

## **Parte 2**

### **A. Análisis exploratorio**

En primer lugar, se realiza un análisis de los datos originales obtenidos del Portal Nacional de Datos.

Luego con los datos filtrados, según se detalla a continuación, se evalúan la cantidad de nombres registrados por año y cuales fueron usados como nombres únicos en el período.

Además se presenta un ranking con los nombres más utilizados, y se comparan los nombres de moda y figuras destacadas del futbol argentino, durante ciertos eventos destacados.

#### **1) Datos originales**

El set de datos original consta de 9.761.609 registros, con información sobre los nombres registrados en Argentina históricamente. Consta de 3 campos, donde se indica para cada *Nombre* en que *año* se utilizó y en qué *Cantidad*.

Se observan 13 datos faltantes en el campo Nombre, mientras que en los otros – cantidad y año – no hay datos nulos.

En cuanto a los nombres registrados se encontraron 627 que presentan en alguna parte del mismo algún número y además 4.482 con caracteres no alfabéticos, como ser puntos, comas, guiones, entre otros. No hay registros duplicados, considerando la combinación nombre y año.

Se decidió borrar los registros cuyos nombres contuvieran dígitos o sean nulos. Además, se eliminan los caracteres no alfabéticos, por considerarlos errores de carga, manteniendo los registros, si luego de la limpieza, los nombres seguían teniendo al menos dos caracteres de largo.

No se realizó ninguna modificación en relación a los caracteres acentuados. Por esta razón se encuentran datos como “María”, “María” y “Maria”, que serán considerados distintos.

En adelante se realiza todo el análisis sobre los datos originales filtrados, los cuales representan 9.760.961 registros.

Los datos presentan 3.210.988 nombres distintos (sin tener en cuenta la ponderación del mismo, indicada por el campo cantidad). Mostrando al nombre simple “Ana” con la mayor frecuencia (187).

Los años informados son 94 y van desde 1922 a 2015. La mayor frecuencia se asocia al 2015 con 209.628 registros, mientras que en el otro extremo la menor se registra en 1922 con 10.333.

## 2) Nombres registrados por año

Al considerar la cantidad de nombres registrados por año, se aprecia el incremento de los mismos conforme se llega a comienzos de la década del 90. En el período 1994-2011 los totales por año disminuyen, encontrando un nuevo pico de crecimiento en 2012 (Figura 1).

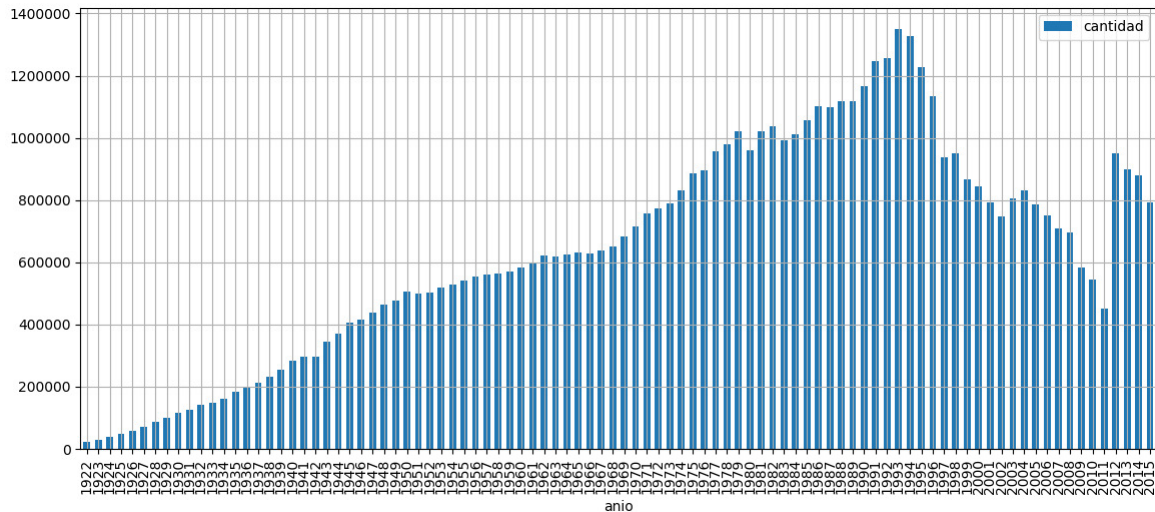


Figura 1 – Nombres registrados por año (barras)

