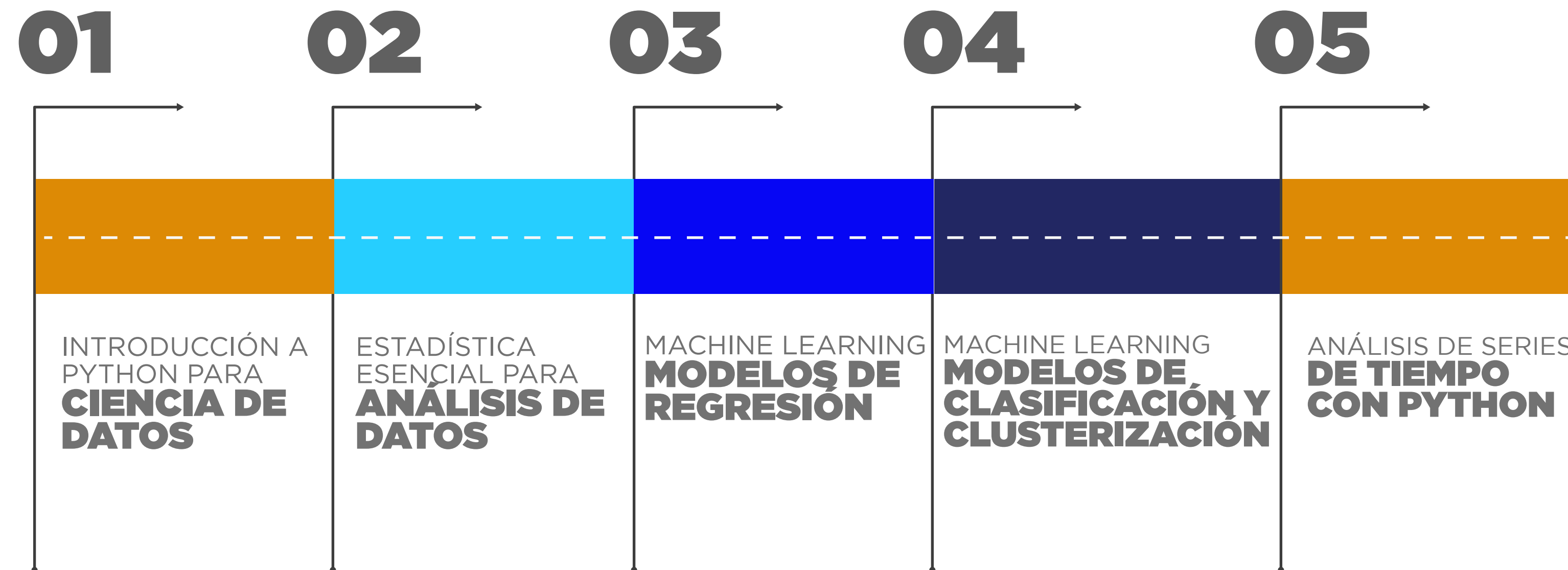


ESDPY

**ESPECIALIZACIÓN EN PYTHON PARA
CIENCIA DE DATOS Y MACHINE LEARNING**

CURSOS QUE COMPONEN LA ESPECIALIDAD



Este curso es la puerta de entrada a Python, uno de los lenguajes más populares en ciencia de datos. Iniciando desde los fundamentos, exploraremos la historia de la ciencia de datos y su relevancia en la actualidad. A través de clases prácticas, aprenderás a manejar estructuras de datos (listas, tuplas, diccionarios), realizar operaciones en álgebra lineal, y desarrollar habilidades en control de flujo y funciones personalizadas. También nos adentraremos en el manejo de DataFrames y la manipulación avanzada de datos: importación desde diversas fuentes, limpieza de datos, y visualización mediante gráficos dinámicos. Este curso establece una base sólida para todas las siguientes etapas de la especialidad.

TEMARIO A CUBRIR:

1. Introducción a Python y Ciencia de Datos

- a. Historia de la Ciencia de Datos
- b. Historia de Python y su aplicación a la Ciencia de Datos

2. Fundamentos de Programación con Python

- a. Variables y sus tipos
- b. Estructuras de Datos
 - i. Listas
 - ii. Tuplas
 - iii. Diccionarios
 - iv. Conjuntos
- c. Algebra Lineal
 - i. Escalares
 - ii. Vectores
 - iii. Matrices
 - iv. Tensores
- d. Funciones Personalizadas
- e. Estructuras de Control de Flujo
 - i. Condicionales (If, elif, else)
 - ii. Bucles (for, while)

- iii. Control de Bucles (break, continue)
- iv. Control de Excepciones (try, except, finally)
- f. DataFrames

