

# Análisis de datos: Validación de cédulas panameñas

Luis D. Ortiz

[github.com/luisortiz11](https://github.com/luisortiz11)

**Abstract.** El proceso de validación de una cédula se puede realizar considerando principalmente el contenido de entropía de una entrada cualquiera. Es decir, la medida de incertidumbre de la entrada nos permite filtrar las posibles cédulas validas, considerando solamente el contenido de dicha entrada que no es parte de un formato específico. Este proceso es ventajoso porque todas las cédulas panameñas se agrupan en formatos específicos y comparten características en cada uno de tales grupos.

## 1. Introducción

Las cédulas de identidad personal (C.I.P) emitidas en Panamá comparten características principalmente porque se debe identificar inequívocamente a cada ciudadano y se debe agrupar a la población bajo alguna métrica o disposición, para facilitar la implementación e imposición de leyes. En sentido general, las cédulas transmiten información de procedencia, edad aproximada, estatus migratorio y hasta cuestiones de etnia, en el caso de la comunidad autóctona panameña. Todas estas características permiten limitar el formato de las cédulas y agruparlas de tal forma que, para una entrada cualquiera, se puede filtrar el contenido de información y separarlo del contenido de formato.

## 2. Estrategia

Consideremos una entrada cualquiera, esta entrada puede contener una cédula valida o no. Considerando que dicha entrada contiene una cédula, debemos decidir si el contenido adicional de dicha entrada no causa “confusión” en la determinación de la validez de la cédula. Habiendo dicho esto, para garantizar la funcionalidad de validador decidimos que una entrada debe tener exactamente la misma forma que alguno de los formatos posibles.

En esta ocasión, los formatos utilizados para comparar con la entrada son:

- Regular (provincia-libro-tomo). Ej: 1-1234-12345
- Panameño nacido en el extranjero (PE-libro-tomo). Ej: PE-1234-123456
- Extranjero con cédula (E-libro-tomo). Ej: E-1234-12345
- Naturalizado (N-libro-tomo). Ej: N-1234-12345
- Panameños nacidos antes de la vigencia (provinciaAV-libro-tomo). Ej: 1AV-1234-12345
- Población indígena (provinciaPI-libro-tomo). Ej: 1PI-1234-12345

Como podemos notar, utilizaremos 6 formatos. Automáticamente, una solución directa para crear nuestro validador sería comparar la entrada con cada uno de los posibles formatos y luego realizar comparaciones adicionales de ser necesarias. Consecuentemente, debemos determinar si esta solución es óptima.

### 3. Analizando datos

Por simple inspección podemos aseverar que los formatos comparten características. Enumerando dichas similitudes obtenemos que:

- Todos los formatos terminan en libro y tomo con sus respectivos guiones. (-1234-12345 o -1234-123456)
- Las cédulas regulares, AV y población indígena comienzan por provincia y existe un número asociado a cada provincia.
- Las cédulas de panameño nacido en el extranjero, extranjero y naturalizado comienzan por PE, E o N.

Datos adicionales incluyen las limitantes de la longitud de los grupos de números y sus posibles valores.

### 4. Solución optima

Considerando la solución directa de realizar 6 comparaciones y comparaciones subsecuentes, queremos hallar una solución para validar cédulas con menos complejidad. Utilizando el análisis de datos, obtenemos que 2 comparaciones son suficientes para determinar si la cédula es parte de uno de los siguientes grupos:

- Grupo 1: cédula regular, AV, población indígena. (comienza por provincia, numero de 1 hasta 10)
- Grupo 2: cédula de panameño nacido en el extranjero, extranjero, naturalizado. (comienza por PE, E o N).

Dentro de cada grupo, la siguiente comparación sería determinar cuál elemento del grupo se trata y determinar si es válida la longitud y el valor del número de libro y tomo. Considerando que dichos datos están separados por guiones.

### 5. Implementación

Para implementar nuestra solución, utilizaremos el lenguaje Python y la librería “re” de expresiones regulares, que contiene funciones para realizar comparaciones, dado que tenemos formatos específicos (patrones). Cabe destacar que nuestra solución optima nos permite:

- Utilizar las expresiones regulares de forma austera (en solo 2 comparaciones). Las demás comparaciones se pueden realizar por estructuras de decisión if.
- Se reduce significativamente la complejidad de nuestro código, por el gran costo por unidad de longitud de las expresiones regulares.
- Se garantiza la validación de cédulas.
- Se distingue información de libro, tomo, provincia y tipo de cédula.