



Màster Universitari en Ciència de Dades
PAC 4

Autor: Guillem de la Calle Vicente

Professor: **Yago Diez Donoso**

Assignatura: 22.403 - Programació per a la ciència de dades

13 de juny de 2022

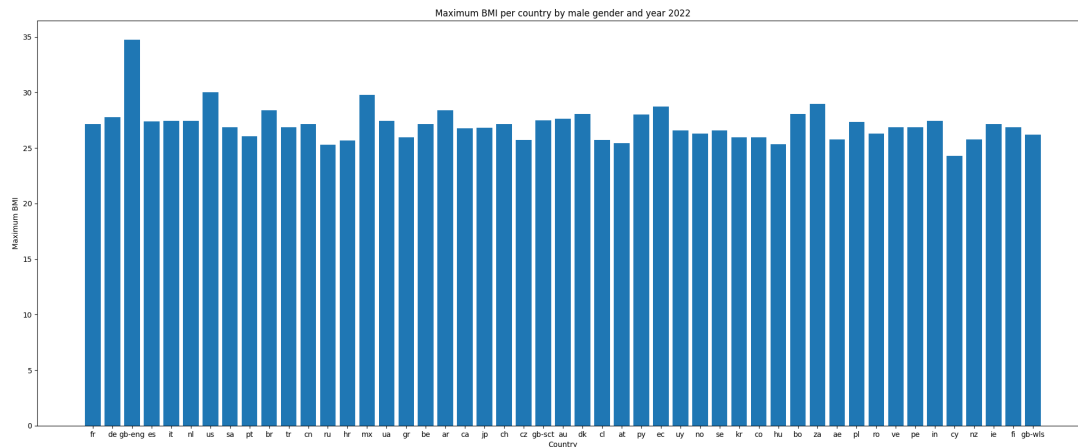
Índex

1	Exercici 3	2
2	Exercici 6	4

1 Exercici 3

- a) Comentem aquest apartat per tal de respondre a la pregunta que ens fan. Primer de tot voldríem comentar que al fitxer .csv apareixen alguna fila buida (respecte al camp *club_flag_url*) i per això els hem tret mitjançant `dropna()`.

Hem considerat que una bona manera de representar aquesta informació era amb un diagrama de barres. El mostrem a continuació:



Es pot veure amb més detall executant el main. Cal tenir en compte que hem representat el valor màxim del BMI per cada país, pel que és normal que ens apareixin les barres amb valors tan alts: ens surten classificats a la categoria *overweight*, que sembla estrany per jugadors professionals (tot i que sempre hi pot haver algun jugador amb un BMI més alt del normal). Però de ben segur que si es fes la mitjana del BMI, en comptes del màxim, ens sortirien els resultats esperats.

- b) Per fer aquest apartats, considerem els jugadors que van jugar a Espanya l'any 2020 (ja que les dades de l'INE són d'aquest any) juntament amb els homes d'edat compresa entre 18 i 34 anys (rang d'edat dels futbolistes). També farem un diagrama de barres per tal d'observar-ho, però ara ho farem diferent que abans per tal de basar-nos en les dades de l'INE. Calcularem la proporció de jugadors que estan a cadascuna de les categories de BMI i ho denotarem com (u, n, o, ob) i al costat ficarem d'homes (M). Per exemple, nM voldrà dir *normal weight men*.

Les dades extretes de l'INE són:

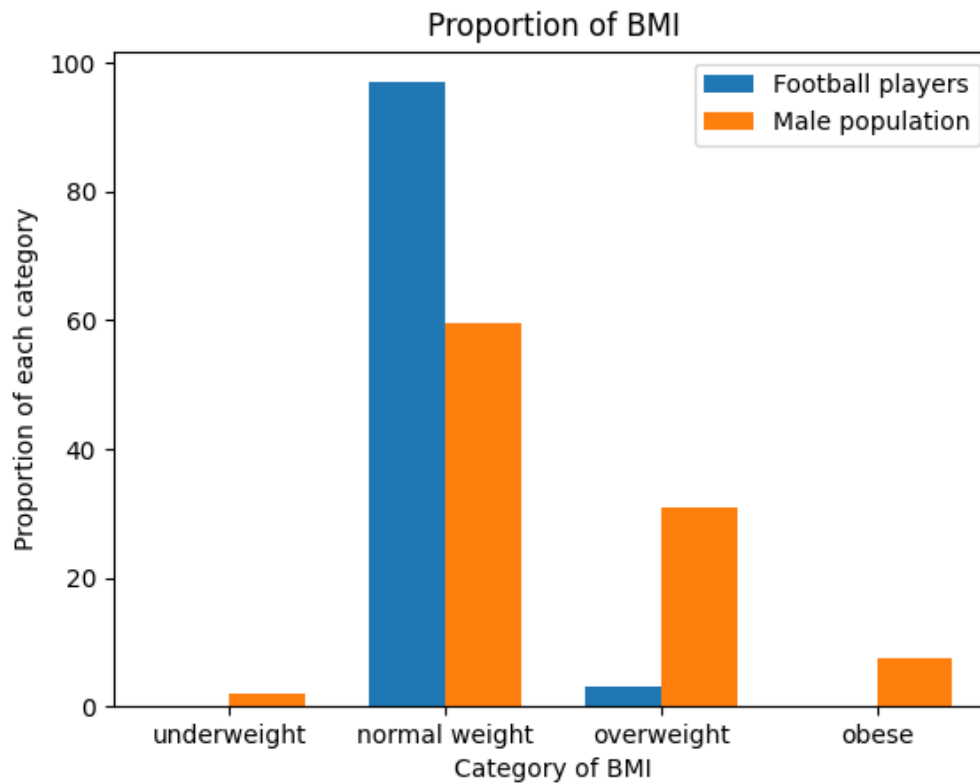
	Underweight (BMI < 18.5 kg/m ²)	Normal weight (18.5 kg/m ² <= BMI < 25 kg/m ²)	Overweight (25 kg/m ² <= BMI < 30 kg/m ²)	Obesity (BMI >= 30 kg/m ²)
Men				
From 18 to 24 years old	53.2%	1,123.8%	380.6%	74.6%
From 25 to 34 years old	31.9%	1,408.3%	929.5%	248.0%

on les unitats són milers de persones. Per tant el que farem és trobar la proporció d'homes a cada categoria.

- uM: 2%
- nM: 59,58%

- oM: 30,83%
- obM: 7,59%

Amb Python calcularem aquestes dades pels futbolistes que juguen a Espanya i ho representarem amb un diagrama de barres agrupat:



d'on veiem el que s'esperava i que hem comentat a l'apartat anterior: els futbolistes professionals estan gairebé tots a *normal weight* i hi ha algun outlier a *overweight*.

2 Exercici 6

El primer que fem per tal de trobar la millor defensa del món és aconseguir 10 jugadors/es que juguin a cadascuna de les quatre posicions (és a dir, en total tindrem 40 jugadors amb algun que apareix a dues posicions, potser). Tindrem 10 jugadors disponibles per ocupar la posició de RB, 10 jugadors disponibles per ocupar la posició de LB i 20 jugadors per ocupar les dues posicions de CB. Per fer-ho, ho farem observant la puntuació global dels jugadors a cadascuna d'aquestes posicions (overall). Després, ja ens centrarem en els criteris per cadascuna de les posicions per tal d'acabar de decidir la nostra línia defensiva tenint en compte els millors jugadors disponibles en aquella posició (que seran els 10 que agafarem en un principi per tal de filtrar i reduir el nostre conjunt de dades). Això ho realitzem a la funció *players_best_position* que ens retornarà un diccionari amb quatre claus (una per a cada posició) i els respectius *sofifa_id* dels jugadors a cada posició en una llista. Aprofitarem que el fitxer .csv ja està ordenat pels *overall* i per tant haurem d'agafar els 10 millors jugadors de cada posició i ja està.

Un cop tenim els 10 o els 20 millors jugadors de cada posició, el que fem és crear un diccionari per a cada posició que tingui com a clau un identificador (*identifier*) i com a valor una llista on s'inclouran els criteris (entre 3 i 5 com se'ns demana) que considerem més adients per tal que un jugador jugui a una posició determinada. Això ho farem a la funció *qualities_players_position*. Els criteris que hem pensat que defineixen millor cada línia defensiva són:

- CB:
 - a) Defensa: *mentality_interceptions*, *mentality_aggression*, *defending_standing_tackle*.
 - b) Control de pilota (possessió): *passing*.
 - c) Atac: *attacking_heading_accuracy*.
- LB:
 - a) Defensa: *defending_marking_awareness*, *mentality_aggression*, *mentality_interceptions*.
 - b) Control de pilota (possessió): *passing*.
 - c) Atac: *pace*.
- RB:
 - a) Defensa: *defending_marking_awareness*, *mentality_aggression*, *mentality_interceptions*.
 - b) Control de pilota (possessió): *passing*.
 - c) Atac: *pace*.