La frecuencia máxima está en 1993, con 1.351.198 personas, en tanto que la mínima ocurre en 1922, anotándose 23.667, seguramente influenciado por la digitalización posterior de la información (Figura 2).

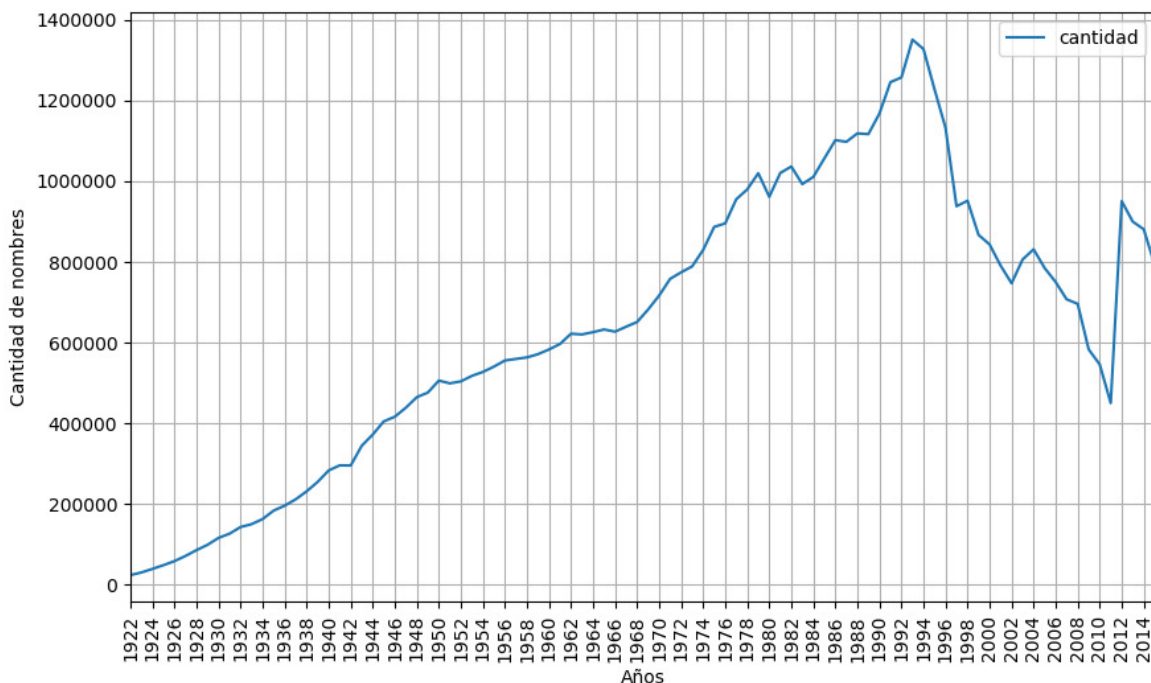


Figura 2 – Nombres registrados por año

La baja cantidad de nombres registrados en 2011 evidencian alguna anomalía, puede responder a faltantes de datos o inconsistencias en los mismos.

### 3) Nombres simples usados por año

Analizando la cantidad de nombres registrados por persona, se puede decir que predominan los nombres formados por dos simples, en todo el set de datos, como muestra la Figura 3. Un ejemplo sería “Juan Manuel”. Teniendo en cuenta que no se consideran los denominados “conectores” (formados por las palabras “de”, “del”, “lo”, “los”, “la” y “las”). Es decir, el nombre completo “María del Carmen”, estará formado en este caso por dos nombres simples.

