

# ▶ TALENTO TECH



Universidad  
Tecnológica  
de Bolívar

## Pandas 2.1

Ejecutor técnico: Jorge Luis Villalba Acevedo

[www.utb.edu.co/talento-tech](http://www.utb.edu.co/talento-tech)

# Categorical Data

# Resumen General

Los datos categóricos son variables que toman un número limitado de valores posibles, como géneros, categorías de productos, colores, etc. pandas proporciona herramientas poderosas para representar y manejar eficientemente estos tipos de datos.

En pandas, puedes usar el tipo `Categorical` para representar datos categóricos, lo que mejora el rendimiento y la claridad del código.

# 1. Crear Datos Categóricos

**Funciones clave:**

Función	Descripción	Ejemplo de uso
<code>pd.Categorical()</code>	Convierte una Serie o lista en un tipo de dato categórico	<code>pd.Categorical(['apple', 'banana', 'apple'])</code>
<code>astype('category')</code>	Convierte una columna en categórica	<code>df['col'].astype('category')</code>

# Ejemplo

```
import pandas as pd

categories = pd.Categorical(['apple', 'banana',
                             'apple', 'orange'])

# Convertir una columna de DataFrame en categórica
df = pd.DataFrame({
    'fruit': ['apple', 'banana', 'apple', 'orange',
             'banana']
})

df['fruit_cat'] = df['fruit'].astype('category')
```

## 2. Ordenar Categorías

**Funciones clave:**

Función	Descripción	Ejemplo de uso
<code>CategoricalDtype()</code>	Define un tipo categórico ordenado	<code>pd.Categorical(df['col'], categories=..., ordered=True)</code>

# Ejemplo

```
# Crear categorías ordenadas
categories_ordered = pd.Categorical(['low',
'medium', 'high', 'medium'], categories=['low',
'medium', 'high'], ordered=True)
```

### 3. Codificación de Datos Categóricos

Funciones clave:

Función	Descripción	Ejemplo de uso
<code>cat.codes</code>	Devuelve los códigos numéricos de las categorías	<code>df['fruit_cat'].cat.codes</code>



# Ejemplo

```
# Obtener los códigos numéricos de las  
categorías  
df['fruit_codes'] = df['fruit_cat'].cat.codes
```

## 4. Operaciones con Datos Categóricos

Funciones clave:

Función	Descripción	Ejemplo de uso
<code>cat.add_categories()</code>	Agrega nuevas categorías	<code>df['fruit_cat'].cat.add_categories(['grape'])</code>
<code>cat.remove_categories()</code>	Elimina categorías	<code>df['fruit_cat'].cat.remove_categories(['orange'])</code>
<code>cat.rename_categories()</code>	Renombra categorías	<code>df['fruit_cat'].cat.rename_categories({'apple': 'APPLE'})</code>
<code>cat.set_categories()</code>	Establece un nuevo conjunto de categorías	<code>df['fruit_cat'].cat.set_categories(['banana', 'apple', 'orange'])</code>

# Ejemplo

```
# Renombrar una categoría  
df['fruit_cat'] =  
df['fruit_cat'].cat.rename_categories({'apple'  
: 'APPLE'})
```

## 5. Verificar si una Columna es Categórica

**Funciones clave:**

Función	Descripción	Ejemplo de uso
<code>df.dtypes</code>	Devuelve los tipos de datos de las columnas del DataFrame	<code>df.dtypes</code>
<code>df['col'].dtype == 'category'</code>	Verifica si una columna es de tipo categórico	<code>df['fruit_cat'].dtype == 'category'</code>

# Ejemplo

```
# Verificar si una columna es categórica  
df['fruit_cat'].dtype == 'category'
```

## 6. Filtrar y Agrupar por Categorías

**Funciones clave:**

Función	Descripción	Ejemplo de uso
<code>groupby()</code>	Agrupar datos por categorías	<code>df.groupby('fruit_cat').sum()</code>

# Ejemplo

```
# Agrupar por categorías  
grouped = df.groupby('fruit_cat').size()
```

# Aplicación

Retomando la problemática **Desafíos en la Educación en Colombia** aplica la función `groupby` para dar respuesta a los siguientes interrogantes.

1. ¿Cuál es la calificación promedio de los estudiantes por género?
2. ¿Cuál es el porcentaje de asistencia promedio por grado?
3. ¿Cuál es la calificación promedio por región?



# Aplicación

Retomando la problemática **Desafíos en la Educación en Colombia** aplica la función `groupby` para dar respuesta a los siguientes interrogantes.

4. ¿Cuál es la edad promedio de los estudiantes por género y por grado?
5. ¿Cuál es la asistencia promedio por región y por grado?
6. ¿Cuántos estudiantes hay en cada región?
7. ¿Cuál es la calificación promedio por región y por género?

# Aplicación

Retomando la problemática **Desafíos en la Educación en Colombia** aplica la función `groupby` para dar respuesta a los siguientes interrogantes.

8. ¿Cuál es la edad promedio de los estudiantes por región?
9. ¿Cuál es la asistencia promedio de los estudiantes por género?
10. ¿Cuántos estudiantes hay en cada combinación de grado y región?

# Referencias

1. **McKinney, W. (2017).** *Python for data analysis: Data wrangling with pandas, numpy, and ipython* (2nd ed.). O'Reilly Media.
2. **Sweigart, A. (2019).** *Automate the boring stuff with Python: Practical programming for total beginners* (2nd ed.).
3. **González Duque, R. (2015).** *Python para todos*. Independently published.
4. **Bosch, J. (2020).** *Introducción a la programación en Python: Algoritmos y lógica para principiantes*. Independently published.