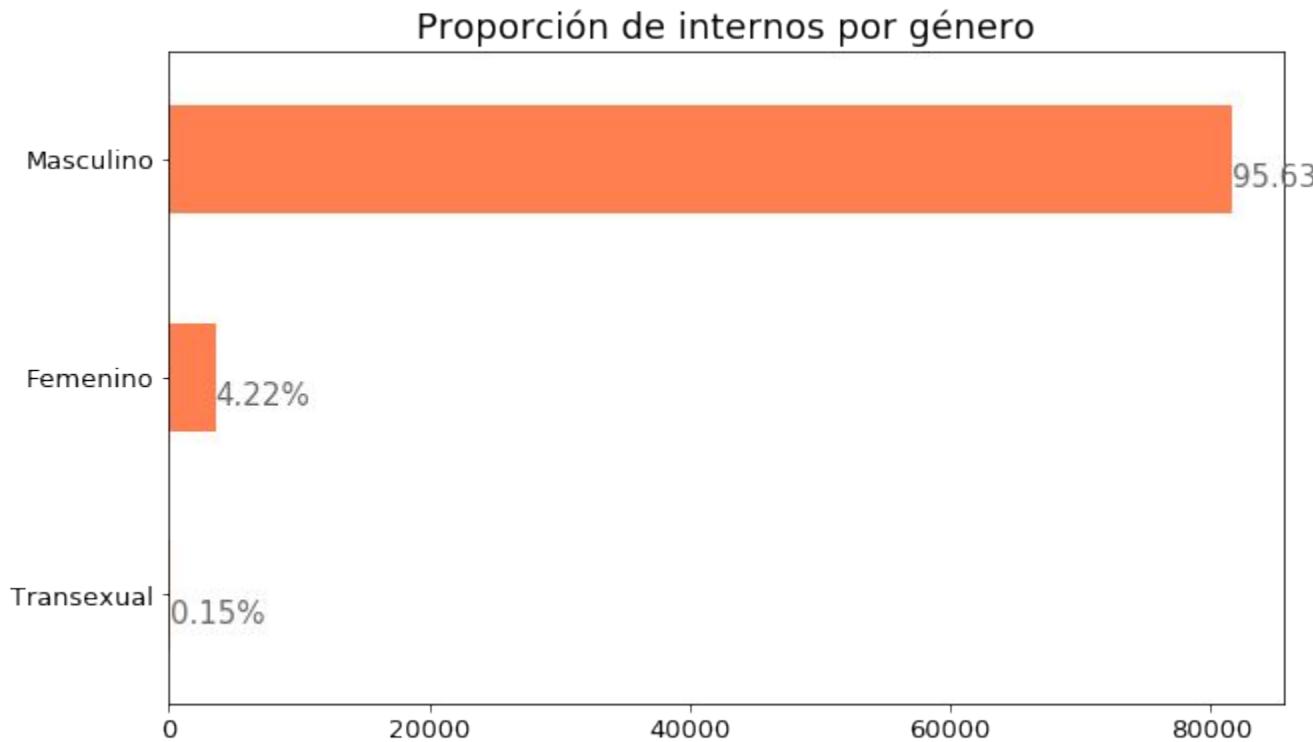
## Delito



# Población Penitenciaria de Argentina 2017

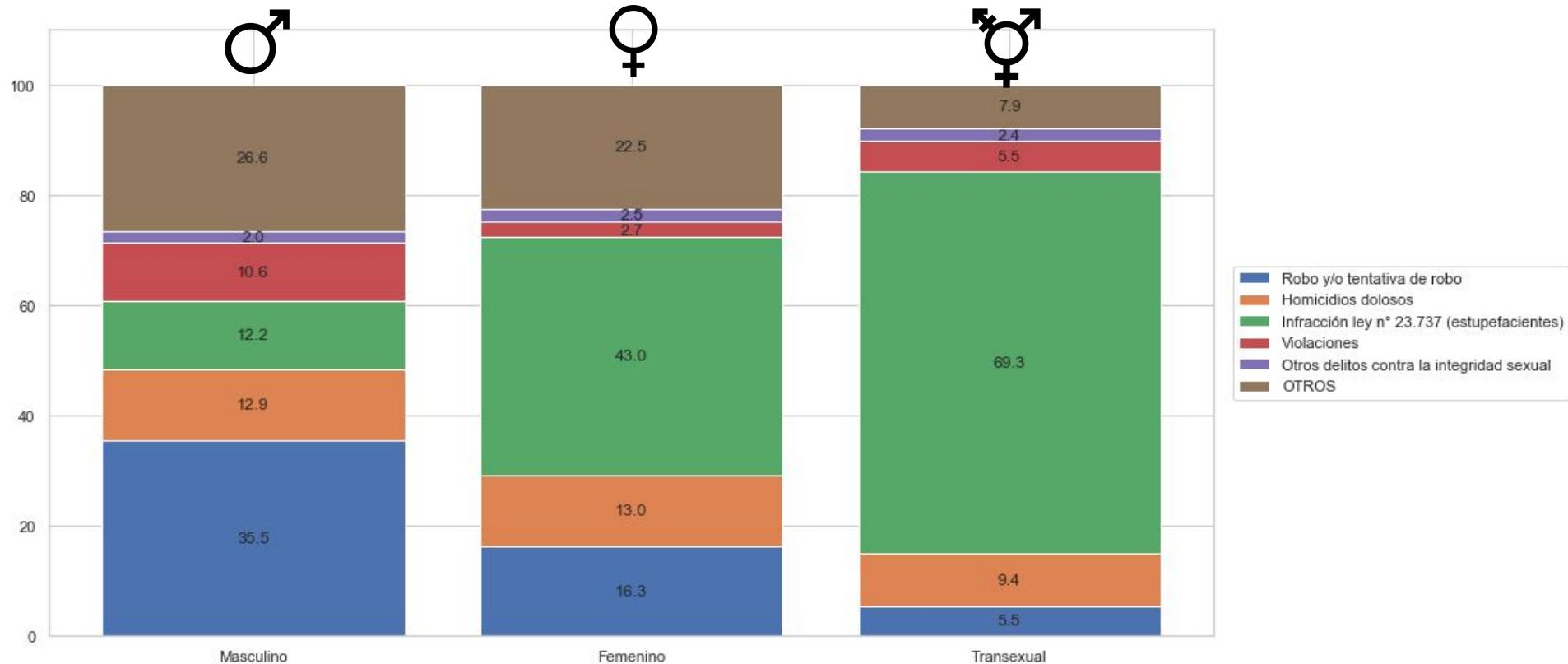
## Género

	counts	percentage
<b>Masculino</b>	81553	95.626326
<b>Femenino</b>	3602	4.223585
<b>Transexual</b>	128	0.150089



# Población Penitenciaria de Argentina 2017

## Delito | Género



# Análisis y Curación de Datos

# Análisis y Curación de Datos

# Análisis y Curación de Datos

**Censo nacional sobre el total de la población detenida** al día 31 de diciembre de cada año y en cada establecimiento.

*Implementado por primera vez en el año 2002*

# Análisis y Curación de Datos

**Censo nacional sobre el total de la población detenida al día 31 de diciembre de cada año y en cada establecimiento.**

*Implementado por primera vez en el año 2002*

**Cantidad de registros: 939727, Features: 50**

Año, Provincia, Establecimiento, Edad, Género, Nacionalidad, Estado Civil, Nivel Instrucción, Última situación laboral, Capacitación laboral al ingresar, Último lugar de residencia, Última Provincia de residencia, Jurisdicción, Situación legal, Fecha de detención, Fecha de condena, Establecimiento de procedencia, Delitos (1, 2, 3, 4 y 5),

Horas de trabajo remunerado, Participación de programa laboral, Participación de programa educativo, Participación de actividades deportivas, Recibió atención médica en el último año, Recibió visitas en el último año, Participó de la alteración al orden en el último año, Tipo de infracción disciplinaria, Sanción aplicada, Calificación de conducta, Tentativa de fugas y evasiones, Tentativa de suicidio, Fue lesionado, Duración de condena en años, Duración de condena en meses, Rango de la duración de condena, Tipo de condena, Tiene medidas de seguridad, Es reincidente, Tiene período de progresividad, Repartición, Tuvo salidas transitorias, Incorporado por régimen de semi libertad, Participa de programa pre libertad, Participa de programa de prisión discontinua, Participa programa semi detención, Tuvo reducción de pena, Mujer tiene hijos intramuro, Provincia INDEC, Última provincida de residencia INDEC

# Análisis y Curación de Datos

**Censo nacional sobre el total de la población detenida al día 31 de diciembre de cada año y en cada establecimiento.**

*Implementado por primera vez en el año 2002*

**Cantidad de registros: 939727, Features: 50**

Año, Provincia, Establecimiento, Edad, Género, Nacionalidad, Estado Civil, Nivel Instrucción, Última situación laboral, Capacitación laboral al ingresar, Último lugar de residencia, Última Provincia de residencia, Jurisdicción, Situación legal, Fecha de detención, Fecha de condena, Establecimiento de procedencia, Delitos (1, 2, 3, 4 y 5),

Horas de trabajo remunerado, Participación de programa laboral, Participación de programa educativo, Participación de actividades deportivas, Recibió atención médica en el último año, Recibió visitas en el último año, Participó de la alteración al orden en el último año, Tipo de infracción disciplinaria, Sanción aplicada, Calificación de conducta, Tentativa de fugas y evasiones, Tentativa de suicidio, Fue lesionado, Duración de condena en años, Duración de condena en meses, Rango de la duración de condena, Tipo de condena, Tiene medidas de seguridad, Es reincidente, Tiene período de progresividad, Repartición, Tuvo salidas transitorias, Incorporado por régimen de semi libertad, Participa de programa pre libertad, Participa de programa de prisión discontinua, Participa programa semi detención, Tuvo reducción de pena, Mujer tiene hijos intramuro, Provincia INDEC, Última provincia de residencia INDEC

Informe “Una Gestión Penitenciaria Integral”

## SNEEP 1

### CENSO EN ESTABLECIMIENTOS DE EJECUCIÓN DE PENAS PRIVATIVAS DE LA LIBERTAD

Sistema Nacional de Estadística sobre Ejecución de la Pena (SNEEP) - Ministerio de Justicia, Seguridad y Derechos Humanos de la Nación

Provincia:	Repartición:	Nombre del Establecimiento:	Tipo de Establecimiento: (Art.176, Ley N° 24.660)	Año:
Código Identificatorio o Legajo	Edad (Indicar años)	Sexo (M o F)		
(1) Nacionalidad	(2) Estado Civil	(3) Nivel de instrucción (4) Situación Laboral al momento del ingreso (5) Capacitación Laboral al momento del ingreso (6) Lugar de Residencia al momento del ingreso (7) Provincia de Residencia al momento del ingreso (8) Jurisdicción Judicial	(9) Situación Legal	
		FECHA DE DETENCIÓN (fecha de privación de libertad) Indicar dia/mes/año		
		FECHA DE CONDENA (para condenados) Indicar dia/mes/año		
		(10) Establecimiento de Procedencia		
		(11) Tipo de Delito por el cual ingresó en el establecimiento (indicar hasta cinco delitos de acuerdo con los códigos del instructivo)		
		(12) Trabajo Remunerado según cantidad de horas a la fecha de (13) ¿Participó en un Programa de Form. o Capacit. laboral durante el último año?		
		(14) ¿Participó regularmente de algún Programa Educativo durante el último año?		
		(15) ¿Participó regularmente de Actividades Recreativas y Deportivas en el último año?		
		(16) ¿Recibió Asistencia Médica durante el último año?		
		(17) ¿Recibió visitas durante el último año?		
		(18) Partició de alteraciones del orden en el último año (tipo de alteración)		
		(19) Tipo de Infracción Disciplinaria		
		(20) Sanciones Disciplinarias impuestas al interno		
		(21) Última Clasificación de conducta		
		(22) Tentativa de Fugas y Evasiones en el último año		
		(23) Intentos de Suicidio en el último año		
		(24) ¿Fue lesionado en el establecimiento, en el último año?		
		DURACIÓN DE LA CONDENA (indicar años y meses)		
		(25) ¿Está condenado a prisión oclusión perpetua?		
		(26) ¿Tiene medida de seguridad (art. 52 CP)?		
		(27) Reincidencia		
		(28) Período de Progresividad del Régimen Penitenciario		
		(29) ¿Tiene o tuvo Salidas Transitorias en el último año?		
		(30) ¿Fue incorporado al Régimen de Semilibertad?		
		(31) ¿Participa en un programa de Prelibertad?		
		(32) Prisión discontinua		
		(33) Semilibertad		
		(34) ¿Tuvo reducción de pena en el último año?		
		(35) Para Mujeres: ¿Tiene hijos alojados con ella?		

# Análisis y Curación de Datos

# Análisis y Curación de Datos

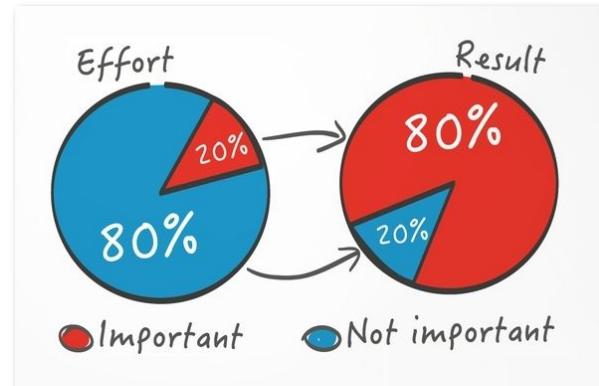
Informe “Una Gestión Penitenciaria Integral”

Es importante aclarar que la aplicación del sistema tiene sus limitaciones. En primer lugar, la falta de recursos y de personal capacitado, conjuntamente con una poca tradición en el manejo y producción de información por parte de las instituciones penitenciarias, complica la tarea de recolección de datos. Esto provoca, entre otras complicaciones, el retraso en el envío de datos o inconsistencias en la información remitida en algunos casos.

# Análisis y Curación de Datos

Informe "Una Gestión Penitenciaria Integral"

Es importante aclarar que la aplicación del sistema tiene sus limitaciones. En primer lugar, la falta de recursos y de personal capacitado, conjuntamente con una poca tradición en el manejo y producción de información por parte de las instituciones penitenciarias, complica la tarea de recolección de datos. Esto provoca, entre otras complicaciones, el retraso en el envío de datos o inconsistencias en la información remitida en algunos casos.

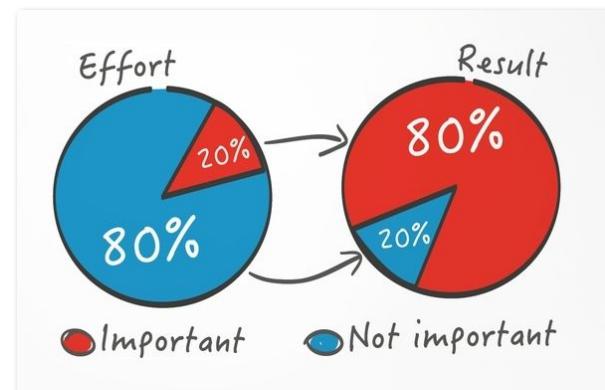


# Análisis y Curación de Datos

Informe "Una Gestión Penitenciaria Integral"

Es importante aclarar que la aplicación del sistema tiene sus limitaciones. En primer lugar, la falta de recursos y de personal capacitado, conjuntamente con una poca tradición en el manejo y producción de información por parte de las instituciones penitenciarias, complica la tarea de recolección de datos. Esto provoca, entre otras complicaciones, el retraso en el envío de datos o inconsistencias en la información remitida en algunos casos.

- Análisis e imputación de valores nulos
- Parseo de fechas
- Análisis y consistencia de datos
  - Edad
  - Mujer tiene hijos/as intramuro
- Generación de nuevas features
  - Edad al ser condenado
  - Cantidad de delitos

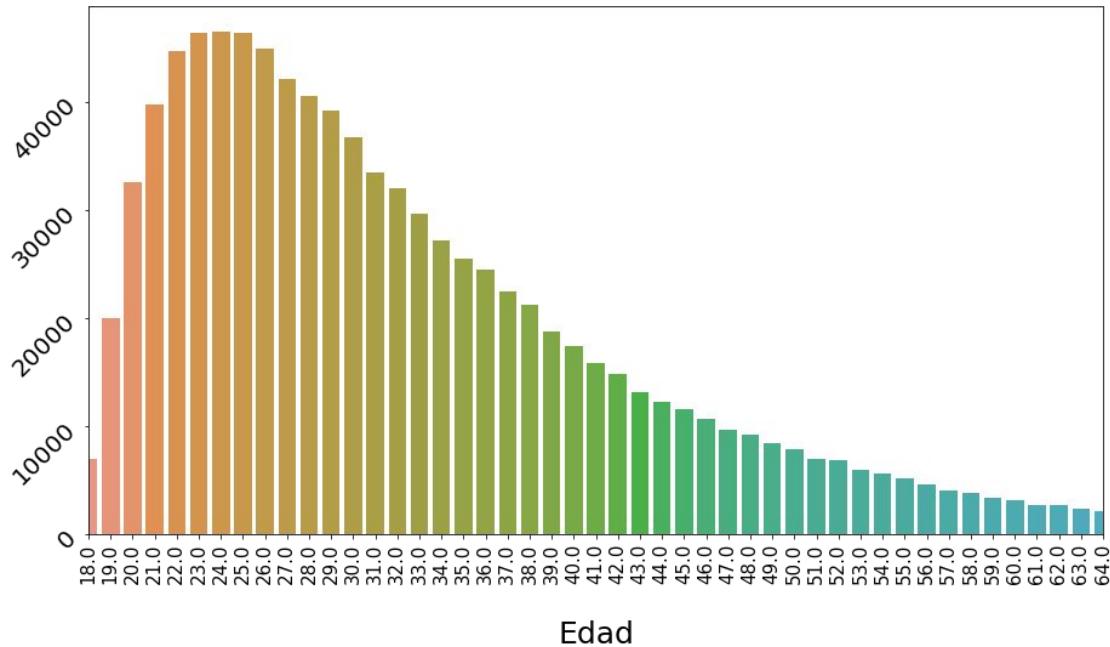


# Análisis y Curación de Datos

## Análisis e Imputación de valores nulos

# Análisis y Curación de Datos

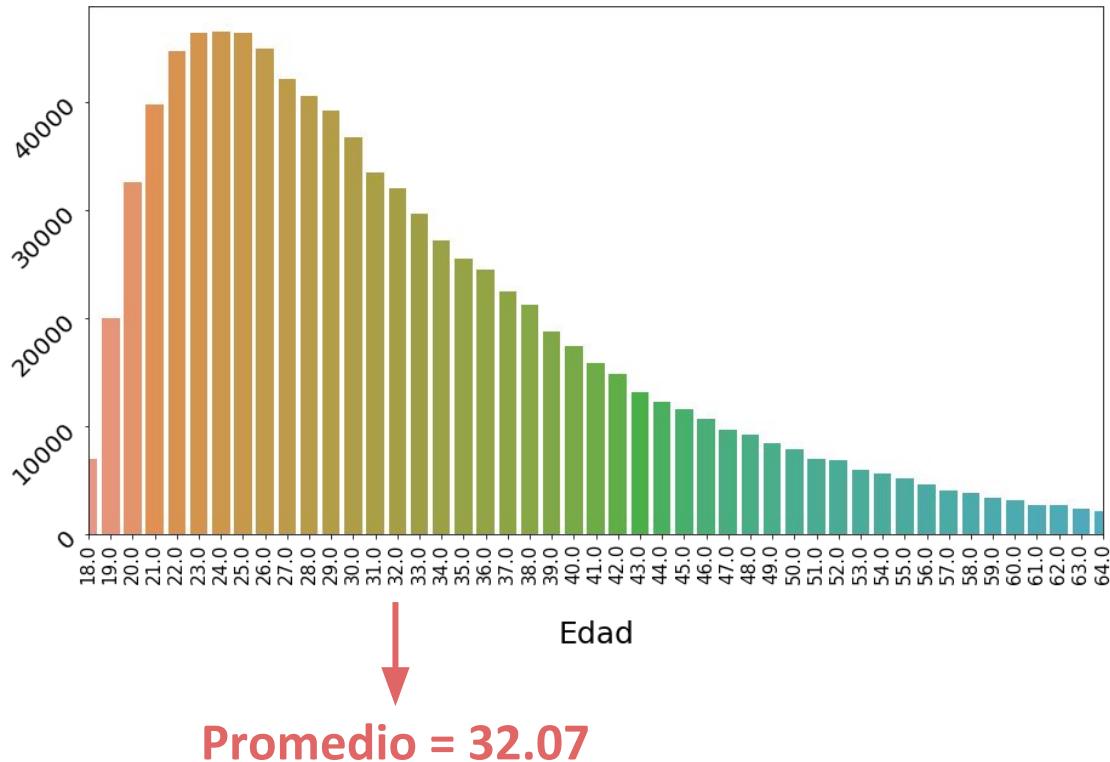
## Análisis e Imputación de valores nulos



