
Análisis Exploratorio

AUTOR

GARAGIOLA MARIANO

Córdoba

En este informe presento un breve análisis exploratorio del *dataframe: historico-nombre.csv*.

El dataframe contiene los nombres registrados en el Registro Nacional de las Personas agrupados por año, desde el año 1922 hasta el año 2015. Quiero mencionar que debido a la cantidad total de nombres registrados no pude realizar un análisis del conjunto completo debido a las limitaciones computacionales. El análisis que presento lo realicé con los primeros cinco millones de datos (5432786 datos exactamente), de esta forma, cubro el periodo temporal comprendido entre 1922 y 1988.

Inicié contando la cantidad de personas registradas por año en el Registro Nacional de las Personas. El gráfico está presentado en la figura 1, donde se puede observar un crecimiento en el número de personas registradas a medida que se avanza en el tiempo, mostrando un claro crecimiento en la población.

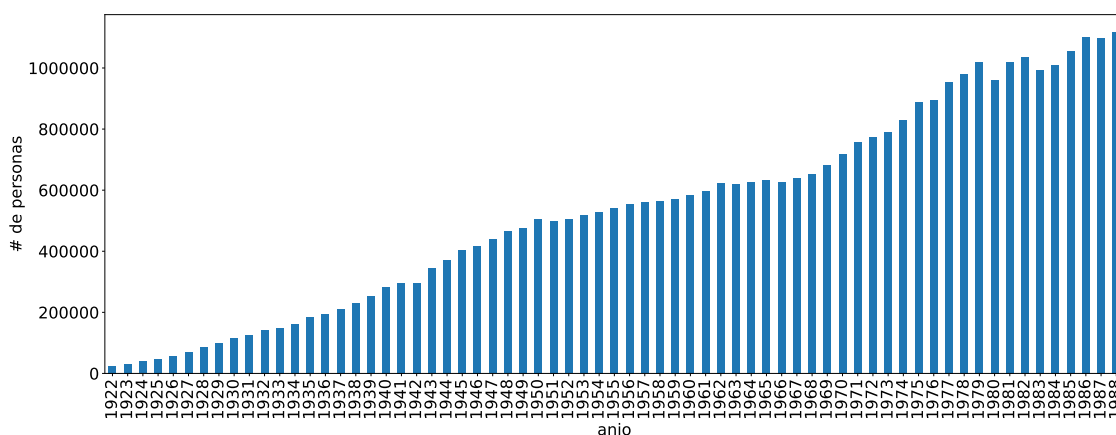


Figura 1: Número de personas registradas por año en el Registro Nacional de las Pesonas.

Si bien, en general, se muestra un crecimiento en la cantidad de personas registradas, hay algunos años en los cuales dicho número disminuye respecto al año anterior o bien no sigue con la tendencia, esto probablemente es debido a condiciones políticas-sociales del país durante esos años.

A continuación, decidí analizar la cantidad de nombres distintos registrados a lo largo de los años. Dicho gráfico, figura 2, muestra que también ha aumentado la variedad de nombres a lo largo del tiempo, esta cantidad tiene un comportamiento similar al mostrando en la figura 1. Nuevamente, hay años en los cuales la cantidad de nombres distintos registrados no sigue la tendencia de los años anteriores, como por ejemplo el año 1980, pero esto no quiere decir que haya disminuido la cantidad de personas registradas. Sería interesante poder tener los registros de la cantidad de personas que nacieron en cada año, de esa forma sería posible estimar el porcentaje de personas nacidas en Argentina y así también estimar el porcentaje de personas resgistradas que provienen de otros países, ya que hubo años en los cuales inmigraron un gran cantidad de personas.

