

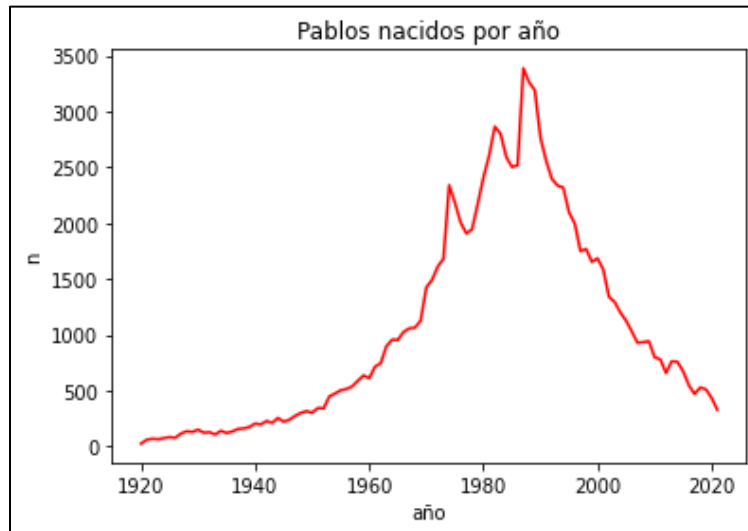
### Part A: How common is your name?

Para esta parte de la tarea, se trabajará con Python 3.9, principalmente con las librerías pandas, numpy y matplotlib, analizando una base de datos con registros de nombres, con las columnas "anio", "nombre", "sexo", "n", y "proporción", con casi 859.000 entradas.

Cabe destacar que se está investigando la popularidad del nombre **Pablo**.

El primer acercamiento a la popularidad de este nombre se muestra en el gráfico 1.

Figura 1: Cantidad de veces que el nombre Pablo fue registrado por año.

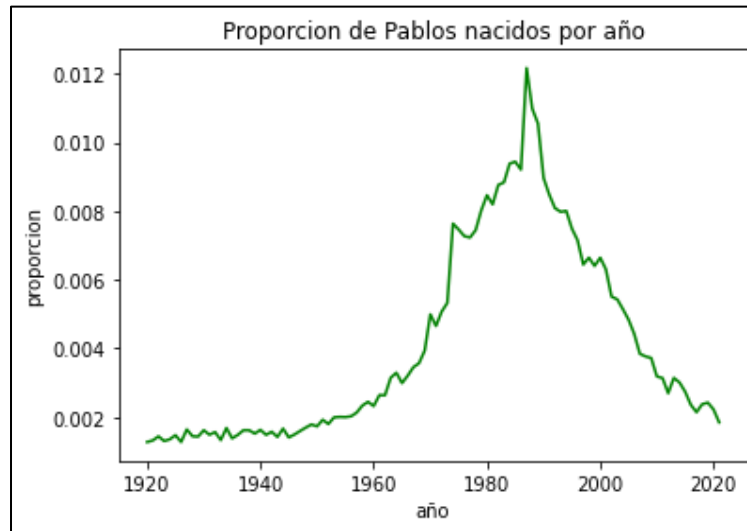


Fuente: Elaboración propia

A partir de la base de datos se puede observar que el nombre Pablo estuvo en aumento durante varios años, llegando a su tope en el año 1987, con un total de 3387 registros, a partir de ese año empezó a decrecer su popularidad.

El mismo análisis se puede hacer trabajando con la proporción del nombre, como en el gráfico que se muestra a continuación (Figura 2), llegando a la misma conclusión, que el año 1987 fue cuando se registró la mayor proporción del nombre, con un 1,21%

Figura 2: Proporción del nombre Pablo en la población total.



Fuente: Elaboración propia.

En la Tabla 2 se puede apreciar el registro del nombre Pablo para el año 1986, 1987 y 1988, con su respectivo total de registros (n) y proporción.

Tabla 1: Extracto de la cantidad total de registros y proporción para el nombre Pablo

Año	Sexo	n (#)	Proporción (%)
1986	M	2519	0,92%
1987	M	3387	1,21%
1988	M	3262	1,09%

Fuente: Elaboración propia

El segundo acercamiento que podría explicar la popularidad del nombre Pablo, es hacer un ranking de los nombres que más veces se registraron, particularmente en el año 1999, correspondiente al año de nacimiento del autor.

