

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)

La delincuencia en España es un tema de conversación latente. Aunque no lo es tanto, pero sigue siendo importante, es hablar de la edad del delincuente. En este estudio analizaremos los diferentes tipos de delito ocasionados por menores de entre 14 y 17 años durante los 10 últimos años. Los objetivos del estudio es responder a las siguientes preguntas:

1. ¿Cuáles son los delitos más frecuentes y que edades lo frecuentan? ¿Siguen siendo recientes o han disminuido?
2. Analizar los picos de delitos y relacionarlos con hechos reales. (Buscar sucesos, leyes)
3. Encontrar tendencias entre los años.
4. ...

WIP: Definir el problema y los objetivos del estudio

1.2. Carga de librerías y datos

```
# 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(
  file = "../data/delitos_menores_2013_2023.csv",
  delim = ";",
  escape_double = FALSE,
  trim_ws = TRUE,
  show_col_types = FALSE,
  # Espacios y tabulaciones eliminados.
  # Omitir mensajes en carga de datos.
```

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/>).

```
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.
```

1.3. Estudio y acondicionamiento de los datos

WIP (TODOS, #01-Introduccion): - 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.

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

```
##   delitos_n1      delitos_n2      delitos_n3      delitos_n4
## Length:2915      Length:2915      Length:2915      Length:2915
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##   delitos_n5      edad      año      total_delitos
## Length:2915      Length:2915      Min.   :2013      Min.   :    0.0
## Class :character Class :character 1st Qu.:2015      1st Qu.:   17.0
## Mode  :character Mode  :character Median :2018      Median :   78.0
##                                     Mean  :2018      Mean  :  756.2
##                                     3rd Qu.:2021      3rd Qu.:  410.0
##                                     Max.   :2023      Max.   :26349.0
##                                     NA's    :210
```

```
# str(delitos_menores_raw) # STR no nos aporta información adicional
# glimpse(delitos_menores_raw) # Lo mismo sucede con GLIMPSE
```

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_delitos**, podemos apreciar que tiene un número relevante de valores faltantes (210) 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” o “glimpse” en este caso no nos ha aportado información adicional.

```
# 1. delitos_n1
table(delitos_menores_raw$delitos_n1, useNA = "always")
```

```
##
```

```
## Total Infracciones      <NA>
##           2915           0
```

```
# 2. delitos_n2
table(delitos_menores_raw$delitos_n2, useNA = "always")
```

```
##
## A Delitos B Faltas      <NA>
##    2585    275      55
```

```
# 3. delitos_n3, N4, N5
tail(table(delitos_menores_raw$delitos_n3, useNA = "always"), 2)
```

```
##
## R Resto de delitos      <NA>
##           55          165
```

```
tail(table(delitos_menores_raw$delitos_n4, useNA = "always"), 2)
```

```
##
## 8.99 Otros delitos contra la libertad e indemnidad sexuales
##                                     55
##                                     <NA>
##                                     1045
```

```
tail(table(delitos_menores_raw$delitos_n5, useNA = "always"), 2)
```

```
##
## 22.2.2 Resistencia y desobediencia      <NA>
##           55          2695
```

```
# 4. edad
table(delitos_menores_raw$edad, useNA = "always")
```

```
##
## 14 años 15 años 16 años 17 años  Total      <NA>
##    583    583    583    583    583      0
```

```
# 5. año
table(delitos_menores_raw$año, useNA = "always")
```

```
##
## 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 <NA>
##   265  265  265  265  265  265  265  265  265  265  265   0
```

De las diferentes columnas podemos decir que:

- **delitos_n1:** la primera columna no nos aporta valor alguno al solo tener un valor. Más adelante será eliminada.
- **delitos_n2:** esta columna nos servirá para hacer el primer filtrado y quedarnos con solo los delitos y faltas cometidas. Omitiendo los totales calculados. Sus valores pueden ser reemplazados en valores más claros.

- **delitos_n3, n4, y n5:** ayudan a clasificar los delitos en subtipos de infracción. Sin embargo, podemos ver como a mayor nivel de detalle más valores faltantes. Esto se debe a que un nivel de clasificación tan preciso no es necesario para describir todos los delitos.

La columna **delitos_n5** no será necesaria. Aporta muy poca información y el análisis no requerirá tanto nivel de detalle para una infracción.