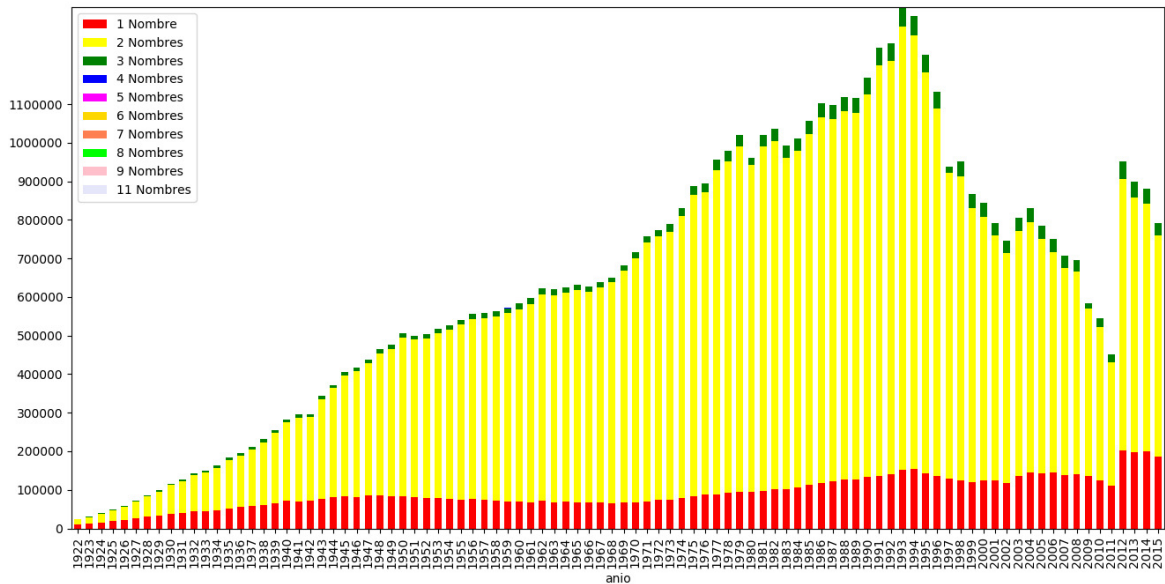


Figura 3 – Cantidad de nombres simples registrados por año

Solamente menos del 3% tiene un tercer nombre, número que va disminuyendo conforme se suman palabras al nombre completo.

Durante los primeros años censados las personas con dos nombres representan entre el 55 y 65%. Luego, desde 1930 comienza a aumentar, llegando a pasar el 75% en 1945. Dos décadas después ya supera el 85%, dejando los nombres simples solamente para algo más del 10% de la población registrada. Iniciando el 2000 la moda de tener dos nombres comienza a bajar, observándose en el 80% de los casos. Caída que se sostiene en el 2010, perdiendo otro 10%. Al parecer, la tendencia en los últimos años sería a volver a los nombres únicos para los recién nacidos.

