



SEMINARIO DE TRABAJO FINAL

Profesor: Alejandro Mario Hernandez

Alumno: Cristian Fabián Pesse

ÁREA: SEGURIDAD

TEMA:

Estadísticas criminales en la República Argentina por provincias

Problema:

El problema a resolver es la necesidad de comprender y abordar la incidencia delictiva en Argentina.

Solución:

Realizar un análisis detallado de los datos para informar el desarrollo de políticas y programas de seguridad pública efectivos.

ARCHIVOS ADJUNTOS:

Csv drive:

<https://drive.google.com/file/d/1frJxI5rWDak9KqdB8wbKI2xkd3PBiWFr/view?usp=sharing>

Google colab:

<https://colab.research.google.com/drive/1W49DBtGdsIVedeQNYp2feH1cQivNJhKs>

Fuente del dataset:

https://datos.gob.ar/dataset/seguridad-snic---provincial-estadisticas-criminales-republica-argentina-por-provincias/archivo/seguridad_2.2

Página Web del proyecto:

<https://cfp-dev8.github.io/seminario.github.io/>

PASOS A SEGUIR:

Obtener el conjunto de datos:

Se buscó fuentes confiables que proporcionen estadísticas criminales desglosadas por provincias en Argentina.

En la exploración de sitios web gubernamentales, como los portales de datos abiertos del Ministerio de Seguridad o de las agencias de estadísticas nacionales o provinciales.

También se buscó en organizaciones internacionales y en instituciones académicas que recopilan datos sobre delitos en Argentina.

Se decidió usar un dataset que encontré en datos.gob.ar:

https://datos.gob.ar/dataset/seguridad-snic---provincial-estadisticas-criminales-republica-argentina-por-provincias/archivo/seguridad_2.2

Evaluación la calidad de los datos:

Antes de comenzar el análisis, es importante evaluar la calidad y la integridad del conjunto de datos.

Así que se revisó la documentación asociada con los datos para comprender cómo se recopilaban, cuáles son las definiciones de los diferentes tipos de delitos, y si hay alguna limitación o sesgo conocido en los datos.

Exploración de la estructura del conjunto de datos:

Primero se hace una exploración con la estructura del conjunto de datos, examinando las variables disponibles, cómo están registrados los datos y cualquier metadato asociada. Esto ayudará a comprender la información que se tiene a disposición y a planificar el análisis.

Análisis para ver las primeras filas del dataset:

index	provincia_id	provincia_nombre	año	codigo_delito_snic_id	codigo_delito_snic_nombre	cantidad_hechos	cantidad_victimas	cantidad_victimas_masc	cantidad_victimas_fem	cantidad_victimas_sd	tasa_hechos	tasa_victimas	tasa_victimas_masc	tasa_victimas_fem
0	2	Ciudad Autónoma de Buenos Aires	2000	1	Homicidios dolosos	149	158.0	121.0	37.0	0.0	4,9750013	5,2755051	8,8257008	2,2703601
1	2	Ciudad Autónoma de Buenos Aires	2000	2	Homicidios dolosos en grado de tentativa	54	58.0	40.0	18.0	0.0	1,8030207	1,9365778	2,9175068	1,1083914
2	2	Ciudad Autónoma de Buenos Aires	2000	3	Muertes en accidentes viales	121	126.0	97.0	29.0	0.0	4,040102	4,2070404	7,0751481	1,7057418
3	2	Ciudad Autónoma de Buenos Aires	2000	4	Homicidios culposos por otros hechos	35	36.0	20.0	13.0	3.0	1,1686245	1,2020137	1,4587934	80050492
4	2	Ciudad Autónoma de Buenos Aires	2000	5	Lesiones dolosas	13154	13732.0	8105.0	5405.0	142.0	439,20248	458,50146	591,17603	337,7515