Un hecho que me llamó la atención del *dataset* es que algunos nombres están formados por muchas palabras, por este motivo decidí determinar el número de palabras que forma cada nombre. De esta forma, realicé el gráfico del número de personas con una dada cantidad de palabras en su respectivo nombre, el gráfico se puede ver en la figura 3. Como es de esperarse, la mayor cantidad de nombres está compuesto por dos palabras, pero hay un número considerable de nombres con una, tres y cuatro palabras, lo cual es entendible. En el registro hay personas con más de cinco palabras en el nombre, si bien no está prohibido tener más de cinco palabras en el nombre no deja de ser extraño. En particular, existen cuatro registros que contienen diez palabras en el nombre y dos registros que contienen once palabras en el nombre. Viendo en detalle cada uno, encontré que uno de los registros tiene el nombre: *Maria De La Concepcion Trinidad Maria Amalia Fernanda*

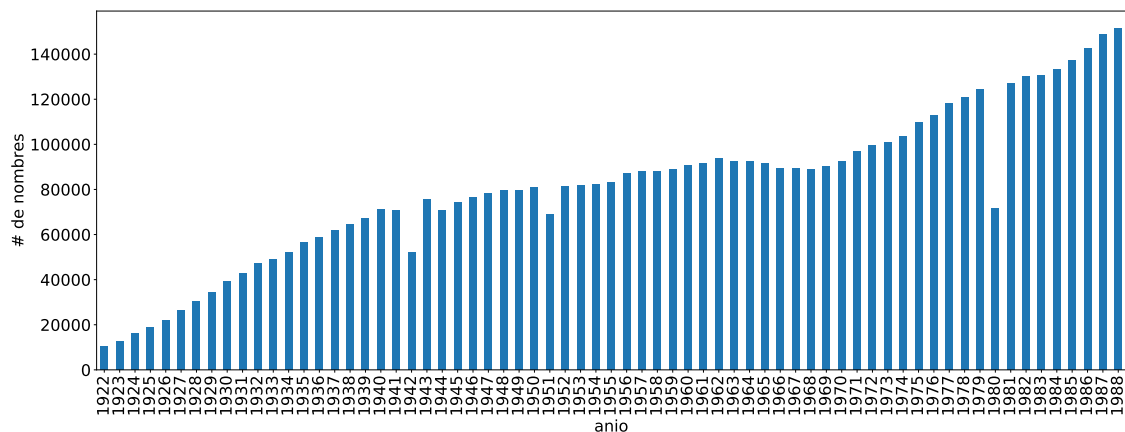


Figura 2: Número de nombres distintos registradas por año en el Registro Nacional de las Pesonas.

Felisa Ignacia Adelaida y es del año 1922, lo que me hace dudar de que sea una misma persona. Lo mismo sucede con el segundo registro que posee once palabras, el nombre es *Josina Livina Dolores Lucia Carmen Balduina Yolanda Maria Inmaculada Gislana Javiera*.

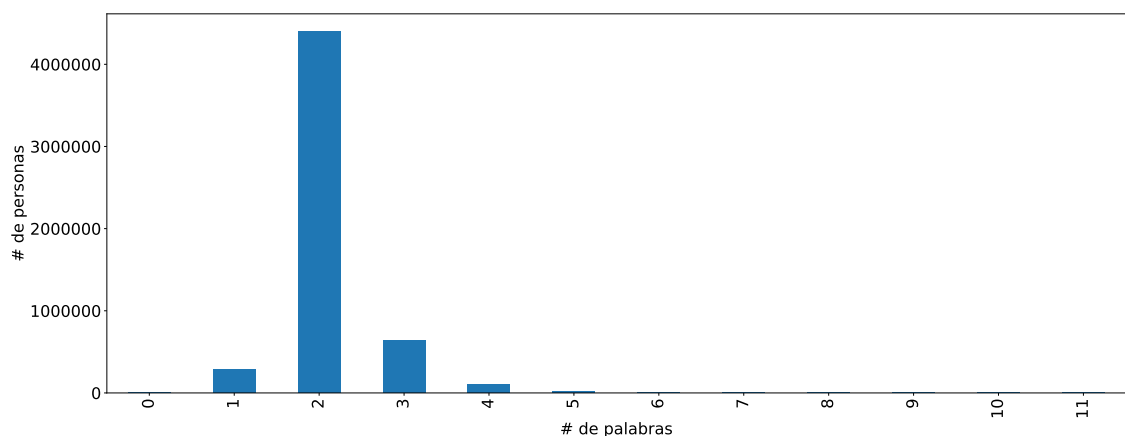


Figura 3: Número de personas con un dado número de palabras en su nombre.

A partir del *dataset* es posible determinar cuál fue el nombre más usado en cada año, por ejemplo, Juan Carlos fue el nombre más usado durante los años de 1933 al 1951.