

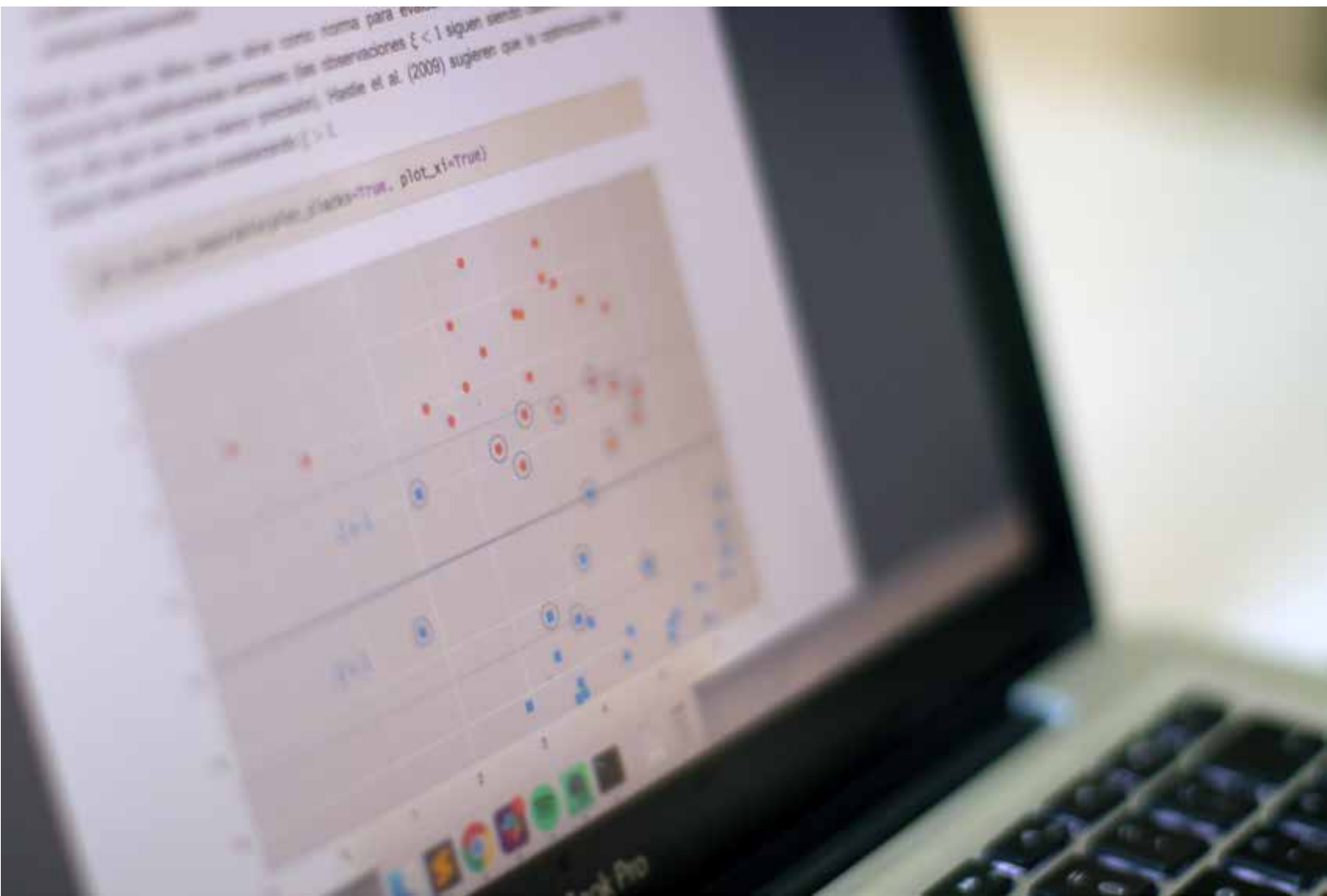
Curso
Introducción a
Data Science

{desafío}
latam_

Descripción del Programa

Introducción a Data Science es un curso práctico, diseñado para manejar las principales técnicas de análisis, buscando introducir a los estudiantes en el mundo de la programación orientada a la estadística, y en el desarrollo de capacidades críticas de interpretación de resultados con el lenguaje de programación Python. Se revisarán las técnicas de análisis e interpretación y visualización de datos, manipulación de tablas y construcción de modelos predictivos desde la econometría y el Machine Learning.