# Análisis y Curación de Datos

## Análisis e Imputación de valores nulos

reemplazamos  
valores nulos por  
el valor  
promedio de la  
*feature* 'edad'



# Análisis y Curación de Datos

## Parseo de Fechas

# Análisis y Curación de Datos

## Parseo de Fechas

```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

```
object  
object
```

# Análisis y Curación de Datos

## Parseo de Fechas

```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

object  
object



# Análisis y Curación de Datos

## Parseo de Fechas

```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

object  
object

```
#importamos 'datetime':  
from datetime import datetime
```

```
df.loc[:, 'fecha_detencion'] = pd.to_datetime(df['fecha_detencion'], errors = 'coerce')  
df.loc[:, 'fecha_condenado'] = pd.to_datetime(df['fecha_condenado'], errors = 'coerce')
```

```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

datetime64[ns]  
datetime64[ns]

# Análisis y Curación de Datos

## Parseo de Fechas



```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

object  
object

```
#importamos 'datetime':  
from datetime import datetime
```

```
df.loc[:, 'fecha_detencion'] = pd.to_datetime(df['fecha_detencion'], errors = 'coerce')  
df.loc[:, 'fecha_condenado'] = pd.to_datetime(df['fecha_condenado'], errors = 'coerce')
```

```
# tipo de variables de 'fecha_detencion' y 'fecha_condenado':  
print(df['fecha_detencion'].dtypes)  
print(df['fecha_condenado'].dtypes)
```

datetime64[ns]  
datetime64[ns]

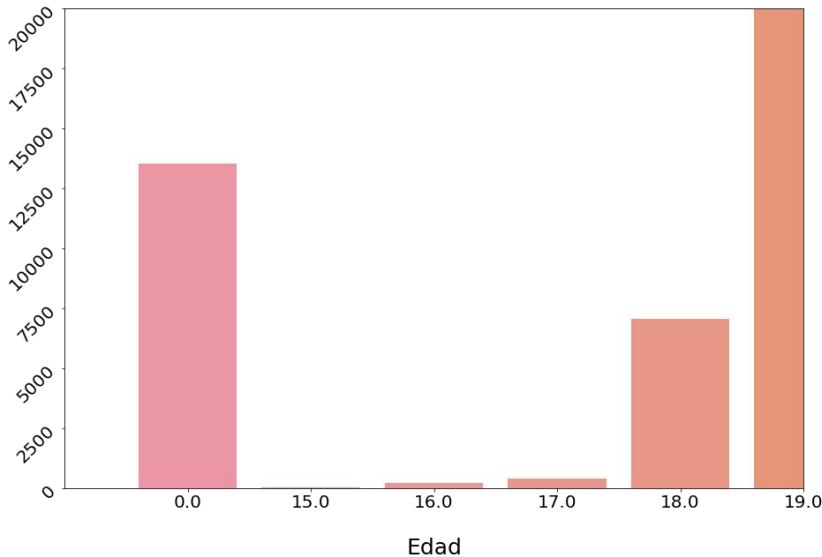


# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

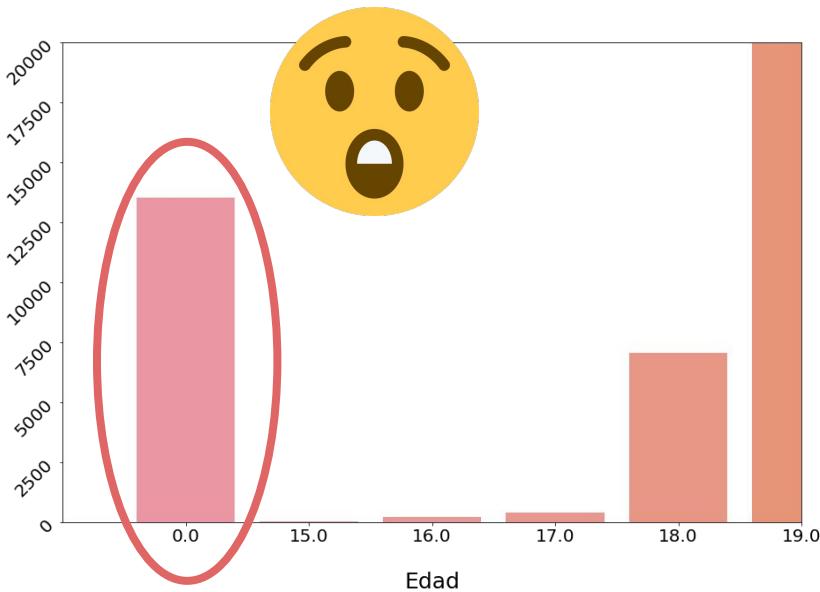
# Análisis y Curación de Datos

## Análisis y Consistencia de Datos



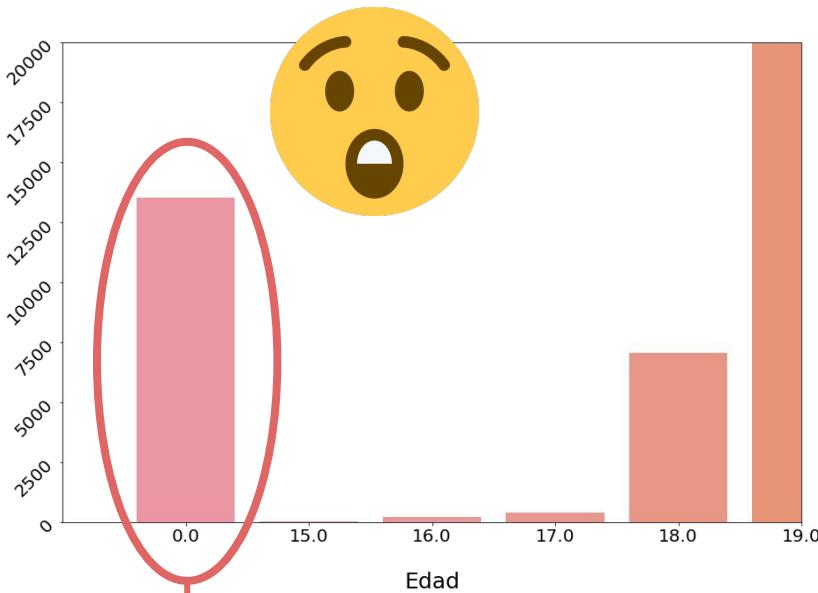
# Análisis y Curación de Datos

## Análisis y Consistencia de Datos



# Análisis y Curación de Datos

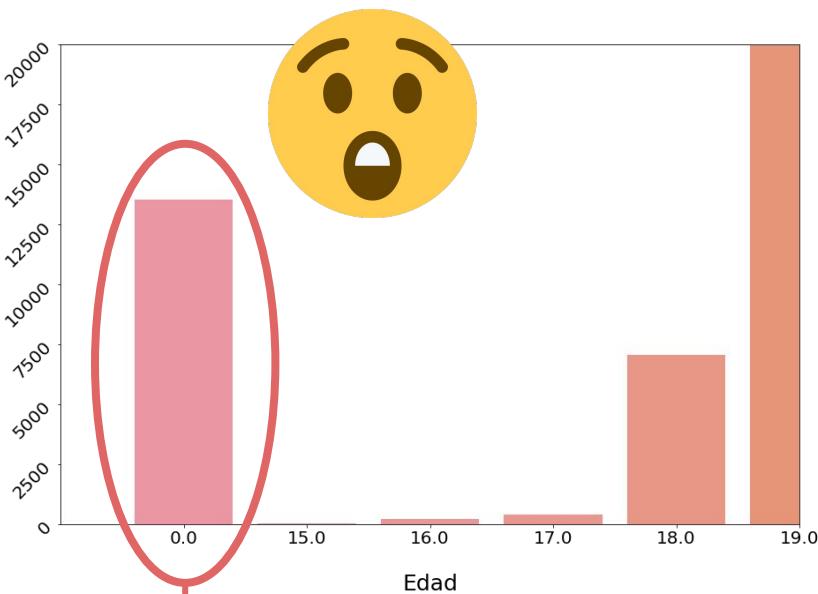
## Análisis y Consistencia de Datos



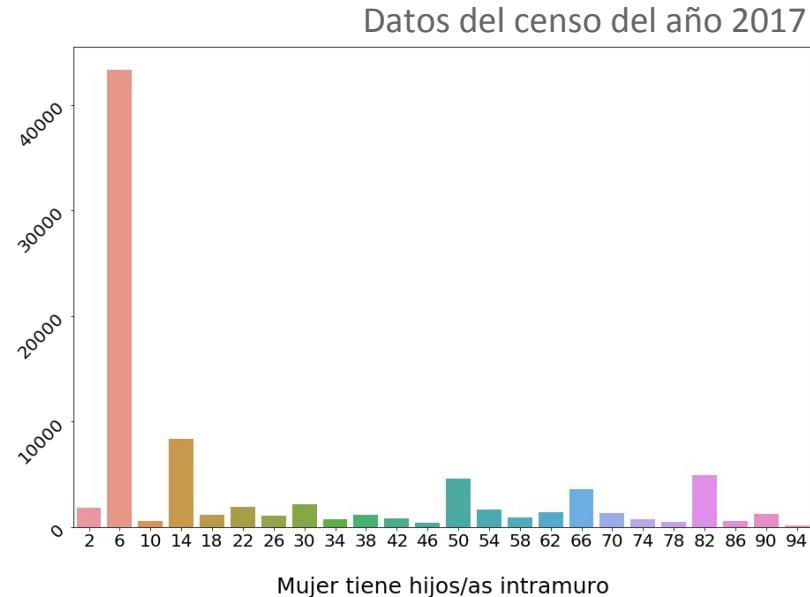
Reemplazamos estos números  
por el valor promedio de la  
feature 'edad'

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

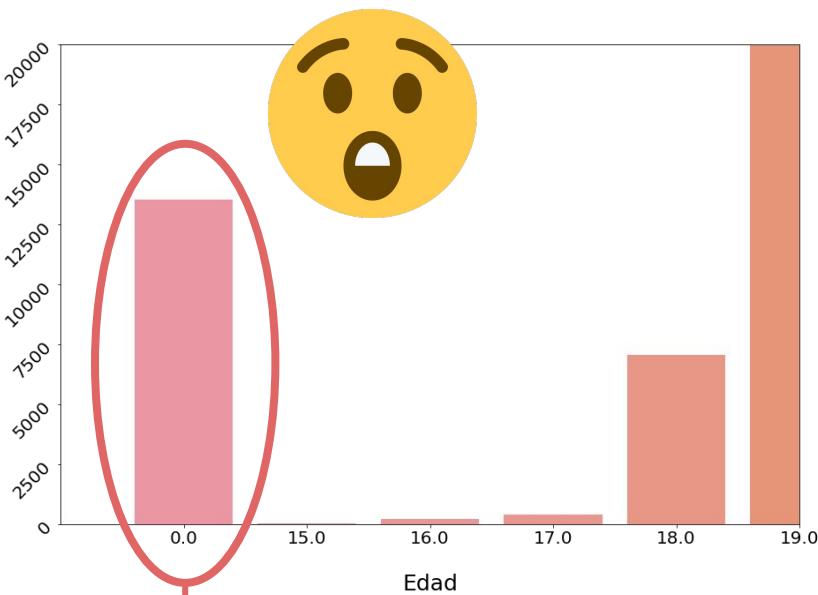


Reemplazamos estos números  
por el valor promedio de la  
feature 'edad'

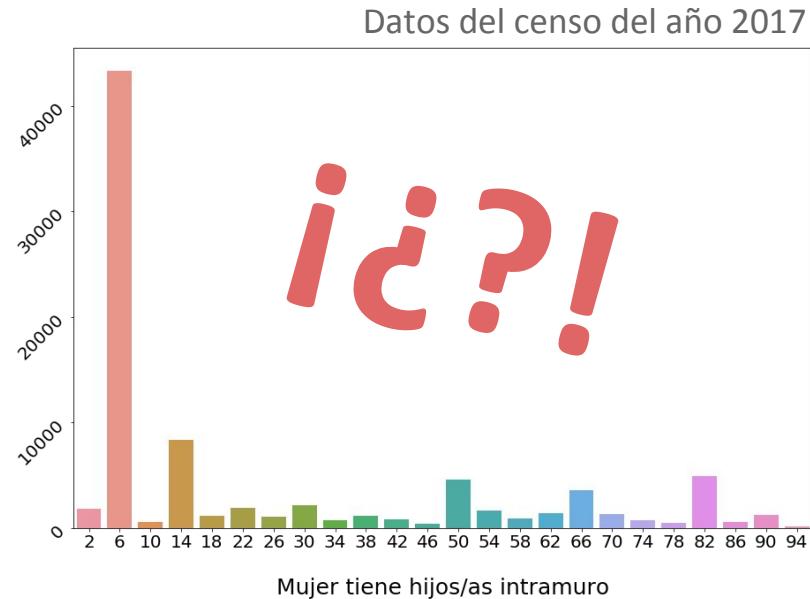


# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

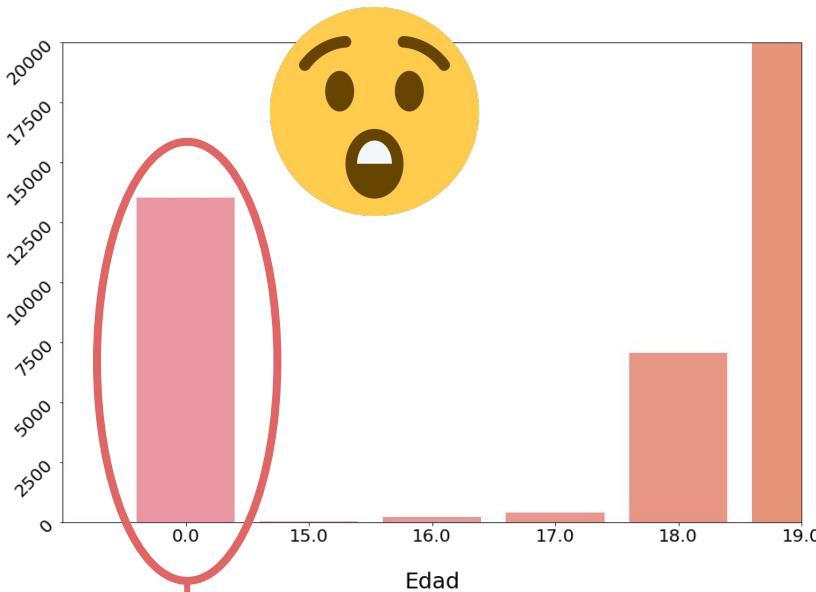


Reemplazamos estos números  
por el valor promedio de la  
feature 'edad'

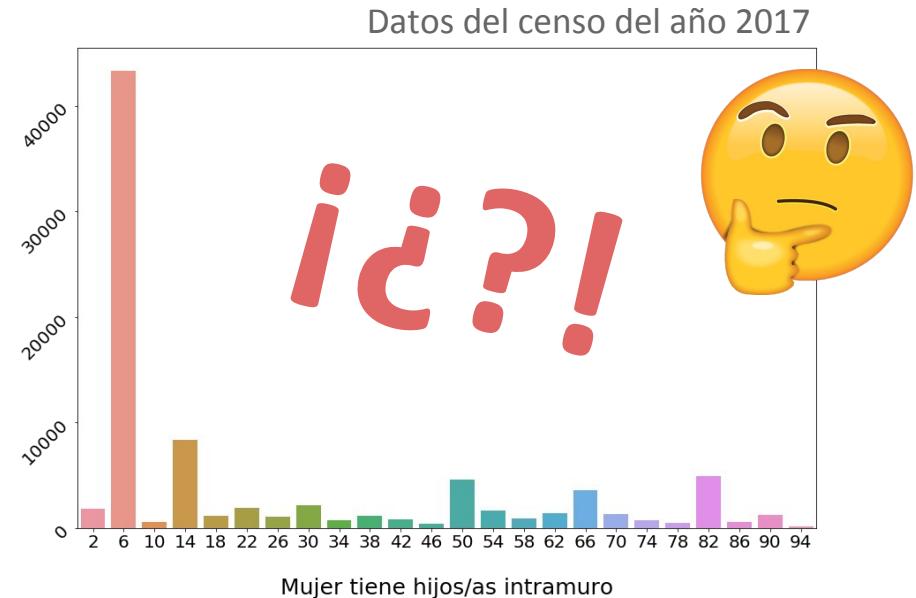


# Análisis y Curación de Datos

## Análisis y Consistencia de Datos



Reemplazamos estos números  
por el valor promedio de la  
feature 'edad'



Vemos en profundidad el  
dataset ...

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64
63 duracion_condena_rango int64
64 tipo_condena float64
65 tiene_medidas_seguridad float64
66 es_reincidente_id object
67 es_reincidente_descripcion float64
68 tiene_periodo_progresividad_id object
69 tiene_periodo_progresividad_descripcion int64
70 reparticion_id object
71 reparticion_descripcion float64
72 tuvo_salidas_transitorias_id object
73 tuvo_salidas_transitorias_descripcion float64
74 incorporado_reg_semi_libertad_id object
75 incorporado_reg_semi_libertad_descripcion float64
76 participa_programa_pre_libertad float64
77 participa_programa_prision_discontinua_id object
78 participa_programa_prision_discontinua_descripcion float64
79 participa_programa_semi_detencion_id object
80 participa_programa_semi_detencion_descripcion float64
81 tuvo_reduccion_pena_id object
82 tuvo_reduccion_pena_descripcion float64
83 mujer_tiene_hijos_intramuro int64
84 provincia_indec_id float64
85 ultima_provincia_residencia_indec_id float64
```

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64
63 duracion_condena_rango int64
64 tipo_condena float64
65 tiene_medidas_seguridad float64
66 es_reincidente id object
67 es_reincidente_descripcion float64
68 tiene_periodo_progresividad_id object
69 tiene_periodo_progresividad_descripcion int64
70 reparticion_id object
71 reparticion_descripcion float64
72 tuvo_salidas_transitorias_id object
73 tuvo_salidas_transitorias_descripcion float64
74 incorporado_reg_semi_libertad_id object
75 incorporado_reg_semi_libertad_descripcion float64
76 participa_programa_pre_libertad float64
77 participa_programa_prision_discontinua_id object
78 participa_programa_prision_discontinua_descripcion float64
79 participa_programa_semi_detencion_id object
80 participa_programa_semi_detencion_descripcion float64
81 tuvo_reduccion_pena_id object
82 tuvo_reduccion_pena_descripcion float64
83 mujer_tiene_hijos_intramuro int64
84 provincia_indec_id float64
85 ultima_provincia_residencia_indec_id float64
```

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64
63 duracion_condena_rango int64
64 tipo_condena float64
65 tiene_medidas_seguridad float64
66 es_reincidente id object
67 es_reincidente_descripcion float64
68 tiene_periodo_progresividad_id object
69 tiene_periodo_progresividad_descripcion int64
70 reparticion_id object
71 reparticion_descripcion float64
72 tuvo_salidas_transitorias_id object
73 tuvo_salidas_transitorias_descripcion float64
74 incorporado_reg_semi_libertad_id object
75 incorporado_reg_semi_libertad_descripcion float64
76 participa_programa_pre_libertad float64
77 participa_programa_prision_discontinua_id object
78 participa_programa_prision_discontinua_descripcion float64
79 participa_programa_semi_detencion_id object
80 participa_programa_semi_detencion_descripcion float64
81 tuvo_reduccion_pena_id object
82 tuvo_reducción_pena_descripcion float64
83 mujer_tiene_hijos_intramuro int64
84 provincia_indec_id float64
85 ultima_provincia_residencia_indec_id float64
```