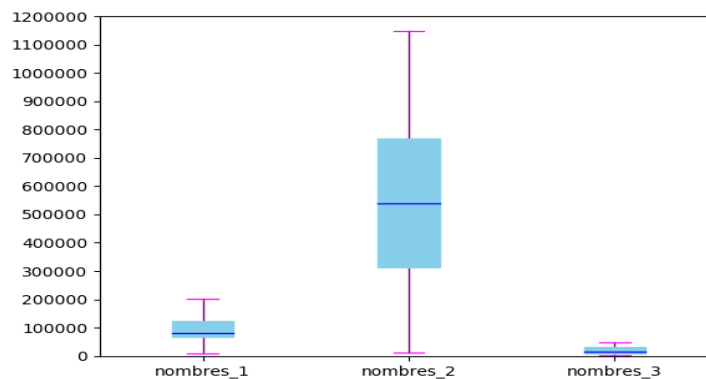
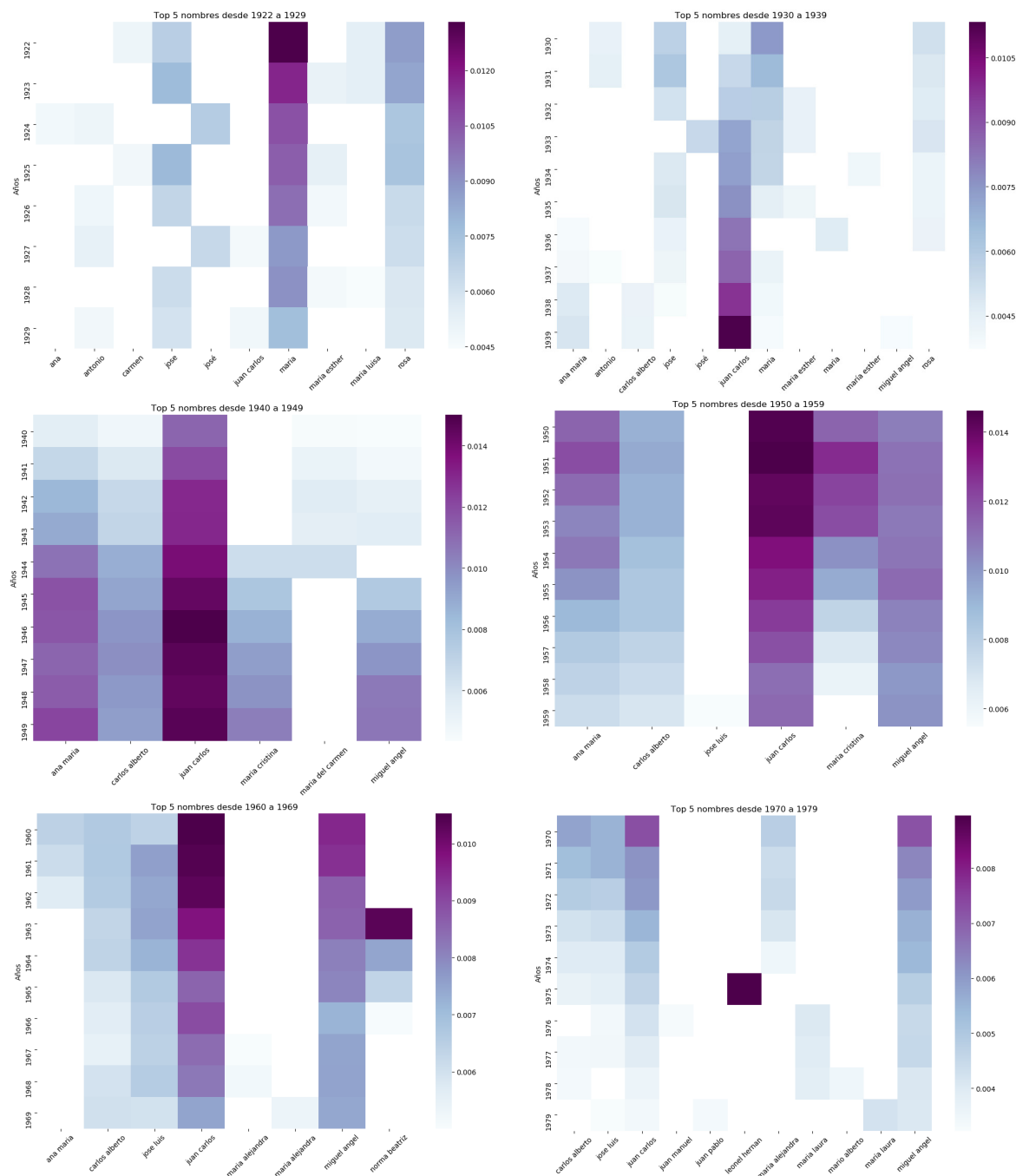


