

INTRODUCCIÓN A LA CIENCIA DE DATOS PARA SU USO EN POLÍTICAS PÚBLICAS

Herramientas para el análisis, visualización y comunicación de datos

Oficina de la CEPAL en Buenos Aires. Modalidad virtual

Presentación

En la actualidad estamos presenciando una verdadera revolución en la cantidad de datos generados. El incremento exponencial en la capacidad de procesamiento, la capacidad de almacenamiento y la velocidad de transmisión dieron origen a la disciplina conocida como Ciencia de Datos. Este nuevo paradigma abre un sinfín de posibilidades en la mayoría de los ámbitos y las industrias, incluyendo la ciencia, la salud, la agricultura y las políticas públicas.

Este taller tiene como objetivo introducir a los participantes en los fundamentos teóricos y las principales herramientas de la Ciencia de Datos, aplicado al diseño, gestión y evaluación de políticas públicas. La modalidad de trabajo será teórico-práctica, implementando código en el lenguaje de programación R. Se espera que al finalizar el programa los participantes tengan una competencia básica en algunas de las principales herramientas para el análisis, visualización y comunicación de datos. El lenguaje de programación utilizado es el R, de acceso libre y gratuito.

El programa de formación constará de dos módulos. El primero es un módulo introductorio, donde se presentará el lenguaje de programación R y se repasarán algunas nociones básicas de estadística. Este primer módulo estará acompañado de contenido asincrónico para reforzar la parte práctica. El segundo módulo cubrirá distintas herramientas para comunicar los resultados del análisis de datos en forma inteligente y dinámica.

Equipo docente

Francisco Valentini

Economista de la Universidad de Buenos Aires y Especialista en Explotación de Datos y Descubrimiento de Conocimiento en la Universidad de Buenos Aires. Actualmente es Data Scientist en Equifax. Anteriormente trabajó como analista técnico en la Dirección de Modelos y Proyecciones del Ministerio de Hacienda de la República Argentina y como docente de Estadística en la Universidad de Buenos Aires y la Universidad Nacional de Moreno.

Sebastián Guiñazú

Economista de la Universidad de Buenos Aires, magistrando en Desarrollo Económico en la Universidad Nacional de San Martín. Es Líder de Ciencia de Datos en Equifax. Anteriormente trabajó como consultor en la Dirección de Evaluación de Impacto del Ministerio de Producción de la Nación Argentina. También realizó trabajos como consultor externo en el Banco Interamericano de Desarrollo (BID), Banco de Desarrollo de América Latina (CAF) y Banco Mundial. Es autor de artículos académicos y se desempeña como docente de Estadística y Econometría en la Universidad de Buenos Aires.

Contenidos

Módulo I – Introducción al análisis de datos con R

Clase 1: Introducción al universo R

- Introducción a la consola R
- Paradigmas de funciones: R nativo, Tidyverse.
- Estructura de datos: vectores, listas, data frames
- Importar y exportar datos. Opciones de bases de datos disponibles para trabajar
- Introducción a la programación funcional (loops y funciones)

Clase 2: Transformación de datos y exploración

- Modelo tidy
- Limpieza de una base de datos
- Estadística descriptiva
- Principales tipos de gráficos
- Visualización con ggplot2

Módulo II – Análisis avanzado de datos con R

Clase 3: Regresiones en R

- Nociones básicas de estadística clásica
- Implementación de un modelo regresión lineal
- Implementación de regresión logística
- Interpretación de resultados

Clase 4: Exploración avanzada de datos

- Modelos lineales con interacciones
- Árboles de decisión
- Algoritmos de clustering

Módulo III – Reportería con R

Clase 5: Comunicación con reportes y apps

- Notebooks y documentos html
- Dashboards y widgets interactivos
- Introducción al diseño de apps

Clase 6: Scraping y análisis de texto

- Estructura de un sitio web
- Extracción de componentes de documentos html
- Automatización de la extracción de texto
- Introducción al text mining (expresiones regulares, normalización de texto, bag-of-words)