\_id → object

\_descripcion → float64/int64

# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64
63 duracion_condena_rango int64
64 tipo_condena float64
65 tiene_medidas_seguridad float64
66 es_reincidente id object
67 es_reincidente descripcion float64
68 tiene_periodo_progresividad_id object
69 tiene_periodo_progresividad_descripcion int64
70 reparticion_id object
71 reparticion_descripcion float64
72 tuvo_salidas_transitorias_id object
73 tuvo_salidas_transitorias_descripcion float64
74 incorporado_reg_semi_libertad_id object
75 incorporado_reg_semi_libertad_descripcion float64
76 participa_programa_pre_libertad float64
77 participa_programa_prision_discontinua_id object
78 participa_programa_prision_discontinua_descripcion float64
79 participa_programa_semi_detencion_id object
80 participa_programa_semi_detencion_descripcion float64
81 tuvo_reduccion_pena_id object
82 tuvo_reducción_pena_descripcion float64
83 mujer_tiene_hijos_intramuro int64
84 provincia_indec_id float64
85 ultima_provincia_residencia_indec_id float64
```

Eliminamos feature ‘tipo\_condena’ y renombramos las columnas con el nombre *de abajo*

**\_id → object**  
**\_descripcion → float64/int64**

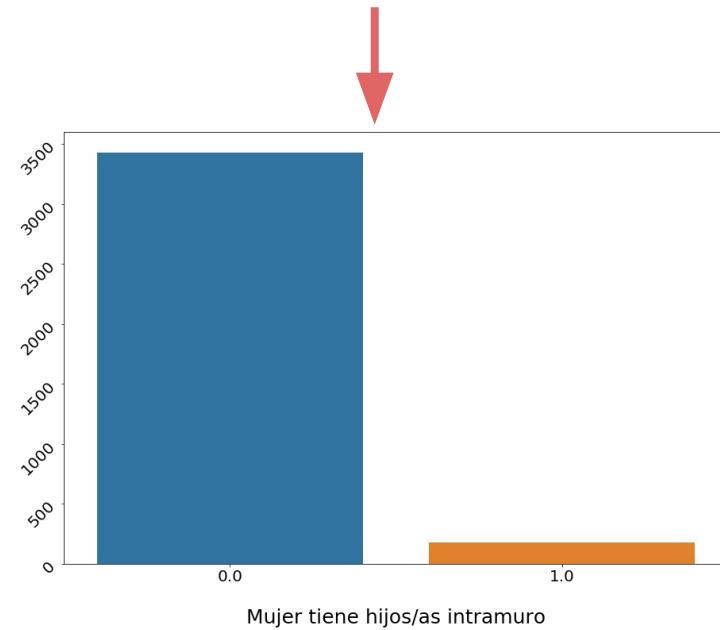
# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64  
63 duracion_condena_rango int64  
64 tipo_condena float64  
65 tiene_medidas_seguridad float64  
66 es_reincidente id object  
67 es_reincidente_descripcion float64  
68 tiene_periodo_progresividad_id object  
69 tiene_periodo_progresividad_descripcion int64  
70 reparticion_id object  
71 reparticion_descripcion float64  
72 tuvo_salidas_transitorias_id object  
73 tuvo_salidas_transitorias_descripcion float64  
74 incorporado_reg_semi_libertad_id object  
75 incorporado_reg_semi_libertad_descripcion float64  
76 participa_programa_pre_libertad float64  
77 participa_programa_prision_discontinua_id object  
78 participa_programa_prision_discontinua_descripcion float64  
79 participa_programa_semi_detencion_id object  
80 participa_programa_semi_detencion_descripcion float64  
81 tuvo_reduccion_pena_id object  
82 tuvo_reducción_pena_descripcion float64  
83 mujer_tiene_hijos_intramuro int64  
84 provincia_indec_id float64  
85 ultima_provincia_residencia_indec_id float64
```

\_id → object  
\_descripcion → float64/int64

Eliminamos feature ‘tipo\_condena’ y renombramos las columnas con el nombre de abajo



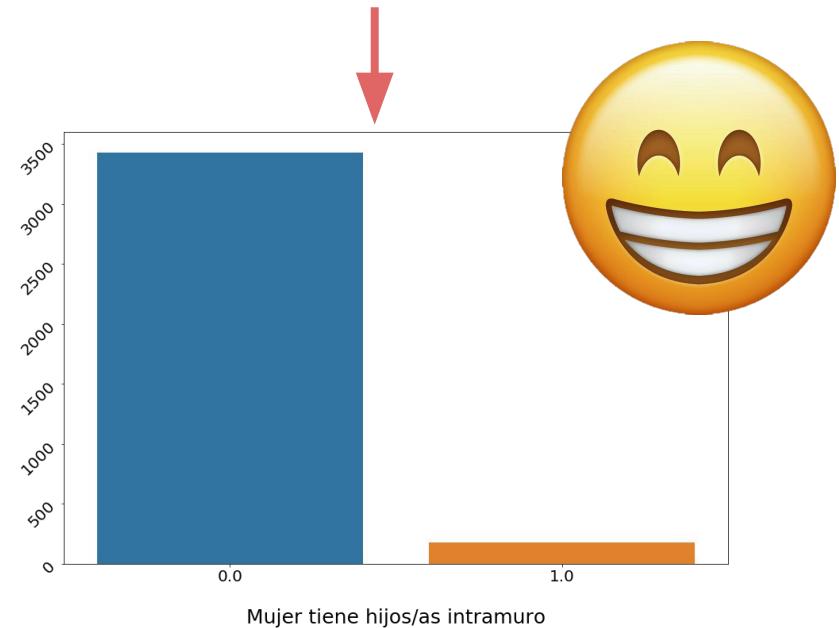
# Análisis y Curación de Datos

## Análisis y Consistencia de Datos

```
62 duracion_condena_meses float64  
63 duracion_condena_rango int64  
64 tipo_condena float64  
65 tiene_medidas_seguridad float64  
66 es_reincidente id object  
67 es_reincidente_descripcion float64  
68 tiene_periodo_progresividad_id object  
69 tiene_periodo_progresividad_descripcion int64  
70 reparticion_id object  
71 reparticion_descripcion float64  
72 tuvo_salidas_transitorias_id object  
73 tuvo_salidas_transitorias_descripcion float64  
74 incorporado_reg_semi_libertad_id object  
75 incorporado_reg_semi_libertad_descripcion float64  
76 participa_programa_pre_libertad float64  
77 participa_programa_prision_discontinua_id object  
78 participa_programa_prision_discontinua_descripcion float64  
79 participa_programa_semi_detencion_id object  
80 participa_programa_semi_detencion_descripcion float64  
81 tuvo_reduccion_pena_id object  
82 tuvo_reducción_pena_descripcion float64  
83 mujer_tiene_hijos_intramuro int64  
84 provincia_indec_id float64  
85 ultima_provincia_residencia_indec_id float64
```

\_id → object  
\_descripcion → float64/int64

Eliminamos feature ‘tipo\_condena’ y renombramos las columnas con el nombre de abajo



# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenando = row["edad"] - anios
    return edad_al_ser_condenando

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenando = row["edad"] - anios
    return edad_al_ser_condenando

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenando = row["edad"] - anios
    return edad_al_ser_condenando

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

Edad al ser condenado

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenado = row["edad"] - anios
    return edad_al_ser_condenado

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

Edad al ser condenado

Nos quedamos con los registros cuya edad al ser condenado sea mayor a 15 años y cuyo año al ser condenado sea menor al año del censo

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenando = row["edad"] - anios
    return edad_al_ser_condenando

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

Edad al ser condenado

Nos quedamos con los registros cuya edad al ser condenado sea mayor a 15 años y cuyo año al ser condenado sea menor al año del censo

```
def add_cantidad_delitos(df):
    delito_cols = ['delito{}_descripcion'.format(i) for i in range(1, 6)]
    df['delitos_cantidad'] = df[delito_cols].count(axis=1)
    return df

def filter_delitos(df):
    data = df[df["delitos_cantidad"] == 1]
    return data
```

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenado = row["edad"] - anios
    return edad_al_ser_condenado

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

Edad al ser condenado

Nos quedamos con los registros cuya edad al ser condenado sea mayor a 15 años y cuyo año al ser condenado sea menor al año del censo

```
def add_cantidad_delitos(df):
    delito_cols = ['delito{}_descripcion'.format(i) for i in range(1, 6)]
    df['delitos_cantidad'] = df[delito_cols].count(axis=1)
    return df

def filter_delitos(df):
    data = df[df["delitos_cantidad"] == 1]
    return data
```

Cantidad de delitos

# Análisis y Curación de Datos

## Generación de nuevas *features*

```
def add_anio_condenado(df):
    df["anio_condenado"] = df.apply(lambda row: calculate_anio_condenado(row) if row['fecha_condenado'] is pd.NaT
                                    else row['fecha_condenado'].year, axis=1)
    return df

def calculate_edad_al_ser_condenado(row):
    anios = row["anio_censo"] - row["anio_condenado"]
    edad_al_ser_condenado = row["edad"] - anios
    return edad_al_ser_condenado

def add_edad_al_ser_condenado(df):
    df["edad_al_ser_condenado"] = df.apply(lambda row: calculate_edad_al_ser_condenado(row), axis=1)
    return df

def remove_wrong_anio_o_edad(df):
    data = df[df["edad_al_ser_condenado"] > 15]
    data = data[data["anio_condenado"] <= data["anio_censo"]]
    return data
```

Año de condena

Edad al ser condenado

Nos quedamos con los registros cuya edad al ser condenado sea mayor a 15 años y cuyo año al ser condenado sea menor al año del censo

```
def add_cantidad_delitos(df):
    delito_cols = ['delito{}_descripcion'.format(i) for i in range(1, 6)]
    df['delitos_cantidad'] = df[delito_cols].count(axis=1)
    return df

def filter_delitos(df):
    data = df[df["delitos_cantidad"] == 1]
    return data
```

Cantidad de delitos

Nos quedamos con los registros cuya cantidad de delitos sea uno

# Análisis y Curación de Datos

## Filtrado Final

# Análisis y Curación de Datos

## Filtrado Final

Cantidad de registros inicial: 939727

Cantidad de registros de condenados: 385086

Cantidad de registros sin duplicados: 102935

Cantidad de registros con un solo delito: 82988

Cantidad de registros quitando año de condena o edad equivocado: 77024

Cantidad de registros quitando rango sin datos 68755

# Análisis y Curación de Datos

## Filtrado Final

### Registros inicial

Cantidad de registros inicial: 939727

Cantidad de registros de condenados: 385086

Cantidad de registros sin duplicados: 102935

Cantidad de registros con un solo delito: 82988

Cantidad de registros quitando anio de condena o edad equivocado: 77024

Cantidad de registros quitando rango sin datos: 68755

### Registros final

# Modelos de Aprendizaje Automático

Variable objetivo:

Duración de la condena

# Duración de condena

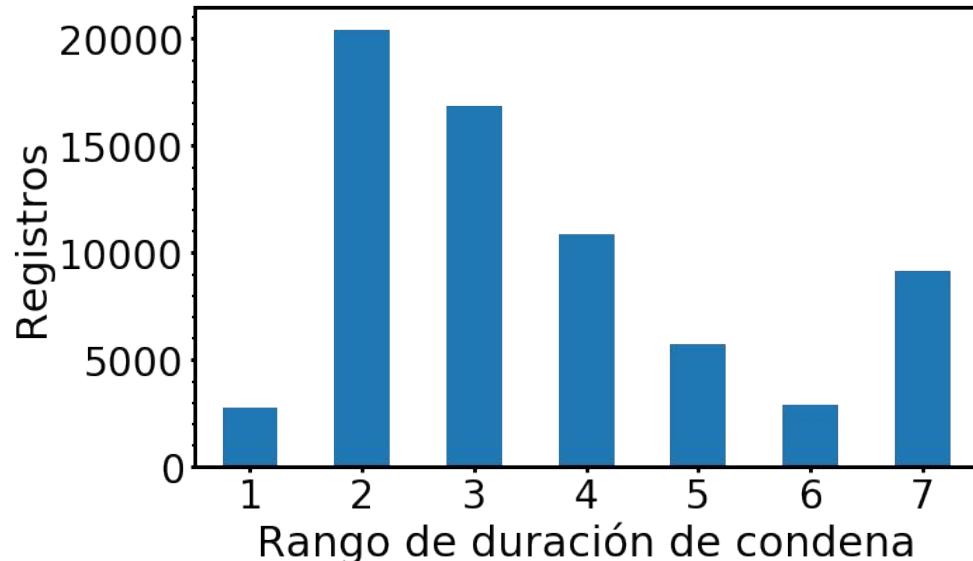
## Variable objetivo

Duración	Label
Hasta 3 años	1
De 3 a 6 años	2
De 6 a 9 años	3
De 9 a 12 años	4
De 12 a 15 años	5
De 15 a 18 años	6
Más de 18 años	7

# Duración de condena

Variable objetivo

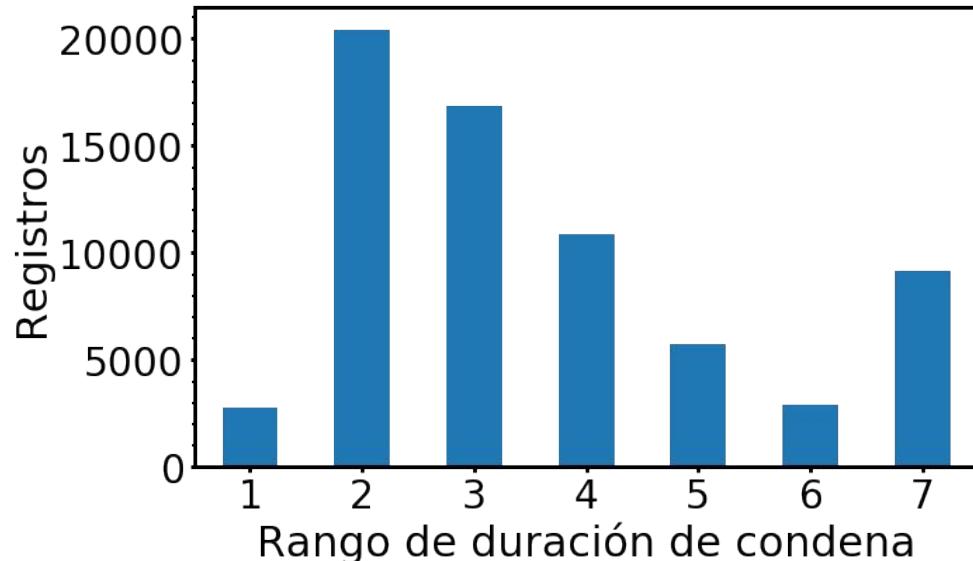
Duración	Label
Hasta 3 años	1
De 3 a 6 años	2
De 6 a 9 años	3
De 9 a 12 años	4
De 12 a 15 años	5
De 15 a 18 años	6
Más de 18 años	7



# Duración de condena

Variable objetivo

Duración	Label
Hasta 3 años	1
De 3 a 6 años	2
De 6 a 9 años	3
De 9 a 12 años	4
De 12 a 15 años	5
De 15 a 18 años	6
Más de 18 años	7



**Métrica:** accuracy

# Duración de condena

## Features

Delito	Es reincidente	Año condenado	Edad al ser condenado	Nivel de instrucción	Provincia	Estado civil	Género	Nacionalidad	Rango de duración de condena
Robo y/o tentativa de robo	Primario	2010	28	Primario incompleto	Mendoza	Concubino	Masculino	Argentina	7
Violaciones	Reiterante	2004	52	Primario completo	Córdoba	Separado	Masculino	Argentina	5
Infracción ley nº 23.737 (estupefacientes)	Primario	2002	42	Primario incompleto	Buenos Aires	Viudo	Femenino	Argentina	2
Homicidio doloso	Primario	2001	30	Secundario incompleto	Entre Ríos	Casado	Masculino	Argentina	3
Robo y/o tentativa de robo	Reincidente	2003	28	Primario incompleto	Santa Fe	Soltero	Masculino	Argentina	2
Infracción ley nº 23.737 (estupefacientes)	Primario	2001	30	Primario incompleto	La Pampa	Soltero	Masculino	Boliviana	2
Robo y/o tentativa de robo	Primario	2002	20	Primario completo	Córdoba	Soltero	Masculino	Argentina	2

# Duración de condena

## Features

Delito	Es reincidente	Año condenado	Edad al ser condenado	Nivel de instrucción	Provincia	Estado civil	Género	Nacionalidad	Rango de duración de condena
Robo y/o tentativa de robo	Primario	2010	28	Primario incompleto	Mendoza	Concubino	Masculino	Argentina	7
Violaciones	Reiterante	2004	52	Primario completo	Córdoba	Separado	Masculino	Argentina	5
Infracción ley nº 23.737 (estupefacientes)	Primario	2002	42	Primario incompleto	Buenos Aires	Viudo	Femenino	Argentina	2
Homicidio doloso	Primario	2001	30	Secundario incompleto	Entre Ríos	Casado	Masculino	Argentina	3
Robo y/o tentativa de robo	Reincidente	2003	28	Primario incompleto	Santa Fe	Soltero	Masculino	Argentina	2
Infracción ley nº 23.737 (estupefacientes)	Primario	2001	30	Primario incompleto	La Pampa	Soltero	Masculino	Boliviana	2
Robo y/o tentativa de robo	Primario	2002	20	Primario completo	Córdoba	Soltero	Masculino	Argentina	2

