
INFORME TAREA 1

Visualización de datos

10 DE ABRIL DE 2025
CARLES CARBONELL SALES
MIARFID

1. Introducción:

En este trabajo se va a realizar una investigación de la evolución de la población española a lo largo del tiempo y como ha sido afectada por diferentes factores como la tasa de natalidad, defunción o la inmigración extranjera. Para ello, se ha hecho uso de una serie de *datasets* provenientes de la página web del Instituto nacional de estadística. Concretamente, se han empleado los siguientes:

- 1.1. Defunciones1975: *Dataset* de 147 columnas y 54 filas que contiene la cantidad de defunciones organizada por provincias y sexos desde el 1975 hasta el 2023.
- 1.2. EdadPob1971-PAños: *Dataset* de 107 filas y 2 columnas que contiene la cifra de personas con cierta edad en el 1971 diferenciando entre hombres y mujeres.
- 1.3. EdadPob2024-PAños: *Dataset* de 107 filas y 2 columnas que contiene la cifra de personas con cierta edad en el 2024 diferenciando entre hombres y mujeres.
- 1.4. Flujo de inmigración procedente del extranjero por año, sexo y edad2008: *Dataset* con 276 filas y 14 columnas que representa la cantidad de inmigración clasificada por edad y sexo desde el 2008 hasta el 2021.
- 1.5. Nacimientos1975: *Dataset* de 147 columnas y 54 filas que contiene la cantidad de nacimientos organizada por provincias y sexos desde el 1975 hasta el 2023.
- 1.6. Población residente por fecha, sexo y edad1971: *Dataset* de 104 columnas y 327 filas con la población española clasificada por sexo y edad desde el 1971 hasta el 2022.

2. Preparación de los datos:

Por lo que hace a la lectura y al preprocesamiento de los diferentes datos de los *datasets*, primero se han cargado en un *dataFrame* mediante la función *read_excel* de *pandas*, después se han transpuesto para que los años representen filas y no columnas y finalmente, se han transformado estos años (que a veces eran fechas) a tipo *datetime*. Al tratar con 6 *datasets* con formatos diferentes, el proceso comentado puede diferir levemente.

A partir de estos *dataFrames* generados, se han creado conjuntos de datos generales seleccionando las diferentes columnas de interés de cada *dataFrame* y se han concatenado o unido en uno nuevo. En cuanto al relleno de los datos faltantes, se han asignado a 0 o al valor de la entrada anterior dependiendo del caso.

3. Visualización de datos:

3.1. Visualización de Natalidad

Como se puede apreciar en la Figura 1 la natalidad en España presenta actualmente una tendencia altamente bajista. Esta, desde el 1975 hasta el 2023 se ha llegado a reducir a la mitad y únicamente ha presentado un crecimiento hasta un máximo local en la franja entre los años 1996 y el 2008 (antes del comienzo de la crisis económica).

