Periodismo de datos

# PEC3: Reportaje periodístico con datos [Encontrar el hilo argumental utilizando herramientas y aplicando habilidades].

Paula León Gil-Gibernau

Contenido

[PEC3: Reportaje periodístico con datos [Encontrar el hilo argumental utilizando herramientas y aplicando habilidades]. 1](#_Toc25769515)

[Metodología 3](#_Toc25769516)

[Extracción de los datos 4](#_Toc25769517)

[Análisis de los datos 5](#_Toc25769518)

[Visualizaciones interactivas 5](#_Toc25769519)

[Conclusiones y resultados 7](#_Toc25769520)

[Hilo argumental 7](#_Toc25769521)

## 

## Metodología

Para realizar el reportaje y que sea sencillo acceder a los datos originales y reproducir el análisis, se ha creado un repositorio público en GitHub con el siguiente enlace:

<https://github.com/pleongi/Periodismo-de-datos>.

En este repositorio se puede encontrar:

* **Código fuente:**
  + **code/analize\_data.ipynb:** El código fuente usado para extraer los datos y analizarlos con las explicaciones de casa paso.
  + **code/analize\_data.html:** La ejecución del código fuente usado para extraer los datos y analizarlos con las explicaciones de casa paso.
* **Dataset original:**
  + **data\_sources/25721.xls**: fichero Excel con el dataset original, que contiene los menores condenados anuales según el sexo, la edad y la nacionalidad de los años 2018 a 2013, se ha extraído del Instituto Nacional de Estadística, en el siguiente link: <https://www.ine.es/jaxiT3/Tabla.htm?t=25721&L=0> .
  + **data\_sources/9674.xls**: fichero Excel con el dataset original, que contiene la nacionalidad de la población de España, por sexo, fecha y grupo de edad y de los años 2019 a 2012. Se ha extraído del Instituto Nacional de Estadística, en el siguiente link: <https://www.ine.es/jaxiT3/Tabla.htm?t=9674&L=0> .
* **Datasets generados:**
  + **generated\_output/output\_condenas\_menores.xlsx**: fichero Excel con el dataset final que contiene los menores condenados anuales según el sexo, la edad y la nacionalidad de los años 2018 a 2013.
  + **generated\_output/output\_poblacion.xlsx**: fichero Excel con el dataset final que contiene la población de España con nacionalidad española, africana o total, por sexo y por grupo de edad menor de 19 años, y de los años 2019 a 2013.
* **Data visualizations**:
  + **tableau/Periodismo de datos\_ Menores Condenados.twbx:** fichero de tipo tableau con el que se ha creado el dashboard para la visualización de los menores condenados en España. Dicha visualización está publicada en Tableau Public en el siguiente enlace:

<https://public.tableau.com/shared/KPTZ74WGN?:display_count=y&:origin=viz_share_link>.

* + **tableau/La población en España.twbx**: fichero de tipo Tableau con el que se ha creado el dashboard para la visualización de la población en España. Dicha visualización está publicada en Tableau Public en el siguiente enlace: <https://public.tableau.com/views/LapoblacinenEspaa/Dashboard1?:display_count=y&publish=yes&:origin=viz_share_link>

Tal y como se muestra en los contenidos del repositorio se han creado dos visualizaciones con Tableau, y se han hecho públicas en Tableu Public para que todo el mundo pueda acceder.

## Extracción de los datos

Se han usado dos datasets, ambos extraídos de la página del Instituto Nacional de Estadística.

En concreto se han extraído:

* Un primer dataset en formato Excel, llamado 25721.xls, con el número de menores condenados en España, anualmente según el sexo, la edad y la nacionalidad de 2013 a 2018. El enlace a este es

<https://www.ine.es/jaxiT3/Tabla.htm?t=25721&L=0> .

* Un segundo dataset en formato Excel, llamado 9674.xls, con la nacionalidad de la población de España, por sexo, fecha y grupo de edad, de 2002 a 2019. El enlace a este es

<https://www.ine.es/jaxiT3/Tabla.htm?t=9674&L=0> .

El primer paso realizado en el script de Python **analize\_data.ipynb,** es cargar y preparar dichos datos para luego ser analizados.

Para el primer dataset, número de menores condenados en España, se han realizado los siguientes pasos:

1. En primer lugar cargar el Excel y crear un dataset con pandas.
2. Puesto que el formato del Excel era complejo de interpretar por Python, se han tenido que hacer diversas cosas como remplazar nombres, quitar filas vacías, y hacer transformaciones de filas a columnas…
3. El resultado final es un Excel, este Excel generado se encuentra también en el repositorio de GitHub. Contiene las siguientes features:
   1. **Age**: la edad de los menores condenados en dicha observación. Un número entero por debajo de 18.
   2. **Nationality**: la nacionalidad de los menores condenados en dicha observación. Española, País de la Unión Europea sin España, País del resto de Europa, América, África, Asia u Oceanía.
   3. Sex: el sexo de los menores condenados en dicha observación. Mujer u Hombre.
   4. **Year**: El año en el que se tomó la métrica de número de menores condenados.
   5. **Menores\_condenados**: el número de menores condenados.