Figura 4 – Distribución de las personas con uno, dos o tres nombres

#### 4) Ranking de nombres

Tomando los nombres completos, se confeccionó un ranking con los cinco más frecuentes de cada año, como se muestra en la Figura 5, por períodos de diez años. El valor representado con la escala de colores es la frecuencia del nombre (cantidad de personas del total registrado en el año).



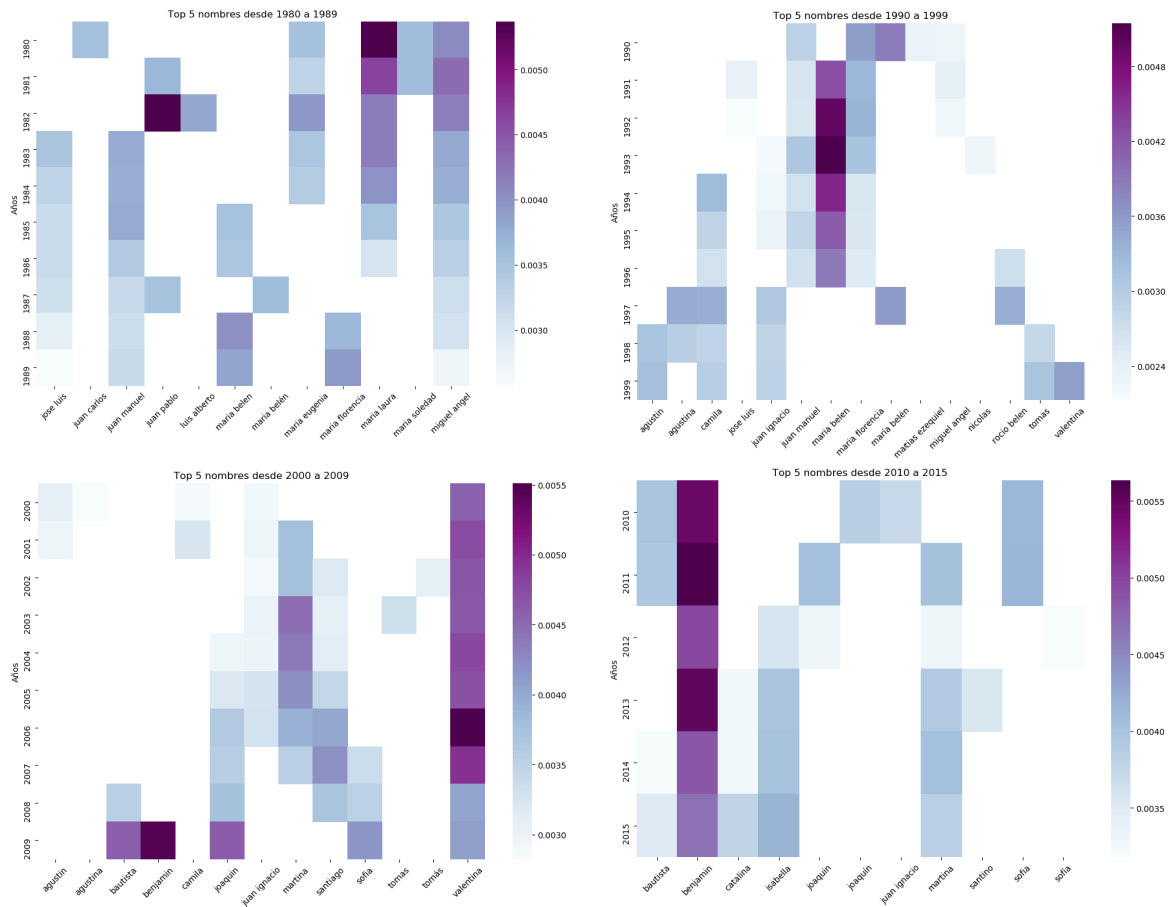


Figura 5 – Los cinco nombres completos más frecuentes por año

En la década del 20 “María” fue el nombre más elegido, desplazado del podio recién en 1932 cuando “Juan Carlos” tomó protagonismo. Lugar que mantuvo hasta los 70, registrándose en todos esos años un solo cambio, con la aparición fugaz de “Norma Beatriz” en 1963.

Durante el período 1970-1979 los preferidos se fueron alternando entre “Miguel Ángel” y “Juan Carlos”, a excepción de 1975, donde se observan más “Leonel Hernán” que cualquier otro nombre completo.

El nombre compuesto “María Laura” ocupó el primer lugar del Ranking en los '80, apenas opacado en 1982 por “Juan Pablo”, lo que marca la influencia de la visita del máximo referente de la iglesia católica al país. Sobre el final de la década fue superado por “María Belén”, el nombre representativo de la década del 90 en su conjunto.

Entrando al nuevo siglo, los más utilizados fueron nombre simples, como “Valentina” los primeros años, o “Benjamín” desde el 2009 en adelante.

## 5) ¿Moda o fanatismo?

La elección de los nombres populares puede responder a un gran número de causas: moda o tradición, podrían ser las principales. Sin embargo, en un país donde el fútbol es el deporte que más jugadores (profesionales o amateurs) tiene y el que despierta mayor fanatismo, no sería raro que los ídolos de una época se vean immortalizados en los nombres de las generaciones futuras.