Este programa contempla los dos primeros módulos de la carrera de Data Science que imparte Desafío Latam.



Unidades y Contenidos



Módulo 1

Introducción a la programación con Python

Unidad 1

Introducción a la programación

- Instalar herramientas: Editor, Python, Python interactivo, Pip.
- Ejecutar Python desde el terminal.
- Ejecutar Python desde el editor de texto.
- Realizar diagramas de flujo y pseudocódigo.
- Construir aplicaciones tipo calculadora.
- Manejar flujos y operadores lógicos.

Unidad 2

Ciclos y métodos

- Leer y transcribir diagramas de flujo con interacciones a código Python.
- Analizar e implementar diagramas de flujo con repeticiones.
- Identificar componentes de un flujo for .
- Validar entradas de un iterador.
- Conocer el concepto de complejidad algorítmica y sus implicaciones para el desarrollo de flujos.
- Conocer la notación Big-O para el cálculo de complejidades en función a la cantidad de ciclos.
- Identificar los elementos que componen una función.
- Conocer el scope de una función.
- Identificar el alcance de variables globales y locales.
- Seguir la orden de ejecución de una función mediante pdb .
- Entender el retorno implícito return None en Python.
- Implementar retornos explícitos en la función.
- Conocer el principio Don't Repeat Yourself y su relevancia para la implementación de código.
- Identificar los argumentos por defecto y opcionales.
- Conocer args y kwargs .

- Manejar excepciones.
- Conocer la diferencia entre error y excepción.
- Conocer los antipatrones utilizados en las excepciones.
- Implementar excepciones mediante try y except .
- Realizar debugging.
- Dividir un proyecto en varios archivos.
- Instalar componentes vía PIP

Unidad 3

Estructuras de datos

- Diferenciar las estructuras de datos de datos básicos de Python: list , tuple y set .
- Manejar listas.
- Resolver problemas típicos de listas: reducciones, transformaciones, filtros, selecciones y lectura de datos desde archivos utilizando pandas .
- Implementar funciones vectorizadas utilizando numpy .
- Realizar persistencia.

Unidad 4

API

- Guardar datos en archivos.
- Entender el objetivo de una API.
- Conocer los principales componentes de una API.
- Conocer las herramientas que utilicen API para la interacción (Postman).
- Conocer la lógica de consumo de API.
- Utilizar Postman para realizar requests a una API.
- Conocer los endpoints.
- Conocer y entender los verbos REST.
- Conocer el formato de archivos JSON.
- Utilizar Python para realizar un request a una API.
- Consumir los datos desde Python.
- Guardar los requests de una API en un JSON y procesarlos dentro de Python.
- Conocer y comprender la importancia del protocolo HTTPS.
- Utilizar Python para realizar un request mediante HTTPS.



Módulo 2

Fundamentos de Data Science

Unidad 1

Estadística Univariada y Control de Flujo

- Conocer los principales modos de trabajo con Jupyter Notebook.
- Utilizar las estructuras de datos de `pd.Series` y `pd.DataFrame`.
- Analizar datos de forma univariada con `pandas`.
- Utilizar control flujos para obtener medidas estadísticas.

Unidad 2

Probabilidades y Funciones

- Utilizar funciones para reutilizar código. (Principio D.R.Y)
- Convertir una fórmula matemática a una función en Python.
- Construir y utilizar funciones orientadas al análisis de datos.
- Optimizar funciones reemplazandolas por funciones vectorizadas.
- Utilizar conceptos básicos de probabilidad.
- Generar segmentaciones de un `pd.DataFrame` en base a indexación y selección.