Tabla 2: Ranking de nombres registrados en el año 1999

Posición	Nombre	n
1	Nicolás	5999
2	Matías	5177
3	Sebastián	4712
4	Diego	4046
5	José	3611
....	...	...
19	Rodrigo	1656
<b>20</b>	<b>Pablo</b>	<b>1654</b>
21	Joaquín	1643

Fuente: Elaboración propia

De la tabla 2 se puede observar que el nombre más popular el año 1999 fue Nicolás, seguido por Matías. En la posición 19 se encuentra Rodrigo y con solo 2 registros más el nombre Pablo en la posición 20. Considerando que hubo un total de 3908 nombres distintos registrados ese año, en jóvenes de 22 años Pablo es un nombre bastante popular.

El tercer y último enfoque que se va a estudiar en este análisis es la popularidad de los nombres a nivel histórico.

Para hacer esto se ordenó la base de datos según la cantidad de veces que el nombre se registró según el registro civil (N), ordenados de forma decreciente. Luego se filtraron solo los nombres masculinos, y posteriormente se seleccionaron solo las primeras 50.000 entradas, esto debido a que si se trabajaba con las más de 800.000 entradas el tiempo computacional crecía de una forma significativa. Además, solo contemplando las primeras 50.000 entradas ya se contemplan **por lo menos** los primeros 30 nombres más populares en el periodo 1920-2020 considerando todos sus registros.

La base de datos construida para el tercer enfoque contempla el nombre, un “n” que representa la cantidad total de registros del nombre en la base de datos reducida, una media de la proporción, su respectiva desviación estándar, un “N” que representa la cantidad de veces que el nombre fue inscrito en el periodo, la posición del nombre en el ranking, y finalmente un intervalo de confianza construido al 95% de confianza. Un extracto de la tabla se muestra en la tabla 3.

Tabla 3: Ranking de nombres masculinos (Periodo 1920-2020) ordenados según N.

Ranking	Nombre	n (base de datos)	Proporción (media)	Desviación estándar	N (registro civil)	Límite inferior	Límite superior
1	José	102	0.0304	0.0160	588.512	0.0273	0.0335
2	Juan	102	0.0276	0.0133	581.458	0.0251	0.0302
3	Luis	102	0.0250	0.0134	505.769	0.0224	0.0276
...							
25	Claudio	102	0.0038	0.0034	106.873	0.0032	0.0045
<b>26</b>	<b>Pablo</b>	<b>102</b>	<b>0.0040</b>	<b>0.0028</b>	<b>106.418</b>	<b>0.0035</b>	<b>0.0046</b>
27	Ricardo	102	0.0040	0.0019	98.827	0.0036	0.0044

Fuente: Elaboración propia

Considerando que en la base de datos existe un total de 76450 nombres masculinos, que el nombre Pablo esté en la posición 26 indica que es un nombre MUY popular, con un total de 106.418 registros, y una proporción media que se encuentra entre [0,35% ; 0,46%] con un 95% de confianza, razón por la que en el curso de Bias, Lies and Causal Inference existen 4 personas con el nombre.

## Part B: Did The Beatles (the English rock band formed in 1960) impact the registration of Chilean babies' names?

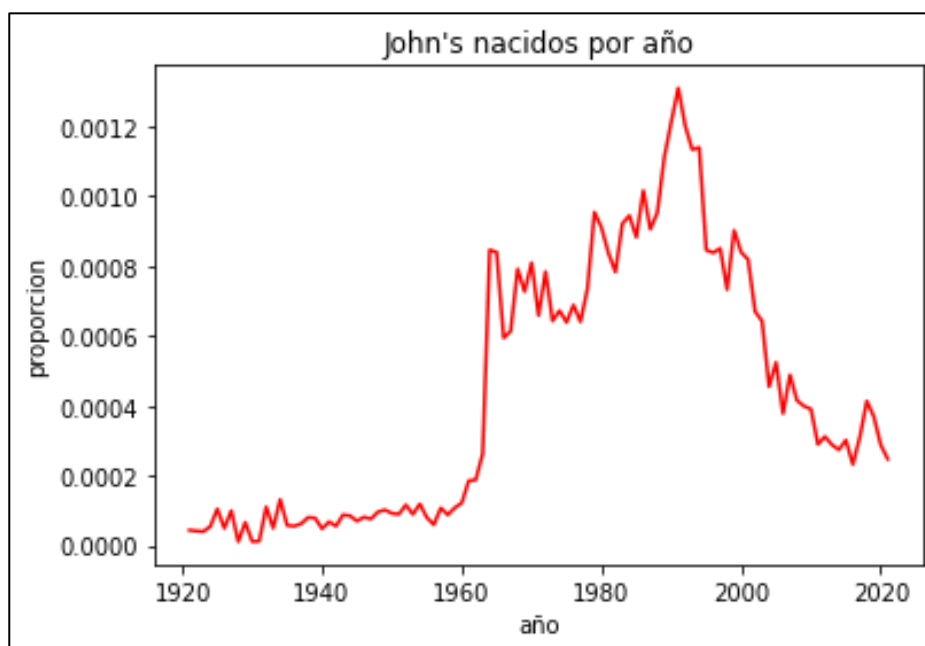
Lo primero que hay que saber para estudiar si el boom de Los Beatles afectó el registro de nombres en Chile, es saber los nombres de los integrantes. A continuación, los integrantes de la banda:

1. John Lennon
2. Paul McCartney
3. George Harrison
4. Ringo Starr

Para analizar el impacto, se procede a analizar cada uno de los nombres, por separado.

