



# CIENCIA DE DATOS Y MACHINE LEARNING CON PYTHON

PROGRAMA ESPECIALIZADO



## CONTACTO

[www.datayanalytics.com](http://www.datayanalytics.com) 

[cursos@datayanalytics.com](mailto:cursos@datayanalytics.com) 

(+51) 976 760 803 

# NUESTROS ALUMNOS NOS RECOMIENDAN



**Katy Pacheco**

*Me gusto, sobre todo porque el material es bastante actualizado.*

**José Taco Sarayasi**

*Es un curso extenso y detallado, por que hay mucha dedicación en el contenido del curso.*



**Diana Castillo Ramírez**

*Lo recomendaría porque fue un buen curso, nos llevó de la mano la instructora.*

**Lloyd Laura**

*Lo recomendaría a mis amigos o colegas que recién empiezan ya que el curso es económico y muy didáctico.*



**Oswaldo Villegas**

*Me agrada el curso, considero que el enfoque que lleva está bien definido.*

**Jordan Capa Zamata**

*Clases didácticas y bastante profesionalismo, en definitiva no será el único curso que lleve aquí, lo recomiendo bastante.*



# PRESENTACIÓN

La vasta generación de datos en todos los sectores, ha generado que el dominio de la ciencia de datos y en principal el dominio de Python sea una de las habilidades con mayor demanda a nivel mundial. Con este Programa Especializado Ciencia de Datos con Python desarrollarás las habilidades para desarrollar proyectos de ciencia de datos y analizar datos con Python desde cero; así, alcanzarás tus objetivos y pondrás en práctica lo aprendido en tu empresa y tus proyectos.

Estas son habilidades que te permitirán diferenciarte del resto, permitiéndote crecer profesionalmente, siendo más competitivo y lograr mejores oportunidades laborales.

## DÍAS

2 veces por semana

## DURACIÓN

3.5 meses

120 horas académicas

## DIRIGIDO A

- Estudiantes, egresados y profesionales de diversos rubros o áreas.
- Público en general que desea trabajar en Ciencia de Datos con Python.

## METODOLOGÍA

- Teórico - práctico con casos reales aplicados.
- Código en Python + Data + PPT's + Sesiones en vivo por ZOOM.
- Experiencia en la resolución de proyectos de ciencia de datos.
- Aprendizaje basado en resolución de casos y ejemplos.
- Desarrollarás 05 proyectos que te permitirán consolidar tus conocimientos.















## PROYECTOS

- Desarrollarás 05 proyectos (*personal y grupal*) que te permitirán consolidar tus conocimientos.

*Nota mínima aprobatoria: 15 (escala del 0 a 20)*

### HORARIO POR PAÍSES

Sesiones de 2 horas 30 min

	Nicaragua	6:00 p.m.
	El Salvador	6:00 p.m.
	Costa Rica	6:00 p.m.
	Perú	7:00 p.m.
	Panamá	7:00 p.m.
	Colombia	7:00 p.m.
	México CDMX	7:00 p.m.
	Ecuador	7:00 p.m.
	Cuba	8:00 p.m.
	Chile	8:00 p.m.
	Bolivia	8:00 p.m.
	Paraguay	8:00 p.m.
	Brasil	9:00 p.m.
	Argentina	9:00 p.m.



1

#### SOFTWARE PYTHON

#### 2 ENTIENDE EL NEGOCIO

- Los objetivos empresariales
- Posibles escenarios de solución
- Identificar fuentes de datos.
- Elaborar el plan de proyecto

#### 3 COMPRESIÓN DE DATOS

- Cargar la data de cada fuente.
- Realizar la exploración de los datos.
- Verificar cada fuente de datos y su estatus de calidad.