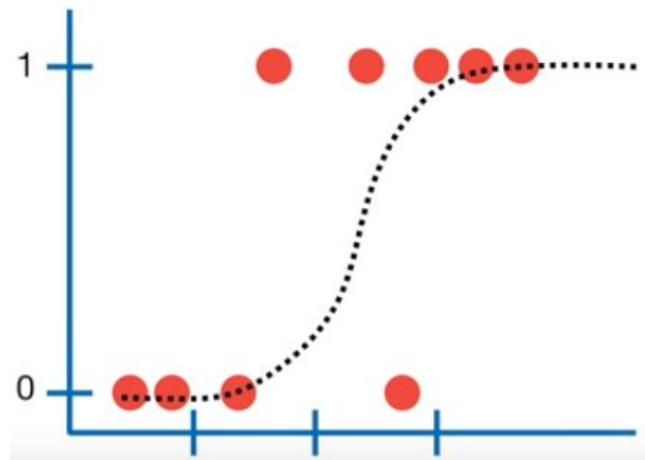
# Duración de condena

## Elección del método predictivo

# Duración de condena

## Elección del método predictivo

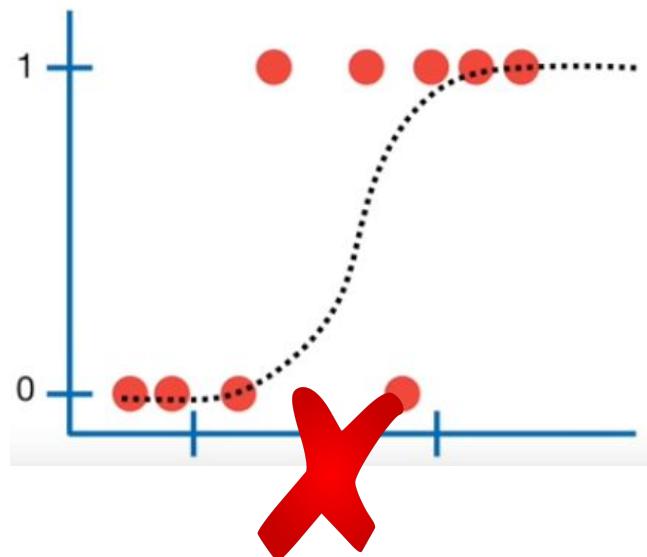
Logistic Regression



# Duración de condena

## Elección del método predictivo

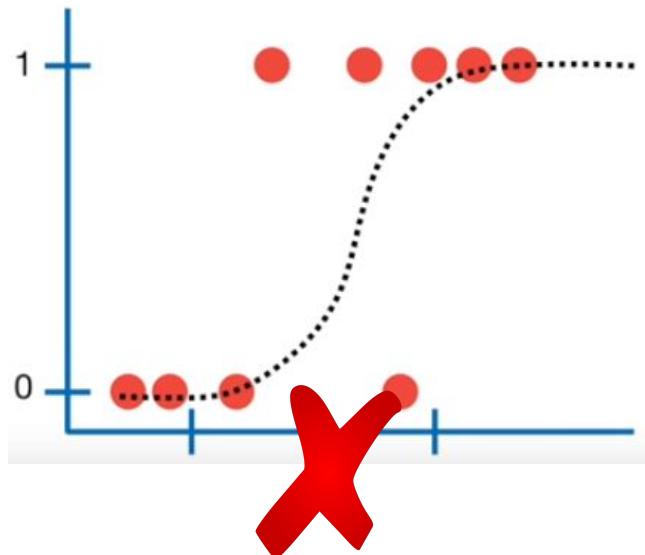
Logistic Regression



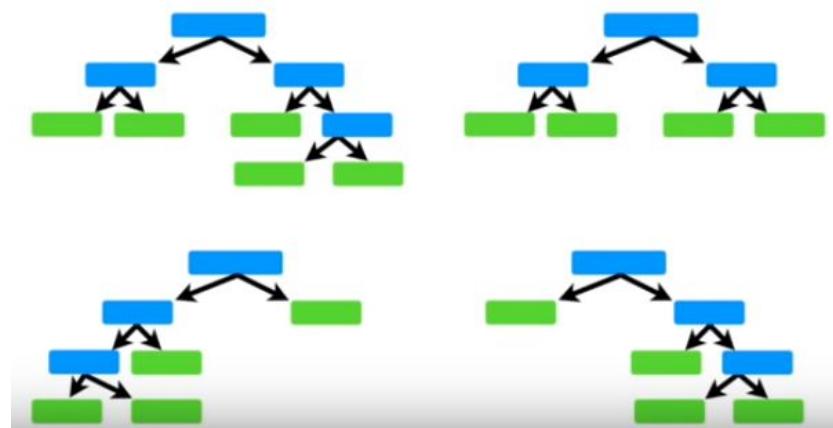
# Duración de condena

## Elección del método predictivo

Logistic Regression



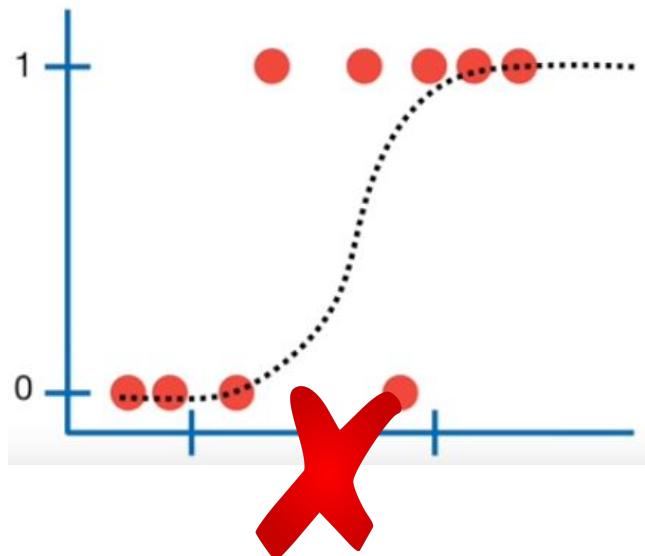
Random Forest



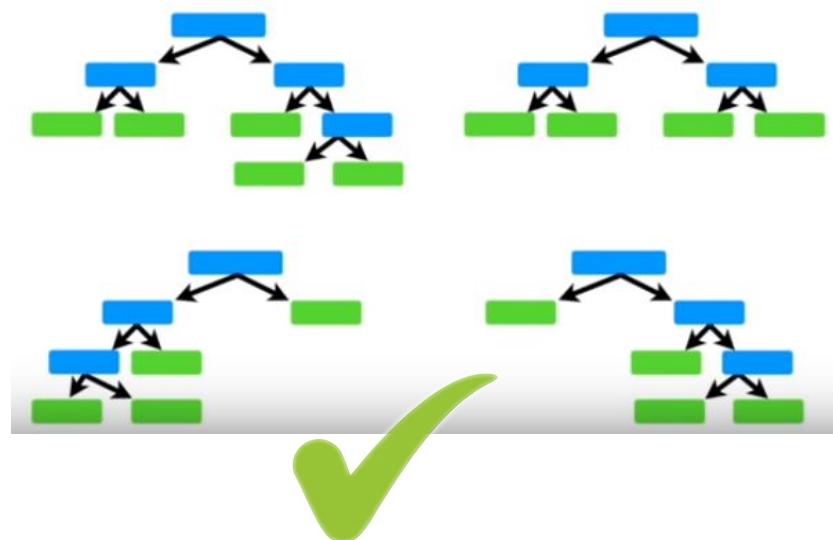
# Duración de condena

## Elección del método predictivo

Logistic Regression

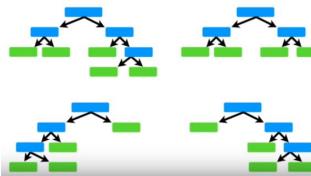


Random Forest



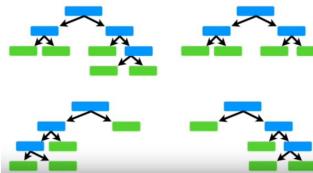
# Duración de condena

## Evaluación del modelo



# Duración de condena

## Evaluación del modelo



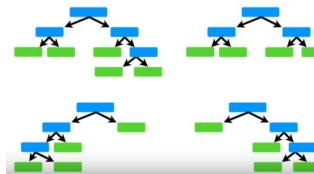
Accuracy

**Baseline = 0.38**  
**Train = 0.85**  
**Test = 0.57**

Label	Precision	Recall	F1-score	Support
1	0.44	0.13	0.2	288
2	0.56	0.77	0.65	2012
3	0.51	0.47	0.49	1658
4	0.57	0.51	0.54	1143
5	0.61	0.41	0.49	561
6	0.69	0.43	0.53	285
7	0.67	0.67	0.67	929

# Duración de condena

## Evaluación del modelo



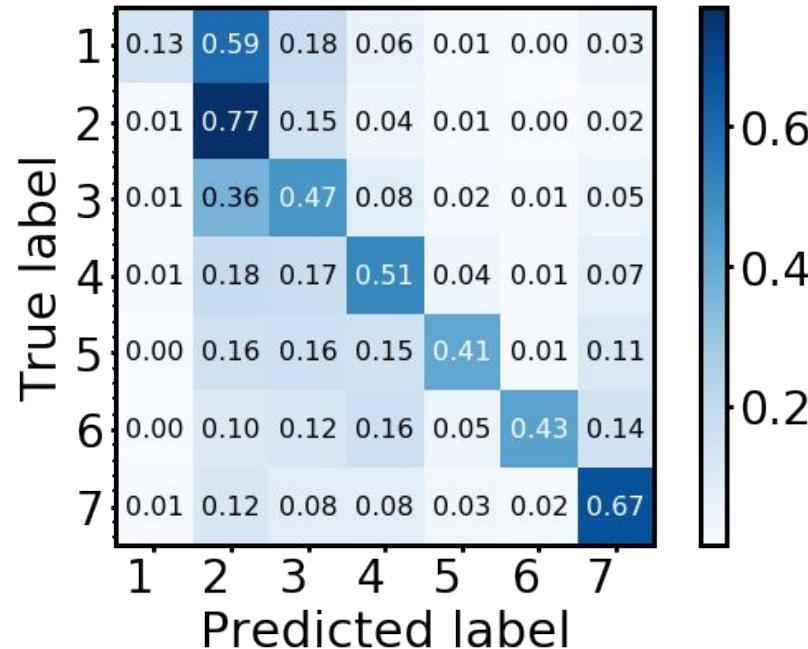
Accuracy

**Baseline = 0.38**

**Train = 0.85**

**Test = 0.57**

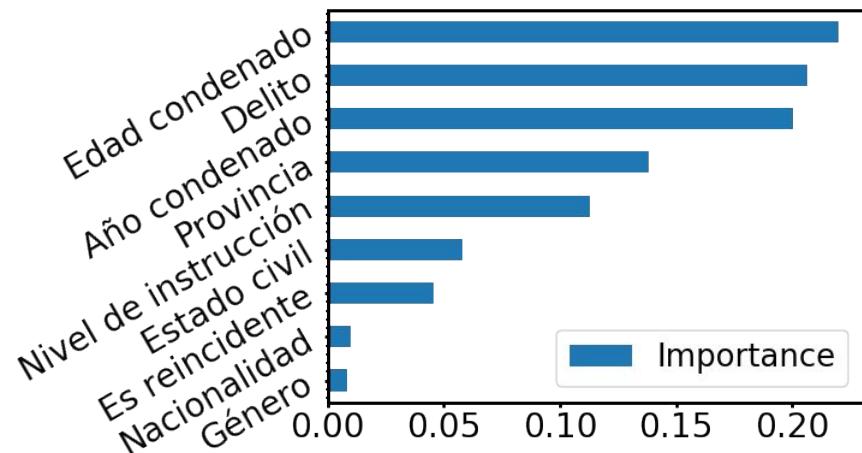
Label	Precision	Recall	F1-score	Support
1	0.44	0.13	0.2	288
2	0.56	0.77	0.65	2012
3	0.51	0.47	0.49	1658
4	0.57	0.51	0.54	1143
5	0.61	0.41	0.49	561
6	0.69	0.43	0.53	285
7	0.67	0.67	0.67	929



# Duración de condena

## Interpretabilidad

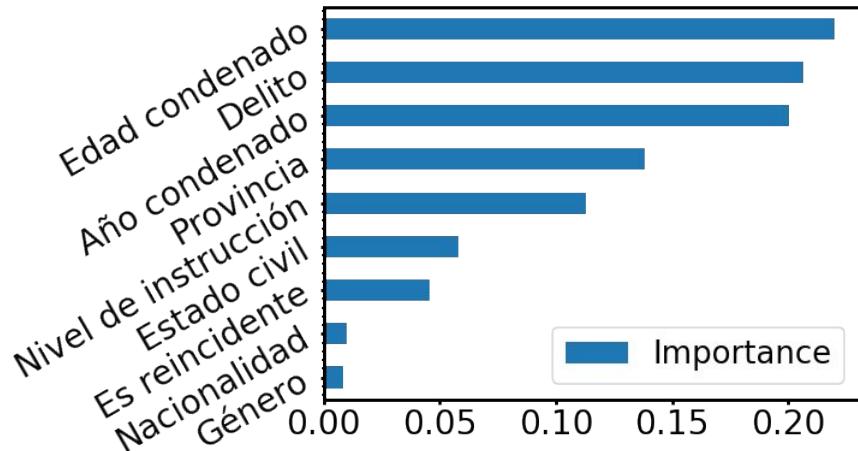
MDI importance



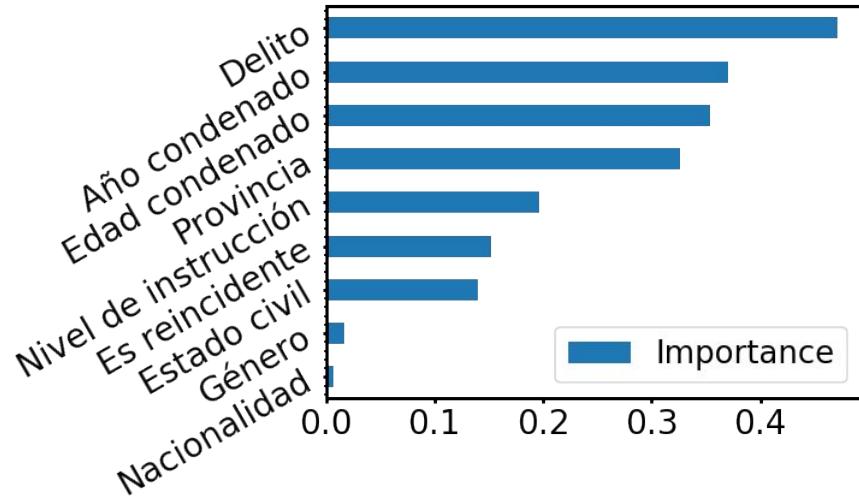
# Duración de condena

## Interpretabilidad

MDI importance



Permutation importance



# Duración de condena

## Mejoras propuestas

- **Eliminar features**
  - Género
  - Nacionalidad
- **Crear nuevos features**
  - Tipo de establecimiento
  - Buenos Aires vs. otras provincias
- **Modificar codificación**
  - One-hot encoding

# Duración de condena

## Limitaciones del modelo y de los datos

- **Inconsistencia en los datos**
  - Datos mal cargados
- **Información incompleta**
  - Poca especificidad en la categorización de los delitos.
    - Ley nº 23737 (estupefacientes)
    - Antecedentes, atenuantes, agravantes, ...

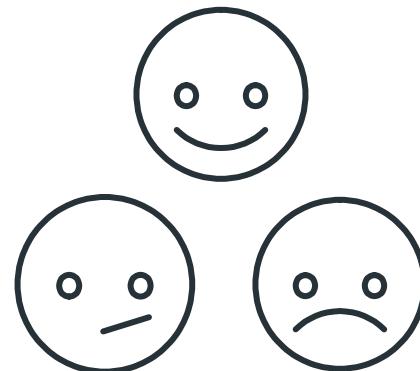
# Comentarios finales y conclusiones

# Comentarios finales

- **Más interrogantes para atacar**
  - Predicción de tentativas de suicidio (detección de outliers)
  - Evolución temporal de condenas por delito
- **Empoderamiento ciudadano a través del análisis de datos públicos**
  - Chequeo del discurso público

# ¡Muchas gracias!

## ¿Preguntas?



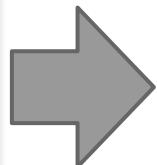
# Más Visualización y Análisis

Descripción de más variables del data set  
para el final y si queda tiempo o si a  
alguien le interesa o pregunta...

# Visualización y Análisis

## Variables del dataset: *(50 en total)*

año - provincia - establecimiento - edad - género - nacionalidad - estado civil  
- nivel de instrucción - última situación laboral - capacitación laboral al ingresar - último lugar de residencia - jurisdicción - situación legal - fecha detención - fecha condena - establecimiento de procedencia - delito 1 - delito 2 - delito 3 - delito 4 - delito 5 - horas de trabajo remunerado - participación en programa laboral - participación en programa educativo - participación en actividades deportivas - recibió atención médica en el último año - recibió visitas en el último año - participó de alteración al orden en el último año - tipo de infracción disciplinaria - sanción aplicada - calificación de conducta - tentativa de fugas - tentativa de suicidio - fue lesionado - duración de condena en años - duración de condena en meses - duración de condena rango - tipo de condena - tiene medidas de seguridad - es reincidente - tiene período de progresividad - repartición - tuvo salidas transitorias - incorporado al régimen de semi libertad - participa en programa pre libertad - participa de programa prisión discontinua - participa de programa - participa de programa semi detención - tuvo reducción de pena - mujer tiene hijos intramuro



## Utilizamos:

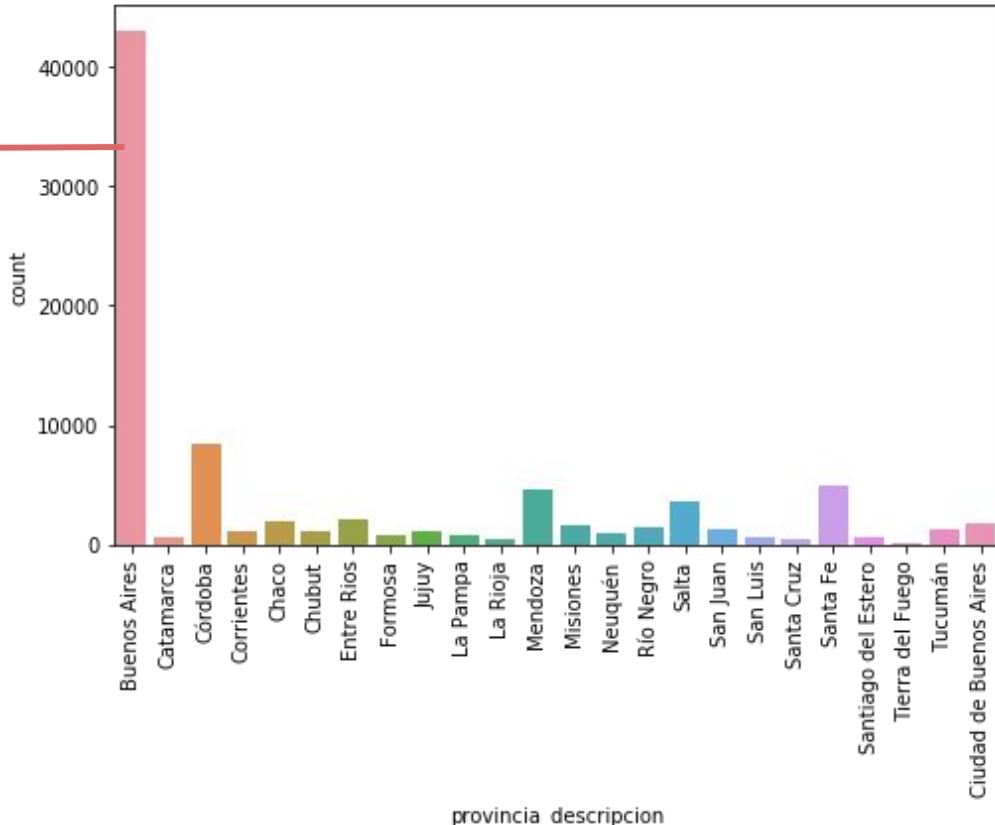
año - provincia - establecimiento - edad - género - nacionalidad - situación legal - fecha detención - fecha condena - delito 1 - delito 2 - delito 3 - delito 4 - delito 5 - participación en programa educativo - duración de condena en años - duración de condena en meses - ...

Censo 2017  
85.283  
internos

# Población Penitenciaria de Argentina 2017

## Provincia

La provincia de Buenos Aires es la que mayores cantidad de internos tiene



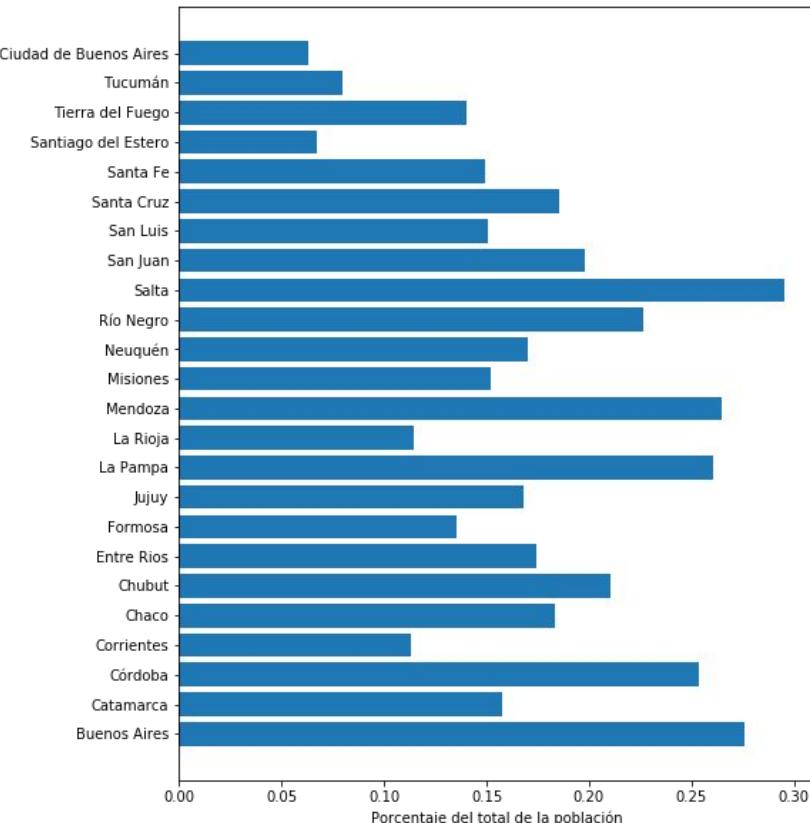
Estudiamos la población penal de cada provincia teniendo en cuenta la población total de la provincia