Luego veo información sobre columnas, tipos de datos, etc.

```

1 print(df.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21344 entries, 0 to 21343
Data columns (total 14 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   provincia_id                         21344 non-null  int64
1   provincia_nombre                     21344 non-null  object
2   año                                  21344 non-null  int64
3   codigo_delito_snic_id                21344 non-null  object
4   codigo_delito_snic_nombre            21344 non-null  object
5   cantidad_hechos                      21344 non-null  int64
6   cantidad_victimas                    7608 non-null   float64
7   cantidad_victimas_masc                7608 non-null   float64
8   cantidad_victimas_fem                 7608 non-null   float64
9   cantidad_victimas_sd                  7608 non-null   float64
10  tasa_hechos                          21344 non-null  object
11  tasa_victimas                        7608 non-null   object
12  tasa_victimas_masc                    7608 non-null   object
13  tasa_victimas_fem                     7608 non-null   object
dtypes: float64(4), int64(3), object(7)
memory usage: 2.3+ MB
None

```

Con la función describe() podemos ver algunas de las estadísticas, para ver el promedio, desviación estándar, mínima, máxima, primer quartir, etc.

	provincia_id	anio	cantidad_hechos	cantidad_victimas	\
count	21344.000000	21344.000000	21344.000000	7608.000000	
mean	48.008996	2012.676162	1696.281906	1012.169558	
std	27.685422	6.873185	6509.622936	3848.662761	
min	2.000000	2000.000000	0.000000	0.000000	
25%	25.000000	2007.000000	6.000000	19.000000	
50%	48.000000	2014.000000	80.000000	84.000000	
75%	71.000000	2019.000000	638.250000	420.250000	
max	94.000000	2022.000000	143840.000000	68450.000000	

	cantidad_victimas_masc	cantidad_victimas_fem	cantidad_victimas_sd
count	7608.000000	7608.000000	7608.000000
mean	425.355021	356.239353	230.576630
std	1857.924786	1611.364764	1705.029415
min	0.000000	0.000000	0.000000
25%	1.000000	0.000000	0.000000
50%	18.000000	13.000000	2.000000
75%	109.000000	124.000000	37.000000
max	29567.000000	41317.000000	61164.000000

Interpretación de los resultados:

Cantidad de hechos: La cantidad media de hechos delictivos por entrada es de aproximadamente 1031.68, con una gran variabilidad evidente en los datos. El valor mínimo es 0, lo que sugiere que algunas entradas no tienen registros de delitos. El valor máximo alcanza los 37599, lo que indica la presencia de entradas con un número significativamente alto de hechos delictivos.

Cantidad de víctimas: La cantidad media de víctimas por entrada es de alrededor de 98.93. Al igual que con la cantidad de hechos, hay una gran variabilidad en los datos, con valores que van desde 0 hasta 6099. Nuevamente, la presencia de valores mínimos en 0 sugiere que algunas entradas no tienen víctimas registradas.

Cantidad de víctimas masculinas y femeninas: Se observa una mayor cantidad media de víctimas masculinas (62.78) en comparación con las víctimas femeninas (35.90) por entrada. La variabilidad en los datos es evidente en ambas variables.

Cantidad de víctimas sin distinguir sexo: La cantidad media de víctimas sin distinguir sexo es muy baja, con un promedio de 0.24 por entrada. Esto sugiere que la mayoría de las entradas tienen víctimas registradas con un género específico.

Tasas de hechos y víctimas: Las tasas medias de hechos y víctimas por entrada son 51.18 y 4.57 respectivamente. Estas cifras proporcionan una medida de la incidencia de delitos y víctimas en relación con el número de entradas.