3. Importación de Datos

- a. Excel
- b. Archivos Planos
- c. Carpetas
- d. SQL Server
- e. Web Scrapping

4. Manipulación y transformación de Datos

- a. Gestión de Índices
- b. (LOC - ILOC)
- b. Selección de Variables
- c. Filtros de Datos
- d. Operadores Lógicos - Búsqueda
- e. Condicionales
- f. Ordenamiento de Datos
- g. Agrupaciones

- h. Funciones de Agregación
- i. Pivot Table
- j. Combinación de Tables (Merge - JOIN)
- k. Fuzzy Wuzzy (Combinación de similitud)
- l. Unión de Tablas
- m. Funciones de Fecha
- n. Expresiones Regulares
- o. Funciones de Texto

5. Limpieza de Datos

- a. Validación y Gestión de Duplicados
- b. Coerción de Formato en los Datos
- c. Validación Cruzada (Integridad de los datos)
- d. Identificación de Valores Nulos
- e. Eliminación - Reemplazar Vacíos

- f. Función Anónima Lambda para rellenar valores

6. Visualización de Datos

- a. Gráfico de líneas
- b. Gráficos de Barras
- c. Gráficos Apilados
- d. Gráficos Combinados
- e. Gráficos de Dispersión
- f. Grafico Circular y de Anillo
- g. MitoSheet (Tablas y Gráficos Dinámicos)
- h. Gráfico de Pirámide



INTRODUCCIÓN A PYTHON PARA CIENCIA DE DATOS

Duración: 24 horas / **Requisitos:** Manejo de Excel Avanzado o experiencia previa en otras herramientas o lenguajes para análisis de datos.



ESTADÍSTICA ESENCIAL PARA ANÁLISIS DE DATOS CON PYTHON

La estadística es el corazón de cualquier análisis de datos, y en este curso aprenderás cómo aplicar tanto estadística descriptiva como inferencial en proyectos de ciencia de datos. Desde medidas de tendencia central hasta técnicas avanzadas para detectar valores atípicos (como Isolation Forest y DBSCAN), este módulo te prepara para explorar y comprender patrones en los datos. Además, cubriremos métodos de muestreo, intervalos de confianza, y el uso de pruebas de hipótesis (como ANOVA y T-Student), fundamentales para tomar decisiones basadas en datos. Con esta capacitación, serás capaz de analizar datos de manera crítica y estructurada.

TEMARIO A CUBRIR:

1. Estadística

Descriptiva

- Medidas de Rango
- Medidas de Tendencia Central
- Medidas de Dispersión
- Covarianza
- Datos Cuantitativos y Cualitativos

2. Análisis

Exploratorio de Datos

- Histograma
- Densidad
- Boxplot
- Matriz de Correlación

3. Valores Atípicos

- Técnicas detección Univariante
 - Método Rango Intercuartílico
 - Método Desviación Estándar
 - Método Puntuación Z
- Técnicas detección Multivariante
 - Isolation Forest
 - DBSCAN

- Factor Atípico Local

4. Análisis de Normalidad

- Representaciones Graficas
 - Histograma
 - QQ-Plot
- Contraste de Hipótesis
 - Shapiro Wilk
 - D'Angostino K-Squared
- Métodos Analíticos
 - Asimetría
 - Curtosis

5. Estadística Inferencial

- Muestreos
 - Probabilísticos
 - No Probabilísticos
- Estimación de Parámetros
 - Estimación Puntual
 - Estimación por Intervalo
- Remuestreo

- Resampling
- Bootstrap
- Permutacion
- Pruebas de Hipótesis
 - Contraste de Hipótesis
 - Conformidad
 - Homogeneidad
 - Independencia
 - Prueba T-Student
 - Prueba ANOVA
 - Prueba Chi-Cuadrado
 - Prueba Kolmogorov-Smirnov
 - Prueba Mann-Whitney U
 - Prueba Wilcoxon Signed Rank
 - Prueba Kruskal-Wallis H
 - Prueba Correlación Spearman
- Intervalos de Confianza

- IC para Media
- IC para Proporción
- IC diferencia de Medias
- IC para Varianza
- Probabilidad Estadística
 - Distribuciones Discretas
 - Distribuciones Continuas
 - Distribución Multivariada
 - Generación de Variables Aleatorias
 - Simulación de Datos según Distribución
 - Modelo Monte Carlo
 - Teorema Central del Limite
 - Ajuste de Distribución Paramétrica y No Paramétrica

Duración: 24 horas / **Requisitos:** Introducción a Python para Ciencia de Datos

Enfocado en el aprendizaje supervisado, este curso te guía en la construcción de modelos de regresión para predecir variables continuas. Comenzaremos con un análisis exploratorio para identificar relaciones en los datos, luego aprenderás técnicas de preprocesamiento de datos como la estandarización y la imputación. Exploraremos algoritmos de regresión lineal (Ridge, Lasso, ElasticNet) y métodos no paramétricos como KNN y Random Forest. Además, trabajaremos en la optimización de hiperparámetros y la validación de modelos para lograr resultados precisos y confiables. Este curso es ideal para quienes buscan una comprensión profunda de la predicción en ciencia de datos.