# Población Penitenciaria de Argentina 2017

## Provincia

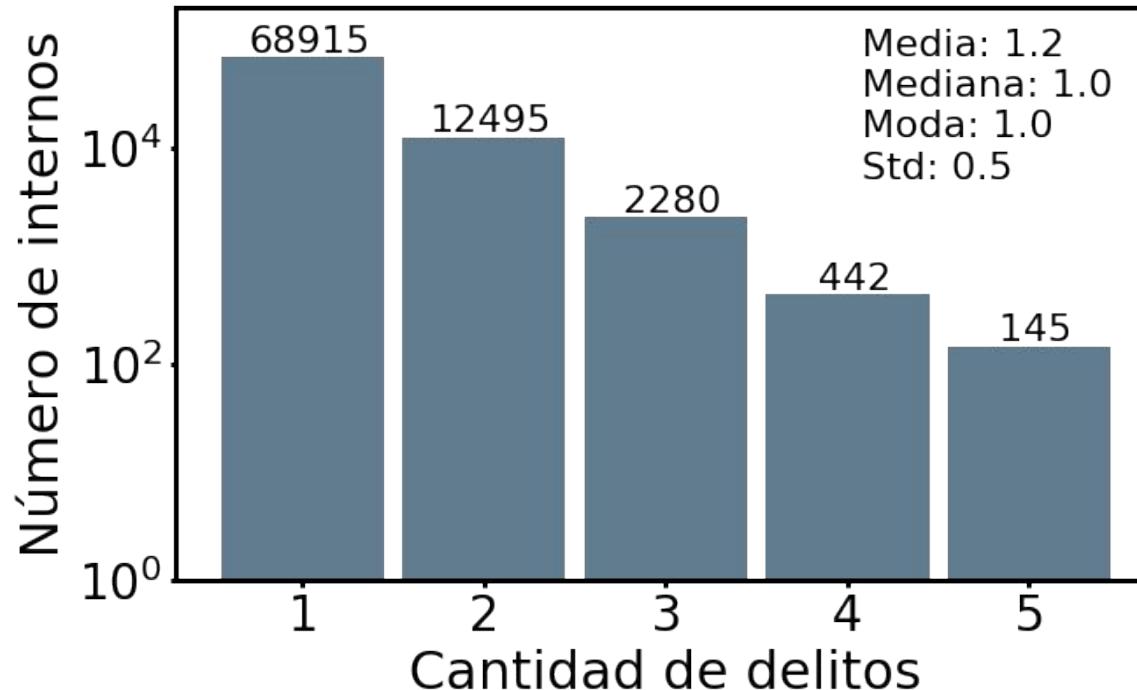
### Población de las provincias (censo 2010)<sup>6</sup>

1. Provincia de Buenos Aires	•	15.594.428
2. Córdoba	•	3.304.825
3. Santa Fe	•	3.300.736
4. Ciudad Autónoma de Buenos Aires	•	2.891.082
5. Mendoza	•	1.741.610
6. Tucumán	•	1.592.878
7. Entre Ríos	•	1.236.300
8. Salta	•	1.215.207
9. Misiones	•	1.097.829
10. Chaco	•	1.053.466
11. Corrientes	•	993.338
12. Santiago del Estero	•	896.461
13. San Juan	•	680.427
14. Jujuy	•	672.260
15. Río Negro	•	633.374
16. Neuquén	•	550.344
17. Formosa	•	527.895
18. Chubut	•	506.668
19. San Luis	•	461.588
20. Catamarca	•	367.820
21. La Rioja	•	331.847
22. La Pampa	•	316.940
23. Santa Cruz	•	272.524
24. Tierra del Fuego, Antártida e Islas del Atlántico Sur	•	126.190



# Población Penitenciaria de Argentina 2017

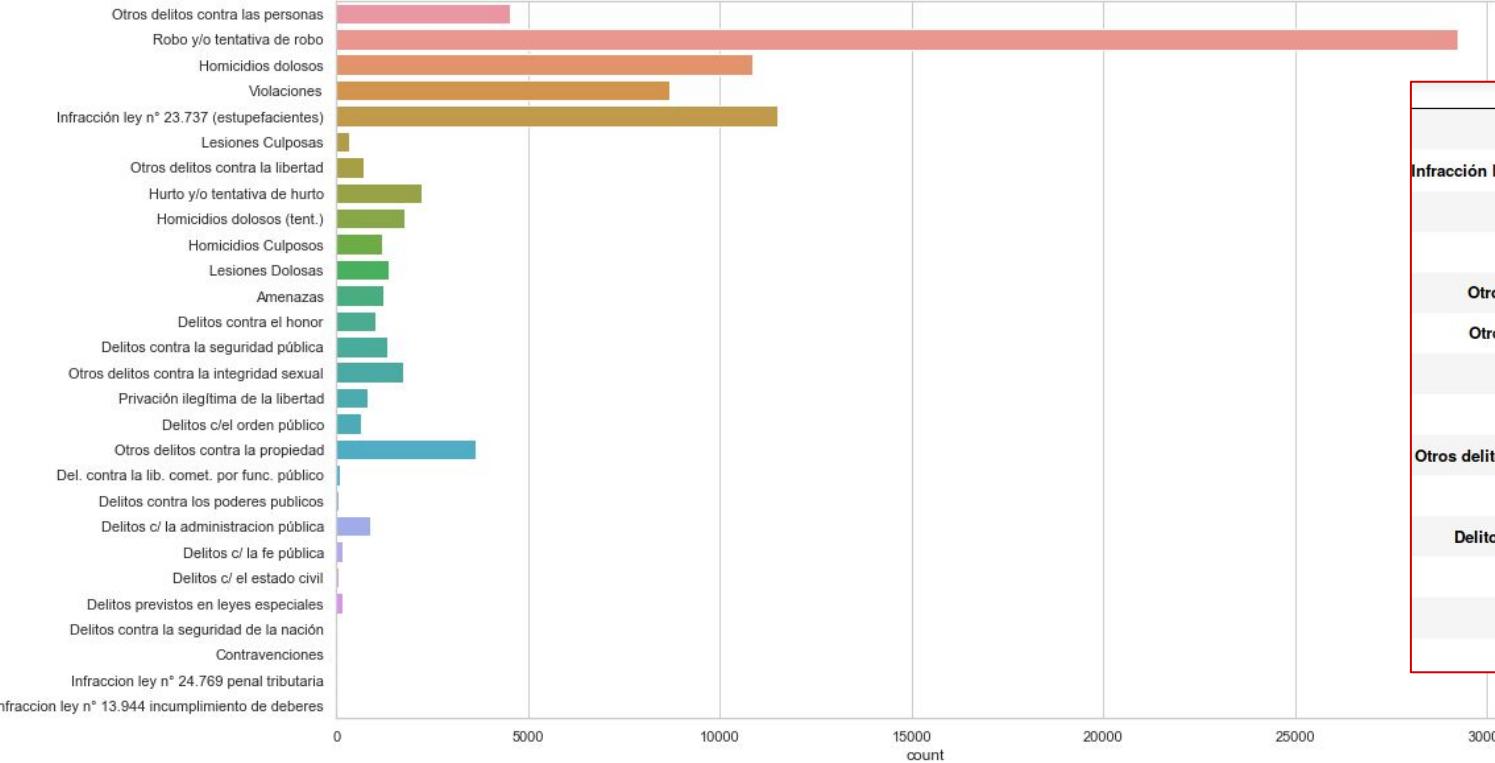
## Cantidad de Delitos



# Población Penitenciaria de Argentina 2017

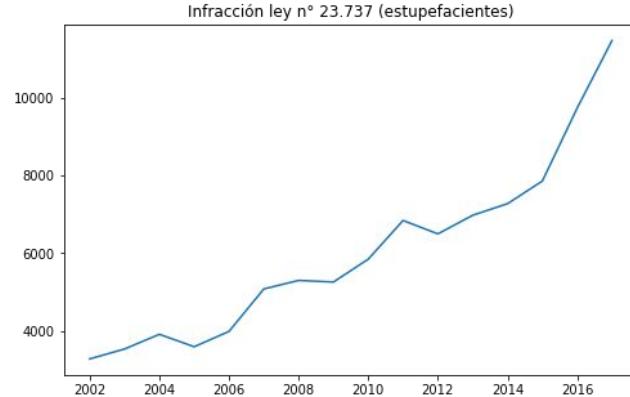
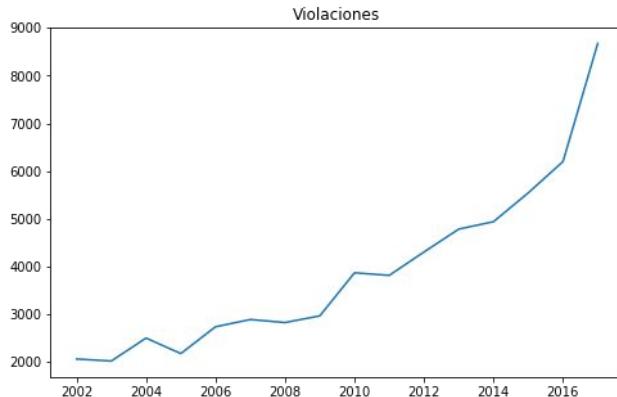
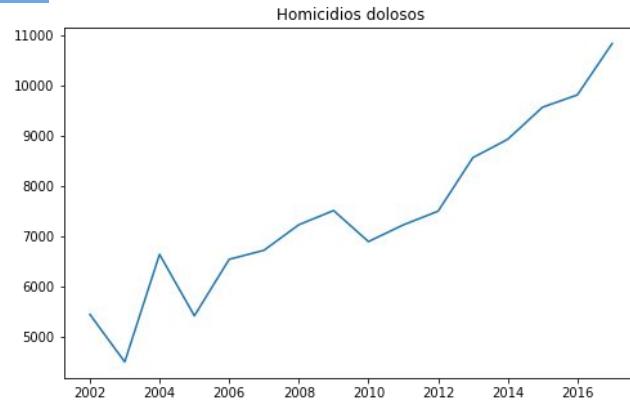
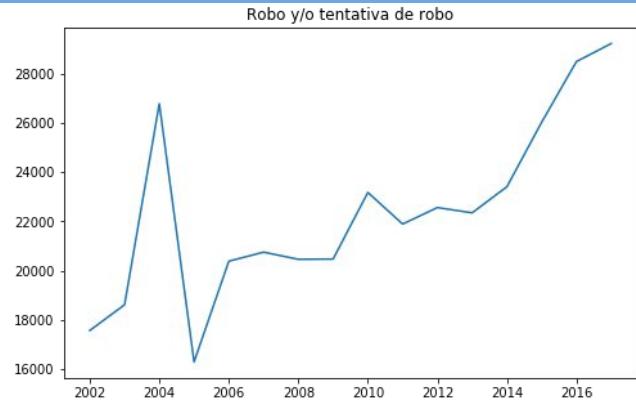
## Delito

Frecuencia delitos, población general



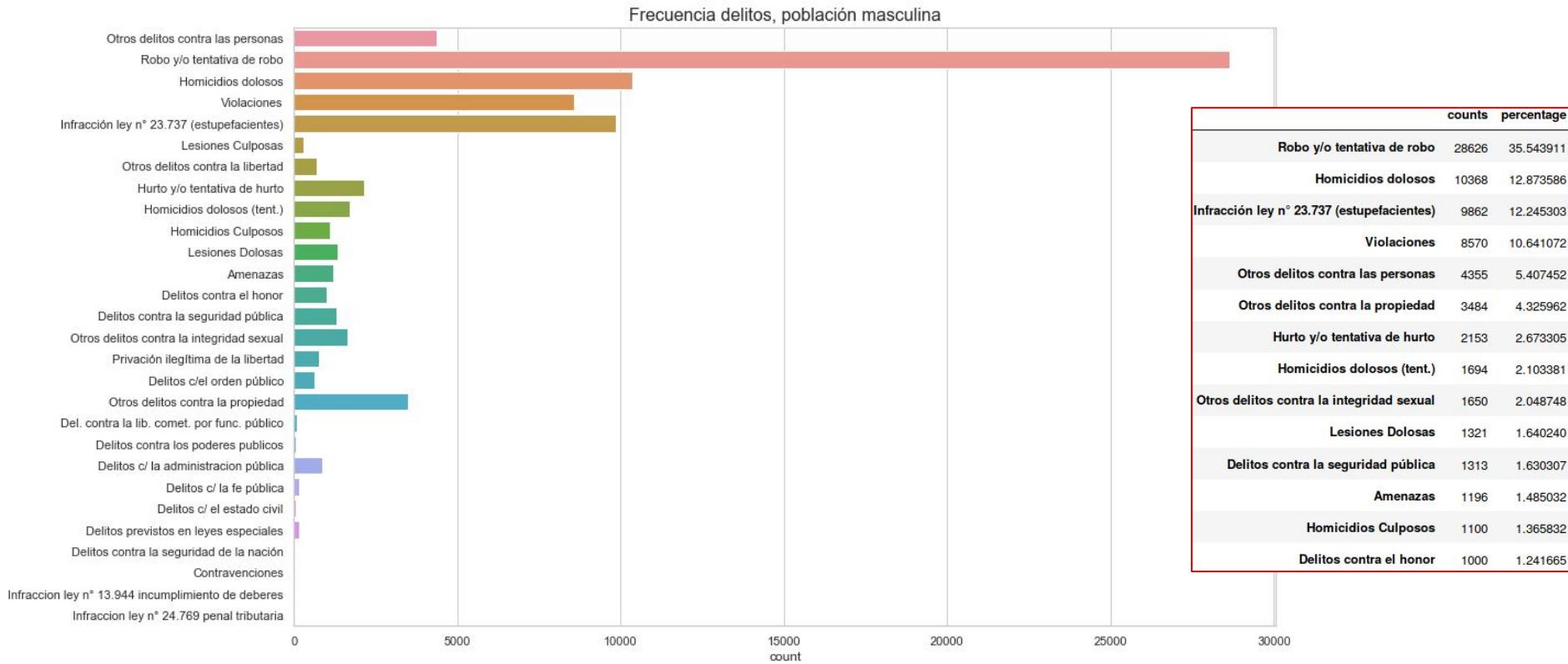
# Evolución Temporal

## Delitos más frecuentes



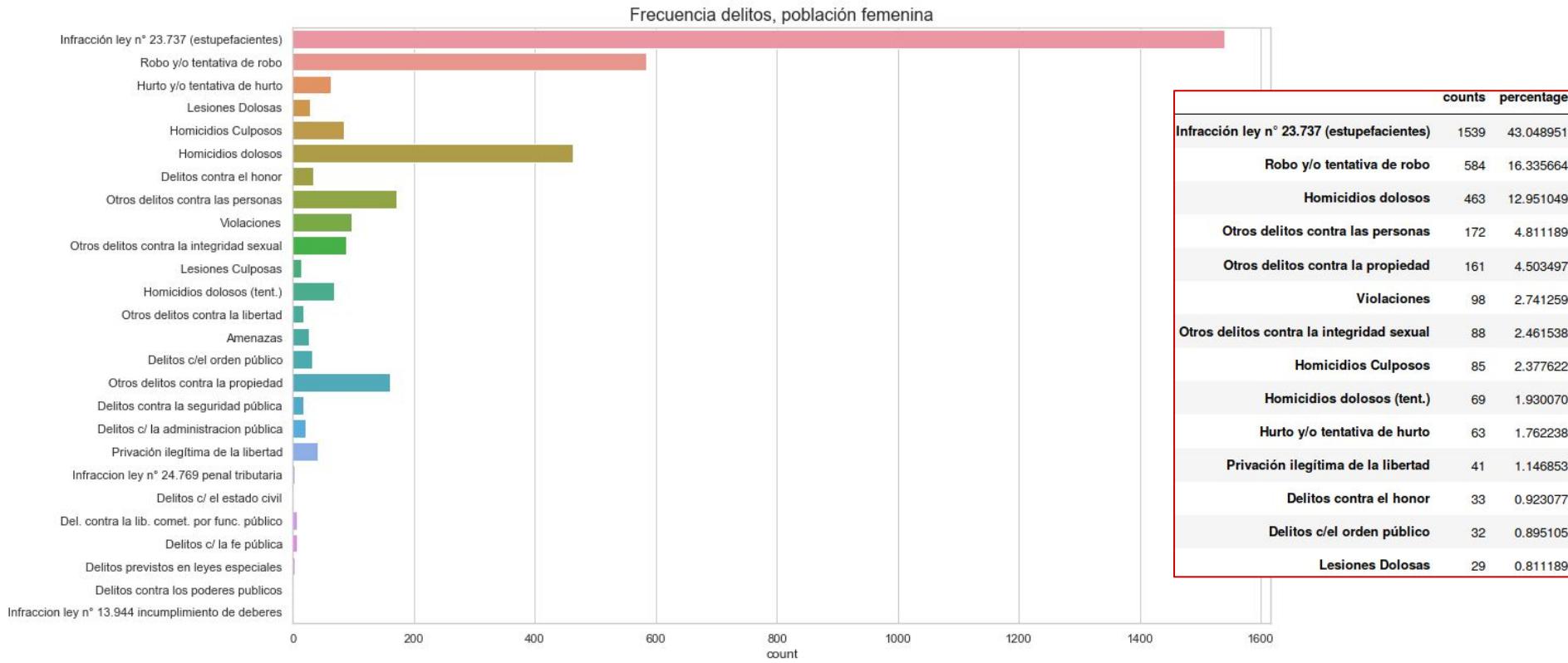
# Población Penitenciaria de Argentina 2017

## Delito1 | Hombres



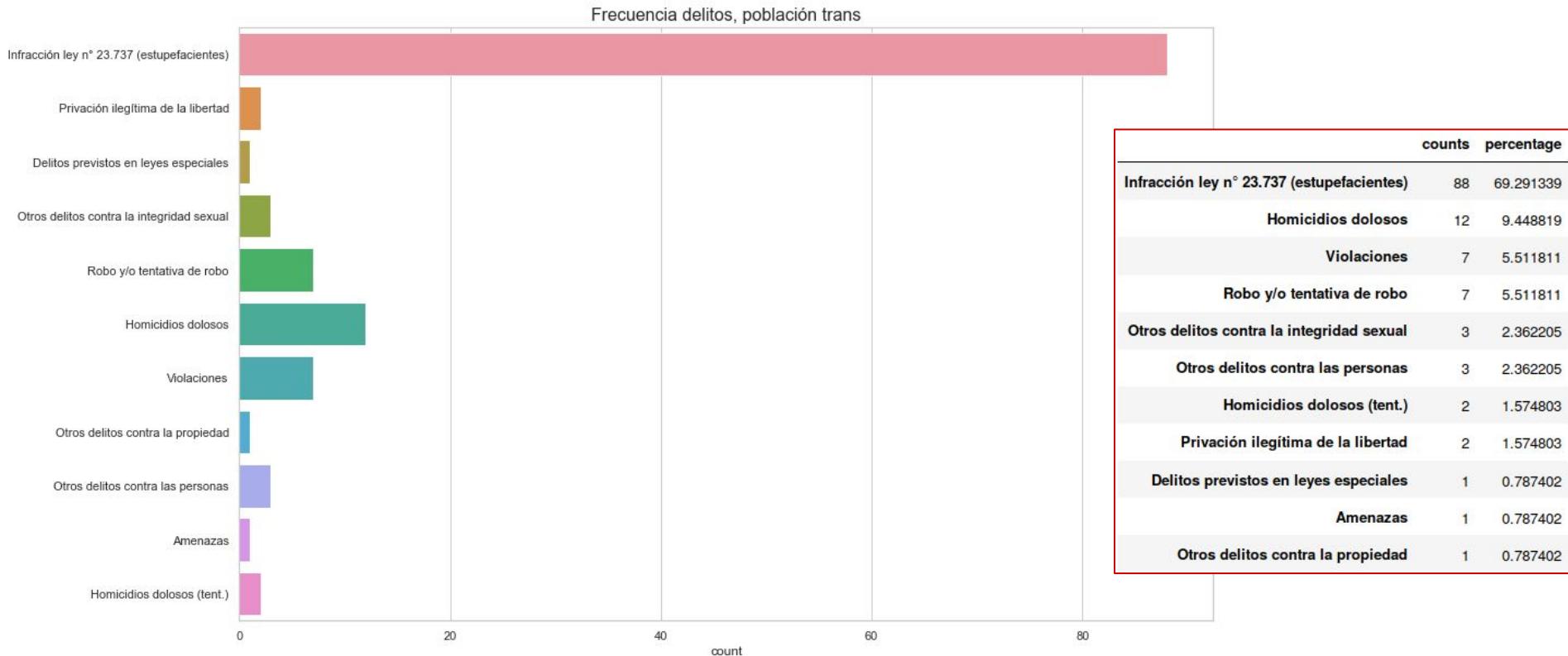
# Población Penitenciaria de Argentina 2017

