

Programa

CIENCIA DE DATOS PARA EL DISEÑO Y GESTIÓN DE POLÍTICAS PÚBLICAS

Herramientas para el análisis, visualización y comunicación de datos

Oficina de la CEPAL en Buenos Aires. Modalidad virtual.



NACIONES UNIDAS

CEPAL

Presentación

En la actualidad estamos presenciando una verdadera revolución en la cantidad de datos generados. Algunas estimaciones sostienen que el 90% de los datos disponibles se generó en los últimos dos años. El incremento exponencial en la capacidad computacional 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 el transporte, la salud, la ciencia, la agricultura y las políticas públicas.

Este taller, realizado en el marco del Acuerdo de Cooperación entre el Ministerio de Economía de la Nación y la CEPAL, 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, de acceso libre y gratuito. 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 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. El segundo módulo cubrirá distintas herramientas para comunicar los resultados del análisis de datos en forma inteligente y dinámica.

Requisito: Conocimiento de estadística y de programación en R (excluyente).

Fecha y horario: Martes y jueves de 9.30 a 13hs, entre el 18 de agosto y el 10 de septiembre (inclusive).

Equipo docente

Mariano Aizpurua. Médico especialista en Pediatría. Realizó maestría en Efectividad Clínica con orientación en Evaluación de Tecnologías Sanitarias y Evaluaciones Económicas en UBA - IECS. Se desempeña como coordinador en el Área de Ciencia de Datos de la Gerencia Operativa de Gestión de Información y Estadísticas en Salud del Ministerio de Salud de la Ciudad de Buenos Aires. Lleva adelante el Programa de Desarrollo Profesional de la Gerencia para la formación continua y desarrollo de capacidades en el equipo. Además, trabaja como Médico Pediatra realizando tareas asistenciales en efectores de salud de la CABA.

Sebastián Guiñazú. Economista de la Universidad de Buenos Aires, magistrando en Desarrollo Económico en la Universidad Nacional de San Martín. Es Data Scientist 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.

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.

Contenidos

Módulo I – Introducción al análisis de datos con R

Clase 1: Introducción al universo R

- Introducción a la consola R
- Diferencias y compatibilidades con otros lenguajes.
- 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
- Diferencias y ventajas entre R scripts y R Markdown

Clase 2: Transformación de datos

- Modelo tidy
- Principales verbos de transformación
- Evaluación de la calidad de los datos
- Limpieza de una base de datos

Clase 3: Exploración y reportes

- Estadística descriptiva

- Principales tipos de gráficos
- Visualización con ggplot2
- Introducción a visualización con mapas
- Programación Funcional (loops)

Clase 4: Visualización de la información

- El paradigma 'ggplot'
- Gráficos comunes y extensiones
- Introducción al uso de datos espaciales
- Diseño de mapas

Módulo II – Análisis avanzado de datos y reportería con R

Clase 5: Regresiones en R

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

Clase 6: Exploración avanzada de datos

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

Clase 7: Comunicación con reportes y apps

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

Clase 8: 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)