- **edad:** todos los datos de esta columna son uniformes y esperados (ningún valor faltante). Finalmente se eliminarán aquellas instancias que representen un total y se convertirá la columna a tipo numérico.

```
# 6. total_delitos
df_total_delitos <- delitos_menores_raw %>%
  filter(startsWith(delitos_n3, "13")) %>%
  filter(edad=="14 años") %>%
  filter(año==2013) %>%
  mutate(delitos_n4 = ifelse(!is.na(delitos_n4),
                             substring(delitos_n4, 1, 4),
                             NA))

knitr::kable(df_total_delitos[c(4,5,8)], format = "latex",
             booktabs = TRUE,
             caption = "Caso delitos 'Contra el patrimonio y el orden
                        socioeconómico' en 2013 por delincuentes de 14 años",
             align = 'l', centering = TRUE,
             table.envir = "table", position = "H")
```

Table 1. Caso delitos ‘Contra el patrimonio y el orden socioeconómico’ en 2013 por delincuentes de 14 años

delitos_n4	delitos_n5	total_delitos
NA	NA	1433
13.1	NA	142
13.2	NA	1108
13.2	13.2.1 Robo con violencia	602
13.2	13.2.2 Robo con fuerza	506
13.4	NA	55
13.5	NA	9
13.6	NA	8
13.9	NA	97
13.1	NA	13
13.9	NA	1

- **total:** respecto a esta columna pese a que no hayamos utilizado ninguna instrucción, hemos mediante visualización del fichero analizado un caso en concreto para definir la naturaleza de esta columna. Ya vimos anteriormente, gracias a otras columnas, que hay agrupaciones internas de los datos calculando los totales. Pero eso no es todo. Podemos apreciar que el total de un delito, es la suma de todas sus subclasificaciones.

Es decir, el total de “13.2 Robos”, es igual a la suma de “13.2.1” y “13.2.2”, a su vez el delito contra el patrimonio es la suma de los delitos “13.1”, “13.2”, ... Este hecho agrega redundancia a los datos. Por ende no solo tendremos que eliminar las filas que representen

totales pero también aquellas filas que son las agrupaciones de otras con el objeto de tener un conjunto **Tidy**.

En resumen, las columnas de delitos n1 y n5 no son necesarias y pueden ser eliminadas sin problemas. Hemos averiguado que los delitos de N5 se suman para los delitos de N4. Las columnas de delitos restantes serán convertidas a factor. La columna edad será modificada para mayor fidelidad al dato.

1.3.2. Acondicionamiento y preparación de los datos

```
df_delitos_menores <- delitos_menores_raw %>%
  # Retiramos las filas que se dividan/clasifiquen hasta el nivel 5
  filter(is.na(delitos_n5)) %>%

  # Selección de características/variables relevantes
  select(-c(delitos_n1, delitos_n5)) %>%

  # Renombramiento de las columnas
  rename(delito_falta = delitos_n2,          # Delito o Falta
         clasi_infrac = delitos_n3,          # Primer nivel de clasificación
         sub_clasi_infrac = delitos_n4) %>% # Segundo nivel de clasificación

  # Filtrado, retiramos las columnas que ya están pre-calculadas.
  # Caso edad = "Total"
  # Caso "nivel_n1" era un "total" y los demás niveles de delito "NA"
  filter(edad != "Total") %>%
  filter(!is.na(delito_falta)) %>%

  # Reemplazo de valores
  mutate(
    # valores entendibles para "delito_falta"
    delito_falta = case_when(delito_falta == "A Delitos" ~ "delito",
                             delito_falta == "B Faltas" ~ "falta"),
    # Eliminamos "años" y pasamos a numérico
    edad = as.numeric(substring(edad, 1, 2))
  ) %>%

  # Convertimos a factores
  mutate(
    delito_falta = as.factor(delito_falta),
    clasi_infrac = as.factor(clasi_infrac),
    sub_clasi_infrac = as.factor(sub_clasi_infrac)
  )

# WIP. Eliminacion aquellas columnas que tienen subclasisificaciones.
```

1.3.3. Validación datos acondicionados

Para validar que los datos se han cargado correctamente vamos a sacar las columnas totales precaculadas y compararlas con

```
# Obtener los totales
df_totales <- delitos_menores_raw %>%
  filter(edad == "Total")
```

2. Análisis110

2.1. Análisis de las variables111

```
# 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 perdidos112

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

2.3. Visualización Explicatoria113

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

3. Conclusiones114

(TODOS) WIP: - Conclusiones finales. - Objetivos. - Proyecto. - NOTA: Visualizaciones115

interactivas para finalizar/resumir, puede ser un buen punto.116

Author Contributions: WIP. En que hemos contribuido cada uno.117

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual118

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 to119

people or property resulting from any ideas, methods, instructions or products referred to in the content.120