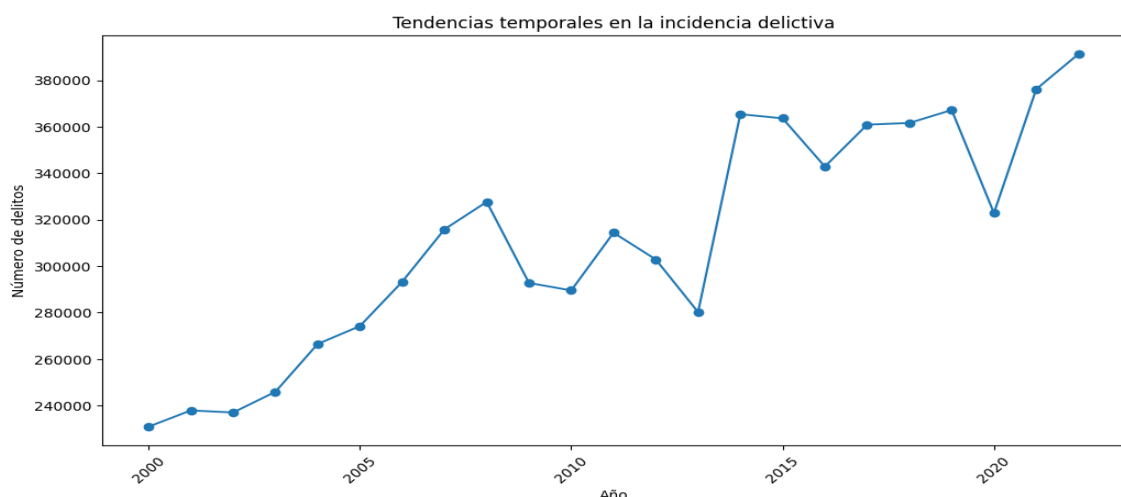
Tasas de víctimas masculinas y femeninas: Las tasas medias de víctimas masculinas y femeninas son 2.90 y 1.65 respectivamente. Esto indica la incidencia de víctimas masculinas y femeninas en relación con el número de entradas.