Empezando por John:

Figura 3: Proporción de la población con el nombre John en el periodo 1920-2020

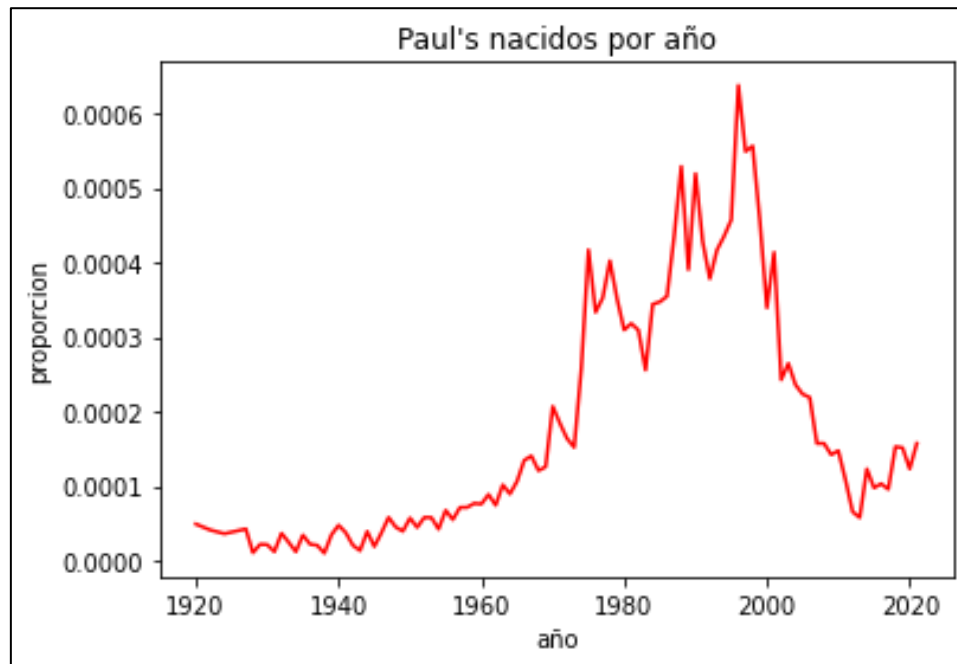


Fuente: Elaboración propia

Se puede ver claramente antes del 1960 el nombre no era muy popular en el país, sin embargo, a partir del 1964 creció significativamente su popularidad, en 1964 llegando a una proporción del 0,0846%, con 246 registros. Comparando el año 1964 contra el año 1963 (0,0263% y 75 registros) se puede decir que aumentó en un 221% (aproximadamente 3 veces más). A partir de ahí el registro se mantuvo en niveles parecidos, registrando su máxima en el año 1991, con 395 registros y una proporción del 0,131% de la población nombrada John.

Para el nombre Paul, el gráfico se muestra en la figura 4.