Para el segundo dataset, la nacionalidad de la población en España, se han realizado los siguientes pasos:

1. En primer lugar cargar el Excel y crear un dataset con pandas.
2. En el caso del segundo dataset se ha seleccionado/filtrado únicamente aquellos datos relativos a menores de edad, en este caso nos hemos quedado con los grupos de edad inferiores a 19 años, y se han cogido solo aquellos datos entre 2019 y 2013, para compararlos con el otro dataset, y reducir el tamaño del mismo. Además, solo se han cogido los datos que aplican a la nacionalidad (país de origen) africana o española y la población total. Estos eran los datos indispensables y suficientes para hacer las comparativas y los análisis necesarios.
3. El resultado final es un Excel, este Excel generado se encuentra también en el repositorio de GitHub. Contiene las siguientes features:
   1. **Age**: la edad del recuento de la población en dicha observación. Menores de edad.
   2. **Nationality**: la nacionalidad de la población en dicha observación. Española, Africana o Total. Se entiende por total, la población total que cumple dichas features.
   3. **Sex**: el sexo del recuento de la población en dicha observación. Mujer u Hombre.
   4. **Date**: La fecha en la que se calculó el recuento de la población.
   5. **Poblacion**: el número de la población total para esa observación.

Para más detalles se puede consultar en el repositorio de GitHub, la ejecución y explicación de cada paso en el fichero **analize\_data.html.**

## Análisis de los datos

En primer lugar se ha realizado un analissi descriptivo visual.

Se han mostrado las variables cualitativas que pueden tomar pocos valores mediante diagramas de barras. Para evaluar variaciones en cuanto a al recuento de menores condenados. Parecen existir grandes variaciones en:

Las variables numéricas como la edad, que no toma demasiado valores también pueden representarse con un diagrama de barras. Donde están la mayoría de menores condenados?

No observamos outliers?

Correlaciones?

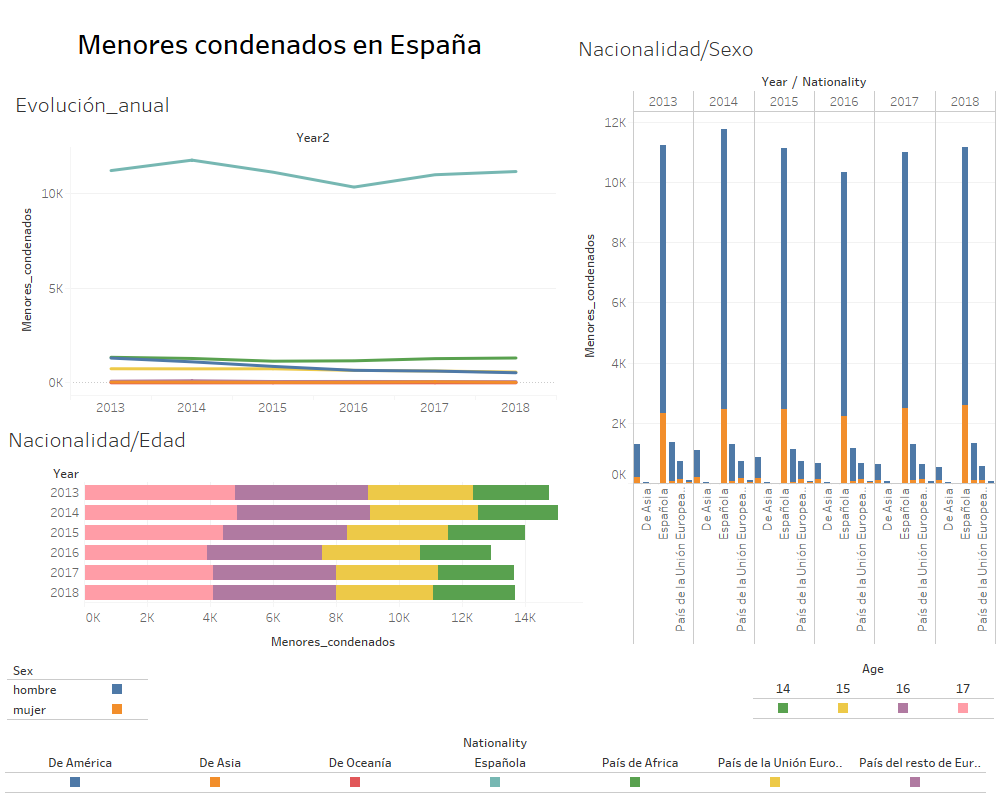
Max, min, mean, porcentaje, rango?