#### 4 PREPARACIÓN DE DATOS

- Selección y limpieza de datos
- Tratamiento de datos perdidos
- Tratamiento de datos outliers
- Construcción de datos
- Integración de dato
- Formato de datos

#### 5 MODELADO Y EVALUACIÓN

- Selección de algoritmos (mín 03).
- Selección de variables más importantes.
- Partición de muestras.
- Entrenamiento del modelo
- Afinamiento del modelo

#### 6 IMPLEMENTACIÓN

- Plan de implementación
- Monitoreo y mantenimiento del modelo
- Publicar modelo



*Basado en la metodología de CRISP-DM*

## APRENDE HACIENDO – MÉTODOLÓGÍA PARA SOLUCIONAR PROBLEMAS DE CIENCIA DE DATOS –



# CONTENIDO

## 01. Python + Colab

- Introducción
  - Introducción al lenguaje
  - Instalación y configuración de mi workspace en Colab
- Python básico
- Hello Python
- La interfaz con Python
- Variables y tipos
- Asignación de variables
- Cálculos con variables
- Otros tipos de variables
- Operaciones con otro tipo de variables
- Tipo de conversión

### Casos prácticos

## 02. Listas y estructuras

- Listas Python
- Crear una lista
- Crear listas con diferentes tipos
- Listas de listas
- Subconjunto de listas
- Subconjunto y calculo
- Cortado y tajado
- Manipulación de listas
- Reemplazo de elementos de listas
- Extensión de listas
- Estructuras de datos integradas
  - Tuplas
  - Diccionarios

### Casos prácticos

## 03. Numpy

- NumPy
- Primera matriz NumPy
- Subconjunto matriz NumPy
- Matriz 2D Numpy
- Estadísticas básicas en NumPy
- Manejo de errores y excepciones
- Excepciones, tipos y control de excepciones
- Programación orientada a objetos
- Clases y objetos
- Atributos y métodos

### Casos prácticos

## 04. Condicionales y Panda

- Condicionales
- Sentencias
- Expresiones
- Pandas
  - Data
  - Indexación
  - Órdenes
  - Datos perdidos
  - Operaciones
  - Agrupaciones
  - Estadísticas
  - Uniones



### Casos prácticos

## 05. Funciones y paquetes

- Funciones
- Funciones familiares
- Múltiples argumentos
- Métodos de cadena
- Métodos de lista
- Paquetes
- Importación de paquetes
- Importación selectiva
- Funciones lambda

### Casos prácticos

## 06. Visualización y Analytics

- Visualización de datos
- Análisis de datos
- Análisis univariado
- Análisis bivariado
- Random forest

### Casos prácticos Proyecto

## 07. Introducción a la estadística

- Conceptos básicos: Estadística, población, muestra, elemento muestral
- Tipos de variables: Categóricas (Nominales y ordinales) y Numéricas (Discretas y continuas)
- Secuencia / Metodología de análisis de datos
- Introducción a Colab
- Full introducción a Python

### Resolución de ejercicios

## 08. Análisis descriptivo

- Medidas de tendencia central: moda, mediana, media. (Aritmética, geométrica, armónica, recortada)
- Medias de dispersión: desviación estandar, varianza, rango, rango intercuartilico, coeficiente de variación.
- Coeficiente de asimetría y kurtosis
- Medidas de asociación: correlación
- Pruebas Chi Cuadrado

## 09. Tratamiento de datos nulos y outliers

- Tratamiento de datos nulos
- Tratamiento de outliers
- Imputación paramétrica y no paramétrica

*Casos aplicados*

## 10. Inferencia estadística

- Teoría de la inferencia estadística (paramétricas y no paramétricas)
- Teorema del límite central
- Pruebas de hipótesis
- Pruebas de normalidad
- Pruebas de correlación (paramétricas y no paramétricas)
- Prueba chi cuadrado
- Prueba de la media de 1 muestra
- Prueba de la media de 2 muestras
- ANOVA
- Prueba de Wilcoxon
- Prueba de Mann Whitney

*Resolución de ejercicios*  
*Proyecto*

## 11. Aprendizaje supervisado

- Creación de nuevas variables (atributos)
- Regresión lineal simple
- Regresión lineal múltiple
- Regresión logística
- Random forest
- Análisis de multicolinealidad
- Estimaciones
- Métricas: ROC, gini, accuray, KS, log-loss, entre otros.