## Delito1 | Mujeres



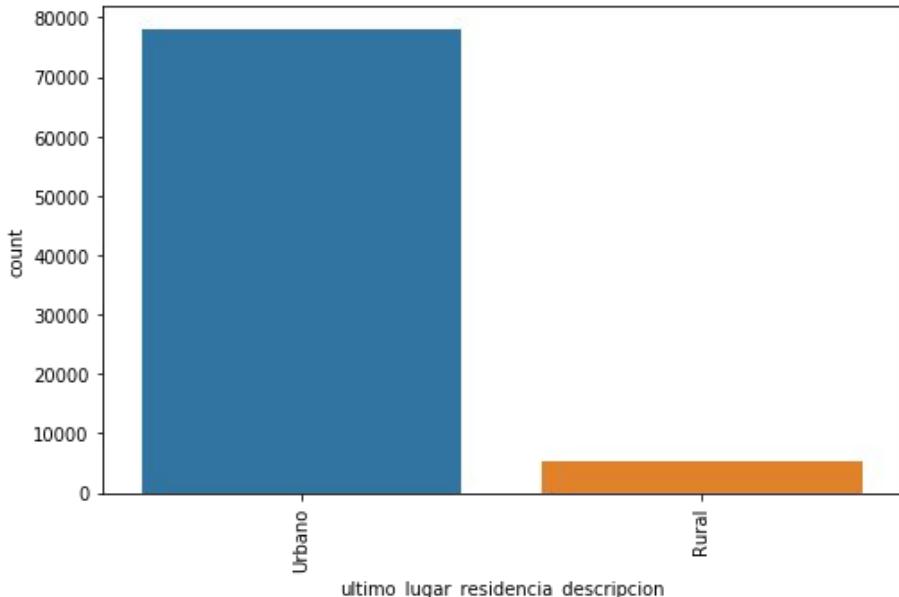
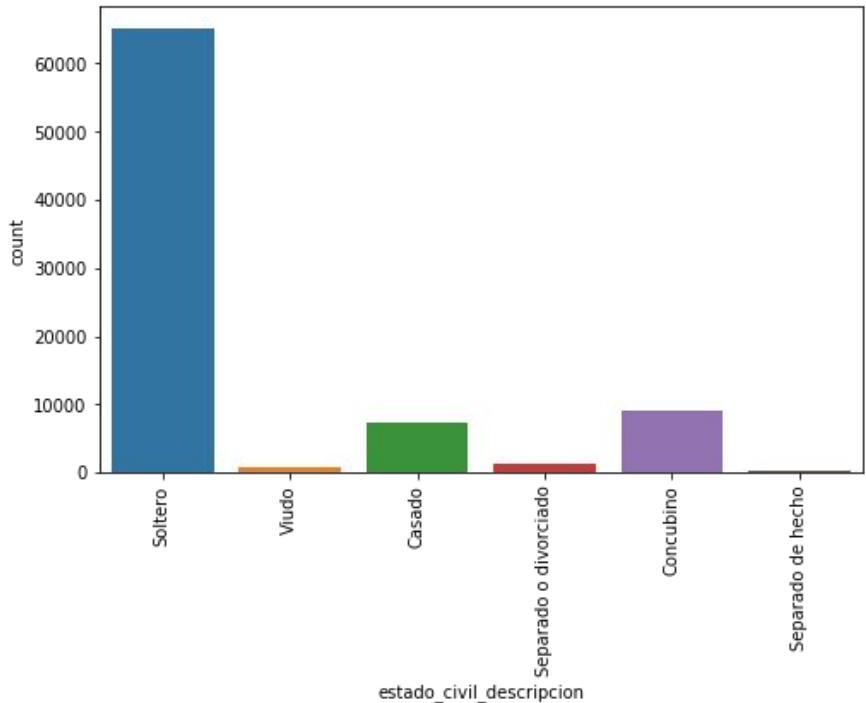
# Población Penitenciaria de Argentina 2017

## Delito1 | Trans



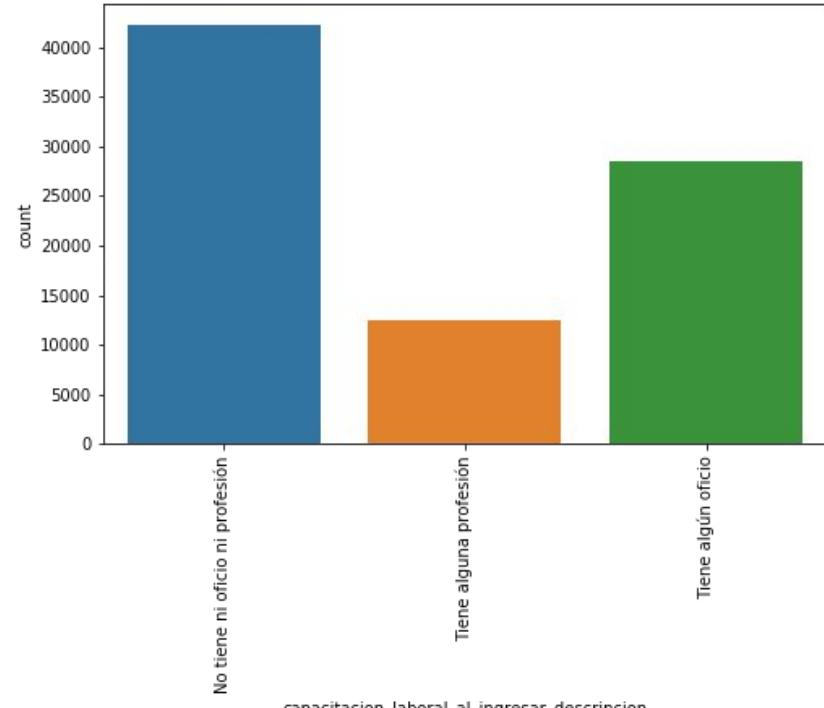
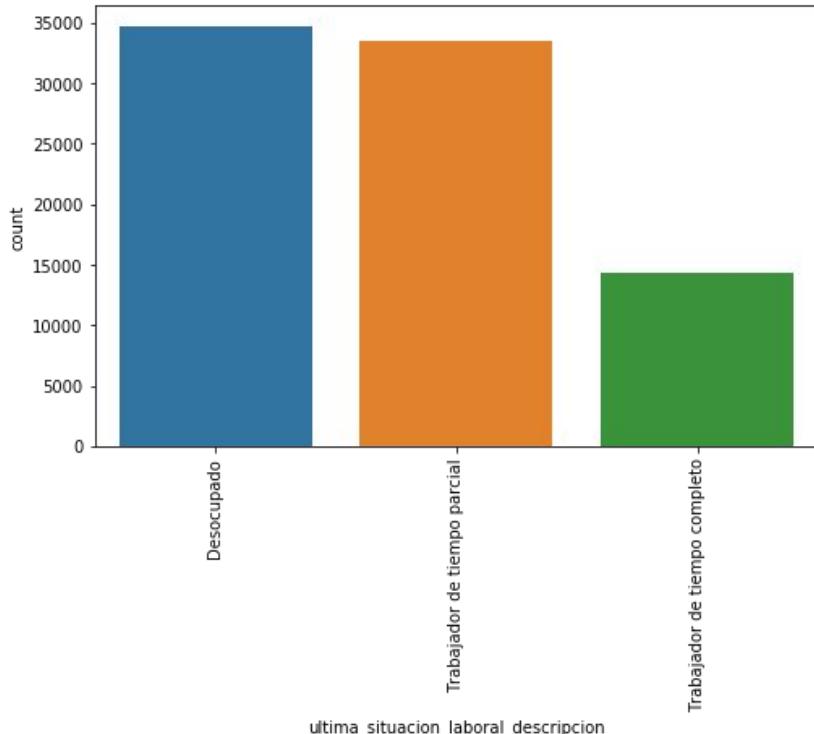
# Población Penitenciaria de Argentina 2017

## Estado Civil - Último lugar de residencia



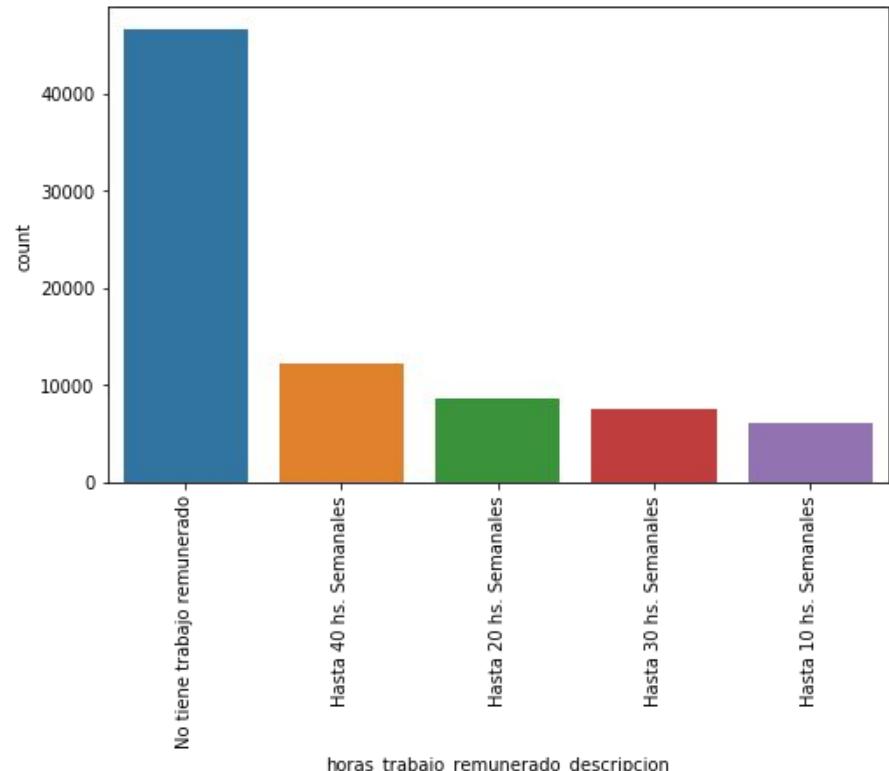
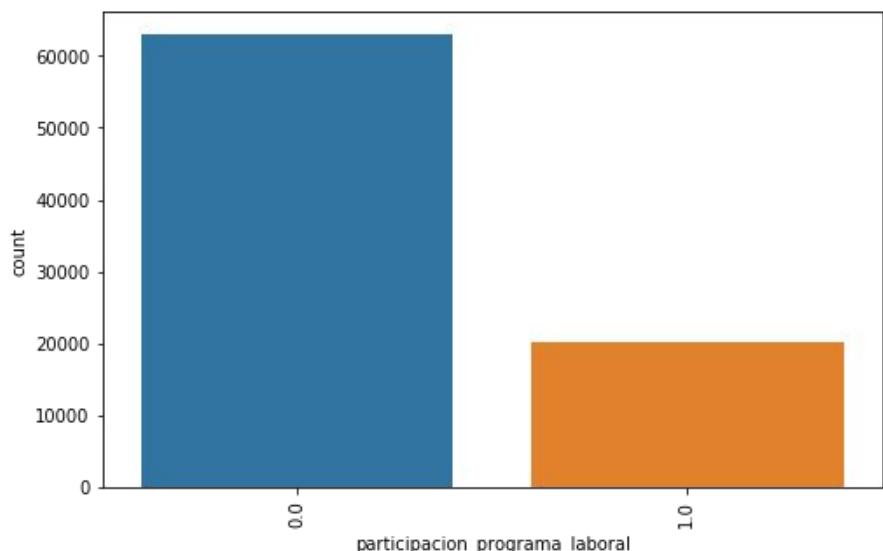
# Población Penitenciaria de Argentina 2017

## Trabajo



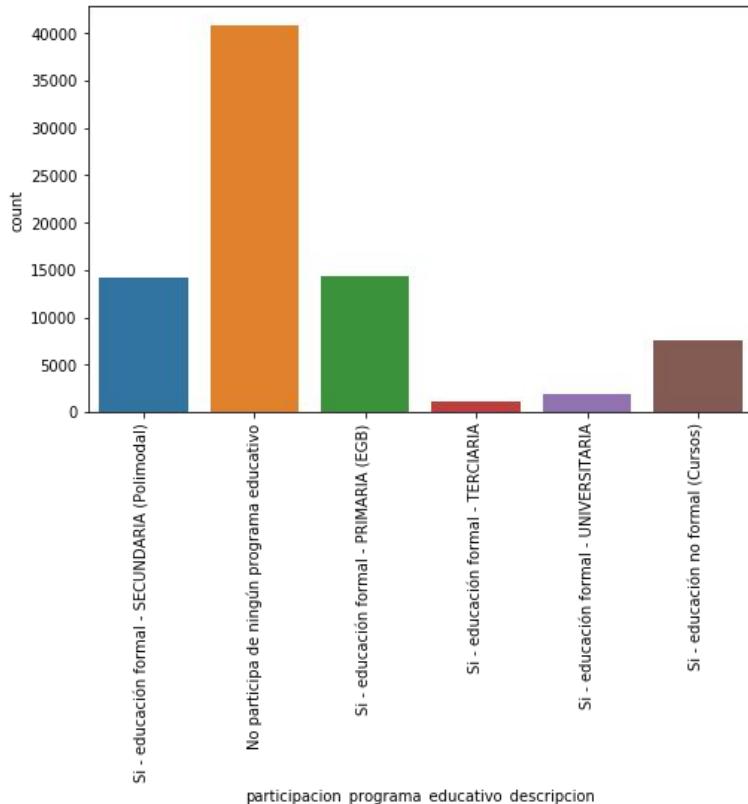
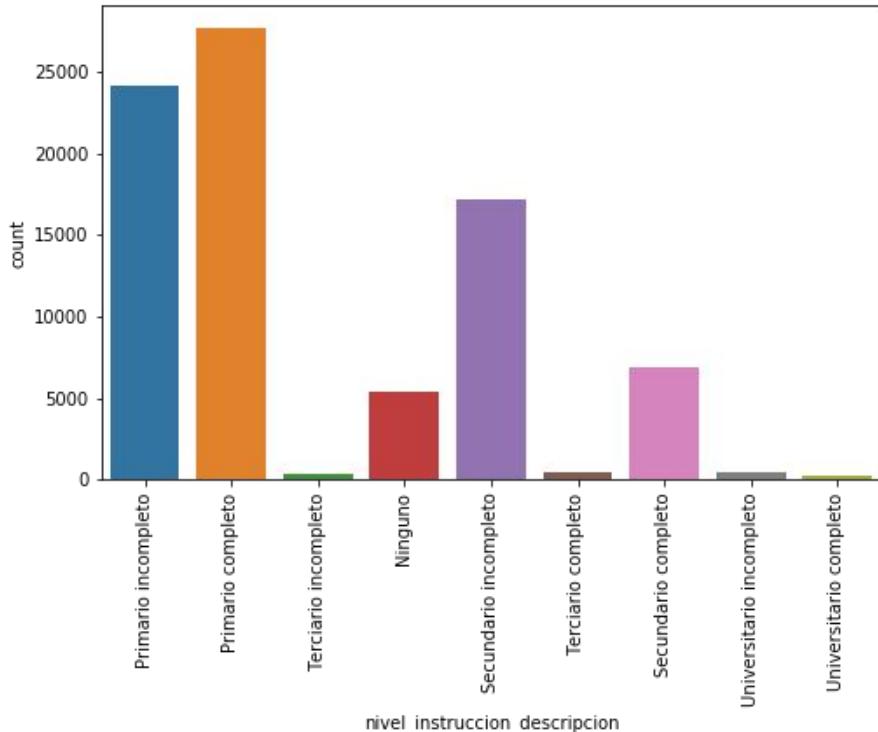
# Población Penitenciaria de Argentina 2017

## Trabajo



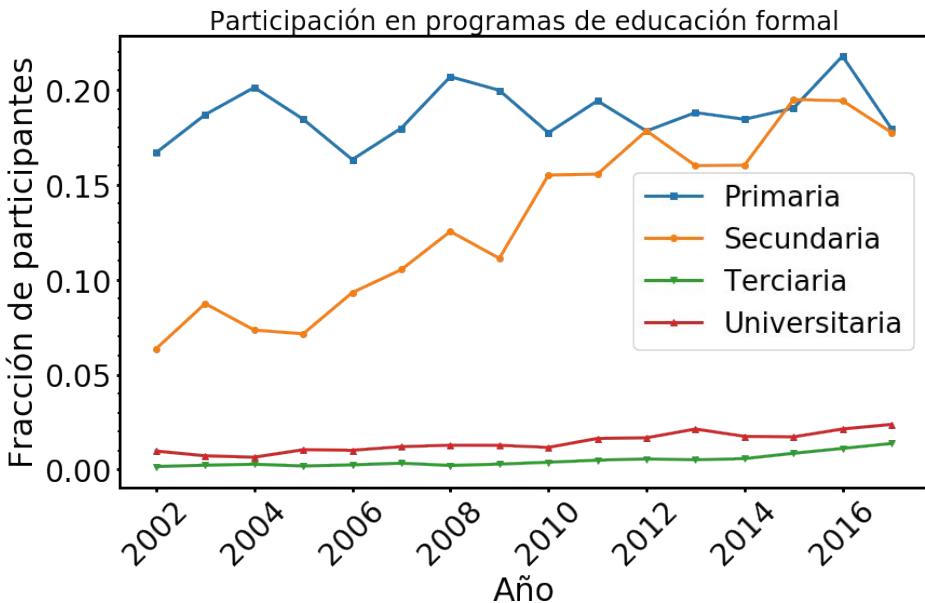
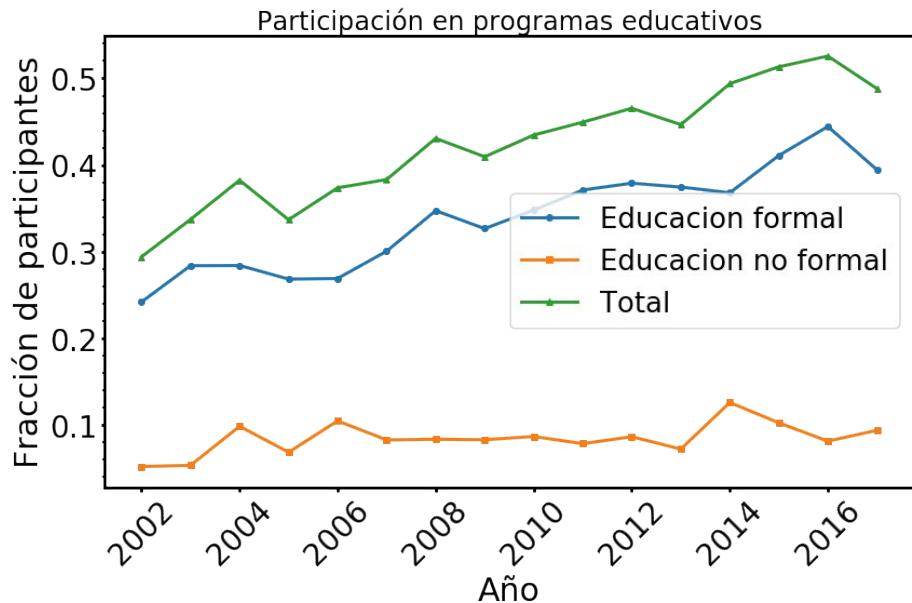
# Población Penitenciaria de Argentina 2017