Figura 4: Proporción de la población con el nombre Paul en el periodo 1920-2020

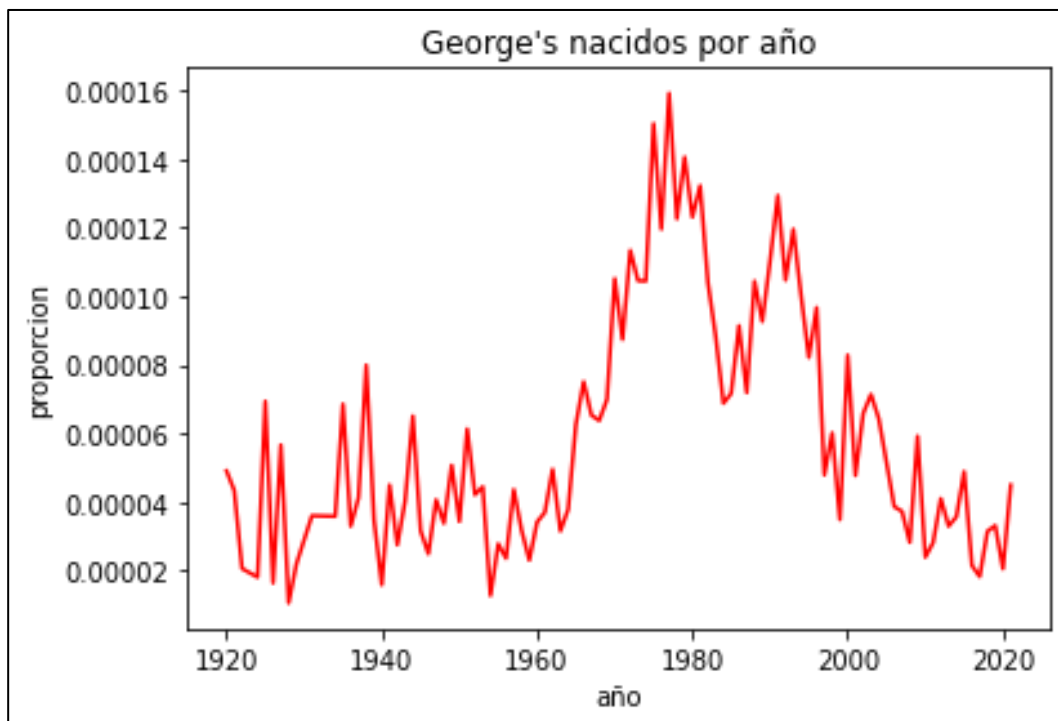


Fuente: Elaboración propia.

Al igual que el nombre anterior, en los inicios de la base de dato este nombre no fue para nada popular, y al contrario del nombre anterior, en el año 1964 no tuvo un cambio tan significativo, sin embargo, estaba en crecimiento. No fue hasta el año 1975 que representó un pico de popularidad, llegando a 122 registros y una proporción del 0,0417% de la población de ese año. Continuando en niveles parecidos, en el año 1996 ocurrió su máxima cantidad de registros, con un total de 178 y una proporción de 0,0637%.

Para el nombre George:

Figura 5: Proporción de la población con el nombre George en el periodo 1920-2020

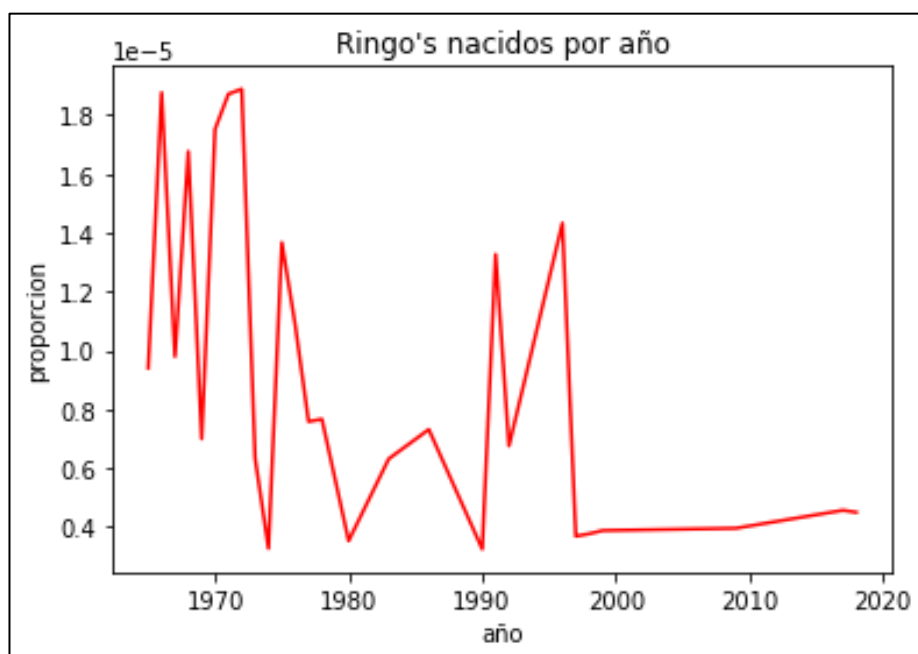


Fuente: Elaboración propia

Antes de 1960 no se aprecia una proporción elevada, con un promedio de 5 registros por año, sin embargo, se nota una pendiente positiva a partir de ese año, llegando en 1970 a un total de 30 registros, y en 1975 registrando su máxima con 44 registros y una proporción de 0,015% de la población para ese año.