*Resolución de ejercicios*

## 12. Aprendizaje no supervisado

- Anál isis de componentes principales
- Análisis clúster
- Validación de clústers
- Perfilamiento de clústers

*Proyecto*

## 13. EDA y tratamiento para datos

- Análisis exploratorio de datos
- Tratamiento de datos, OneHotEncoder, TargerEncoder
- Tratamientos de valores atípicos
- Imputación paramétrica (*media, mediana, moda*)
- Imputación no paramétrica para la imputación de valores perdidos
- División de los datos categóricos en números
- Creación de Driver
- División de los datos en conjunto de entrenamiento y prueba
- Normalización de los datos
- Blanceo de data: SMOTE, OverSampling, UnderSampling
- Selección de variables

*Resolución de ejercicios*

## 14. Aprendizaje supervisado para regresiones

- Regresión lineal
- KNN
- Árbol de decisión
- Suport Vector
- Ensemble methods
- AdaBoost, GBM, XGBoost
- Gridsearch de parámetros, RandomGridsearch
- Optimizacipon bayesiana
- Aplicación de un modelo TensorFlow

*Resolución de ejercicios*

## 15. Aprendizaje supervisado para clasificación

- Regresión logística
- Naive bayes
- Árbol de decisión
- KNN
- Random forest
- Support vector
- AdaBoost, GBM, XGBoost
- Gridsearch de parámetros, RandomGridsearch
- Optimización bayesiana

*Resolución de ejercicios*

- Validación de modelos
- Validación cruzada
- Matriz de confusión
- GINI, ROC-AUC, K-S, sensibilidad, recall, F-Score, Log-Loss

## 16. Aprendizaje no supervisado

- Aprendizaje no supervisado
  - K-means clustering
  - Análisis de cesta de canasta de mercado
  - Segmentación o puntuación RFM
- Reducción de dimensiones
  - PCA
  - LDA

## 17. Introducción a series de tiempo

- Introducción a series de tiempo
- Entendiendo series de tiempo
- Método baso en suavizamiento exponencial

*Resolución de ejercicios  
5to Proyecto*

# INSTRUCTORES



### Director de Proyectos en Data & Analytics

Ingeniero Estadístico e Informático, cuenta con más de 16 años de experiencia en Analytics, Modelos supervisados y no supervisados, Dirección de proyectos de ciencia de datos, investigación de mercados, diseño de estrategias de gestión y desempeño, desarrollo de modelos predictivos, calidad de datos y capacitación de equipos. Máster en Administración y Dirección de Marketing con Especialización en Machine Learning, Big Data, Analytics, Business Intelligence, Metodología de la Investigación, entre otros. Experiencia en el sector privado y público: consumo masivo, telecomunicaciones, educación, banca, AFP, previsión, hoteles, minería, finanzas, retail, buró, microfinanzas, contraloría, registros públicos, MINEDU, SUNAT, entre otros.

Herramientas favoritas: R Studio, Python, Julia, SPSS, Modeler, Minitab, Rapidminer, Oracle, SQL Server, Qlikview, Power BI, Google Cloud, Office Avanzado, MS Project, entre otros.



### Instructor en Data & Analytics

Magíster y Licenciado en Estadística, cuenta con más de 6 años de experiencia en análisis y modelamiento de datos bajo el enfoque clásico y bayesiano. Experiencia en el sector público y privado: INEI, MINEDU, INDECOPI, electricidad y AFP. Además cuenta con 3 años de experiencia en docencia universitaria en temas de Data Mining, Machine Learning, Estadística Bayesiana, Modelos de pronóstico, y Muestreo estadístico en las principales universidades del Perú (UNI, UNMSM, UNICA, UPC, UTEPSA, Universidad Continental, etc).