## Educación



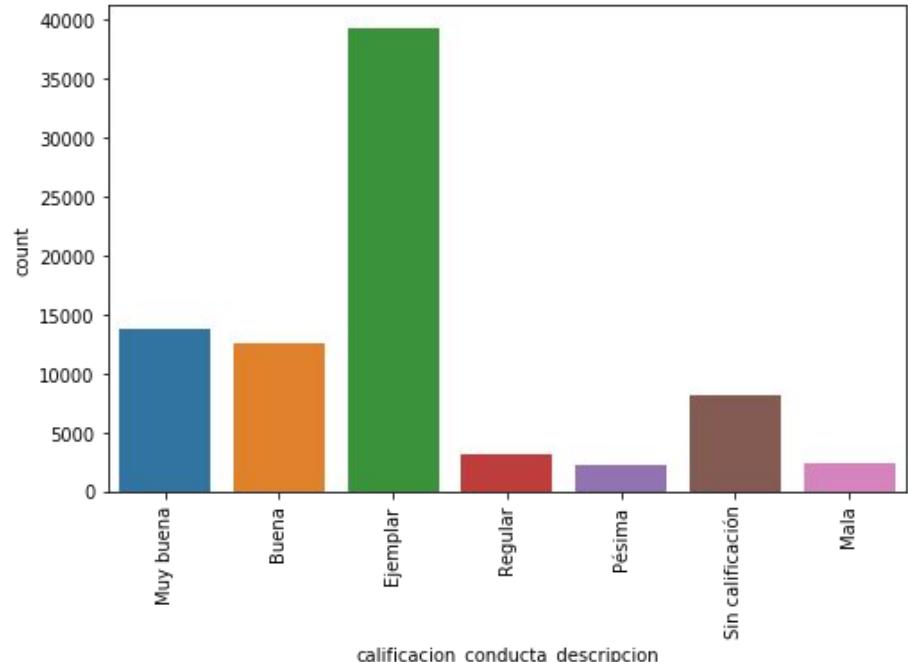
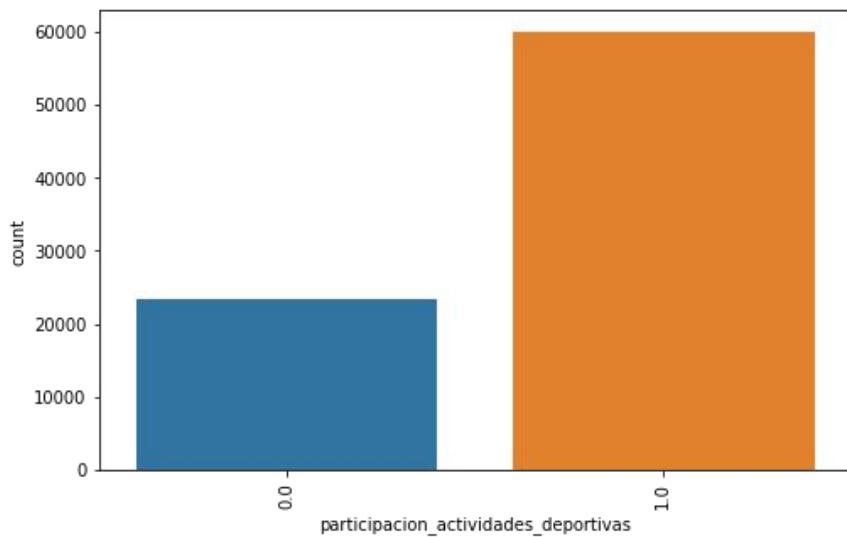
# Evolución Temporal

## Participación en programa educativo



# Población Penitenciaria de Argentina 2017

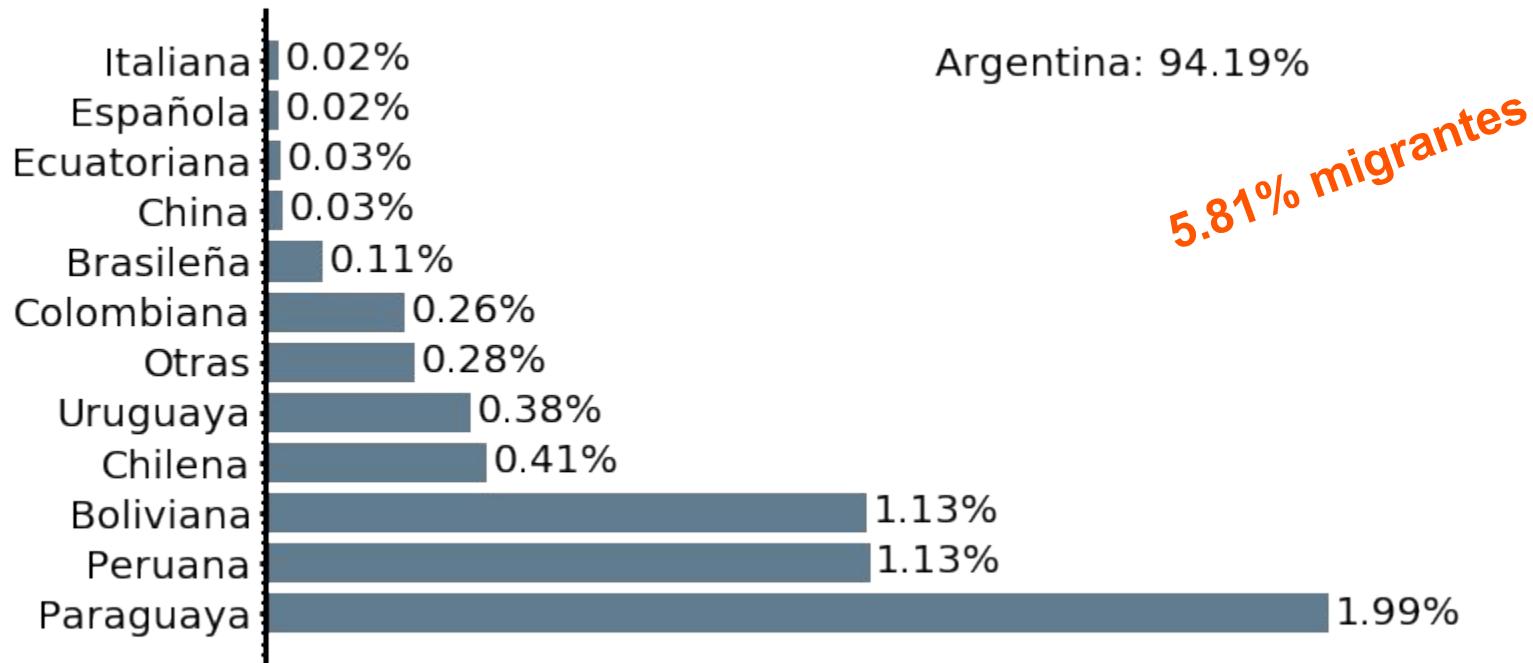
## Deportes - Conducta



# Población Penitenciaria de Argentina 2017

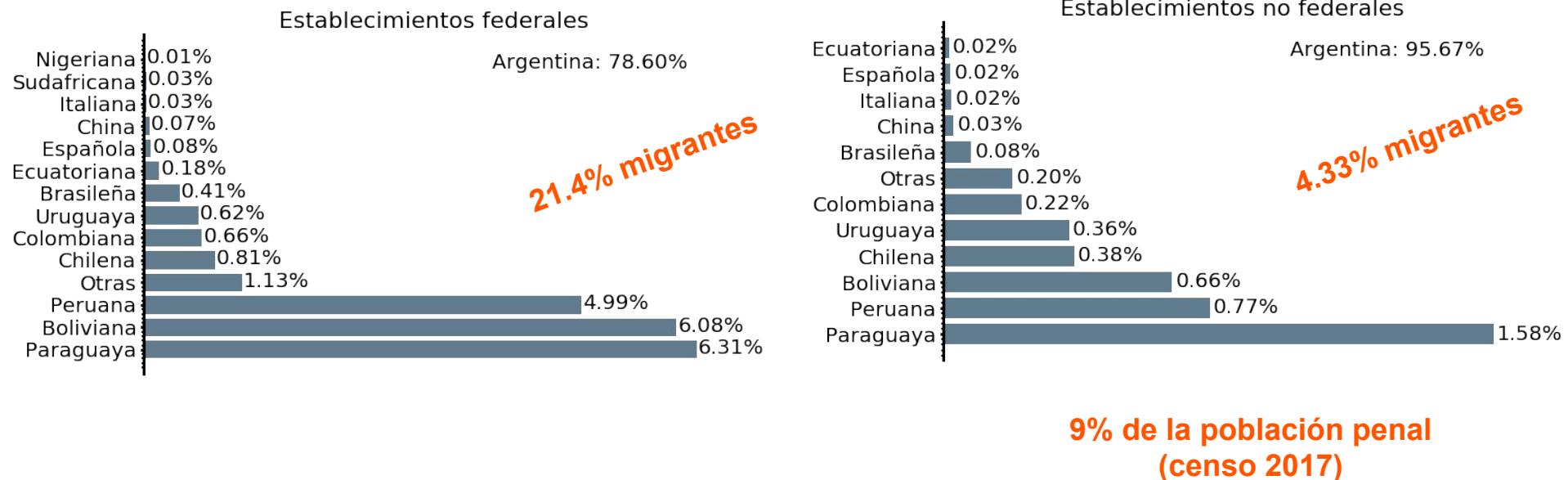
## Nacionalidad

Todos los establecimientos



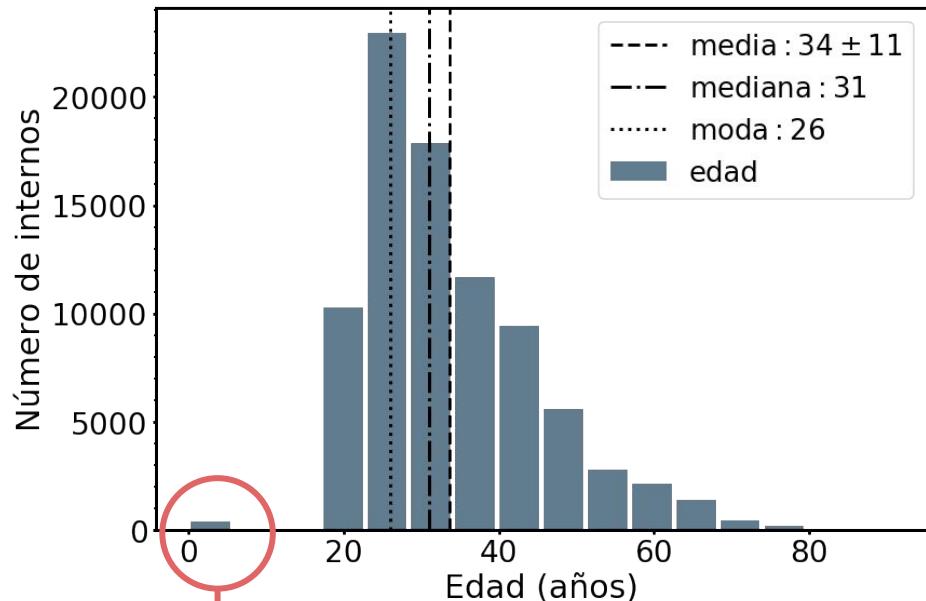
# Población Penitenciaria de Argentina 2017

## Nacionalidad | Establecimiento

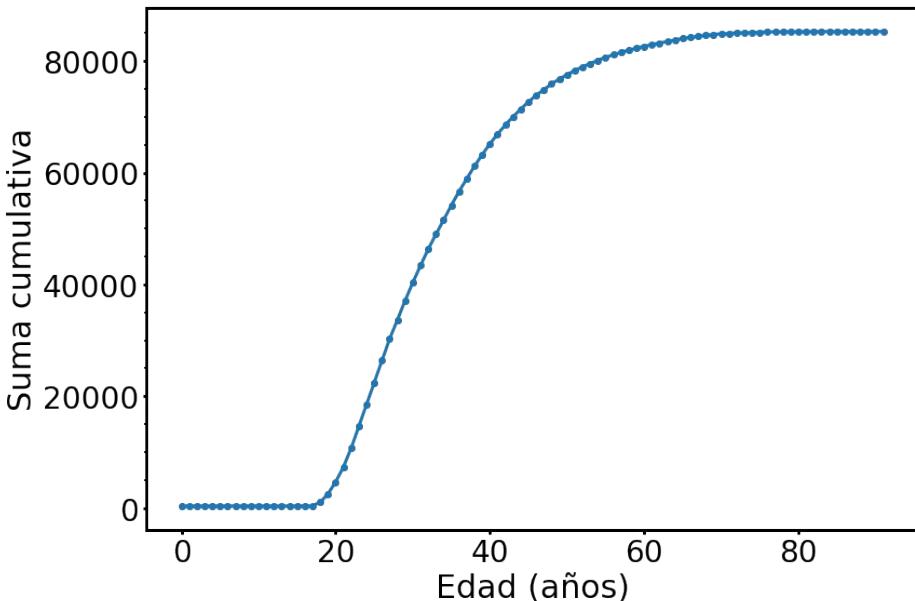


# Población Penitenciaria de Argentina 2017

## Edad



outliers!



**Eliminamos las filas con 0 años de edad como dato**  
(pueden ser hijos/as de internos o datos mal cargados)

# Población Penitenciaria de Argentina 2017

## Edad

Cantidad total de internos:

84880

Estadísticos de la distribución:

Moda: 26

Media:  $\approx$  34

Desviación estándar:  $\approx$  11

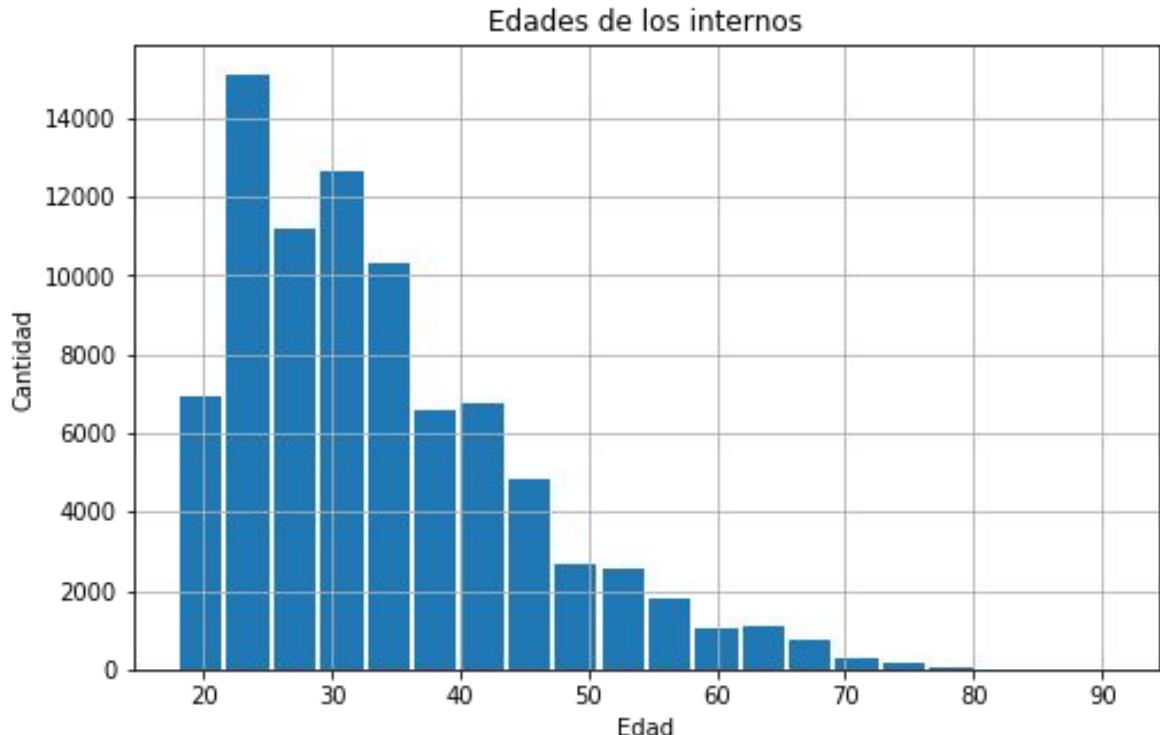
Valor mínimo: 18

1er cuartil: 25

2do cuartil (mediana): 31

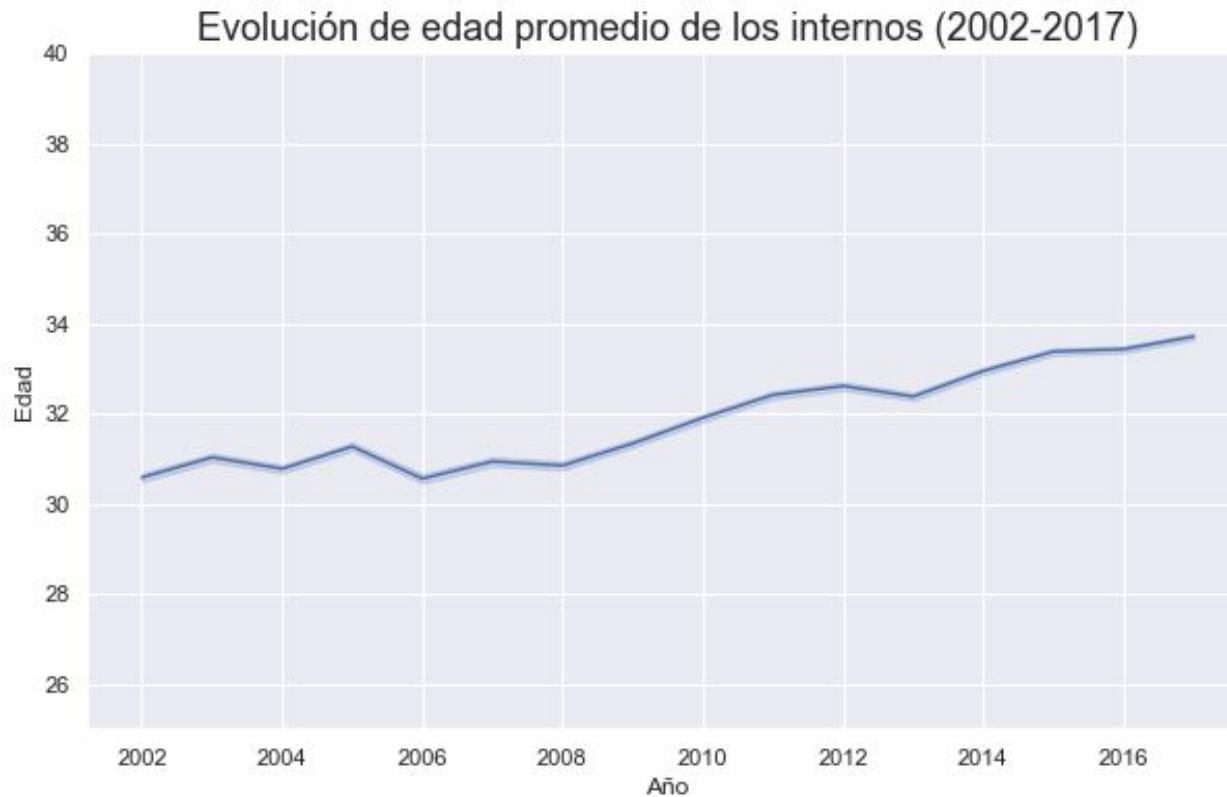
3er cuartil: 40

Valor máximo: 91



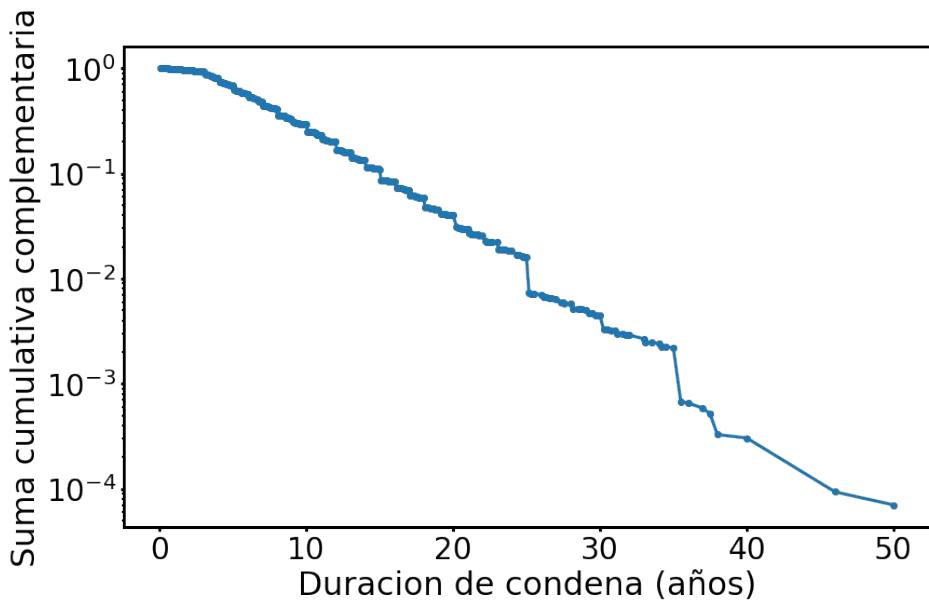
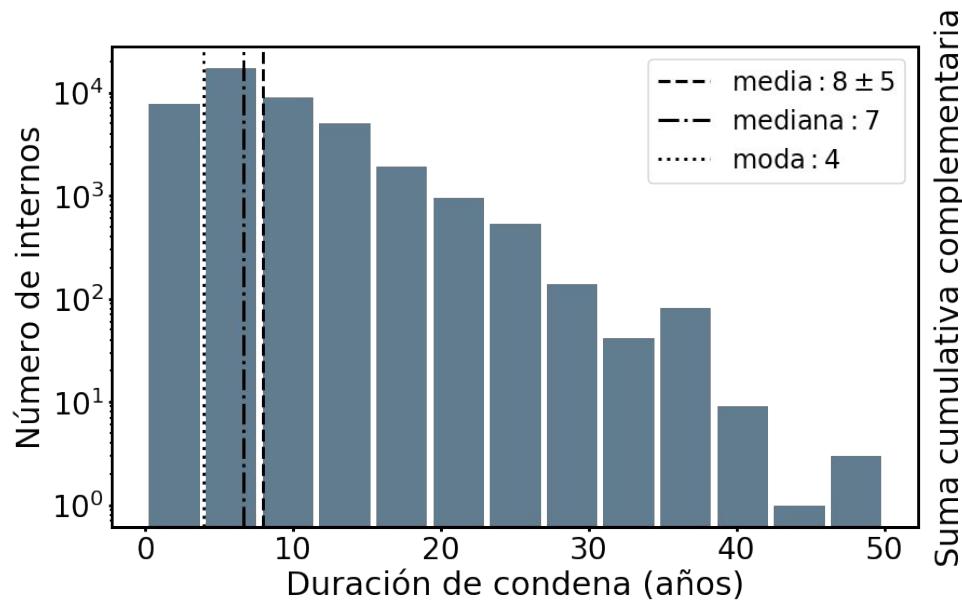
# Evolución temporal