Finalmente, para el nombre Ringo:

Figura 6: Proporción de la población con el nombre George en el periodo 1920-2020



Fuente: Elaboración propia

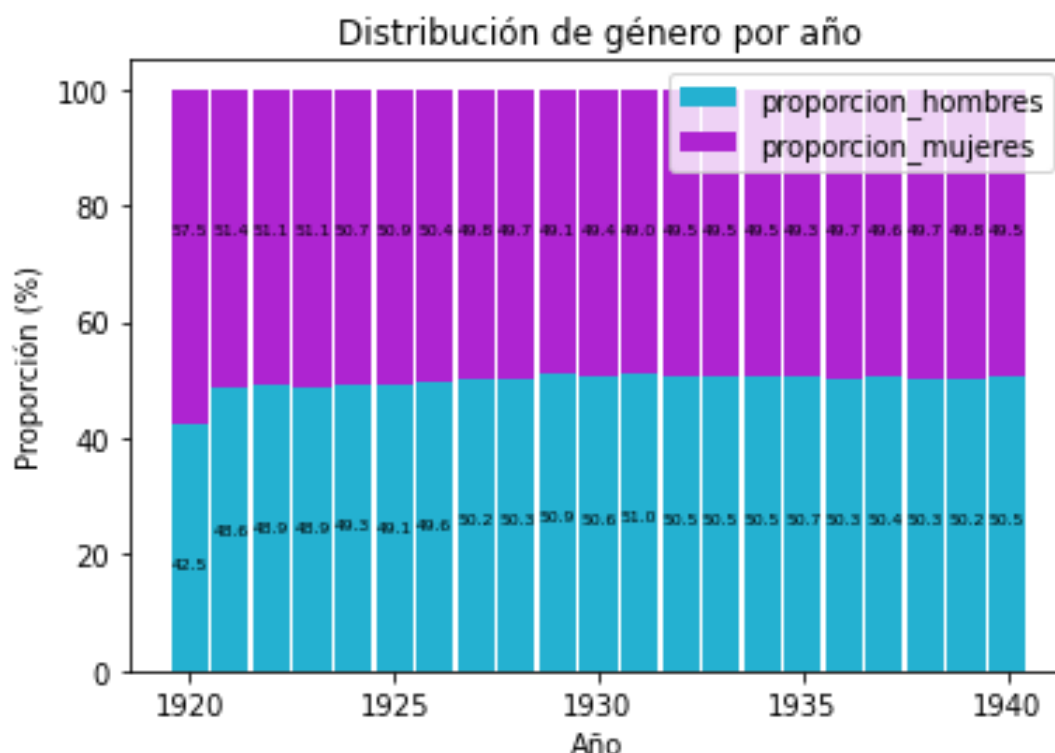
Para el caso de Ringo, la situación no es muy esperanzadora, sin embargo, el año 1965 se registraron 3 personas con este nombre, coincidiendo con el boom de la banda. Cabe destacar que el eje de la proporción está todo multiplicado por el factor  $10^{-5}$ , por lo tanto, la proporción es muy baja.

A modo de resumen, para el nombre John el boom de Los Beatles sí afectó en el registro de nombres, lo mismo se puede decir para Ringo, a pesar de que no contempla muchos registros, sin embargo, antes de 1965 no existía ninguna persona con este nombre. Para George y Paul también se notó un incremento en el periodo post formación de los Beatles.

**Part C: What can you say about the gender distribution in Chile across the years? Do you see any trends and/or singularities?**

El primer acercamiento a esta respuesta se ve en el siguiente gráfico (sólo se muestra el primer segmento de años, el resto en los anexos por temas de resolución)

Figura 7.1: Distribución de género para el periodo 1920-1940



Fuente: Elaboración propia

A simple vista no se puede decir nada, solo observando la serie de gráficos 7.1, 7.2, 7.3, 7.4, 7.5 y 7.6 se ve una leve tendencia de que el sexo masculino es el predominante en cuanto a porcentaje, lo cual contradice a toda lógica y estudios que avalan que el porcentaje de mujeres siempre es mayor respecto al de hombres, resultando en un 51% contra un 49% respectivamente<sup>1</sup> (INE, Censo 2017). Pero no hay que alarmarse, ya que estos estudios contemplan muestras muy grandes y totalmente confiables, sin embargo, lo que se está estudiando en este informe es la cantidad de bebés nacidos para el periodo en cuestión, y todo indica que nacen más bebés del sexo masculino.

Para corroborar esto, se realiza un intervalo de confianza, para comprobar estadísticamente que el porcentaje de hombres nacidos es realmente mayor al de mujeres.