Herramientas favoritas: Rstudio, Python, Julia, Matlab, SPSS, Stata, Minitab, SQL, Power BI, entre otros.

# CERTIFICADO

- El alumno obtendrá 4 certificados a nombre de Data & Analytics, una vez aprobado con la nota mínima establecida.
- Duración del curso en general: **3.5 meses**  
(120 horas académicas: 96 horas pedagógicas + 24 horas solución de proyectos)
- Los certificados serán por cada curso establecido dentro del programa:



Programa Especializado en  
Ciencia de datos y Machine  
Learning con Python

120 horas académicas



Curso Especializado en  
Estadística para Ciencia de  
Datos con Python

40 horas académicas



Curso Especializado en  
Python para Ciencia de Datos

40 horas académicas



Curso Especializado en  
Machine Learning con Python

40 horas académicas

## MATERIALES

Se entregará:

- Scripts en Python, PPT's, Data.
- Sesiones en vivo por ZOOM.

## BENEFICIOS

- Acceso directo a los eventos gratuitos organizados por nosotros: Foros, Talleres, Webinars, Mini cursos
- Bolsa de trabajo.



(+51) 976 760 803



cursos@datayanalytics.com



www.datayanalytics.com





# PROCESO DE INSCRIPCIÓN

01

Enviar un correo a [cursos@datayanalytics.com](mailto: cursos@datayanalytics.com) con copia a [datayanalyticspe@gmail.com](mailto: datayanalyticspe@gmail.com) la siguiente información:

- Nombres y Apellidos
- Fecha de nacimiento
- Nro de documento de identidad
- Nro de celular
- Gmail
- Empresa donde labora
- Foto del comprobante de pago

**Asunto:** PROGRAMA PYTHON



Si desea solicitar factura, enviar:

- \* Nro de RUC
- \* Razón social
- \* Dirección (País)

Una vez registrado el alumno, recibirá 2 correos con la siguiente información:

- **ZOOM** para las clases en vivo.
- Acceso a los materiales y videos, mediante **Google Classroom**.

02

## MÉTODOS DE PAGO

**R&M INGENIEROS ASOCIADOS GROUP**  
RUC: 20600323441



Cuenta Soles  
CC: **488 - 3002002636**  
CCI: **003-488-003002002636-42**



**960 876 044**



Nombre: **Yaneth Ccahuantico Mendoza**  
DNI: **41809490**  
País de destino: **Perú**



Tarjeta de crédito hasta 6 cuotas sin intereses:



Solicitar **link de pago**

Otros medios de pago, contactar al (+51) 916 591 245 ó (+51) 976 760 803



(+51) 976 760 803



[cursos@datayanalytics.com](mailto: cursos@datayanalytics.com)



[www.datayanalytics.com](http://www.datayanalytics.com)



# PRECIO

**PRECIO REGULAR**



<b>EN SOLES</b>	<b>EN DÓLARES</b>
<b>s/1850</b>	<b>us\$ 510</b>

- Estudiantes pre-grado universidad o institutos con descuento del 10%
- Grupos de 2 y más 10% de descuento.
- Los precios son para personas naturales.
- Las cuotas deben ser canceladas hasta un día antes del inicio del curso.
- Para matrículas corporativas, solicitar una cotización a [cursos@datayanalytics.com](mailto:cursos@datayanalytics.com)

## NUESTROS PARTNER



ID.MPN 6504783

**Microsoft**  
Partner Network



## CONTÁCTANOS

Erika Ayala | Coordinación  
[cursos@datayanalytics.com](mailto:cursos@datayanalytics.com)  
Tel.: (+51) 916 591 245

Mishell Bardales | Coordinación  
[mishell.bardales@datayanalytics.com](mailto:mishell.bardales@datayanalytics.com)  
Tel. : (+51) 976 760 803



[www.datayanalytics.com](http://www.datayanalytics.com)



60%