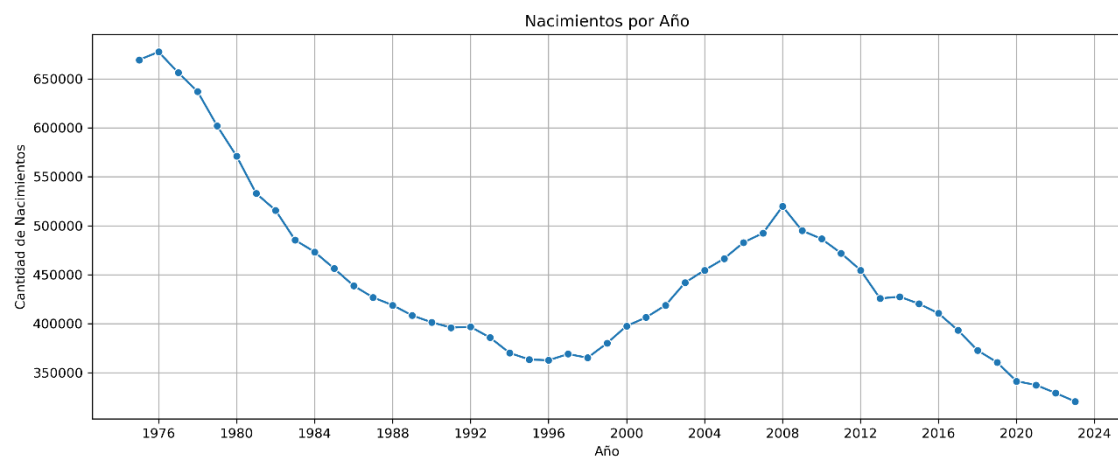


Figura 1. Line Plot de los nacimientos por año

Esta variación y decremento de la natalidad se vislumbra mejor al compararse porcentualmente; donde la natalidad hasta los principios de los 2000 ya representa más de $\frac{3}{4}$ de la natalidad global de la población y donde hasta los 2000 se presenta más de la mitad de la natalidad absoluta. Estas diferencias, se verían incluso aún más acentuadas si se contaran con los datos de los años 1970 hasta el 75 y con los del 2023 hasta el 2030.

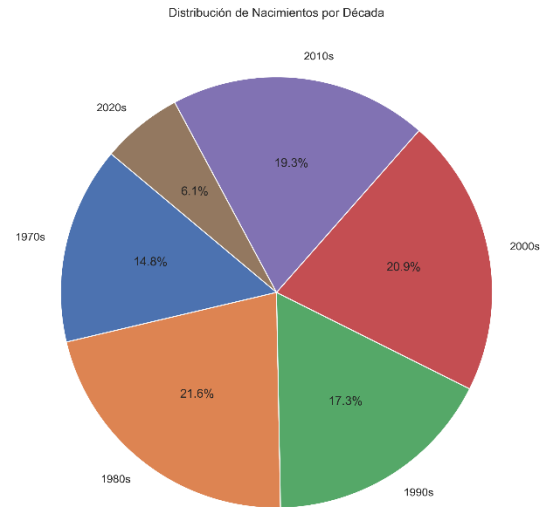


Figura 2. Pie de los nacimientos por década

3.2. Visualización de Defunción

Como se puede apreciar en la Figura 3 la mortalidad de la población ha seguido una linealidad creciente. Este resultado refleja como poco a poco la cantidad de las defunciones de población ha ido elevándose, partiendo de unas 300000 personas anuales en 1975 hasta las 430000 de 2023 y pasando por un máximo de casi 500000 en 2020 (probablemente debido a la pandemia de COVID). Esta subida constante de la mortalidad de la población se puede deber a un posible crecimiento de esta.

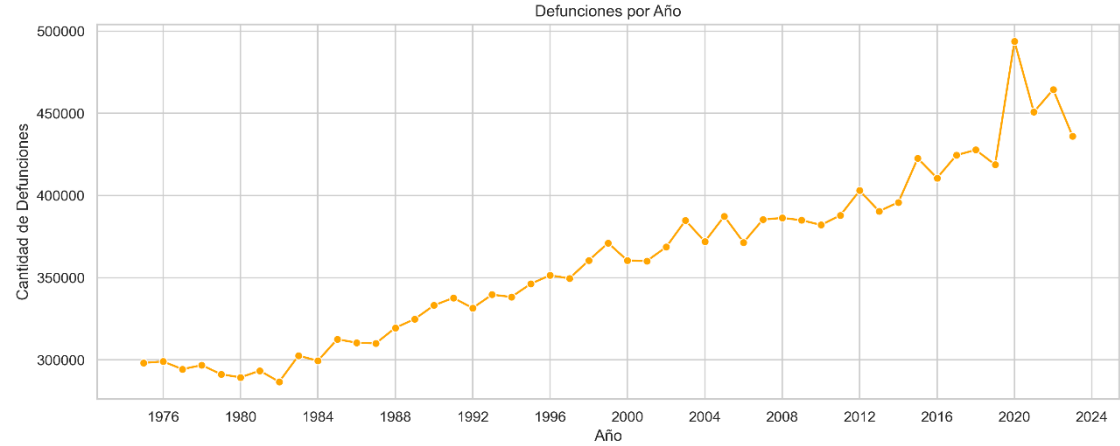


Figura 3. *Line Plot* de las defunciones por año

De igual forma, este análisis se puede apreciar en el gráfico porcentual de pastel de la Figura 4, donde cada década se percibe un ligero aumento de la mortalidad respecto a la anterior. Al igual que se ha comentado en la sección anterior, el hecho de que no se cuenta con los datos completos de las décadas de los 1970 y los 2020 afectan ligeramente a la concordancia con los resultados analizados. Si se contara estos, seguirían el patrón especificado.