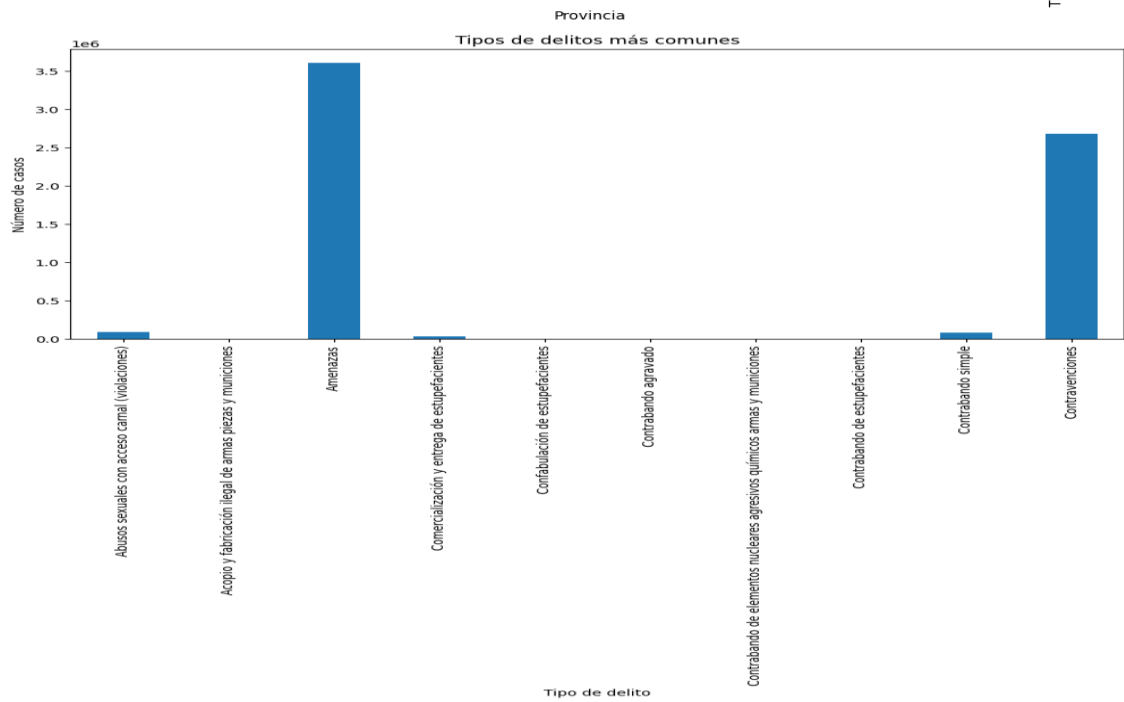
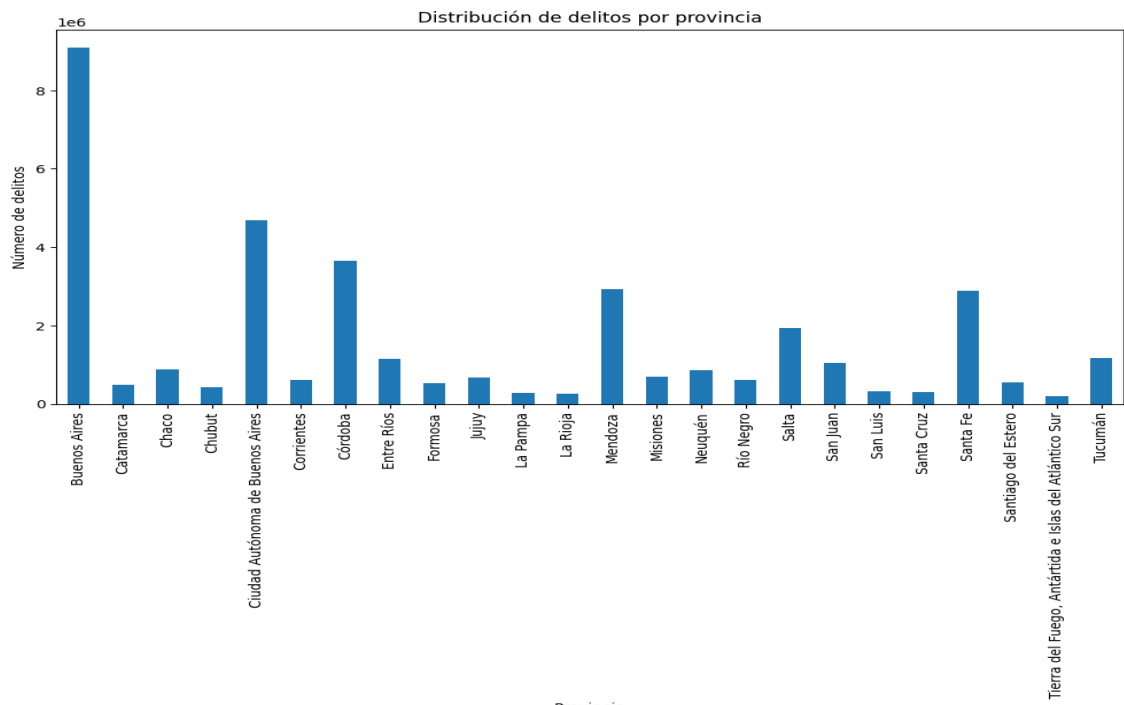
Estas estadísticas descriptivas proporcionan una visión general de la distribución y características de los datos, lo que es fundamental para comprender el conjunto de datos antes de realizar análisis más detallados y específicos.

