



TEAM_CHALLANGE:TOOLBOX

INTEGRANTES: CÉSAR SÁNCHEZ, NICHOLAS BROWN, NATALIA Y SOFÍA GÓMEZ

FUNCIONES



DESCRIBE_DF

- Resume las características de cada variable del dataset: tipo, porcentaje de nulos, número de valores únicos y cardinalidad.
- Nos ayuda a entender qué tipo de datos tenemos antes de aplicar modelos.

```
describe_df(df_titanic)
```

	DATA_TYPE	MISSINGS (%)	UNIQUE_VALUES	CARDIN (%)
fare	float64	0.00	248	27.83
age	float64	19.87	88	9.88
sibsp	int64	0.00	7	0.79
parch	int64	0.00	7	0.79
deck	object	77.22	7	0.79
embarked	object	0.22	3	0.34
pclass	int64	0.00	3	0.34
embark_town	object	0.22	3	0.34
class	object	0.00	3	0.34
who	object	0.00	3	0.34
sex	object	0.00	2	0.22
survived	int64	0.00	2	0.22
adult_male	bool	0.00	2	0.22
alive	object	0.00	2	0.22
alone	bool	0.00	2	0.22

TIPIFICA_VARIABLES

- Clasifica automáticamente cada variable como binaria, categórica, numérica continua o discreta, según su cardinalidad.
- Es útil para saber qué tratamiento darle a cada variable en el modelo.

tipifica_variables(df_titanic)		
	nombre_variable	tipo_sugerido
0	survived	Binaria
1	pclass	Numerica Discreta
2	sex	Binaria
3	age	Numerica Continua
4	sibsp	Numerica Discreta
5	parch	Numerica Discreta
6	fare	Numerica Continua
7	embarked	Categórica
8	class	Categórica
9	who	Categórica
10	adult_male	Binaria
11	deck	Categórica
12	embark_town	Categórica
13	alive	Binaria
14	alone	Binaria