<sup>1</sup> [https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda#:~:text=Resumen%20de%20censos%20de%20poblaci3n%20y%20vivienda&text=Sus%20resultados%20indican%20que%20la,51%2C1%25\)%2C%20mujeres.](https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda#:~:text=Resumen%20de%20censos%20de%20poblaci3n%20y%20vivienda&text=Sus%20resultados%20indican%20que%20la,51%2C1%25)%2C%20mujeres.)



A partir de los datos, se llega a que el intervalo de confianza en un 95% para hombres es:

$[0,502;0,506]$

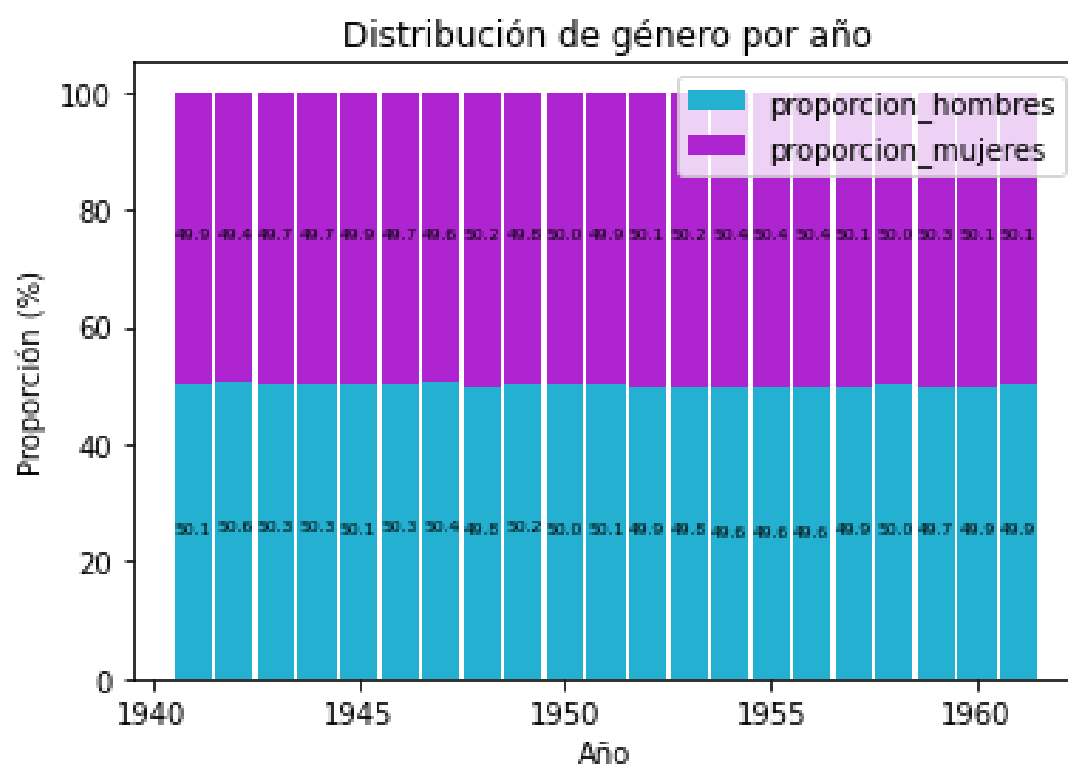
Al mismo nivel de confianza, el intervalo de confianza para mujeres es:

$[0,493;0,497]$

Donde se puede apreciar que el IC para hombres es mayor en su totalidad al de mujeres, es decir, nacen más hombres que mujeres en Chile.

