

explore

June 26, 2023

Analyse exploratoire

1 Intro

```
[1]: # Import des bibliothèques
import os
os.chdir("..")
import pandas as pd
import numpy as np
import re
from app.utils.helper import parse_name, give_short_name

# Import des données
from app.data.cleaning import df, athlet, games, region, result
```

2 Athlètes

```
[2]: n = 20
(athlet
 .join(result.groupby("athlete").sport.first())
 .join(result.groupby("athlete").medal.count())
 .rename(columns={"medal": "n_medals"})
 .sort_values(by="n_medals", ascending=False))[:n]
```

```
[2]:
```

	first_name	last_name	gender	birth_year	latteest_noc	\
id						
207089493835	Michael	PHELPS	M	1985	USA	
027503151640	Larysa	DIRIY	F	1935	URS	
779052604582	Nikolay	ANDRIANOV	M	1953	URS	
206027787982	Edoardo	MANGIAROTTI	M	1919	ITA	
354705440133	Gustaf	CARLBERG	M	1880	SWE	
473622409020	Ole	BJRNDALLEN	M	1974	NOR	
027629514478	Borys	SHAKHLIN	M	1932	URS	
189908206539	Aleksey	NEMOV	M	1976	RUS	
747407348886	Natalie	HALL	F	1983	USA	
632866972353	Dara	MINAS	F	1967	USA	

692484174312	Birgit	SCHMIDT	F	1962	GER
024676404810	Ryan	LOCHTE	M	1984	USA
470290582797	Paavo	NURMI	M	1898	FIN
429998356546	Sawao	KATO	M	1947	JPN
123462359385	Jennifer	CUMPELIK	F	1973	USA
151171364282	Takashi	ONO	M	1931	JPN
897714493094	Mark	SPITZ	M	1950	USA
948616223040	Matthew	BIONDI	M	1966	USA
317143000542	Vra	ODLOILOV	F	1942	TCH
858118795892	Carl	OSBURN	M	1885	USA

	sport	n_medals
id		
207089493835	Swimming	28
027503151640	Gymnastics	18
779052604582	Gymnastics	15
206027787982	Fencing	13
354705440133	Shooting	13
473622409020	Biathlon	13
027629514478	Gymnastics	13
189908206539	Gymnastics	12
747407348886	Swimming	12
632866972353	Swimming	12
692484174312	Canoeing	12
024676404810	Swimming	12
470290582797	Athletics	12
429998356546	Gymnastics	12
123462359385	Swimming	12
151171364282	Gymnastics	12
897714493094	Swimming	11
948616223040	Swimming	11
317143000542	Gymnastics	11
858118795892	Shooting	11