

{desafío}
latam_

Data Analysis _



Itinerario

Introducción al
framework de trabajo

Flujo de trabajo

Desafío

Introducción al framework de trabajo

Jupyter Notebook

- Jupyter funciona como un punto intermedio entre el kernel de Python y Markdown
- Existen dos modos:
 - Markdown
 - Code Chunks
- Existen kernels en múltiples lenguajes

Iniciar Jupyter Notebook

```
cd <nuestro_directorio>  
  
jupyter notebook
```

Nos debe redirigir a localhost:8888 de forma automática

Python

- Lenguaje objeto orientado con énfasis a la claridad del código
- Ampliamente utilizado por las comunidades científicas y privadas
- Cuenta con un amplio catálogo de librerías, algunas de ellas son:
 - numpy
 - pandas
 - matplotlib

Markdown

- Lenguaje de marcas extremadamente flexible
- Permite la exportación a múltiples formatos, entre los que destacan:
 - .html
 - .doc
 - .pdf
- Su objetivo es permitir utilizar un formato fácil de leer, fácil de escribir y fácil de convertir.

Flujo de trabajo

- Identificar la naturaleza de los datos
- Identificar la estructura de las variables

Enfoque analítico

Pasos obligatorios

1. Análisis exploratorio
2. Feature engineering

Estrategia de modelación

	Supervisado	No Supervisado
Continúa	Regresión	Dimensionalidad
Discreta	Clasificación	Clustering

- Para problemas supervisados: $y = f(x) \sim \text{Distribucion}(\cdot)$
- Para problemas no supervisados: $y^* = f(\mathbf{X})$

Desafío

Objetivo:

Explorar la base de datos **nations.csv** que contiene indicadores de salud y economía de una muestra de países. Con ello aprenderás a manipular, graficar y analizar datos sin la necesidad de tener conocimientos previos de Data Science.

Análisis exploratorio

Gráficos → Simples y bonitos

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Problema de regresión

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

- **Econometría:** La regresión busca la recta de mejor ajuste.
- **Machine Learning:** La regresión es un problema de predicción de cantidades.

{desafío}
latam_

*Academia de
talentos digitales*

www.desafiolatam.com