

Ejercicio

Para el ejercicio vas a necesitar bajar el dataset de [nombres argentinos](#) del portal nacional de datos. Vamos a trabajar con el [consolidado histórico](#). La tabla tiene 3 campos: Nombre, Cantidad y Año.

El ejercicio consta de dos partes, una primera parte guiada donde hay que responder a consignas cerradas y una segunda parte más libre donde se pide un análisis exploratorio del dataset según tu propio criterio.

Los ejercicios se pueden hacer en el lenguaje que quieras, usando cualquier librería de ese lenguaje. También se pueden sumar otros datasets de la fuente que sea, siempre y cuando incluyas el link para poder bajarlo.

Para hacernos llegar tu respuesta, te pedimos que crees un repositorio público en github y nos compartas el link.

Parte 1

Para cada ítem te pedimos que entregues:

- Un breve texto explicando en tus palabras cómo lo resolviste
- El código en el lenguaje que hayas elegido
- El resultado

- A) ¿Cuántos nombres distintos hay en todo el dataset, considerando cada nombre por separado? Es decir, un nombre compuesto como *María Inés* cuenta una vez para *María* y otra para *Inés*.¹
- B) Construí un clasificador que separe los nombres en masculinos y femeninos² ¿Cuán bien funciona? Elegí una métrica apropiada para evaluar el desempeño del clasificador y reportala.³
- C) Usá el clasificador que construiste en el punto anterior y graficá la cantidad estimada de nombres de varón y nombres de mujer para cada año.

¹ Vas a notar que hay variantes entre nombres (por ejemplo Inés e Ines), algunas que son errores de digitalización y otras que son variantes válidas. No hace falta que intentes solucionar este problema para realizar el conteo.

² Este clasificador tendrá necesariamente limitaciones, entre otras cosas, porque existen nombres que son usados tanto para varones como para mujeres. Para este punto nombre es el nombre completo es decir "Juan Carlos" se considera todo junto.

³ Es probable que necesites generar (o descargar de algún lado) un conjunto de entrenamiento para el clasificador.

Parte 2

En esta parte te pedimos que realices un pequeño análisis exploratorio del dataset y que lo presentes en forma de reporte breve en un archivo PDF. Tanto la diagramación del informe como el contenido queda a tu criterio. Solo a modo disparador incluimos algunas preguntas: ¿Qué preguntas interesantes se pueden responder con los datos? ¿Qué cosas en el dataset son anómalas y llevan a pensar en posibles errores? ¿Con qué otra fuente de datos sería interesante integrar este dataset?