

# CHEATSHEET PANDAS – CALIDAD DE DATOS – CORRECCION

## CAMBIAR TIPOS DE DATOS

- **Método general:** `df.Var1.astype('category')`
- **Tipos más frecuentes a convertir con `astype`:** int, float, category, object, datetime
- **Convertir a tipo numérico:** `pd.to_numeric(df['Var1'])`
- **Convertir a tipo fecha:** `pd.to_datetime(df['Var1'])`
- **Link a los códigos de formato para las fechas:**  
<https://docs.python.org/3/library/datetime.html#strptime-and-strftime-behavior>
- **Cambiar el formato de una fecha:** `df['Fecha'].dt.strftime('%d/%m/%Y')`
- **Convertir a categórica ordenada:**
  1. Definir el orden: `orden_Status = pd.CategoricalDtype(['valor1', 'valor2'])`
  2. Aplicar el orden: `df[Var1] = df[Var1].astype(orden)`
- **Cambiar el orden de una categórica ordenada:**  
`df.Var1.cat.set_categories(['valor2', 'valor1'], ordered = True, inplace = True)`

## OPERACIONES DE CORRECCION

- **Eliminar variables:** `df.drop(columns = [Var1, 'Var2'])`
- **Eliminar registros:** `df.drop(index = df.index[0:5])`
- **Insertar variables:** `df.insert(2, 'Var3', df['Var1'] - df['Var2'])`
- **Eliminar variables con algún nulo:** `df.dropna(axis = 1)`
- **Eliminar registros con algún nulo:** `df.dropna(axis = 0)`
- **Reemplazar nulos:** `df.fillna()`
- **Renombrar variables:** `df.rename(columns = {'Nombre viejo': 'Nombre nuevo'})`
- **Reemplazar valores:** `df.Var1.replace('Valor viejo', 'Valor nuevo')`
- **Recodificar valores:** `df.Var1.map(diccionario valores viejos y nuevos)`

## OPERACIONES DE TEXTO (CON ACCESSOR STR)

- **A mayúsculas:** `texto.str.upper()`
- **A minúsculas:** `texto.str.lower()`
- **A estilo frase:** `texto.str.capitalize()`
- **A estilo titular:** `texto.str.title()`
- **Concatenar texto:** `texto.str.cat()`
- **Concatenar texto:** `texto.str.split()`
- **Longitud del texto:** `texto.str.len()`
- **Reemplazar valor:** `df.columns.str.replace(pat = ' ', repl = '_')`
- **Empieza por:** `df.columns.str.startswith(' texto')`
- **Termina por:** `df.columns.str.endswith(' texto')`
- **Contiene:** `df.columns.str.contains(' texto ')`