

Vas conociendo las esencia del uso de Python y su programación.

Ahora quiero que practiques lo más importante que has visto en el apartado de manipular data frames.

Para ello te he preparado una serie de ejercicios :)

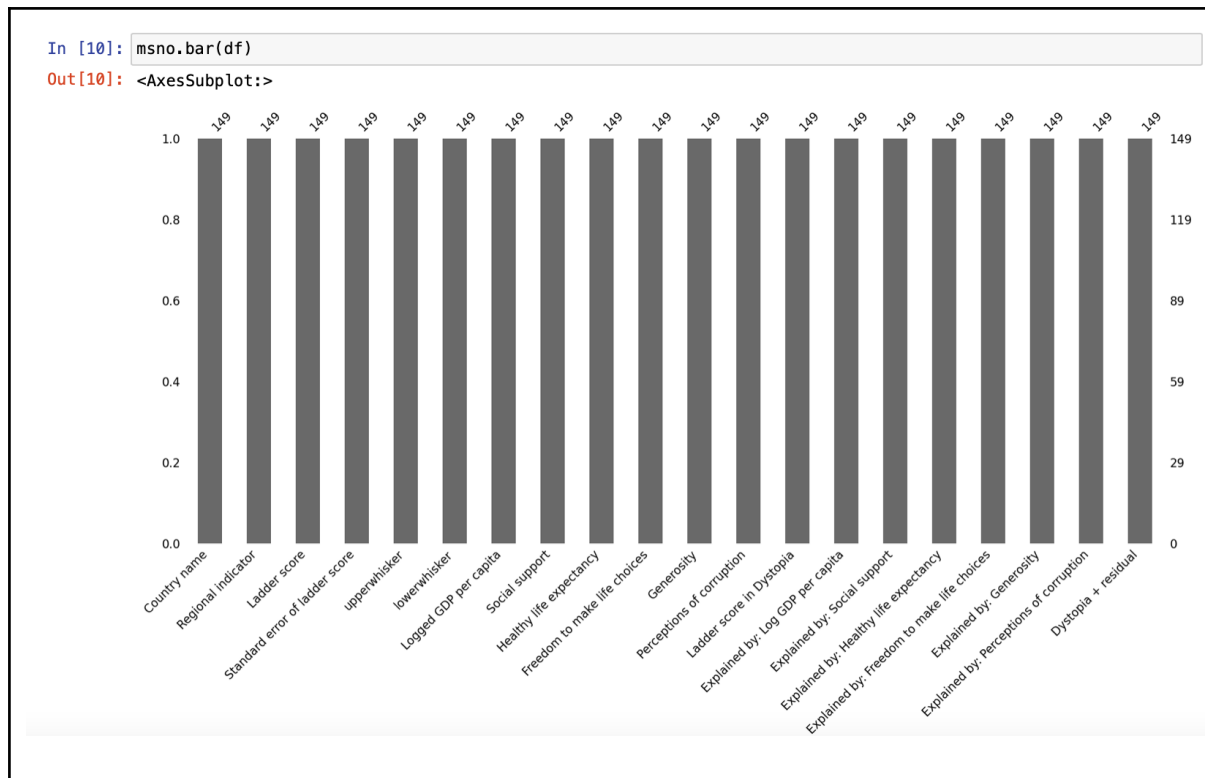
¡ A por ello!

1. Lee y limpia los datos

- Lee los siguientes datos con read.csv:
<https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021>
- Corrige el header si hace falta
- Limpia los datos para tener las variables con el tipo correcto
- Modifica las variables con el tipo correcto
- Realiza un gráfico con la estructura de nans del data frame

```
In [8]: df.info()

0    Country name                149 non-null    object
1    Regional indicator          149 non-null    object
2    Ladder score                149 non-null    float64
3    Standard error of ladder score 149 non-null    float64
4    upperwhisker                149 non-null    float64
5    lowerwhisker                149 non-null    float64
6    Logged GDP per capita        149 non-null    float64
7    Social support               149 non-null    float64
8    Healthy life expectancy      149 non-null    float64
9    Freedom to make life choices 149 non-null    float64
10   Generosity                  149 non-null    float64
11   Perceptions of corruption    149 non-null    float64
12   Ladder score in Dystopia      149 non-null    float64
13   Explained by: Log GDP per capita 149 non-null    float64
14   Explained by: Social support   149 non-null    float64
15   Explained by: Healthy life expectancy 149 non-null    float64
16   Explained by: Freedom to make life choices 149 non-null    float64
17   Explained by: Generosity       149 non-null    float64
18   Explained by: Perceptions of corruption 149 non-null    float64
19   Dystopia + residual           149 non-null    float64
dtypes: float64(18), object(2)
```



2. Manipula los data frames:

- Cuenta el número de países por Regional indicator.
- Calcula el promedio de Logged GDP y el promedio de Healthy life expectancy por Regional indicator.
- Calcula el promedio de todas las variables numéricas por Regional indicator
- De Western Europe lista los 10 países con más Perceptions of corruption
- Lista los países con mayor Freedom to make life choices por cada Regional Indicator
- Para los países con Logged GDP per capita menor que 10 calcula el promedio de Ladder score por Regional Indicator
- Crea una nueva variable factor llamada LADDER_6 que nos indique valores de Ladder score mayores que 6 con un SI y con un NO menores que 6
- ¿Cuántos países tenemos en cada grupo de LADDER_6?
- Cuántos países de cada región tenemos en LADDER_6.
- Calcula el promedio de las variables numéricas teniendo en cuenta Regional Indicator y LADDER_6

Pinta aquí la última tabla de promedio para la variable Healthy life expectancy

```
LADDER_6 Regional indicator
NO      Central and Eastern Europe      68.152714
        Commonwealth of Independent States 64.965900
        East Asia                      71.582600
        Latin America and Caribbean    65.931000
        Middle East and North Africa    64.293417
        South Asia                     62.681000
        Southeast Asia                  63.380375
        Sub-Saharan Africa              55.577486
        Western Europe                  73.032667
SI      Central and Eastern Europe      68.468400
        Commonwealth of Independent States 65.227500
        East Asia                      69.600000
        Latin America and Caribbean    68.221100
        Middle East and North Africa    68.766800
        North America and ANZ          72.325000
        Southeast Asia                  76.953000
        Sub-Saharan Africa              66.701000
        Western Europe                  73.033167
Name: Healthy life expectancy, dtype: float64
```

¡Genial!

Celébralo como más te guste :)

Tómate una cerveza, un zumo, un té o lo que te guste. Son tus primeros pasos con Python.

1 abrazo y hasta pronto.
Brindo por tus éxitos :)

Jordi Ollé