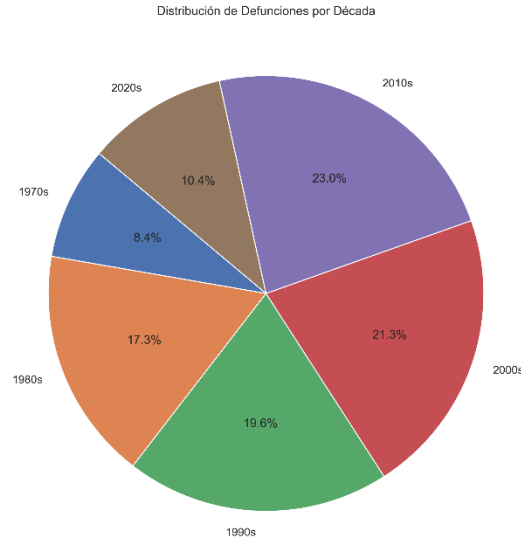


Figura 4. *Pie* de las defunciones por década

3.3. Visualización de Población

Como se puede observar en la Figura 5 la evolución de la población de España a partir del año 1971 presenta un crecimiento constante hasta la entrada de los 2000 donde, posiblemente por la mejora de la economía y la situación social, se percibe un mayor aumento de la población. Este, termina en 2008 donde, por la crisis surgida, se crea un estancamiento que se mantiene hasta la actualidad. En esta figura también se puede contemplar que a lo largo del crecimiento de la población me mantiene cierta paridad entre el número de mujeres y hombres.

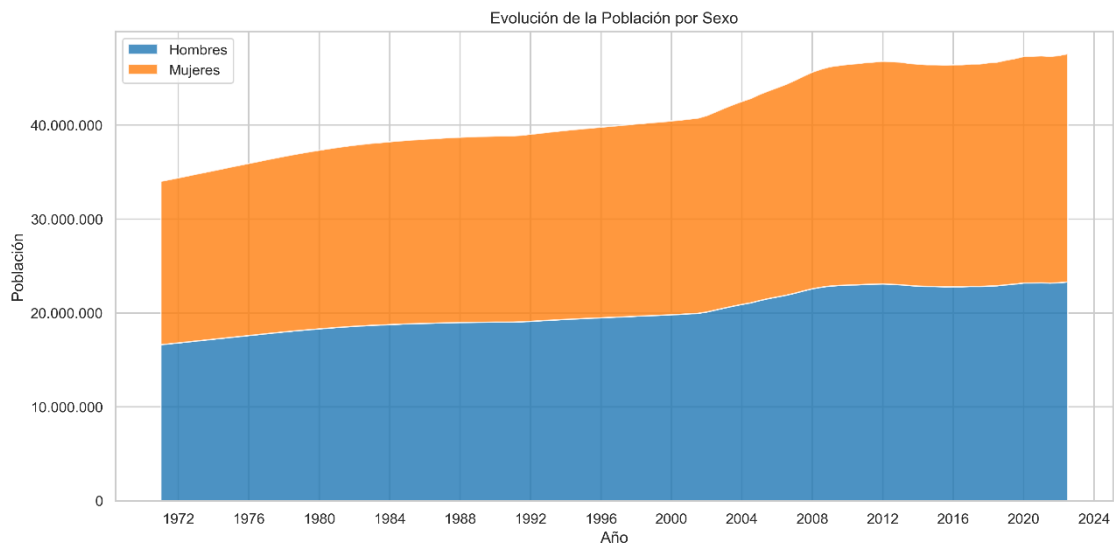


Figura 5. *Stack Plot* de la evolución de la población por sexo

Con el crecimiento poblacional ya definido, se ha creado una gráfica de doble eje Y (Figura 6) para observar cómo afecta la variación de la natalidad y mortalidad. Como se aprecia, la diferencia entre la natalidad y la mortalidad permanece positiva de forma relevante hasta el 2016, repercutiendo altamente en el crecimiento de la población. A partir de este año se invierte la tendencia; la tasa de mortalidad supera gravemente a la de la natalidad. No obstante, no se aprecia un decremento de la población sino más bien un estancamiento seguido de un muy pequeño incremento. En el siguiente apartado se va a explorar la posible causa de este efecto.

