**Pandas** es una de las bibliotecas más importantes para el análisis de datos en Python, proporcionando estructuras de datos y herramientas de análisis de alto rendimiento. Aquí te doy un resumen de las funciones y métodos más importantes de Pandas a lo largo de las distintas fases del análisis de datos, desde la exploración inicial hasta la preparación para *machine learning*.

**1. Importación y carga de datos**

**- pd.read\_csv()**: Lee un archivo CSV y lo convierte en un *DataFrame*.

**- pd.read\_excel()**: Lee un archivo Excel y lo convierte en un *DataFrame*.

**- pd.read\_sql()**: Ejecuta una consulta SQL y devuelve los resultados en un *DataFrame*.

**- pd.read\_json()**: Lee un archivo JSON y lo convierte en un *DataFrame*.

**- pd.read\_html()**: Extrae tablas de HTML y las convierte en *DataFrames*.

**2. Exploración de datos**

**- df.head()**: Muestra las primeras filas del *DataFrame*.

**- df.tail()**: Muestra las últimas filas del *DataFrame*.

**- df.info()**: Proporciona un resumen de la estructura del *DataFrame*, incluyendo el tipo de datos de cada columna y la cantidad de valores no nulos.

**- df.describe()**: Genera estadísticas descriptivas que resumen la tendencia central, dispersión y forma de la distribución de un conjunto de datos.

**- df.shape**: Devuelve una tupla que representa la dimensionalidad del *DataFrame*.

**- df.columns**: Lista las columnas del *DataFrame*.

**- df.index**: Devuelve el índice (etiquetas) del *DataFrame*.

**3. Selección y manipulación de datos**

**- df['col']:** Selecciona una columna.

**- df[['col1', 'col2']]:** Selecciona múltiples columnas.

**- df.iloc[]:** Selección por posiciones enteras (basada en índices).

**- df.loc[]:** Selección por etiquetas (basada en nombres de filas/columnas).

**- df[df['col'] > value]:** Filtrado de datos.

**- df.drop():** Elimina filas o columnas.

**- df.rename():** Renombra las etiquetas de filas o columnas.

**- df.sort\_values():** Ordena los datos por el valor de una o más columnas.

**4. Limpieza de datos**

**- df.isnull():** Identifica valores nulos.

**- df.dropna():** Elimina filas o columnas con valores nulos.

**- df.fillna():** Rellena valores nulos con un valor específico.

**- df.replace():** Reemplaza valores en el *DataFrame*.

**- df.duplicated():** Detecta filas duplicadas.

**- df.drop\_duplicates():** Elimina filas duplicadas.

**5. Transformación de datos**

**- df.apply():** Aplica una función a lo largo de un eje del *DataFrame*.

**- df.applymap():** Aplica una función a cada elemento del *DataFrame*.

**- df.map():** Aplica una función a cada elemento de una Serie.

**- df.groupby():** Agrupa el *DataFrame* usando un criterio específico.

**- df.pivot\_table():** Crea una tabla dinámica.

**- df.melt():** Convierte un *DataFrame* de formato ancho a formato largo.

**- pd.concat():** Concatenación de *DataFrames*.

**- pd.merge():** Fusiona dos *DataFrames* en uno solo, similar a las uniones SQL.

**6. Visualización de datos**

**- df.plot():** Genera varios tipos de gráficos usando *Matplotlib*.

**- df.hist():** Crea un histograma para cada columna numérica.

**7. Preparación de datos para Machine Learning**

**- df.sample():** Muestra una fracción aleatoria del *DataFrame*.

**- pd.get\_dummies():** Convierte variables categóricas en variables dummy/indicadoras.

**- df.corr():** Calcula la correlación de columnas entre sí.

**- df.to\_numpy():** Convierte el *DataFrame* en una matriz *NumPy*, útil para muchas bibliotecas de machine learning.

**- df.to\_csv():** Guarda el *DataFrame* como un archivo CSV.

import pandas as pd

# Cargar datos

df = pd.read\_csv('datos.csv')

# Exploración inicial

print(df.head())

print(df.info())

print(df.describe())

# Limpieza de datos

df.dropna(inplace=True)

df.drop\_duplicates(inplace=True)

# Transformación de datos

df['nueva\_columna'] = df['columna\_existente'].apply(lambda x: x \* 2)

# Preparación para machine learning

X = df.drop('target', axis=1)

y = df['target']

X = pd.get\_dummies(X)

Este conjunto de funciones y métodos de Pandas cubre gran parte del flujo de trabajo en ciencia de datos, desde la importación y limpieza de datos hasta su preparación para modelado y *machine learning*.