Poblacion según el sexo

## Visualizaciones interactivas

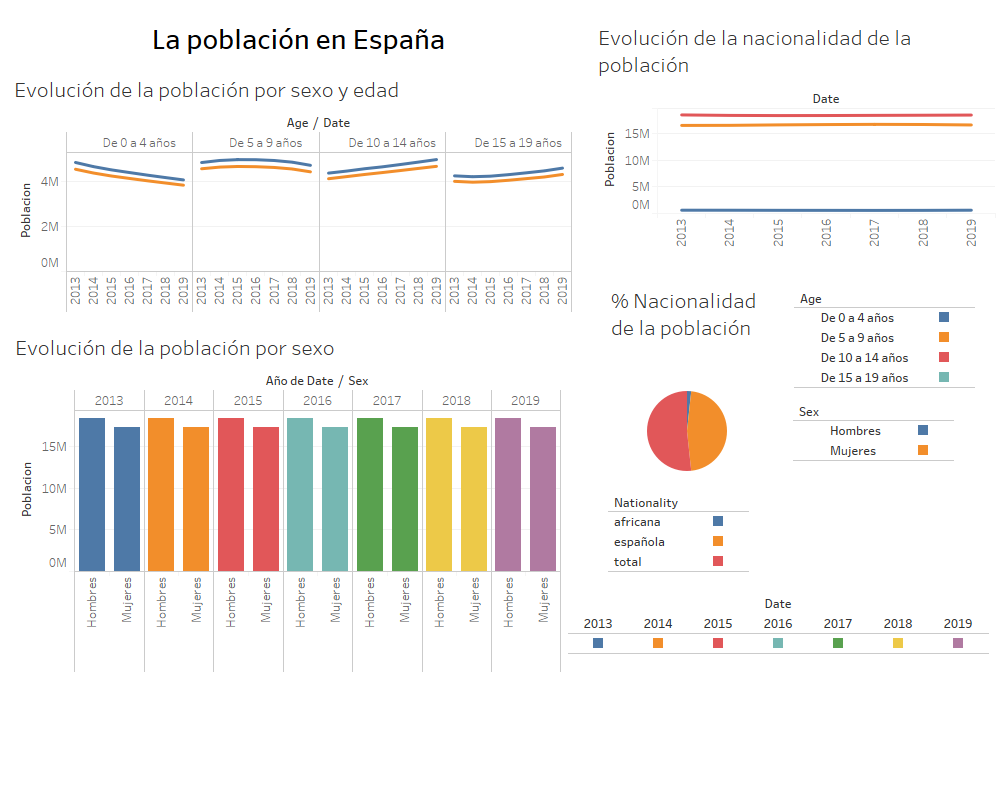
Además de las visualizaciones generadas durante el análisis de los datos. Se han creado dos visualizaciones interactivas con Tableau. Que han ayudado también a extraer resultados y conclusiones de los datos.

Una que permite ver el recuento de menores condenados en España, permitiendo filtrar por edad, sexo y nacionalidad. Y viendo su evolución anual. Se puede ver en la siguiente captura o en el link mencionado anteriormente:



Una que permite ver el recuento de menores condenados en España, permitiendo filtrar por edad, sexo y nacionalidad. Y viendo su evolución anual. Se puede ver en la siguiente captura o en el link mencionado anteriormente:

Otra que permite ver el recuento de la población de España, permitiendo filtrar por edad, sexo, nacionalidad y fecha. Viendo si evolución temporal. Se puede ver en la siguiente captura o en el link mencionado anteriormente:



## Conclusiones y resultados

Responder las dos preguntas, justificar con números y graficos!

## Hilo argumental

Título: Los MENAS en Barcelona!

El hilo argumental que se pretende seguir es:

* **Contexto:** dar respuesta a: qué, quién, cuándo, dónde, por qué y cómo. Expicando qué son los MENAS y la situación actual en Barcelona con la creación de varios centros donde residen menores extranjeros no tutelados, que además suelen tener muy pocas mujeres. Y qué preguntas pretende responder este reportaje. Que son:
  + ¿Es cierto que a raíz de la cantidad de menores no acompañados que residen en Barcelona y el resto de España ha habido un aumento de altercados?
  + ¿Es cierto que la cantidad de mujeres inmigrantes es más baja que la de los hombres?
* **Pregunta 1:** Después se procederá a dar respuesta a la primera pregunta exponiendo los resultados y conclusiones que se han extraído de los datos. Se mostraran visualizaciones para apoyar los resultados. Se indicara la procedencia de los datos y donde poder interactuar con ellos Tableau Public.
* **Pregunta 2:** Después se procederá a dar respuesta a la segunda pregunta exponiendo los resultados y conclusiones que se han extraído de los datos. Se mostraran visualizaciones para apoyar los resultados. Se indicara la procedencia de los datos y donde poder interactuar con ellos Tableau Public.
* **Conclusiones:** Se resumirán las conclusiones del reportaje.