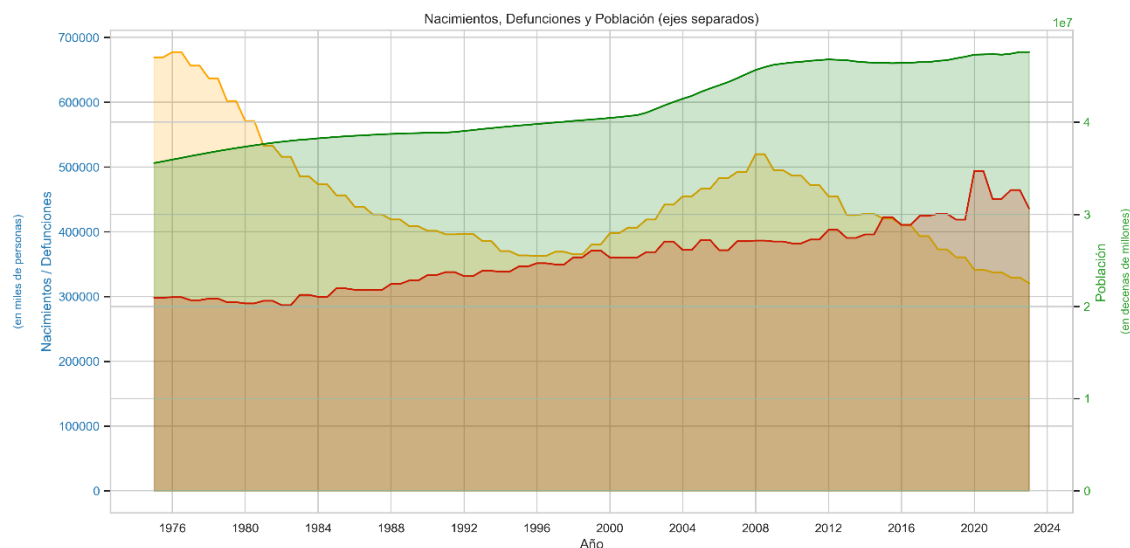


Figura 6. *Area Plot* de dos ejes separados de la población, nacimientos y defunciones

3.4. Visualización de Inmigración

La gran causa de que no haya un gran decremento en la población española se sitúa en la inmigración de países exteriores (Figura 7). Esta, en el período de inversión de natalidad y mortalidad anteriormente comentado, ha presentado un gran aumento, llegando en 2019 hasta una cantidad de más de 700000 personas.

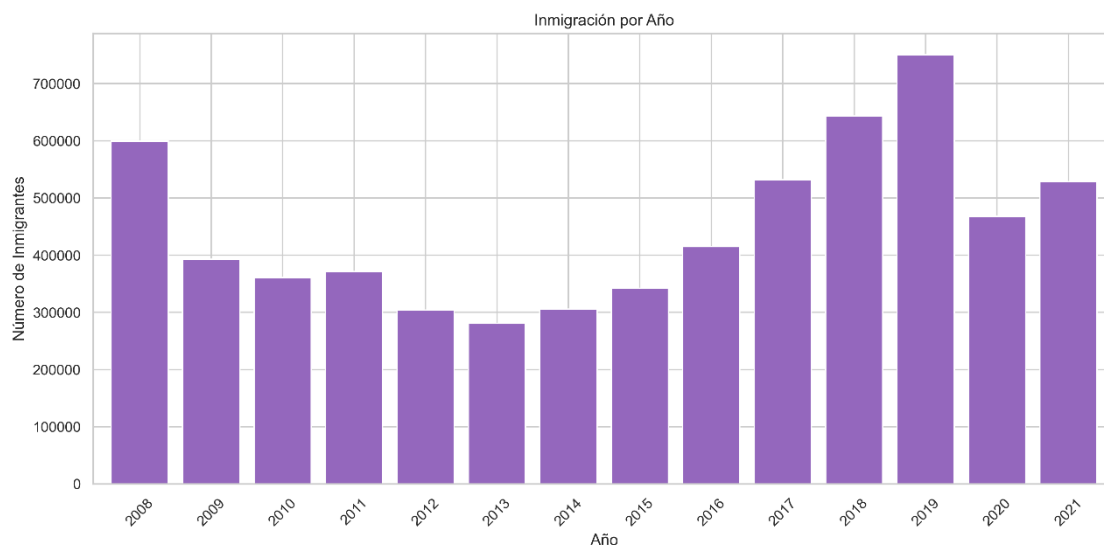


Figura 7. Barras apiladas de la inmigración por año

En este caso, un recuento de probabilidades como el mostrado en la Figura 8 no resulta muy fiable debido a la falta de datos en las diferentes décadas.

