



CENTRO DE
INTELIGENCIA
ARTIFICIAL

DESARROLLO de Talento 2021



FundaciónAxcel A.C.
Conectando huManos



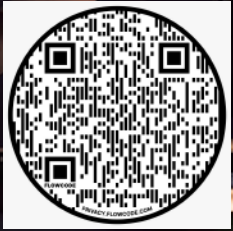
Microsoft



CENALTEC
CENTRO DE ENTRENAMIENTO EN ALTA TECNOLOGÍA



INADET
INSTITUTO DE APOYO
AL DESARROLLO
TECNOLÓGICO



Ciencia de datos usando R y Rstudio Nivel Básico



En español



Fecha de inicio
7 junio 2021



Duración
8 semanas



Dedicación
5 hrs semana



Certificación
IA Center y
SEP



40 horas



Solicita Beca
Microsoft



Modalidad
En línea vía
Discord



Horario
Mar y Jue
11:30 a.m - 14:00 p.m

Descripción

El objetivo de este curso es adquirir los conocimientos básicos del lenguaje R usando el entorno de desarrollo RStudio en los procesos de importar, limpiar, transformar, visualizar y modelar datos.

Objetivo

El objetivo de este curso es adquirir conocimiento sobre que es la ciencia de datos y su aplicación interdisciplinaria. De igual manera, conocer el cómo el lenguaje de programación R implementa diferentes estructuras de datos para el manejo y transformación de datos en información. En este curso se implementará el conjunto básico de librerías que componen el ecosistema de Tidyverse para apoyarnos en los procesos de importar, limpiar, visualizar, y analizar datos que servirán de base para la futura implementación de algoritmos de machine o deep learning.



Requisitos

Los interesados en participar en esta capacitación deberán tener conocimiento básico de estadística (descriptiva) y del uso de dispositivos y herramientas digitales (equipo de cómputo, internet, paquetes estadísticos como SPSS, Eviews, Stata)

Contar con un equipo de cómputo, acceso a internet, micrófono y cámara.



Dirigido a

Estudiantes y profesionistas que en sus actividades académicas o laborales usen datos para los procesos de análisis y toma de decisiones.



Perfil de ingreso

La persona tiene interés en conocer que es la ciencia de datos y su alcance de aplicación, así como interés en desarrollar habilidades en el manejo y transformación de datos usando el lenguaje de programación R y el entorno de trabajo RStudio



Perfil de egreso

La persona tiene el conocimiento de identificar las diferentes estructuras de los datos usando el lenguaje de programación R, así como tiene el conocimiento de importar diversos tipos de formatos (tsv, csv, Excel) para su limpieza, visualización y análisis usando el entorno de trabajo Rstudio.



Instructor

Dr. Alan Ponce

El instructor es investigador y profesor de tiempo completo e en la Universidad Autónoma de Cd. Juárez (UACJ), tiene un doctorado en la Universidad de Southampton en el Reino Unido, cuenta con varias certificaciones técnicas en el área de ciencia de datos y ha trabajado en el sector público, privado y académico en la transformación y manejo de datos para la toma de decisiones y publicaciones académicas.

train@ia.center

www.ia.center/desarrollo_de_talento_2021

1.○— Introducción a la Ciencia de datos

- Presentación del temario del curso, de los participantes y de la metodología de trabajo durante el curso.
- Identificar el rol de la ciencia de datos desde una perspectiva técnica e interdisciplinaria

2.○— Introducción a R

- Presentar el entorno de trabajo de RStudio
- Identificar y practicar con las diferentes estructuras de datos en R
 - o Vectores,
 - o Matrices,
 - o Factores
 - o Dataframes
 - o Listas

3.○— Introducción a Tidyverse

- Reconocer e implementar las diferentes librerías que componen el ecosystema de Tidyverse:
 - Estructura de datos “tidy”.
- Identificar y practicar los verbos que componen la gramática de la manipulación de los datos.

4.○— Importar datos en R

- Conocer y practicar con los diferentes formatos que R puede importar datos
 - o tsv.
 - o csv
 - o Excel
 - o Json
 - o Rds
 - o Sav
 - o Dta
 - o Mtp

5.○— Limpieza de datos en R

- Identificar problemas comunes en la limpieza de datos:
 - o Conversión de tipo de datos
 - o Identificación de duplicados
 - o Problemas con datos categóricos
 - o Reemplazar caracteres
 - o Remover caracteres
 - o Datos nulos

6.○— Visualización de datos I

- Introducción a la gramática de gráficos con ggplot:
 - o Identificación y practica con mapeos estéticos
 - o Introducción y practica con las geometrías más comunes como gráficos de dispersión, de barras y gráficos de líneas
 - o Conocer y practicar el concepto de la capa de temas en las visualizaciones

7.○— Visualización de datos II

- En este curso se complementarán cálculos estadísticos apoyándose en conceptos como las capas de coordenadas y las facetas que apoyan la comunicación visual con datos.
- Capa de coordenadas, “smoothing”, posición de objetos
- Capa de facetas, manejo de diferentes variables, etiquetado de facetas, márgenes.
- Identificación de mejores prácticas en la visualización de datos.

8.○— Análisis exploratorio

- Identificar y practicar con variables categóricas.
 - o Selección de variables
 - o Conteo vs Proporción
 - o Distribución de las variables
- Identificar y practicar con variables numéricas.
 - o Histogramas con capas facet
 - o Boxplots
 - o Gráficos de densidad
 - o Outliers
- Resúmenes numéricos
 - o Medidas de tendencia central y dispersión
 - o Calcular la medida de propagación
 - o Transformaciones

9.○— Análisis exploratorio: Caso de Uso

- Recolección y limpieza de datos en datos de interés aplicadas a un área de dominio.
- Unión y análisis de diferentes fuentes de datos.
- Visualización de datos.
- Modelado de datos para análisis estadístico
- Interpretación de resultados

10.○— Aprendizaje automático para todos

- Que es el aprendizaje automático.
- Flujo de trabajo del aprendizaje automático.
- Aprendizaje supervisado y no supervisado
 - o Regresión
 - o Clasificación
 - o Agrupamiento
- Tipos de algoritmos para datos estructurados y no estructurados.