

ProyectoAED2024

- ¹ Universitat de València. ETSE.; alcaluis@alumni.uv.es
² Universitat de València. ETSE.; carigar4@alumni.uv.es
³ Universitat de València. ETSE.; jopebrui@alumni.uv.es

Simple Summary: Estudio de los delitos causados por menores en España entre 2013 y 2023.

Abstract: Proyecto de tratamiento de datos para la asignatura de Análisis exploratorio de datos. El tópico es la delincuencia en España provocado por menores de entre 14 y 17 años (ambos inclusivos). Donde se estudiarán los delitos causados con mayor frecuencia.

Keywords: Delito; Delincuencia; Menores; Datos; Ciencia de Datos;

1. Introducción

1.1. Definición proyecto y planteamiento de preguntas (objetivos)

WIP: Definir el problema y los objetivos del estudio

1.2. Carga de librerías y datos

```
# WIP (TODOS):  
# Carga de librerías  
librerias <- c("readr",      # Lectura de ficheros con formato  
              "dplyr",      # Gramática de manipulación de datos  
              "ggplot2")    # Visualización mediante gráficas elegantes  
pacman::p_load(char=librerias)  
  
# Carga de datos  
col_names <- c("Delitos N1", "Delitos N2", "Delitos N3",  
              "Delitos N4", "Delitos N5", "Edad",  
              "Año", "Total Delitos")  
  
delitos_menores_raw <- read_delim(  
  "../data/delitos_menores_2013_2023.csv",  
  delim = ";",  
  escape_double = FALSE,  
  trim_ws = TRUE,          # Espacios y tabulaciones eliminados.  
  show_col_types = FALSE,  # Omitir mensajes en carga de datos.  
  col_names = col_names,   # Si proporcionamos nombres de columnas h  
                           # saltarnos la primera fila (skip=1).  
  skip=1,  
  locale = locale(grouping_mark = ".")) # Locale nos permitirá leer debidamente  
                                         # los millares.
```

Citation: Albacete, L; Ribes, C; Pedro, J.
ProyectoAED2024. *Journal Not Specified* **2023**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2024 by the author.
Submitted to *Journal Not Specified*
for possible open access publication
under the terms and conditions
of the Creative Commons Attribution
(CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1.3. Estudio y acondicionamiento de los datos

WIP (TODOS, #01-Introducción): - Análisis de la estructura y valores iniciales de los datos. - Acondicionamiento y preparación de los datos. - Validación datos acondicionados

1.3.1. Análisis de la estructura y valores iniciales de los datos.

13

```
# Análisis de la estructura y valores iniciales de los datos.
summary(delitos_menores_raw)
```

```
## Delitos N1      Delitos N2      Delitos N3      Delitos N4      14
## Length:2915    Length:2915    Length:2915    Length:2915    15
## Class :character Class :character Class :character Class :character 16
## Mode  :character Mode  :character Mode  :character Mode  :character 17
##                                                    18
##                                                    19
##                                                    20
##                                                    21
## Delitos N5      Edad      Año      Total Delitos  22
## Length:2915    Length:2915    Min.   :2013    Min.   :    0.0  23
## Class :character Class :character 1st Qu.:2015    1st Qu.:   17.0  24
## Mode  :character Mode  :character Median :2018    Median :   78.0  25
##                                                    Mean  :2018    Mean  :  756.2  26
##                                                    3rd Qu.:2021   3rd Qu.:  410.0  27
##                                                    Max.   :2023   Max.   :26349.0  28
##                                                    NA's   :210      29
```

```
str(delitos_menores_raw)
```

```
## spc_tbl_ [2,915 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame) 30
## $ Delitos N1 : chr [1:2915] "Total Infracciones" "Total Infracciones" "Total 31
## $ Delitos N2 : chr [1:2915] NA NA NA NA ... 32
## $ Delitos N3 : chr [1:2915] NA NA NA NA ... 33
## $ Delitos N4 : chr [1:2915] NA NA NA NA ... 34
## $ Delitos N5 : chr [1:2915] NA NA NA NA ... 35
## $ Edad       : chr [1:2915] "Total" "Total" "Total" "Total" ... 36
## $ Año        : num [1:2915] 2023 2022 2021 2020 2019 ... 37
## $ Total Delitos: num [1:2915] 23662 25822 26349 20366 26049 ... 38
## - attr(*, "spec")= 39
## .. cols( 40
## .. 'Delitos N1' = col_character(), 41
## .. 'Delitos N2' = col_character(), 42
## .. 'Delitos N3' = col_character(), 43
## .. 'Delitos N4' = col_character(), 44
## .. 'Delitos N5' = col_character(), 45
## .. Edad = col_character(), 46
## .. Año = col_double(), 47
## .. 'Total Delitos' = col_number() 48
## .. ) 49
## - attr(*, "problems")=<externalptr> 50
```

Gracias a “summary” podemos apreciar qué columnas contienen NAs y si los valores (de las columnas numéricas) entran dentro de lo esperado (min, max, median, ...).

Teniendo para la columna **Año** valores esperados, desde 2013 hasta 2023 y la media obviamente en 2018. Respecto a la columna **Total**, podemos apreciar que tiene un número importante de valores faltantes y confirmamos que los valores se han importado correctamente (manteniendo los millares).

Además hemos identificado que la columna **Edad** se ha importado como tipo “character”. Tras hacer una visualización del archivo nos damos cuenta que los años se han

importado como “14 años” agregando el sustantivo para todos los valores y en algunos casos agregando un “Total”. Acondicionamiento de la columna será necesario.

Las columnas restantes, que son de tipo “character”, representan el delito en sus distintos niveles. Habremos de factorizar estas variable categóricas.

El uso de la función “str” en este caso no nos ha aportado información adicional.

```
# 1. Delitos N1
# Esta columna no nos aporta valor alguno al solo tener un valor.
# Más adelante será eliminada.
unique(delitos_menores_raw$'Delitos N1')
```

```
## [1] "Total Infracciones"
```

```
# 2. Delitos N2
# Esta columna nos servirá para hacer el primer filtrado y quedarnos
# con solo los delitos cometidos y no las faltas o los totales calculados.
unique(delitos_menores_raw$'Delitos N2')
```

```
## [1] NA          "A Delitos" "B Faltas"
```

```
# 3. Delitos N3, N4, N5
```

```
# 4. Edad
```

```
# 5. Año
```

```
# 6. Total
```

2. Análisis

2.1. Análisis de las variables

```
# WIP:
# - Análisis de las variables (individualmente, explicación de los datos,
#                               estadísticos.) (CARLOS)
# - Análisis de las variables (relaciones) (JOAN)
#
#
# - NOTA: Se debe hacer uso de una "Visualización temprana (exploratoria)"
#         de los datos.
# - NOTA: Junto a la visualización se debe aportar una explicación.
```

2.2. Análisis de ‘outliers’, métodos de imputación aplicados y análisis de datos perdidos

```
# WIP:
# - Análisis de outliers, métodos de imputación y datos perdidos.
# (si procede) (LUIS)
```

2.3. Visualización Explicatoria

```
# WIP:
# - Visualización enfocada a los objetivos (encontrar una respuesta al problema)
#
# - NOTA: Junto a la visualización se debe aportar una explicación.
```

3. Conclusiones70

(TODOS) WIP: - Conclusiones finales. - Objetivos. - Proyecto. - NOTA: Visualizaciones71
interactivas para finalizar/resumir, puede ser un buen punto.72

Author Contributions: WIP. En que hemos contribuido cada uno.73

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual74
author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to75
people or property resulting from any ideas, methods, instructions or products referred to in the content.76