Unidad 3

VARIABLES ALEATORIAS Y GRÁFICOS

- Hacer uso de métodos de `pandas` para segmentar columnas y filas.
- Hacer uso de los métodos `iterrows` e `iteritems` para implementar loops en `pandas`.
- Implementar `enumerate` en loops.
- Conocer las convenciones y principios rectores de la visualización de gráficos.
- Conocer las principales convenciones en la visualización de resultados en histogramas, gráficos de punto y barras.
- Generar simulaciones de la distribución normal.
- Conocer las principales aplicaciones de las distribuciones.
- Calcular e interpretar puntajes z .
- Describir la Ley de los Grandes Números y Teorema del Límite Central y su importancia en la inferencia estadística.

Unidad 4

Hipótesis y Correlación

- Conocer las funcionalidades avanzadas de gráficos estáticos mediante seaborn.
- Aprender a segmentar datos y los principales criterios de estratificación.
- Conocer los principales criterios de transformación de variables.
- Aplicar funciones a columnas de datos mediante ufuncs, map-reduce-filter.
- Entender e interpretar la correlación a partir de diagramas de dispersión.
- Entender el marco inferencial frecuentista de las hipótesis.
- Conocer la distribución t de Student y su aplicación.
- Aplicar pruebas de hipótesis simples en el contexto de la inferencia.

Unidad 5

Regresión

- Reconocer la terminología asociada a la modelación estadística.
- Conocer la regresión lineal y sus fundamentos.
- Interpretar los parámetros estimados en la regresión.
- Conocer y ser capaz de interpretar estadísticos de bondad de ajuste y coeficientes.
- Reconocer los supuestos en los que la regresión tiene sustento teórico.
- Implementar un modelo de regresión con statsmodels.
- Utilizar transformaciones simples en las variables independientes.
- Implementar un modelo predictivo con scikit-learn.

Unidad 6

Clasificación

- Conocer la regresión logística y sus fundamentos.
- Conocer y ser capaz de interpretar estadísticos de bondad de ajuste y coeficientes.
- Reconocer los supuestos en que tiene sustento teórico.
- Implementar un modelo de regresión con statsmodels.
- Implementar un modelo predictivo con scikit-learn.
- Conocer los conceptos de validación cruzada y medidas de desempeño.

Unidad 7

Dimensionalidad y Agrupación

- Entender el problema de la "maldición de la dimensionalidad" y sus implicancias para el modelo.
- Conocer la aproximación psicométrica del
- Principal Component Analysis y el Análisis Factorial.
- Implementar algoritmos de reducción de dimensiones (Principal Components Analysis) y de reconocimiento de estructuras latentes (Análisis Factorial) con scikit-learn.
- Utilizar técnicas para identificar patrones de datos perdidos.
- Implementar algoritmos de agrupación (k-Means).

Unidad 8

Modelos Generalizados

- Conocer los componentes del marco analítico de los Modelos Lineales
- Generalizados (Componentes estocásticos, sistemáticos y funciones de enlace).
- Conocer el método de estimación por Máxima Verosimilitud con el que se estiman los Modelos Lineales Generalizados.
- Identificar la correcta implementación de los modelos en base a la naturaleza del problema.
- Implementar modelos mediante la librería statsmodels acorde a la naturaleza del problema.
- Interpretar las estimaciones de manera correcta tomando en cuenta las funciones de enlace asociadas a cada modelo.

Duración

- **13 semanas**
- **Sesión online:**
78 horas (13 sesiones de 6 horas cada una)
- **Sesión presencial:**
78 horas (26 sesiones de 3 horas cada una)

Requerimientos

Características de tu notebook*

Sistema Operativo: Windows, Linux o Mac
Procesador Intel Core i3, 8GB RAM, 128 Disco SSD

Plataformas y Software

- Empieza (<https://empieza.desafiolatam.com>)
- Jupyter Notebook
- Python Kernel
- Anaconda

** El notebook es por cuenta de todos los participantes: docente, ayudante y alumnos.
**Programas open source, por lo que el estudiante no necesita incurrir en gastos de licencias.*



Curso

Introducción a Data Science

{desafío}
latam_

www.desafiolatam.com



/DesafioLatam



/DesafioLatam



/DesafioLatam



/DesafioLatam