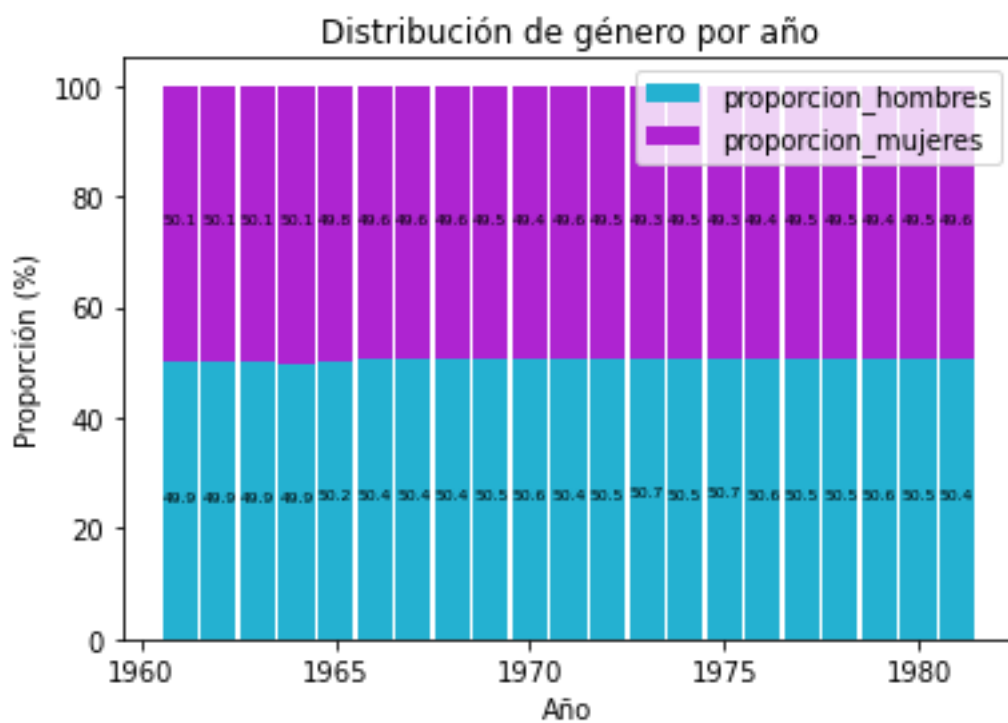
## Anexos

Figura 7.2: Distribución de género para el periodo 1920-1940



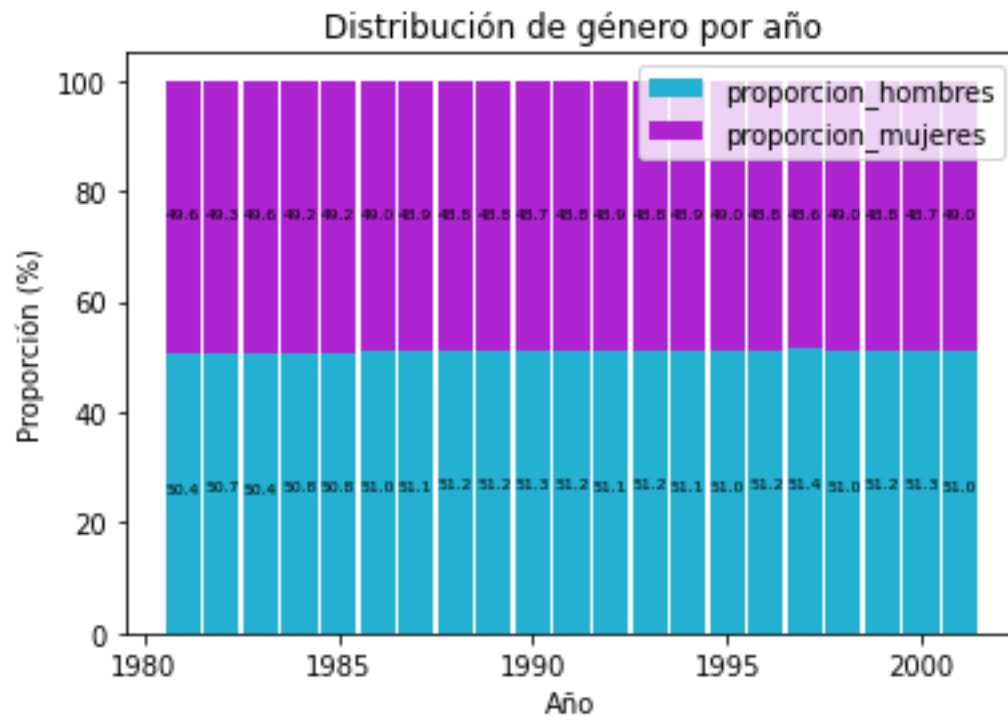
Fuente: Elaboración propia

Figura 7.3: Distribución de género para el periodo 1941-1961



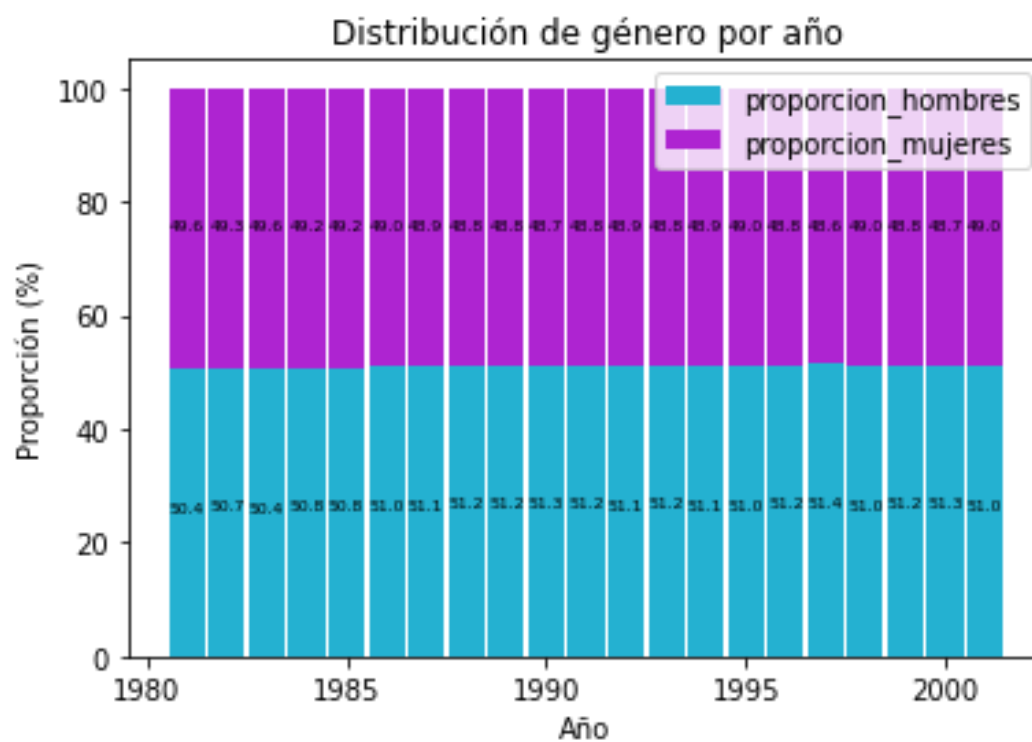
Fuente: Elaboración propia