Tomando los datos de 1978, cuando Argentina obtiene por primera vez el título de campeón del mundo, y los dos años posteriores, el nombre “Mario Alberto” se encuentra séptimo entre los preferidos para hombres. Casualmente el mismo que lleva Kempes, goleador y autor de dos de los goles en la final contra Holanda.

Más adelante, en el puesto 18, la aparición de “Daniel Alberto” hace referencia al capitán de aquel equipo Passarella. Primer nombre que además se repite en otros 7 de los top 50 de esos años.

Para cuando Argentina festeja en México 86, la segunda copa obtenida en este deporte, y en los siguientes dos años, ya son varias más las menciones a nombres de pila de los jugadores, entre los 50 más frecuentes. En tercer lugar estaba de moda “José Luis”, como marcaba el documento de uno de los artilleros de la final, “Tata” Brown o el defensor Cuciuffo. El sexto lugar entre los frecuentes, era para “Jorge Luis”, al igual que Burruchaga.

Más abajo en el listado hay lugar para “Sergio Daniel” (Batista) y nuevamente aparece “Daniel Alberto” (Passarella), a pesar de no haber tenido minutos en cancha. Sobre el final del top 50, se encuentra un rezagado, “Diego Armando” (Maradona), artífice de la hazaña y también “Carlos Daniel” (Tapia).

## B. Implementación

Toda la implementación fue realizada en Python 3.6. Se utilizaron las librerías: pandas, numpy, matplotlib y seaborn.

## C. Referencias

Dataset	Nombres argentinos históricos.
Descripción	Contiene nombre y cantidad por año desde 1922 al 2015.
Fuente	Portal nacional de datos
URL	<a href="http://datos.gob.ar/dataset/nombres-personas-fisicas">http://datos.gob.ar/dataset/nombres-personas-fisicas</a>
Dataset	Selecciones futbol
Descripción	Jugadores argentinos de cada plantel participantes de los mundiales de futbol
Fuente	Losmundialesdefutbol.com
URL	<a href="http://www.losmundialesdefutbol.com/selecciones/argentina_planteles.php">http://www.losmundialesdefutbol.com/selecciones/argentina_planteles.php</a>