



Profundiza más

Recurso de Profundización

Ejemplo completo de *data handling* usando el *dataset iris*

Utilizando VSCODE, con Python instalado cree un archivo de extensión .py o .ipynb y escriba el siguiente código para poder hacer un ejercicio de *data handling*.

```
# Importamos librerías
import Pandas as pd
from sklearn.datasets import load_iris

# Cargar el dataset de iris
iris = load_iris()

# Convertir a un DataFrame de Pandas
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)

# Agregar la columna de etiquetas (target)
df['target'] = iris.target

# Mostrar las primeras filas
print(df.head())

# Información general del DataFrame
print(df.info())

# Estadísticas descriptivas
print(df.describe())

# Valores únicos de la variable target
print(df['target'].unique())

# Renombrar columnas
df = df.rename(columns={
'sepal length (cm)': 'sepal_length',
'sepal width (cm)': 'sepal_width',
'petal length (cm)': 'petal_length',
'petal width (cm)': 'petal_width'
})
print(df.head())

# Verificar valores faltantes
```



Profundiza más

```
print(df.isnull().sum())

# Filtrar datos
df_filtrado = df[df['sepal_length'] > 5]
print(df_filtrado.head())

# Crear una nueva columna
df['petal_area'] = df['petal_length'] * df['petal_width']
print(df.head())

# Agrupar datos
promedios = df.groupby('target')['sepal_length'].mean()
print(promedios)

# Visualizar datos
import matplotlib.pyplot as plt

# Gráfico de barras
promedios.plot(kind='bar', color=['blue', 'green', 'red'])
plt.title('Promedio de Longitudes de Sépalo por Clase')
plt.xlabel('Clase (target)')
plt.ylabel('Longitud de Sépalo')
plt.xticks(ticks=[0, 1, 2], labels=iris.target_names)
plt.show()

# Gráfico de dispersión
plt.scatter(df['sepal_length'], df['sepal_width'], c=df['target'], cmap='viridis')
plt.title('Relación entre Largo y Ancho del Sépalo')
plt.xlabel('Largo del Sépalo')
plt.ylabel('Ancho del Sépalo')
plt.colorbar(label='Clase (target)')
plt.show()
```



Profundiza más

Resultados:

```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  target
0                5.1                3.5                1.4                0.2         0
1                4.9                3.0                1.4                0.2         0
2                4.7                3.2                1.3                0.2         0
3                4.6                3.1                1.5                0.2         0
4                5.0                3.6                1.4                0.2         0

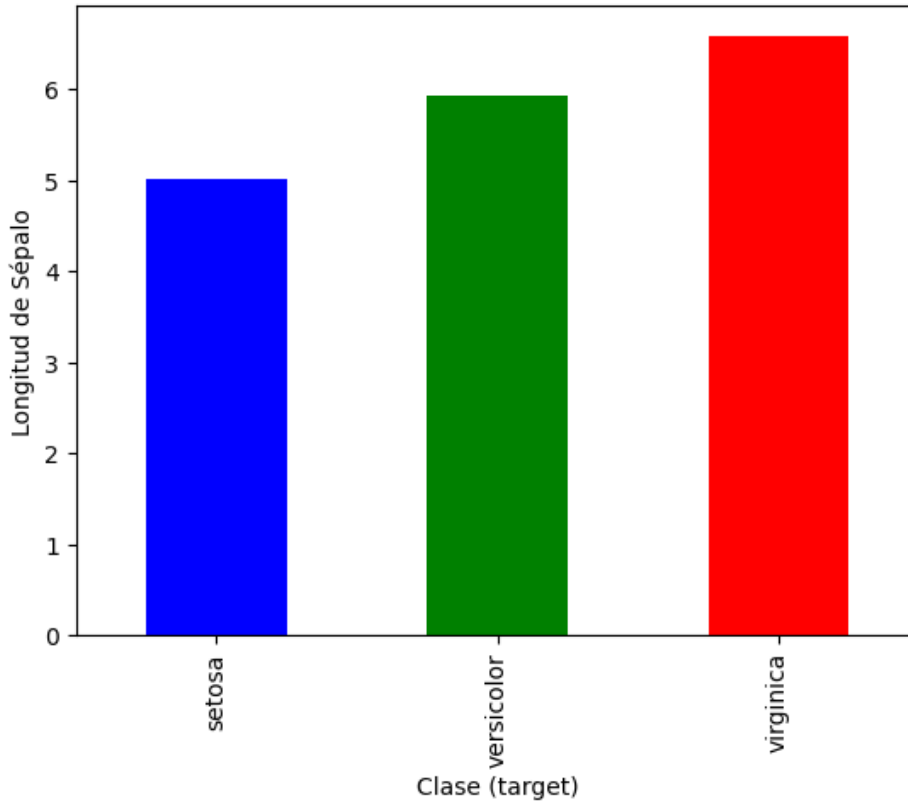
   target
0         0
1         0
2         0
3         0
4         0
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   sepal length (cm)     150 non-null   float64
1   sepal width (cm)      150 non-null   float64
2   petal length (cm)     150 non-null   float64
3   petal width (cm)      150 non-null   float64
4   target                150 non-null   int64
dtypes: float64(4), int64(1)
memory usage: 6.0 KB
...
0    5.006
1    5.936
2    6.588
Name: sepal_length, dtype: float64
```



Profundiza más

Promedio de Longitudes de Sépalo por Clase



Relación entre Largo y Ancho del Sépalo