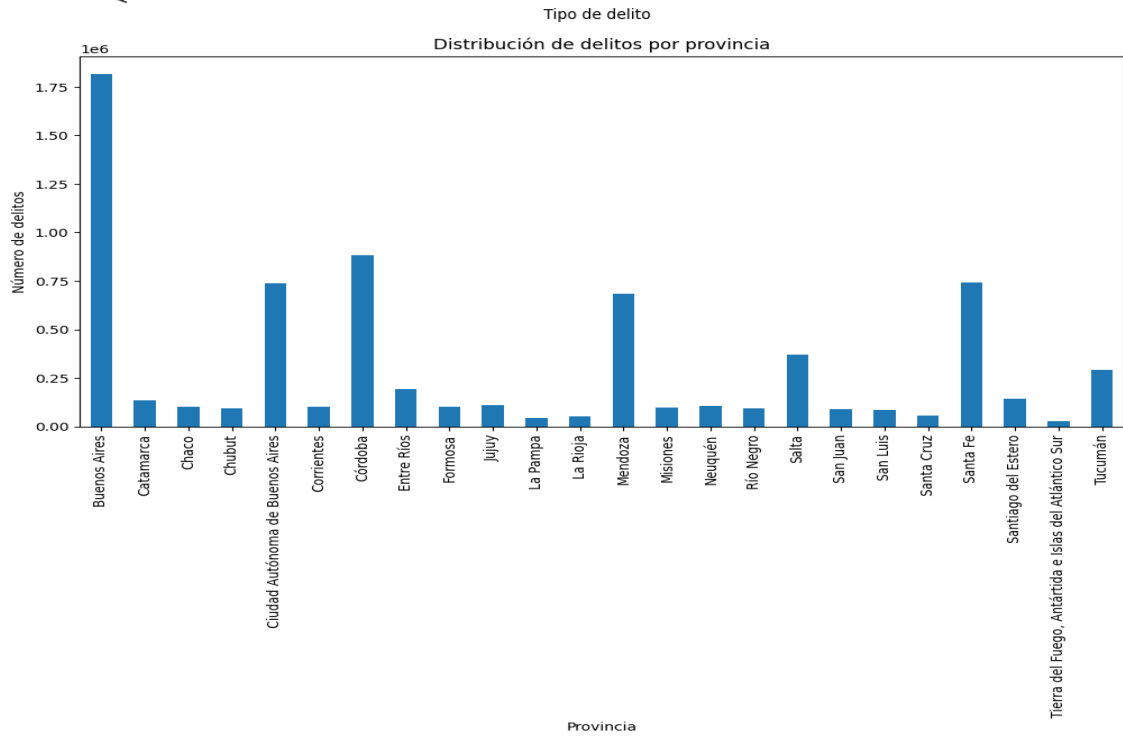
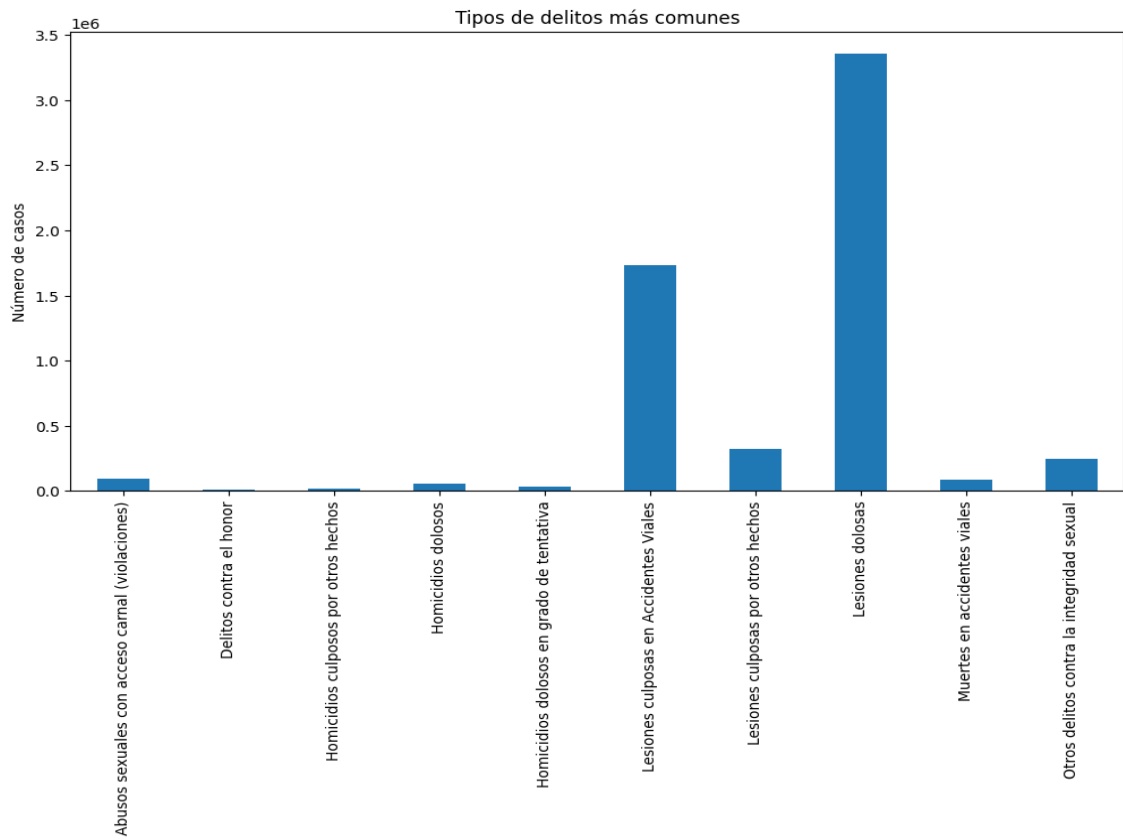
GRAFICOS DE BARRA:

Con los gráficos podemos ver las provincias con más delitos, el tipo de delito también podemos incluir algunos gráficos para saber sobre los horarios de los delitos.

También podemos ver los delitos por años por ejemplo:







CONCLUSIONES:

1. **Provincia con mayor incidencia delictiva:**

Se identificó que la provincia de Buenos Aires presenta la mayor cantidad de delitos registrados en el conjunto de datos, con un total de 1,817,159 delitos reportados.

Esto sugiere la necesidad de medidas adicionales de seguridad y prevención del delito en esta región para abordar la alta incidencia delictiva.

2. **Necesidad de datos adicionales:**

Para futuros análisis, es importante considerar la recopilación y el análisis de datos adicionales, como la hora exacta de los delitos y otros factores socioeconómicos y demográficos relevantes.

Estos datos complementarios podrían proporcionar una comprensión más completa de los patrones delictivos y ayudar a informar políticas y programas de seguridad más efectivos.

3. **Enfoque en la prevención del delito:**

Dada la alta incidencia delictiva en la provincia de Buenos Aires y en otras regiones identificadas en el análisis, es crucial adoptar un enfoque integral de prevención del delito que aborde las causas subyacentes y promueva la colaboración entre el gobierno, la sociedad civil y otras partes interesadas. Esto podría incluir iniciativas de intervención temprana, programas de rehabilitación, mejoras en la iluminación pública y el aumento de la vigilancia policial en áreas de alto riesgo.

4. **Variabilidad geográfica en la incidencia delictiva:**

Además de Buenos Aires, se observa una variabilidad significativa en la cantidad de delitos registrados en otras provincias. Esto sugiere que los factores contextuales y socioeconómicos pueden influir en los niveles de criminalidad en diferentes regiones.

5. **Enfoque en la colaboración interinstitucional:**

Dada la complejidad de los desafíos relacionados con la seguridad pública, es esencial promover las colaboraciones entre diversas instituciones y actores, incluidas las fuerzas del orden, las autoridades gubernamentales, las organizaciones comunitarias y los expertos en políticas públicas. Una coordinación efectiva entre estas entidades puede facilitar la implementación de estrategias integrales de prevención del delito y mejorar la respuesta ante situaciones de emergencia.