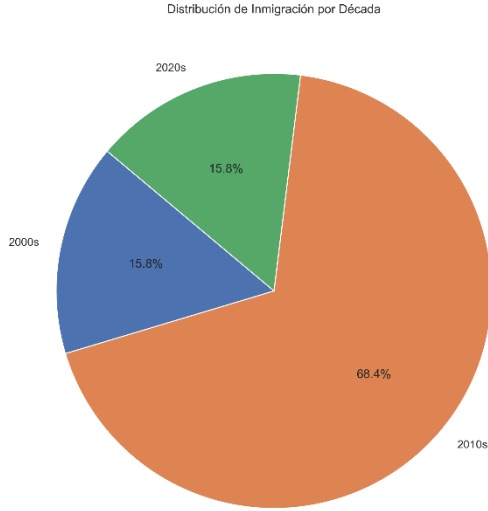


Figura 8. *Pie* de la inmigración por Década

3.5. Visualización de relaciones entre Natalidad, defunción, población e inmigración

A continuación, se muestran una serie de gráficas para visualizar más sencillamente cómo la inmigración, las defunciones y los nacimientos afectan a la evolución de la población.

La primera gráfica (Figura 9) muestra mediante un *heatmap* como las defunciones y la inmigración aumentan a lo largo del tiempo, como la natalidad decremента y, como se ha comentado a lo largo del trabajo, como estas variables afectan al aumento y estancamiento de la población.

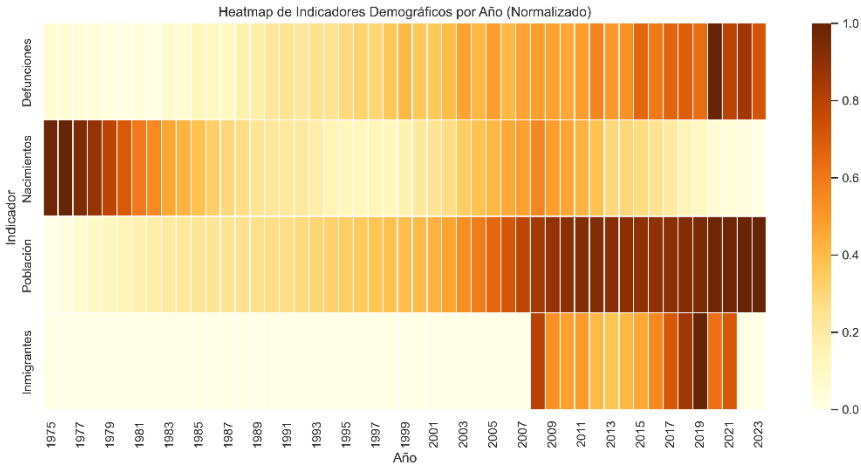


Figura 9. *Heatmap* de indicadores demográficos por año normalizado

De forma homónima, el *Bubble Chart* de la Figura 10 muestra de forma más sencilla este estancamiento y leve crecimiento a través de la representación de la inmigración mediante el tamaño de las burbujas y la diferencia entre nacimiento y defunciones (Saldo Natural) mediante su color.

