

Programador Junior en Machine Learning I



**Argentina
programa
4.0**



**Ministerio de
Desarrollo Productivo
Argentina**

**Secretaría de
Economía del Conocimiento**

Semana 5

+ Introducción al Análisis Exploratorio de Datos (AED)

- + ¿Qué es el AED y por qué es importante?
- + Descubrimiento de patrones, tendencias y relaciones ocultas

+ Librería Pandas en Python

- + ¿Qué es Pandas y por qué es popular en análisis de datos?
- + Estructuras de datos principales: Series y DataFrames
- + Importación de Pandas en Python



Introducción al Análisis Exploratorio de Datos (AED)

¿Qué es el AED y por qué es importante?

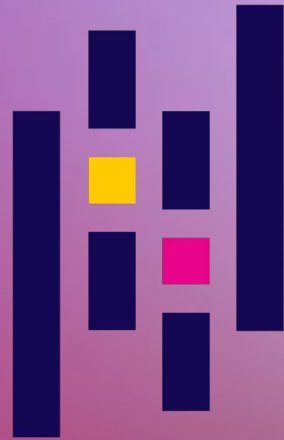
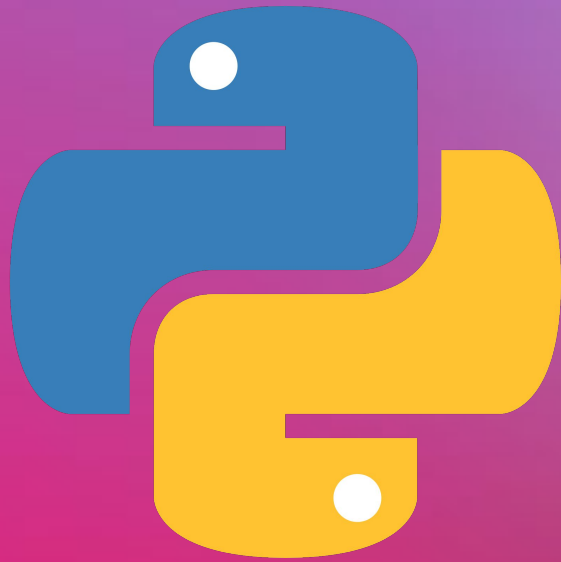
El Análisis Exploratorio de Datos (AED) es una etapa fundamental en cualquier proyecto de análisis de datos. Se trata de un enfoque que nos permite comprender y familiarizarnos con los datos antes de realizar análisis más avanzados.

El objetivo principal del AED es descubrir patrones, tendencias y relaciones ocultas en los datos. A través de técnicas estadísticas y visualizaciones, podemos explorar los datos de manera sistemática y obtener información valiosa.

El AED nos ayuda a responder preguntas clave como: ¿Qué información contienen los datos? ¿Qué variables están presentes y cómo se relacionan entre sí? ¿Existen valores atípicos o datos faltantes? Estas respuestas nos permiten tomar decisiones informadas y diseñar estrategias adecuadas para nuestro análisis.

Algunas de las técnicas comunes utilizadas en el AED incluyen: análisis descriptivo, visualización de datos, identificación de valores atípicos, análisis de correlación, agrupación y segmentación de datos, entre otros.

Librería Pandas en Python



pandas

¿Qué es Pandas y por qué es popular en análisis de datos?

La biblioteca Pandas es una de las herramientas más populares y poderosas para el análisis de datos en el lenguaje de programación Python. Proporciona estructuras de datos flexibles y eficientes, así como una amplia gama de funciones para manipular y analizar datos.

La estructura de datos principal en Pandas es el DataFrame, que es una tabla bidimensional con etiquetas en filas y columnas. Los DataFrames son ideales para organizar y manipular datos tabulares, similares a una hoja de cálculo o una base de datos.

Pandas es ampliamente utilizado en la comunidad de ciencia de datos debido a su capacidad para manejar y manipular datos de manera eficiente. Puede leer y escribir datos en varios formatos, incluyendo CSV, Excel, SQL y más. Además de la manipulación básica de datos, Pandas ofrece funcionalidades avanzadas, como la agrupación y agregación de datos, la fusión y unión de conjuntos de datos, la transformación de datos y el manejo de datos faltantes. La combinación de Pandas con otras bibliotecas populares de Python, como NumPy, Matplotlib y Scikit-learn, brinda un ecosistema completo para el análisis de datos, la visualización y el aprendizaje automático.

Sitio oficial: <https://pandas.pydata.org/>

Importación de Pandas en Python

Importar Pandas en Python es muy sencillo. Utilizamos la siguiente convención de importación:

```
import pandas as pd
```

Esto nos permite acceder a las funciones y estructuras de datos de Pandas utilizando el alias "pd".