

**INSTITUTO DATA SCIENCE ARGENTINA**

Educación a distancia

**DIRECCIÓN ACADÉMICA**

Riobamba 1059 C1116ABE

C.A.B.A., Argentina

**CONTACTO**

Mail: [info@institutodatascience.org](mailto:info@institutodatascience.org)

Teléfono: +54 9 11 3914 8021

# Diplomatura en Data Analytics Con R y Python

[www.institutodatascience.org](http://www.institutodatascience.org)

# Diplomatura en Data Analytics con R y Python

**Duración: 24 Semanas.**

**Carga Horaria: 2 clases por semana.**

**Modalidad: A distancia, en vivo.**

## ¿Por qué aprender Data Analytics?

La analítica de datos es el proceso de extraer información de los datos para tomar decisiones. Se utiliza en una variedad de industrias, tales como las finanzas, la salud y el gobierno. Data analytics implica el análisis de datos para mejorar

la eficiencia operativa de las organizaciones y perfeccionar su modelo de negocio. Permitiendo tomar las mejores decisiones al gestionar los riesgos y lograr un crecimiento rentable y sostenible.

LA CALIDAD  
ACADÉMICA DE LA  
DIPLOMATURA EN  
CIENCIA DE DATOS  
CON R Y PYTHON  
HA SIDO AVALADA  
POR LA RESOLUCIÓN  
NRO 290/20 DE  
LA UNIVERSIDAD  
TECNOLÓGICA  
NACIONAL UTN FRRQ.



### **Clases personalizadas live online**

Todas las clases en vivo se dictan por Zoom. Quedan grabadas a disposición de los estudiantes y no se exige asistencia a clases. También se dispone de clases de ediciones anteriores para que puedan reforzar los conocimientos.

### **Consultas online**

Acceso a los profesores, tutores por chat grupal. Se coordinan videollamadas.

### **Aprende haciendo**

Un fuerte componente práctico orientado a proyectos para que aprendas haciendo. Al final de la cursada puedes realizar tus propios proyectos.

# Modalidad Online

LA DIPLOMATURA HA SIDO DISEÑADA Y DESARROLLADA POR UN EQUIPO DE PROFESIONALES CON AMPLIA EXPERIENCIA, TE PROPORCIONARÁ LAS PRINCIPALES HERRAMIENTAS ACTUALES DE LOS LENGUAJES DE PROGRAMACIÓN.

## Data Analyitcs para Todos

La cursada contiene tres etapas. Cada etapa es acompañada con clases en vivo personalizadas, materiales y prácticas.

Dependiendo de la experiencia previa de cada estudiante es la etapa por la que debe iniciar su cursada.

Las etapas son: Nivelación

positiva en Programación y Bases de Datos.

Introducción a los conceptos de data science y sus herramientas.

La tercer etapa es de aplicación de los algoritmos. Durante la cursada también se ven conceptos de Estadística y Oratoria.

## Objetivos

- a. Entiendan los conceptos del campo.
- b. Manejen los principales algoritmos.
- c. Sepan medir con criterio del negocio los resultados de sus modelos.
- d. Participen de las experiencias aportadas por los instructores.
- e. Realicen proyectos enfocados a sus áreas de interés.

# Programa Académico

1

## Introducción Conceptual

Introducción a Data Analytics. Niveles a los que opera la ciencia de datos. Introducción a Data Warehouse. Introducción a Data Mining. Introducción a Knowledge Discovery. Introducción a Herramientas OLAP y Tableros de comando. Repaso de herramientas disponibles. Taxonomía de las competencias de un científico de datos.

2

## Introducción a R

Variables, vectores y matrices. Dataframes. Paquetes y librerías. Manejo de archivos csv. Conversiones de tipos. Conexión a bases de datos. Ejecución condicional. Bucles. Funciones en R. Medidas estadísticas. Correlaciones. Funciones estadísticas.

3

## Introducción a Python

Descarga e instalación. Principales librerías. Variables y tipos de datos. Listas, tuplas y diccionarios. Ejecución condicional. Ciclos definidos e indefinidos. Manejo de Excepciones. Funciones y Generadores. Clases y objetos. Manejo de archivos y directorios. Conexión a archivos planos. Conexión a Excel. Conexión a Bases de Datos.

4

## Análisis Exploratorio

Uso de NumPy. Uso de Pandas. Uso de Matplotlib. Módulo matplotlib. Módulo math. Módulo numpy. Librería Mito. Librería SweetViz.

5

## Test básicos

Funciones de distribución en R. Histogramas en R. Gráficos de líneas en R. Gráficos de áreas en R. Uso de NumPy. Uso de Pandas. Uso de Matplotlib. Módulo matplotlib. Módulo math. Módulo numpy. Módulo yt. Módulo mayavi.

6

## Árboles de decisión

Algoritmo básico en Excel. División en entrenamiento y prueba. Uso de Rpart y cp. Predicción y valoración de la solución. Uso de Party, Aplicación al problema del call center. Instalación y uso del paquete en Python. Ejemplo en Python. Principales parámetros de ajuste y control.

7

## Clusters

Algoritmo básico en Excel. Uso de kmeans en R. Ejemplo de aplicación real en R. Otros algoritmos de agrupamiento en R. Ejercicio de aplicación de agrupamiento de mascotas. Uso de Scikit-Learn en Python. Análisis de componentes principales en Python. K-Medias en Python. Clustering jerárquico en Python.