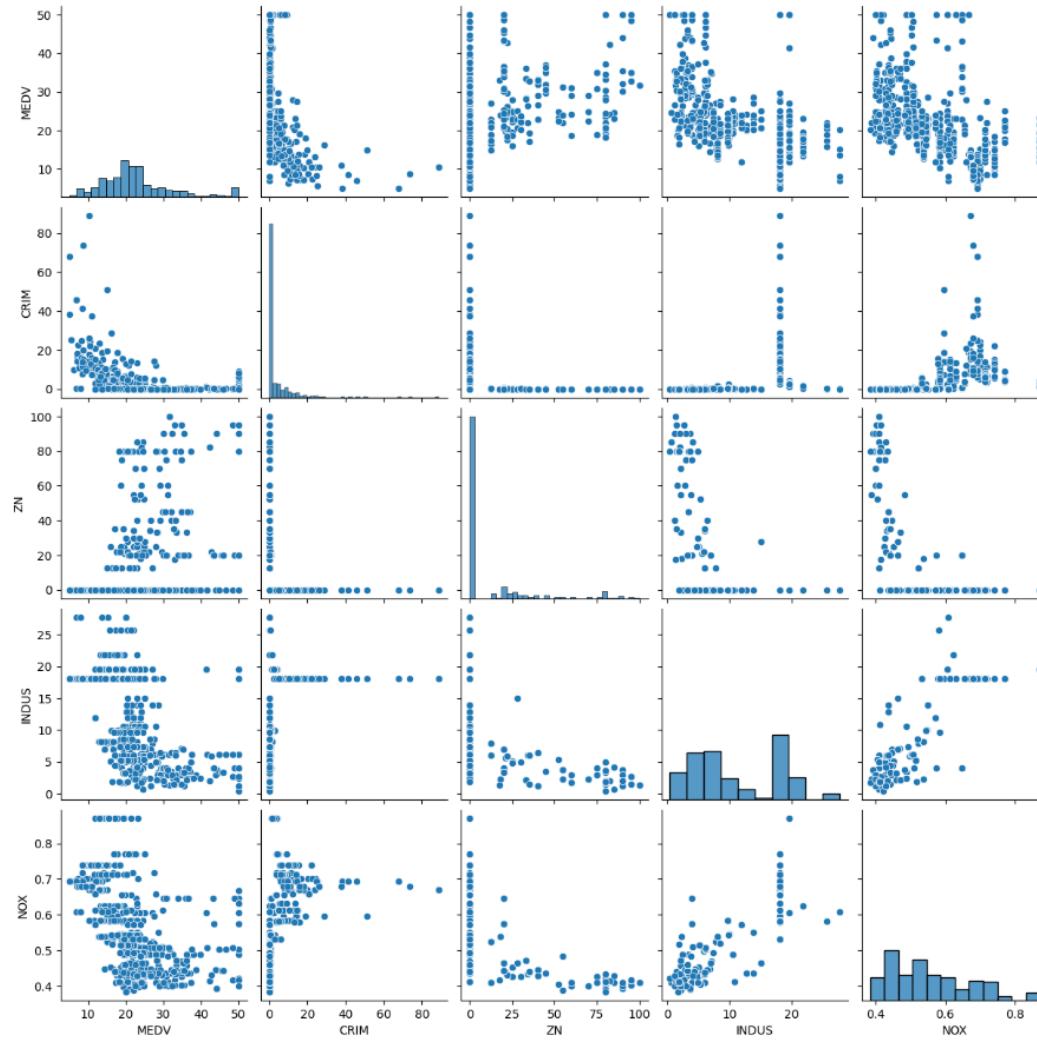
GET_FEATURES_NUM_REGRESSION

- Selecciona las variables numéricas que tienen alta correlación y significancia estadística con el target.
- Nos ayuda a elegir buenas variables predictoras para regresión.

```
get_features_num_regression(df_boston, target_col= "MEDV" , umbral_corr= 0.4)  
['INDUS', 'NOX', 'RM', 'TAX', 'PTRATIO', 'LSTAT']
```

PLOT_FEATURES_NUM_REGRESSION

- Genera gráficos pairplot para visualizar la relación entre el target y las variables numéricas seleccionadas.
- Sirve para validar visualmente si hay patrones útiles.



GET_FEATURES_CAT_REGRESSION

- Selecciona las variables categóricas que tienen una relación estadísticamente significativa con el target.
- Nos permite incluir variables categóricas relevantes en el modelo.

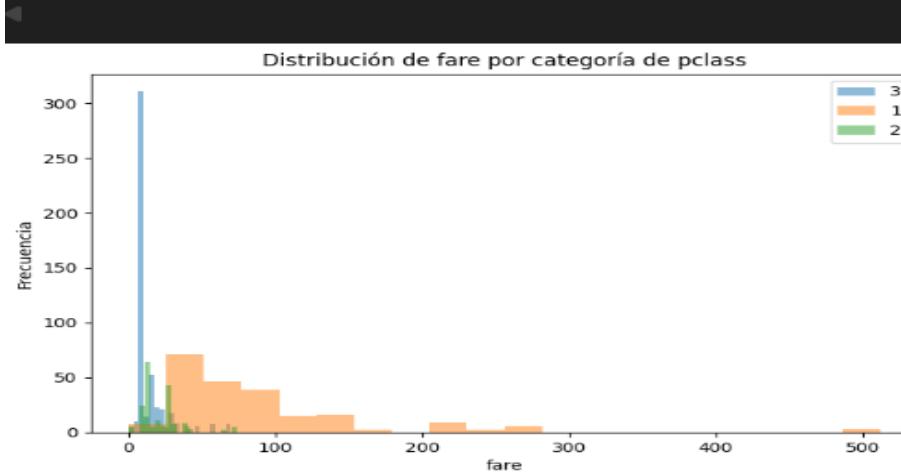
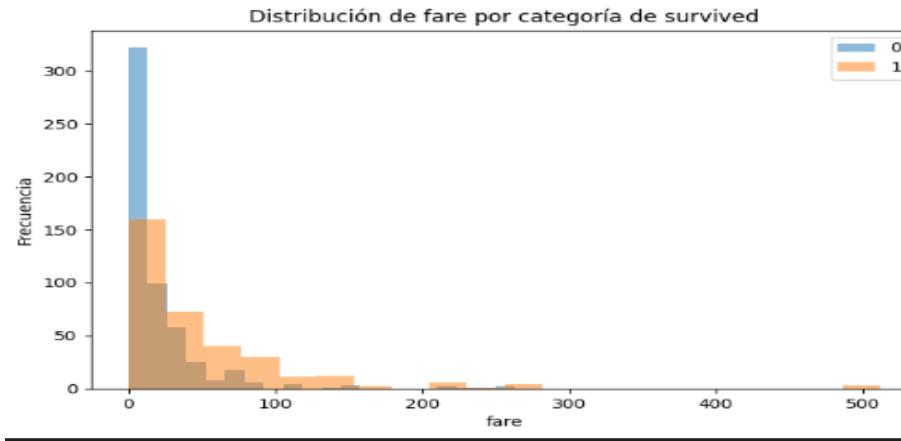
```
cols_significativas = get_features_cat_regression(  
    df=df_titanic,  
    target_col="fare",  
    pvalue=0.05,  
    with_individual_plot=False  
)
```

```
Columnas significativas: ['sex', 'embarked', 'class', 'who', 'embark_town', 'alive']
```

PLOT_FEATURES_CAT_REGRESSION

- Crea gráficos que muestran cómo cambia el valor del target según cada categoría. Ayuda a identificar si las variables categóricas generan diferencias relevantes y pueden aportar información al modelo.

```
plot_features_cat_regression(  
    df=df_titanic,  
    target_col="fare",  
    columns=[],           # lista vacía → analiza NUMÉRICAS  
    pvalue=0.05,  
    with_individual_plot=True  
)
```





GRACIAS