TEMARIO A CUBRIR:

1. EDA para Machine Learning

- a. Distribución Variables Cualitativas
- b. Correlaciones de Variables Numéricas
- c. Transformación Logarítmica o Raíz Cuadrática
- d. Ajuste Distribución AIC-BIC - DIC
- e. Relación de Variables Independiente contra la Dependiente

2. Preprocesamiento de Datos

- a. División Train-Test
- b. Estandarización y Escalado de Variables

- c. Imputación de Valores Vacíos
- d. Binarización de Variables Cualitativas

3. Algoritmos de regresión lineal

- a. Ridge
- b. Lasso
- c. ElasticNet
- d. OLS

4. Algoritmos No Paramétricos

- a. KNN
- b. Árbol de Decisión
- c. Random Forest
- d. Gradient Boosting
- e. Super Learner

5. Validación

- a. Validación Cruzada
- b. Validación Cruzada Repetida
- c. Diagnóstico de Residuos

6. Métricas de Precisión

- a. MAE
- b. MSE
- c. RMSE
- d. Coeficiente Determinación

7. Búsqueda de HiperParámetros

- a. GridSearch
- b. RandomSearch
- c. Optimización Bayesiana

Duración: 20 horas / **Requisitos:** Introducción a Python para Ciencia de Datos

MACHINE LEARNING MODELOS DE
REGRESIÓN CON PYTHON

Este curso profundiza en los modelos de machine learning que permiten clasificar datos en categorías. Cubriremos desde algoritmos de clasificación binaria y multiclase (como Regresión Logística y Random Forest) hasta técnicas de agrupamiento (k-Means y DBSCAN). El curso incluye el desarrollo de técnicas de ingeniería de características, balanceo de muestras, y selección de características. También aprenderás a evaluar el rendimiento del modelo usando métricas como precisión, F1-Score, y curvas ROC/AUC. Es una capacitación esencial para quienes desean desarrollar habilidades prácticas en clasificación y segmentación de datos.

TEMARIO A CUBRIR:

1. Feature Engineering

- a. Imputación de Valores Faltantes
- b. Normalización y Estandarización
- c. Pipelines
- d. Generación de Características Temporales
- e. Binning (Agrupamiento de Variables)
- f. Tratamiento de Outliers en Pipelines
- g. Descomposición de Componentes Principales
- h. Feature Selection
- i. Encoding
- j. Balance de Muestras (Oversampling y Undersampling)

2. Modelos Clasificación Binaria y Multiclase

- a. Regresión Logística
- b. Máquinas de Soporte Vectorial
- c. K Vecinos Mas Cercanos
- d. Árbol de Decisión
- e. Bosques Aleatorios
- f. Naive Bayes
- i. Categórico
- ii. Bernoulli
- iii. Gaussiano

3. Precision del Modelo

- a. Exactitud
- b. Precision
- c. Recall
- d. F1-Score
- e. Matriz de Confusion
- f. Reporte de Clasificación

- g. AUC/ROC
- h. Cohen's Kappa
- i. Log Loss

4. Búsqueda de HiperParámetros

- a. GridSearch
- b. RandomSearch

5. Serialización del modelo

- a. Guardar el Modelo
- b. Predicciones a datos nuevos

6. Modelos de Clusterización

- a. Cluster Particional y Basados en Densidad
- i. K-Means
- ii. DBSCAN
- iii. Affinity Propagation
- iv. Spectral Clustering

- b. Cluster Jerárquicos
- i. Jerárquico Aglomerativo
- ii. Jerárquico Divisivo
- iii. HDBSCAN
- c. Técnicas de Validación
- i. Metodo del Codo
- ii. Coeficiente de Silueta
- iii. Dendrograma

CURSO MACHINE LEARNING MODELOS DE CLASIFICACIÓN Y CLUSTERIZACIÓN

Duración: 20 horas / **Requisitos:** Introducción a Python para Ciencia de Datos



ANÁLISIS DE SERIES DE TIEMPO CON PYTHON

En este curso, aprenderás a trabajar con datos secuenciales en el tiempo, ideal para sectores que necesitan identificar patrones y hacer predicciones basadas en datos históricos. Abordaremos propiedades clave como estacionalidad y tendencia, y aplicarás técnicas de preprocesamiento como suavizado y resampling. Explorarás modelos clásicos como ARIMA y SARIMA, así como métodos avanzados como GARCH y modelos de regresión para series temporales. Además, aprenderás a evaluar la precisión del modelo y a presentar resultados en visualizaciones dinámicas. Este curso es ideal para cualquier persona interesada en predicciones a lo largo del tiempo.

