

## Objetivos

### Unidad 1 – Lenguajes Regulares y Autómatas

Definir formalmente los distintos tipos de autómatas finitos (deterministas, no deterministas, no deterministas con transiciones lambda, transductores), dar ejemplos, identificar sus elementos y conocer sus aplicaciones.

Entender y aplicar, tanto las nociones de lenguajes y expresiones regulares, como de autómatas de estado finito, para el reconocimiento de patrones, procesamiento, validación y extracción de texto usando un lenguaje de programación.

A usted se le entrega un archivo de excel transacciones.xls que contiene 100 transacciones. Cada una de ellas tiene un consecutivo, identificador, fecha, valor, banco y medio.

Como primer paso deberá cargar el archivo en un dataframe de pandas, a partir de dicho dataframe usted tendrá que generar 3 dataframes distintos que denominará taxes, public\_services y shopping. Cada uno de estos los obtendrá haciendo uso de autómatas que servirán como filtros sobre la variable identificador.

La columna identificador es bastante significativa y es aquella que le permitirá reconocer el tipo de transacción (impuestos, servicios públicos y compras).

La transacción asociada a impuestos tiene un identificador que empieza con im o tx (mayúsculas o minúsculas o combinaciones), sigue con tres ceros consecutivos, un dígito que no sea cero, luego cinco dígitos y terminará con z minúscula o mayúscula. Para este filtro se le solicita utilizar un autómata determinista.

La transacción asociada a servicios públicos tiene un identificador que empieza con s (mayúscula o minúscula), sigue con la cadena 11 o 10, luego la cadena 22 o 33, posteriormente 6 dígitos y termina con una s (mayúscula o minúscula). Para este filtro se le solicita utilizar un autómata no determinista (al menos un estado deberá tener dos transiciones con una misma etiqueta).

La transición asociada a compras tiene un identificador que empieza con c (mayúscula o minúscula) o la cadena shp (mayúsculas o minúsculas o combinaciones), luego un caracter especial guión, slash o dos puntos (-,/, :), posteriormente **dos ceros, un dígito no cero, luego ocho dígitos** y finaliza con Y mayúscula. Para este filtro debe utilizar un autómata no determinista con transiciones lambda. Debe utilizar al menos dos transiciones lambda en el autómata que construya.

Deben entregar diagramas de los tres autómatas y el notebook con la implementación de los autómatas en pyformlang y los dataframes resultantes.