

# CHEATSHEET PANDAS – CALIDAD DE DATOS – DIAGNOSTICO

## VISION GLOBAL

- **Información general:** `df.info()`
- **Tamaño en memoria:** `df.info(memory_usage = 'deep')`
- **Dimensión del dataset:** `pd.shape`
- **Nombres de las variables:** `df.columns`
- **Tipos de las variables:** `df.dtypes`
- **Estadísticos:** `df.describe().T`

## NULOS, DUPLICADOS Y VALORES UNICOS

- **Conteo de nulos por variable:** `df.isna().sum().sort_values(ascending = False)`
- **Porcentaje de nulos por variable:**  
`df.isna().mean().sort_values(ascending = False) * 100`
- **Conteo de duplicados:** `df.duplicated().sum()`
- **Localizar los duplicados:** `df[df.duplicated()]`
- **Número de valores únicos:** `df.nunique()`
- **Valores únicos diferentes:** `df.Var1.unique()`

## ESTADISTICOS BASICOS

- **Conteo de frecuencias:** `df.Var1.value_counts()`
- **Conteo de frecuencias (en porcentaje):**  
`df.Var1.value_counts(normalize=True)*100`
- **Moda:** `df.Var1.mode()`
- **Media:** `df.Var1.mean()`
- **Mediana:** `df.Var1.median()`
- **Máximo:** `df.Var1.max()`
- **Mínimo:** `df.Var1.min()`
- **Índice del máximo:** `df.Var1.idxmax()`
- **Índice del mínimo:** `df.Var1.idxmin()`
- **Correlación variables continuas:** `df.Var1.corr(Var2)`
- **Correlación variables ordinales:** `df.Var1.corr(Var2, method = 'kendall')`