8

## Reglas de Asociación

Algoritmo básico en Access. División en entrenamiento y prueba en R. Uso de arules en R. Ejemplo de una aplicación real votación de reglas en R. Ejemplo de una aplicación real a datos de ventas en supermercados en R. Instalación y uso del paquete en Python. Ejemplo en Python. Principales parámetros de ajuste y control. Esquema de votación de reglas en Python.

9

## Redes Neuronales

Algoritmo básico en Excel. División en entrenamiento y prueba en R. Uso de neuralnet. Predicción y valoración de la solución en R. Discusión del problema de legibilidad de los resultados desde el negocio. Competencia entre árboles y redes en un caso concreto en R. Instalación y uso del paquete en Python. Ejemplo en Python. Principales parámetros de ajuste y control en Python. Problema concreto en Python.

11

## Algoritmos genéticos

Algoritmo básico en Excel. Discusión del tipo de problemas en los

que se aplica. Discusión de otros mecanismos de optimización. Implementación en R del uso de algoritmos genéticos como valores reales. Binarios. Permutaciones. Implementación en Python.

**10**

### **Series temporales**

Taxonomía. Separación de componentes. Predicciones. ARIMA implementado en R. Predicción en series con un único período. Predicción en series con múltiples períodos. Predicción en series con períodos variables: renormalización. Implementación en Python.

**11**

### **Minería de textos**

Clasificación supervisada de piezas de texto. Construcción de una red semántica. Discusión de las posibilidades de reconocimiento de voz. Ejemplos de aplicación real del algoritmo de clasificación. Aplicación a la detección de sentimientos. Ejemplo de implementación en Python.

**12**

### **Vecinos Cercanos (Knn)**

Instalación y uso del paquete en R. Ejemplo conceptual en Excel. Ejemplo en R. Instalación y uso del paquete en Python. Ejemplo en Python. Principales parámetros de ajuste y control.

**13**

### **Bayes Ingenuo**

Implementación en Excel. Ejemplo de aplicación en R. Comparación de 4 algoritmos predictivos (Bayes ingenuo, árboles, redes neuronales y regresión logística). Ejemplo de aplicación en Python.

**14**

### **Random Forest**

Bootstrap y bagging. Descripción conceptual del método random forest. Paquete randomForest para R. Ejemplo de aplicación. Comparación con otras técnicas. Implementación en Python.

15

**Métodos bayesianos avanzados**

Descripción conceptual del método. Regresión lineal bayesiana. Regresión logística bayesiana. Inferencia bayesiana. Red bayesiana. Ejemplos de aplicación en R. Paquete BAS. Paquete brms. Paquete arm. Paquete bnlearn. Comparación con otras técnicas. Ejemplos de aplicación en Python.

16

**Máquina de soporte vectorial**

Descripción conceptual del método. Paquete e1071 para R. Ejemplo de aplicación en R. Ejemplo de aplicación en Python. Comparación con otras técnicas.

17

**Discriminante lineal y cuadrático**

Instalación y uso del paquete. Ejemplo conceptual en Excel. Ejemplo en R. Ejemplo en Python. Principales parámetros de ajuste y control en R. Principales parámetros de ajuste y control en Python. Problema concreto.

18

**Diseño de Datawarehouses**

Diferencias entre los DW y los OLTP. Tipos de datos y soportes. Dimensiones y jerarquías. Estimación de recursos y tiempos según tipos de datos. Ejemplos de staging. Interacción entre equipos (DB-ETL-Soporte-FrontEnd-Usuarios). Ejercicios de diseño de DW.

19

**Diseño y construcción de ETL**

Procesos ETL: generalidades, buenas prácticas, esquemas de ETL. Detalles de cada etapa: extracción, limpieza, normalización, transformación, carga. Estrategias de update. Ejercicios de diseño de ETL para la carga inicial. Ejercicios de diseño de ETL para la actualización. Licencia, descarga e instalación de Open Refine. Uso general como herramienta de limpieza de datos. Pre-procesado de los datos. Casos de interés y ejemplos. Licencias, descarga e instalación de Pentaho-Kettle. Pentaho Kettle: funciones, características, utilización. Auditoría y documentación. Ejemplos de uso.

# Proyectos de la cursada

---

## **Selección de prospectos**

Aprenderás a seleccionar los prospectos de venta que tengan la máxima rentabilidad o máxima ganancia.

## **Selección de personal**

Aprenderás a seleccionar personal dependiendo de ciertos factores.

## **Cartera de Acciones**

Distribuir la cartera de acciones de manera de maximizar la rentabilidad y minimizar la variabilidad.

## **Optimización de Cajas**

Cómo optimizamos la cantidad de cajas que tiene que haber en un supermercado para controlar la demora.

## **Muchos más proyectos**

Durante tu cursada seguirás viendo más proyectos que te permitirán preparar el camino para tu proyecto final de cursada.

## **Proyecto final de cursada**

Como condición de aprobación el estudiante debe presentar su propio proyecto final que realizará con tutoría docente.



# Instancias de evaluación

Autoevaluación recursos de nivelación positiva (opcional).

Presentación aprobada del proyecto final de cursada.

Examen final de la diplomatura (Mínimo 60 % de respuestas correctas).

## Equipo docente

---



**Ignacio Urteaga**



**Nicolás del Frate**



**Anahí Romo**



**Cesar Procopio**



**Claudia Piaggio**



**Jerónimo Pardo**





# **INSTITUTO**

Data Science

[www.institutodatascience.org](http://www.institutodatascience.org)