Figura 7.4: Distribución de género para el periodo 1961-1981



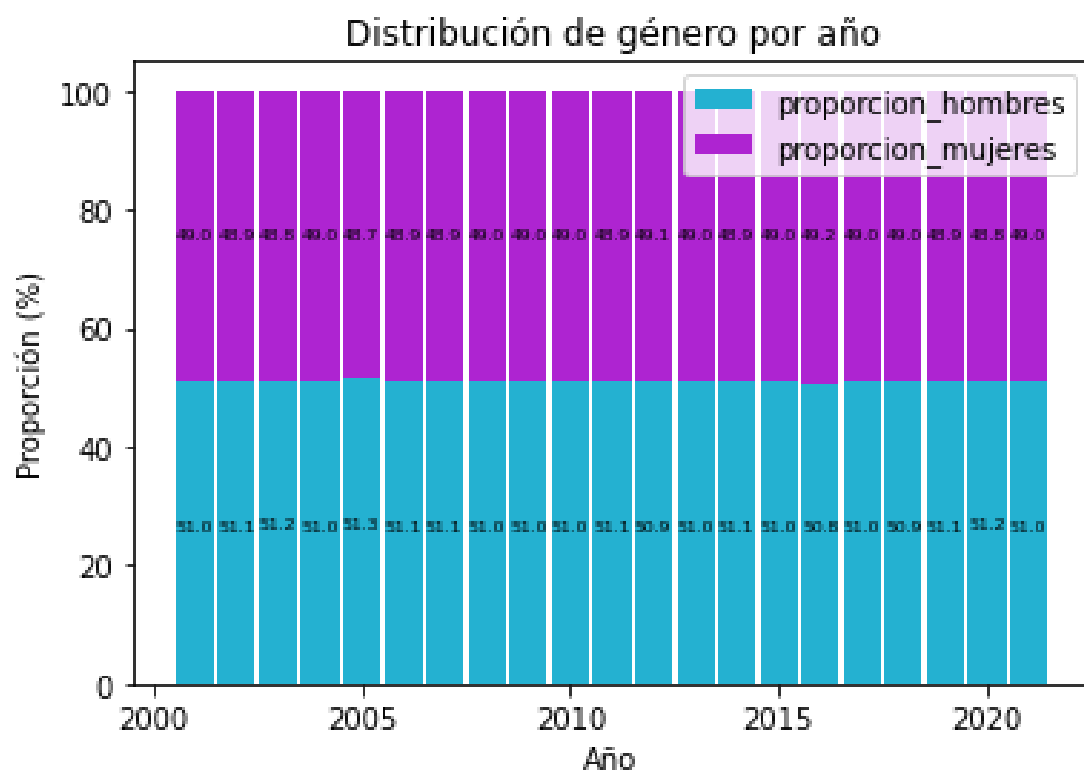
Fuente: Elaboración propia

Figura 7.5: Distribución de género por año para el periodo 1980-2001



Fuente: Elaboración propia

Figura 7.6: Distribución de género por año para el periodo 2001-2021



Fuente: Elaboración propia