



CENTRO DE  
INTELIGENCIA  
ARTIFICIAL

# DESARROLLO de Talento 2021



FundaciónAxcel A.C.  
Conectando huManos



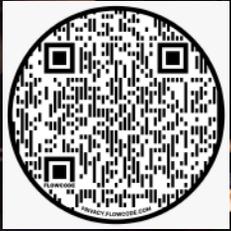
Microsoft



CENALTEC  
CENTRO DE ENTRENAMIENTO EN ALTA TECNOLOGIA



INADET  
INSTITUTO DE APOYO  
AL DESARROLLO  
TECNOLÓGICO



## Ciencia de datos usando Python y Anaconda Nivel Básico



En español



Fecha de inicio  
7 junio 2021



Duración  
8 semanas



Dedicación  
5 hrs semana



Software  
indicado



40 horas



Solicita Beca  
Microsoft



Modalidad  
En línea vía  
Discord



Horario  
Mar y Jue  
16-18:30 p.m



Certificación  
IA Center y  
SEP

### Descripción

El objetivo de este curso es adquirir los conocimientos básicos del lenguaje Python usando el ambiente de trabajo Anaconda en los procesos de importar, limpiar, transformar, visualizar y modelar datos.

### Objetivo

Adquirir conocimiento sobre que es la ciencia de datos y su aplicación interdisciplinaria. De igual manera, conocer el cómo el lenguaje de programación Python implementa diferentes estructuras de datos para el manejo y transformación de datos en información. En este curso se implementará el conjunto básico de librerías como pandas, matplotlib, numpy, seaborn, entre otras 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

El estudiante debe contar con equipo de cómputo, conocimientos básicos de programación, 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

Con interés por conocer qué 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 Python y el ambiente de trabajo Anaconda.



### Perfil de egreso

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



### Instructor

Dr. Alan Ponce

Investigador y profesor de tiempo completo e en la Universidad Autónoma de Ciudad Juárez (UACJ), tiene un doctorado en la Universidad de Southampton en 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

## 1.0— Introducción a la Ciencia de datos I

- 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

## 2.0— Python para la Ciencia de datos I

- Presentar el entorno de trabajo de Anaconda
- Identificar y practicar con las diferentes estructuras de datos en Python
  - Listas,
  - Diccionarios,
  - Tuples
  - Conjuntos
- Identificar y practicar con funciones en Python
  - Argumentos y alcance

## 3.0— Python para la Ciencia de datos II

- Identificar y practicar con funciones y paquetes en Python
  - Pandas
  - Numpy
  - Matplotlib
- Identificar y practicar con diferentes formas de manejar los datos como.
  - Seleccionar variables
  - Subconjuntos
  - Manejo de listas
  - Reemplazar / Eliminar elementos

## 4.0— Importar datos en Python

- Conocer y practicar con los diferentes formatos que Python puede importar datos
  - tsv.
  - csv
  - Excell
  - Json
  - Rds
  - Sav
  - Dta
  - Mat

## 5.0— Limpieza de datos en Python

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

## 6.0— Visualización de datos I

- Introducción a la gramática de graficos con Matplotlib/Seaborn:
  - Identificación y practica de uso y personalización de ejes
  - Leyendas anotaciones y estilos
  - Modificando colormaps
  - Introducción a Seaborn
    - Graficar regresión lineal
    - Visualizar distribuciones
    - Visualizar correlaciones

## 7.0— Visualización de datos II

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

## 8.0— Análisis exploratorio

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

## 9.0— 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

## 10.0— Aprendizaje automático para todos

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