TEMARIO A CUBRIR:

1. Propiedades de las Series de Tiempo

- a. Estacionalidad
- b. Tendencia
- c. Estacionariedad
- d. Ciclos

2. Preprocesamiento y Limpieza de Datos

- a. Imputación de Datos
- b. Identificar y tratamiento de Outliers
- c. Diferenciación
- d. Logaritmos
- e. Detrending y desestacionalización
- f. Resampling y Smoothing
 - i. Resampling
 - ii. Suavizado Exponencial Simple y Doble

- iii. Suavizado Holt-Winters

3. Exploración de Datos y Análisis de la Serie Temporal

- a. Gráficos de Líneas
- b. Descomposición de Series de Tiempo
- c. Métodos Aditivos y Multiplicativos
- d. Análisis de Componentes
- e. ACF y PACF

4. Modelos Clásicos de Series de Tiempo

- a. Suavizado Exponencial Simple
- b. Suavizado Exponencial Doble
- c. Suavizado Exponencial Holt-Winters

- d. Modelos AR – MA – ARMA – ARIMA
- e. Modelos SARIMA

5. Modelos Avanzados de Series de Tiempo

- a. ARIMAX
- b. GARCH
- c. LGMRegressor

6. Evaluación y Validación

- a. MAE
- b. MSE
- c. RMSE
- d. MAPE
- e. Validación Cruzada
- f. Backtesting

Duración: 20 horas / **Requisitos:** Introducción a Python para Ciencia de Datos



INFORMACIÓN GENERAL

ORDEN DE LOS CURSOS:

- Introducción a Python para Ciencia de Datos
Precio: \$126,480 / Duración: 24 horas
- Estadística Esencial para Análisis de Datos con Python
Precio: \$126,480 / Duración: 24 horas
- Machine Learning Modelos de Regresión con Python
Precio: \$110,160 / Duración: 20 horas
- Machine Learning Modelos de Clasificación y Clusterización con Python
Precio: \$110,160 / Duración: 20 horas
- Análisis de Series de Tiempo con Python
Precio: \$110,160 / Duración: 20 horas

PRECIO TOTAL: \$583.440 por persona IVA incluido

PRECIO EQUIVALENTE EN DÓLARES:

DURACIÓN TOTAL: 108 horas

REQUISITOS DE APROBACIÓN:

Al finalizar cada curso con una nota mayor a 70 se entrega un certificado de aprovechamiento. Si se culminaron todos los cursos con nota mayor a 70 se entrega el certificado final como Especialista en Python para Ciencia de Datos y Machine Learning.

EXAMEN DE UBICACIÓN VS EXAMEN DE CONVALIDACIÓN:

La persona puede realizar una prueba de ubicación para determinar si tiene el nivel adecuado para entrar directamente a uno de los cursos en específico, sin embargo, para completar el programa de especialidad debe hacer examen de convalidación para validar los conocimientos en este tema, el cual tiene un costo de \$30 y una duración de 2 horas.

CERTIFICACIÓN INTERNACIONAL:

Al cursar esta especialidad tendrás una excelente preparación en conceptos y técnicas de ciencia de datos y machine learning. Si deseas además obtener la certificación internacional de Microsoft como Azure Data Scientist Associate, debes aprobar el examen DP-100 Azure Machine Learning (costo no incluido en esta especialidad). Para prepararte apropiadamente para aprobar este examen, te recomendamos adquirir adicionalmente el curso DP-100 Azure Machine Learning de Grow Up Data Analytics.

MODALIDAD:

Virtual sincrónica (con instructor en vivo)

REQUISITOS:

- Manejo de Herramientas de Office (Excel)
- Equipo de Computadora con conexión a internet y capacidad de instalar herramientas de análisis de datos y programación como Python, Jupyter Notebook y Anaconda.
- Cuenta de Google y acceso a Google Colab
- Deseable conocimiento técnico básico de análisis de datos en herramientas como Power BI, Tableau, SQL, etc.
- Inglés a nivel de lectura

Más información al: [\(506\) 8414-4646](tel:50684144646)

Programa respaldado por la política de calidad de cursos Grow Up:

<https://www.growupcr.com/politicadecalidad>

growup
DATA ANALYTICS

MECÁNICA DE LA ESPECIALIDAD