També podríem haver considerat els criteris més generals: *pace*, *shooting*, *passing*, *dribbling*, *defending* i *physic*; però com que només en podíem triar entre 3 i 5, hem preferit centrar-nos en els que pensem que són més importants per a cadascuna de les posicions.

En aquesta funció, retornarem un diccionari amb clau l'identificador i com a valors una llista que estarà formada per les cinc qualitats anteriors (depenent de la posició), per la suma de totes les qualitats, per la suma de les qualitats de defensa, per la suma de les qualitats de control de pilota i per la suma de les qualitats d'atac. És a dir, la llista estarà formada per $5 + 1 + 3 = 9$ elements.

Tot seguit, utilitzem la funció *product* del mòdul *itertools* per tal de calcular totes les combinacions per tal de formar la defensa de 4 jugadors i eliminem les repeticions agafant els elements amb 4 elements (els elements de la llista de les combinacions són diccionaris amb clau *short_name* i per tant ho podem fer). Llavors, creem una funció que ens ordena la llista tenint en compte el valor de la suma de totes les qualitats comentades anteriorment (és a dir, l'element de la posició 6 dels valors del diccionari) i mostra a la pantalla les 3 millors alineacions segons el nostre criteri (en podríem mostrar més, però hi ha tantes combinacions, que pensem que mostrant aquestes ja és suficient). Aquesta funció s'anomena *order_combinations*. Ho imprimim per pantalla amb la funció *print_combinations*.

Les 3 millors línies defensives per cada categoria són:

```
3 millors línies defensives masculina
ALINEACIÓ 1:
CB 1: V. van Dijk
CB 2: Sergio Ramos
LB: A. Robertson
RB: J. Kimmich
Contribució a la defensa: 1040
Contribució a la possessió: 314
Contribució a l'atac: 333

ALINEACIÓ 2:
CB 1: V. van Dijk
CB 2: Sergio Ramos
LB: A. Robertson
RB: T. Alexander-Arnold
Contribució a la defensa: 1018
Contribució a la possessió: 316
Contribució a l'atac: 342

ALINEACIÓ 3:
CB 1: V. van Dijk
CB 2: Sergio Ramos
LB: A. Robertson
RB: João Cancelo
Contribució a la defensa: 1014
Contribució a la possessió: 311
Contribució a l'atac: 348
```

```
ALINEACIÓ 1:
CB 1: W. Renard
CB 2: J. Ertz
LB: A. Majri
RB: L. Bronze
Contribució a la defensa: 1039
Contribució a la possessió: 297
Contribució a l'atac: 354

ALINEACIÓ 2:
CB 1: W. Renard
CB 2: J. Ertz
LB: A. Majri
RB: B. Sauerbrunn
Contribució a la defensa: 1047
Contribució a la possessió: 289
Contribució a l'atac: 341

ALINEACIÓ 3:
CB 1: W. Renard
CB 2: J. Ertz
LB: A. Majri
RB: C. Dunn
Contribució a la defensa: 1024
Contribució a la possessió: 307
Contribució a l'atac: 351
```

```
3 millors línies defensives de jugadors veterans
ALINEACIÓ 1:
CB 1: Sergio Ramos
CB 2: G. Chiellini
LB: Jordi Alba
RB: K. Walker
Contribució a la defensa: 1003
Contribució a la possessió: 293
Contribució a l'atac: 347

ALINEACIÓ 2:
CB 1: Sergio Ramos
CB 2: G. Chiellini
LB: Jordi Alba
RB: Jesús Navas
Contribució a la defensa: 999
Contribució a la possessió: 298
Contribució a l'atac: 342

ALINEACIÓ 3:
CB 1: Sergio Ramos
CB 2: G. Chiellini
LB: Jordi Alba
RB: K. Trippier
Contribució a la defensa: 994
Contribució a la possessió: 299
Contribució a l'atac: 328
```