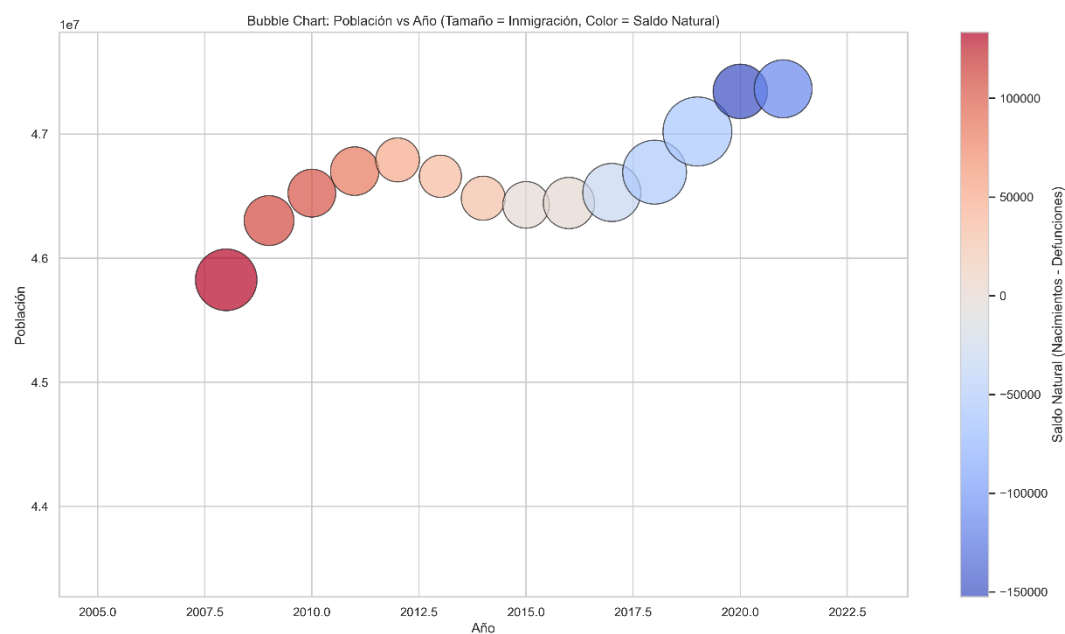


Figura 10. *Bubble Chart* de la población vs año

3.6. Visualización de pirámide poblacional

Un último aspecto interesante que analizar se trata de diferencia entre las distribuciones de franjas de edades de la población o también llamadas pirámides poblacionales. En este caso, se han creado dos pirámides (Figura 11 y 12) mediante la librería *plotly* (Apéndice 1) que representan la distribución poblacional de los años 1971 y 2024 respectivamente.

En cuanto a la pirámide poblacional del año 1971 (Figura 11), se puede observar que se cuenta con una población muy joven concentrada en la franja de los 0 hasta los 29 años. Esto se puede deber a que, debido a la baja calidad de vida y la limitada esperanza de vida en ese periodo, era común que la mayor parte de la población estuviera compuesta por personas jóvenes. Además, las altas tasas de natalidad propias de mediados del siglo XX contribuían a una base ancha en la pirámide, reflejando una sociedad en crecimiento, aunque con una notable disminución de población conforme aumenta la edad. Esta forma piramidal clásica indica un modelo demográfico aún en transición, con una mortalidad elevada en edades avanzadas y un fuerte peso de las generaciones jóvenes.

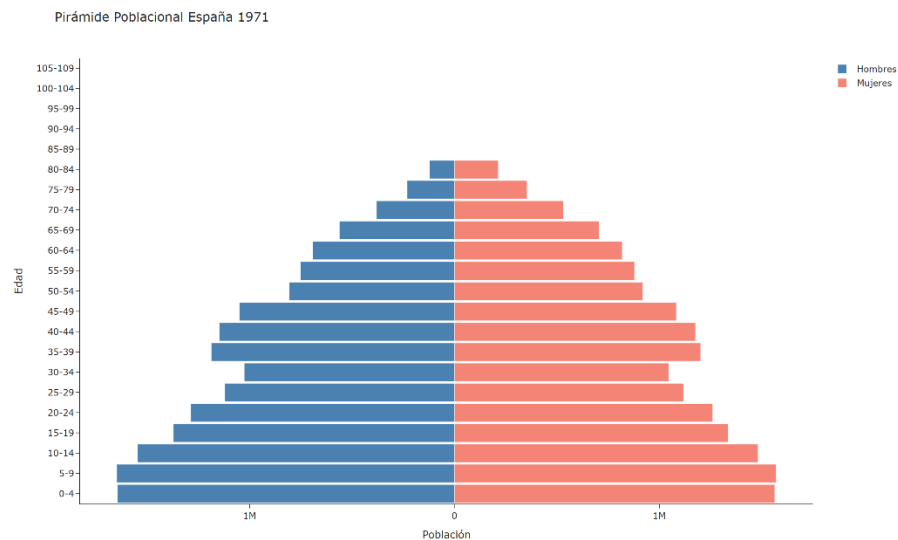


Figura 11. Pirámide poblacional España 1971

En la pirámide del año 2024 (Figura 12) se observa una clara inversión en la estructura demográfica en comparación con la de 1971. La base de la pirámide, correspondiente a las edades más jóvenes (0-14 años), es notablemente más estrecha, lo que indica una disminución sostenida de la natalidad en las últimas décadas. Esta tendencia puede deberse al encarecimiento general de la vida, que ha provocado un aplazamiento en la decisión de tener hijos por parte de muchas parejas, o incluso un rechazo directo a la procreación debido a la imposibilidad de mantener económicamente a la descendencia. En contraste, las franjas de edad comprendidas entre los 45 y los 59 años son las más anchas, reflejando el envejecimiento de las generaciones nacidas durante el baby boom de mediados del siglo XX.