## Edad



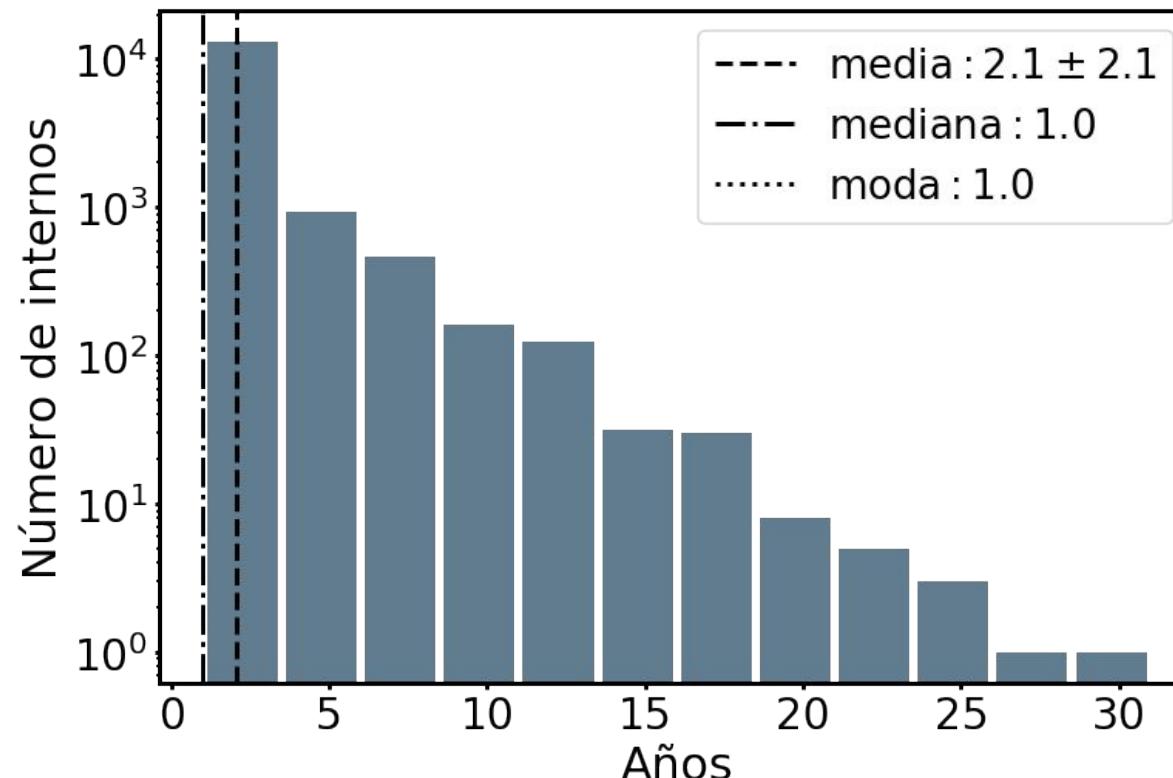
# Población Penitenciaria de Argentina 2017

## Duración condena



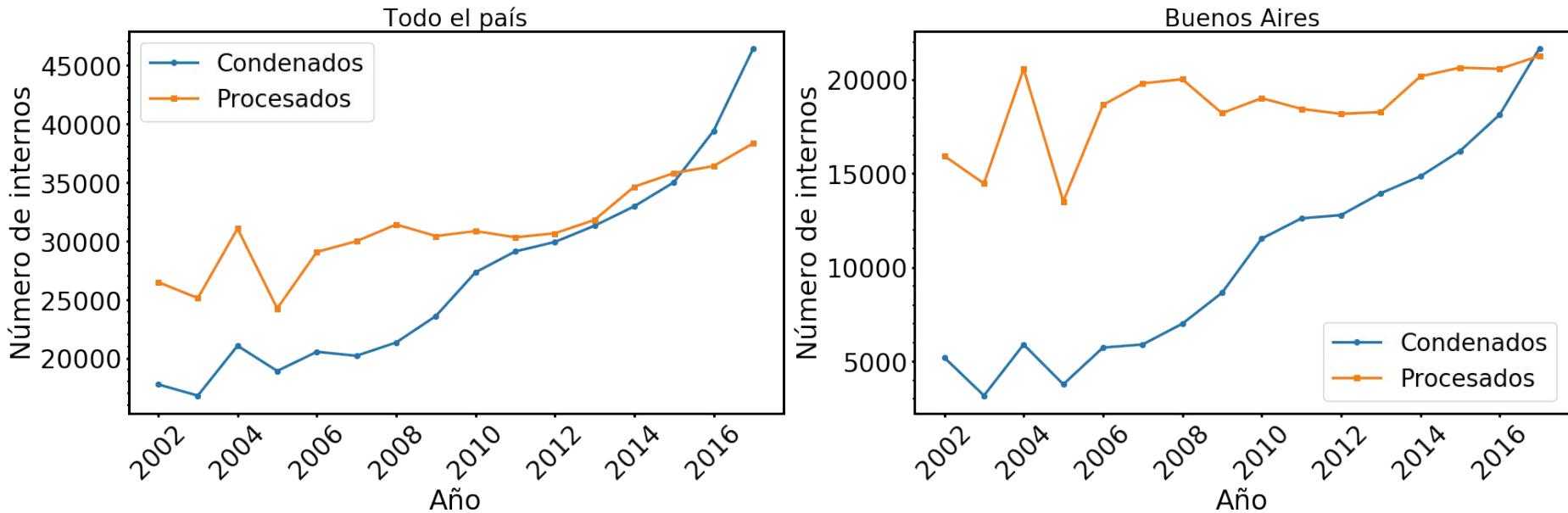
# Población Penitenciaria de Argentina 2017

## Diferencia fecha condena y fecha detención



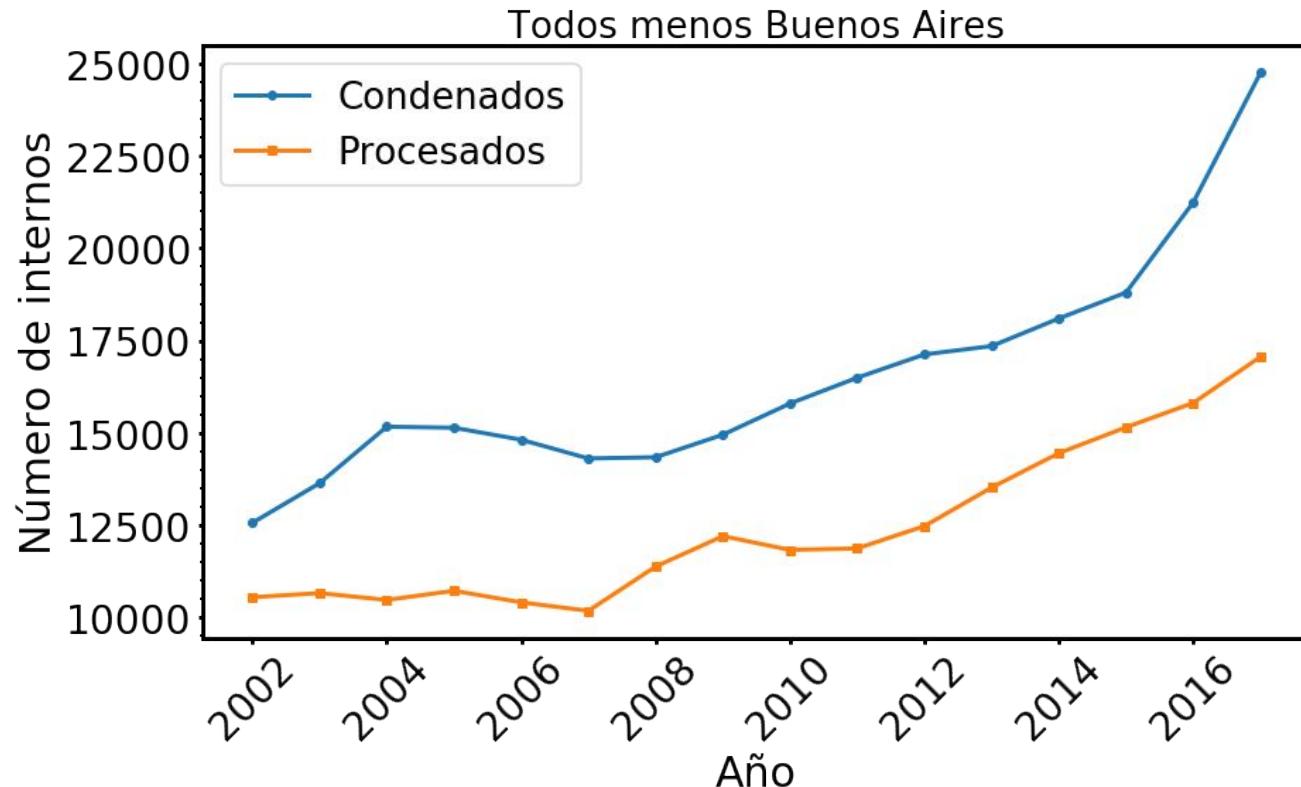
# Evolución temporal

## Situación Legal



# Evolución temporal

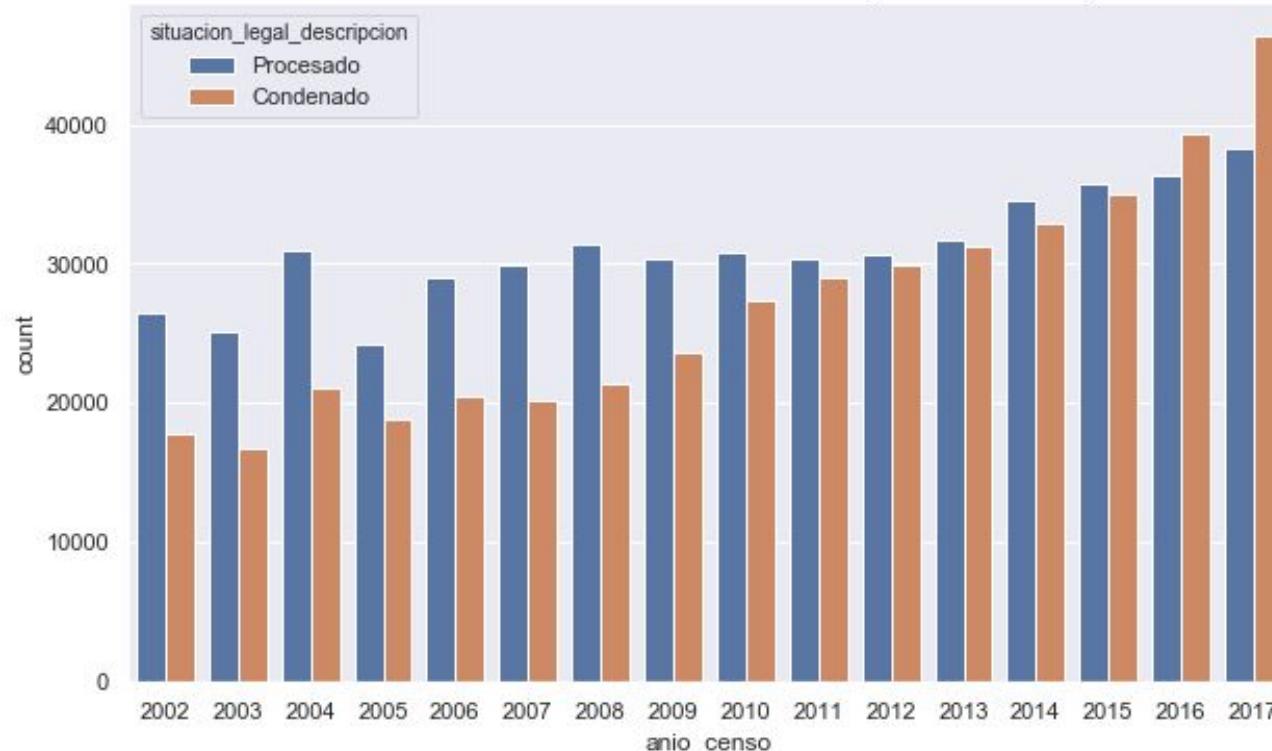
## Situación Legal



# Evolución temporal

## Situación Legal

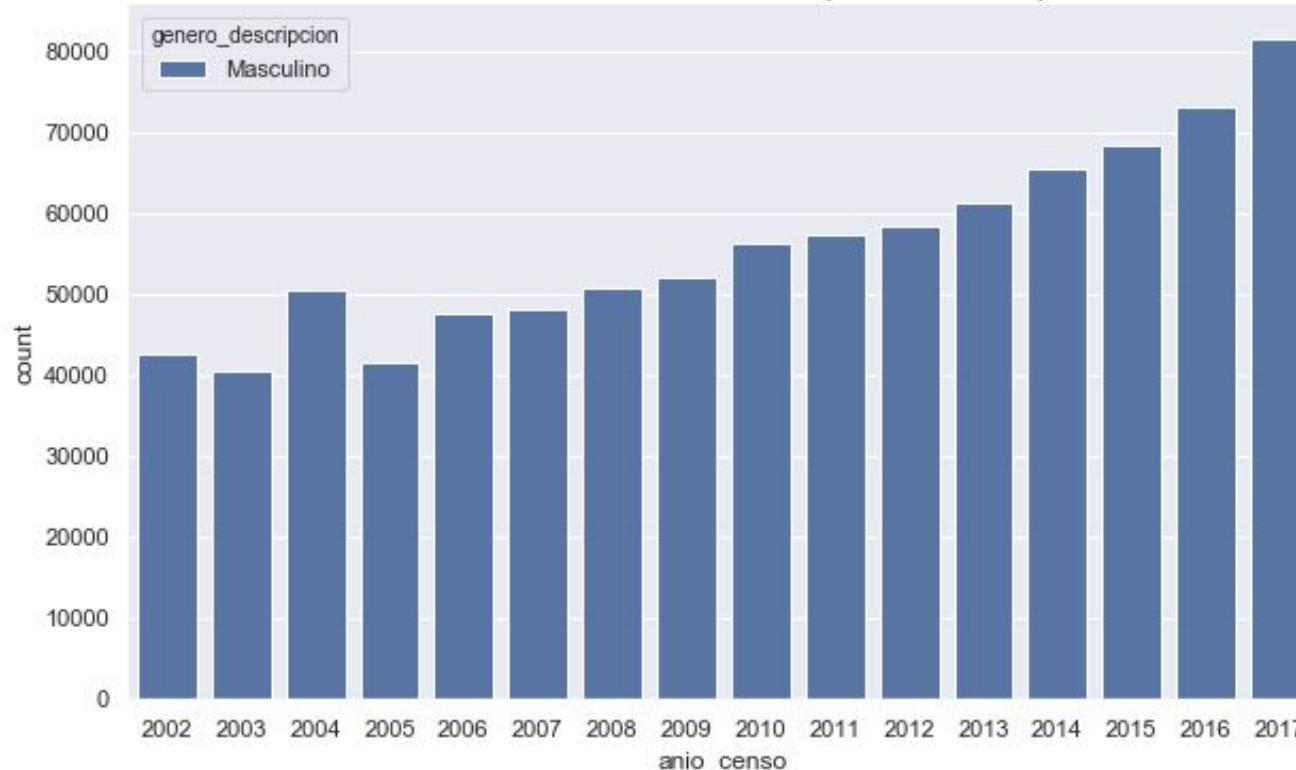
Condenados vs. Procesados (2002-2017)



# Evolución Temporal

## Género

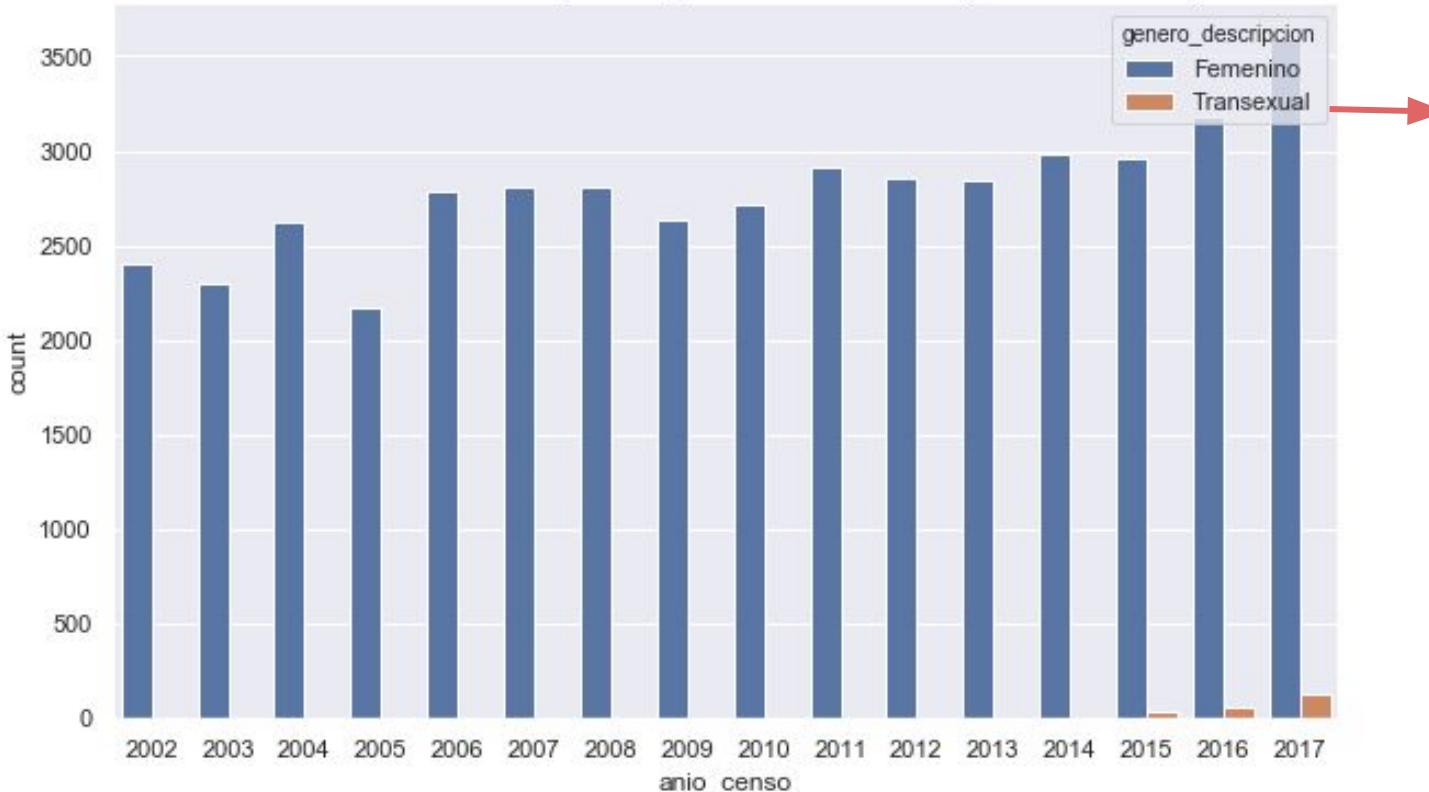
Cantidad de hombres (2002-2017)



# Evolución Temporal

## Género

Cantidad de mujeres y transexuales (2002-2017)



a partir del año 2015 el SNEEP comienza a incluir en su registro a la población trans, cambiando la denominación 'Sexo' por la de 'Género'