

WORKING WITH CATE. DATA

Datos categóricos $\begin{cases} \text{Ordinales (Tienen un orden)} \\ \text{Nominales} \end{cases}$

Para envolver un tipo objeto a Categorical \rightarrow `df[col] = df[col].astype('category')`

¿Porque guardar la columna como categorías? \rightarrow AHORRA MEMORIA

Puedes ver los bytes que ocupa una columna \rightarrow `df[col].nbytes`

Para configurar una columna categórica $\begin{cases} \text{Nuevas categorías} \rightarrow \text{df['col'] = df['col'].cat.set_categories(new_categories=['...'])} \\ \text{Para pasar que sea en el orden} \rightarrow \text{ordered=True} \\ \text{Para añadir categorías} \rightarrow \text{df['col'] = df['col'].cat.add_categories(new_categories=['...'])} \\ \text{Para borrar categorías} \rightarrow \text{df['col'] = df['col'].cat.remove_categories(remove=['...'])} \\ \text{Para renombrar categorías} \rightarrow \text{df['col'] = df['col'].cat.rename_categories(['Nombre viejo', 'Nombre nuevo', ...])} \\ \text{Para cambiar las categorías en 3} \rightarrow \text{df['col'] = df['col'].replace(['X1', 'X2', 'X3'], ['Y1', 'Y2', 'Y3'])} \end{cases}$

Después hay que \rightarrow renombrar a categoría

Para reordenar una categoría \rightarrow `df['col'].cat.reorder_categories(new_categories=['...', '...', ordered=True, replace=True)`

Para convertir categorías a labels numéricos \rightarrow `df['col'].cat.codes`

\rightarrow Para obtener el dict de asignación código: categoría \rightarrow

`cod = df['col'].cat.codes
cat = df['col'].cat.categories
mapa = dict(zip(cod, cat))`

Para ver el código booleano \rightarrow `df['cat.col'] = np.where(condicion, 1, 0)`

`used_cols = pd.get_dummies(df, columns=['col'], prefix='...')`

One-hot encoding \rightarrow