Además, se aprecia un ensanchamiento progresivo de los grupos de mayor edad, especialmente a partir de los 70 años, con una clara predominancia femenina, lo que evidencia una mayor esperanza de vida para las mujeres. Esta forma más parecida a un barril o un rombo que a una pirámide clásica refleja un país con baja natalidad, longevidad elevada y envejecimiento demográfico, típico de sociedades avanzadas.

Este tipo de pirámides plantean una gran problemática a futuro, ya que la poca tasa de natalidad y la gran vejez de la población imposibilita el relevo generacional necesario para mantener el equilibrio entre cotizantes y beneficiarios de un sistema de bienestar como lo es el español.

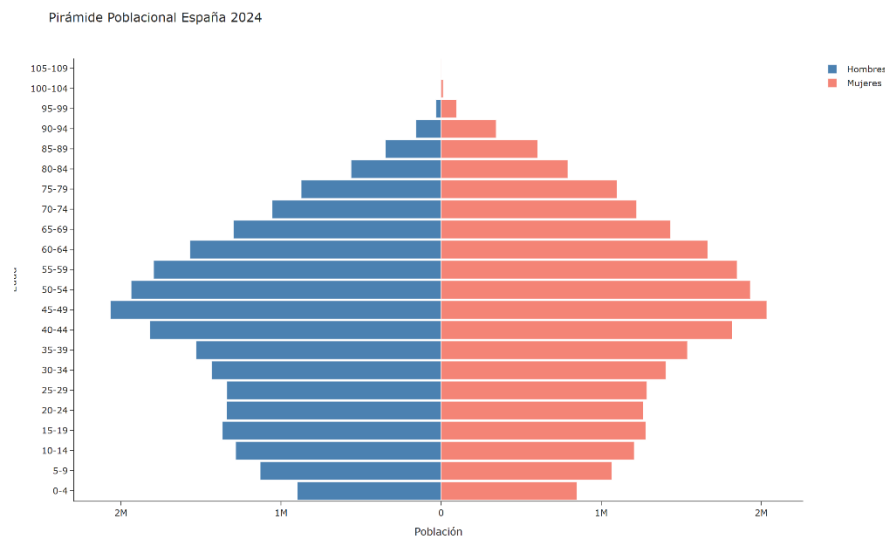


Figura 11. Pirámide poblacional España 2024

4. Conclusiones:

En conclusión, en este trabajo se ha realizado, a través del tratamiento de diversos *datasets* y la creación de ciertas gráficas, un análisis de cómo importantes variables como la tasa de nacimiento, inmigración o defunción afectan a la evolución de una población. Concretamente, se ha observado como en el caso de España, frente al gran decremento de la natalidad y el aumento de las defunciones en la actualidad que podrían causar un gran decremento en población, el crecimiento de la inmigración ha causado su estabilización e incluso un leve incremento. Además, también se ha investigado como a lo largo de la evolución del país se ha invertido completamente su pirámide poblacional, pasando de una población mayoritariamente joven a una mayor y con una muy baja natalidad (con los problemas que ello conlleva).

5. Apéndice 1: *Plotly*

Plotly es una biblioteca de visualización de datos interactiva y de código abierto para Python (también disponible en R, Julia y JavaScript). Permite crear gráficos dinámicos y altamente personalizables.

En este caso se ha empleado para crear los gráficos de las pirámides poblacionales (Figura 11 y 12). Para ello, se han cargado los dos *datasets* poblacionales de 1971 y de 2024 mediante *pandas* y se han agrupado por franjas de 5 años con la función *agrupar_por_franjas_separado*. Posteriormente, por cada *dataset* se han añadido dos *goBars* de *plotly*, una para los datos de las mujeres y otra para los hombres y, finalmente se ha personalizado el *layout* y se han mostrado por pantalla.